

Article

Anomaly Analysis of Alzheimer's Disease in PET Images Using an Unsupervised Adversarial Deep Learning Model

Husnu Baris Baydargil ¹, Jang-Sik Park ^{1,*} and Do-Young Kang ^{2,3,4,*}

¹ Department of Electronics Engineering, Kyungsoo University, Busan 48434, Korea; baris.baydargil@gmail.com

² Institute of Convergence Bio-Health, Dong-A University, Busan 49315, Korea

³ Department of Nuclear Medicine, College of Medicine, Dong-A University, Busan 49201, Korea

⁴ Department of Translational Biomedical Sciences, College of Medicine, Dong-A University, Busan 49201, Korea

* Correspondence: jsipark@ks.ac.kr (J.-S.P.); dykang@dau.ac.kr (D.-Y.K.)

Abstract: In this study, the anomaly analysis of Alzheimer's disease using positron emission tomography (PET) images using an unsupervised proposed adversarial model is investigated. The model consists of three parts: a parallel-network encoder, which is comprised of a convolutional pipeline and a dilated convolutional pipeline that extracts global and local features and concatenates them, a decoder that reconstructs the input image from the obtained feature vector, and a discriminator that distinguishes if the input image is real or fake. The hypothesis is that if the proposed model is trained with only normal brain images, the corresponding construction loss for normal images should be minimal. However, if the input image belongs to a class that is designated as an anomaly that which the model is not trained with, then the construction loss will be high. This will reflect during the anomaly score comparison between the normal and the anomalous image. A multi-case analysis is performed for three major classes using the Alzheimer's Disease Neuroimaging Initiative dataset, Alzheimer's disease, mild cognitive impairment, and normal control. The base parallel-encoder network shows better classification accuracy than the benchmark models, and the proposed model that is built on the parallel model outperforms the benchmark anomaly detection models. The proposed model gave out 96.03% and 75.21% in classification and area under the curve score, respectively. Additionally, a qualitative evaluation done by using Fréchet inception distance gave a better score than the state-of-the-art by three points.

Keywords: computer vision; Alzheimer's disease; deep learning; anomaly detection; deep adversarial networks



Citation: Baydargil, H.B.; Park, J.-S.; Kang, D.-Y. Anomaly Analysis of Alzheimer's Disease in PET Images Using an Unsupervised Adversarial Deep Learning Model. *Appl. Sci.* **2021**, *11*, 2187. <https://doi.org/10.3390/app11052187>

Academic Editor: José Manuel Ferreira Machado

Received: 26 January 2021

Accepted: 25 February 2021

Published: 2 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Alzheimer's disease (AD) is by far the most common type of dementia, generally seen in elderly people. For the majority of cases, AD symptoms begin to appear during the mid-60s, although the early-onset may occur in ages as early as the 30s. It is estimated that around 106 million people will be diagnosed in the world by 2050 due to the increase in the aging population [1].

During the progression of AD, the brain structure changes due to the deposition of amyloid- β ($A\beta$) plaques and hyperphosphorylated tau. The initial damage starts at the hippocampus [2], which handles episodic and spatial memory as well as working as a relay between the brain and the rest of the body, so the damage disrupts these pathways. The symptoms of AD are the result of these $A\beta$ plaques and intracellular neurofibrillary tangles [3,4]. This decreases the brain metabolism of both glucose and oxygen, leading to progressive memory loss and inability to move in the late stages [5]. These changes start to form years before the initial clinical symptoms of AD are seen. Mild cognitive impairment (MCI) causes cognitive changes to the person that is noticeable by the family

members and friends, but the person may still carry out daily activities. It's been estimated that 15% to 20% of people over the age of 65 have MCI [6]. Not every case progresses into dementia; in some individuals, MCI remains stable throughout their lives [7]. The stages of AD progress can be seen as low-intensity brain cell activity in medical images. Positron emission tomography (PET) is an imaging technique that uses radiotracers to measure the metabolic processes, and magnetic resonance imaging (MRI) is another imaging technique that uses strong magnetic fields, magnetic field gradients, and radio waves. Resting-state functional magnetic resonance imaging (fMRI) is also considered a promising biomarker; however, in more severe cases, may cause problems due to the fMRI being very sensitive to head motion, mitigating its usefulness [8]. PET imaging provides substantially sensitive assays, easily detecting a very low concentration of molecules of interest labeled through positron emitters, and MRI suffers from very low molar sensitivity for different metabolites and probes. This, coupled with the absolute quantification of substrate concentration being more challenging with MRI, PET imaging is generally chosen for AD imaging [9]. Different types of nuclear tracers are developed for various purposes to highlight different parts of the body; some of the radiotracers that are used to detect $A\beta$ plaques in the brain are florbetapir 18F, flutemetamol F18, PiB, and florbetaben 18F [10]. Unfortunately, at the time of writing this paper, there is no definite cure for AD, and most of the treatments aim at alleviating the disease-related symptoms.

Beginning from the early 1970s, multiple computer-aided diagnostics (CAD) systems were proposed. Their early versions were usually based on manually extracted feature vectors [11]. Later, these vectors were trained in a supervised method with a machine-learning model such as a support vector machine (SVM) [12] to classify the input vector. It was soon realized that there were many shortcomings to such systems [13]. Researchers turned towards data mining approaches in the 1980 and 1990s to develop more robust and flexible systems to overcome these limitations.

With the improvements to imaging methodologies, machine learning methodologies have been proposed to detect AD. These studies generally have multiple steps: feature extraction, feature selection, dimensionality reduction, and feature-based classification algorithm selection [14]. The main issue with such works has been the reproducibility of these approaches [15]. As an example, during the feature selection process, features that are related to AD are chosen from various modalities which may include cortical thickness, brain glucose metabolism, subcortical volumes, and cerebral $A\beta$ accumulation in regions of interest (ROI), such as the hippocampus [16]. Recent works make use of multi modalities of MRI, PET, and fMRI images, and use various preprocessing techniques such as segmentation [17] to improve the accuracy of their proposed models [18].

Recent developments in deep learning technology have led to a massive surge of constantly-improving models, especially in the computer vision area, mainly due to the improvements brought by convolutional neural networks (CNNs). These models learn to extract meaningful features from inputs and produce highly accurate results. The replacement of human experts with automated systems in the near future is being discussed due to the reduced cost and relatively similar performance [19].

Anomaly detection is defined as the recognition of an unexpected pattern that is significantly different from the rest of the data. The majority of the challenges include a proper feature extraction method, imbalance in the data distribution, variance in anomalous cases, and environmental conditions. In the field of computer vision, anomaly detection is recently gaining more attention due to the advancements in the machine learning area. A key handicap with the public datasets is that most of these datasets have an imbalance among the classes they contain. The bias makes using such datasets difficult, especially in medical imaging, causing trained models to have less than desired performance [20]. Furthermore, manual labeling of the medical imaging data is a costly and labor-intensive process, and considering the deep learning models thriving with large amounts of data, developed systems may end up with limited utility and poor generalization [21]. In such scenarios, an unsupervised approach to data distribution modeling can be feasible, where

the majority of the normal case data is used for training, and abnormal case data and the remainder of the normal case data are used to find outlier cases during the inference. For a variety of domains [22–24], different approaches to anomaly detection [25–27] have been proposed in the past. It is generally assumed that the anomalies differ in lower dimensions as well as in high-dimensional space, meaning the latent space mapping is a vital point in anomaly detection. Recent studies that include generative adversarial networks (GANs) [28] in their models are highly effective in mapping the data distribution. GANs being efficient in mapping both high-dimensional and low-dimensional features with very little information loss has sparked a new interest in anomaly detection works [29].

In medical imaging, GANs are generally used for AD diagnosis [30], and data generation to be used in training the deep learning models [31]. Anomaly detection can be used to differentiate normal brain images and the abnormal brain images is not very well researched.

In this study, the analysis of Alzheimer’s disease as an anomaly using the proposed model is researched. In the Alzheimer’s Disease Neuroimaging Initiative (ADNI) dataset, there are three classes which are AD, MCI, and normal control (NC). The proposed model is trained using the unsupervised method. The anomaly analysis is performed on these classes in the following cases: AD-NC, MCI-NC, AD-MCI, the initial class is the anomaly class, and the latter class, the normal class. The area under the curve (AUC) is calculated for each of these comparisons to evaluate the performance of the model, as per previous work in the field [32,33], as well as Fréchet Inception Distance as a qualitative evaluation. Contributions of this paper are as follows:

- Novelty—the Alzheimer’s disease anomaly analysis of PET images using a proposed unsupervised adversarially trained model with a unique feature extractor model. To the authors’ knowledge, there are no anomaly detection studies in Alzheimer’s disease cases using adversarial deep learning models.
- Effectiveness—the proposed model quantitatively and qualitatively outperforms the state-of-the-art models.

2. Related Anomaly Detection Works

For better observation of the changes in the brain caused by AD, there has been a tremendous effort in the medical imaging field. Multiple machine learning applications have been proposed to classify different stages of AD using brain images [34–36]. Different imaging techniques such as PET [37,38], structural magnetic resonance imaging (sMRI) [39,40], functional magnetic resonance imaging (fMRI) [18,41,42] have been used in AD diagnosis. It has also been shown that multi-modal use may increase the performance compared to a single modal [43]. For example, using the ADNI dataset, Lie et al. [44] propose a multi-modality CNN model for binary classification of AD vs. NC with a classification score of 93.26%. Singh et al. [45] achieved a remarkable 97.37% F1 score with their proposed method. Various implementations of GANs in medical imaging have also opened new possibilities from image segmentation [46] to image translation [47].

GANs are a type of generative machine learning framework that includes a generator that creates realistic images, usually from random noise, and a discriminator that identifies if the input image is real or fake. The generator is usually a decoder that learns the input data distribution from the latent vector, and the discriminator generally has a classic architecture that reads the input image and discriminates fake images from real images. The comparison between the novel GAN framework and the proposed model can be seen in Figure 1.

Due to their potential uses, GANs have been examined in great depth [48] and many approaches have been proposed to improve their stability [49]. A well-known work, called Deep Convolutional GAN (DCGAN) [50], removes fully-connected layers and makes use of convolutional layers and batch normalization [51]. With the use of Wasserstein loss, the performance is improved even further [52].

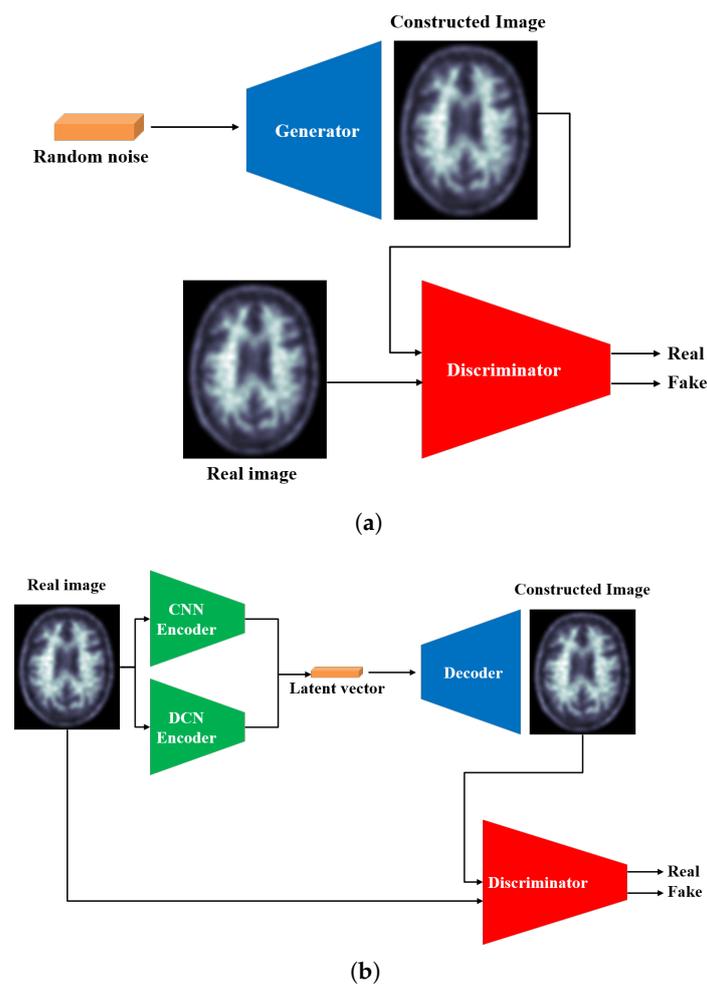


Figure 1. The comparison between the novel generative adversarial network (GAN) and the proposed model architecture. (a) GAN takes a random noise as an input, reconstructs the image using a decoder and the discriminator distinguishes between the normal input image and the reconstructed image. (b) The proposed model takes a real input image, obtains the latent vector using the embedded parallel network that is comprised of a conventional convolutional neural network (CNN) pipeline, and a dilated convolutional neural network (DCN) pipeline, and reconstructs the image. The discriminator then distinguishes between the normal input image and the reconstructed image. More detailed information about the proposed model can be found in Section 3 and the details of the proposed architecture can be seen in Figure 2. (a) The novel GAN architecture. The input is a random noise vector that is built into an image with the generator. The constructed image and a real image from the dataset are then fed to the discriminator where they can be classified as real or fake. (b) The proposed model architecture. The input is an actual image that is sent through the parallel network to generate a latent vector which will then be used to reconstruct the image. This reconstructed image and the corresponding real image are fed to the discriminator where the discrimination between the real and the fake image can be made.

Recent attention on anomaly detection tends to be reconstruction-based. One study from Ravanbaksh et al. [53] takes advantage of image-to-image translation [54] to detect abnormality in crowded scenes. The approach uses two conditional-GANs; the first generator constructs an optical flow from input frames, and the second one generates frames from the generated flow. The main breakthrough comes from Schlegl et al.'s work [23] where it is hypothesized that the end latent vector is a direct representation of the data distribution; however, mapping it is not a straightforward process. The first step is training a generator and a discriminator using only normal images, and afterward, remapping to the latent vector by freezing the weights based on the z vector. The model

shows the anomaly score during the inference by pinpointing an anomaly. Furthering this study, Zenati et al. [32] (EGBAD) uses BiGAN [55], examining joint training to map the data distribution end from the image and latent space, respectively. Akcay et al. [29] propose the use of a conventional autoencoder (GANomaly) with the addition of another encoder the decoder to jointly train the model and the discriminator with the additional latent space loss. Furthering their work, Akcay et al. [56] propose a U-Net-like [57] autoencoder model (Skip-GANomaly) with skip-connections trained jointly with a discriminator.

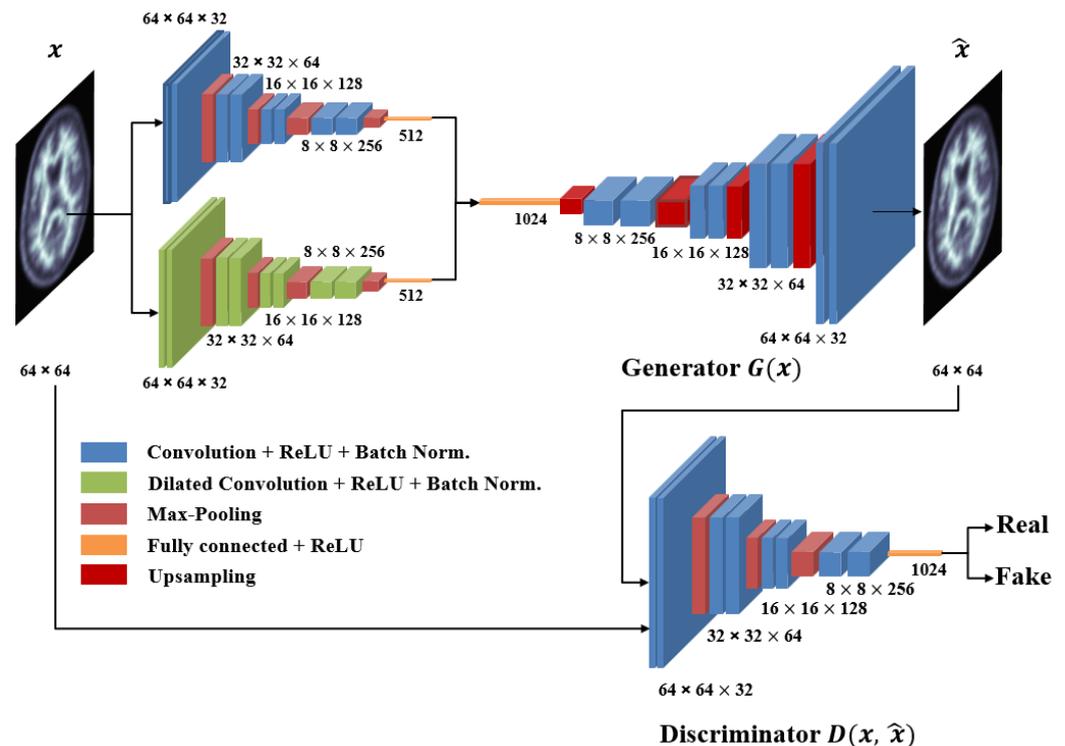


Figure 2. The architecture of the proposed model. The encoder part of the generator G is built on a unique parallel model that is capable of extracting local features from an input image x through the convolutional neural network (CNN), and global features through the DCN. Concatenation of these features z is used as the input for the generator which in turn reconstructs the original input image as \hat{x} . The discriminator D is an encoder that inputs both the original and the reconstructed image and outputs a class label if the image is real or fake. Its fully connected layer \hat{z} is also used as a part of the main loss function during training.

The human brain has a structure with many unique features that can be extracted by different CNN models. However, the majority of the works focus on creating a single complicated pipeline to extract these features. This study uses a parallel model that has been proven [58] to extract more features than a single pipeline, which are reflected in its class activation maps during the inference. Furthermore, while the conventional GANs generate images from random input noise, the proposed model generates images from an input brain image, resulting in more realistic images. The findings of the experiment are explained and shown in the following sections.

3. Proposed Model

The proposed model uses an unsupervised adversarial training scheme. It has two major components:

- The generator (G) learns the dataset distribution from the input image, encodes it into a latent vector, and reconstructs the image by upsampling. The uniqueness of the generator is that the encoder uses a parallel model that is comprised of a convolutional pipeline (CNN) and a dilated convolutional end network (DCN) that is 8 layers deep,

each layer uses 3×3 convolutional filters, a Rectified linear unit (ReLU) activation function, and a batch normalization operation. After two identical layers, a max-pooling operation is used for spatial dimension reduction and doubling the depth of the tensor. A latent vector of the input image is then generated.

The DCN is eight layers deep, each layer uses 3×3 convolutional filters with a dilation factor of 2, a ReLU (Rectified linear units) activation function, and a batch normalization operation. After 2 identical layers, a max-pooling operation is used for spatial dimension reduction and increasing the depth of the tensor. A latent vector of the input image is then generated. Concatenation of these features gives the optimal feature vector of the input image [58]. The class activation map for the given input image is shown in Figure 3.

- The discriminator (D) predicts the class of the input (whether it is fake or not) based on learned features. The discriminator generally uses an encoder-type architecture.

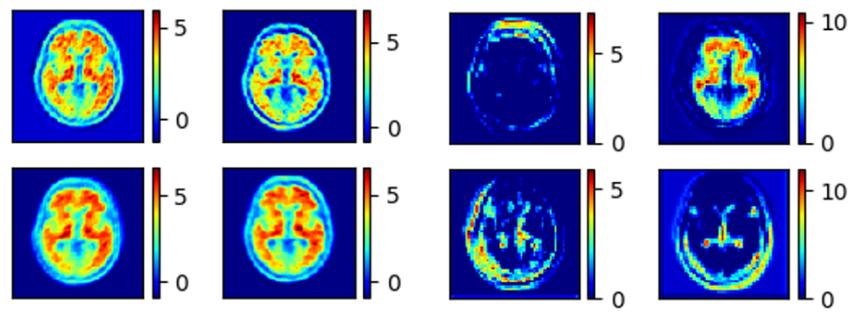
The mathematical definition and the formulation of the problem are the following:

The dataset is split into a training set D that is comprised of N normal images where $D = \{x_1, \dots, x_N\}$, and a testing set \hat{D} of A normal and abnormal images combined $\hat{D} = \{\hat{X}_1, y_1, (\hat{X}_2, y_2), \dots, (\hat{X}_A, y_A)\}$ where $y_A \in [0, 1]$ denotes normal and abnormal class labels, respectively. The task is training the proposed model f on D and perform inference on A . Ideally, the training set should be larger than the testing set, $D \gg A$. Training helps to map the distribution of the dataset D in all vector spaces. This enables the network to learn both higher and lower-level features that are different from abnormal samples.

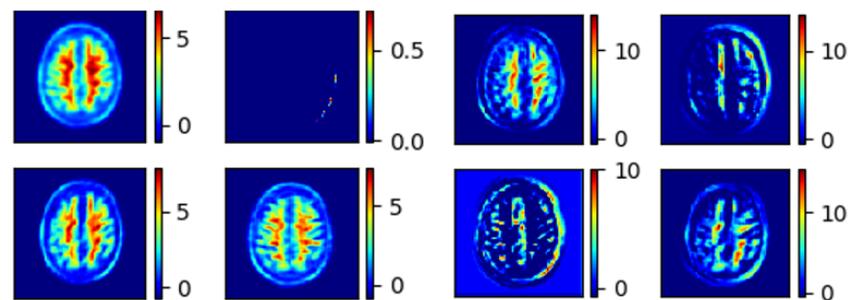
As shown in Figure 2, the proposed model consists of a generator G and a discriminator D . The generator uses an autoencoder-type structure to generate an image \hat{x} through extracted latent vector z from the input image x such that $G : x \rightarrow z$ where $x \in \mathbb{R}^{(w \times h \times c)}$ and $z \in \mathbb{R}^d$. The input image x is fed to both pipelines where the conventional CNN extracts local features and the DCN extracts global features. Concatenation of these features creates the latent vector z . The convolutional network pipeline consists of eight convolutional layers, each layer is created using 3×3 convolutional filters, a ReLU activation, and a batch normalization operation. The DCN consists of 8 dilated convolutional layers with each layer being comprised of a 3×3 convolutional filters with a dilation factor of 2, a ReLU activation, and a batch normalization operation. At the end of each pipeline, extracted image features are concatenated to create the latent vector z .

The decoder network consists of four upsampling layers and eight convolutional layers with 3×3 convolutional filters on each layer, and a ReLU activation. Its task is to upsample the latent vector z back to its original input image dimension which is denoted as \hat{x} .

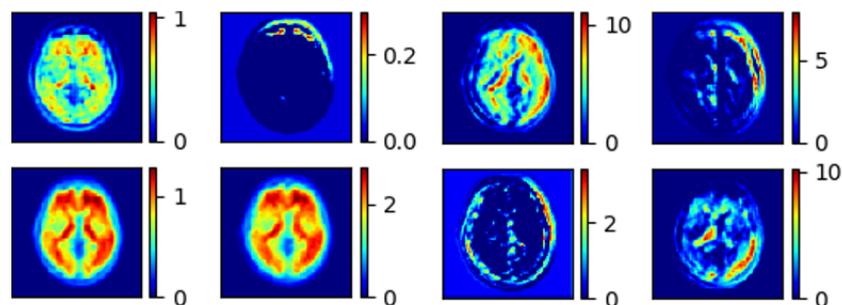
Unsupervised training is performed with the majority of the normal class in the proposed GAN-based anomaly detection. In all three cases, the train-test split is the same. Eighty percent of the NC data is used to train the model, and the remaining 20% is used together with a similar number of AD images to detect the AD as the anomaly, or MCI as the anomaly with the NC, or the AD as the anomaly with the MCI dataset. The model is trained with only MCI data to distinguish AD, and MCI and AD data are considered abnormal data.



(a) Class activation maps for a single Alzheimer's Disease (AD) image. The left 2×2 block shows the CNN activations, and the right 2×2 block shows the DCN activations.



(b) Class activation maps for a single Mild Cognitive Impairment (MCI) image. The left 2×2 block shows the CNN activations, and the right 2×2 block shows the DCN activations.



(c) Class activation maps for a single NC image. The left 2×2 block shows the CNN activations, and the right 2×2 block shows the DCN activations.

Figure 3. Class activation maps of all three classes; (a) AD, (b) MCI, and (c) NC. Despite the CNN in all cases cover a larger activation area, the activation values are lower than the DCN activations. Furthermore, it can be seen that the class activation maps highlight the different parts of the brain.

4. Experimental Environment and Results

4.1. The Dataset

The dataset used in this study was obtained from ADNI (<http://adni.loni.usc.edu>) (accessed on 22 December 2020). The ADNI study contains AD, MCI, and NC subjects and its main purpose is to establish clinical, imaging, biochemical, and genetic markers for the early detection of AD. For this study, participants with baseline, plus a-year follow-up 18F-fluorodeoxyglucose (FDG)-PET images were used. Of 256 subjects in total, there are 148 NC subjects, 83 MCI subjects, and 25 AD subjects. Demographic and clinical information (Clinical Dementia Rating—CRD, Mini-Mental State Examination—MMSE) is shown in Table 1. For the training stage, the normal class dataset is split 80-20% for training and inference (e.g., AD vs NC, MCI vs NC, AD vs MCI). One hundred percent of the abnormal class is used for inference, meaning, the model is not trained with the abnormal class.

The detailed information about the dataset split and the numbers can be seen in Table 2. Each patient's axial scan has 96 images, and each image is grayscale has the size of 160×160 pixels. The images were downsampled to 64×64 to reduce the computational cost.

Table 1. The clinical information and the demographic of the population (mean \pm standard deviation).

Group	NC	MCI	AD	Total
No. of subjects	148	83	25	256
Gender (M/F)	82/66	45/38	15/10	
No. of Images	14,208	7968	2400	24,576
Age, mean \pm SD	75.9 \pm 4.5	74.8 \pm 7.1	76.5 \pm 6.7	
CDR, mean \pm SD	0.0 \pm 0.0	0.5 \pm 0.0	0.8 \pm 0.2	
MMSE, mean \pm SD	29.1 \pm 1.0	27.2 \pm 1.6	23.4 \pm 2.0	

Table 2. The information about the train-test split for three cases. In training the case, 80% of the normal class data is split and the proposed model is trained with it. The remaining 20% is mixed with the corresponding abnormal class. The number of abnormal class data is either the same as the normal class or the highest possible number of usable images.

Cases	Training			Testing			Total
	AD	MCI	NC	AD	MCI	NC	
AD-NC			11,366	2400		2842	16,608
MCI-NC			11,366		2842	2842	17,050
AD-MCI		6374		2400	1594		10,368

4.2. Training the Model

The recent trend of anomaly detection models such as skip-GAN [56] and AnomalyGAN (AnoGAN) [23] is to train the model on the normal dataset by separating it into train-test fashion, and then perform inference with both unseen normal and unseen abnormal samples. Ideally, normal sample inference images will have similar latent vectors and the reconstructed image will be similar to the input image. On the other hand, abnormal samples that the model has not been trained for are expected to fail in both cases. A higher loss value can be expected in such cases, which will be used for anomaly score calculation. In case of training, the proposed model has three loss functions for optimization which are contextual loss, latent vector loss, and adversarial loss.

1. Contextual Loss: To learn the distribution of the dataset, \mathcal{L}_1 normalization is applied to the input x and the output \hat{x} . This helps the generation of contextually similar images from the normal samples and is proven to produce less blurry images than \mathcal{L}_2 normalization [54]. The loss formula is given below as:

$$\mathcal{L}_{context} = E_{x \sim p_x} |x - \hat{x}|_1. \quad (1)$$

2. Adversarial Loss: Taken from [28], this loss ensures that the generator G can reconstruct an image x as realistically as possible while the discriminator D can differentiate between the normal and fake images. The task is to minimize this objective for G and maximize it for D to achieve the min-max equilibrium \mathcal{L}_1 where it is defined as:

$$\mathcal{L}_{adv} = E_{x \sim p_x} [\log D(x)] + E_{x \sim p_x} [\log(1 - D(\hat{x}))]. \quad (2)$$

3. Latent Loss: This \mathcal{L}_2 loss is used in obtaining the latent representations of the input x and the generated output \hat{x} as similar as possible. This ensures that the network can

produce similar latent representations for sampling. Using the concatenated features $z = f(x)$ and the fully connected layer of the discriminator $\hat{z} = f(\hat{x})$. The loss becomes:

$$\mathcal{L}_{latent} = E_{x \sim p_x} |f(x) - f(\hat{x})|_2. \quad (3)$$

Thus, the total loss of the model becomes:

$$\mathcal{L}_{total} = \mathcal{L}_{context} + \mathcal{L}_{adv} + \mathcal{L}_{latent}. \quad (4)$$

For training the proposed model f , stochastic gradient descent (SGD) with lambda decay, and Nesterov momentum optimizer function with the initial learning rate 10^{-2} is used. The model is trained until the convergence, which took 60 epochs. To perform a patient-wise training and inference, the batch size is taken as 96 which is the total number of axial brain images of a single patient. Same hyperparameters were used to train the benchmark models and the ablation study, with the same dataset, and the same 80-20% split. The implementation is done using TensorFlow v2.0, Python 3.6.8, CUDA 10.0, and CuDNN 7.6.5. The experiments are conducted on an NVIDIA P5000 GPU.

4.3. Model Evaluation

Model performance is evaluated through multiple methods. First, a classification inference is performed with the parallel model to check whether the parallel encoder module is working accurately. Trained parallel model weights are frozen, a classification layer (softmax) is added, and fine-tuning is performed with the same training hyperparameters, with the categorical cross-entropy loss. Same hyperparameters are used to train the benchmark classification models. The benchmark result with other classification models is given in Table 3, and the confusion matrix of the parallel model can be seen in Figure 4.

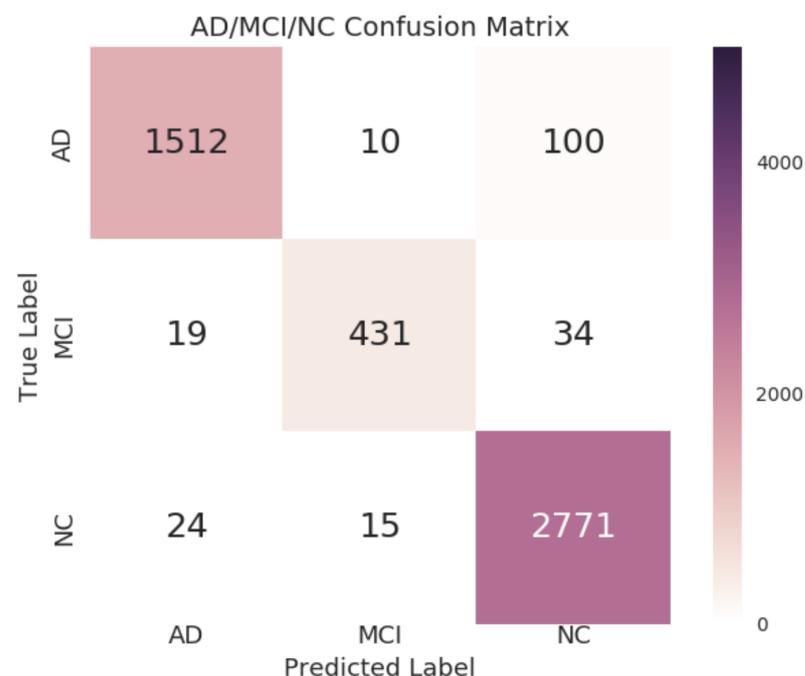


Figure 4. Confusion matrix for the parallel model inference. The confusion matrix is drawn with the results from the inference using true labels and the model's predicted labels.

Class activation maps are helpful to highlight the regions where the model's attention is focused on in data visualization. These regions shown in Figure 3 are relevant to each different class. Thus, the greater the variance between these regions is, the more accurate the model performs. Looking at the model's class activation maps, both CNN and DCN

activation areas are complementary to each other in a given brain image. The four activation images on the left-hand side are from the CNN, whereas the four activation images on the right-hand side are from the DCN. Although the DCN has a lower activation area overall, the values in these areas are higher than the CNN activation values, indicating that the DCN features are distinct and unique.

Table 3. Classification performance comparison between image classification models such as DenseNet169, Inception-V3, ResNet50, VGG16, and the parallel model.

Performance	Accuracy	Precision	Recall	F1-Score
Parallel Model	96.03	95.70	93.73	94.67
DenseNet169 [59]	95.44	93.98	92.31	94.20
Inception-V3 [60]	90.06	88.73	87.12	89.44
ResNet50 [61]	89.41	86.81	87.29	88.52
VGG16 [62]	85.13	83.98	82.31	85.01

The second evaluation method is the area under the curve (AUC) [63]. It is a function that uses true positive rates (TPR) and false-positive rates (FPR) with varying threshold values from the inference data.

$$TPR = \frac{TP}{TP + FN} \quad , \quad FPR = \frac{FP}{FP + TN} \quad (5)$$

where TP is a true positive, FN is a false negative, FP is a false positive, and TN is a true negative. The model's performance is compared with recent anomaly detection works such as GANomaly [29], Skip-GANomaly [56], EGBAD [32], and AnoGAN [23]. Additionally, an ablation study is performed on CNN pipeline and DCN pipeline networks. The comparison can be seen in Table 4.

Table 4. Area under the curve (AUC) results of benchmark models GANomaly, EGBAD, AnoGAN, Skip-GAN, and ablation study of the proposed model, and the proposed model itself.

Model	AD-NC	MCI-NC	AD-MCI
GANomaly [29]	59.31	57.33	48.34
EGBAD [32]	58.46	60.31	54.71
AnoGAN [23]	55.71	52.93	48.11
Skip-GANomaly [56]	71.42	67.02	68.56
Ablation-Conv.	68.71	65.34	61.25
Ablation-Dilated Conv.	60.82	54.02	52.40
Proposed Model	75.21	70.84	71.85

A qualitative evaluation method called Fréchet Inception Distance (FID) [64] is applied as a metric. This method is used to evaluate the quality of the generated images by calculating the distance of feature vectors between the real and the generated images. The estimation is done via the Inception-V3 [60] model which was built on the original GoogLeNet architecture [65] classifies the generated image. The conditional class probability and the confidence of each image are combined. This is defined as [64]:

$$FID = |\mu_r - \mu_g|^2 + Tr\left(\sum_r + \sum_g - 2\left(\sum_r \sum_r \sum_g\right)^{-1/2}\right) \quad (6)$$

where $X_r \sim N(\mu_r, \Sigma_r)$ and $X_g \sim N(\mu_g, \Sigma_g)$ are 2048-dimensional activations of the pool3 layer for real and generated samples. If both images are identical, the score should ideally be 0. Since there is no clear given metric value that is acceptable for unsupervised learning in the literature, there might be problems with what the acceptable value. A total of 1000 constructed images gathered from each class were compared with the correspond-

ing 1000 real images to obtain an FID score for each class separately. The score comparison for different classes and models can be seen in Table 5 and reconstructed sample images for all cases and the ablation study results can be seen in Figure 5.

Table 5. Fréchet Inception Distance (FID) scores comparison for three different classes; AD, MCI, and NC using GANomaly, Skip-GANomaly, EGBAD, and AnoGAN along with the ablation study. Note that a lower score generally means better performance.

Models	AD	MCI	NC
GANomaly [29]	18.216	17.353	17.885
EGBAD [32]	20.815	21.621	19.134
AnoGan [23]	19.908	17.991	21.759
Skip-GANomaly [56]	11.892	12.516	11.908
Ablation Conv.	16.175	18.023	17.659
Ablation Dilated Conv.	21.762	18.895	21.342
Proposed Model	9.481	10.024	10.858

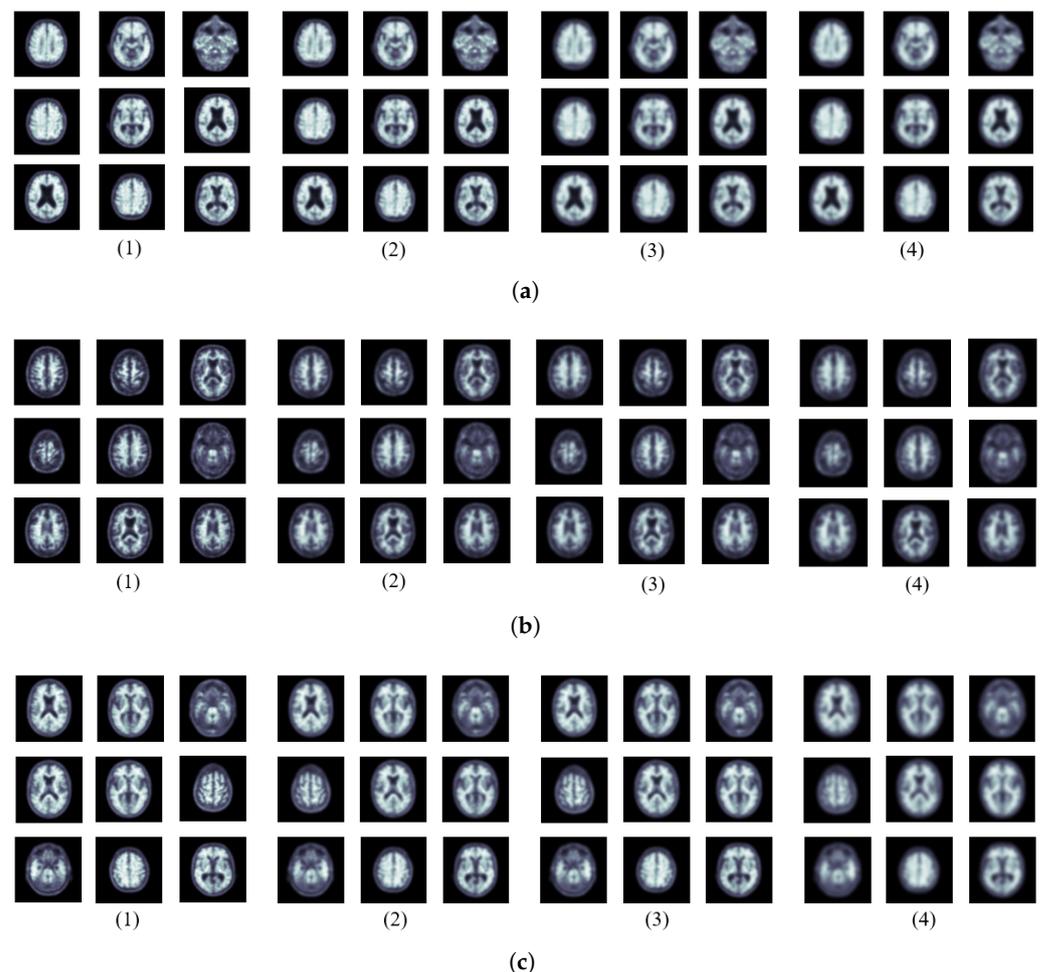
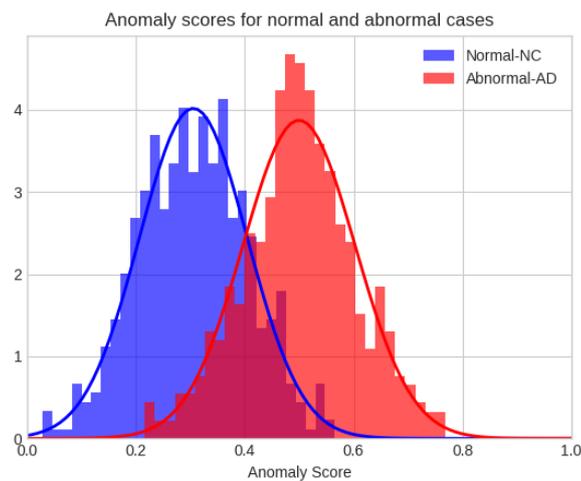


Figure 5. 9 original images for each class AD (a), MCI (b), and NC (c), respectively. (1) The original image, (2) the reconstructed image by the proposed model, (3) the CNN reconstruction, and (4) the DCN reconstruction images are given. (a) An original batch of AD images and the reconstructed counterparts by the proposed model and its sub-networks. (b) An original batch of MCI images and the reconstructed counterparts by the proposed model and its sub-networks. (c) An original batch of NC images and the reconstructed counterparts by the proposed model and its sub-networks.

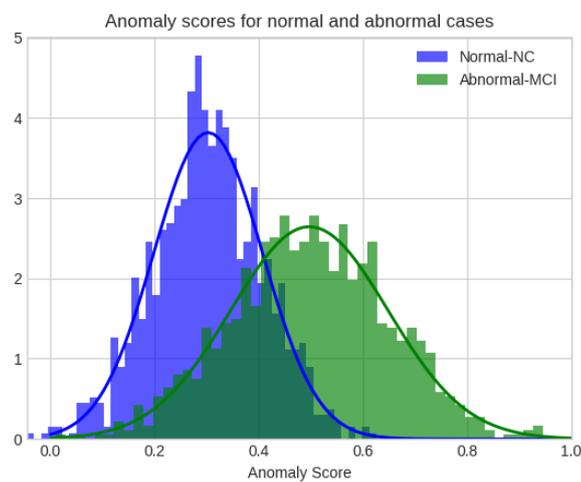
An anomaly score [32] is used to detect the anomalies in a given test image. For an input image \hat{x} the corresponding anomaly score is calculated as:

$$A(\hat{x}) = (1 - \lambda) \cdot R(\hat{x}) + \lambda \cdot D(\hat{x}) \quad (7)$$

where $R(\hat{x})$ is the \mathcal{L}_1 loss between the input image and the corresponding reconstructed image. $D(\hat{x})$ is the latent vector score given in Equation (3). λ is the weighting parameter on emphasizing the importance of the scores. This experiment was performed based on $\lambda = 0.5$. The anomaly scores were then normalized. The anomaly score comparison for three different scenarios with the frequency graphs are shown in Figure 6.



(a)



(b)

Figure 6. Cont.

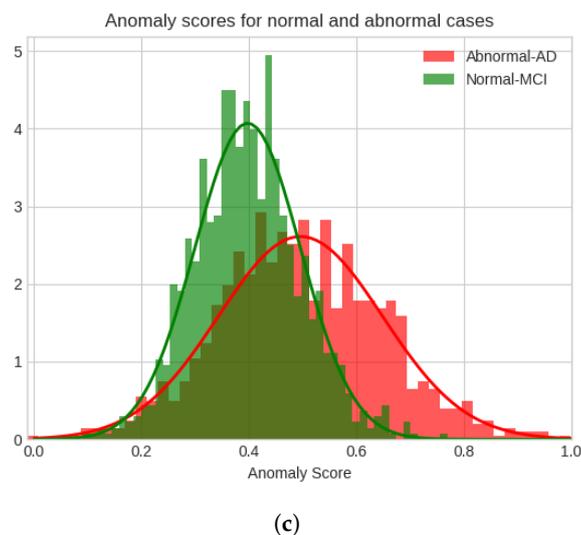


Figure 6. Anomaly score distributions for test images. Note that the anomaly scores are normalized to $[0, 1]$, and the images with higher anomaly scores are reflected with values closer to 1 while the normal images are closer to 0. (a) AD-NC case. AD is considered the anomaly, which reflects a higher anomaly score in the distribution. (b) MCI-NC case. MCI is considered the anomaly, which reflects a higher anomaly score in the distribution. (c) AD-MCI case. AD is considered the anomaly, which reflects a higher anomaly score in the distribution.

5. Conclusions

In this work, an analysis of Alzheimer’s Disease as an anomaly in PET images using an unsupervised anomaly detection model featuring a parallel feature extractor is conducted. The parallel feature extractor consists of a CNN and a DCN. Latent vectors obtained from both pipelines are concatenated and used as the main feature vector to reconstruct the input image. The discriminator takes both the original input image and the reconstructed image as input and labels them as either real or fake. As it is shown in Tables 4 and 5, the proposed model outperforms the previous anomaly models both quantitatively and qualitatively by having a classification accuracy of 96.03%, 0.59% higher than DenseNet169, and an AUC score of 75.21%, 3.79% higher than Skip-GANomaly. The ablation study shows that the proposed model also outperforms ablation study sub-networks. The narrow activation areas with high values seen in the DCN class activation maps provide a boosting effect to the CNN’s wider, but much lower value areas. The justification of the superior performance in the classification of the parallel feature extractor model is given in Table 3 and the proof is given in the form of class activation maps, as shown in Figure 3. Although there are no solid metric values of the optimal performance for the unsupervised models, comparison based on the FID score shows that the model produces images similar to the real ones. The score comparisons are given in Table 5, and a sample batch of images is shown in Figure 5. There are multiple discussions to be made in future studies:

- Among the three loss functions, the importance of each loss function can be evaluated. A genetic algorithm or a grid search can be used to assign weights for each loss function to observe its effect on the model.
- The depth of the parallel model can be altered by using other parameter search algorithms. A deeper model may increase the computational cost while improving the accuracy.
- The skip connections used in autoencoders have been showing promising results [57]. The possibility and feasibility of skip connections will be investigated to further improve the performance of the model.
- Probable ways to improve the AUC score further will be investigated.

Author Contributions: Conceptualization, H.B.B., J.-S.P., and D.-Y.K.; methodology, H.B.B. and J.-S.P.; software, H.B.B.; validation, H.B.B., J.-S.P., and D.-Y.K.; formal analysis, H.B.B. and J.-S.P.; investigation, H.B.B., J.-S.P., and D.-Y.K.; resources, ADNI; data curation, ADNI; writing—original draft preparation, H.B.B.; writing—review and editing, J.-S.P., and D.-Y.K.; visualization, H.B.B.; supervision, J.-S.P., and D.-Y.K.; project administration, J.-S.P., and D.-Y.K.; funding acquisition, J.-S.P., and D.-Y.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Research Foundation (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2018R1A2B2008178) & Brain Busan 21+ Program.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Boards of both Kyungseong University and Dong-A University.

Informed Consent Statement: All subjects gave their informed consent for inclusion before they participated in the study.

Data Availability Statement: Data used in the preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.usc.edu>, accessed on 22 December 2020). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. 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). For up-to-date information, see <http://www.adni-info.org> (accessed on 22 December 2020).

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, accessed on 22 December 2020). 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.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AD	Alzheimer’s Disease
MCI	Mild Cognitive Impairment
NC	Normal Control
A β	Amyloid beta
PET	Positron emission tomography
MRI	Magnetic Resonance Imaging
MMSE	Mini Mental State Examination
CDR	Clinical Dementia Rating
GANs	Generative Adversarial Networks
DCCGAN	Deep Convolutional Generative Adversarial Networks

AnoGAN	Anomaly GAN
EGBAD	Efficient GAN-Based Anomaly Detection
BiGAN	Bidirectional GAN
sMRI	structural Magnetic Resonance Imaging
fMRI	functional Magnetic Resonance Imaging
CNN	Convolutional Neural Network
ReLU	Rectified Linear Units
ROI	Region of Interest
DCN	Dilated Convolutional Network
AUC	Area Under the Curve
FID	Fréchet Inception Distance
FPR	False Positive Rate
TPR	True Positive Rate
SVM	Support Vector Machine
SGD	Stochastic Gradient Descent
ADNI	Alzheimer's Disease Neuroimaging Initiative

References

1. Johns Hopkins Bloomberg School of Public Health. Alzheimer's Disease to Quadruple Worldwide by 2050. Available online: <https://www.jhsph.edu/news/news-releases/2007/brookmeyer-alzheimers-2050.html> (accessed on 17 January 2021).
2. Yangling, M.; Gage, F. H. Adult hippocampal neurogenesis and its role in Alzheimer's disease. *Mol. Neurodegener.* **2011**, *6*, 85. [[CrossRef](#)]
3. Villemagne, V.L.; Rowe, C.C.; Macfarlane, S.; Novakovic, K.; Masters, C.L. Imaginem oblivionis: The prospects of neuroimaging for early detection of Alzheimer's disease. *J. Clin. Neurosci.* **2005**, *12*, 221–230. [[CrossRef](#)] [[PubMed](#)]
4. Haass, C.; Selkoe, D.J. Soluble protein oligomers in neurodegeneration: Lessons from the Alzheimer's amyloid β -peptide. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 101. [[CrossRef](#)] [[PubMed](#)]
5. Chen, X.; Yan, S.D. Mitochondrial A β : a potential cause of metabolic dysfunction in Alzheimer's disease. *IUBMB Life* **2006**, *52*, 686–694. [[CrossRef](#)] [[PubMed](#)]
6. Alzheimer's Association. Mild Cognitive Impairment (MCI). Available online: https://www.alz.org/alzheimers-dementia/what-is-dementia/related_conditions/mild-cognitive-impairment (accessed on 17 January 2021).
7. Sabri, O.; Sabbagh, M.N.; Seibyl, J.; Barthel, H.; Akatsu, H.; Ouchi, Y.; Senda, K.; Murayama, S.; Ishii, K.; Takao, M.; et al. Florbetaben PET imaging to detect amyloid beta plaques in Alzheimer's disease: Phase 3 study. *Alzheimer's Dement.* **2015**, *11*, 964–974. [[CrossRef](#)] [[PubMed](#)]
8. Johnson, K.A.; Fox, N.C.; Sperling, R.A.; Klunk, W.E. Brain Imaging in Alzheimer Disease. *Cold Spring Harb. Perspect. Med.* **2012**, *4*. [[CrossRef](#)] [[PubMed](#)]
9. Catana, C.; Guimaraes, A.R.; Rosen, B.R. PET and MRI: The Odd Couple or a Match Made in Heaven? *J. Nucl. Med.* **2013**, *54*, 815–824. [[CrossRef](#)] [[PubMed](#)]
10. Desai, B.; Sheth, S.; Conti, P. Review of novel radiotracers for positron emission tomography (PET) imaging. *J. Nucl. Med.* **2013**, *54*, 815–824. [[CrossRef](#)]
11. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.; Van Ginneken, B.; Sanchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [[CrossRef](#)]
12. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
13. Yanase, J.; Triantaphyllou, E. A systematic survey of computer-aided diagnosis in medicine: Past and present developments. *Expert Syst. Appl.* **2019**, *138*, 112821. [[CrossRef](#)]
14. Jo, T.; Nho, K.; Saykin, A.J. Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data. *Front. Aging Neurosci.* **2019**, *11*, 220. [[CrossRef](#)] [[PubMed](#)]
15. Samper-Gonzalez, J.; Burgos, N.; Bottani, S.; Fontanella, S.; Lu, P.; Marcoux, A.; Routier, A.; Guillon, J.; Bacci, M.; Wen, J.; et al. Reproducible evaluation of classification methods in Alzheimer's disease: framework and application to MRI and PET data. *Neuroimage* **2018**, *183*, 504–521. [[CrossRef](#)]
16. Riedel, B.C.; Daianu, M.; Ver Steeg, G.; Mezher, A.; Salminen, L.E.; Galstyan, A.; Thompson, P.M. Uncovering biologically coherent peripheral signatures of health and risk for Alzheimer's disease in the aging brain. *Front. Aging Neurosci.* **2019**, *10*, 390. [[CrossRef](#)] [[PubMed](#)]
17. Shen, S.; Sandham, W.A.; Granat, M.H. Preprocessing and segmentation of brain magnetic resonance images. In Proceedings of the IEEE EMBS International Conference on Information Technology Applications in Biomedicine, Birmingham, UK, 24–26 April 2003.
18. Dachena, C.; Casu, S.; Fanti, A.; Lodi, M.B.; Mazzarella, G. Combined Use of MRI, fMRI and Cognitive Data for Alzheimer's Disease: Preliminary Results. *Appl. Sci.* **2019**, *9*, 3156. [[CrossRef](#)]
19. Tufail, J.A.; Rudisill, C.; Egan, C.; Kapetanakis, V.V.; Vega-Salas S.; Owen, C.G.; Lee, A.; Louw, V.; Anderson, J.; Liew, G.; et al. Automated Diabetic Retinopathy Image Assessment Software: Diagnostic Accuracy and Cost-Effectiveness Compared with Human Graders. *Ophthalmology* **2017**, *124*, 343–351. [[CrossRef](#)]

20. Johnson, J.M.; Khoshgoftaar, T.M. Survey on deep learning with class imbalance. *J. Big Data* **2019**, *6*, 1. [[CrossRef](#)]
21. Willeminck, M.J.; Koszek, W.A.; Hardell, C.; Wu, J.; Fleischmann, D.; Harvey, H.; Folio, L.R.; Summers, R.M.; Rubin, D.L.; Lungren, M.P. Preparing Medical Imaging Data for Machine Learning. *Radiology* **2020**, *295*, 4–15. [[CrossRef](#)]
22. Ahmed, M.; Naser Mahmood, A.; Hu, J. A survey of network anomaly detection techniques. *J. Netw. Comput. Appl.* **2016**, *60*, 19–31. [[CrossRef](#)]
23. Schlegl, T.; Seeböck, P.; Waldstein, S.M.; Schmidt-Erfurth, U.; Langs, G. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2017; Volume 10265, pp. 146–147. [[CrossRef](#)]
24. Kiran, B.R.; Thomas D.M.; Parakkal, R. An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos. *J. Imaging* **2018**, *4*, 36. [[CrossRef](#)]
25. Hodge, V.; Austin, J. A Survey of Outlier Detection Methodologies. *Artif. Intell. Rev.* **2004**, *22*, 85–126. [[CrossRef](#)]
26. Pimentel, M.A.; Clifton, D.A.; Clifton, L.; Tarassenko, L. A review of novelty detection. *Signal Process.* **2014**, *99*, 215–249. [[CrossRef](#)]
27. Chandola, V.; Banerjee, A.; Kumar, V. Anomaly detection: A Survey. *ACM Comput. Surv.* **2009**, *41*, 1–58. [[CrossRef](#)]
28. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the Conference on Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 2672–2680.
29. Akcay, S.; Atapour-Abarghouei, A.; Breckon, T.P. GANomaly: Semisupervised anomaly detection via adversarial training. In Proceedings of the Asian Conference on Computer Vision, Perth, Australia, 2–6 December 2018; pp. 622–637.
30. Shin, H.C.; Ihsani, A.; Xu, Z.; Mandava, S.; Sreenivas, S.T.; Forster, C.; Cha, J.; Alzheimer’s Disease Neuroimaging Initiative. GANDALF: Generative Adversarial Networks with Discriminator-Adaptive Loss Fine-Tuning for Alzheimer’s Disease Diagnosis from MRI. In Proceedings of the Conference on Medical Image Computing and Computer Assisted Intervention, Lima, Peru, 4–8 October 2020; pp. 688–697.
31. Islam, J.; Zhang, Y. GAN-based synthetic brain PET image generation. *Brain Inform.* **2020**, *7*. [[CrossRef](#)]
32. Zenati, H.; Foo, C.S.; Lecouat, B.; Manek, G.; Chandrasekhar, V.R. Efficient gan-based anomaly detection. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018.
33. Sabokrou, M.; Khalooei, M.; Fathy, M.; Adeli, E. Adversarially learned one-class classifier for novelty detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 30 April–3 May 2018; pp. 3379–3388.
34. Fan, Y.; Resnick, S.M.; Wu, X.; Davatzikos, C. Structural and functional biomarkers of prodromal Alzheimer’s disease: A high-dimensional pattern classification study. *Neuroimage* **2008**, *41*, 277–285. [[CrossRef](#)] [[PubMed](#)]
35. Hu, K.; Wang, Y.; Chen, K.; Hou, L.; Zhang, X. Multi-scale features extraction from baseline structure MRI for MCI patient classification and AD early diagnosis. *Neurocomputing* **2016**, *175*, 132–145. [[CrossRef](#)]
36. Filiphovych, R.; Davatzikos, C.; Alzheimer’s Disease Neuroimaging Initiative. Semi-supervised pattern classification of medical images: Application to mild cognitive impairment (MCI). *Neuroimage* **2011**, *55*, 1109–1119. [[CrossRef](#)]
37. Gray, K.R.; Wolz, R.; Heckermann, R.A.; Aljabar, P.; Hammers, A.; Rueckert, D.; The Alzheimer’s Disease Neuroimaging Initiative. Multi-region analysis of longitudinal FDG-PET for classification for the Alzheimer’s disease. *Neuroimage* **2012**, *60*, 221–229. [[CrossRef](#)]
38. Liu, M.; Cheng, D.; Yan, W.; The Alzheimer’s Disease Neuroimaging Initiative. Classification of Alzheimer’s Disease by Combination of Convolutional and Recurrent Neural Networks Using FDG-PET Images. *Neuroimage* **2012**, *60*, 221–229. [[CrossRef](#)]
39. Davatzikos, C.; Bhatt, P.; Shaw, L.M.; Batmanghelich, K.N.; Trojanowski, J.Q. Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. *Neurobiol. Aging* **2011**, *32*, 19–27. [[CrossRef](#)]
40. Aderghal, K.; Benois-Pineau, J.; Afdel, K.; Gwenaëlle, C. FuseMe: Classification of sMRI images by fusion of Deep CNNs in 2D + ϵ projections. In Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing, Firenze, Italy, 19–21 June 2017; pp. 1–7. [[CrossRef](#)]
41. Suk, H.; Wee, C.Y.; Shen, D. Discriminative group sparse representation for mild cognitive impairment classification. In Proceedings of the International Workshop on Machine Learning in Medical Imaging, Nagoya, Japan, 22 September 2013; pp. 131–138.
42. Ramzan, F.; Khan, M.U.G.; Rehmat, A.; Iqbal, S.; Saba, T.; Rehman, A.; Mehmood, Z. A Deep Learning Approach for Automated Diagnosis and Multi-Class Classification of Alzheimer’s Disease Stages Using Resting-State fMRI and Residual Neural Networks. *J. Med. Syst.* **2020**, *44*, 916–922. [[CrossRef](#)]
43. Perrin, R.J.; Fagan, A.M.; Holtzman, D.M. Multimodal techniques for diagnosis and prognosis of Alzheimer’s disease. *Nature* **2009**, *461*, 916–922. [[CrossRef](#)]
44. Liu, M.; Cheng, D.; Wang, K.; Wang, Y.; The Alzheimer’s Disease Neuroimaging Initiative. Multi-Modality Cascaded Convolutional Neural Networks for Alzheimer’s Disease Diagnosis. *Neuroinformatics* **2009**, *16*, 295–308. [[CrossRef](#)] [[PubMed](#)]
45. Singh, S.; Srivastava, A.; Mi, L.; Chen, K. Deep Learning based Classification of FDG-PET Data for Alzheimers Disease Categories. In Proceedings of the International Conference on Medical Information Processing and Analysis, San Andres Island, Colombia, 17 November 2017. [[CrossRef](#)]
46. Xue, Y.; Xu, T.; Zhang, H.; Long, L.R.; Huang, X. SegAN: Adversarial Network with Multi-scale L_1 Loss for Medical Image Segmentation. *Neuroinformatics* **2017**, *16*, 383–392. [[CrossRef](#)] [[PubMed](#)]

47. Armanious, K.; Jiang, C.; Fischer, M.; Kustner, T.; Hepp, T.; Nikolau, K.; Gatidis, S.; Yang, B. MedGAN: Medical image translation using GANs. *Comput. Med. Imaging Graph.* **2019**, *79*. [[CrossRef](#)] [[PubMed](#)]
48. Creswell, A.; White, T.; Dumoulin, V.; Arulkumaran, K.; Sengupta, B.; Bharath, A.A. Generative adversarial networks: An overview. *IEEE Signal Process. Mag.* **2018**, *35*, 53–65. [[CrossRef](#)]
49. Salimans, T.; Goodfellow, I.; Zaremba, W.; Cheung, V.; Radford, A.; Chen, X. Improved techniques for training gans. In *Advances in Neural Information Processing*; MIT Press: Cambridge, MA, USA, 2016; pp. 2234–2242.
50. Radford, A.; Metz, L.; Chintala, S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. In *Proceedings of the International Conference on Learning Representations*, San Juan, Puerto Rico, 2–4 May 2016.
51. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the International Conference on Machine Learning*, Lille, France, 7–9 July 2015; pp. 448–456.
52. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein generative adversarial networks. In *Proceedings of the International Conference on Machine Learning*, Sydney, Australia, 6–11 August 2017; pp. 214–223.
53. Ravanbakhsh, M.; Sangineto, E.; Nabi, M.; Sebe, N. Training Adversarial Discriminators for Cross-channel Abnormal Event Detection in Crowds. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, Waikoloa, HI, USA, 7–11 January 2019. [[CrossRef](#)]
54. Isola, P.; Zhu, J.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 21–26 July 2017; pp. 5967–5976. [[CrossRef](#)]
55. Donahue, J.; Krahenbuhl, P.; Darrell, T. Adversarial Feature Learning. In *Proceedings of the International Conference on Learning Representations*, Toulon, France, 24–26 April 2017.
56. Akcay, S.; Atapour-Abarghouei, A.; Breckon, T.P. Skip-GANomaly: Skip Connected and Adversarially Trained Encoder-Decoder Anomaly Detection. In *Proceedings of the International Joint Conference on Neural Networks*, Budapest, Hungary, 14–19 July 2019. [[CrossRef](#)]
57. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Proceedings of the Medical Image Computing and Computer-Assisted Intervention*, Munich, Germany, 5–9 October 2017; pp. 234–241. [[CrossRef](#)]
58. Baydargil, H.B.; Park, J.S.; Kang, D.Y.; Kang, H.; Cho, K. A Parallel Deep Convolutional Neural Network for Alzheimer’s disease classification on PET/CT brain images. *KSII Trans. Internet Inf. Syst.* **2020**, *14*. [[CrossRef](#)]
59. Huang, G.; Liu, Z.J.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 21–26 July 2017. [[CrossRef](#)]
60. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 26 June–1 July 2016. [[CrossRef](#)]
61. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 26 June–1 July 2016. [[CrossRef](#)]
62. Simonyan, K.; Sizzerman, A. Very deep convolutional networks for large-scale image recognition. In *Proceedings of the International Conference on Machine Learning*, Lille, France, 6–11 July 2016. [[CrossRef](#)]
63. Ling, C. X.; Huang, J.; Zhang, H. Auc: A statistically consistent and more discriminating measure than accuracy. In *Proceedings of the International Joint Conferences on Artificial Intelligence*, Acapulco, Mexico, 9–15 August 2003; pp. 519–524.
64. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B. VGANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. In *Proceedings of the International Conference on Neural Information Processing Systems*, Long Beach, CA, USA, 4–9 December 2017; pp. 6629–6640.
65. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 7–12 June 2015. [[CrossRef](#)]