

Brain Lesion Segmentation from MRI

Final Report

Peter Paik*, *Student Member, IEEE*, and John Varun Paul Konaiahgari*

Abstract—Magnetic Resonance Imaging (MRI) is a non-invasive method to capture the internal and external structures of the brain which is used by expert radiologists to make prognoses and diagnoses of brain-related abnormalities. Using expert-labeled MRI images, image processing techniques, and deep-learning techniques, computers can be equipped to detect brain abnormalities in what is called Brain Lesion Segmentation. Lesions, which are areas of damaged brain tissue, can be difficult to find without the help of experts, and these judgments can also vary between experts. Therefore, the capability of using computers for lesion detection is a helpful resource to the medical community. This project will compare three deep-learning convolutional neural networks (CNN) models for brain lesion segmentation. The popular and widely used 2016 MICCAI challenge dataset is used [1]–[3], which contains 3D MRI images of different MRI modalities. Another purpose of this project is to study the data pre-processing steps used for preparing MRI images, particularly denoising techniques. In this regard, the performance of the non-local means for denoising and anisotropic diffusion filtering for denoising is examined.

Index Terms—MRI, Segmentation, Neural Networks

I. PROJECT OVERVIEW

This project covers the following topics: a literature review, testing and comparison of different denoising techniques for pre-processing, and also testing and comparison of three different U-net architectures for brain lesion segmentation.

The literature review covers different types of Magnetic Resonance Imaging (MRI) modalities, existing labeled brain lesion datasets, data pre-processing steps, and existing deep learning pipelines for brain lesions segmentation. The literature review is conducted by both authors and is used to make informed decisions on the type of dataset to use, pre-processing steps to compare, and existing deep-learning pipelines to explore.

The testing and comparison of different denoising techniques are conducted by Peter. In this step, denoising techniques and skull segmentation techniques are explored. In addition, different denoising techniques are compared using the peak-signal-to-noise ratio (PSNR). The other pre-processing steps are also discussed and reviewed heavily in the literature review to investigate how they improve the quality of the MRI images. Data pre-processing contains crucial steps to ensure

Peter Paik and John Varun Paul Konaiahgari is with the Department of Electrical and Computer Engineering, New York University, New York, NY, 11201, USA.

*Paik and Konaiahgari contributed equally to this work.

This work is completed for the Image and Video Processing Course (EL-GY 6123) of Spring 2023 at NYU under the Department of Electrical and Computer Engineering.

the dataset is of good quality, especially since deep-learning techniques are only as good as the data provided.

The testing and comparison of three different U-net architectures are conducted by both authors. In this step, three U-net models are trained and tested using the same dataset [1]. Specifically, a 2D U-net with a single modality of MRI, a 2D U-net with four modalities of MRI, and a 3D U-net with a single modality of MRI are compared. The performance is evaluated using the soft DICE loss function.

The rest of this report is structured as follows: Section II reports the findings from the literature review, Section III reports the data pre-processing results, Section IV presents the brain lesion segmentation pipeline and results, and Section V summarizes the project findings.

II. LITERATURE REVIEW

This section provides the details of the literature review, starting with the different types of MRI Modalities, which helped to guide our decisions as we advance in the project.

A. Types of MRI Modalities

Although other non-invasive methods for capturing brain structures exist, MRI provides more detailed information and is typically safer [4]. The popular types of MRI modalities include T1-weighted MRI, T2-weighted MRI, T1-weighted MRI with contrast enhancement, and Fluid-Attenuated Inversion Recovery (FLAIR) [4]–[6]. T1-weighted MRI is used to easily detect healthy brain tissue, T1-weighted images with contrast enhancement is used to detect the brain lesion border easily, and T2-weighted MRI and FLAIR are used to easily detect swollen regions of the brain, which are possible areas of brain lesions. In addition to finding swollen regions, FLAIR helps separate the cerebrospinal fluid from these swollen regions [4]–[6]. Typically, deep-learning techniques combine all MRI modalities for brain lesion segmentation. However, since the FLAIR contains the most distinctive features of the brain lesion, single modality models exist using the FLAIR. Thus, we have also opted for FLAIR in our single-modality U-nets.

B. Brain MRI Datasets

To train deep-learning algorithms to have good accuracy in detecting brain lesions, it is necessary to have a good dataset containing expert-labeled ground truth segmentation of MRI. Scientists and researchers have created many open-source datasets to improve the techniques and methods for automated brain lesion segmentation. These include AANLIB

by Harvard Medical School, ADNI1 by Alzheimer's Disease neuroimaging initiative, ATLAS by Allen Institute Publications for Brain Science, BRATS 2016 by MICCAI 2016 Challenge, and RIDER by TCIA among many others [4], [7].

Some important considerations made when determining the dataset to use were the number of modalities provided, the number of images in the dataset, and the method by which the ground truth data was labeled.

Many of these datasets provide T1-weighted and T2-weighted MRI; however, only a few provide the FLAIR, and even fewer provide T1-weighted with contrast enhancement MRI. The only dataset that provided the four modalities was the MICCAI dataset. In addition, this same dataset contained 300 images from 300 patients. These images are 3D images, meaning that they contain the 2D splices of the brain. This allows for creating a desired 2D or 3D convolutional neural network (CNN) design. However, the images are segmented manually by only one to four experts.

A group from the National Institute for Research and Digital Science and Technology improved on the MICCAI 2016 dataset by having manual segmentation from 7 experts [1]¹. However, only 53 of the 300 images were labeled by the experts. Regardless, we have decided that using a dataset with high-quality ground truth data and multiple modalities will provide a stronger foundation for designing a method that can 'expertly' detect brain lesions from typical MRI modalities.

C. Data Pre-processing

It is necessary to pre-process the images to produce quality images for training deep-learning models. Standard MRI images contain noise artifacts, non-cerebral regions of tissue (i.e., skull, scalp), intensity variations, and possible deviations in the angle due to the scanning angle [6], [8]. Therefore, many pre-processing steps are typically required to produce a quality image for training a neural network.

The first pre-processing step is to remove noise artifacts through denoising techniques such as Anisotropic Diffusion Filtering (ADF), Discrete-time Wavelet Transforms (DWT), Non-Local Means (NLM), or Independent Component Analysis (ICA) [6], [8]. ADF has been a very popular algorithm used in MRI denoising. ADF removes noise from an image by smoothening areas of pixels that do not have edges by filtering in the directions parallel to the edges. This method computes the gradient of the pixels along each direction and applies an iterative approach to reducing noisy pixels through a weighted Laplacian operator [9].

Non-cerebral regions of the brain can be removed in a process called skull stripping which, in itself, is another type of segmentation task. Some existing algorithms include BSE, BET, HWA, and ROBEX [10]. This is necessary since the non-cerebral regions do not contain any relevant information for brain lesion segmentation.

Intensity normalization or standardization is used to achieve uniformity in the intensity of the splices of the MRI image through histogram matching of the intensities [11]. The non-parametric nonuniform intensity normalization (N3) is one

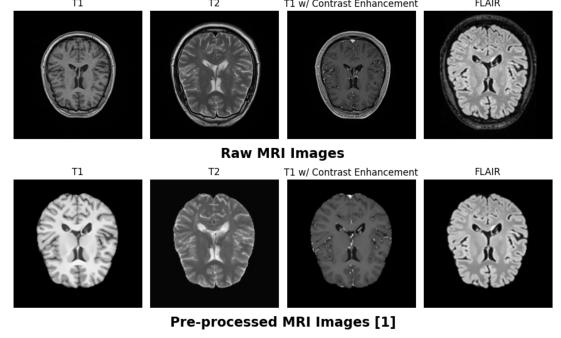


Fig. 1. MRI data of the four modalities (T1, T2, T1 w/ contrast enhancement and FLAIR). The top row is the unprocessed raw data, and the bottom row is the pre-processed data performed by [1].

such intensity normalization scheme [12] which has been widely used. This method was improved in 2010 by Tustison et. al and called N4ITK [13]. Today, this is one of the standard intensity normalization algorithms for MRI images.

Another crucial pre-processing step, if using multiple modalities of MRI, is block matching [14]. This ensures that the pixels in each location of the different MRI modalities are matched to the same brain locations. This works by resampling the other MRI modalities according to the geometry of a chosen MRI modality through interpolation [14].

D. Deep-learning Methods

Deep-learning methods for brain lesion segmentation is a fully automatic segmentation task where the trained model on a computer performs the segmentation task without any expert consultation [5]. These methods require excellent prior knowledge of segmentation through large datasets and expertly labeled data.

Convolutional Neural Networks (CNNs) are a popular choice since they are good at identifying spatial patterns (such as lesions) and also because the working data are images; in fact, most deep-learning methods are based on a CNN framework. 2D CNNs and 3D CNNs have been established for classifying brain lesions from 2D and 3D MRI images [15]; examples include ResNet, VGG, GoogleNet, AlexNet [8]. A more recent CNN-based algorithm for image segmentation used in brain lesion segmentation is the U-net [16]. In this regard, we have decided to create and evaluate three different U-net models for brain lesion segmentation.

III. DATA PRE-PROCESSING

As mentioned, MRI images require several data pre-processing steps before they can be used to train a neural network. The MRI images need to be pre-processed to have any form of consistency; several environmental factors can affect the noise, intensity, and sizes of the recorded MRI.

As shown in Fig. 1, in the raw unprocessed data (top row), the different modalities have different resolutions, different intensity ranges, and also contain parts of the skull. In the pre-processed data (bottom row), all modalities have the same image resolution through block matching, the intensities have been normalized, and the skulls are removed.

¹<https://shanoir.irisa.fr/shanoir-ng/welcome>

In this section, we implement the popular anisotropic diffusion filtering for denoising [9] and compare this with the non-local means denoising technique [17]. The two are compared using the peak-signal-to-noise ratio (PSNR) as a metric. In addition, skull stripping is performed using the open-source system volBrain [18]².

A. Approach

Anisotropic Diffusion Filtering involves computing the Laplacian of each voxel (i.e., pixels in a 3D image) $\nabla I(x, y, z, t)$. The Laplacian is scaled according to a conductivity factor $c(x, y, z, t)$ determined by the magnitude of the gradients and a defined diffusivity parameter K [9], [17].

$$c(x, t) = e^{-\frac{||\nabla I(x, y, z, t)||}{K^2}} \quad (1)$$

The conductivity factor is inversely proportional to the gradient magnitude and, therefore, smallest along edges.

$$\frac{\delta I(x, y, z, t)}{\delta t} = \text{div}(c(x, t)\nabla I(x, y, z, t)) \quad (2)$$

The original image is denoised iteratively by computing the weighted Laplacian of each pixel and adding the weighted Laplacian to the original image in each iteration. By doing this, the image is smoothed while preserving the edges. A larger diffusivity parameter, the number of iterations k , and iteration time step γ will increase smoothening at the risk of removing important structures. Therefore, these parameters need to be tuned.

Therefore, the intensity of each voxel in the image is updated for k iterations according to:

$$ADF(I(x, y, z, t)) = I(x, y, z, t-1) + \gamma * \frac{\delta I(x, y, z, t)}{\delta t} \quad (3)$$

Non-local means is a more computationally heavy approach that denoises the image while keeping the edges intact. This approach works by using blocks, a volume of voxels centered around a voxel of interest, and updating each voxel in the 3D image based on the similarity between the block of interest and surrounding blocks within a search radius (as a weighted average). The weights are assigned according to a Gaussian distribution. In [17], the algorithm was improved by (1) updating the voxels by blocks and not individually, (2) only comparing blocks that have high similarity according to the mean and variances of blocks, and (3) adaptively updating the smoothening parameter 'h'. See equations (8)-(11) of [17].

B. Results

The results for the ADF and NLM denoising techniques on the same 3D MRI FLAIR image are shown in Fig. 2. The FLAIR was chosen since it visually contained the most noise (see Fig. 1). The ADF and NLM denoising techniques were each performed on the raw 3D FLAIR images and the PSNR was calculated between the denoised FLAIR volume and the raw FLAIR volume for each method.

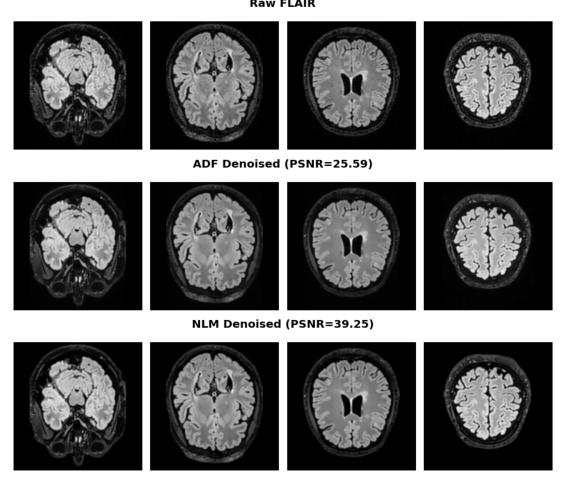


Fig. 2. Raw FLAIR Image slices (top), ADF-Denoised FLAIR Image slices (middle), and NLM-Denoised FLAIR Image slices (bottom). The related PSNR between the denoised and original 3D MRI is labeled.

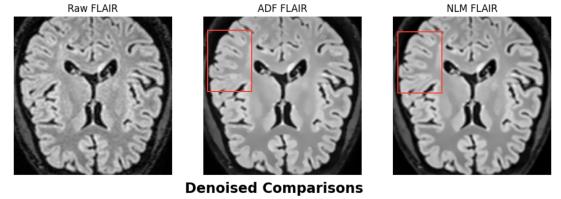


Fig. 3. Close-up comparison of one of the slices of the FLAIR image using the denoising techniques.

The PSNR of the NLM was higher than the ADF, meaning the signal-to-noise ratio of the NLM-denoised FLAIR is better than that of the ADF-denoised FLAIR. To get a closer look Fig. 3 is provided.

It can be seen that the intensities at the edges of the brain in the ADF are more smudged, while in the NLM, the intensities at the edges are more well-kept and distinguished. Although the ADF minimizes the iterative smoothening to the edges, it still contains non-zero changes. And this smoothening can increase with the number of iterations and diffusivity parameter. Meanwhile, NLM iterates through each voxel once, and the changes to the intensities are based only on the similarity between the intensity blocks. Since edges have a unique and directional intensity block compared to noise, these edges are kept intact very well for NLM. However, the NLM takes significantly longer to run compared to the ADF across the entire volume of the MRI.

The volBrain open-source algorithm provides a trained algorithm for segmenting the non-cerebral regions of the brain in a 3D volumetric image. This algorithm takes a T1-weighted image and outputs the resulting segmentation map. The result of the mask applied to the Denoised (using ADF) T1-weighted image is shown in Fig. 4. Since block matching has not been performed, it is not yet possible to apply this mask to the other modalities. But, once block matching is performed, the mask can easily be applied to the other modalities.

In Fig. 4, each slice of the T1-weighted image has been stripped of the non-cerebral regions. Therefore, the entire volume of the cerebral region is extracted using the mask.

²<https://www.volbrain.upv.es/>

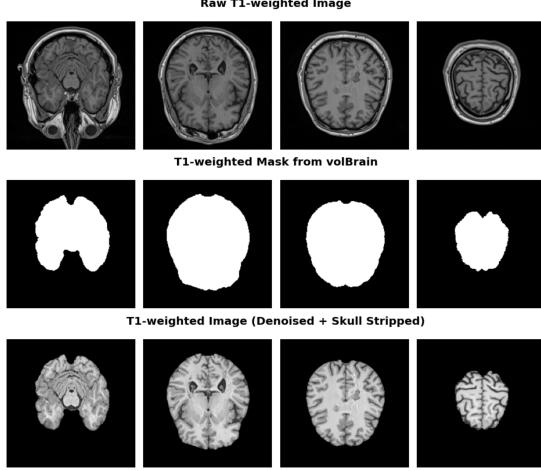


Fig. 4. Raw T1-weighted image slices (top), Mask slices of Cerebral Region from volBrain (middle) [18], and Denoised and Skull-stripped T1-weighted image slices (bottom).

As a result, the intensity range is adjusted in the stripped T1-weighted image slices since the bony regions (higher intensities) have been removed. Notice that these intensities are quite arbitrary; therefore, intensity normalization must be performed as a final step in the data pre-processing.

IV. DEEP LEARNING METHODS FOR LESION SEGMENTATION

The dataset from [1] contains 53 patient 3D MRI images (T1-weighted, T2-weighted, FLAIR, and T1-weighted with contrast enhancement (GADO)) from three different centers. Seven different experts segmented the FLAIR MRI from each patient. In addition, the seven expert segmentation maps have been used to create a consensus segmentation mask used as the ground truth mask. This section describes the approach used for the three U-net architectures.

A. Resource

New York University's High-Performance Computing (NYU HPC) environment called Greene was used to train and test the models. A Miniconda environment was set up using Singularity, and we used Open On-Demand to utilize the Jupyter Lab interface. Each model was trained on the V100 GPU with 32GB of memory. The dataset was uploaded to the endpoint using Globus. More details can be found on the Google site for NYU HPC³.

B. U-net Architectures

Three different U-net architectures were created: a 2D U-net using FLAIR slices, a 2D U-net using multimodal MRI slices (FLAIR, T1-weighted, T2-weighted, and T1-weight w/ contrast enhancement), and a 3D U-net using FLAIR volume.

The two 2D U-nets are identical in structure except that the multimodal model takes an input of 4 channels. The 3D U-net follows the same structure as the 2D U-net, except all functions are changed to be three-dimensional (i.e., 3D

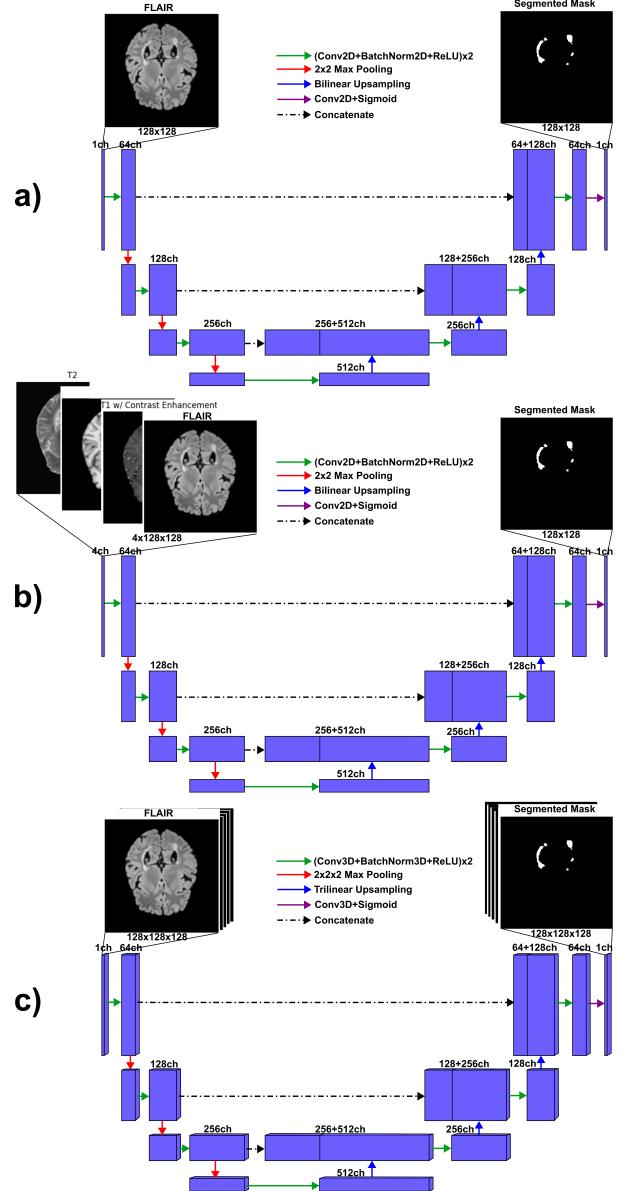


Fig. 5. a) 2D U-net Architecture using FLAIR slices only. b) 2D U-net Architecture using multimodal MRI slices. c) 3D U-net Architecture using FLAIR implemented using 3D FLAIR volume.

convolution instead of 2D convolution, trilinear upsampling instead of bilinear upsampling, etc.). All three models have 7 hidden layers, and 1 output layer. Each hidden layer comprises two sequences of 2D/3D convolution, 2D/3D batch normalization, and ReLU activation function. The first four hidden layers encompass the encoding path, where their outputs are downsampled by max pooling. The last three hidden layers encompass the decoding path, where their outputs are upsampled by bilinear/trilinear upsampling. The output layer is a 2D/3D convolution followed by a sigmoid activation function. As with U-net architectures, skip connections exist between the encoding and decoding paths where the outputs of the encoding layers are concatenated with the outputs of the decoding layers. All three model architectures can be seen in Fig. 5. The similarities in model architectures are done intentionally, so the models are comparable.

³<https://sites.google.com/nyu.edu/nyu-hpc/home>

TABLE I
MODEL SIZE

Model	Trainable Parameters (#)	Model Size (MB)
2D U-net (FLAIR)	7,785,089	235.51
2D U-net (Multi)	7,786,817	235.70
3D U-net (FLAIR)	23,344,001	19209.05

A comparison of model sizes is provided in Table I. As shown, the 2D U-nets have a similar model size, and the 3D U-net has triple the number of trainable parameters and a much larger model size.

1) *2D U-net Data Preparation:* Many slices of the MRI volume contained null layers/masks. Therefore, to avoid class imbalance issues, the top 25% and bottom 50% of the MRI were cropped. This provided us with a middle section of the MRI containing equal portions of the brain with and without lesions. For the 2D U-nets, only 15 patient MRIs were used; after cropping, 1380 slices of MRI were kept. These image slices were resized to 128-by-128.

The slices were then split into train-validation-test sets using an 80-10-10 split. Data augmentations were applied to the training set, which included: random flip, random rotation, random scale, and random shift.

The only difference between the single modality and multimodality 2D U-nets is that the four modality image slices are concatenated along the channel dimension before the data augmentations for the multimodal U-net.

2) *3D U-net Data Preparation:* For the 3D U-net, all null layers of the MRI are cropped, giving us just the slices containing parts of the brain. This also allows for uniformity in the brain shape and volume. Following the cropping, the volume is resized to 128-by-128-by-128. For the 3D U-net, 38 patient FLAIR MRIs were used.

These volumes were split into train-validation-test sets using an 80-10-10 split. Only the random flip data augmentation was applied to the training set.

3) *Training Hyperparameters:* All three models used the Adam optimizer, SoftDice Loss function, and 100 epochs. The 2D U-nets used a batch size of 8, while the 3D U-net uses a batch size of 2. The initial learning rate for the 2D U-net is 0.002. The initial learning rate for the 3D U-net is 0.0001. These values were chosen after observing the convergence/divergence of the models during training. All three use a scheduler that linearly reduces the learning rate to 10% at the final epoch. The best model based on the validation SoftDice score is saved during training. The best model is loaded and tested on the testing set.

C. Results

The results of the 'best' saved model for each of the architectures are presented in this section. For the 2D U-net using FLAIR, we achieved an 80.12% Dice score on the training set, 83.54% Dice score on the validation set, and 83.23% Dice score on the testing set. For the 2D U-net using multimodal MRI, we achieved an 82.98% Dice score on the training set, 86.04% Dice score on the validation set, and 86.15% Dice score on the testing set. For the 3D U-net using FLAIR, we achieved a 76.26% Dice score on the training set,

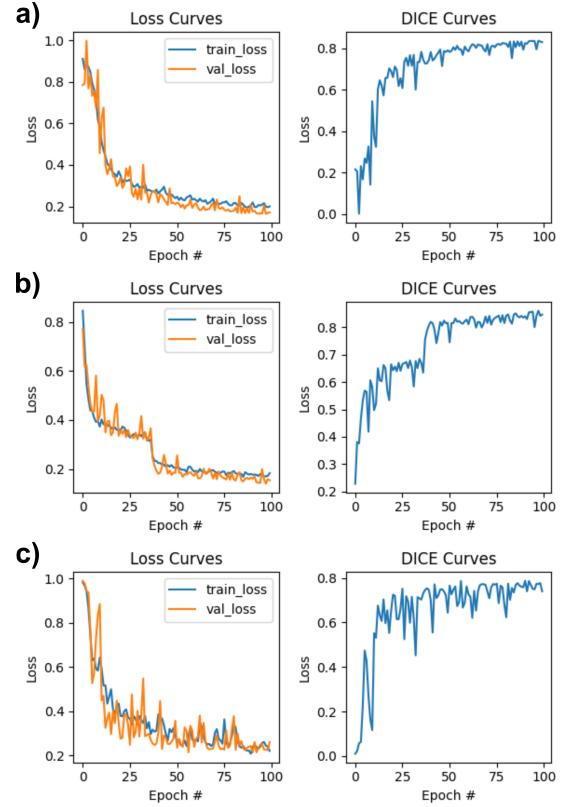


Fig. 6. Training Loss Curves and SoftDice Score curves for a) 2D U-net using FLAIR, b) 2D U-net using multimodal MRI, and c) 3D U-net using FLAIR.

TABLE II
MODEL RESULTS

Model	Test Dice Score (%)	Training Time (s)
2D U-net (FLAIR)	83.23	569
2D U-net (Multi)	86.04	552
3D U-net (FLAIR)	84.13	6041

78.62% Dice score on the validation set, and 84.13% Dice score on the testing set.

The SoftDice Loss curves and SoftDice curves during the training are shown in Fig. 6. For the 2D U-net using multimodal MRI, there is another increase in the Dice score around epoch 40 for both the training and validation sets. We are not quite sure what is the exact cause but believe that due to the multimodal nature, the model slowly transitions from one locally optimum solution to a globally optimum solution.

The SoftDice Test score results are summarized in Table II along with the training times. The 2D U-net using multimodal MRI boasts the highest Testing set Dice Score and also the fastest training time (albeit by a small amount compared to the other 2D U-net). One disclaimer is that the comparison between the 2D U-nets and the 3D U-net is not perfectly comparable since the testing sets are different. Regardless, the 2D U-net and 3D U-net using only the FLAIR have similar Dice scores on their own testing sets. However, the 3D U-net has a much longer training time compared to the 2D U-nets.

Example results from the testing set of the 2D U-nets are shown in Fig. 7. Likewise, example slices taken from the results of 3D U-net testing set are shown in Fig. 8. As can be seen, the segmented masks generated by the U-nets are able to

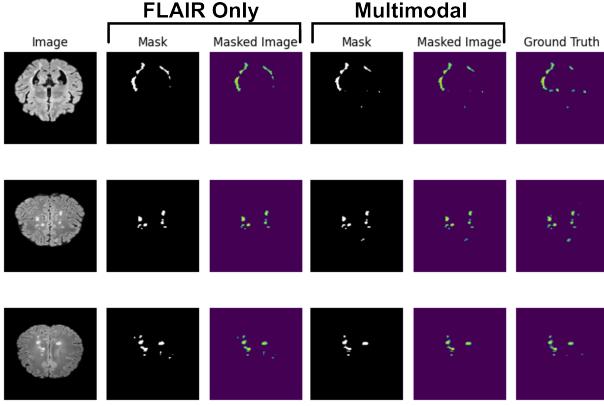


Fig. 7. Example Segmented Masks from the testing set of the 2D U-nets. The ground truth segmentation is provided as well.

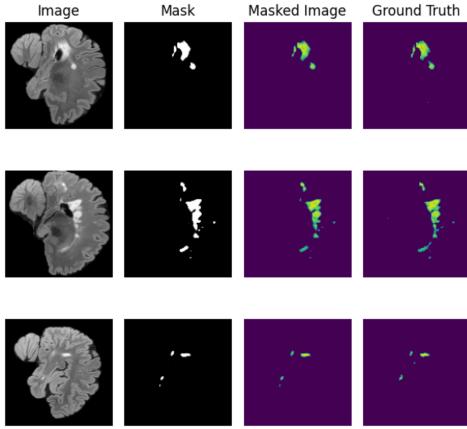


Fig. 8. Example Segmented Masks from the testing set of the 3D U-net. The ground truth segmentation is provided as well.

capture some of the lesion portions but not perfectly. However, the results are decent with all having Dice scores above 80%.

V. SUMMARY

In this project, we have explored several data pre-processing steps necessary for preparing an MRI for brain lesion segmentation. These data pre-processing steps include 3D denoising, skull stripping, intensity normalization or bias correction, and block matching. The first three are critical in any segmentation algorithm; block matching is crucial when using multiple modalities. In addition, we have trained, tested, and compared three different U-net architectures for brain lesion segmentation. We have found that the 2D U-net using multimodal MRI slices has the highest Dice Score of 86.04% and also the lowest training time of 552 seconds. However, the other two single modality U-nets (2D U-net and 3D U-net using FLAIR) perform with Dice scores above 80% as well. Through this project, we have learned about two image-filtering techniques called Anisotropic Diffusion Filtering and Non-local Means filtering. We have seen that the NLM outperforms the ADF but takes significantly more computational effort. We have also further explored different U-net architectures, their advantages, and nuances (i.e., hyperparameters). And, we also learned the importance of dataset preparation to avoid class imbalance or inconsistencies in brain shape/volume.

In the future, we hope to explore other convolutional neural network architectures for image segmentation. Additionally, we would try different loss functions and combinations of loss functions such as the focal loss (cross-entropy loss) with the Dice loss. Finally, we would like to explore other existing datasets, especially those without multiple modalities or block matching. Another interesting question is to see whether the network would be able to apply its own pre-processing to the raw MRI; therefore, relaxing some of the data pre-processing steps that we explored in the project.

All relevant files are available on a Google Drive folder⁴.

REFERENCES

- [1] O. Commowick *et al.*, “Multiple sclerosis lesions segmentation from multiple experts: The miccai 2016 challenge dataset,” *Neuroimage*, vol. 244, p. 118589, 2021.
- [2] U. Baid *et al.*, “The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification,” *arXiv preprint arXiv:2107.02314*, 2021.
- [3] S. Bakas *et al.*, “Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features,” *Scientific data*, vol. 4, no. 1, pp. 1–13, 2017.
- [4] M. K. Abd-Ellah *et al.*, “A review on brain tumor diagnosis from mri images: Practical implications, key achievements, and lessons learned,” *Magnetic resonance imaging*, vol. 61, pp. 300–318, 2019.
- [5] A. İşin *et al.*, “Review of mri-based brain tumor image segmentation using deep learning methods,” *Procedia Computer Science*, vol. 102, pp. 317–324, 2016.
- [6] J. Liu *et al.*, “A survey of mri-based brain tumor segmentation methods,” *Tsinghua science and technology*, vol. 19, no. 6, pp. 578–595, 2014.
- [7] B. H. Menze *et al.*, “The multimodal brain tumor image segmentation benchmark (brats),” *IEEE transactions on medical imaging*, vol. 34, no. 10, pp. 1993–2024, 2014.
- [8] D. Sadeghi *et al.*, “An overview of artificial intelligence techniques for diagnosis of schizophrenia based on magnetic resonance imaging modalities: Methods, challenges, and future works,” *Computers in Biology and Medicine*, p. 105554, 2022.
- [9] P. Perona and J. Malik, “Scale-space and edge detection using anisotropic diffusion,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.
- [10] W. Speier *et al.*, “Robust skull stripping of clinical glioblastoma multiforme data,” in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2011: 14th International Conference, Toronto, Canada, September 18–22, 2011, Proceedings, Part III 14*. Springer, 2011, pp. 659–666.
- [11] L. G. Nyúl and J. K. Udupa, “On standardizing the mr image intensity scale,” *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 42, no. 6, pp. 1072–1081, 1999.
- [12] J. Sled *et al.*, “A nonparametric method for automatic correction of intensity nonuniformity in mri data,” *IEEE Transactions on Medical Imaging*, vol. 17, no. 1, pp. 87–97, 1998.
- [13] N. J. Tustison *et al.*, “N4itk: improved n3 bias correction,” *IEEE transactions on medical imaging*, vol. 29, no. 6, pp. 1310–1320, 2010.
- [14] O. Commowick *et al.*, “Block-matching strategies for rigid registration of multimodal medical images,” in *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2012, pp. 700–703.
- [15] R. E. Gabr *et al.*, “Brain and lesion segmentation in multiple sclerosis using fully convolutional neural networks: A large-scale study,” *Multiple Sclerosis Journal*, vol. 26, no. 10, pp. 1217–1226, 2020.
- [16] O. Ronneberger *et al.*, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab *et al.*, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.
- [17] P. Coupé *et al.*, “An optimized blockwise nonlocal means denoising filter for 3-d magnetic resonance images,” *IEEE transactions on medical imaging*, vol. 27, no. 4, pp. 425–441, 2008.
- [18] J. V. Manjón and P. Coupé, “volbrain: an online mri brain volumetry system,” *Frontiers in neuroinformatics*, vol. 10, p. 30, 2016.

⁴https://drive.google.com/drive/folders/1-1wfy5EV_Yhgj0EIHMymD8RTbJAH_NfK?usp=share_link