UiA University of Agder

Master

3D Convolution Neural Networks for Medical Imaging; Classification and Segmentation

A Doctor's Third Eye

CHARUL GIRI

SUPERVISOR

Morten Goodwin

University of Agder, 2020
Faculty of Engineering and Science
Department of Information and Communication Technology

Acknowledgements

I wish to express my sincere gratitude to my supervisor, Associate Professor Morten Goodwin, without whose brilliant guidance and encouragement the goal of this thesis would not have been realized. I truly appreciate his suggestions and advice from, and the Tuesday meetings, even in the times of Corona pandemic from skype.

I would like to thank CAIR for allowing me to use their Nvidia servers to run my code, and for all other resources like the PhD candidate Jivitesh Sharma who also happens to be my husband and whose unconditional love and support kept me going on, and my parents who constantly kept asking when I am finishing this thesis, and other family and friends.

And a big thanks to all my colleagues from Anzyz Technologies for their faith and support during the writing of this thesis. A special thanks to Svein Olaf and Haakon Lagnes for being very understanding and competitive, keeping me in the race.

Finally to the efforts of the reviewers of this thesis, who have to go through this every year.

P.S. And to Turtle, my dog.

Abstract

In this thesis, we studied and developed 3D classification and segmentation models for medical imaging. The classification is done for Alzheimer's Disease and segmentation is for brain tumor sub-regions. For the medical imaging classification task we worked towards developing a novel deep architecture which can accomplish the the complex task of classifying Alzheimer's Disease volumetrically from the MRI scans without the need of any transfer learning. The experiments were performed for both binary classification of Alzheimer's Disease (AD) from Normal Cognitive (NC), as well as multi class classification between the three stages of Alzheimer's called NC, AD and Mild cognitive impairment (MCI). We tested our model on the ADNI dataset and achieved mean accuracy of 94.17% and 89.14% for binary classification and multiclass classification respectively.

In the second part of this thesis which is segmentation of tumors sub-regions in brain MRI images we studied some popular architecture for segmentation of medical imaging and inspired from them, proposed our architecture of end-to-end trainable fully convolutional neural network which uses attention block to learn the localization of different features of the multiple sub-regions of tumor. Also experiments were done to see the effect of weighted cross-entropy loss function and dice loss function on the performance of the model and the quality of the output segmented labels. The results of evaluation of our model are received through BraTS'19 dataset challenge. The model is able to achieve a dice score of 0.80 for the segmentation of whole tumor, and a dice scores of 0.639 and 0.536 for other two sub-regions within the tumor on validation data.

In this thesis we successfully applied computer vision techniques for medical imaging analysis. We show the huge potential and numerous benefits of deep learning to combat and detect diseases opens up more avenues for research and application for automating medical imaging analysis.

List of Figures

List of Tables

Contents

Introduction

Chapter 1

Introduction

The early detection and diagnosis of the various disease is made possible thanks to the development in the medical image acquisition devices and medical imaging technology, such as MRIs, CT scans, PET scans, etc. As a result of this growth huge volumes of complex and heterogeneous medical data is generated for which extensive and tedious efforts and a large force of human experts such as radiologists and physicians is required who can interpret this data for the diagnosis and treatment of the disease. With the increase in the workload, the diagnosis task becomes prone to human errors. Applying a deep learning-based medical image analysis to computer-aided diagnosis system can significantly decrease the efforts required for medical diagnosis and with more accurate results improving the efficiency and quality of the treatment.

The wave of technological advancement hit the medical industry in a very positive way. Fast and robust detection and segmentation of medical imaging can provide great support to pathology and can assist doctors and medical labs. AI is a very fast-growing field and the rise in popularity of computer vision algorithms like CNN has caught the attention of researchers to develop systems for the automation of medical image analysis.

An AI based medical imaging system tries to find the relevant patterns in the data to identify specific anatomical markers. AI is playing an important role in diagnosing neuroimaging data. In this thesis we have focused on two task of brain imaging diagnosis. In part one the goal is to develop a 3D CNN architecture which can efficiently classify Alzheimer's Disease using the brain MRIs and in part two, we have worked on deep end-to-end fully convolutional 3D CNN model to create segmented masks of different sub-regions of gliomas present in the brain.

1.0.1 Alzheimer's Disease

The challenging task of detection of Alzheimer's in early stages has been a focus of many studies as it contributes 60%-70% of the total dementia cases worldwide. In Alzheimer's Disease brain cells involved in cognitive functioning starts to irrevocably deteriorate at a

gradual pace. Detection of Alzheimer's in early stag is really difficult because the disease is thought to begin 20 years before the symptoms arrive, with very small and unnoticeable changes in the brain. Common symptoms of Alzheimer's include memory loss, reduced ability to reason and think. As the disease spread to further parts in brain it leads to loss of the ability to perform everyday activities.

1.0.2 Brain Tumor

A brain tumor is an abnormal growth of a mass of cells in parts of brain. A tumor can be benign or cancerous. Among various kinds of brain tumors gliomas are the most common one. In the segmentation part of this thesis, the focus is on gliomas. Common symptoms of gliomas include headaches, memory loss, confusion and decline in brain functioning. The task of brain tumour segmentation is to separate healthy tissues from tumour regions such as advancing tumour, necrotic core and surrounding edema.

1.0.3 Motivation and Goals

Alzheimer's disease as well as malignant brain tumours have a high mortality rate. In these cases, early diagnosis plays a very important role in increasing the chances of survival. Also, diagnosis is prone to human error that could be lethal. In order to circumvent these issues, many technological innovations have been proposed to assist medical professionals. These systems have proven to be vital in early diagnosis and resulted in a reduction in medical errors.

After the advent of AI, intelligent decision support systems have captured the market of medical imaging. Various computer vision techniques have been used to assist early diagnosis. And, after the emergence of deep learning in early 2000s, neural networks have outperformed most of the core medical imagining techniques. The go-to method for computer vision using deep learning is the Convolutional Neural Network (CNN). Today, most of the state-of-the-art models in medical imagining are based on CNNs. However, there are some flaws and shortcomings of these models that we try to address in this thesis.

We propose novel 3D-CNN architectures for the task of Alzheimer's detection and brain tumour segmentation. Transfer learning has been used extensively in the task of Alzheimer's detection. However, pretrained weights might contain irrelevant information that could hinder performance and cause higher false positive and false negative rates. For Alzheimer's detection, we aim to build a model that is able to achieve high detection accuracy without using any pretrained weights with less number of parameters.

Another challenging task that we address in this thesis is brain tumour segmentation. It entails a complex computer vision application of semantic segmentation. As pointed out

precisely by [1] that manual segmentation of tumors could be bias based on the relevant experience of the person and their subjective decision making since there are yet no standard protocols to be followed. This gives the rise to the need of having automatic segmentation systems. This task becomes even more complex for medical image segmentation as it consists of 3D images and highly unbalanced class distribution. The irregular shapes, size, location, heterogeneous appearance and heterogeneous appearances of the tumors adds up the challenges to the task. Brain tumour segmentation models are computationally and architecturally complex, but obtain high performance. We aim to further boost performance of a state-of-the-art 3D segmentation model called V-net. To do this, we propose a novel additive self-attention module with a modified V-net architecture. This module encourages the model to focus on relevant sub-regions of the 3D brain scans.

We are able to achieve the above mentioned objectives.

1.0.4 Thesis Outline

This thesis is divided into the following chapters:

2 Background

- This chapter provides preliminary information and introduces to the fundamental building blocks of this thesis.

3 Related Work

- This chapter sheds light on the previous researches done in this area.

4 Methods

- This chapter describes the method approach we have taken to attain the goals of this thesis.
- Detailed description of the experiments performed
- Discussion the data and the the different model architectures we have studied for classification task as well as segmentation task.

5 Results

- This chapter records our observation and results from the experiments done in Chapter 4

6 Conclusion and Future work

- Conclusion of the experiments and results of this thesis and potential future work.

Background

Chapter 2

Background

This chapter walks through the fundamental theory and mechanics of this thesis. In the following sections we present a few concepts of Deep Learning and Bio-medical Imaging that are pertinent to the rest of the thesis. In this chapter we have briefly introduced Convolutional neural networks and related topics that are used throughout the thesis.

2.1 Deep Learning

2.1.1 Convolutional Neural Network

The Convolutional Neural Network is one of the most popular computer vision methods that was proposed in 1980 [18], which was called the Neocognitron. The inspiration behind the CNN was the visual cortex in animals and humans that is responsible for decoding features extracted from the visual environment that are sent to the brain for interpretation. The CNN is designed to extract features from n-dimensional structures like 2D and 3D Images and capture local spatial patterns. This is achieved with local receptive field connections and shared weights. The network consists of one or more convolution layers followed by pooling layers in-between, and connected to one or more fully connected layers, as in a standard feedforward neural network. The main advantage of CNNs over fully connected Neural Networks is that they have a lot fewer trainable parameters.
In CNNs, kernels/filters are used to extract particular features that are present in an image by convolving the kernel weights with an image patch. The size of the filters gives rise to locally connected structure which are each convolved with the image to produce feature maps. The features extracted by kernel maps are aggregated by using mean or max pooling. The kernel weights are shared at all locations in the image, resulting in a reduction in the number of parameters. Each kernel's parameters can be shared across the whole image because the statistics of a patch of a natural image are the same as any other patch of the image, which suggests that features learned at a location are also relevant for other locations in the image. So, we can apply a learned kernel feature extractor anywhere in the image to extract the same feature across the image. This makes CNNs ideal for extracting relevant features from images. Deep CNNs consisting of many consecutive layers have been used

for various computer vision applications such as image classification, object detection and semantic segmentation

2.1.2 Convolution operation

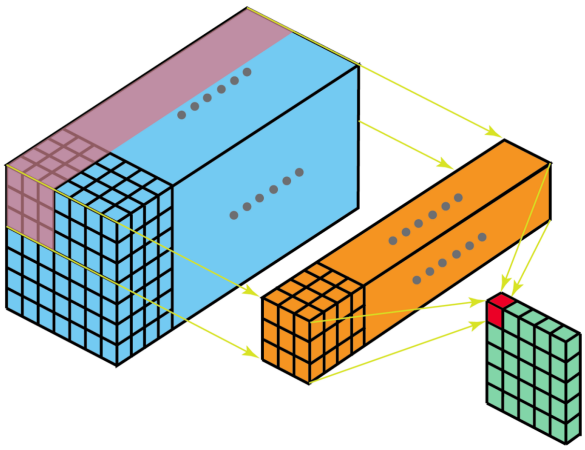
The convolution operation is basically used to calculate the correlation between the kernel weights and the image pixels (or previous layer output). It is an element-wise multiplication operation followed by a summation between the kernel feature map and the input, as shown in the equation below.
Given an input of size N,H,W,D,C_in where H=height, W=width, D=depth, C=no.of channels, N=batch size, the output of the convolution layer is produced as:

$$Out(N_i, C_{out_j}) = b(C_{out_j}) + \sum_{k=0}^{C_{in}-1} weight(C_{out_j}, k) * input(N_i, k) \tag{2.1}$$

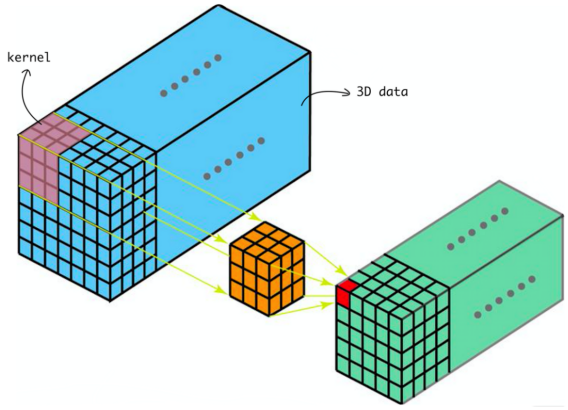
where * is the cross correlation operation (3D cross correlation in our case) between two signals. The learnable kernels are $l \times l \times l$ matrices, which slide over the large input to detect relevant patterns creating new feature maps and convolving feature maps from previous layer.
The idea behind this operation is that the kernel weights are learned using backpropagation so that they extract relevant features from the input. A kernel's weights are spatially shared for all locations in the image. This is because a feature extracted by a kernel at a particular part in an image is relevant for other parts of the image as well. This is an important property of CNNs as it reduces the number of parameters. It is performed as a matrix operation rather than an individual element-wise multiplication and summation in practice in almost all widely used deep learning libraries.

2.1.3 2D vs 3D CNN

In a typical 2D CNN, where input is something like image data, the convolution is performed in 2 dimensions using 2D kernels. While 2D CNNs are great at capturing spatial features they lack the potential of capturing the temporal information present in 3D data like MRI images.
3D CNNs apply convolution in 3 dimensions hence capturing the temporal as well as the spatial features present in the data describing the relationship of instances in 3D space. The 3D convolution is achieved by convolving a 3D kernel to the cube formed by stacking multiple contiguous frames together.
Fig 2.1 shows the 2D and 3D convolution operations



(a) 2D convolution



(b) 3D convolution

Figure 2.1: Difference between 2D and 3D CNN

2.1.4 ReLU

The Rectified Linear Unit[19] is an activation function most widely used in deep neural networks. It can be regarded as a truncation operator applied element-wise on the input. The rectified linear unit (ReLU) is defined as

$$f(x) = max(0, x) \tag{2.2}$$

The derivative of ReLU is:

$$f'(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0, & otherwise \end{cases} \tag{2.3}$$

There is no parameter inside a ReLU layer, hence no need for parameter learning in this layer. It is used as a nonlinear activation function in neural networks, as shown in Fig. x. In the case of images, the ReLU function will let through features that result positive for a certain patterns and will render other negative patterns to zero.
In this way, the ReLU function induces sparsity in the model which reduces complexity and memory consumption. Other popular activation functions such as sigmoid and tanh have a tendency of saturating at their extremas. ReLU has many variants but we have used

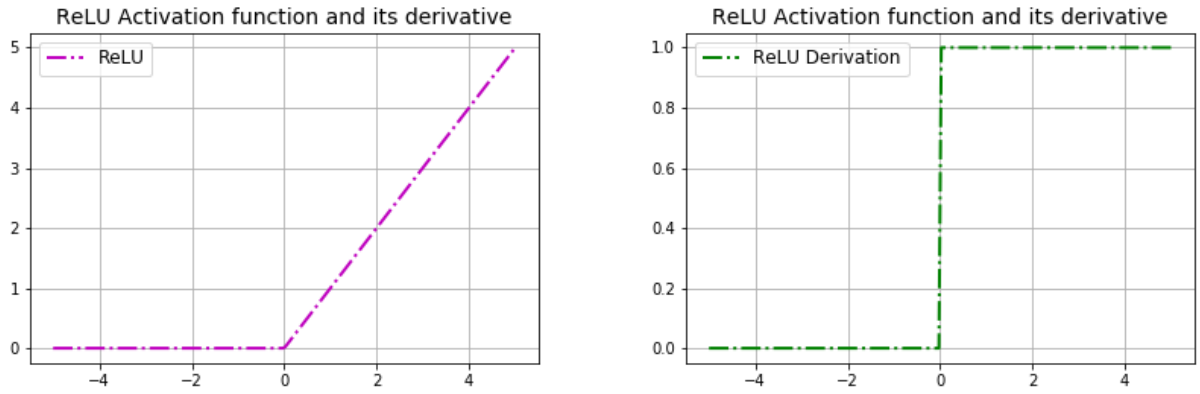


Figure 2.2: Plot of ReLU and it's derivative

ReLU because it encourages the model to learn sparse features early in training. It clips the activation at value 6. It is given as

$$y = \min(\max(x, 0), 6) \tag{2.4}$$

2.1.5 Batch Normalization

During the forward pass, the multiplication between weights and the input shifts the distribution along the direction of the weights. In a deep neural network, subsequent operations result in a huge shift in the distribution, which leads to loss in performance. This is called internal covariate shift. In order to tackle this problem, the batch normalization layer was proposed in [31]. Just like the input to the network must be normalized with zero mean and unit variance, the batchnorm layer normalizes the input to each layer by calculating the mean and variance in a batch-wise manner.

$$y_i = \gamma \hat{x_i} + \beta \tag{2.5}$$

$$where,\ \hat{x_i} = \frac{x_i - \mu_B^k}{\sqrt{\sigma_B^{(k)^2} + \epsilon}} \tag{2.6}$$

for a d-dimensional input, where $k \in [1, d]$ *and* $i \in [1, m]$, μ_B^k and $\sigma_B^{(k)^2}$ are the per-dimension mean and variance of the mini batch B. γ *and* β are shifting and translation parameters learned during training.

2.1.6 Pooling Layers

Pooling layers are used to downsample the feature maps produced by the convolution layer by summarizing the features. In this way the resultant feature maps contain more global information even at lower resolution making them robust to the location changes in the input image. Two ways to do pooling operation is

1. Maxpooling, where the maximum value is taken from the patch of the feature map.

2. Average pooling, where the average of the patch is taken as the new value in the resultant feature map.

2.2 Medical Imaging

Medical Imaging can be described as a tool-set in medical science for diagnosis and clinical intervention of the internal organs of a living body. It is a way to visualize the insides of an organ or a body without opening it surgically or with any other invasive procedures; which makes it one of the most safest and powerful resources available for diagnostic purposes. The utilization of imaging technology has increased drastically with all the improvement in the science and technology in the last couple of decades, as a result of which today a number of different imaging modalities are available suiting the need of the diagnostic measure. Various types of imaging techniques used in medical radiation are MRI(Magnetic Resonance Imaging), PET(positron emission tomography) Scan, Ultrasounds, CT (Computer Tomography) Scans, X-Rays etc. All these imaging techniques work slightly differently from each other. The following section give a brief description of some of these technologies.

2.2.1 Types of Medical Imaging

Magnetic Resonance Imaging (MRI)

Magnetic Resonance Imaging is the one of the medical diagnostic measures used in radiology to form pictures of the anatomy and the physiological processes of the body. MRI uses strong magnetic field projected through MRI scanners to generate the medical images.

Positron Emission Tomography (PET)

PET scans are used to plot the activity and functioning of the organs and tissues using a radioactive drug called tracer which glows up in highly chemically active areas when scanned through a PET scanner. Certain diseases have high chemical activity so in this way it helps in identifying the disease

Computer Tomography (CT)

CT scan uses X-rays projected on the body from different angles and then processed by computer to create cross-sectional images of the internal body parts. These are more-detailed than normal X-ray scans.

2.2.2 Magnetic Resonance Imaging (MRI)

In this study we have used MRI images for both classification and segmentation task as the data was readily available in this format. Fig 2.3 shows the view of MRI in all three axis. To create MRI image, strong and uniform magnetic fields are required. The strength is

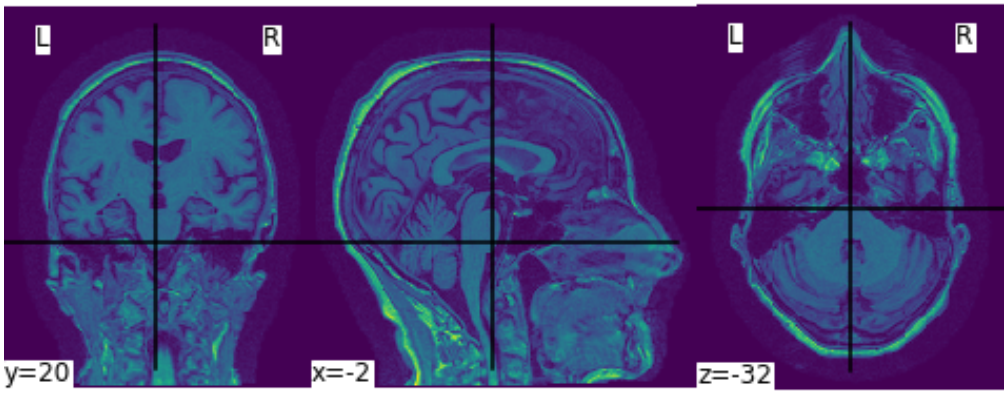


Figure 2.3: Coronal View, Sagittal View and Axial View (left to right)

measured in Tesla(T). A MRI sequence is a particular set of setting of pulse sequences and pulsed field gradients to create MRI images with specific properties. A brief introduction to the properties of MRI sequences we have used is given below.

1. T1: Lower signal for high water content (edema, tumor, infarction etc) and high signal for fat

2. T2: High signal for more water content, low signal for fat.

3. T1-Gd: It is type of contrast based sequence. It uses gadolinium-based contrast agent to improve the visibility of the MRI. Often T1 MRI is used as a base MRI, and it's visibility is improved using gadolinium-base agent.

4. FLAIR: It is an inversion recovery based sequence, where signals from the cerebrospinal fluid (CSF) are removed. It results in a T2 MRI where grey matter is brighter than white matter but CSF is dark instead of bright

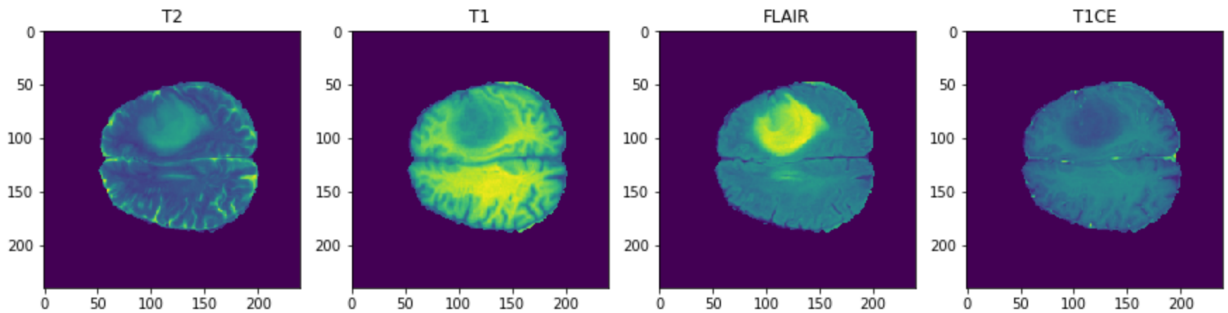


Figure 2.4: MRI Sequences

2.2.3 Identifying Alzheimer's Disease on MRI

The area earliest affected by the Alzheimer's disease is the hippocampus and its connected structures. Hippocampus is needed for memory retrieval, hence person affected with Alzheimer's struggle to remember things. As the disease spread the cortex become thinner

and brain starts to shrink. Other areas damaged by the disease are parietal lobe, frontal lobe and temporal lobe. MRI image in Fig 2.5 shows the different lobes on the brain

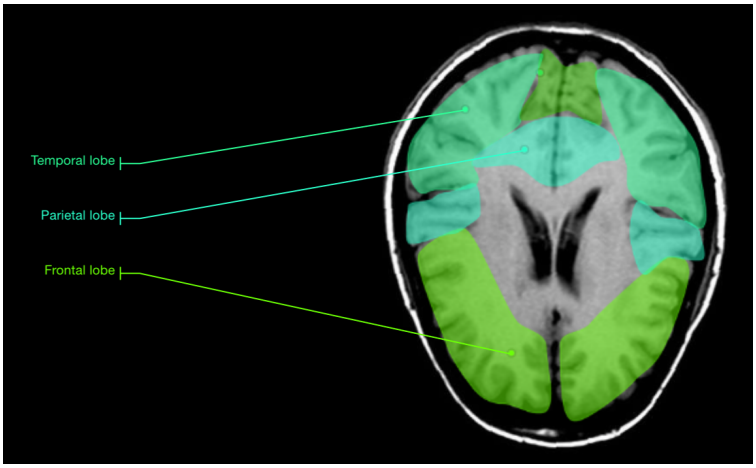


Figure 2.5: Alzherimer's Affected parts

2.2.4 Identifying Tumors on MRI

Gliomas are the most common type of brain tumors and can be further divided into High Grade Gliomas (HGG) and Low Grade Gliomas (LGG) based on the growth rate. In the segmentation section of this study, we have focused on the following tumor sub-regions[1]:

1. NCR: Necrotic Core, is present in the enhancing core of the high-grade gliomas
2. ED: Edema are peritumoral edematous and invaded tissues, present in tentacle shapes and are easily visible on T2-weighted MRI images.
3. NET:Non-enhancing gross abnormality.
4. AT: Enhancing regions within the gross tumor but not in the necrotic center.

Fig 2.6 shows the different tumor sub-regions explained above.

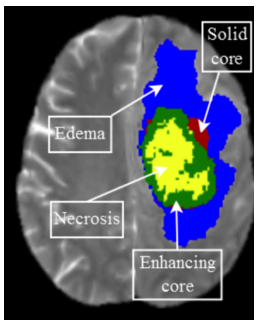


Figure 2.6: Tumor sub-regions

[1]These definitions are taken from **(author?)** [6]

Related Work

Chapter 3

Related Work

The need of automation of biomedical imaging analysis has been there for long. In early times low-level pixel processing and mathematical model based systems were used to solve a set tasks. Gradually the focus shifted to more modern approaches which includes utilizing the potential of rapid-growing filed of Artificial Intelligence. By the end of 1990, supervised machine learning techniques became popular especially in the field of biomedical image analysis. Applications included active shape models (for segmentation), atlas methods (where the atlases that are fit to new data form the training data)[37]. General reason for the popularity of machine learning techniques is their ability to learn and recognize pattern with great accuracy and much efficiently if compare to humans. There has been a huge upgrade in the field of AI since then. The related work revolves around medical image classification and segmentation in for Alzheimer and brain tumors. In this section we will study some of the work that has been done so far for the classification and segmentation of medical imaging using deep learning for Alzheimer's Disease and Brain Tumor respectively.

3.1 Classification of Alzheimer's Disease

Several studies have been conducted in the recent years to develop a computer-aided diagnosis system for Alzheimer's detection. Traditional methods included researcher trying to handcraft features through voxel-based methods, ROI based methods, hippocampal shape and volume or patch-based methods.
[44] have attempted to compute region of interest (ROI) to detect AD. ROI is a section of image in which a binary mask is used to carry out various operations like filtering. [8] used voxel based morphometry (VBM) and MRI to investigate gray matter change in medial temporal structures and volume changes in several other brain regions.
[58] have analysed regional brain atrophy for example in the hippocampus to detect patterns of neuron death by segmenting different types of brain tissues such as grey-matter (GM), white-matter (WM) against cerebrospinal fluid (CSF) in the MRI. They segmented the images using watershed transformation algorithm[50] with marker image, and then calculating the shrinkage happened in the whole brain through Tissue Atrophy Ratio (AT) for early

detection of AD.

[39] has used an inherent structure-based multi-view learning (ISML) method in which they have extracted multi-view features based on multiple selected templates. They then employed a subclass clustering algorithm for feature selection in order to eliminate the redundant features. A SVM-based ensemble classifier is used to classify subjects into AD, MCI and NC.

[47] have used Hu moments invariants[29], calculating a set of seven invariant moments to extract features in the brain images (MRI) of all subjects. They also showed that normalizing these moments results in better feature extraction which makes it easier for the classifier to distinguish. The extracted features are then used as inputs to SVM and KNN classifiers to classify the subjects. They compared the classifiers, showing that SVM performed far better than KNN.

In [21] feature extraction is done by using ROI on three sMRI biomarkers, named as Voxel-based morphometry, Cortical and sub-cortical volume and Hippocampus volume. They used Principal component analysis (PCA) [4] for feature selection. PCA is a dimensionality reduction method simplifying a high dimension data into smaller dimension without losing the important patterns or trend in the data. Using PCA, they selected 61 features for the classification of AD. They studied three different classifiers: SVM, Random Forest and KNN, and evaluated their performance in two stages. First stage included individual features from s VBM-extracted ROI volumes, CSC-extracted feature volumes, and HV extracted features and second staged is evaluating classifiers using the combination of all 61 features. They concluded that SVM outperformed KNN and Random Forest in all cases.

Multi-modal data fusion using MRI and PET scans was proposed in [42]. They used stacked auto encoders and and a sigmoidal decoder to discover the synergy between MRI and PET scans for high level feature extraction with a softmax classifier. A zero-masking technique (SAE-ZEROMASK) is used in contrast to simple feature concatenation (SAE-CONCAT) technique. They randomly hide one modality and trained the hidden layers to reconstruct the multi-modal using inputs mixed with hidden modality.

However, SAE-CONCAT usually fails to captures the non-linear co-relation between two different modalities[59]. That's why authors in [59] proposed Multi-modal Stacked Deep Polynomial Networks algorithm (MM-SDPN) which uses multi-modalities like [42] but they have used two stages of SDPN to learn high-level features.

Various other machine learning algorithms have proven to be efficient when it comes to extracting high level features. Artificial neural networks were used by [22] for Nephropathy Detection and Classification. The drawback of using Feed Forward Neural Network (as usually called ANN) for computer vision is that they are computationally expensive. The number of learning parameters in ANN exponentially increase with respect to the size of the image. Thus to counter this problem, the use of convolution neural network (CNN) to automate feature learning in images has become popular because of their ability to generalize well to high dimensional data, without losing important patterns.

A 2-D CNN is presented by [2] where they used VGG16[60] as a base model, and treated

a 3-D MRI image as a stack of 2-D MRI slices. Other variants of CNNs have been used in the researches such as in [35] authors used a ROI focused 3-D CNN with multi-modality. Each modality and ROI region was assigned a dedicated pipeline of a CNN block, whose output was flattened. The flattened outputs are the extracted features from each modality and region of interests(ROI). These feature outputs were then concatenated, resulting into late data fusion and were passed to a softmax classifier. Problem with this approach is similar to what was described earlier, that while performing late data fusion using simple concatenation, it ignores the variance in the nature of multi-modalities and fails to learn the non-linear co-relation between modalities[59].

A few other research works employ pre-training 3-D CNN with auto encoders such as [53].They used sparse auto encoders for feature extraction and also compared the performance of 2-D CNN against 3-D CNN. Authors in [26] took a two stage approach where they first used a convolutional auto encoder in place of conventional unsupervised auto encoder to extract local features with possibly long voxel-wise signal vectors. These features are used to perform task-specific classification with a target-domain-adaptable 3D-CNN using transfer learning with Net2Net weight initialization. They later proposed [25] in which they trained the same model with deep supervision, which resultant in an improvement.

[15] also build a 3-D convolutional neural network for an end-to-end classification of subjects with AD. They added metadata (sex and age of subjects) to the first fully connected layer in their model. The downside of using metadata in the neural network is that the network will try to find the correlation based on the metadata that might be biased towards the predilection of meta-data, for e.g., older patient are more likely to be affected by Alzheimer's Disease, so the network might bias towards assigning older people to the Alzheimer's Class.

[67] studied various paradigms of 3-D CNNs like patch-level 3D CNN, ROI based 3D CNN, subject-level 3D CNN, along with exploring transfer learning using auto encoder pre-training and ImageNet pre-training. They also reviewed studies done on AD classification using Deep Learning from January 1990 to the 15^{th} of January 2019, which proved very helpful in the proposed research.

3.2 Segmentation of Brain Tumors

[1] proposed a 2D CNN architecture to segment tumors in MRI images on BraTS'15 dataset. [62] proposed an Enhanced Convolutional Neural Networks where they divided the segmentation task into two parts, first preprocessing with image enhancement and second is to calculate the segmented mask using Hybrid Convolutional Neural Networks. They also introduced a novel loss optimization function called Novel BAT optimization algorithm (NOBA) which uses the concept of echolocation mechanism to calculate the difference between an optimal and non-optimal error. [65] introduced a scribble-based segmentation pipeline incorporating the CNN into a bounding box. They used P-Net as a base model for bounding box-based binary segmentation. They made a model compatible with both 2D data and 3D data with some minor changes.

[57] proposed a 2D architecture called U-Net. It is a key architecture in medical image segmentation. The model has great ability to localize the features and inspired a lot more models. For example [17] proposed a memory efficient version of U-Net regularized using auto-encoders to segment brain tumors. A combination of mean-squared error and Kullback-Leibler loss was used to as a loss function for the encoder, while SoftDice loss was used to train the U-Net inspired network. [32] also carefully modified the architecture of U-NET to boost the performance and accuracy of the biomedical imaging segmentation task.

While most of the studies focused on medical image segmentation using patches or 2D slices, authors in [46] proposed an architecture capable of performing volumetric segmentation of medical imaging called V-Net. The applied their model to the segmentation task on the Promise 2012 dataset. The introduction of residual blocks in their network ensured the convergence of the model much faster compared to other studies. [10] in their work made use of two cascaded CNNs inspired from V-Net architecture and modified residual connections for brain tumor segmentation problem. Their first network segments the overall tumor and the second network then delineation of the different tumor regions using the output of first network as input. Their study contributed toward BraTS'17 challenge. Inspired from U-Net, [12] propose a 3D version of it, called 3D U-Net, which became another one of the key architectures for biomedical imaging segmentation.[68] also heavily used U-Net architecture in their study for segmentation of brain tumors in BraTS'18 challenge. They exploits a 3D U-Net based model to first locate the tumors in the brain and an another but more complex and smaller 3D U-Net to further segment the localized tumor into its sub-regions. [64] also create a 2 staged 3D U-Net framework, where they utilized the potential of image super-resolution CNN (SRCNN) to process the MRI images at full resolution. They first detected the ROI from the full volumes and predicted the segmented masks from these ROIs.

Since 3D CNNs have large memory consumption, and 2D CNNs while having low memory requirement ignore the 3D context in the data, [66] proposed a novel framework to use 2.5D CNN, that is a trade-off between memory consumption, model complexity and receptive field. They evaluated their model on BraTS'17 data and ranked second in the challenge. They propose a test-time augmentation technique claiming to improve segmentation accuracy.

DeepMedic[34] is a multi-scale deep 3D CNN for lesion segmentation. The architecture consists of two parallel convolutional pathways. The pathways process images at different resolutions giving it a better receptive field for the final classification. Inspired from DeepMedic, [34] proposed their architecture as extended DeepMedic with residual connections. They tested their results on BraTS'15. The residual connections gave a modest but consistent improvement.

Many researches revolved around the idea of using encoder-decoders to segment lesions in brain. [49] proposed encoder-decoder based architecture to segment tumors in 3D MRI images. They also used a variational-autoencoder joined to the main architecture, to reconstruct the input image and regularize the shared decoder. The architecture ranked 1^{st} place in the BraTS 2018 challenge.[54] has combined the model of normal brain tissues and the

spatial information provided by model of tumor and edema. They have used a three step approach where they first compute the region of interest which is any abnormal regions, in second stage they find whether the computed region contains both tumor and edema, and in the final stage they create the segmented labels using the proper sample locations determined through the knowledge of spatial and geometric properties discovered in the second stage.

[14] trained a regression network with a fully 3D convolutional architecture for lesion detection in from weak labels only.[28] proposed a novel segmentation method based on multi-cascaded convolutional neural network (MCCNN) and fully connected conditional random fields (CRFs). They trained three segmentation models each dedicated to one of the three axial, coronal, and sagittal MRI patches, and finally combined them to obtain the final segmentation. They take the advantage of multi-cascasded CNN to account the local dependencies of the labels by combining the results of the various intermediate components and have used CRF to extract the spatial contextual information. Similarly, [13] has also CRF to achieve the spatial and appearance accuracy. They have trained their model on 2D patches and the proposed model is trained in three steps where they first train a Heterogeneous Convolution Neural Networks (HCNN) to identify the tumor in the patches, then they trained CRF with CRF-Recurrent Regression based Neural Network (RRNN) and in the last step they fine tuned the with HCNN and CRF-RRNN image slices. Similar to [28], they also trained three segmentation model one for each view, and then combined the results using voting fusion technique.

[48] proposed an ensemble of multi-dimension and multi-resolution networks for brain tumor segmentation. The developed 2D and 3D segmentation models and ensembled them to create robust segmentation maps. They developed their ensembled models based on DenseNET-169[30], SE-RESNEXT-101 and SE-Net-154[27]. They segment the tumor components seperately and then comibine them followed by post processing. A novel reinforcement based approach is used by [7] where they propose to use the thermal information present in tumor regions of the brain scans to improve the results of the segmentation task. They calculated the thermal maps of the tumor regions by solving the Pennes bioheat equation. They extracted tumor contours from the calculated thermal maps using Canny edge detector[9].

Methods

Chapter 4

Methods for 3D Convolutional Neural Network for Medical Diagnosis

The study is divided into two sections:

1. Classification of Diseases from MRI images
2. Segmentation of tumors in MRI scans

Task 1 is specific to the detection of Alzheimer's Disease and in task 2 we have performed segmentation of different kinds of tumour sub-regions in Brain MRIs with computer vision. This chapter gives a detailed explanation of the tasks, and the proposed methods to solve the respective tasks. In this chapter we will also walk through the description of the datasets used in this thesis, and training methods and experiments we have performed.

4.1 Classification of Alzheimer's Disease

4.1.1 Overview

For the classification task we have worked with Alzheimer's Disease Data the algorithm we have considered for Alzheimer's Disease detection is 3D Convolutional Neural Network. We have used 1.5T brain MRI dataset collected from the Alzheimer's Disease Neuroimaging Initiative(ADNI). ADNI is a research collaboration started in 2004 under the leadership of Dr. Michael W. Weiner. It is private-public partnership between with $27 million contributed by 20 companies and two foundations through the Foundation for the National Institutes of Health and $40 million from the National Institute on Aging, designed to collect and study clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of Alzheimer's disease (AD). The goals of ADNI include early detection of Alzheimer's in the stage of pre-dementia and track the progression of the disease by tracking its biomarker. The timeline of ADNI initiative consists of 4 phases called ADNI-1, ADNI-GO, ADNI-2 and ADNI-3. A brief introduction of all the phases is given below.

ADNI-1

Started in 2004 as phase-1, ADNI-1 continued for a duration of 5 years. The goal of this phase was to develop the biomarkers as outcome measures for clinical trials. The study in this phase gathered and examined thousands of brain scans including structural MRI and PET (both FDG-PET, genetic profiles, and blood and cerebrospinal fluid biomarkers. Many subjects were included in the study where 400 were diagnosed with mild cognitive impairment (MCI), 200 subjects with the early AD, and 200 elderly control subjects.

ADNI-GO

ADNI-GO where GO stands for "Grand Opportunities" was started in 2009 after ADNI-1 and continued for another 2 years. The goal of this phase was to examine the biomarkers developed in Phase-1 in the earlier stage of Alzheimer's Disease.

ADNI-2

ADNI-2 was started in 2011 with an addition of some new participant groups. 150 elderly controls, 100 EMCI subjects, 150 late mild cognitive impairment (LMCI) subjects, and 150 mild AD patients were added to the study during ADNI-1. 107 subjects with Significant Memory Concern(SMC) also participated in the study. This phases lasted for 5 more years

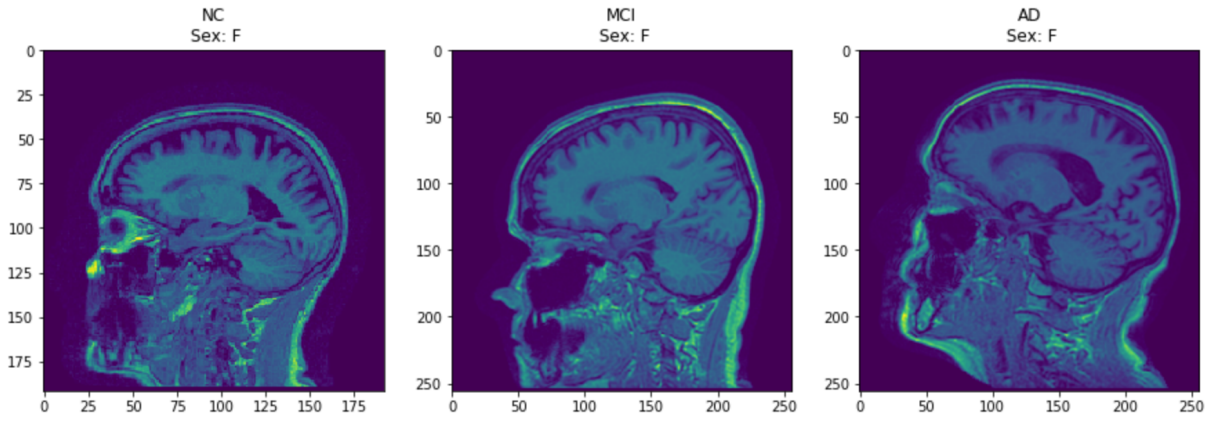


Figure 4.1: MRI images from the dataset for three different stages of Alzheimer's in Females

ADNI-3

ADNI-3 is the current ongoing phase of the ADNI initiative. It started in September 2016 and is aimed to complete in 2022. In ADNI-3 some new brain scans has been added which could detect the presence of tau protein tangles (tau PET), which is a key indicator of Alzheimer's Disease. The goal of this phase is to study the use of tau PET and functional imaging techniques in clinical trials.

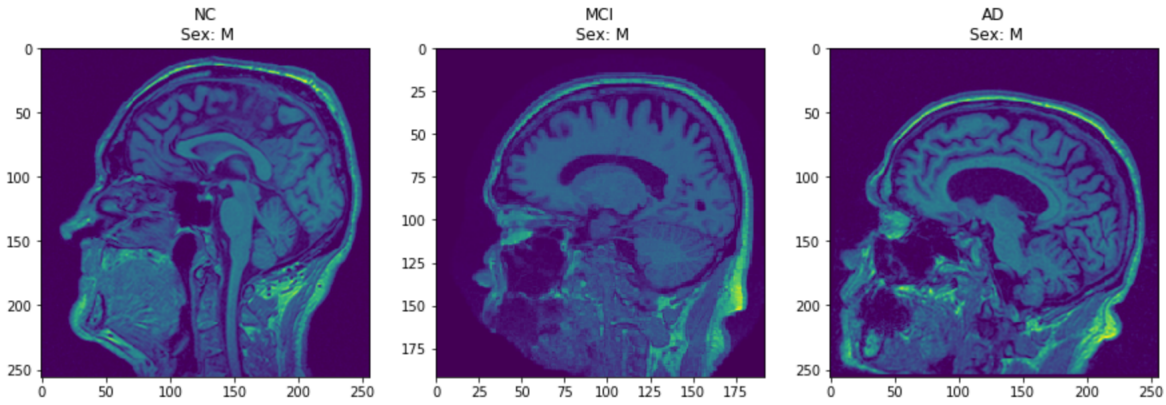


Figure 4.2: MRI images from the dataset for three different stages of Alzheimer's in Males

4.1.2 Dataset

The data used in this study consists of 1075 1.5T Screening MRI records of 817 subjects collected from the ADNI-1 Standardized Data Collections. The demographic information of the subjects is given in the table 4.1 .

Table 4.1: Demographic data for 817 subjects from the ADNI database (STD - standard deviation))

Diagnosis	Subjects	Age($mean_{\pm std}$)	Gender (F/M)
AD	188	$75.36_{\pm 7.5}$	89/99
MCI	401	$74.84_{\pm 7.3}$	143/258
NC	228	$75.96_{\pm 5.0}$	110/118

The MR images were downloaded in NIfTI format. NIfTI is a medical image file format initially developed for neuroimaging analysis by Neuroimaging Informatics Technology Initiative (NIFTI). The configuration of NIfTI format contains the metadata and image data in a single file, with metadata stored in the beginning of the file.
The original resolution of MRI images was inconsistent so we downscaled them a resolution of $120 \times 90 \times 130$, to have a standard size which can be process by our CNN. We followed the common practice of downscaling for the deep learning model, to process the large images. We have applied min-max normalization to the data for feature scaling. Min-max normalization transforms the features in such a way that every feature will have exact same scale between 0 and 1. Mathematically it is given as:

$$X = \frac{x - X_{min}}{X_{max} - X_{min}}, x \in X \tag{4.1}$$

where X is our dataset of size $1075 \times 120 \times 90 \times 130$, and x is each individual MRI image. Alzheimer's is a disease which can emerge in any of the hemisphere of the brain, so

augmenting the data by flipping the MRI images horizontally is a good notion. We have used only left and right flip augmentation of the data, which doubled the size of our dataset to train the neural network. We have chosen axial view as it helps in avoiding the motion artifacts from eyeball which can appear in other views [20].

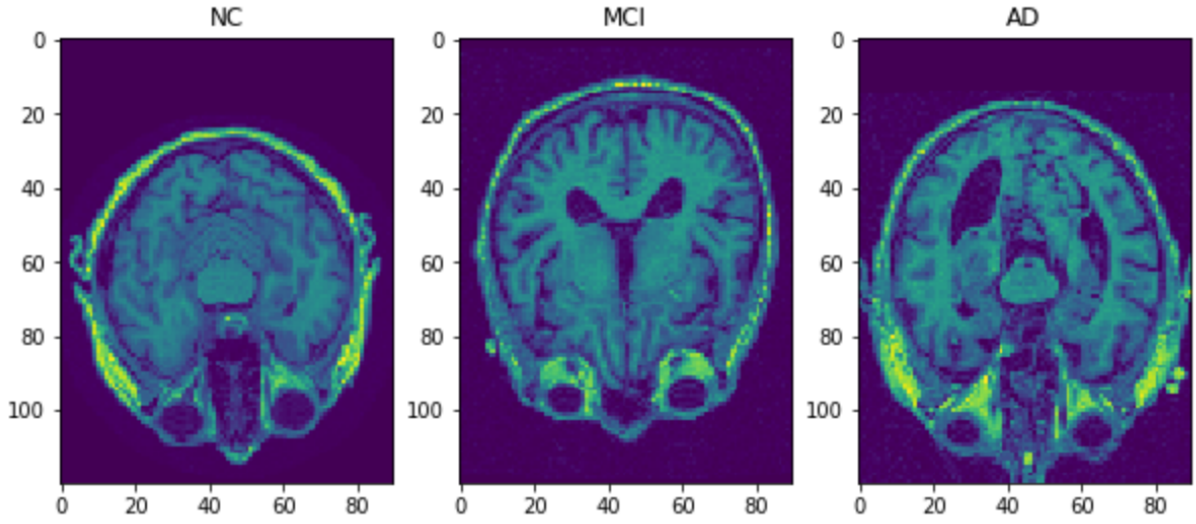


Figure 4.3: Processed MRI images from the dataset showing three different stages of Alzheimer's

4.1.3 Architecture

Figure 4.4 shows the architecture of the 3D CNN used to detect Alzheimer's. The architecture shown is for binary classification of Alzheimer's between two stages of Normal Cognitive(NC) and Alzheimer Disease(AD). The architecture of the model for multi-class classification among three stages (NC, AD and Mild Cognitive(MC) is exactly similar except the number of output neuron.

The 3D CNN we have designed for the classification task is a deep convolutional neural network consisting of 12 convolutional layers with batch-norm and maxpool layers in between as shown in the architecture figure. The architecture takes an inverted bottle-neck shape as the number of kernels increase at every layer, but at 8^{th} convolutional layer a small wave shape at the end occurs where the number of kernels in the layer is smaller than the previous layer. The network contains two blocks like this. The objective behind experimenting with this kind of architecture was to design a network efficiently extracting relevant features with lesser number of parameters resulting in increased robustness of the model.

The model uses 3D kernels for feature extraction which are $l \times l \times l$ matrices. 3D kernels also called filters create feature maps by sliding over the larger input to detect relevant patterns in the data, later on convolving feature maps from previous layer.
Table 4.2 gives a thorough insight of the model. The model consists of an input convolution

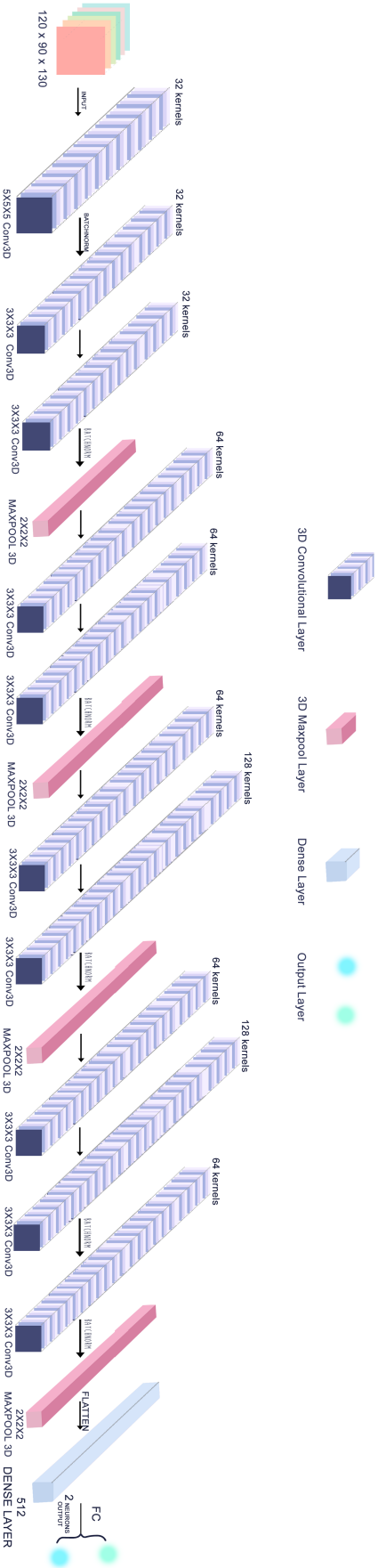


Figure 4.4: Architecture for binary classification of Alzheimer's disease

block, a rearmost convolution block, a concluding classifier block and in-between four identical convolution block with two 3D convolution layers stacked up on a batch-normalization layer in each block. Initial three blocks from this group goes through a 3D maxpool layer with a pool size of (2, 2, 2). Maxpooling reduces the spatial size of the resulting feature map by accumulating the generated features. Strided maxpooling is much faster, computationally efficient and provides the same amount of translation invariance a convolution layer would have provided if used inplace of it for downsampling. The input convolution block is made up of one 3D convolution layer followed by batch normalization layer. The rearmost convolution block consists one unit of each 3D convolution layer, batch normalization layer and maxpool layer in the respective manner. The classifier block takes the flatten input from the preceding block, sends it to fully connected layer followed by a softmax classifier. All convolution blocks, except the input block exercise kernels of size $3 \times 3 \times 3$ focused on learning the small details of the patterns in the Alzheimer's affected parts of the brain.

Table 4.2: Model Architecture

Layer Name	Kernel size	No of kernels/neurons
Conv3D	$5 \times 5 \times 5$	32
BatchNorm	-	-
Conv3D	$3 \times 3 \times 3$	32
Conv3D	$3 \times 3 \times 3$	32
BatchNorm	-	-
MaxPool3D	$2 \times 2 \times 2$	-
Conv3D	$3 \times 3 \times 3$	64
Conv3D	$3 \times 3 \times 3$	64
BatchNorm	-	-
MaxPool3D	$2 \times 2 \times 2$	-
Conv3D	$3 \times 3 \times 3$	64
Conv3D	$3 \times 3 \times 3$	128
BatchNorm	-	-
MaxPool3D	$2 \times 2 \times 2$	-
Conv3D	$3 \times 3 \times 3$	64
Conv3D	$3 \times 3 \times 3$	128
BatchNorm	-	-
Conv3D	$3 \times 3 \times 3$	64
BatchNorm	-	-
MaxPool3D	$2 \times 2 \times 2$	-
Flatten	-	-
Dense	-	512
Dropout(0.1)	-	-
Dense	-	2

4.1.4 Training

Weight initialization is a very important step in training deep learning models. If the parameters of a neural network are initialized correctly it helps in achieving the optimization in relatively short period of time as well as prevents prevents the layer activation outputs from exploding or vanishing during the forward pass of the training. There are many techniques for weight initialization. General practice is zero initialization of bias and random initialization of weights. The problem with this approach is that since weights are randomly initialized, they can be initialized with either too large or too small values. This often leads to exploding and vanishing gradients respectively. To overcome this issue, some weight initialization techniques have been developed in the recent years. For example Xavier initialization and He initialization. These are activation aware initialization. Both are almost identical, but

Xavier initialization works better for tanh activation function while He initialization is used for ReLU activation function.

For the training of our model we have used He initialization to achieve optimization in relatively shorter time. In He initialization, we use the random initialization multiplied by a factor given as,

$$\sqrt{\frac{2}{n}} \tag{4.2}$$

$$W^{[l]} = random^{[l,n]} \times \sqrt{\frac{2}{n}}$$

where $W^{[l]}$ is the weight initialization, n is the size of last layer/number of connections coming from last layer, l is current layer and $random^{[l,n]}$ is a random vector of size $l \times n$
The model takes a batch size of 2 samples because of memory constraints. One common problem faced during the training of deep neural networks is overfitting, when the model performs too well on the training data but not on unseen data. We have used l2 regularization and drop out on the last fully connected layer to regularize our model and tackle overfitting. We have used batch normalization layer in the convolution blocks to reduce the covariance shift in the hidden units/kernels of the network. It increases the stability of the network by normalizing the output of the previous layer using the current batch statistics. In this way the kernels of our model can adapt to any change in the distribution of the data. Batch normalization also gives a slight regularization effect as it scales the weights based on batch mean and variance, which makes the weights not to fluctuate too much.

We have used k-fold cross validation with k=10 to validate the training of our model. K-fold cross validation results in a less biased model because the model is trained and tested on various subsets of the data. It ensures that the model will perform well on any data point and not on just some lucky random set. To perform k-fold we have divided the data into 10 parts, and trained the network on 9 parts and tested on the remaining one part. From the 9 parts we used 10% data for validation during training. We repeated this process with all the 10 parts.

4.1.5 Experiments

We experimented with various depths for the model and different hyperparameters. The hyperparameter estimation is done through randomized search. For initial learning rate, values were to be chosen from set [0.01, 0.001, 0.0001,], dropout from [0.1, 0.15, 0.2] and the *l2* regularization parameter for the classification layers from [0.001, 0.0001]. The search picks the best parameters to be 0.0001 for initial learning rate, a dropout factor of 0.1 and regularization parameter to be 0.001.

Random Search Methods often picks the best results much faster as compared to Grid

Table 4.3: Hyper parameter search grid

Learning Rate	Drop out	L2
0.01	0.1	0.001
0.001	0.15	0.0001
0.0001	0.2	-

Search method. In Grid Search method we try all the possible combinations for all the values for each hyperparameter, which can be computationally expensive if the size of the dataset is large. By contrast Random Search method selects random combination to train the model. In this way it gives us more control over how many attempts and number of parameter combinations we want to try to get an optimal value for the hyperparameter.

We have trained the model for 100 epochs in each fold. We also experimented with longer training sessions of 500 epochs and 200 epochs, but found that the model converges at 100 epochs efficiently so we restricted the training to 100 epochs only. While training, the model makes use of schedule learning rate based on step decay where the learning rate is decreased by a factor of 0.1 after every 40^{th} epoch.

We observed the effect of Batch normalization Layer and L2 regularization on the performance of our model. We have used Batch normalization layer in our final model as mentioned earlier. We initially tried L2 regularization on convolution layers along with Batch normalization layer, and observed no significant regularizing effect on our model as claimed in [36] too. To experiment further we tried L2 regularization and Batch normalization layer separately and found an increment of 3% in the model accuracy in binary classification and an increment of 2.34% in multi class classification.

We have used a *"build and train from scratch"* approach which means we have designed the architecture of our model from scratch and trained it from scratch without any transfer learning as opposed to other Deep Learning based Alzheimer Detection models, which have either used some well-known CNN architecture as a backbone or have used pre-trained models for transfer learning or some kind of additional data. For example [16] have used demographic

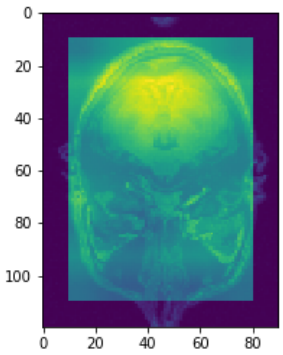
information like age and sex of the subjects as additional features in their network for the Alzheimer's classification task. This additional data will make the model learn some new features based on the correlation present in the statistics of demographic data. At first it might look like a good and useful approach. But if the model starts to give more weightage to the information present in the demographic data than the MRI data, chances are the model will be highly influenced by the bias present in the demographic information.

To clear this further, we can take the example of age bias present in Alzheimer's patients. According to [24] out of 4.7 M people aged over 65 with Alzheimer's disease in the United States, around 81% of them are of or over 75 years old. And according to [5] about two-thirds of clinically diagnosed cases of dementia and AD are women. Such statistics if provided to the model as additional features, can result in a model highly biased towards a particular age or sex group. Hence to eliminate such bias, we have made use of only imaging data inform of MRI images. It results in model which is able to classify Alzheimer's disease based only on the features it has learned from the anatomy of the brain from the MRI scans.

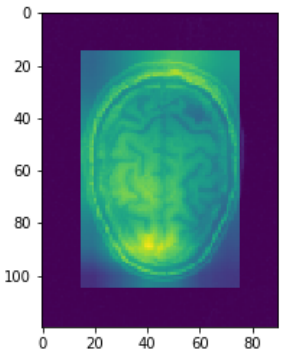
Visualizing Model

To estimate where the model is looking at for the classification, we have create class activation heatmaps using Gram-CAM algorithm. Gram-CAM algorithm helps us to visually validate our model. Fig 4.5 shows heatmaps for two random slice of two random samples taken from AD class and NC class respectively.

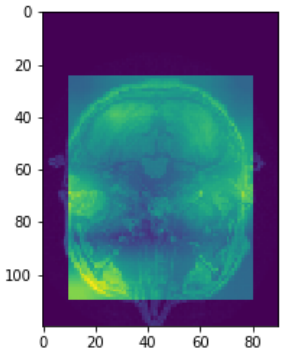
Fig 4.6 shows the some sample heatmaps averaged over AD and NC samples in the validation set of a random fold. We can see that to classify Alzheimer's the model focuses on the relevant areas of the brain. In the slices we can see the highlighted areas around the frontal, temporal and parietal lobe which are the most affected areas during Alzheimer's. This proves the model is classifying based on relevant feature learning.



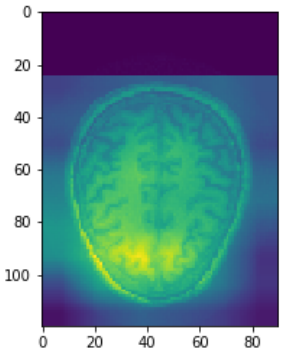
(a) AD class sample slice 44



(b) AD class sample slice 88

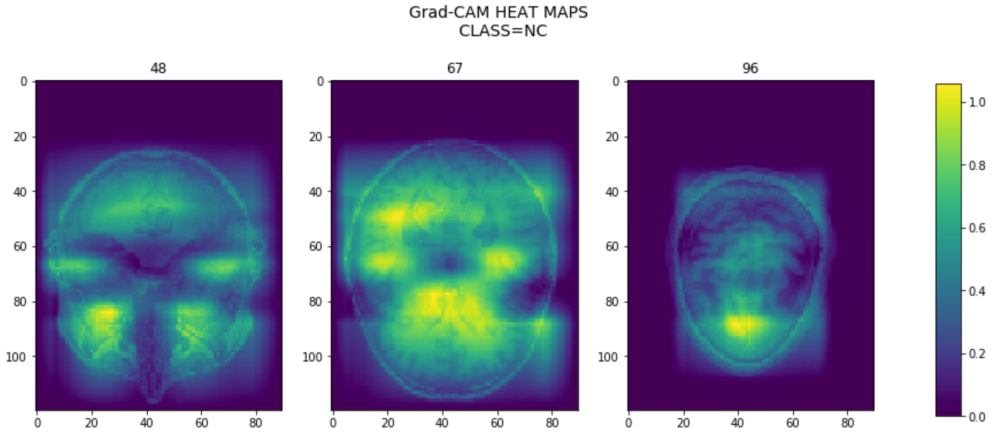


(c) NC class sample slice 44

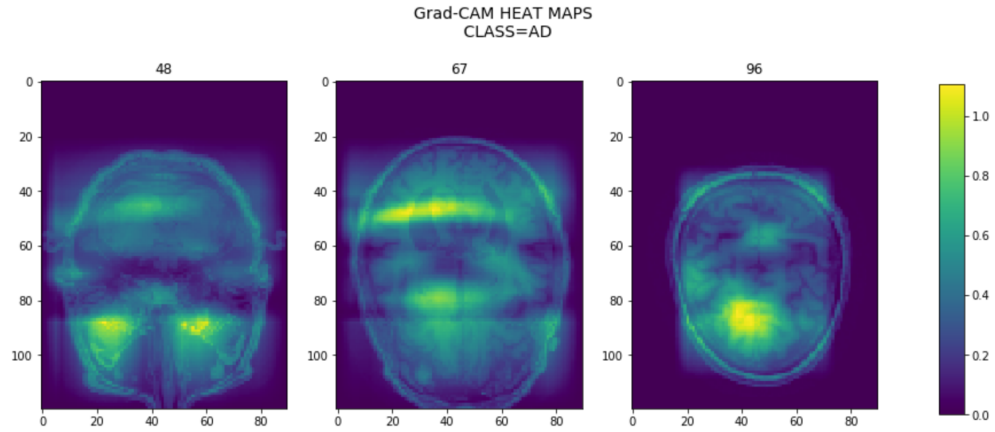


(d) NC class sample slice 88)

Figure 4.5: Heatmaps for sample images on different slices



(a)



(b)

Figure 4.6: Visualizing CNNs for MRI-based Diagnosis of Alzheimer's Disease through Grad-CAM Heatmaps averaged over test samples of (a) NC class (b) AD class.

4.2 Biomedical Image Segmentation

4.2.1 Dataset and Preprocessing

BraTS

The Brain Tumor Segmentation (BraTS) challenge focuses on the evaluation of the methods for brain tumor segmentation in multi-parametric magnetic resonance imaging (mpMRI) scans. It started in 2012 and takes place every year since then. Every year it provides a publicly available, on request, dataset of mpMRI scans for segmentation of intrinsically heterogeneous brain tumors called gliomas. BraTS later on expanded it's focus to overall survival prediction of patient, but we have focused only on the segmentation part of BraTS challenge.

In this study we have used the latest dataset available at the time from BraTS called BraTS'19. The dataset consists of mpMRI scans of 335 subjects in training set and 125 subjects in validation set, where each mpMRI consists of a) a native T1-weighted scan (T1), b) a post-contrast T1-weighted scan (T1Gd), c) a native T2-weighted scan (T2), and d) a T2 Fluid Attenuated Inversion Recovery (T2-FLAIR) scan, so in total we have 4 different kinds of MRI scans available for each case. Out of 335 subjects in training set, 259 subjects have High Grade Gliomas (HGG) and the rest 76 have Low Grade Gliomas (LGG).

The BraTS segmentation task originally focused on four glioma sub-regions in the MRI scans delineating NCR, ED, NET, and AT and were labeled as 1,2,3 and 4 respectively. But in later BraTS segmentation tasks, since BraTS'17 to the present BraTS'19 the four sub-regions were converted to three sub-regions, eliminating label 3 of NET sub-region and combining it with the label of NCR sub-region, which is label 1. Therefore, in the BraTS'19 dataset that we have used, there are three sub-regions, namely NCR, ED and AT for segmentation. Fig 4.7 gives a full overview of the tumor sub-regions in the dataset.

The standard preprocessing done by BraTS on the MRI images in the dataset includes co-registering the MRI images to a common anatomical template from [56], skull stripping and resampling to a uniform isotropic resolution of 1 mm^3.

The further preprocessing done on the dataset is the intensity normalization, N4 bias correction and downsampling of the MRI images as explained in the subsections below.

Intensity Normalization

Intensity normalization is an important data pre-processing step as MR images do not have a consistent intensity scale. There are various intensity normalization techniques to process MRIs such as Z-score normalization, Least squares (LSQ), Gaussian Mixture Model(GMM) based WM mean normalization, Tissue mean normalization. [55] showed the vital role in-

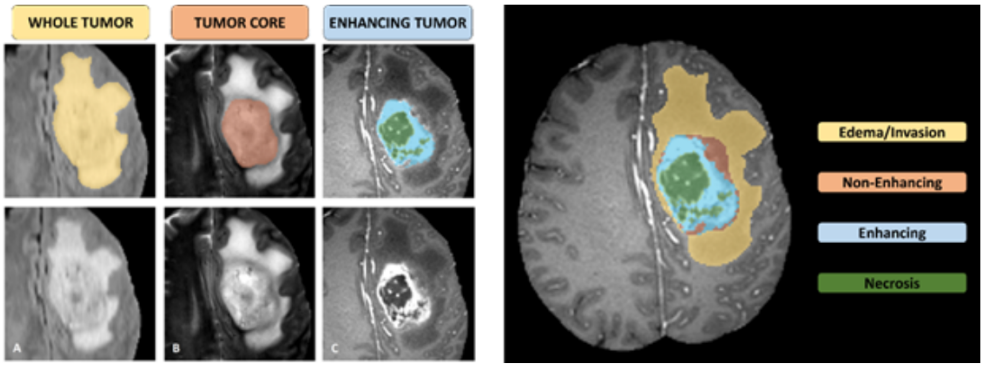


Figure 4.7: Glioma sub-regions. The image patches show from left to right: the whole tumor (yellow) visible in T2-FLAIR (A), the tumor core (red) visible in T2 (B), the active tumor structures (light blue) visible in T1Gd, surrounding the cystic/necrotic components of the core (green) (C). The segmentations are combined to generate the final labels of the tumor sub-regions (D): ED (yellow), NET (red), NCR cores (green), AT (blue). Figure taken from [45]

tensity normalization plays in image synthesis using deep learning. Their study also gives the evidence that results of the deep learning model are robust to the choice of normalization technique, as they did not see much statistical difference in the performances of the normalizing algorithms. That's why we decided to used Z-score normalization for the intensity normalization of the MRI images as a pre-processing technique, because of its simpler implementation and competency with other algorithms.

As the name suggests Z-score normalization uses the statistical measures of the MRI image data and converts the intensities of the MRI images to a common scale of zero mean and standard deviation of one. Mathematically it is given as,

$$I_{z-score}(X) = \frac{I(X) - \mu_{zs}}{\sigma_{zs}} \tag{4.3}$$

N4 bias correction

A bias field is a low frequency smooth corrupting signal present in MRI image data due to intensity non-uniformity in the magnetic fields of the MRI machines. It often results in blurred images, changed intensity value of the same tissue present in different place within an image degrading the performance of the image processing algorithms [33].

So a preprocessing step to remove this bias is a must for the optimal performance of our model. There are various techniques to remove this bias from the MRI images. Most popular among them are N3 bias correction and N4 bias correction. N3 bias correction is an intensity distribution based bias correction method. It iteratively calculates a total bias field by maximizing the high frequency component in the intensity distribution of MRI image. N4 bias correction is an improved version of N3 bias correction method, where bias correction

is computed based on the results of previous iteration. To eliminate the bias present in our data we have use N4 bias correction. Fig 4.8 shows an example of N4 bias corrected image. A bias field is difficult to be observed by human eyes but it can greatly degrade the piece-wise constant property of medical images[61].

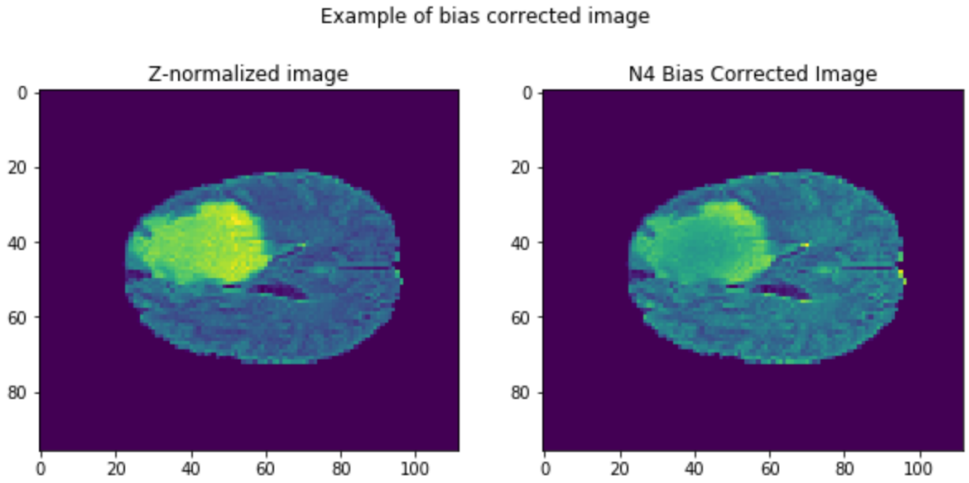


Figure 4.8: Bias Corrected Image

4.2.2 Architecture

In this section we explain three different popular deep learning models for medical image segmentation and the proposed architecture for brain tumor segmentation dataset.

U-Net

U-Net[57] was first proposed in 2015 as *"fully convolutional network"*[43]. As the name suggests, U-Net has a symmetric U shaped architecture. It consists of a contraction path(left side) to learn feature maps from the whole image for segmentation, and an expensive path (right side) to produce a full-resolution segmentation. The contraction path follows a typical 2D convolution neural network architecture, where every block consists of two 2D convolution layers with kernels of size 3×3 followed by a ReLU for feature learning and 2×2 max pooling operation with stride 2 for downsampling. The expensive path upsamples the feature maps by using up-convolution layers of size 2×2. The upsample feature maps are concatenated with the corresponding cropped feature maps from the contracting path. A 3×3 convolution is then used on the resultant feature maps. This concludes the operations performed at each step taken by the expansive path. Fig 4.9 gives a clear overview of the architecture of a U-Net model.

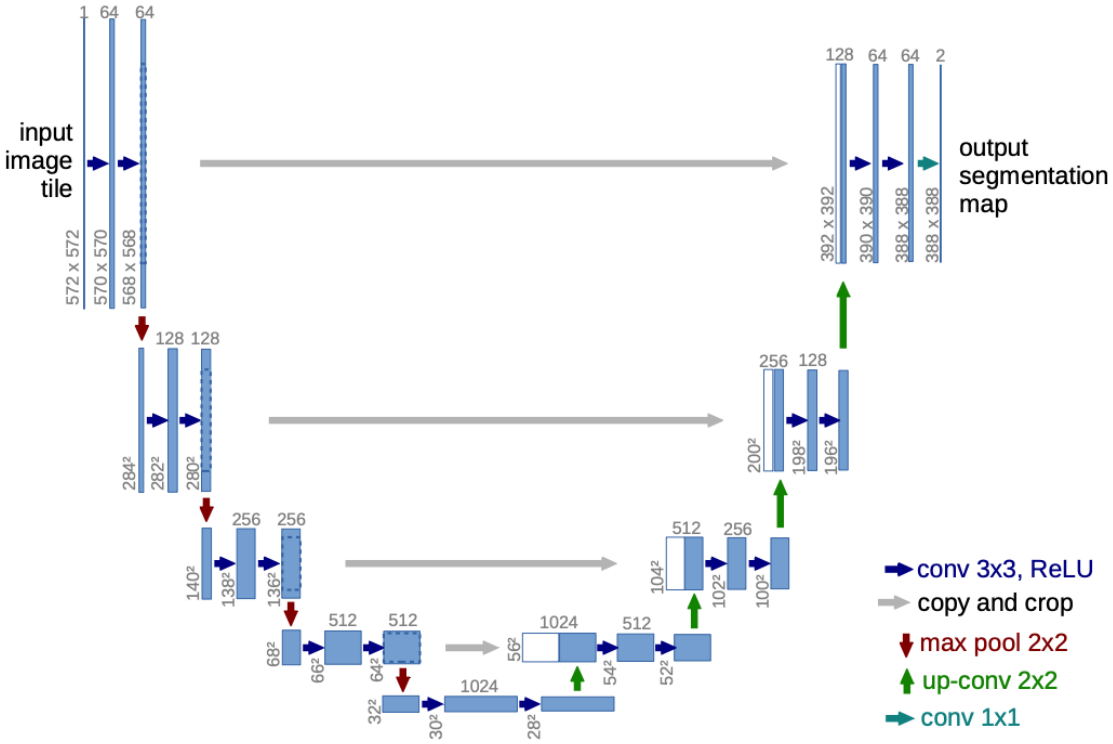


Figure 4.9: Architecture of U-net. Fig taken from [57]

The skip connections used in U-Net gives it the ability to better localize because the high-resolution features from the contracting path are combined with the upsampled output of the expansive path, yielding more precise segmentations.

3D U-Net

One drawback of U-Net is that it doesn't work with 3D data as whole. To segment 3D MRI images, we either give 2D slices or patches as input. 3D U-Net[12] overcomes this drawback. A 3D U-Net follows the same architecture as U-Net, but uses 3D convolution layers, 3D max pooling, and 3D up-convolutional layers. Fig 4.10 shows the architecture of a 3D U-Net model.

Task it was used for first?

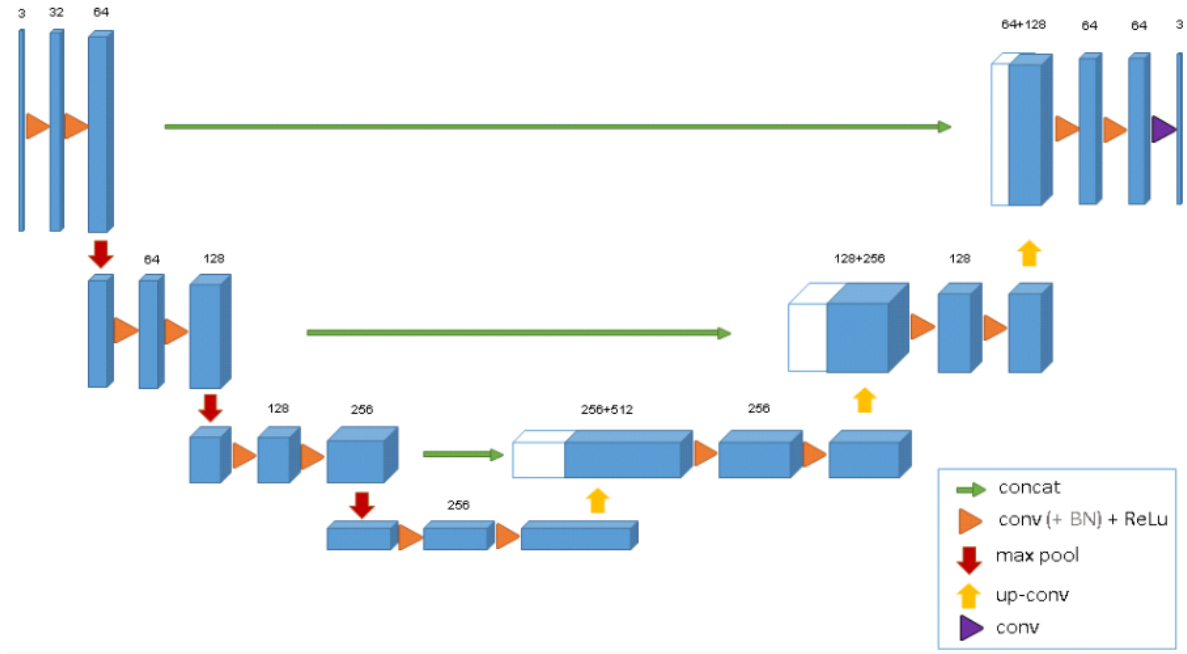


Figure 4.10: Architecture of 3D U-Net. Fig taken from [12]

V-Net

V-Net[46] was first proposed in 2016. It gained popularity because of it's efficiency and accuracy in "PROMISE 2012"[38] dataset challenge for volumetric binary segmentation of prostate gland. V-Net also uses volumetric convolutions as opposed to 2D slices or patches used by other deep learning models for medical imaging. The V-Net architecture also follows a similar contraction and expansion path like 3D U-Net but with some improvements. Fig 4.11 gives a schematic representation of the V-Net architecture. The left side like U-Net and 3D U-Net analyzes the image. Each stage consists of one to three 3D convolution layers and at each stage the network learns a residual function. A residual is defined as a combination of the input of each stage and the output of the last convolution layer of that stage processed through the non-linearities. At each stage 3D kernels of size $5 \times 5 \times 5$ are applied to the input for feature learning and convolutions with $2 \times 2 \times 2$ kernels with stride 2 are used to reduce the resolution, hence contracting.

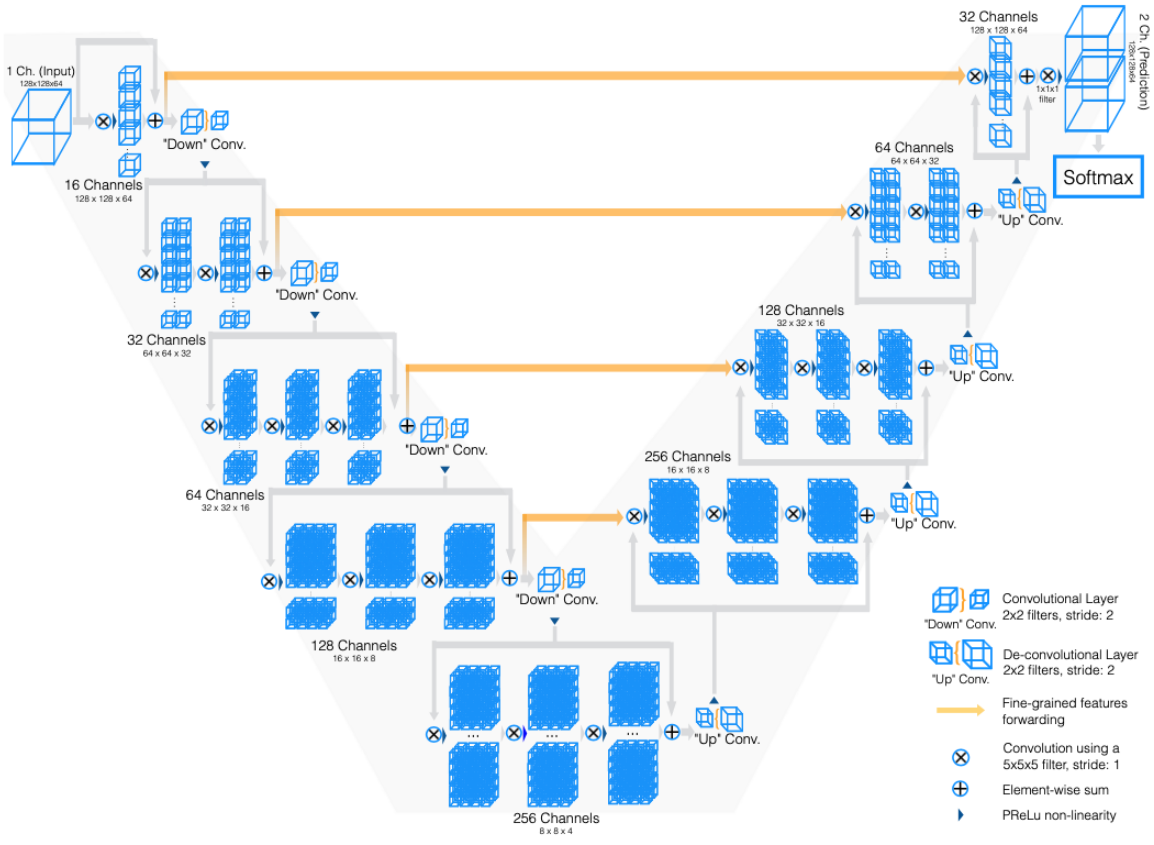


Figure 4.11: Architecture of V-Net. Fig taken from [46]

In the right part of expansion, the low resolution feature maps are expanded, gathering the necessary information to create the segmentation masks as output. The architecture of the right part is similar as left with difference of using de-convolution layers to upsample the image in place of downsampling convolution layers. A residual connection is learnt at

each step just like in the contraction path. There are several advantages of using residual connections. They propagate features from previous layers onto deeper layers that increases feature richness. Another advantage is it alleviates the problem of vanishing gradients by letting gradients pass through residual connections, improving gradient flow.
Similar to U-Net, V-Net also uses skip connections to preserve the location information of the learnt features from the contraction path, hence improving the localization.

4.2.2.1 Proposed Architecture: *Attention V-Net*

In this chapter we will discuss in depth our proposed architecture for volumetric segmentation of medical imaging. Our model is very much inspired from the V-Net architecture, leveraging the potential of end-to-end trained fully convolutional neural networks to process MRI images. All the models explained above focused on binary segmentation but in this study our focus is on segmenting multiple classes of tumor sub-regions. We take the full advantage of skip connections, residual connections and have added novel attention blocks in our model to tackle the complexity of this task. The proposed model uses volumetric convolutions to process MRI images.

The model comprises of an analysis path (contraction) and a synthesis path (expansion). In the analysis path there are 4 down-transition stages (including one input transition stage) where at each stage downsampling and convolutions happen on the input. In the synthesis path there are 5 up-transition stages.

At each stage in the analysis path, there are 3D convolutional layers sequentially increasing from one to four. The input is passed to these convolution layers, followed by ReLU6 non-linearity and then a batch normalization layer. Each convolution layer uses 3D kernels with size $3 \times 3 \times 3$. A residual function is learnt by adding the input of each stage with output of the last convolution layer of that stage. At the end of each analysis stage a downsampling operation is performed by convolution with $2 \times 2 \times 2$ voxels wide kernels applied with stride 2 to reduce the resolution of the feature maps. Downsampling helps helps to increase the receptive field. Each stage outputs double the number of feature maps with half of the input resolution. For downsampling use of convolution layers with strides, over pooling layers is inspired from [46]. As suggested in [46] switching pooling layers to convolution layers also results in a smaller memory footprint.

The right part of the model called synthesis path decompresses the signal until the original size is reached. At each stage in the synthesis path, the number of 3D convolution layers decrease from five to one sequentially, making it a mirror image of the analysis path. At each stage the input is upsampled using de-convolution layers to increase the size of the feature maps. Inspired from [57] and [12] we have implemented horizontal skip connections in our model as well. The output of the respective skip connection and the output of the last stage are used to create an attention block which is explained in more details below. The output of this attention block is then concatenated with the output of the last stage again. This is treated as the input for the current stage in the synthesis path. This input is then processed by the convolution layers present in the block followed by ReLU6 layer again. The convolution layers use 3D kernels of size $3 \times 3 \times 3$. The number of kernels in each convolution layer at a stage is half the number of kernels in the convolution layers present in the last stage. The learnt residual is added to this output similar to the analysis block. This process is repeated at each stage. In the final output stage, the last convolution layer computes the four feature maps, one for the background and the rest three for the sub-regions of the tumor. In this way the the synthesis path gathers the necessary information and assembles

it produce the final output of four channel volumetric segmentation of the tumors. Fig 4.12 presents our proposed model.

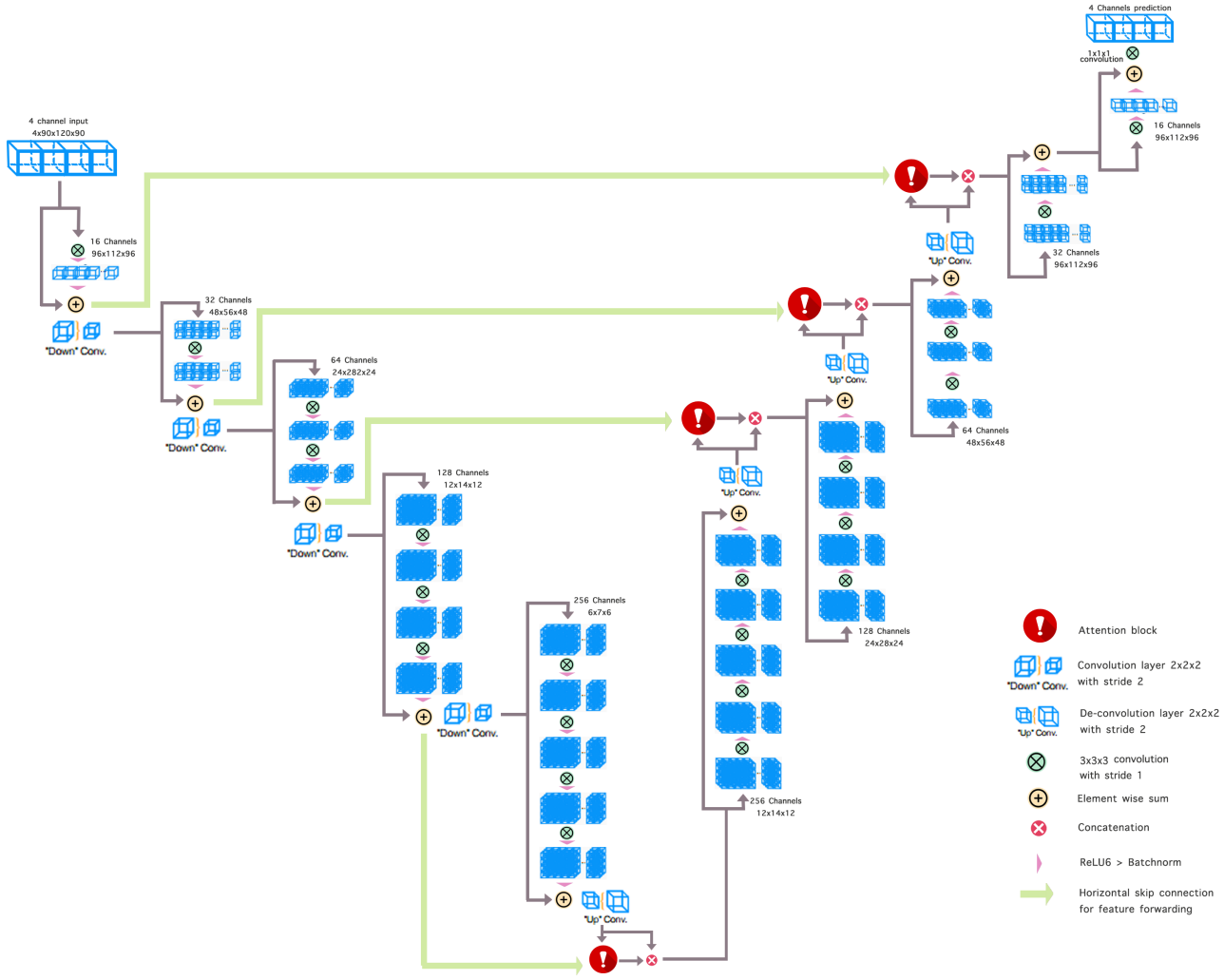


Figure 4.12: Proposed Attention VNet

1. **Horizontal Skip Connections and Residual Connections**

   Skip connections as the name suggests skip some connections in the network and feeds or adds the output of one layer to the output of the connecting layer in the skip connection. In our model, we have used two kinds of skip connection, a) a horizontal skip connection and b) a short residual connection.

   In our model the horizontal skip connection helps in the forwarding the extracted features from the analysis path to the synthesis path as shown in fig 4.13. The output from each stage to analysis goes to the next stage, but along with that it also goes to the attention block in the corresponding synthesis stage. In this way it helps our attention block to focus on the localization of the features.

   In each stage of both analysis and synthesis path, a residual connection[23] is also present. A residual connection is a kind of skip connection which allows smooth information flow from one layer to another by bypassing some layers in deep neural

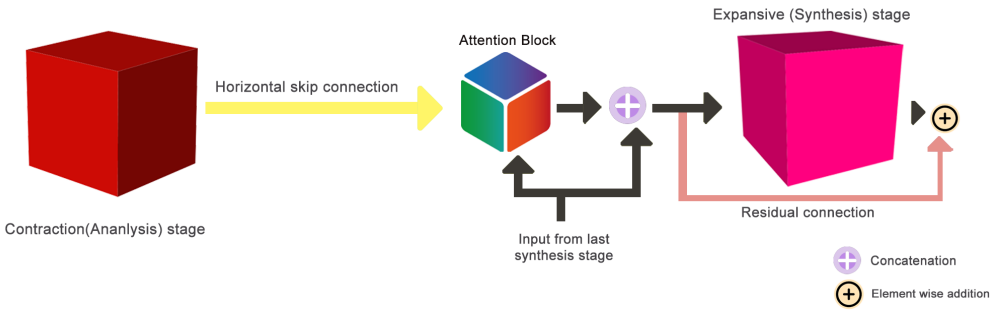


Figure 4.13: Horizontal Skip Connection

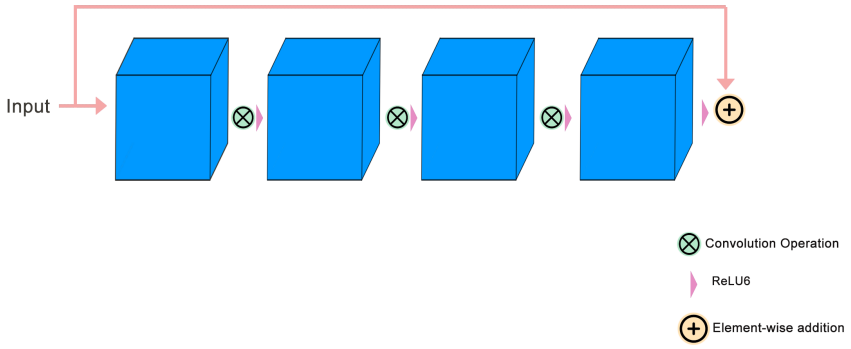


Figure 4.14: Residual Connection

networks. A tradition deep learning model learns a mapping M, from an input x to output y.

$$M(x) = y \tag{4.4}$$

A residual connection tries to learn the difference between a mapping applied to x and the original input x, given by

$$R(x) = M(x) - x \tag{4.5}$$

Transforming eq 4.5, we get the mapping from a residual connection as

$$M(x) = R(x) + x \tag{4.6}$$

Fig 4.14 gives the overview of residual connections from one of the blocks of our network. Since our model is a deep 3D CNN, it is prone to suffer from vanishing gradient problem during backpropagation. Residual connections helps in avoiding this problem because of skipping trait as mentioned earlier.

2. **Attention Blocks**

   The attention block in our model is used to highlight the salient features transferred

through the horizontal skip connection and remove the irrelevant and noisy responses. Fig gives an outline of the attention blocks we have used in our model.

The attention block in our model takes two inputs: 1) from the horizontal skip

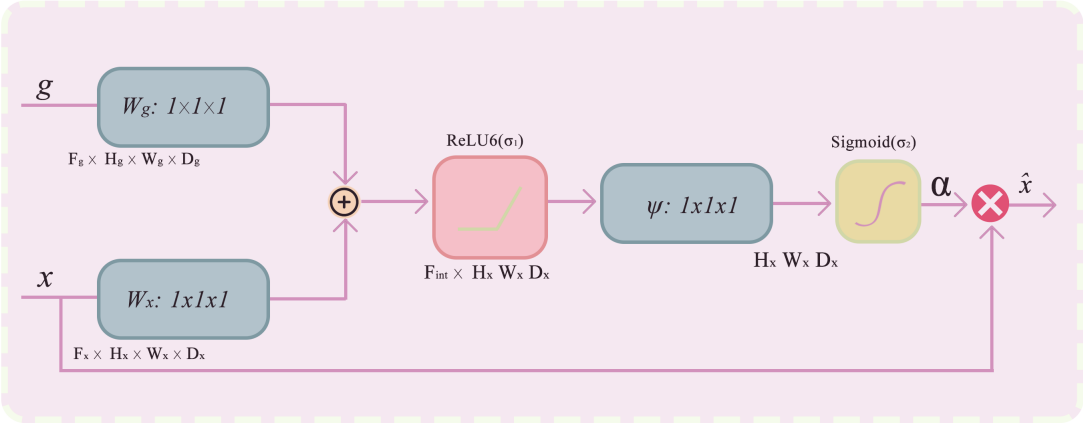


Figure 4.15: Attention Block

connection coming from the corresponding analysis block of the contraction path as mentioned above and 2) from the upsampled output of the previous synthesis block in the expansive path. In the attention block, individual convolutions of $1 \times 1 \times 1$ is applied on both the inputs. We have then applied addition operation on these two vectors following the norm of additive attention, and then passed it through ReLU6 layer to add non-linearity. Then one more $1 \times 1 \times 1$ convolution operation is performed on the resultant with sigmoid activation applied, creating a voxel-wise mask. We then multiply this result with the 2^{nd} input of this attention block. This is the final output of this attention block which is then concatenated with the output of the current stage in the synthesis block, and is passed to the next stage as input and as well as a residual connection. This motivation of this attention block is taken from [51] with some minor changes. Mathematically the attention block is given as:

$$A_{att} = \psi^T(\sigma_1(W_x^T x_i + W_g^T g_i + b_x)) + b_\psi \tag{4.7}$$

$$\alpha_i = \sigma_2(A_{att}(x_i, g_i; \theta_{att})), \tag{4.8}$$

where σ_1 is the *ReLU6* non-linearity in our block, g_i and x_i are the two inputs to the attention blocks, and W_g and W_x are the corresponding convolution layers as explained above. ψ is the last $1 \times 1 \times 1$ convolution transformation followed by the sigmoid activation function (σ_2). α_i is called the activation coefficient. θ_{att} is the set of parameters containing, W_x, W_g, ψ and bias terms b_x, b_ψ

4.2.3 Training and Experiments

The model was trained with 335 MRI images collected from BraTS'19, resized to $96 \times 112 \times 96$ and spatial resolution of $1 \times 1 \times 1\ mm^3$ in axial view. The input to the model consists of the four MRIs stacked upon each other (T1, T1Gd, T2 and T2-FLAIR), resulting in a input of shape $4 \times 96 \times 112 \times 96$. The model takes a batch size of 12. Like most of the medical segmentation task, we also faced the problem of highly unbalanced classification of the labels in our dataset. To get around this problem we experimented with weighted cross-entropy and dice loss functions. In the final experiments we have used the compound loss which a combination of weighted cross-entropy and dice loss. Cross-entropy is a distribution-based loss and dice loss is a region based loss, so combining these two losses we can improve both classification and localization of the labels in our dataset. Weighted cross-entropy is defined as

$$_{W}\text{CE loss}(x, class) = w[class] \left(-x[class] + log \left(\sum_{j} exp(x[j]) \right) \right) \tag{4.9}$$

and Dice loss tries to optimize dice coefficient is given as:

$$D = 1 - \frac{2 \sum_{i}^{N} p_i g_i}{\sum_{i}^{N} p_i^2 + \sum_{i}^{N} g_i^2} \tag{4.10}$$

So the resultant loss we are trying to optimize is

$$loss =_{W} \text{CE loss} + D \tag{4.11}$$

Like classification task, we have initialized our segmentation model with He initialization too. We have trained our model for 10,000 epochs with an initial learning rate of 0.01 with a scheduled drop of factor 0.1 after every 2000 epochs. To optimize the loss we have used Adam optimizer. ReLU6 is used as the activation function with all the convolution layers. A dropout of factor 0.2 is used with all the convolution layers. Nearest interpolation is used to resample the segmented labels to the original size.
Since the labels in the dataset are highly unbalance, the weights we have assigned to the weighted cross-entropy are

$$w(l) = \frac{1}{frequency(l)}, \quad where \ \ l \in label \tag{4.12}$$

in this way, the label which is underrepresented, will be penalized more if classified incorrectly. The distribution of each label in the whole training dataset is given in tab 4.4

Background (0)	NCR(1)	ED(2)	AT(4)
0.989	0.0024	0.0063	0.0020

Table 4.4: Label Distribution

We first trained the model with weighted cross-entropy loss only and then we experimented further with the compound loss of weighted cross-entropy and dice loss. We saw significant

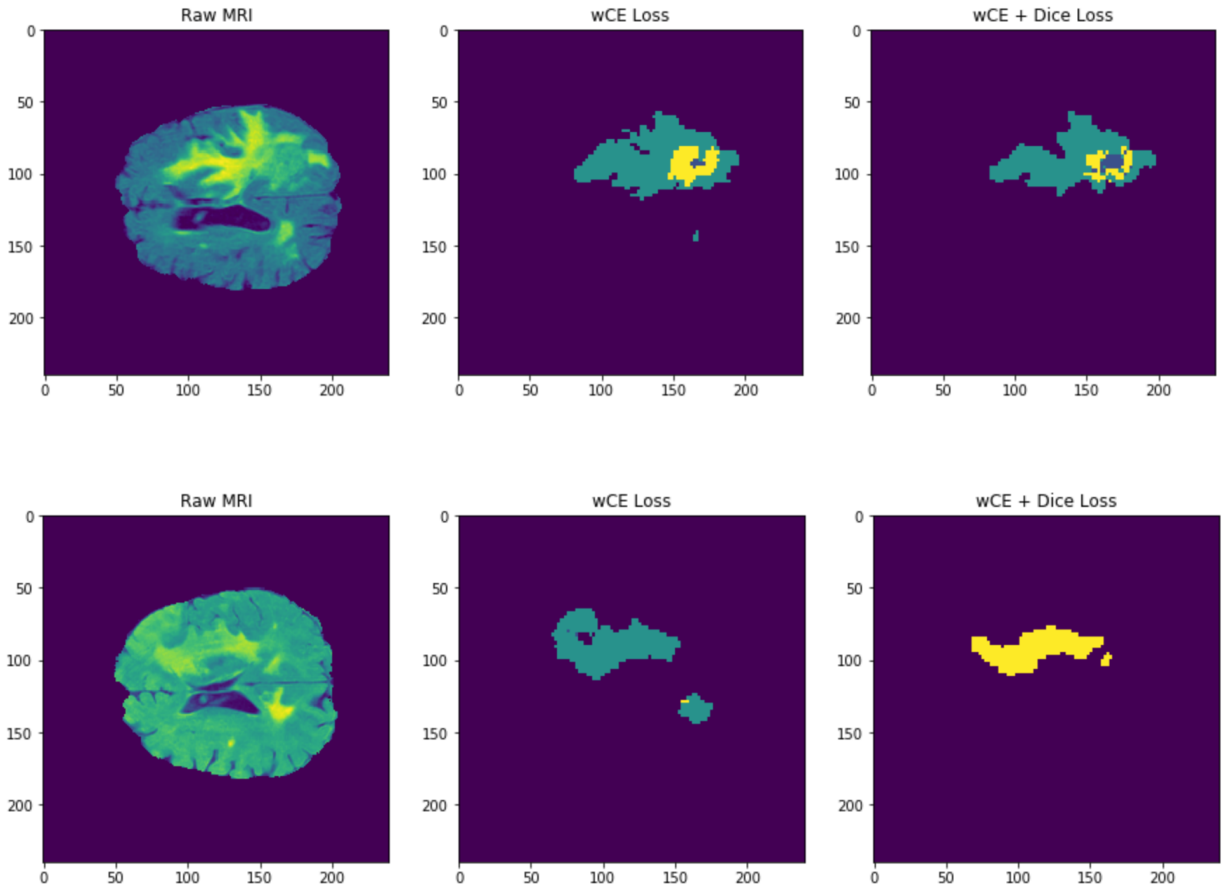


Figure 4.16: Performance of CE vs CE+Dice

improvement in the segmented results especially around the boundary areas of the labels. We have presented some examples showing the difference in the segmentation quality of tumor in Fig 4.16

Along with experimenting with our model, we also performed some experiments with 3D U-Net and V-Net to compare our model performance. We trained all three model with the same parameter settings. We limited the number the training epochs to 1000 only. This was done due to time and memory constraints. We calculated the dice score for each of the label. The comparison results are presented in Fig 5.9 in Chapter 5

Results

Chapter 5

Results

For detecting Alzheimer, a more wholistic approach is needed and the entire brain is examined as a whole, but for brain tumors, the research focuses on pinpointing where the tumors are which makes segmentation most relevant. In this chapter we have presented results of our experiments for the two tasks. Section 5.1 demonstrate the results for classification task where we have provided more heatmaps for better model visualization and have presented the comparison and evaluation matrix for our model. In section 5.2, we have produced the results for tumor segmentation task, where we have given some examples of the segmented labels predicted by our model and have provided the evaluation matrix.

5.1 Classification Task

To evaluate the performance and validity of model we have used accuracy (Acc), precision and recall, and F_2 score and have plotted the Grad-CAM heatmaps for both the Alzheimer's (AD) and NC classes. Fig 5.1 and Fig 5.2 show the every fourth slice out of 130 slices of the heatmaps averaged over the test samples of NC and AD classes respectively. The results shown are here are plotted with different cmap called *rainbow* for better visibility.

For evaluation matrix, following formulas are used.

$$Precision = TP/(TP + FP) \tag{5.1}$$

$$Recall = TP/(TP + FN \tag{5.2}$$

$$F_2 = (5 \times Precision \times Recall)/(4Precision + Recall \tag{5.3}$$

where, TP = True Positive, FP = False Positive, TN = True Negative and FN = False Negative

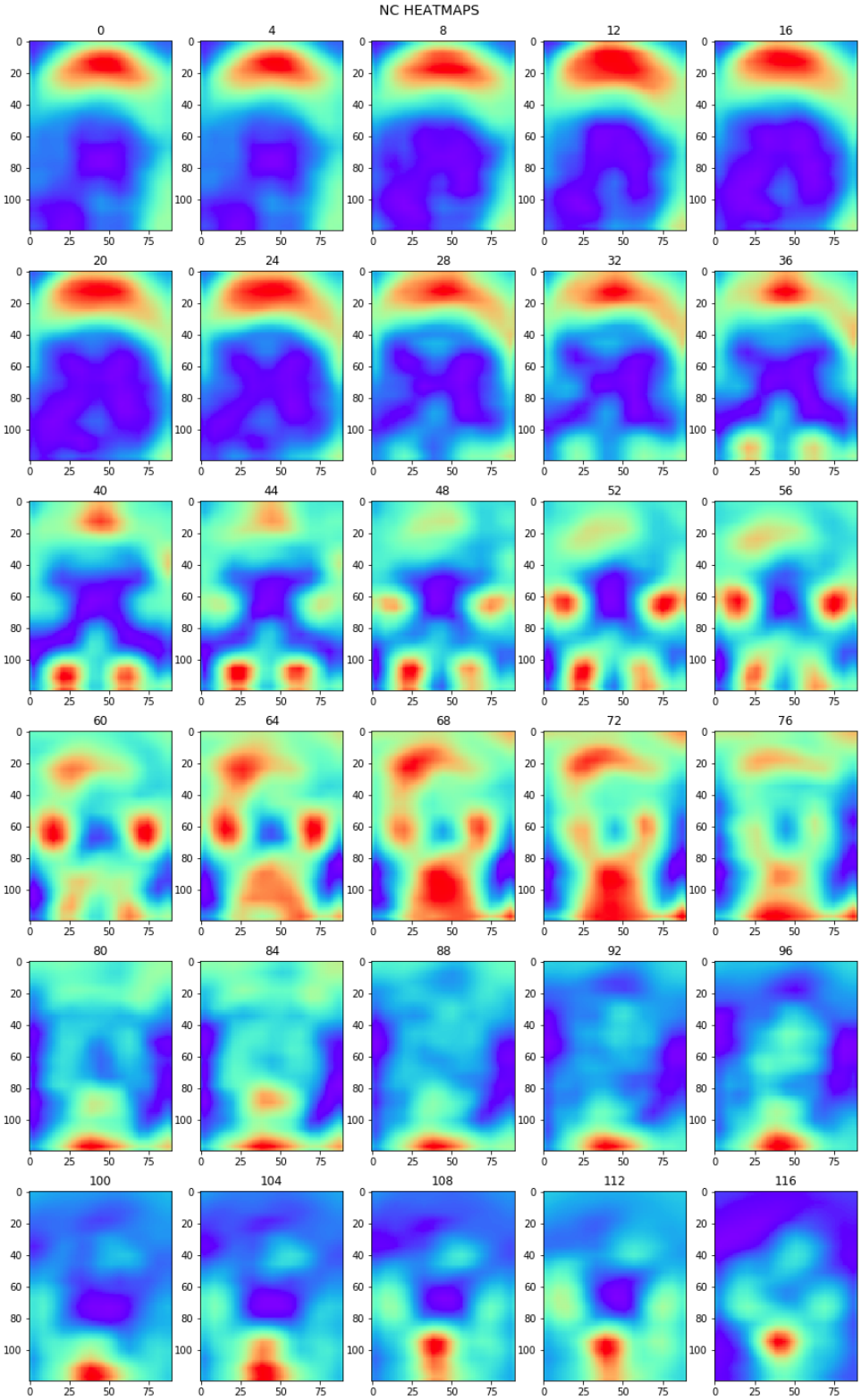


Figure 5.1: Every fourth Grad-CAM Heatmap averaged over test samples of NC class

Table 5.1 presents the evaluation matrix for the three experimented models comparing the effect of batch-norm and *l2* regularization (Last row represents the final model), averaged over 10 folds for binary and multiclass classification. We thoroughly tested and compared the performance of our model against other Deep Learning based Alzheimer's Detection models. Table 5.2 presents the comparison of our model with other studies. As mentioned earlier we

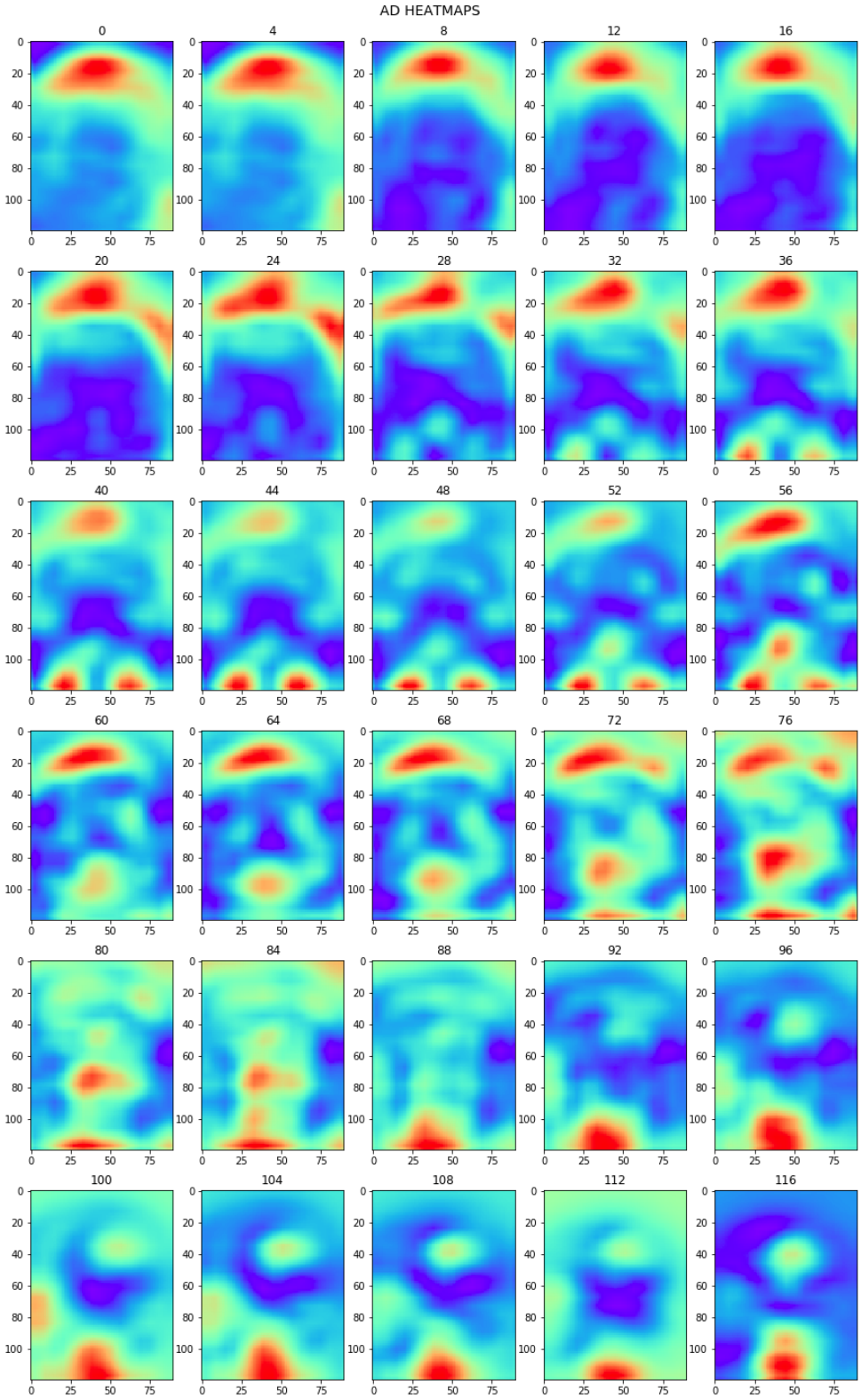


Figure 5.2: Every fourth Grad-CAM Heatmap averaged over test samples of AD class.

have trained our model from scratch and didn't use any kind of transfer learning because unlike other studies because of the fact that pre-trained models like ImageNet don't contain relevent information for transfer learning of medical data classification. Some other studies have used their pre-trained models from binary classification (AD vs NC) task to initialize

Table 5.1: Evaluation Matrix for the Deep 3D CNN Model(L2: L2 regularization, BN: Batch Normalization. Last row represents the configuration of our model)

Model	L2	BN	Binary Class				Multi Class			
			Acc	F_2	Pre	Recall	Acc	F_2	Pre	Recall
Deep 3D CNN	✓	✓	91.11	0.90	0.89	0.94	86.22	0.85	0.88	0.87
Deep 3D CNN	✓	✗	91.20	0.91	0.91	0.92	86.70	0.86	0.84	0.85
Deep 3D CNN	✗	✓	94.17	0.94	0.94	0.94	89.14	0.88	89.21	88.1

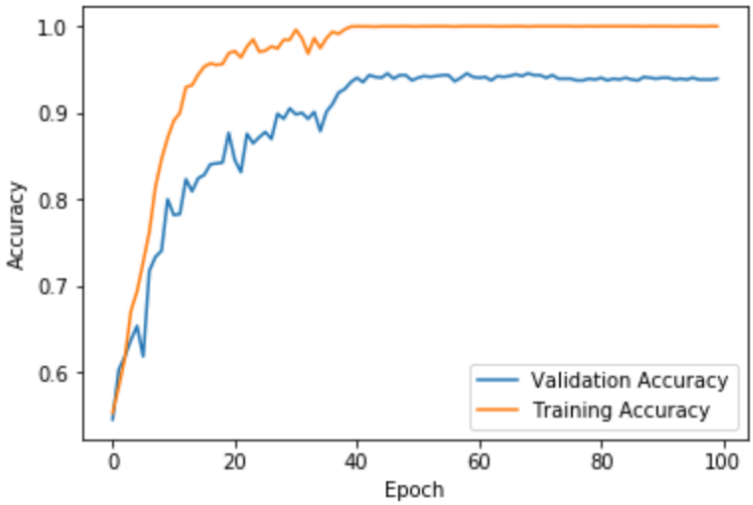
their model for multi-class classification (AD vs NC vs MCI). This approach is very prune to data leakage in case the same data is present for present for some classes in the next classification task. As can be seen from Table 5.2 some studies have used PET scans in combination with MRI scans, and is a very good technique to get more information about the individual data points, our model still outperform them while working with only single modality (MRI only). The training vs validation accuracy and training vs validation loss

Table 5.2: Comparison with previous studies of Alzheimer's Detection using Deep Learning on ADNI Data. Accuracy metrics is used for comparison and represents binary classification between Alzheimer's Disease (AD) vs Normal Cognition(NC) and multi-class classification between Alzheimer's Disease (AD) vs Normal Cognitive(NC) Mild cognitive impairment(MCI)

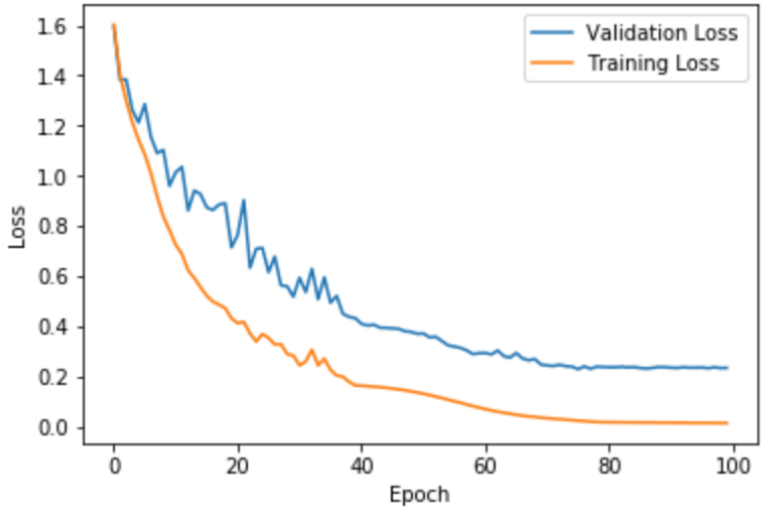
Method	Modality	Dataset Size	Accuracy	
			AD vs NC	AD vs MCI vs NC
[40]	FDG-PET	ADNI (339)	91.20	-
[63]	PET + MRI	ADNI (317)	91.10	-
[41]	MRI+PET	ADNI (331)	91.40	53.84
[3]	MRI	ADNI (815)	91.41	-
[15]	MRI	ADNI (841)	94.10	61.10
Our Approach	MRI	ADNI (817)	94.17	89.14

averaged over 10 folds in presented in Fig 5.3 for binary classification. The smoothing of the graphs can be noticed at every 40^{th} epoch as a result of learning rate decay. The average training accuracy for all the 10 folds in 100% but we can see the average validation accuracy for all the folds is around 94.17% in binary classification. This is due to slight variance in the learned weights across the 10 folds.

Graphs in fig 5.4 depicts the 10-fold average accuracy and loss between training and validation sets for multiclass classification. However we see during multiclass classification a bit of overfitting in our model as the complexity of the task of multiclass classification is higher than that of binary classification.

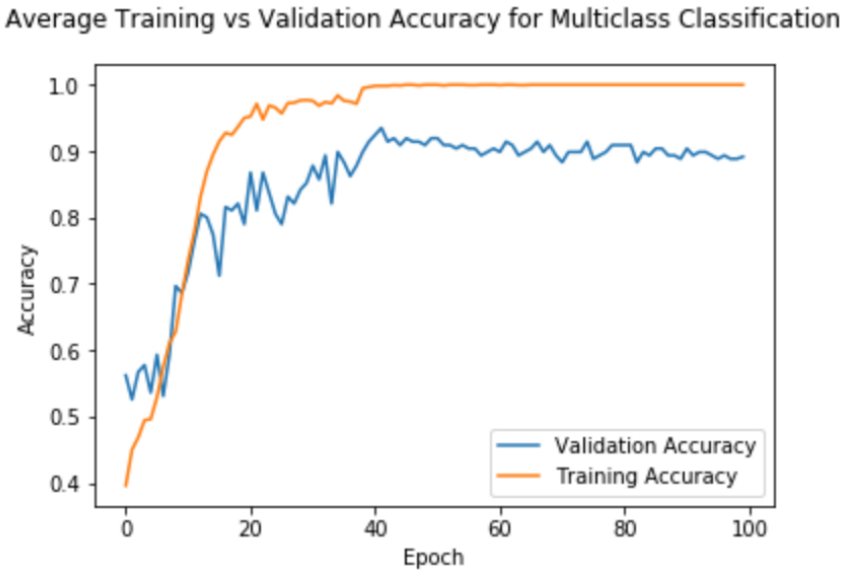


(a)

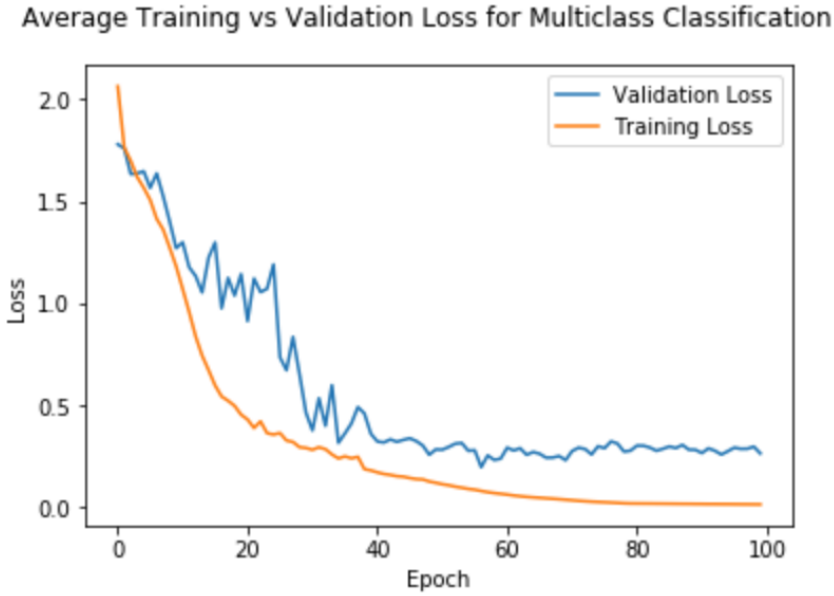


(b)

Figure 5.3: (a) Training vs Validation accuracy (b) Training vs Validation loss for Binary Classification



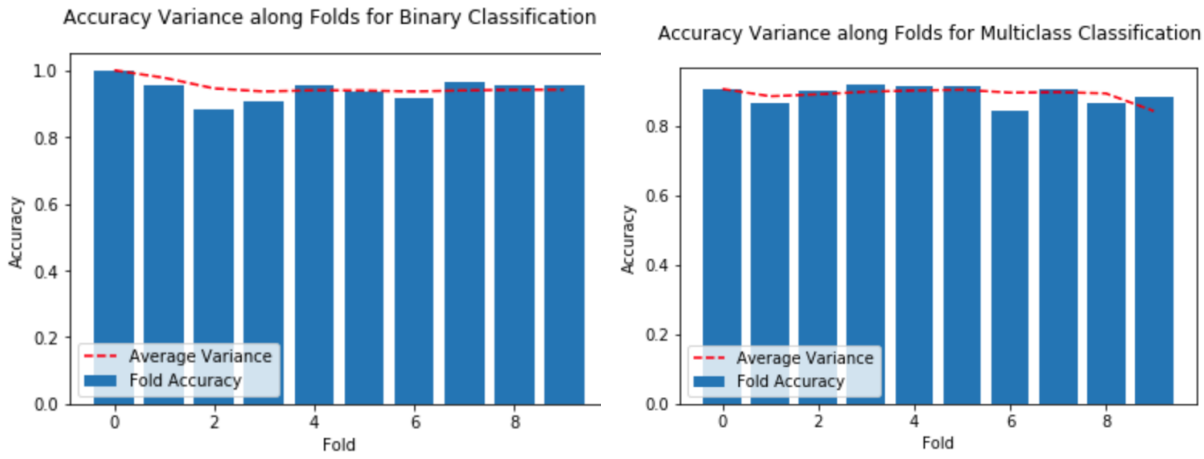
(a)



(b)

Figure 5.4: (a) Training vs Validation accuracy (b) Training vs Validation loss for Multi-class Classification

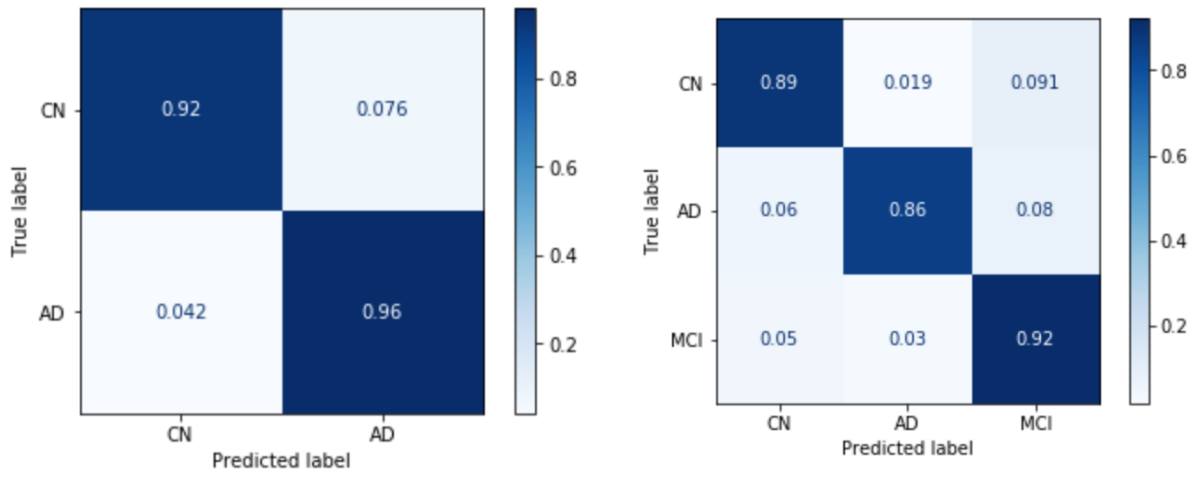
We have presented the average variance in the validation accuracy of the model across all 10 folds for binary classification and multi-class classification in fig 5.5



(a) (b)

Figure 5.5: Average variance in validation accuracy across folds in (a) Binary Classification (b)
Multi-class Classification

The confusion matrix in fig 5.6 for binary and multiclass classification details us with the class-wise performance of our 3D CNN model. The best fold in the binary classification have a very low false negative rate for of 0.016% for AD class and 0.06 in multiclass-classification. The average false negative rate is still very low as can be seen in the confusion matrices.



(a) (b)

Figure 5.6: Confusion Matrix for Binary Classification (a) and Multi-class Classification (b)

5.2 Segmentation

To evaluate the performance of our segmentation model, we have used the validation set of 125 MRIs from BraTS. The ground truths of the validation set are not made available by BraTS. The segmented labels have to uploaded on the challenge website to get the segmented results.

5.2.1 The Evaluation Criteria

As stated earlier the dataset contains three sub-regions of tumors: NCR (Label 1), ED (Label 2) and AT (Label 4). To evaluate the results BraTS has distribute the segmented labels in three classes as follows:

1. **WT: Whole Tumor Extent.** WT class represents the whole tumor and is given by the union of all labels.
2. **TC: Tumor Core.** TC class represents the segmentation of the tumor core outline. It is the union of label 1 and 4
3. **ET: Active/enhancing and the non-enhancing/necrotic tumor regions.** This class is represented by label 4.

The evaluation matrix used by BraTS'19 is class wise DICE score, Sensitivity, Specificity and Hausdorff distance (95%) for the classes mentioned above. Originally a DICE coefficient for two sets X and Y is give as:

$$DSC = \frac{2\,|X \bigcap Y|}{|X| + |Y|} \tag{5.4}$$

Dice score measures the area of overlap between the ground truth and the predicted label. It is similar to F1 score and using the definition of true positive (TP), false positive (FP), and false negative (FN), it can be written as:

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{5.5}$$

Hausdorff distance is a surface distance measure. It measure the distance between two boundaries, in our case boundary of prediction and ground truth segment. A bidirectional Hausdorff distance between two sets X and Y is given as:

$$HD(X, Y) = max(hd(X, Y), hd(Y, X)) \tag{5.6}$$

where,

$$hd(X, Y) = \underset{x \in X}{max}\ \underset{y \in Y}{min} \|x - y\|_2 \tag{5.7}$$

Sensitivity is defined as the number of actual positive cases out of all the cases that are predicted as positive. It is also called Recall and is mathematically defined in sec 5.1, whereas specificity is the proportion of actual negatives, which got predicted as the negative and is given as

$$Specificity = (TrueNegative)/(TrueNegative + FalsePositive) \tag{5.8}$$

5.2.2 Evaluation

We have presented some examples to visualize the performance of our model in Fig 5.7 and Fig 5.8. In Fig 5.7 some examples from training dataset are given showing a raw fMRI image, the corresponding ground truth provided by BraTS and the labels segmented by our model. The official results of the evaluation matrix for our segmented labels received from BraTS are mention in Table 5.3

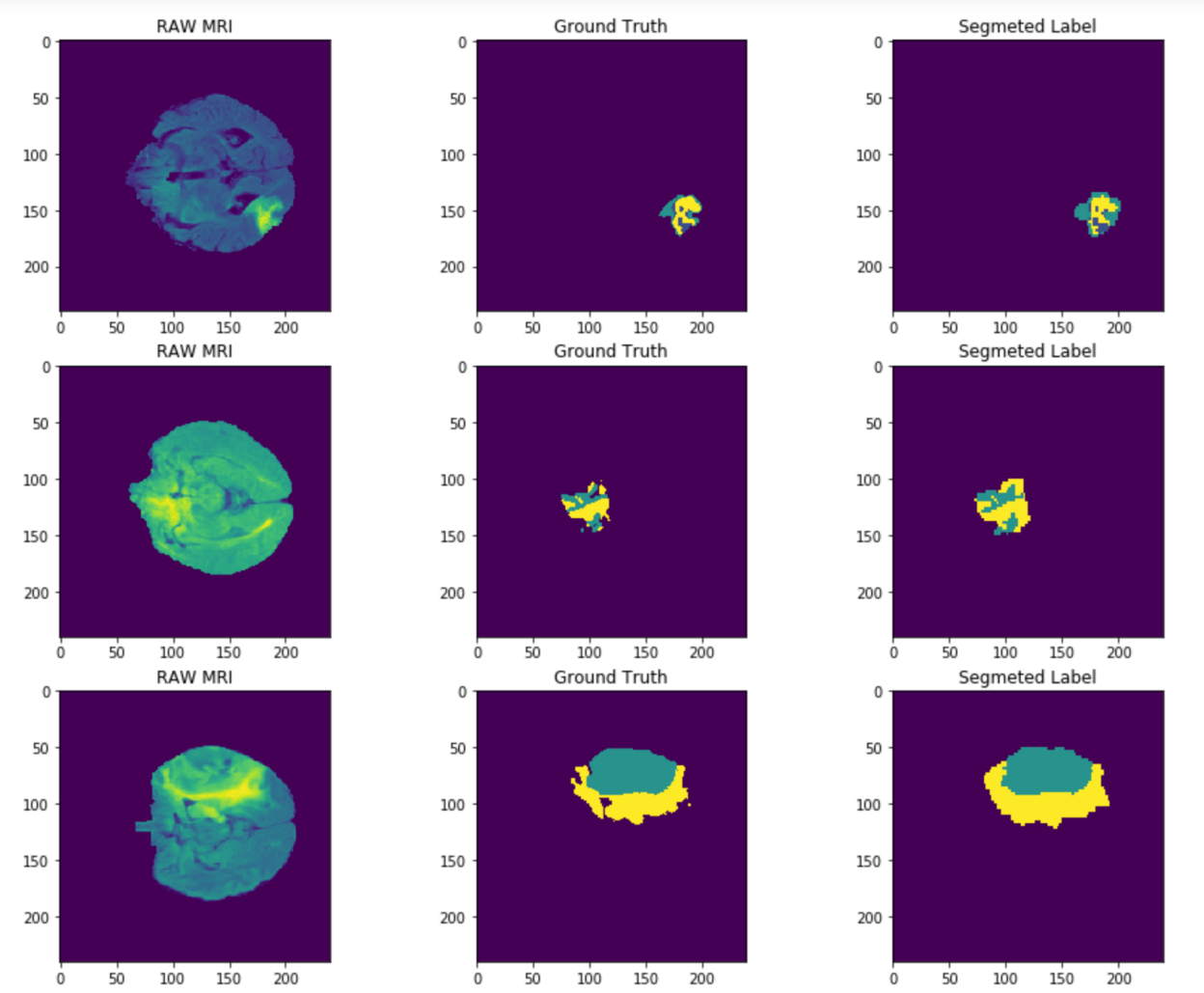


Figure 5.7: Examples from training dataset segmented by our model

Fig 5.8 presents some samples of the segmented labels from validation dataset segmented by our model. As can be seen from both training and validation sample segmented results, the model is able to detect the tumor sub-regions very well especially the whole tumor area. The model pays deep attention on both the shape and location of the tumor regions. We did a thorough analysis of the segmented labels for the localization of the whole tumor from the training dataset (since ground truths aren't available for validation data) and found that the model can effectively localize the tumor. The dice score of whole tumor is 0.843 as can be seen in table 5.3 validates this.

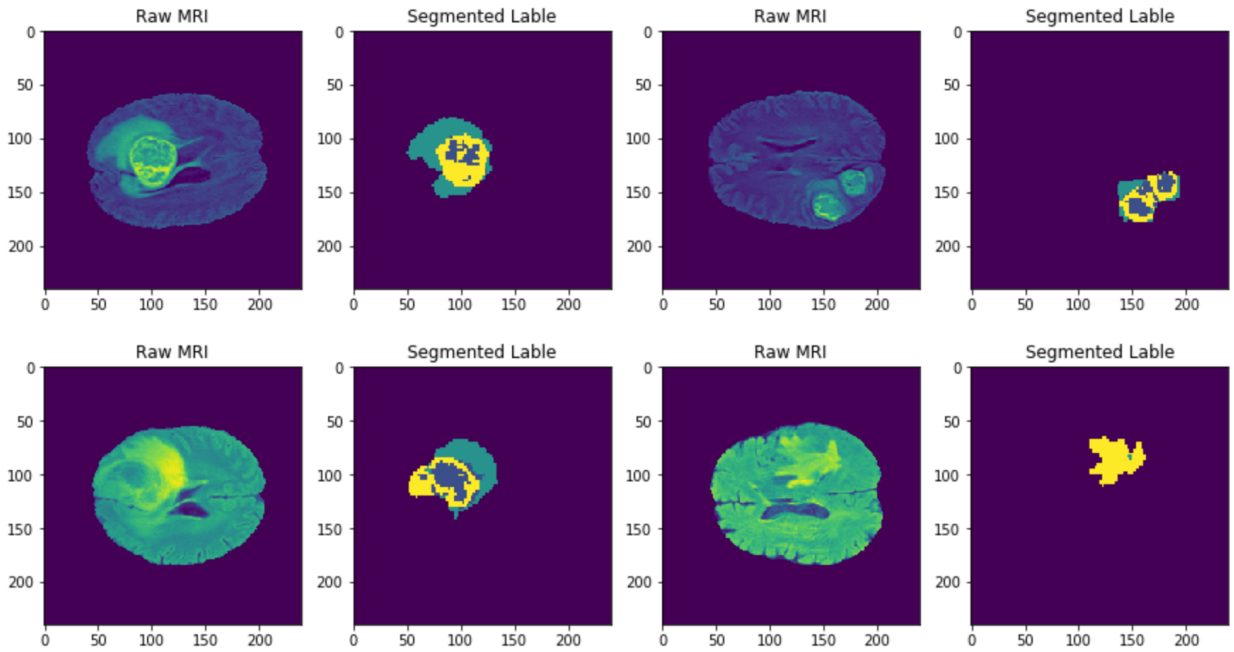


Figure 5.8: Examples from validation dataset segmented by our model

In the fig 5.9 below, we have compared our model with V-Net and 3D U-Net when run with the same setting for 1000 epochs. We had to limit this comparison experiment to 1000 epochs because of memory and time constrains. When run under same environment we can see clearly that our model converges to a much better accuracy compared to the other two models.

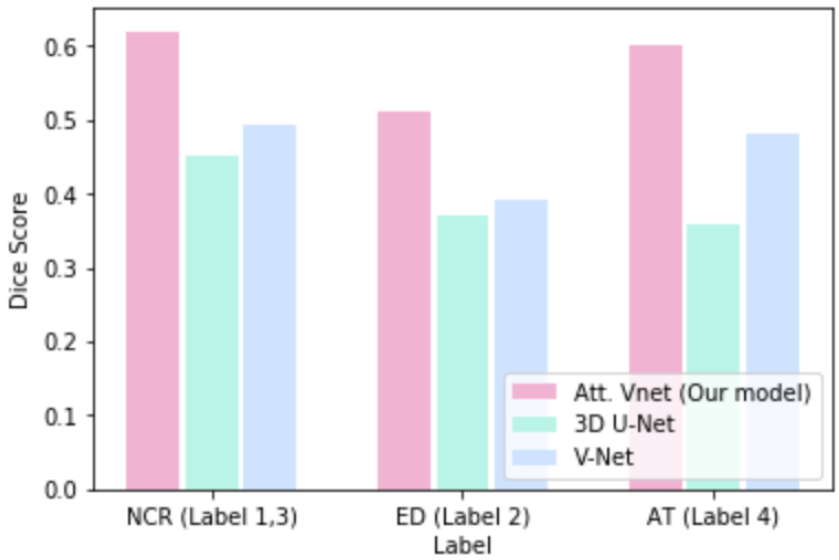


Figure 5.9: Performance comparison on training data for different labels when trained for 1000 epochs

Since the ground truth were not available, segmented labels for the validation set were uploaded on the BraTS website, and were evaluated on their server using the matrices defined above. The official results of our segmented labels received from BraTS are mention in Table 5.4

Measure	WT	TC	ET
Dice	0.843	0.808	0.910
Hausdorff95	4.064	1.892	1.816
Sensitivity	0.996	0.941	0.848
Specificity	0.976	0.996	0.997

Table 5.3: Evaluation Matrices for training data

From these two tables we observed that the model obtains good performance for the segmentation of the whole tumor area but is overfitting for the other two classes, especially ET. This is because of the highly unbalanced number of the classes.

Measure	WT	TC	ET
Dice	0.800	0.639	0.536
Hausdorff95	15.04	20.06	16.23
Sensitivity	0.854	0.621	0.586
Specificity	0.985	0.996	0.996

Table 5.4: Evaluation Matrices for validation data

[0]The results are publically available on BraTS official evaluation website https://www.cbica.upenn.edu/BraTS19/lboardValidation.html under team name CAIR

Chapter 6

Conclusions and Future Work

In this thesis we proposed a novel 3D CNN based architectures for classification of Alzheimer's Detection and segmentation of brain tumors. The classification model showed the ability to learn the relevant features on its own without the use of any transfer learning. From the results of multi-class classification it can be observed that the model has the potential of distinguishing between two very similar looking classes, in our case AD and MCI, and MCI and NC, with a very high confidence. We showed that proper initialization of model and fine parameter tuning leads to superior results. In the segmentation model, the model is able to predict the segmentation mask with good accuracy for the whole tumor. But we saw it overfitting for the other two sub-regions of tumor, named enhancing tumor (ET) and tumor core (CT). This is because of highly unbalanced label classes as these classes contain only 22% and 18% of the whole tumor. The model segmented labels for these classes with high accuracy in training dataset which shows the models ability to learnt these small patterns but lacks in generalization. The future work involves improving the models generalization ability with the use of better data augmentation techniques to increase the size of the training data, hence increasing the times an under-represented class label appears during training. Experimenting with multiple modalities and exploring other feature extraction and feature selection algorithms so that the models will be able to learn more with small amount of data for both Alzheimer's classification and brain tumor segmentation. The models can be improved in further studies by carrying out more sophisticated procedure for feature selection after feature extraction as just mentioned, and using deeply supervised learning. The next step could be to segment the biomarkers of Alzheimer's in the brain, and predicting the pattern of growth of the disease in a very early stage. For segmentation of brain tumors, we would like to contribute towards the new BraTS challenge for 2020 by improving our segmentation model to generalize better. We would also like to explore the machine learning models on MCI subjects to predict who might develop AD at a later stage.

Appendix A

Code

The code is written in python and the full code is made available at https://github.com/turtIe/Master-s-Thesis The library dependencies can also be found under the provided link. Following the are the snippets of the code for the models

A.1 Proposed Classification model

The code is written using Keras[11] library

```
def Create_model():
    model = Sequential()
    model.add(Conv3D(32,(7,7,7), kernel_initializer=keras.initializers.
    he_normal(seed=None) , input_shape=(240, 256, 160,1), activation='
    linear'))
    model.add(LeakyReLU(alpha=0.1))
    model.add(BatchNormalization())
    model.add(MaxPooling3D((2, 2,2),padding='same'))
    model.add(Conv3D(32,(5,5,5), activation='linear'))
    model.add(LeakyReLU(alpha=0.1))
    model.add(BatchNormalization())
    model.add(MaxPooling3D(pool_size=(2, 2,2),padding='same'))
    model.add(Conv3D(64,(3,3,3),activation='linear'))
    model.add(LeakyReLU(alpha=0.1))
    #model.add(BatchNormalization())
    model.add(MaxPooling3D(pool_size=(2, 2,2),padding='same'))
    model.add(Conv3D(64,(3,3,3), activation='linear'))
    model.add(LeakyReLU(alpha=0.1))
    model.add(BatchNormalization())
    model.add(MaxPooling3D(pool_size=(2, 2,2),padding='same'))
    model.add(Flatten())
    model.add(Dense(256, activation='linear'))
    model.add(LeakyReLU(alpha=0.1))
    model.add(Dense(128,activation='linear'))
    model.add(LeakyReLU(alpha=0.1))
#    model.add(Dropout(0.2))
    model.add(Dense(2, activation='softmax'))
```

```
    return model
```

A.2 Proposed Segmentation model

The code is written using PyTorch[52] library

```
class VNet3_seperate_up(nn.Module):
    # the number of convolutions in each layer corresponds
    # to what is in the actual prototxt, not the intent
    def __init__(self, elu=True, nll=False, flat=True,o_chan=2,i_o_chan=1,
    brain=False):
        super(VNet3_seperate_up, self).__init__()
        self.brain = brain
        self.in_tr = InputTransition(16, elu, i_o_chan)
        self.down_tr32 = DownTransition(16, 2, elu)
        self.down_tr64 = DownTransition(32, 3, elu)
        self.down_tr128 = DownTransition(64, 4, elu, dropout=True)
        self.down_tr256 = DownTransition(128, 5, elu, dropout=True)


        self.up_256_128 = Up(256, 128, elu, dropout=True)
        self.up_256_64  = Up(256, 64, elu, dropout=True)
        self.up_128_32  = Up(128, 32, elu, dropout=True)
        self.up_64_16   = Up(64, 16, elu, dropout=True)

        self.up_tr256 = UpTransition3(256, 5, elu, dropout=True)
        self.up_tr128 = UpTransition3(128, 4, elu, dropout=True)
        self.up_tr64  = UpTransition3(64, 3, elu)
        self.up_tr32  = UpTransition3(32, 2, elu)
        if flat:
            self.out_tr   = OutputTransition(32, elu, nll)
        else:
            self.out_tr   = OutputTransition2(32, elu, nll, o_chan)

        self.att_128  = Attention_block3(128,128,128 )
        self.att_64   = Attention_block3(64,64,64)
        self.att_32   = Attention_block3(32,32,32)
        self.att_16   = Attention_block3(16,16,16)

        self.conv_1   = Conv_1_block(512,256, elu, dropout=True)
        self.conv_2   = Conv_1_block(256,128, elu, dropout=True)
        self.conv_3   = Conv_1_block(128,64, elu, dropout=True)


    def forward(self, x):
        out16 = self.in_tr(x, brain=self.brain)
#         print(out16.size())
```

```
        out32 = self.down_tr32(out16)
        out64 = self.down_tr64(out32)
        out128 = self.down_tr128(out64)
        out256 = self.down_tr256(out128)

        up128 = self.up_256_128(out256)
#         print(f"up128: {up128.size()}")
        at128 = self.att_128(out128, up128)
        cat   = torch.cat((up128, at128),1)
        out   = self.up_tr256(cat)

        up64  = self.up_256_64(out)
        at64  = self.att_64(out64, up64)
        cat   = torch.cat((up64, at64),1)
        out   = self.up_tr128(cat)

        up32  = self.up_128_32(out)
        at32  = self.att_32(out32, up32)
        cat   = torch.cat((up32, at32),1)
        out   = self.up_tr64(cat)

        up16  = self.up_64_16(out)
        at16  = self.att_16(out16, up16)
        cat   = torch.cat((up16, at16),1)
        out   = self.up_tr32(cat)

        out = self.out_tr(out)
#         out = F.interpolate(out, (240,240,155), mode='nearest')
        return out
```

Appendix B

Accepted Paper

Deep 3D Convolution Neural Network for Alzheimer's Detection

Charul Giri[1], Morten Goodwin[2], and Ketil Oppedal[3]

Centre for Artificial Intelligence Research (CAIR),
University of Agder, Norway
{charug18[1],morten.goodwin[2]}@uia.no
Stavanger University Hospital
ketil.oppedal@sus.no[3]

Abstract. One of the most well-known and complex applications of artificial intelligence (AI) is Alzheimer's detection, which lies in the field of medical imaging. The complexity in this task lies in the three-dimensional structure of the MRI scan images. In this paper, we propose to use 3D Convolutional Neural Networks (3D-CNN) for Alzheimer's detection. 3D-CNNs have been a popular choice for this task. The novelty in our paper lies in the fact that we use a deeper 3D-CNN consisting of 10 layers. Also, with effectively training our model consisting of Batch Normalization layers that provide a regularizing effect, we don't have to use any transfer learning. We also use the simple data augmentation technique of flipping. Our model is trained for binary classification that distinguishes between Alzheimer's and normal, as well as multiclass classification consisting of Alzheimer's, Mild Cognitive Impairment, and normal classes. We tested our model on the ADNI dataset and achieved 94.17% and 89.14% accuracy for binary classification and multiclass classification, respectively.

Keywords: Convolutional Neural Networks, Alzheimer's Detection, Medical Imaging, Deep Learning, MRI.

1 Introduction

With all the advancement in medical technology, medical error is still a very common factor that contributes to 180,000 deaths every year as of 2008, reported by the US Department of Health and Human Services Office of the Inspector General. This fact makes medical error the third leading causes of deaths in the US according to [22]. Today, more than 50 million people suffer from dementia worldwide and with increasing life expectancy the number will increase. Alzheimer's disease (AD) is one of the most common forms contributing to 60%-70% of the cases (ref: https://www.who.int/news-room/fact-sheets/detail/dementia). AD is an irreversible neurodegenerative disease, in which the human brain cells involved in cognitive functioning are damaged and eventually die. Important symptoms are memory loss, reduced ability to learn new things and think. Additionally,

orientation, comprehension, calculation, language and judgement are affected leading to the loss of the ability to perform everyday activities.

Diagnosing Alzheimer's disease in early stage is really difficult because it is thought to begin 20 years before the symptoms arrive [30]. The changes occurring in the brain during this stage are small and may be untraceable. According to [1], around 5.8 million people in US have Alzheimer's disease and the prediction goes to 13.8 million by 2050. Diagnosing AD is challenging and requires a thorough clinical assessment based on the patient's medical and family history, neuro-psychologic and - psychiatric testing, as well as blood and brain imaging exams such as CT-, PET- and MRI. Even though analysis of brain scans can point the doctors and researchers in the right direction, AD detection is challenging, since similar changes are common during aging as well. Computer-aided diagnostic technology has the potential to improve the challenging task of early and accurate detection of AD diagnosis. The latest research in deep learning has shown promising results in solving a large number of problems from various fields with a very high accuracy, including medical image analysis. Computer vision techniques and machine learning algorithms certainly have the potential to process medical images to efficiently diagnose Alzheimer's disease. The most successful type of model for medical image processing are Convolution Neural Networks (CNN).
In this study we have developed a CNN based pipeline to detect AD and classify it's different stages against the normal cognitive stage (NC), i.e, Mild cognitive impairment (MCI) and AD using MRI as input. MCI is a stage between normal age related forgetfulness due and the development of AD. Not all people with MCI necessarily develop AD, but many of them have a higher risk of developing AD. MRI of the brain can capture structural changes such as the decrease in size of the temporal and parietal lobes which are typically reduced in patients with AD.
CNN is a subclass of Artificial Neural Networks where features are extracted from data using feature maps or kernels that spatially share weights. These feature maps work towards finding distinct features in the images in order to distinguish them into different categories. This gives CNN the ability to extract features from MR images and to classify AD. The multiple building blocks of CNN such as convolution layer, pooling layer efficiently learn feature maps to gain spatial knowledge from image data. That's why CNN have an advantage over other deep learning algorithms in medical image analysis. The novelty of this research inheres within the use of 3D-CNNs to process whole MRI images instead of using MRI slices with 2D CNNS, giving us better performance.

2 Related Work

Several studies have been conducted in the recent years to develop a computer-aided diagnosis system for Alzheimer's detection. Traditional methods included researcher trying to handcraft features through voxel-based methods, ROI based

methods, hippocampal shape and volume or patch-based methods.
[21] have attempted to compute region of interest (ROI) to detect AD. ROI is a section of image in which a binary mask is used to carry out various operations like filtering. [5] used voxel based morphometry (VBM) and MRI to investigate gray matter change in medial temporal structures and volume changes in several other brain regions.
[26] have analysed regional brain atrophy for example in the hippocampus to detect patterns of neuron death by segmenting different types of brain tissues such as grey-matter (GM), white-matter (WM) against cerebrospinal fluid (CSF) in the MRI. They segmented the images using watershed transformation algorithm [24] with marker image, and then calculating the shrinkage happened in the whole brain through Tissue Atrophy Ratio (AT) for early detection of AD.
[17] has used an inherent structure-based multi-view learning (ISML) method in which they have extracted multi-view features based on multiple selected templates. They then employed a subclass clustering algorithm for feature selection in order to eliminate the redundant features. A SVM-based ensemble classifier is used to classify subjects into AD, MCI and NC.
[23] have used Hu moments invariants [13], calculating a set of seven invariant moments to extract features in the brain images (MRI) of all subjects. They also showed that normalizing these moments results in better feature extraction which makes it easier for the classifier to distinguish. The extracted features are then used as inputs to SVM and KNN classifiers to classify the subjects. They compared the classifiers, showing that SVM performed far better than KNN.
In [8] feature extraction is done by using ROI on three sMRI biomarkers, named as Voxel-based morphometry, Cortical and sub-cortical volume and Hippocampus volume. They used Principal component analysis (PCA) [4] for feature selection. PCA is a dimensionality reduction method simplifying a high dimension data into smaller dimension without losing the important patterns or trend in the data. Using PCA, they selected 61 features for the classification of AD. They studied three different classifiers: SVM, Random Forest and KNN, and evaluated their performance in two stages. First stage included individual features from s VBM-extracted ROI volumes, CSC-extracted feature volumes, and HV extracted features and second staged is evaluating classifiers using the combination of all 61 features. They concluded that SVM outperformed KNN and Random Forest in all cases.
Multi-modal data fusion using MRI and PET scans was proposed in [20]. They used stacked auto encoders and and a sigmoidal decoder to discover the synergy between MRI and PET scans for high level feature extraction with a softmax classifier. A zero-masking technique (SAE-ZEROMASK) is used in contrast to simple feature concatenation (SAE-CONCAT) technique. They randomly hide one modality and trained the hidden layers to reconstruct the multi-modal using inputs mixed with hidden modality.
However, SAE-CONCAT usually fails to captures the non-linear co-relation between two different modalities [27]. That's why authors in [27] proposed Multimodal Stacked Deep Polynomial Networks algorithm (MM-SDPN) which uses

multi-modalities like [20] but they have used two stages of SDPN to learn high-level features.

Various other machine learning algorithms have proven to be efficient when it comes to extracting high level features. Artificial neural networks were used by [9] for Nephropathy Detection and Classification. The drawback of using Feed Forward Neural Network (as usually called ANN) for computer vision is that they are computationally expensive. The number of learning parameters in ANN exponentially increase with respect to the size of the image. Thus to counter this problem, the use of convolution neural network (CNN) to automate feature learning in images has become popular because of their ability to generalize well to high dimensional data, without losing important patterns.

A 2-D CNN is presented by [2] where they used VGG16 [28] as a base model, and treated a 3-D MRI image as a stack of 2-D MRI slices. Other variants of CNNs have been used in the researches such as in [15] authors used a ROI focused 3-D CNN with multi-modality. Each modality and ROI region was assigned a dedicated pipeline of a CNN block, whose output was flattened. The flattened outputs are the extracted features from each modality and region of interests(ROI). These feature outputs were then concatenated, resulting into late data fusion and were passed to a softmax classifier. Problem with this approach is similar to what was described earlier, that while performing late data fusion using simple concatenation, it ignores the variance in the nature of multi-modalities and fails to learn the non-linear co-relation between modalities [27].

A few other research works employ pre-training 3-D CNN with auto encoders such as [25].They used sparse auto encoders for feature extraction and also compared the performance of 2-D CNN against 3-D CNN. Authors in [12] took a two stage approach where they first used a convolutional auto encoder in place of conventional unsupervised auto encoder to extract local features with possibly long voxel-wise signal vectors. These features are used to perform task-specific classification with a target-domain-adaptable 3D-CNN using transfer learning with Net2Net weight initialization. They later proposed [11] in which they trained the same model with deep supervision, which resultant in an improvement.

[6] also build a 3-D convolutional neural network for an end-to-end classification of subjects with AD. They added metadata (sex and age of subjects) to the first fully connected layer in their model. The downside of using metadata in the neural network is that the network will try to find the correlation based on the metadata that might be biased towards the predilection of meta-data, for e.g., older patient are more likely to be affected by Alzheimer's Disease, so the network might bias towards assigning older people to the Alzheimer's Class.

[31] studied various paradigms of 3-D CNNs like patch-level 3D CNN, ROI based 3D CNN, subject-level 3D CNN, along with exploring transfer learning using auto encoder pre-training and ImageNet pre-training. They also reviewed studies done on AD classification using Deep Learning from January 1990 to the 15^{th} of January 2019, which proved very helpful in the proposed research.

Table 1: Demographic data for 817 subjects from the ADNI database (STD – standard deviation))

Diagnosis	Subjects	Age($mean_{\pm std}$)	Gender (F/M)
AD	188	$75.36_{\pm 7.5}$	89/99
MCI	401	$74.84_{\pm 7.3}$	143/258
NC	228	$75.96_{\pm 5.0}$	110/118

3 Proposed Work

In this work we have addressed the problem of Alzheimer's Disease Detection, and proposed a novel architecture based on 3D convolution Networks. The advantage of using a 3D convolution Neural Network over 2D convolution Neural Network is that 3D CNN are able to extract features in 3D space. For example, in a video it can derive spatial information from 2 dimensions and as well as temporal information. In our case the MRI scans don't have the temporal dimension, but are 3D images. So by using a 3D CNN we can extract the spatial information from the three dimensional space.
We have performed binary classification between AD and NC (Normal Cognitive) and multi-class classification between AD, MCI and NC. We have used whole brain MRI scans on subject level for the network to focus on and eliminated the need of selecting ROI.

3.1 Dataset and Pre-processing

We collected the data from the Alzheimer's Disease Neuroimaging Initiative (ADNI). ADNI [14] is a collaboration for research in the progression of Alzheimer's Disease, started in 2004 as a private-public partnership among 20 companies. The data consists of 1075 1.5T Screening MRI records of 817 subjects from ADNI1 project. The statistics of subjects is shown in Table 1.
The MR images in the dataset don't have a standard size, so we downscaled the images to a resolution of 120 x 90 x 130 maintaining the axial view. We follow the common practice of downscaling for the deep learning model to process the large images.
Since Alzheimer's can start in any hemisphere of the brain, so it makes sense to augment the data by flipping horizontally. Having larger dataset is crucial for better performance in deep learning models. We only used left and right flip augmentation of the data. We have chosen axial view as it helps in avoiding the motion artifacts from eyeball which can appear in other views [7]. Table 1 gives a summary of the demographic information of the subjects studied in this paper.

3.2 Model

In this paper, we have used a 3-Dimensional Convolution Neural Network (3D CNN). As the name suggests 3D CNN performs convolution operation in 3 di-

Table 2: Model Architecture

Layer Name	Kernel size	No of kernels/neurons
Conv3D	$5 \times 5 \times 5$	32
BatchNorm	-	-
Conv3D	$3 \times 3 \times 3$	32
Conv3D	$3 \times 3 \times 3$	32
BatchNorm	-	-
MaxPool3D	$2 \times 2 \times 2$	-
Conv3D	$3 \times 3 \times 3$	64
Conv3D	$3 \times 3 \times 3$	64
BatchNorm	-	-
MaxPool3D	$2 \times 2 \times 2$	-
Conv3D	$3 \times 3 \times 3$	64
Conv3D	$3 \times 3 \times 3$	128
BatchNorm	-	-
MaxPool3D	$2 \times 2 \times 2$	-
Conv3D	$3 \times 3 \times 3$	64
Conv3D	$3 \times 3 \times 3$	128
BatchNorm	-	-
Conv3D	$3 \times 3 \times 3$	64
BatchNorm	-	-
MaxPool3D	$2 \times 2 \times 2$	-
Flatten	-	-
Dense	-	512
Dropout(0.1)	-	-
Dense	-	2

mensions to extract features as opposed to a traditional 2D Convolution Neural Network, which works in only 2D space. Mathematically a 3D convolution operation in neural network is defined as follows:
Given an input of size N,H,W,D,C_in where H=height, W=width, D=depth, C=no.of channels, N=batch size, the output of the convolution layer is produced as:

$$Out(N_i, C_{out_j}) = b(C_{out_j}) + \sum_{k=0}^{C_{in}-1} weight(C_{out_j}, k) * input(N_i, k) \qquad (1)$$

where * is the 3D cross correlation operation between two signals. The learnable kernels are $l \times l \times l$ matrices, which slide over the large input to detect relevant patterns creating new feature maps and convolving feature maps from previous layer. The architecture of the model is given in Table 2. Our model consists of 4 identical convolution blocks with 2 convolution layer stacked up on a batch normalization layer in each block. Every block goes through 3D maxpooling with a pool size of (2,2,2). Beside this, the model has an input block, last convolution block and a classifier block. The input block consists of one convolution layer and

batch normalization layer. The last convolution block consists of 1 convolution layer followed by a batch normalization and maxpool layer respectively. The classifier block has one fully connected layer followed by a softmax classifier. All convolution blocks contain filters of size 3x3x3, to learn the small details and patterns of Alzheimer's affected parts in the brain.
We have used relu activation function on all layers and categorical crossentropy cost function to optimize the loss, using Adam optimizer. ReLu doesn't activate the neurons with negative input values, which makes it is computationally very efficient over sigmoid and tanh function as it introduces sparsity.

3.3 Training

To train the model we have first initialized all the kernels with He normal initialization to achieve faster convergence. The model takes a batch size of 2 samples at a time. We have trained the model using 10-fold cross-validation to ensure that the model will perform well on all points of data, and not on just some random sets. To prevent the model from overfitting we have used L2 regularization on the last fully connected layer, which is followed by a dropout layer. To stop convolutional layer learning irrelevant features or over-fit to the features, we have used batch-normalization. The batch-normalization layer helps in reducing the co-variance shift of the hidden units in the neural network. It normalizes the output of the previous layer using the current batch statistics. In this way the distribution of the weights of the hidden units or kernels in our case can adapt to any change in the distribution of the data. Batch normalization layer also gave us some regularization effect as it scales the weights based on the batch mean and variance, which makes the weights stable. [16] showed that L2 regularizer has no regularizing effect when it is used along with Normalization, however it can influence the effective learning rate.

4 Experiments

We trained our model using k-fold cross validation with 10 folds. The data distribution of training, validation and testing is 80%, 10% and 10% respectively. The model performs two classification tasks: binary classification between AD and NC, and multiclass classification between AD, NC and MNC. We have seen in other studies that transfer learning using Net2Net and ImageNet has been used extensively before. But our model is able to learn the features for this classification task and achieve superior performance compared to the other models without transfer learning. The results of our approach are presented in Table 3 where we have compared it with other models.
We didn't use any kind of transfer learning because pre-trained models like ImageNet models don't contain relevant information for transfer learning of medical data classification. Other research works have used the weights of their pre-trained model from binary classification (e.g., AD and NC) to initialize the model for multi-class classification (e.g, AD, MCI and NC) or vice versa. The

Table 3: Comparison with previous studies of Alzheimer's Detection using Deep Learning on ADNI Data. Accuracy metrics is used for comparison and represents binary classification between Alzheimer's Disease (AD) vs Normal Cognition (NC) and multiclass classification between Alzheimer's Disease (AD) vs Normal Cognitive (NC) Mild cognitive impairment (MCI)

Method	Modality	Dataset Size	Accuracy	
			AD vs NC	AD vs MCI vs NC
[18]	FDG-PET	ADNI (339)	91.20	-
[29]	PET + MRI	ADNI (317)	91.10	-
[19]	MRI+PET	ADNI (331)	91.40	53.84
[3]	MRI	ADNI (815)	91.41	-
[6]	MRI	ADNI (841)	94.10	61.10
Our Approach	MRI	ADNI (817)	94.17	89.14

drawback of this approach is that if the same data is present for some classes in the next classification, then it will lead to data leakage in the model, causing biased transfer learning. To tackle this issue we initialized our model using He initialization during both classification tasks.

We ran the model for 100 epochs for each fold. We experimented with different hyper-parameters like learning rate, drop out rate and regularization coefficient. We found the best set of parameters for the optimal performance of the model is an initial learning rate of 0.0001, drop out factor of 0.1 and regularization parameter as 0.001. We also used scheduled learning rate based on step decay to optimize the loss curve and avoid divergence. The learning rate is scheduled to drop by a factor of 0.1 every 40^{th} epoch. As mentioned earlier we have used Batch normalization layer with Convolution Layers. We initially tried L2 regularizer. L2 regularizer when used along with Batch Normalization has no regularizing effect [16]. So we tried L2 regularizer and Batch Normalization separately, and found that the model performed better with Batch Normalization than with L2 regularizer, increasing the accuracy by 3% in binary classification and 2.34% in multiclass classification.

We thoroughly tested our model and compared it with previous Deep Learning based Alzheimer Detection model. [6] have used demographic information for the classification task. They merged the age and gender information as additional features in their network. Providing such information to the network can make it bias towards the correlation and pattern present in the information. And if the correlations are strong then the network will rely on these demographic features more than the features learned from the MRI images. For example it's a well know fact that eighty-one percent of people who have Alzheimer's disease are age 75 or older, and almost two-thirds of Americans with Alzheimer's are women [10]. If this statistics is provided then it can make the model biased towards a particular age or sex group.

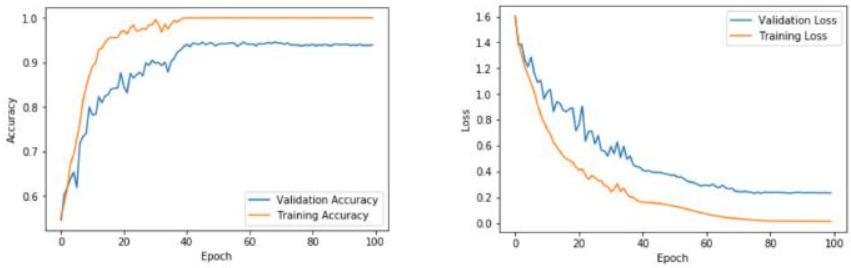
To evaluate the performance of our model we have used accuracy (Acc), precision and recall, and F_2 score. Given True Positive, True Negative, False Positive

Table 4: Evaluation Matrix for the Deep 3D CNN Model (L2: L2 regularization, BN: Batch Normalization. Last row represents the configuration of our model)

Model	L2	BN	Binary Class				Multi Class			
			Acc(%)	F_2	Pre	Recall	Acc(%)	F_2	Pre	Recall
Deep 3D CNN	✓	✓	91.11	0.90	0.89	0.94	86.22	0.85	0.88	0.87
Deep 3D CNN	✓	✗	91.20	0.91	0.91	0.92	86.70	0.86	0.84	0.85
Deep 3D CNN	✗	✓	94.17	0.94	0.94	0.94	89.14	0.88	0.89	0.88

and False Negative as TP, TN, FP, FN recall and precision is calculated as $Precision = TP/(TP + FP)$, $Recall = TP/(TP + FN)$. The F_2 score is calculated as $(5 \times Precision \times Recall)/(4Precision + Recall)$. We have used F_2 score because it gives more emphasis on false negatives making it a suitable option for medical experiments evaluation.

Evaluation Matrix Table 4 presents the evaluation matrix averaged over 10 folds, of our model for the two classification tasks, and comparing the effect of l2 and batch normalization. The training vs validation loss and training vs validation

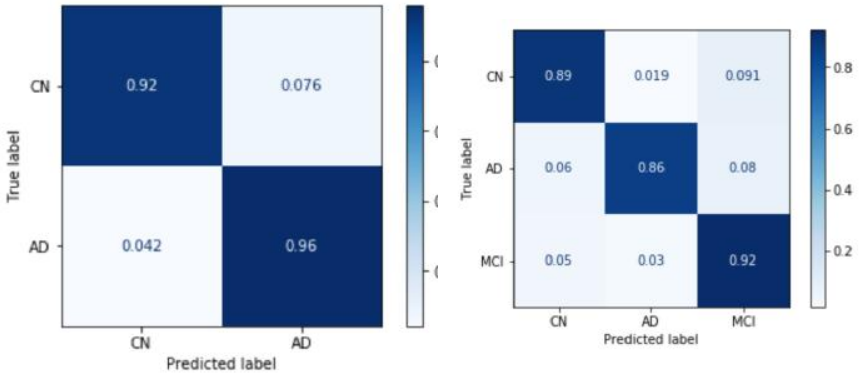


(a) (b)

Fig. 1: (a) training vs validation accuracy (b) training vs validation loss

accuracy of the average of the 10 folds is plotted in Fig. 1 for binary classification. It can be seen that after every 40 epochs the graph becomes smoother. It's due to the learning rate decay occurring every 40^{th} epoch. The learning rate decreases by a factor of 0.1 here, helping the model to converge the optimization and making the learning more stable.

As can be seen from Fig. 2, the confusion matrix details the class-wise performance of our 3D CNN model on binary classification. During the evaluation of all the folds, we noted that the model has a very low false negative rate on its best fold, of 0.016 for AD class in binary classification, and 0.06 in multiclass classification, where it confuses a bit with MCI class. The average false negative rate is still very low at 0.04 for AD class in binary classification. These experi-



(a) (b)

Fig. 2: 10 folds average of Confusion Matrix for (a) Binary Classification and (b) Multi-class Classification for model with best accuracy listed in Table 4

ments show that our model has outperformed the other models with only single modality and training from scratch. The results also demonstrate the robustness and confidence of the Alzheimer's prediction by the proposed model.

5 Conclusion and Future Work

In this paper we proposed a novel 3D CNN based classifier for Alzheimer's Detection. The model showed the ability to learn the relevant features on its own without the use of any transfer learning. From the results of multi-class classification it can be observed that the model has the potential of distinguishing between two very similar looking classes, in our case AD and MCI, and MCI and NC, with a very high confidence. We showed that proper initialization of model and fine parameter tuning leads to superior results. The future work involves experimenting with multiple modalities and exploring other feature extraction and feature selection algorithms so that the model will be able to learn more with small amount of data. The model can be improved in further studies by carrying out more sophisticated procedure for feature selection after feature extraction as just mentioned, and using deeply supervised learning. The next step could be to segment the biomarkers of Alzheimer's in the brain, and predicting the pattern of growth of the disease in a very early stage. We would also like to explore the machine learning models on MCI subjects to predict who might develop AD at a later stage.

References

1. 2019 alzheimer's disease facts and figures. *Alzheimer's and Dementia*, 15(3):321 – 387, 2019.

2. Convolutional neural network based alzheimer's disease classification from magnetic resonance brain images. *Cognitive Systems Research*, 57:147 – 159, 2019.
3. Karim Aderghal, Jenny Benois-Pineau, Karim Afdel, and Gwénaëlle Catheline. Fuseme: Classification of smri images by fusion of deep cnns in 2d+e projections. In *CBMI '17*, 2017.
4. Anders H. Andersen, Don M. Gash, and Malcolm J. Avison. Principal component analysis of the dynamic response measured by fmri: a generalized linear systems framework. *Magnetic Resonance Imaging*, 17(6):795 – 815, 1999.
5. M. Bozzali, M. Filippi, G. Magnani, M. Cercignani, M. Franceschi, E. Schiatti, S. Castiglioni, R. Mossini, M. Falautano, G. Scotti, G. Comi, and A. Falini. The contribution of voxel-based morphometry in staging patients with mild cognitive impairment. *Neurology*, 67(3):453–460, 2006.
6. Soheil Esmaeilzadeh, Dimitrios Belivanis, Kilian Pohl, and Ehsan Adeli. End-to-end alzheimer's disease diagnosis and biomarker identification, miccai 2018, proceedings. pages 337–345, 09 2018.
7. Masami Goto, Shigeki Aoki, Osamu Abe, Tomohiko Masumoto, Yasushi Watanabe, Yoshiroh Satake, Katsuji Nishida, Ino Kenji, Keiichi Yano, Kyohhito Iida, Kazuo Mima, and Kuni Ohtomo. Utility of axial images in an early alzheimer disease diagnosis support system (vsrad). *Nihon Hoshasen Gijutsu Gakkai zasshi*, 62:1339–44, 10 2006.
8. Yubraj Gupta, Kun Ho Lee, Kyu Yeong Choi, Jang Jae Lee, Byeong Chae Kim, Goo Rak Kwon, the National Research Center for Dementia, and Alzheimer's Disease Neuroimaging Initiative. Early diagnosis of alzheimer's disease using combined features from voxel-based morphometry and cortical, subcortical, and hippocampus regions of mri t1 brain images. *PLOS ONE*, 14(10):1–30, 10 2019.
9. Anil Hazarika. A novel technique of neuropathy detection and classification by using artificial neural network (ann). 01 2013.
10. Liesi E. Hebert, Jennifer Weuve, Paul A. Scherr, and Denis A. Evans. Alzheimer disease in the united states (2010–2050) estimated using the 2010 census. *Neurology*, 80(19):1778–1783, 2013.
11. Ehsan Hosseini asl, Mohammed Ghazal, Ali Mahmoud, Ali Aslantas, Ahmed Shalaby, Manual Casanova, Gregory Barnes, Georgy Gimel'farb, Robert Keynton, and Ayman El-Baz. Alzheimer's disease diagnostics by a 3d deeply supervised adaptable convolutional network. *Frontiers in bioscience (Landmark edition)*, 23:584–596, 01 2018.
12. Ehsan Hosseini asl, Robert Keynton, and Ayman El-Baz. Alzheimer's disease diagnostics by adaptation of 3d convolutional network. 09 2016.
13. Ming-Kuei Hu. Visual pattern recognition by moment invariants. *IRE Trans. Information Theory*, 8:179–187, 1962.
14. Clifford R. Jack Jr. et al. The alzheimer's disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging*, 27(4):685–691, 2008.
15. Alexander Khvostikov, Karim Aderghal, Jenny Benois-Pineau, Andrey Krylov, and Gwenaelle Catheline. 3d cnn-based classification using smri and md-dti images for alzheimer disease studies. 01 2018.
16. Twan Laarhoven. L2 regularization versus batch and weight normalization. 06 2017.
17. M. Liu, D. Zhang, E. Adeli, and D. Shen. Inherent structure-based multiview learning with multitemplate feature representation for alzheimer's disease diagnosis. *IEEE Transactions on Biomedical Engineering*, 63(7):1473–1482, July 2016.

18. Manhua Liu, Danni Cheng, Weiwu Yan, and Alzheimer's Disease Neuroimaging Initiative. Classification of alzheimer's disease by combination of convolutional and recurrent neural networks using fdg-pet images. In *Front. Neuroinform.*, 2018.
19. S. Liu, S. Liu, W. Cai, H. Che, S. Pujol, R. Kikinis, D. Feng, M. J. Fulham, and ADNI. Multimodal neuroimaging feature learning for multiclass diagnosis of alzheimer's disease. *IEEE Transactions on Biomedical Engineering*, 62(4):1132–1140, April 2015.
20. Siqi Liu, Sidong Liu, Weidong Cai, Hangyu Che, Sonia Pujol, Ron Kikinis, David Dagan Feng Feng, and Michael Fulham. Multimodal neuroimaging feature learning for multiclass diagnosis of alzheimer's disease. *IEEE Transactions on Biomedical Engineering*, 62:1132–1140, 04 2015.
21. S. Arun M. Latha. Detection of roi for classifying alzheimer's disease using mr. image of brain. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 8, 2019.
22. Martin A Makary and Michael Daniel. Medical error—the third leading cause of death in the us. *BMJ*, 353, 2016.
23. Arwa Mohammed, Fadwa al azzo, and Mariofanna Milanova. Classification of alzheimer disease based on normalized hu moment invariants and multiclassifier. *International Journal of Advanced Computer Science and Applications*, 8, 01 2017.
24. L. Najman and M. Schmitt. Geodesic saliency of watershed contours and hierarchical segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(12):1163–1173, Dec 1996.
25. Adrien Payan and Giovanni Montana. Predicting alzheimer's disease: a neuroimaging study with 3d convolutional neural networks. *ICPRAM 2015 - 4th International Conference on Pattern Recognition Applications and Methods, Proceedings*, 2, 02 2015.
26. Rowayda Sadek. Regional atrophy analysis of mri for early detection of alzheimer's disease. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 6:50–58, 02 2013.
27. Jun Shi, Xiao Zheng, Yan Li, Qi Zhang, and Shihui Ying. Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of alzheimer's disease. *IEEE Journal of Biomedical and Health Informatics*, PP:1–1, 01 2017.
28. Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv 1409.1556*, 09 2014.
29. Tien Duong Vu, Hyung-Jeong Yang, V. Q. Nguyen, A-Ran Oh, and Mi-Sun Kim. Multimodal learning using convolution neural network and sparse autoencoder. In *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pages 309–312, Feb 2017.
30. Victor L Villemagne, Samantha Burnham, Pierrick Bourgeat, Belinda Brown, Kathryn A Ellis, Olivier Salvado, Cassandra Szoeke, S Lance Macaulay, Ralph Martins, Paul Maruff, David Ames, Christopher C Rowe, and Colin L Masters. Amyloid beta deposition, neurodegeneration, and cognitive decline in sporadic alzheimer's disease: a prospective cohort study. *The Lancet Neurology*, 12(4):357 – 367, 2013.
31. Junhao Wen, Elina Thibeau-Sutre, Jorge Samper-Gonzalez, Alexandre Routier, Simona Bottani, Stanley Durrleman, Ninon Burgos, and Olivier Colliot. Convolutional neural networks for classification of alzheimer's disease: Overview and reproducible evaluation, 04 2019.

Bibliography

[1]

[2] Convolutional neural network based alzheimerâ^{TM}s disease classification from magnetic resonance brain images. *Cognitive Systems Research*, 57:147 – 159, 2019.

[3] Karim Aderghal, Jenny Benois-Pineau, Karim Afdel, and Gwénaëlle Catheline. Fuseme: Classification of smri images by fusion of deep cnns in 2d+e projections. In *CBMI '17*, 2017.

[4] Anders H. Andersen, Don M. Gash, and Malcolm J. Avison. Principal component analysis of the dynamic response measured by fmri: a generalized linear systems framework. *Magnetic Resonance Imaging*, 17(6):795 – 815, 1999.

[5] Alzheimer's Association, William Thies, and Laura Bleiler. 2013 alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 9(2):208–245, 2013.

[6] Spyridon Bakas, Mauricio Reyes, Andras Jakab, Stefan Bauer, Markus Rempfler, Alessandro Crimi, Russell Takeshi Shinohara, Christoph Berger, Sung Min Ha, Martin Rozycki, Marcel Prastawa, Esther Alberts, Jana Lipkova, John Freymann, Justin Kirby, Michel Bilello, Hassan Fathallah-Shaykh, Roland Wiest, Jan Kirschke, Benedikt Wiestler, Rivka Colen, Aikaterini Kotrotsou, Pamela Lamontagne, Daniel Marcus, Mikhail Milchenko, Arash Nazeri, Marc-Andre Weber, Abhishek Mahajan, Ujjwal Baid, Elizabeth Gerstner, Dongjin Kwon, Gagan Acharya, Manu Agarwal, Mahbubul Alam, Alberto Albiol, Antonio Albiol, Francisco J. Albiol, Varghese Alex, Nigel Allinson, Pedro H. A. Amorim, Abhijit Amrutkar, Ganesh Anand, Simon Andermatt, Tal Arbel, Pablo Arbelaez, Aaron Avery, Muneeza Azmat, Pranjal B., W Bai, Subhashis Banerjee, Bill Barth, Thomas Batchelder, Kayhan Batmanghelich, Enzo Battistella, Andrew Beers, Mikhail Belyaev, Martin Bendszus, Eze Benson, Jose Bernal, Halandur Nagaraja Bharath, George Biros, Sotirios Bisdas, James Brown, Mariano Cabezas, Shilei Cao, Jorge M. Cardoso, Eric N Carver, AdriÃ Casamitjana, Laura Silvana Castillo, Marcel CatÃ , Philippe Cattin, Albert Cerigues, Vinicius S. Chagas, Siddhartha Chandra, Yi-Ju Chang, Shiyu Chang, Ken Chang, Joseph Chazalon, Shengcong Chen, Wei Chen, Jefferson W Chen, Zhaolin Chen, Kun Cheng, Ahana Roy Choudhury, Roger Chylla, Albert ClÃ©rigues, Steven Colleman, Ramiro German Rodriguez Colmeiro, Marc Combalia, Anthony Costa, Xiaomeng Cui, Zhenzhen Dai, Lutao Dai, Laura Alexandra

Daza, Eric Deutsch, Changxing Ding, Chao Dong, Shidu Dong, Wojciech Dudzik, Zach Eaton-Rosen, Gary Egan, Guilherme Escudero, ThÃ©o Estienne, Richard Everson, Jonathan Fabrizio, Yong Fan, Longwei Fang, Xue Feng, Enzo Ferrante, Lucas Fidon, Martin Fischer, Andrew P. French, Naomi Fridman, Huan Fu, David Fuentes, Yaozong Gao, Evan Gates, David Gering, Amir Gholami, Willi Gierke, Ben Glocker, Mingming Gong, Sandra GonzÃ¡lez-VillÃ¡, T. Grosges, Yuanfang Guan, Sheng Guo, Sudeep Gupta, Woo-Sup Han, Il Song Han, Konstantin Harmuth, Huiguang He, Aura HernÃ¡ndez-SabatÃ©, Evelyn Herrmann, Naveen Himthani, Winston Hsu, Cheyu Hsu, Xiaojun Hu, Xiaobin Hu, Yan Hu, Yifan Hu, Rui Hua, Teng-Yi Huang, Weilin Huang, Sabine Van Huffel, Quan Huo, Vivek HV, Khan M. Iftekharuddin, Fabian Isensee, Mobarakol Islam, Aaron S. Jackson, Sachin R. Jambawalikar, Andrew Jesson, Weijian Jian, Peter Jin, V Jeya Maria Jose, Alain Jungo, B Kainz, Konstantinos Kamnitsas, Po-Yu Kao, Ayush Karnawat, Thomas Kellermeier, Adel Kermi, Kurt Keutzer, Mohamed Tarek Khadir, Mahendra Khened, Philipp Kickingereder, Geena Kim, Nik King, Haley Knapp, Urspeter Knecht, Lisa Kohli, Deren Kong, Xiangmao Kong, Simon Koppers, Avinash Kori, Ganapathy Krishnamurthi, Egor Krivov, Piyush Kumar, Kaisar Kushibar, Dmitrii Lachinov, Tryphon Lambrou, Joon Lee, Chengen Lee, Yuehchou Lee, M Lee, Szidonia Lefkovits, Laszlo Lefkovits, James Levitt, Tengfei Li, Hongwei Li, Wenqi Li, Hongyang Li, Xiaochuan Li, Yuexiang Li, Heng Li, Zhenye Li, Xiaoyu Li, Zeju Li, XiaoGang Li, Wenqi Li, Zheng-Shen Lin, Fengming Lin, Pietro Lio, Chang Liu, Boqiang Liu, Xiang Liu, Mingyuan Liu, Ju Liu, Luyan Liu, Xavier Llado, Marc Moreno Lopez, Pablo Ribalta Lorenzo, Zhentai Lu, Lin Luo, Zhigang Luo, Jun Ma, Kai Ma, Thomas Mackie, Anant Madabushi, Issam Mahmoudi, Klaus H. Maier-Hein, Pradipta Maji, CP Mammen, Andreas Mang, B. S. Manjunath, Michal Marcinkiewicz, S McDonagh, Stephen McKenna, Richard McKinley, Miriam Mehl, Sachin Mehta, Raghav Mehta, Raphael Meier, Christoph Meinel, Dorit Merhof, Craig Meyer, Robert Miller, Sushmita Mitra, Aliasgar Moiyadi, David Molina-Garcia, Miguel A. B. Monteiro, Grzegorz Mrukwa, Andriy Myronenko, Jakub Nalepa, Thuyen Ngo, Dong Nie, Holly Ning, Chen Niu, Nicholas K Nuechterlein, Eric Oermann, Arlindo Oliveira, Diego D. C. Oliveira, Arnau Oliver, Alexander F. I. Osman, Yu-Nian Ou, Sebastien Ourselin, Nikos Paragios, Moo Sung Park, Brad Paschke, J. Gregory Pauloski, Kamlesh Pawar, Nick Pawlowski, Linmin Pei, Suting Peng, Silvio M. Pereira, Julian Perez-Beteta, Victor M. Perez-Garcia, Simon Pezold, Bao Pham, Ashish Phophalia, Gemma Piella, G. N. Pillai, Marie Piraud, Maxim Pisov, Anmol Popli, Michael P. Pound, Reza Pourreza, Prateek Prasanna, Vesna Prkovska, Tony P. Pridmore, Santi Puch, Ã‰lodie Puybareau, Buyue Qian, Xu Qiao, Martin Rajchl, Swapnil Rane, Michael Rebsamen, Hongliang Ren, Xuhua Ren, Karthik Revanuru, Mina Rezaei, Oliver Rippel, Luis Carlos Rivera, Charlotte Robert, Bruce Rosen, Daniel Rueckert, Mohammed Safwan, Mostafa Salem, Joaquim Salvi, Irina Sanchez, Irina SÃ¡nchez, Heitor M. Santos, Emmett Sartor, Dawid Schellingerhout, Klaudius Scheufele, Matthew R. Scott, Artur A. Scussel, Sara Sedlar, Juan Pablo Serrano-Rubio, N. Jon Shah, Nameetha Shah, Mazhar Shaikh, B. Uma

Shankar, Zeina Shboul, Haipeng Shen, Dinggang Shen, Linlin Shen, Haocheng Shen, Varun Shenoy, Feng Shi, Hyung Eun Shin, Hai Shu, Diana Sima, M Sinclair, Orjan Smedby, James M. Snyder, Mohammadreza Soltaninejad, Guidong Song, Mehul Soni, Jean Stawiaski, Shashank Subramanian, Li Sun, Roger Sun, Jiawei Sun, Kay Sun, Yu Sun, Guoxia Sun, Shuang Sun, Yannick R Suter, Laszlo Szilagyi, Sanjay Talbar, Dacheng Tao, Dacheng Tao, Zhongzhao Teng, Siddhesh Thakur, Meenakshi H Thakur, Sameer Tharakan, Pallavi Tiwari, Guillaume Tochon, Tuan Tran, Yuhsiang M. Tsai, Kuan-Lun Tseng, Tran Anh Tuan, Vadim Turlapov, Nicholas Tustison, Maria Vakalopoulou, Sergi Valverde, Rami Vanguri, Evgeny Vasiliev, Jonathan Ventura, Luis Vera, Tom Vercauteren, C. A. Verrastro, Lasitha Vidyaratne, Veronica Vilaplana, Ajeet Vivekanandan, Guotai Wang, Qian Wang, Chiatse J. Wang, Weichung Wang, Duo Wang, Ruixuan Wang, Yuanyuan Wang, Chunliang Wang, Guotai Wang, Ning Wen, Xin Wen, Leon Weninger, Wolfgang Wick, Shaocheng Wu, Qiang Wu, Yihong Wu, Yong Xia, Yanwu Xu, Xiaowen Xu, Peiyuan Xu, Tsai-Ling Yang, Xiaoping Yang, Hao-Yu Yang, Junlin Yang, Haojin Yang, Guang Yang, Hongdou Yao, Xujiong Ye, Changchang Yin, Brett Young-Moxon, Jinhua Yu, Xiangyu Yue, Songtao Zhang, Angela Zhang, Kun Zhang, Xuejie Zhang, Lichi Zhang, Xiaoyue Zhang, Yazhuo Zhang, Lei Zhang, Jianguo Zhang, Xiang Zhang, Tianhao Zhang, Sicheng Zhao, Yu Zhao, Xiaomei Zhao, Liang Zhao, Yefeng Zheng, Liming Zhong, Chenhong Zhou, Xiaobing Zhou, Fan Zhou, Hongtu Zhu, Jin Zhu, Ying Zhuge, Weiwei Zong, Jayashree Kalpathy-Cramer, Keyvan Farahani, Christos Davatzikos, Koen van Leemput, and Bjoern Menze. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge, 2018.

[7] Abdelmajid Bousselham, Omar Bouattane, Youssfi Mohamed, and Abdelhadi Raihani. Towards reinforced brain tumor segmentation on mri images based on temperature changes on pathologic area. *International Journal of Biomedical Imaging*, 2019:18, 03 2019.

[8] M. Bozzali, M. Filippi, G. Magnani, M. Cercignani, M. Franceschi, E. Schiatti, S. Castiglioni, R. Mossini, M. Falautano, G. Scotti, G. Comi, and A. Falini. The contribution of voxel-based morphometry in staging patients with mild cognitive impairment. *Neurology*, 67(3):453–460, 2006.

[9] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, 1986.

[10] Adrià Casamitjana, Marcel Catà, Irina Sánchez, Marc Combalia, and Verónica Vilaplana. Cascaded v-net using roi masks for brain tumor segmentation. In Alessandro Crimi, Spyridon Bakas, Hugo Kuijf, Bjoern Menze, and Mauricio Reyes, editors, *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 381–391, Cham, 2018. Springer International Publishing.

[11] François Chollet et al. Keras. https://keras.io, 2015.

[12] Ozgun Cicek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: Learning dense volumetric segmentation from sparse annotation, 2016.

[13] W. Deng, Q. Shi, M. Wang, B. Zheng, and N. Ning. Deep learning-based hcnn and crf-rrnn model for brain tumor segmentation. *IEEE Access*, 8:26665–26675, 2020.

[14] Florian Dubost, Gerda Bortsova, Hieab Adams, Arfan Ikram, Wiro J. Niessen, Meike Vernooij, and Marleen De Bruijne. Gp-unet: Lesion detection from weak labels with a 3d regression network. In Maxime Descoteaux, Lena Maier-Hein, Alfred Franz, Pierre Jannin, D. Louis Collins, and Simon Duchesne, editors, *Medical Image Computing and Computer Assisted Intervention â^' MICCAI 2017*, pages 214–221, Cham, 2017. Springer International Publishing.

[15] Soheil Esmaeilzadeh, Dimitrios Belivanis, Kilian Pohl, and Ehsan Adeli. End-to-end alzheimerâ^{TM}s disease diagnosis and biomarker identification, miccai 2018, proceedings. pages 337–345, 09 2018.

[16] Soheil Esmaeilzadeh, Dimitrios Belivanis, Kilian Pohl, and Ehsan Adeli. End-to-end alzheimerâ^{TM}s disease diagnosis and biomarker identification, miccai 2018, proceedings. pages 337–345, 09 2018.

[17] Markus Frey and Matthias Nau. Memory efficient brain tumor segmentation using an autoencoder-regularized u-net. *Lecture Notes in Computer Science*, page 388â“396, 2020.

[18] Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4):193–202, apr 1980.

[19] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In Geoffrey Gordon, David Dunson, and Miroslav DudÃk, editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pages 315–323, Fort Lauderdale, FL, USA, 11–13 Apr 2011. PMLR.

[20] Masami Goto, Shigeki Aoki, Osamu Abe, Tomohiko Masumoto, Yasushi Watanabe, Yoshiroh Satake, Katsuji Nishida, Ino Kenji, Keiichi Yano, Kyohhito Iida, Kazuo Mima, and Kuni Ohtomo. Utility of axial images in an early alzheimer disease diagnosis support system (vsrad). *Nihon Hoshasen Gijutsu Gakkai zasshi*, 62:1339–44, 10 2006.

[21] Yubraj Gupta, Kun Ho Lee, Kyu Yeong Choi, Jang Jae Lee, Byeong Chae Kim, Goo Rak Kwon, the National Research Center for Dementia, and Alzheimerâ^{TM}s Disease Neuroimaging Initiative. Early diagnosis of alzheimerâ^{TM}s disease using combined features

from voxel-based morphometry and cortical, subcortical, and hippocampus regions of mri t1 brain images. *PLOS ONE*, 14(10):1–30, 10 2019.

[22] Anil Hazarika. A novel technique of neuropathy detection and classification by using artificial neural network (ann). 01 2013.

[23] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

[24] Liesi E. Hebert, Jennifer Weuve, Paul A. Scherr, and Denis A. Evans. Alzheimer disease in the united states (2010â“2050) estimated using the 2010 census. *Neurology*, 80:1778â“1783, 2013.

[25] Ehsan Hosseini asl, Mohammed Ghazal, Ali Mahmoud, Ali Aslantas, Ahmed Shalaby, Manual Casanova, Gregory Barnes, Georgy Gimelâ™farb, Robert Keynton, and Ayman El-Baz. Alzheimer's disease diagnostics by a 3d deeply supervised adaptable convolutional network. *Frontiers in bioscience (Landmark edition)*, 23:584–596, 01 2018.

[26] Ehsan Hosseini asl, Robert Keynton, and Ayman El-Baz. Alzheimer's disease diagnostics by adaptation of 3d convolutional network. 09 2016.

[27] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7132–7141, 2018.

[28] K. Hu, Q. Gan, Y. Zhang, S. Deng, F. Xiao, W. Huang, C. Cao, and X. Gao. Brain tumor segmentation using multi-cascaded convolutional neural networks and conditional random field. *IEEE Access*, 7:92615–92629, 2019.

[29] Ming-Kuei Hu. Visual pattern recognition by moment invariants. *IRE Trans. Information Theory*, 8:179–187, 1962.

[30] Gao Huang, Zhuang Liu, and Kilian Q. Weinberger. Densely connected convolutional networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269, 2017.

[31] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift, 2015.

[32] Fabian Isensee, Philipp Kickingereder, Wolfgang Wick, Martin Bendszus, and Klaus H. Maier-Hein. Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge. In Alessandro Crimi, Spyridon Bakas, Hugo Kuijf, Bjoern Menze, and Mauricio Reyes, editors, *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 287–297, Cham, 2018. Springer International Publishing.

[33] Jaber Juntu, Jan Sijbers, Dirk Dyck, and Jan Gielen. Bias field correction for mri images. pages 543–551, 01 2005.

[34] Konstantinos Kamnitsas, Enzo Ferrante, Sarah Parisot, Christian Ledig, Aditya V. Nori, Antonio Criminisi, Daniel Rueckert, and Ben Glocker. Deepmedic for brain tumor segmentation. In Alessandro Crimi, Bjoern Menze, Oskar Maier, Mauricio Reyes, Stefan Winzeck, and Heinz Handels, editors, *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 138–149, Cham, 2016. Springer International Publishing.

[35] Alexander Khvostikov, Karim Aderghal, Jenny Benois-Pineau, Andrey Krylov, and Gwenaelle Catheline. 3d cnn-based classification using smri and md-dti images for alzheimer disease studies. 01 2018.

[36] Twan Laarhoven. L2 regularization versus batch and weight normalization. 06 2017.

[37] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. [van der Laak], Bram [van Ginneken], and Clara I. SÃ¡nchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60 – 88, 2017.

[38] Geert Litjens, Robert Toth, Wendy [van de Ven], Caroline Hoeks, Sjoerd Kerkstra, Bram [van Ginneken], Graham Vincent, Gwenael Guillard, Neil Birbeck, Jindang Zhang, Robin Strand, Filip Malmberg, Yangming Ou, Christos Davatzikos, Matthias Kirschner, Florian Jung, Jing Yuan, Wu Qiu, Qinquan Gao, Philip âœEddieâ Edwards, Bianca Maan, Ferdinand [van der Heijden], Soumya Ghose, Jhimli Mitra, Jason Dowling, Dean Barratt, Henkjan Huisman, and Anant Madabhushi. Evaluation of prostate segmentation algorithms for mri: The promise12 challenge. *Medical Image Analysis*, 18(2):359 – 373, 2014.

[39] M. Liu, D. Zhang, E. Adeli, and D. Shen. Inherent structure-based multiview learning with multitemplate feature representation for alzheimer's disease diagnosis. *IEEE Transactions on Biomedical Engineering*, 63(7):1473–1482, July 2016.

[40] Manhua Liu, Danni Cheng, Weiwu Yan, and Alzheimer's Disease Neuroimaging Initiative. Classification of alzheimerâ^{TM}s disease by combination of convolutional and recurrent neural networks using fdg-pet images. In *Front. Neuroinform.*, 2018.

[41] S. Liu, S. Liu, W. Cai, H. Che, S. Pujol, R. Kikinis, D. Feng, M. J. Fulham, and ADNI. Multimodal neuroimaging feature learning for multiclass diagnosis of alzheimer's disease. *IEEE Transactions on Biomedical Engineering*, 62(4):1132–1140, April 2015.

[42] Siqi Liu, Sidong Liu, Weidong Cai, Hangyu Che, Sonia Pujol, Ron Kikinis, David Dagan Feng Feng, and Michael Fulham. Multimodal neuroimaging feature learning for multiclass diagnosis of alzheimer's disease. *IEEE Transactions on Biomedical Engineering*, 62:1132–1140, 04 2015.

[43] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation, 2014.

[44] S. Arun M. Latha. Detection of roi for classifying alzheimerâ^{TM}s disease using mr. image of brain. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 8, 2019.

[45] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, L. Lanczi, E. Gerstner, M. Weber, T. Arbel, B. B. Avants, N. Ayache, P. Buendia, D. L. Collins, N. Cordier, J. J. Corso, A. Criminisi, T. Das, H. Delingette, Ã‡. Demiralp, C. R. Durst, M. Dojat, S. Doyle, J. Festa, F. Forbes, E. Geremia, B. Glocker, P. Golland, X. Guo, A. Hamamci, K. M. Iftekharuddin, R. Jena, N. M. John, E. Konukoglu, D. Lashkari, J. A. Mariz, R. Meier, S. Pereira, D. Precup, S. J. Price, T. R. Raviv, S. M. S. Reza, M. Ryan, D. Sarikaya, L. Schwartz, H. Shin, J. Shotton, C. A. Silva, N. Sousa, N. K. Subbanna, G. Szekely, T. J. Taylor, O. M. Thomas, N. J. Tustison, G. Unal, F. Vasseur, M. Wintermark, D. H. Ye, L. Zhao, B. Zhao, D. Zikic, M. Prastawa, M. Reyes, and K. Van Leemput. The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging*, 34(10):1993–2024, 2015.

[46] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation, 2016.

[47] Arwa Mohammed, Fadwa al azzo, and Mariofanna Milanova. Classification of alzheimer disease based on normalized hu moment invariants and multiclassifier. *International Journal of Advanced Computer Science and Applications*, 8, 01 2017.

[48] Gowtham Krishnan Murugesan, Sahil Nalawade, Chandan Ganesh, Ben Wagner, Fang F. Yu, Baowei Fei, Ananth J. Madhuranthakam, and Joseph A. Maldjian. Multidimensional and multiresolution ensemble networks for brain tumor segmentation. *bioRxiv*, 2019.

[49] Andriy Myronenko. 3d mri brain tumor segmentation using autoencoder regularization. In Alessandro Crimi, Spyridon Bakas, Hugo Kuijf, Farahani Keyvan, Mauricio Reyes, and Theo van Walsum, editors, *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 311–320, Cham, 2019. Springer International Publishing.

[50] L. Najman and M. Schmitt. Geodesic saliency of watershed contours and hierarchical segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(12):1163–1173, Dec 1996.

[51] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben

Glocker, and Daniel Rueckert. Attention u-net: Learning where to look for the pancreas, 2018.

[52] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas KÃ¶pf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library, 2019.

[53] Adrien Payan and Giovanni Montana. Predicting alzheimer's disease: a neuroimaging study with 3d convolutional neural networks. *ICPRAM 2015 - 4th International Conference on Pattern Recognition Applications and Methods, Proceedings*, 2, 02 2015.

[54] Marcel Prastawa, Elizabeth Bullitt, Sean Ho, and Guido Gerig. A brain tumor segmentation framework based on outlier detection. *Medical Image Analysis*, 8(3):275 – 283, 2004. Medical Image Computing and Computer-Assisted Intervention - MICCAI 2003.

[55] Jacob C. Reinhold, Blake E. Dewey, Aaron Carass, and Jerry L. Prince. Evaluating the impact of intensity normalization on mr image synthesis. *Proceedings of SPIE–the International Society for Optical Engineering*, 10949, 2018.

[56] Torsten Rohlfing, Natalie Zahr, Edith Sullivan, and Adolf Pfefferbaum. The sri24 multichannel atlas of normal adult human brain structure. *Human brain mapping*, 31:798–819, 05 2009.

[57] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

[58] Rowayda Sadek. Regional atrophy analysis of mri for early detection of alzheimer's disease. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 6:50–58, 02 2013.

[59] Jun Shi, Xiao Zheng, Yan Li, Qi Zhang, and Shihui Ying. Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of alzheimer's disease. *IEEE Journal of Biomedical and Health Informatics*, PP:1–1, 01 2017.

[60] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv 1409.1556*, 09 2014.

[61] Shuang Song, Yuanjie Zheng, and Yunlong He. A review of methods for bias correction in medical images. 2017.

[62] M. Thaha, Pradeep Kumar, B. Murugan, S. Dhanasekeran, P. Vijayakarthick, and A. Selvi. Brain tumor segmentation using convolutional neural networks in mri images. *Journal of Medical Systems*, 43, 09 2019.

[63] Tien Duong Vu, Hyung-Jeong Yang, V. Q. Nguyen, A-Ran Oh, and Mi-Sun Kim. Multimodal learning using convolution neural network and sparse autoencoder. In *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pages 309–312, Feb 2017.

[64] Chengjia Wang, Tom MacGillivray, Gillian Macnaught, Guang Yang, and David E. Newby. A two-stage 3d unet framework for multi-class segmentation on full resolution image. *ArXiv*, abs/1804.04341, 2018.

[65] G. Wang, W. Li, M. A. Zuluaga, R. Pratt, P. A. Patel, M. Aertsen, T. Doel, A. L. David, J. Deprest, S. Ourselin, and T. Vercauteren. Interactive medical image segmentation using deep learning with image-specific fine tuning. *IEEE Transactions on Medical Imaging*, 37(7):1562–1573, 2018.

[66] Guotai Wang, Wenqi Li, SÃ©bastien Ourselin, and Tom Vercauteren. Automatic brain tumor segmentation based on cascaded convolutional neural networks with uncertainty estimation. *Frontiers in Computational Neuroscience*, 13, 08 2019.

[67] Junhao Wen, Elina Thibeau-Sutre, Jorge Samper-Gonzalez, Alexandre Routier, Simona Bottani, Stanley Durrleman, Ninon Burgos, and Olivier Colliot. Convolutional neural networks for classification of alzheimer's disease: Overview and reproducible evaluation, 04 2019.

[68] Leon Weninger, Oliver Rippel, Simon Koppers, and Dorit Merhof. Segmentation of brain tumors and patient survival prediction: Methods for the brats 2018 challenge. In Alessandro Crimi, Spyridon Bakas, Hugo Kuijf, Farahani Keyvan, Mauricio Reyes, and Theo van Walsum, editors, *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 3–12, Cham, 2019. Springer International Publishing.