MRI-BASED DIAGNOSIS OF ALZHEIMERS DISEASE USING DEEP LEARNING WITH CYCLEGAN FOR DATA AUGMENTATION

Sunny Wang, Allen Huang, for the Alzheimer's Disease Neuroimaging Initiative*
{sunnywang272}@gmail.com, {huanga}@mit.edu

August 1, 2020

Abstract. Alzheimer's disease is a progressive disease causing deterioration of neurons in the brain, leading to dementia and eventually death. Diagnosis of Alzheimer's conventionally consists of a combination of neuropsychological tests and laboratory tests such as MRI scans, and early diagnosis and treatment is critical as it can improve the quality of life of patients. As Alzheimer's is associated with loss in brain mass, which can be discerned from MRI scans, it is a suitable task for deep learning and computer vision. Current clinical diagnosis accuracy lies at around 77%, so an accurate and efficient machine learning model could be of great assistance to physicians, as it could reinforce their diagnosis. However, deep learning typically requires large amounts of data, and medical data is often scarce. A recent breakthrough in machine learning, the generative adversarial network (GAN), is a potential solution to lack of data. GANs use two neural networks in conjunction to generate realistic images. In this study, we construct ResNet50-based convolutional neural networks to perform Alzheimer's disease classification using MRI scans. Furthermore, we demonstrate that GANs, specifically the CycleGAN architecture, can significantly improve classification accuracy when used as a method of data augmentation.

Keywords: Alzheimer's, Magnetic Resonance Imaging (MRI), Generative Adversarial Networks (GANs), Convolutional Neural Networks (CNNs), Data Augmentation, Transfer learning, Deep Learning

*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: `http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf`

1. Introduction

1.1. **Background.** Alzheimer's disease (AD) is a progressive disease characterized by the loss of cognitive ability and is the sixth leading cause of death in the United States. The progression can be categorized by severity, consisting of mild, moderate, and severe AD. These stages are typically classified and diagnosed based on a variety of factors, including cognitive tests, interviews with family members, and laboratory tests. Early diagnosis of AD is critical, as it can greatly improve patients' quality of life and in some cases halt or slow the rate of progression. According to a study conducted by Beach et. al. in 2012, there was wide variation in AD diagnosis accuracy, but the overall accuracy was 77% [1]. This is far from perfect, as many cases are misdiagnosed, with a low true negative rate. This raises the demand for a computer assisted tool to reinforce physicians' diagnosis.

As AD causes the breakdown and death of neurons, these changes in brain mass can be observed through technology such as magnetic resonance imaging (MRI), computerized tomography (CT), and positron emission tomography (PET). These scans can both reveal losses in brain mass and rule out other potential causes for the patient's symptoms, such as cancerous tumors. The use of machine learning techniques for image classification has become a popular option due to their impressive accuracy. For example, convolutional neural networks (CNN) have achieved results in classification that even outperform humans in some specialized cases [2]. Diagnosis of AD using machine learning can serve as a powerful tool for physicians, supplying an additional metric for diagnosis.

However, clean and organized medical data is often scarce and is a prominent obstacle preventing more widespread use of machine learning in medical settings. This raises a demand for data augmentation techniques to improve medical machine learning models. One recently introduced technique is the generative adversarial network (GAN) [3], one of the most influential milestones in machine learning. By having two neural networks, a generator and a discriminator, compete against each other, GANs achieve promising results in image generation, super resolution, and data augmentation, among many other applications. The lack of large amounts of medical data leads to significant potential for the use of GANs for data augmentation. Using a relatively small dataset, GANs can generate similar but original images, as opposed to image modifications used in classical data augmentation. In this study, we investigate the potential for using deep learning in

Alzheimer's disease classification by creating a convolutional neural network model. We also test the feasibility of using GANs for data augmentation, specifically using the CycleGAN architecture.

1.2. **Literature Review.**

1.2.1. *Convolutional Neural Networks.* A convolutional neural network (CNN) is a type of neural network that specializes in dealing with pattern recognition in images. Input usually consists of a three-dimensional array containing width, height, and the pixel values in the image. A diagram of the architecture of a convolutional neural network is shown in Figure 1.

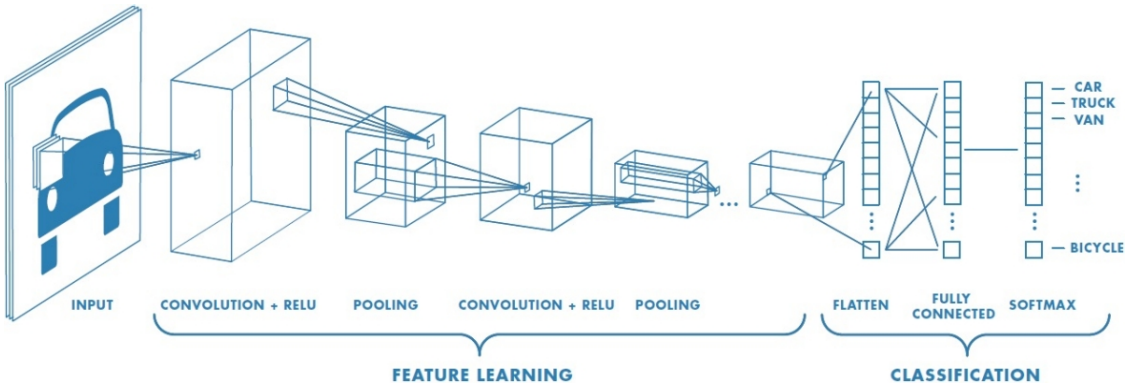

Figure 1. CNN Architecture. Adapted from [4].

CNNs use layers of convolutions, where a filter or kernel is used to create a feature map, allowing the network to sample context within each frame as neighboring pixels are included. These feature maps allow the model to detect low level features within regions. Pooling layers are also added, which create subsamples of the previous layer. In classification tasks, a standard fully-connected feed forward neural network is commonly applied to the flattened final feature map output by the convolutional layers.

1.2.2. *Transfer Learning.* A common method of training image classification networks is through transfer learning, which is the process of using pretrained models to operate on a different task, essentially transferring the knowledge stored in the original network. Common pretrained models include VGG, Inception, or ResNet, all of which have performed

well on datasets such as ImageNet [5]. Transfer learning can often allow for quicker training times along with high accuracy, and is also less prone to overfitting.

1.2.3. *Generative Adversarial Networks.* Generative adversarial networks were first proposed in 2014 by Ian Goodfellow. The paper suggests the simultaneous training of two adversarial networks in the style of a zero-sum minimax game. The goal of the generator is to create images that trick the discriminator into classifying them as real, and does so by taking in random noise and upsampling it through convolutional layers. In contrast, the goal of the discriminator is to correctly classify what is fake data from the generator and what is real data from the training set. This is also usually done through a convolutional neural network, returning a probability that the image is real.

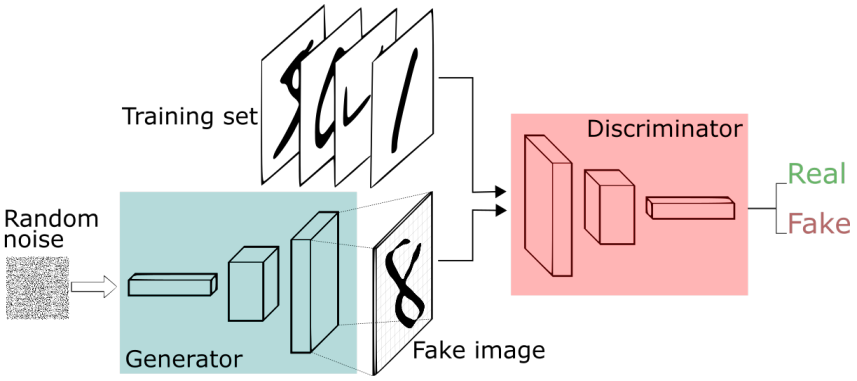

Figure 2. GAN Architecture. Adapted from [6].

The GAN uses the following loss function, where x represents real samples and z represents generated samples.

$$Loss = E_x[\log(D(x))] + E_z[\log(1 - D(G(z)))]$$

The discriminator attempts to maximize D(x), which is the likelihood of a correct sample being classified correctly. The generator wants D(G(z)), the likelihood of the discriminator classifying its generated image as real, to be as high as possible, and thus wants to minimize 1-D(G(z)). This means that the desired G will be minimized and the desired D will be maximized. During each training iteration, the discriminator is updated through gradient ascent in order to maximize the loss function, and the generator is then updated through gradient descent.

1.2.4. *CycleGAN.* Zhu et al. [7] proposed a cycle consistent GAN network that allows for unpaired image to image translation.

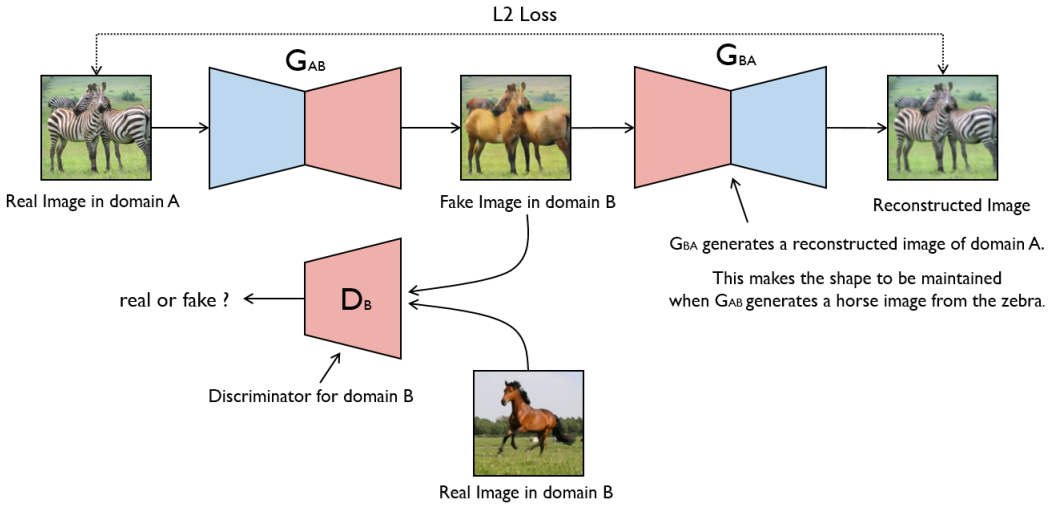

Figure 3. CycleGAN architecture. Adapted from [8].

Consisting of two separate generators and discriminators, CycleGAN allows for discriminatory verification in both directions and uses cycle consistent loss. This constraint essentially allows the generators to learn a spatial transformation from one class to the other. Thus, under ideal conditions, an image that is passed through both generators should return the original image. CycleGAN has been applied to topics such as style transfer and object transfiguration and produces impressive realistic results.

1.2.5. *Data Augmentation.* As stated before, GANs have great potential in data augmentation. Traditional data warping augmentation consists of techniques such as geometric transformations, filters, and random erasing [9]. This can reduce overfitting and improve accuracy. In contrast, GANs are a method of oversampling, as they are able to extrapolate beyond the training set to generate synthetic data, rather than solely modifying existing data within the training set. This is substantially useful in fields lacking large amounts of data, such as in medicine. One instance of GANs being used in augmentation is a study conducted by Frid-Adar et al. proposing the use of synthetic images in liver lesion classification [10]. GAN generated data was used with a CNN for image classification on a dataset

of size 182. Standard DA techniques resulted in 78.6% sensitivity and 88.4% specificity, while the addition of GAN synthesized data yielded 85.7% sensitivity and 92.4%.

1.2.6. *Alzheimer's and Machine Learning.* As MRI scans are an important aspect of the diagnosis of Alzheimer's, there exists substantial literature regarding machine learning methods. The most common model is a convolutional neural network, as they are ideal for image processing. Farooq et al. [11] used a four way CNN classifier between the following classes: normal cognition (NC), early mild cognitive impairment (MCI), late mild cognitive impairment, and Alzheimer's disease (AD). The model was trained on the Alzheimer's disease Neuroimaging Initiative (ADNI) dataset. The only data augmentation performed was flipping images due to the symmetrical nature of MRIs, and specific slices were selected to exclude those without gray matter information. The proposed network based on GoogLeNet and ResNet outperformed other studies done on the same dataset, with about 98% accuracy. Hosseini-Asl et al. [12] used a 3D adaptive convolutional neural network (3D ACNN) on the CADDementia dataset, which could then be generalized to the ADNI dataset, which they used for validation. The model outperformed many state-of-the-art models at task specific classification. Glotzman et al. [13] proposed a network of 2D CNNs, applied to each of three images extracted from each sample in the ADNI dataset. This allowed for the use of two dimensional CNNs on a three dimensional dataset while preserving features. Both MRI scans and PET scans were used, and both two way classification (NC vs AD) and three way classification (NC vs MCI vs AD) were tested. Two way classification with PET-AV 45 scans performed the best, with 83% accuracy.

There have been relatively few applications of generative adversarial networks in classifying AD. A study by Bowles et al. [14] used GANs to model the progression of the disease using MRI data. Using image arithmetic, the model could predict changes in the brain over time and results were comparable to longitudinal examination data. Another study conducted by Pan et al. [15] used GANs to synthesize PET scans from MRI scans in order to fill in gaps in data as many AD patients do not have both due to the high cost of PET scans. A cycle consistent generative adversarial network was used as the first step, which was used in order to learn mappings between the two image domains. The features are then fed into a landmark based multimodal multi-instance learning classifier for diagnosis. Kim et al. [16] conducted a feasibility study on using GANs for slice selective learning on PET

scans. Using a BEGAN, they showed that double slices over the posterior cingulate cortex achieves the best performance, and that two slices performed significantly better than using one slice.

It is also important to consider the practical applications of machine learning in AD diagnosis. A survey conducted by R. Bryan [17] on applying machine learning to AD diagnosis notes that machine learning models are biased by the original population, as it is limited to the dataset that it was trained upon. The overall conclusion reached was that although it is unlikely that machine learning models can replace the skills of radiologists, they can serve as useful tools to complement human skills.

The remainder of this paper is organized as follows. Section 2 describes the methodology and pipeline for model construction. Section 3 presents and discusses the results obtained from using different model architectures and displays the significance of using GAN augmentation. Section 4 and 5 report conclusions and future work respectively.

2. Methodology

Figure 4 outlines the model pipeline, consisting of dataset acquisition, preprocessing (including GAN augmentation), and classification using a CNN.

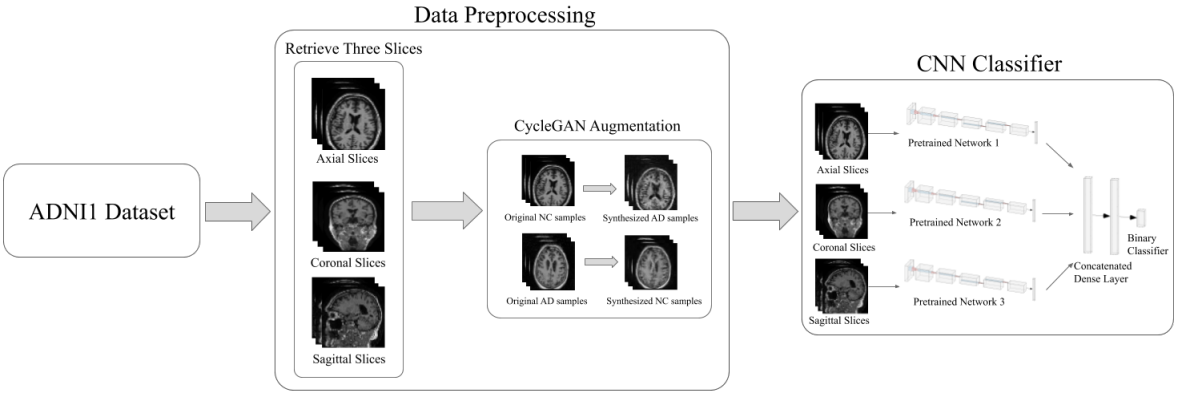

Figure 4. Overview of the model pipeline.

2.1. **Data Acquisition.** Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI),

positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). The specific dataset used for training was the ADNI1 standardized MRI dataset, categorized by severity: AD, MCI, and normal cognition (NC). For the purposes of this study, we trained a network to classify between AD and NC. The dataset contained 705 samples labeled as NC and 476 samples labeled as AD.

2.2. **Data Preprocessing.** The data was stored in NIFTI files, which were converted into three dimensional numpy arrays using nibabel. A csv file containing information about the scans, patients, and ground truth diagnosis labels was also downloaded from the ADNI database. As the NIFTI image data is three dimensional, slicing was required to prepare samples for training. In order to capture as much information from the original image, we extract three slices, one each from the axial, coronal, and sagittal orientations. The slices are taken by retrieving the midpoint of each axis length. Upon taking the slices, the images were resized to 64 x 64.

Two different methods of preprocessing were considered and tested:

(1) Skull stripping was applied using the DeepBrain library.
(2) RAS+ ISO transforms and histogram normalization were performed using the TorchIO library [18].

2.3. **GAN based Data Augmentation.** We constructed the CycleGAN models using the implementation from [7]. This architecture allows for image synthesis of one class from the other, and we used the network to generate sufficient samples to create a balanced dataset. The original dataset was split according to the label and randomly paired. Three individual CycleGAN models were created, where each one was trained on data from a different MRI slice. Each model used the Adam optimizer with learning rate of 2e-4, and were trained for 100 epochs with a batch size of 1, as specified in the CycleGAN paper.

Using CycleGAN, an AD version of each NC sample was generated and vice versa. A total of 705 AD samples and 476 NC samples of each orientation were generated, for a total of 1181 images of each class, as shown in Table 1.

	Normal Samples	Alzheimer's Samples	Total
Dataset	705	476	1181
GAN generated	476	705	1181
Total	1181	1181	2362

TABLE 1. Dataset sizes after GAN augmentation.

2.4. **Convolutional Neural Network Classifier.** We used a transfer learning approach to create the model architecture as it would save training time and is generally effective when datasets are small. We used three well established convolutional neural networks (CNNs) for image classification for the pretrained model: ResNet50, VGG16, and GoogLeNet. Each of these models have achieved state of the art performance on the ImageNet database.

The pretrained ResNet50 architecture takes in 3-channel RGB images while the MRI scans are grayscale. To match the networks, the one channel images were transformed to three channels by stacking the tensor three times across dimension 0. The first layer of each pretrained network was also modified to match the image size of 64 x 64 and the last layer was modified to become a binary classifier. Each neural network was fine-tuned using the Adam optimizer with a learning rate of 0.0001, and trained for 50 epochs with a batch size of 32. A training, validation, and testing split of 80%-10%-10% was used. Both models were evaluated based on accuracy, precision, recall, and F1 score.

The single input model used the middle axial slice from each NIFTI volume as input to a pretrained network. The architecture is shown in Figure 5.

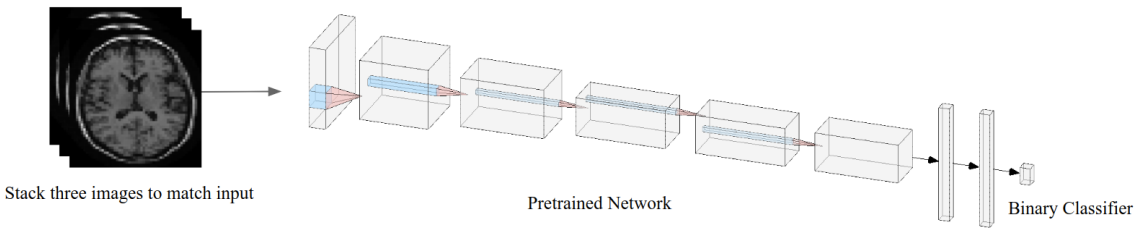

Figure 5. Single CNN architecture.

In order to better encapsulate the volumetric data, we also used a modified CNN with multiple inputs. The model architecture consists of three CNN models, where outputs from each individual CNN are concatenated and passed through fully connected layers, which returns the diagnosis group. The model architecture is shown in Figure 6.

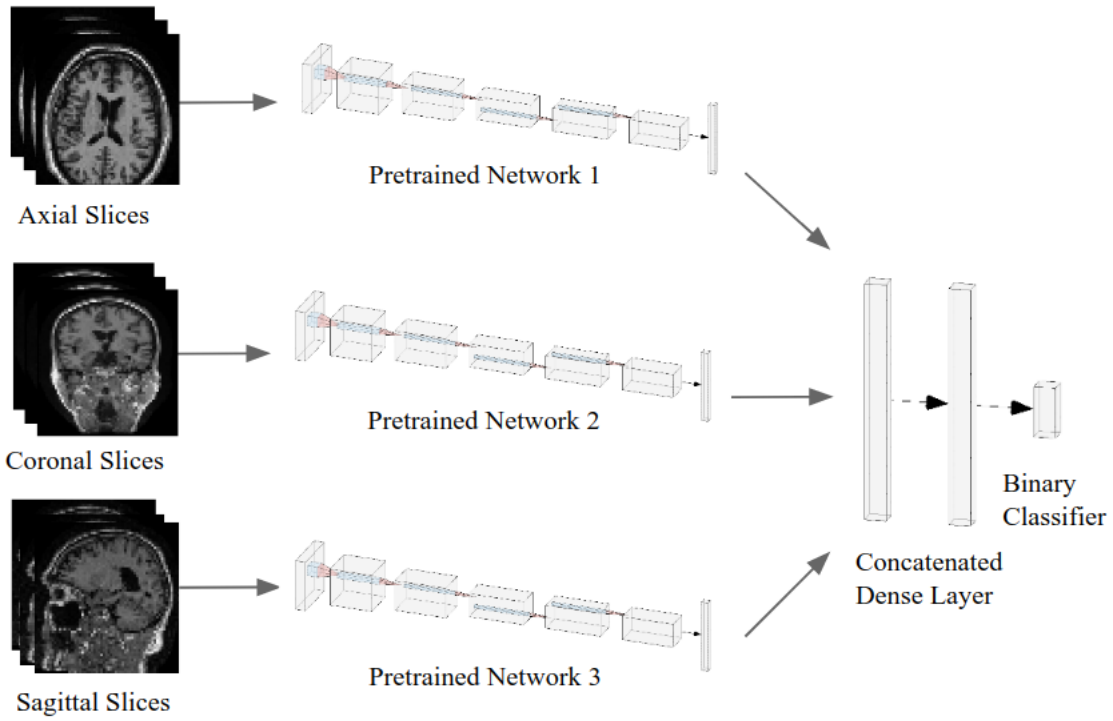

Figure 6. Multiple CNN architecture.

2.5. **Model Implementation.** All discussed networks were implemented using Python 3.7 with the PyTorch library. All training was done on a personal computer with an AMD 3700X CPU and an NVIDIA RTX 2070 GPU.

3. Results and Discussion

3.1. **Single Slice vs Multiple Slice Results.** Table 2 shows the metrics on the test set for both proposed architectures.

Metric	ResNet50 with One Input	ResNet50 with Three Inputs
Accuracy	0.748	**0.815**
Precision	0.660	**0.804**
Recall	0.717	**0.804**
F1 Score	0.688	**0.804**

TABLE 2. Comparison between one input and three input architectures.

The modified ResNet50 model with three inputs outperformed the standard one input ResNet50 model in every metric. As such, the use of multiple slices proves to be effective at encapsulating additional data about the brain.

3.2. **Comparison of Preprocessing Methods.** Two methods of preprocessing was tested, skull stripping and TorchIO transforms. Figures 7 and 8 display the resulting images after applying preprocessing.

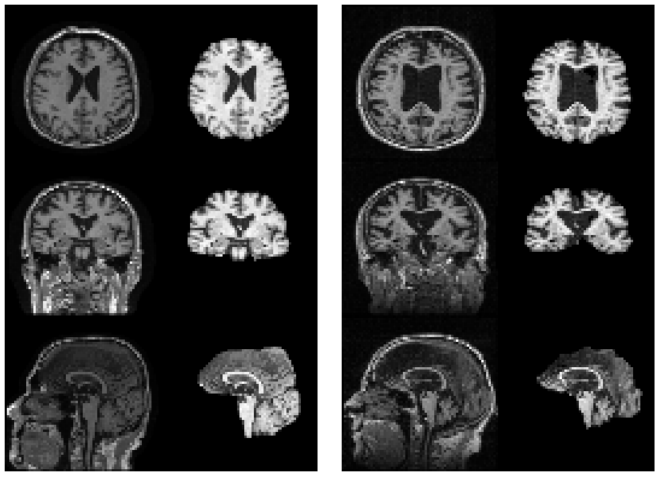

Figure 7. Normal sample (left) and Alzheimer's sample (right). The skull stripped versions are displayed to the right of the original versions.

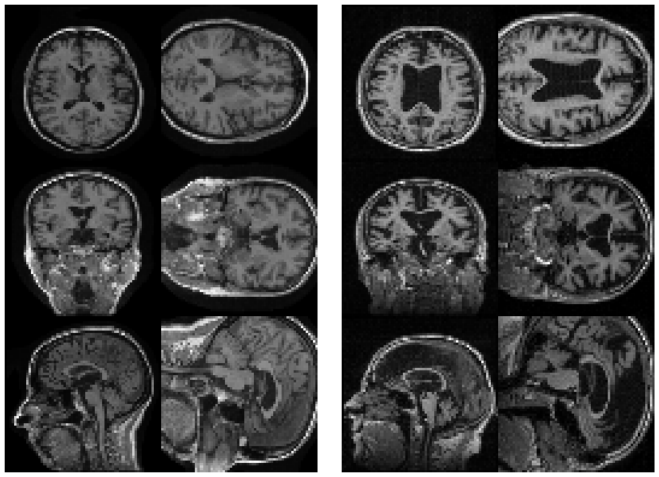

Figure 8. Normal sample (left) and Alzheimer's sample (right). The TorchIO transformed versions are displayed to the right of the original versions.

Table 3 compares the results from applying different methods of preprocessing on the three input ResNet50 network as shown in Figure 6. The different transforms were not compatible with each other, so results were obtained separately. Only the model utilizing skull stripping outperformed the unmodified model, so it was kept for the remainder of the study.

Metric	ResNet50	+ Skull Stripping	+ TorchIO
Accuracy	**0.824**	0.815	0.807
Precision	**0.854**	0.804	0.761
Recall	0.7	**0.804**	0.745
F1 Score	0.769	**0.804**	0.753

TABLE 3. Comparison of ResNet50 networks with different preprocessing.

3.3. **CycleGAN Generation Results.** Examples of CycleGAN generated images are shown in Figure 9.

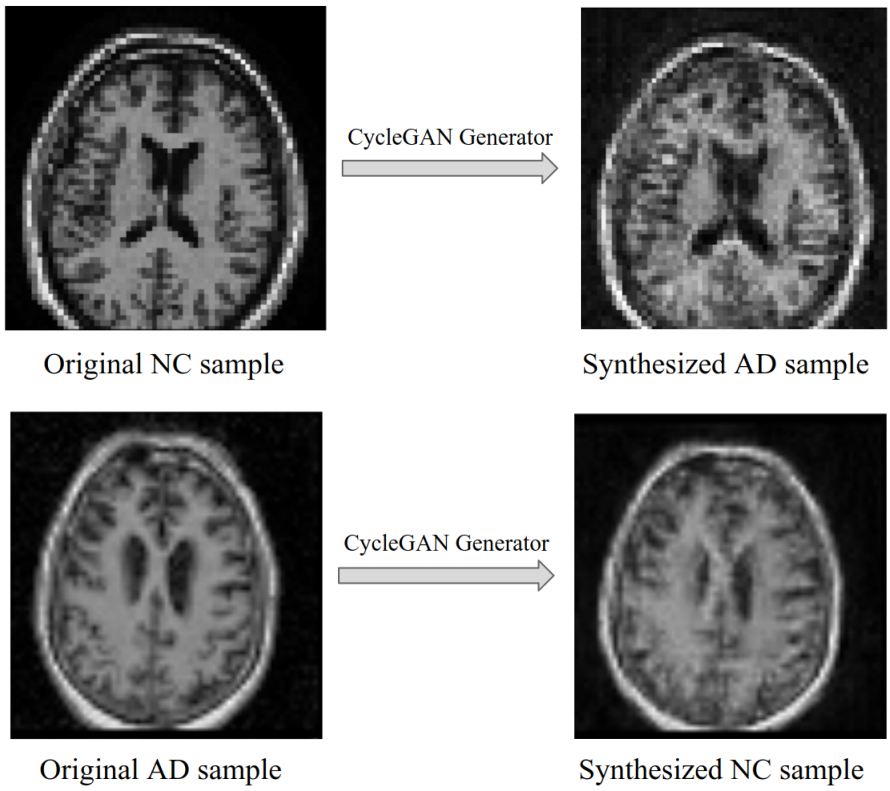

Figure 9. CycleGAN image synthesis.

By observation, the synthesized Alzheimer's sample displays more dark space throughout the brain when compared to the normal sample that it was transformed from, which is an indication of loss of brain mass and a characteristic of Alzheimer's disease. On the contrary, the synthesized normal sample on the bottom right has much less dark space.

3.4. **Comparison with GAN Augmentation.** When using CycleGAN for augmentation, an additional 705 AD samples and 476 NC samples of each orientation were generated, for a total of 1181 images of each class. Tables 4 and 5 display the results when GAN augmentation is applied to the single input and multiple input CNNs.

As shown in Table 4, when GAN augmentation was applied to the single input CNN, there was significant improvement across all metrics. The F1 score increased from 0.688 to 0.81, an 18% increase in performance.

Metric	ResNet50	ResNet50 + GAN
Accuracy	0.748	**0.785**
Precision	0.660	**0.732**
Recall	0.717	**0.908**
F1 Score	0.688	**0.810**

TABLE 4. Comparison of single input CNNs with GAN augmentation.

Table 5 shows that there was also a similar increase in performance for the multiple input CNN model when GAN augmentation is applied. The F1 score for the standard ResNet50 increased from 0.769 to 0.905, an 18% increase. The F1 score for the ResNet50 using skull stripping increased from 0.804 to 0.918, an 14% increase. Overall, these results indicate that CycleGAN can be used effectively as a method of data augmentation.

Metric	ResNet50	ResNet50 + GAN	ResNet50 + SS	ResNet50 + SS + GAN
Accuracy	0.824	0.899	0.815	**0.920**
Precision	0.854	0.870	0.804	**0.939**
Recall	0.7	**0.942**	0.804	0.899
F1 Score	0.769	0.905	0.804	**0.918**

TABLE 5. Comparison of three input CNNs with skull stripping and GAN augmentation.

It is important to note that the results presented were from datasets that were down-sampled from 224 x 224 to 64 x 64 due to computational limitations with training GANs at high resolutions. Thus, only 64 x 64 results were used for comparison with GAN augmentation. However, CNN classification results were obtained for both resolutions for comparison purposes. As displayed in Table 6, when using the same architecture, the images at their a higher 224 x 224 resolution perform significantly better, likely due to their ability to store more information.

Image Dimension	Metric	ResNet50	w/ Skull Stripping	w/ TorchIO	w/ GAN
64x64	Accuracy	0.824	0.815	0.807	**0.920**
	Precision	0.854	0.804	0.761	**0.939**
	Recall	0.7	0.804	0.745	**0.899**
	F1 Score	0.769	0.804	0.753	**0.918**
224x224	Accuracy	0.891	**0.908**	0.907	-
	Precision	**0.932**	0.882	0.897	-
	Recall	0.804	**0.900**	0.854	-
	F1 Score	0.863	**0.891**	0.875	-

TABLE 6. Comparison between different image sizes.

The improvement in performance on the 64 x 64 images can likely be extended to 224 x 224 images, but its feasibility was not directly tested in this study. Data augmentation of full-sized images remains a promising area for future improvement of our results.

4. CONCLUSIONS

In this study, we constructed convolutional neural network models utilizing the ResNet50 architecture to diagnose Alzheimer's disease using MRI scans, with variants using one input and three inputs. We also address the problem of size limitations in medical datasets with the use of generative adversarial networks (GANs). Using the ADNI1 dataset, we demonstrated that the addition of GANs can greatly improve deep learning classification accuracy for Alzheimer's disease diagnosis. Specifically, we used CycleGAN to generate images of one class using the other, balancing the dataset and increasing its overall size. Our results show that in both the single input and three input CNNs, classification accuracy improved substantially. Due to the lack of large datasets in many medical fields, the results obtained in this study can be generalized to many other fields as well. Overall, with promising results in data augmentation, GANs have potential to significantly improve differential classification across a wide variety of applications.

5. Future Work

5.1. **Preprocessing.** Upon inspection, the current methods of preprocessing yield some inconsistency between images, particularly in axial scans, due to subjects having slight variation in brain shape. In the future, other methods of preprocessing could be used to to improve consistency across each slice taken. One method could be to select the slices with the highest average pixel intensity in each direction, which should theoretically select the same slice along each axis. Software such as statistical parametric mapping could also be applied to perform preprocessing such as gray matter segmentation and modulation to reduce variation between images.

5.2. **MRI and PET Fusion.** Diagnosis of Alzheimer's Disease often uses both MRI scans and PET scans. In contrast to MRIs, which provide identification of abnormalities in the brain, PET scans can show areas of low metabolism, allowing for differentiation between Alzheimer's and other types of dementia. Using a fused image with both scans in conjunction would contain more features and will likely improve classification accuracy.

5.3. **External features.** A full Alzheimer's Disease diagnosis requires many elements other than neuroimaging, such as mental status tests and physical exams. Incorporating all of these elements would greatly improve the accuracy of a machine learning system.

Disclosure Statement. The authors have no conflicts of interest to declare.

Acknowledgments. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.;Janssen Alzheimer Immunotherapy Research Development, LLC.; Johnson Johnson Pharmaceutical Research Development LLC.; Lumosity; Lundbeck; Merck Co., Inc.;Meso

Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Appendix A. Supplemental Figures

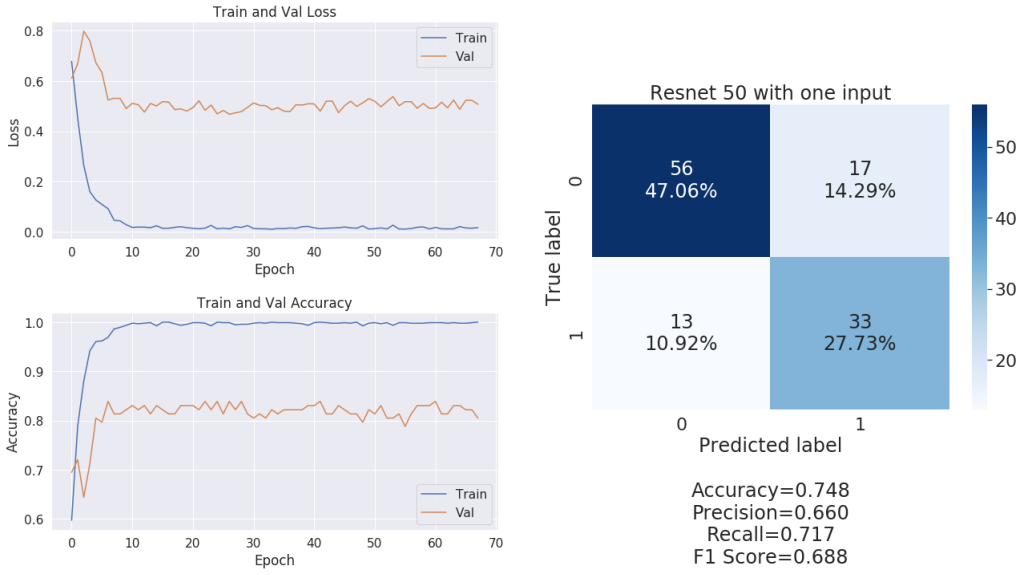

Figure 10. Results for ResNet50 network using a single image as input.

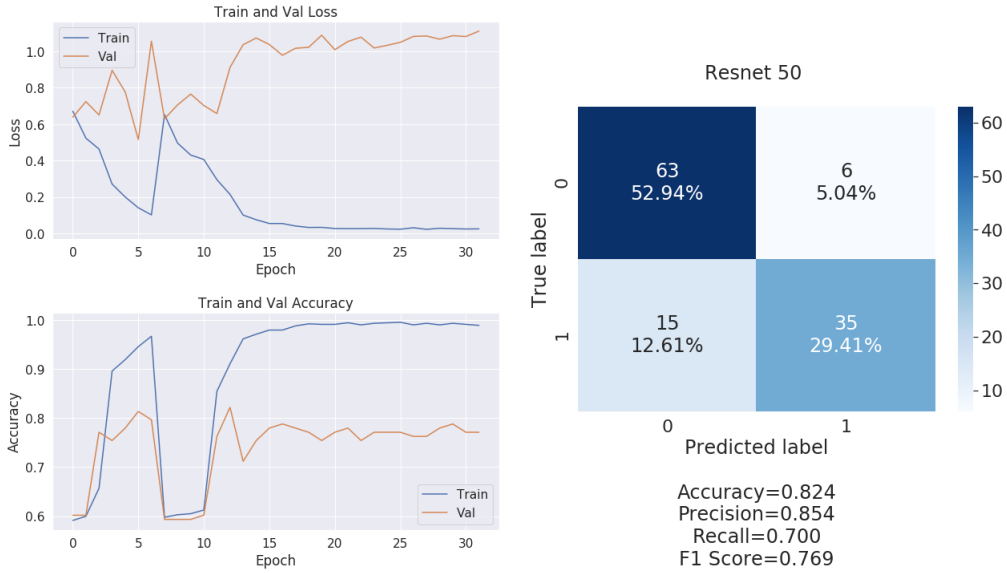

Figure 11. Results for ResNet50 network using a three images as input.

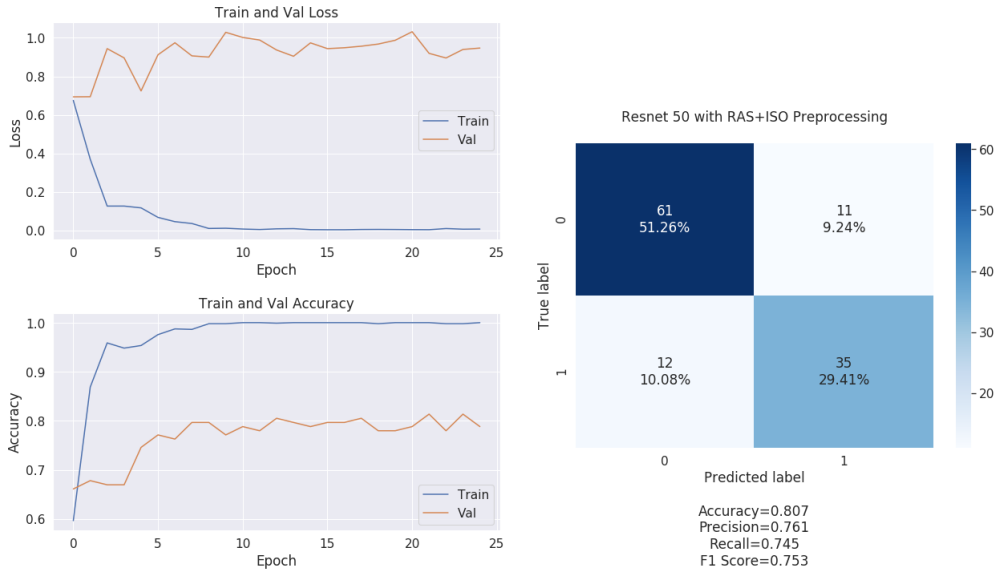

Figure 12. Results for ResNet50 with TorchIO preprocessing.

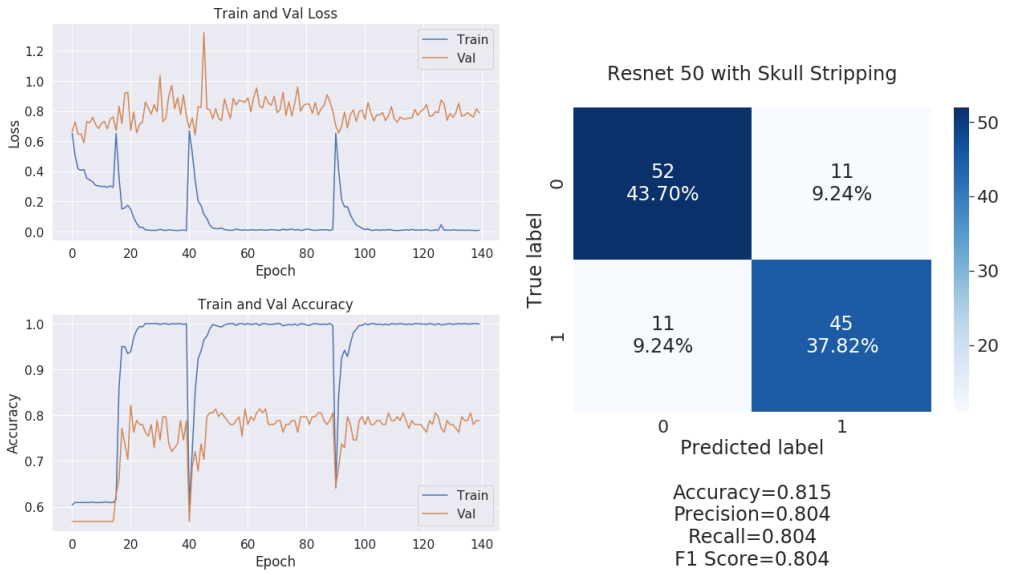

Figure 13. Results for ResNet50 with skull stripping preprocessing.

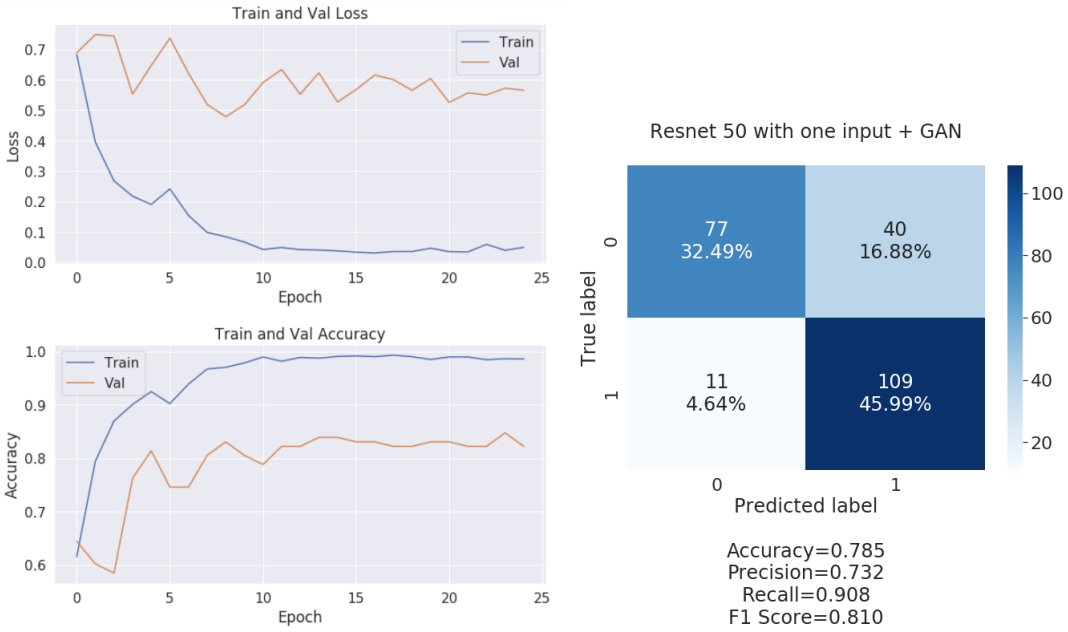

Figure 14. Results for single input ResNet50 with GAN augmentation.

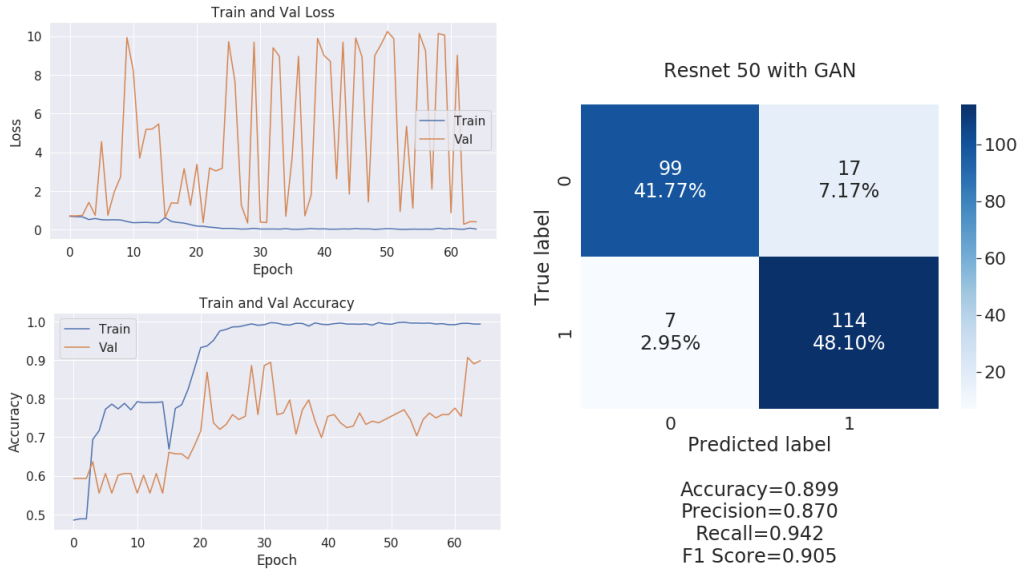

Figure 15. Results for single input ResNet50 with GAN augmentation.

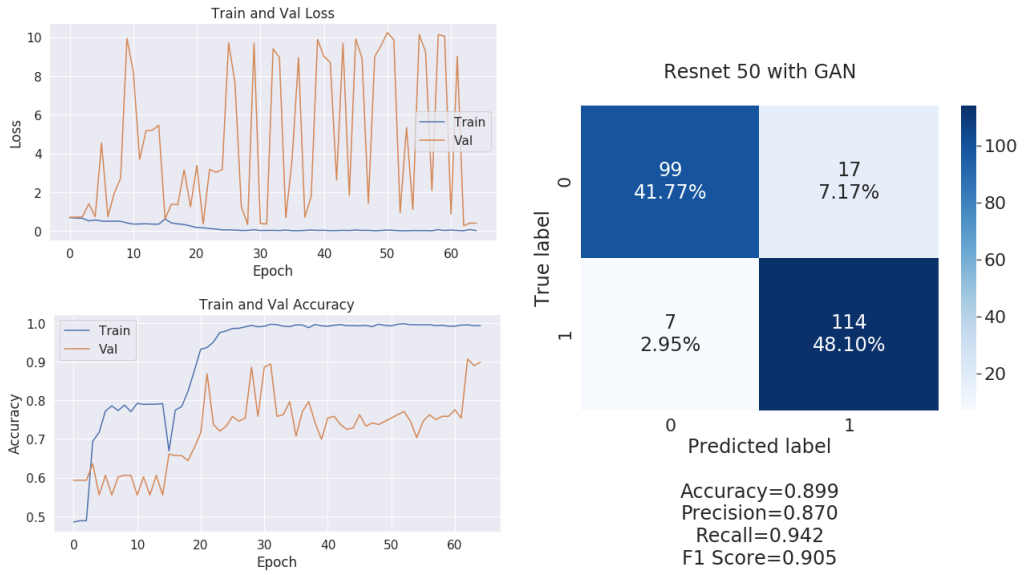

Figure 16. Results for three input ResNet50 with GAN augmentation.

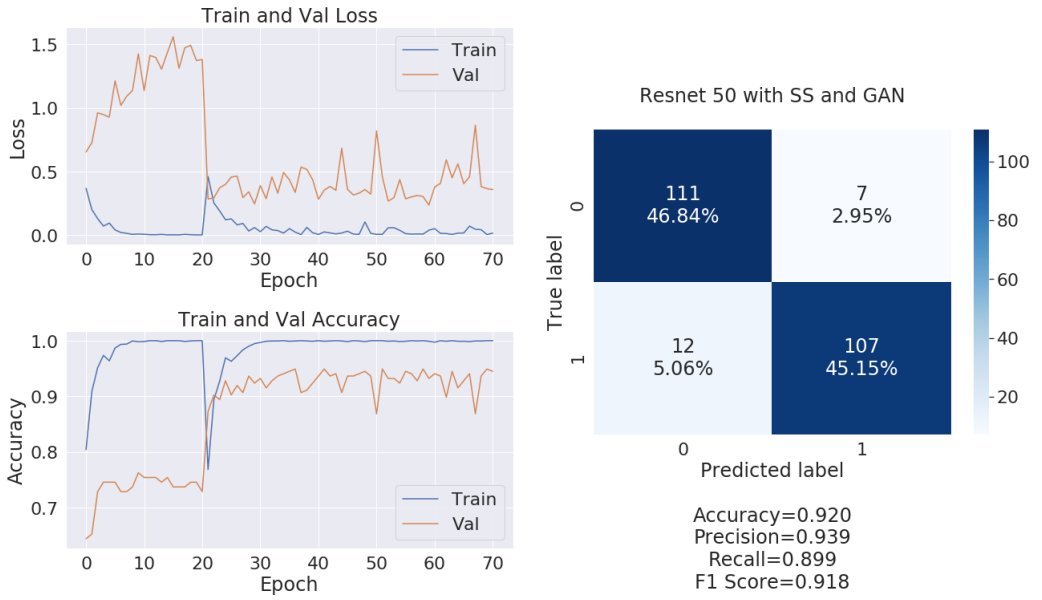

Figure 17. Results for three input ResNet50 with skull stripping and GAN augmentation.

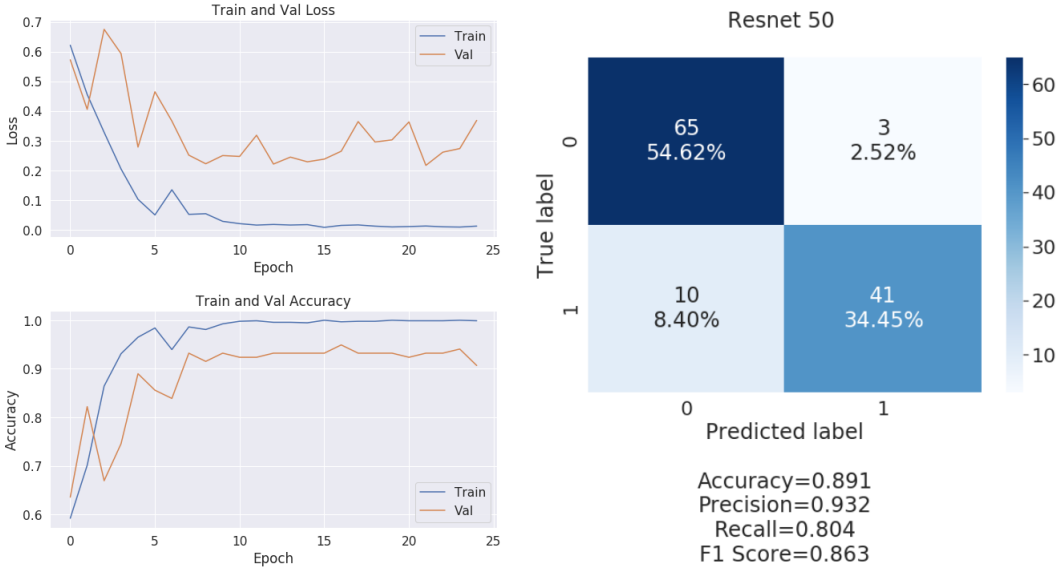

Figure 18. Results for 224 x 224 ResNet50.

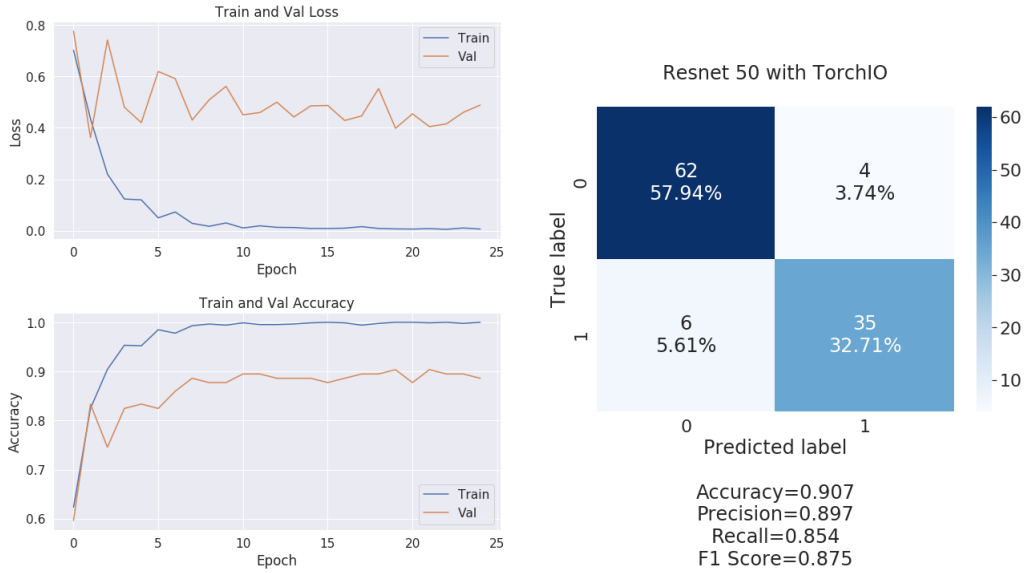

Figure 19. Results for 224 x 224 ResNet50 with TorchIO preprocessing.

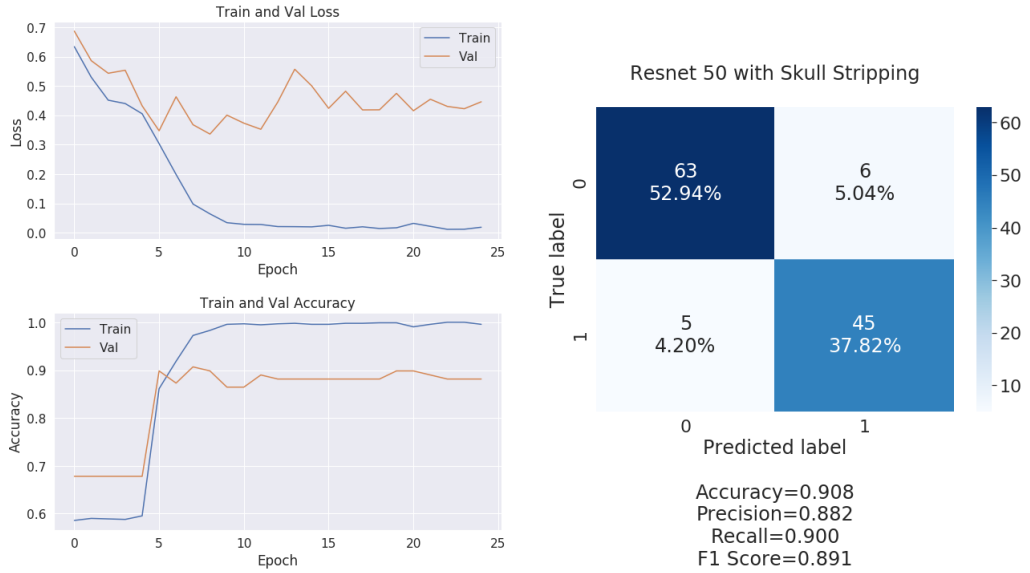

Figure 20. Results for 224 x 224 ResNet50 with skull stripping preprocessing.

References

[1] Thomas G Beach et al. "Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005–2010". In: *Journal of neuropathology and experimental neurology* 71.4 (2012), pp. 266–273.

[2] Antoine Buetti-Dinh et al. "Deep neural networks outperform human expert's capacity in characterizing bioleaching bacterial biofilm composition". In: *Biotechnology Reports* 22 (2019), e00321.

[3] Ian Goodfellow et al. "Generative adversarial nets". In: *Advances in neural information processing systems*. 2014, pp. 2672–2680.

[4] *Convolutional Neural Network*. URL: `https://www.mathworks.com/solutions/deep-learning/convolutional-neural-network.html`.

[5] Rajat Vikram Singh. "ImageNet Winning CNN Architectures–A Review". In: ().

[6] freeCodeCamp.org. *An intuitive introduction to Generative Adversarial Networks (GANs)*. Apr. 2020. URL: `https://www.freecodecamp.org/news/an-intuitive-introduction-to-generative-adversarial-networks-gans-7a2264a81394/`.

[7] Jun-Yan Zhu et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2223–2232.

[8] URL: `https://modelzoo.co/model/mnist-svhn-transfer`.

[9] Connor Shorten and Taghi M Khoshgoftaar. "A survey on image data augmentation for deep learning". In: *Journal of Big Data* 6.1 (2019), p. 60.

[10] Maayan Frid-Adar et al. "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification". In: *Neurocomputing* 321 (2018), pp. 321–331.

[11] Ammarah Farooq et al. "A deep CNN based multi-class classification of Alzheimer's disease using MRI". In: *2017 IEEE International Conference on Imaging systems and techniques (IST)*. IEEE. 2017, pp. 1–6.

[12] Ehsan Hosseini-Asl, Robert Keynton, and Ayman El-Baz. "Alzheimer's disease diagnostics by adaptation of 3D convolutional network". In: *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2016, pp. 126–130.

[13] Tanya Glozman and Orly Liba. *Hidden cues: Deep learning for Alzheimer's disease classification CS331B project final report*. 2016.

[14] Christopher Bowles et al. "Modelling the progression of Alzheimer's disease in MRI using generative adversarial networks". In: *Medical Imaging 2018: Image Processing*. Vol. 10574. International Society for Optics and Photonics. 2018, 105741K.

[15] Yongsheng Pan et al. "Synthesizing missing PET from MRI with cycle-consistent generative adversarial networks for Alzheimer's disease diagnosis". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2018, pp. 455–463.

[16] Han Woong Kim et al. "Slice-selective learning for Alzheimer's disease classification using a generative adversarial network: a feasibility study of external validation". In: *European Journal of Nuclear Medicine and Molecular Imaging* (2020), pp. 1–10.

[17] R Nick Bryan. *Machine learning applied to Alzheimer disease*. 2016.

[18] Fernando Pérez-Garcıa, Rachel Sparks, and Sebastien Ourselin. "TorchIO: a Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning". In: *arXiv preprint arXiv:2003.04696* (2020).