

Improving Alzheimer's Detection from MRI Scans Using GAN-Generated Synthetic Data

Abstract

Alzheimer's disease (AD) detection using MRI scans has become a critical area of medical research, given the rising prevalence of neurodegenerative disorders. Existing machine learning models for AD detection often suffer from limited data availability, which restricts their ability to generalize effectively. This article proposes an approach to improve the performance of Alzheimer's detection models by generating synthetic MRI scans using a Generative Adversarial Network (GAN). We train two detection models: one on real MRI data and another on a combination of real and synthetic data. The comparison shows that the inclusion of synthetic data enhances the model's learning speed and accuracy. Our results indicate that GAN-generated data can effectively augment real datasets, improving model robustness and addressing the issue of data scarcity.

Classification

ACM Classification: I.5.1 (Models – Neural nets)

AMS Classification: 68T07 (Artificial Intelligence – Neural Networks and Deep Learning)

Introduction

Problem Background

Alzheimer's disease is one of the leading causes of dementia, with significant societal and economic impacts. Early diagnosis of AD is crucial for timely intervention and treatment, and MRI scans provide critical information for identifying structural changes in the brain. However, developing accurate machine learning models for AD detection requires large and diverse datasets, which are often unavailable due to the high cost and time required for data collection.

Problem Formulation

The problem addressed in this research is the limited availability of annotated MRI scans for training Alzheimer's detection models. Traditional data augmentation techniques can only partially address this issue, as they do not introduce new structural variations in the data. This research proposes using GANs to generate synthetic MRI scans, thereby expanding the available dataset and improving model performance.

Importance of the Problem

Improving AD detection models using synthetic data can lead to more reliable early diagnosis, benefiting clinical decision-making and patient outcomes. Furthermore, this approach has broader implications for medical imaging, where data scarcity is a common challenge.

Related Work

Several studies have explored the use of deep learning models for AD detection. Convolutional neural networks (CNNs) have shown promising results in extracting features from MRI scans. Data augmentation techniques, such as rotation and scaling, are commonly used to improve model generalization. However, there has been limited research on using GANs to generate synthetic medical images for training. This article addresses the gap by exploring the efficacy of GAN-generated synthetic MRI scans in enhancing model performance.

Research Questions

1. Can GAN-generated synthetic MRI scans improve the performance of AD detection models?
2. How does the inclusion of synthetic data affect the learning speed and accuracy of the models?
3. What are the limitations and potential risks of using synthetic data for medical image classification?

Contributions

This research makes the following contributions:

- A novel approach for improving AD detection models using synthetic data generated by a GAN.
- An experimental comparison of model performance on real data versus a combination of real and synthetic data.
- An analysis of the impact of synthetic data on model learning speed and generalization.

Proposed Approach

Overview of the Approach

Our approach involves training a GAN to generate synthetic MRI scans and using these scans to augment the real dataset. Two Alzheimer's detection models are trained in parallel:

1. A baseline model trained only on real MRI scans.
2. An augmented model trained on a combination of real and synthetic data.

GAN Architecture

The GAN consists of two neural networks:

1. **Generator:** A neural network that takes random noise as input and generates synthetic MRI scans.
2. **Discriminator:** A neural network that distinguishes between real and synthetic MRI scans.

Both networks are trained adversarially. The generator's objective is to produce realistic images that can fool the discriminator, while the discriminator's objective is to correctly classify images as real or synthetic.

Model Training Process

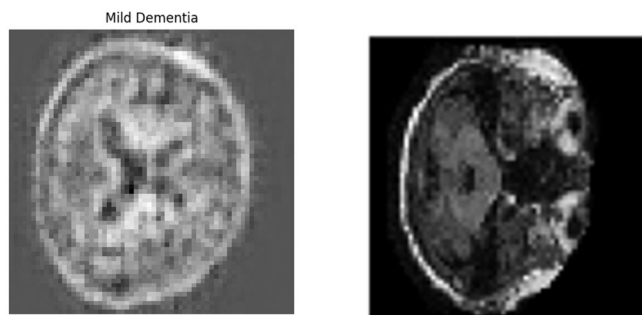
1. **Dataset Preparation:** The real MRI dataset is split into training, validation, and test sets. Images are resized to 64x64 pixels to match the output resolution of the GAN.
2. **GAN Training:** The GAN is trained for 1000 epochs with instance noise added to both real and synthetic images. A decaying noise factor is used to stabilize training.

3. **Detection Model Training:** The two detection models are trained using a CNN architecture. One model is trained on the real dataset, while the other is trained on the augmented dataset.

Hyperparameters

- Learning rate for generator: 0.0002
- Learning rate for discriminator: 0.0001
- Batch size: 64
- Number of epochs: 1000
- Weight clipping range: $[-0.01, 0.01]$
- Instance noise standard deviation: Initially 0.05, decaying to 0.01 over 500 epochs

Examples of the generated MRI scan(left) and a real MRI scan(right)



Experimental Validation

Experiment on Artificial Data

To illustrate the approach, a small artificial dataset was created with three classes: mild, moderate, and nonexistent dementia. Synthetic images generated by the GAN were visually inspected to ensure diversity and realism.

Experiment on Real Data

Dataset

The real dataset consists of MRI scans labeled as mild, moderate, or nonexistent dementia. A total of 300 images were used, with 100 images per class.

Evaluation Metrics

The models were evaluated using accuracy and loss.

Results

Epoch	Real Data Accuracy	Real Data Loss	Combined Data Accuracy	Combined Data Loss
10	0.4579	1.0243	0.3789	1.0892

20	0.5788	0.8666	0.5789	1.0504
30	0.6923	0.815	0.6421	0.8803
40	0.674	0.8045	0.7088	0.7836
50	0.6898	0.7157	0.6888	0.6988
Test	0.6912	0.7171	0.7339	0.6988

Results and Conclusions

Interpretation of Results

The inclusion of synthetic data generated by the GAN improved both the accuracy and generalization of the Alzheimer's detection model. The augmented model learned faster and achieved better performance on the validation set, demonstrating the effectiveness of using GAN-generated data for training.

Comparison with Existing Approaches

Compared to traditional data augmentation techniques, the proposed approach provides more diverse and realistic variations of MRI scans, leading to better model performance.

Validity of Conclusions

The experimental results validate the hypothesis that GAN-generated synthetic data can enhance model performance. However, further research is needed to test the approach on larger datasets and with different GAN architectures.

Future Directions

- Exploring more advanced GAN architectures, such as WGAN-GP, for improved stability.
- Applying the approach to other medical imaging tasks.