Enhancing Parkinson's Disease Diagnosis through Synthetic Image Augmentation and Deep Learning Model Evaluation

Mosarrat Rumman[*], Heidar Davoudi, Mehran Ebrahimi
Faculty of Science, Ontario Tech University
Mosarrat.Rumman@ontariotechu.net
{Heidar.Davoudi, Mehran.Ebrahimi}@ontariotechu.ca

Abstract

Parkinson's disease (PD) is a progressive neurodegenerative disorder that can be clinically diagnosed through various neuroimaging techniques. Single-photon emission computed tomography (SPECT) has proven to be an effective tool for the early detection of PD. Automatic detection of PD from SPECT images, using machine learning or deep learning models is crucial for providing faster, more accurate diagnoses, and facilitating early intervention. While large datasets of SPECT scans for PD are available, they are often highly imbalanced, which can significantly hinder the performance of deep learning models. In this paper, we explore how synthetic image generation can address the dataset imbalance problem and improve the accuracy of deep learning models. We evaluated the performance of several state-of-the-art pre-trained deep learning models, including Vision Transformer (ViT), VGG-16, EfficientNet, and a newly proposed hybrid model, Inception-VGG16. Experimental results demonstrate that augmenting the dataset with synthetic images significantly improves the performance of all models, with ViT achieving the highest test accuracy of 98%. The proposed Inception-VGG16 model performed second best, achieving a test accuracy of 95%. These results suggest that synthetic augmentation can enhance the performance of pre-trained models in detecting Parkinson's disease, presenting a promising approach for enhancing automatic diagnostic tools. The implementation of this work is available at this GitHub Repository.

1 Introduction

Parkinson's disease (PD) is a progressive neurodegenerative disorder caused by the loss of the dopaminergic neurons in the substantia nigra region in the midbrain [1]. The symptoms include tremors, stiffness, bradykinesia (slowness of movement), and hypokinesia (reduced movement), postural instability, falls, orthostatic hypotension, and dementia [2]. Parkinson's disease can be detected using neuroimaging techniques such as Magnetic Resonance Imaging (MRI), Single Photon Emission Computed Tomography (SPECT), Positron Emission Tomography (PET), Computed Tomography (CT), etc [3]. However, SPECT can detect PD at an early stage and can be used in the differential diagnosis between PD and non-degenerative forms of Parkinson's [4]. Computer-aided diagnosis can enhance Parkinson's disease detection through advanced image analysis and machine learning, providing earlier, more accurate diagnosis. Several works have been done on computer-aided detection of Parkinson's disease from brain scans using traditional machine learning models such as Support Vector Machine, Regression, and Decision tree, as well as deep learning models such as Neural networks showing promising results [5, 6]. This work evaluates the performance of some of the state-of-the-art pre-trained deep learning models in diagnosing PD. However, an imbalanced dataset can adversely affect the performance of the diagnostic models. There are various kinds of data augmentation techniques such as oversampling, undersampling, flipping, rotating, etc, but using generative models to produce synthetic images has shown promising results in augmenting datasets [7]. This research investigates how using synthetic images can improve the performance of diagnostic models. The objective of this research is stated as follows:

- To balance an imbalanced dataset by generating

[*] Corresponding Author

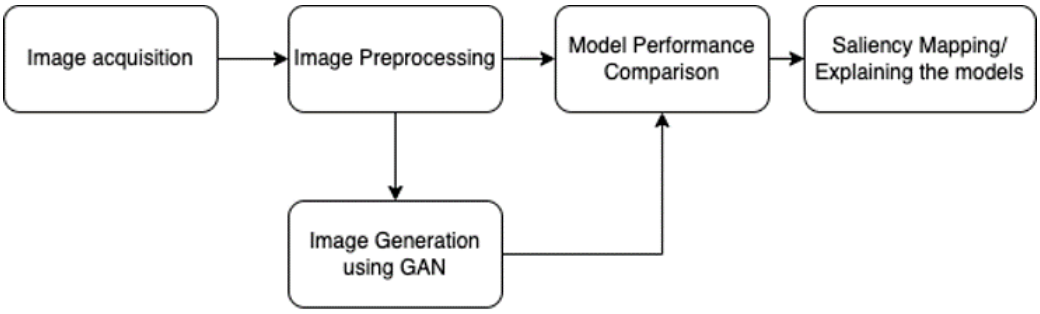
Image acquisition
Image Preprocessing
Model Performance Comparison
Saliency Mapping/ Explaining the models
Image Generation using GAN

Figure 1: Overview of the proposed methodology

synthetic SPECT images and demonstrate the impact of the generated images on the performance of the classification models.

- Leverage transfer learning to classify SPECT images for PD detection using various pre-trained models and comprehensively evaluate the performance of the models.

- To propose a new model- Inception-VGG16 which provides a higher accuracy compared to using VGG-16 alone.

2 Methodology

As shown in Figure 1, the workflow outlines key steps for diagnosing Parkinson's disease using pre-trained deep learning models with synthetic image generation. The process starts with image acquisition from the database, followed by pre-processing to enhance the image quality. A Generative Adversarial Network (GAN) generates synthetic images to balance the dataset. The preprocessed and augmented data is then used for model performance comparison. Finally, saliency mapping explains the models' decisions by identifying key regions in the images. Each step is briefly detailed in the following subsections.

2.1 Data Acquisition and Preprocessing

The dataset has been collected from the Parkinson's Progression Marker's Initiative (PPMI) database. A total of 1442 SPECT images were collected, of which 1,157 belong to the PD class, and 285 belong to the healthy control or non-PD class. Scans were taken between the 12th month and 156th month of the disease progression, and the volunteers aged in the range of 35 to 50 years.

The reconstructed SPECT images collected from PPMI are 3D images in the DICOM format. For standardization and convenience, the 3D images are converted to 2D in the PNG format. From the 96 slices, the 42nd slice was chosen for all images as the Region of Interest (ROI) is most visible in this slice. To further pre-process the images, Horizontal and vertical flips were performed to induce variability in the data. The images were blurred using Gaussian blur to reduce noise and resized to standardize the size of the images, preparing them as training models.

2.2 Synthetic Image Generation

The dataset collected from PPMI was highly imbalanced, with 1,157 images of PD cases and 285 non-PD images. This imbalance can hinder the performance of deep learning models. To address this issue, StyleGAN3, a Generative Adversarial Network (GAN) developed by NVIDIA [8], was employed to generate approximately 700 synthetic images for the non-PD class. These synthetic images were visually similar to the original non-PD images, at least confirmed by 1D histograms showing comparable distributions. Some abnormal images were detected and removed using the Alibi Detect outlier detection model [9]. To prevent the ratio of original to synthetic images from being too large, the PD class was undersampled from 1,157 to 996 images. Balancing the dataset at exactly 960 images for both classes would have resulted in a higher proportion of synthetic images in the non-PD class compared to the PD class. This imbalance could introduce biases in the model, as it might learn features that are more reflective of the synthetic data rather than the true characteristics of the original data. The choice of 996 samples of PD class preserved a closer balance between original and synthetic images. Ultimately, the dataset was balanced, consisting of 996 PD cases and 960 non-PD cases, allowing for improved model performance.

2.3 Model Overview

In this research, four models are utilized for analysis, comprising three existing pre-trained models—Vision Transformer (ViT) [10], VGG-16 [11], and Efficient-

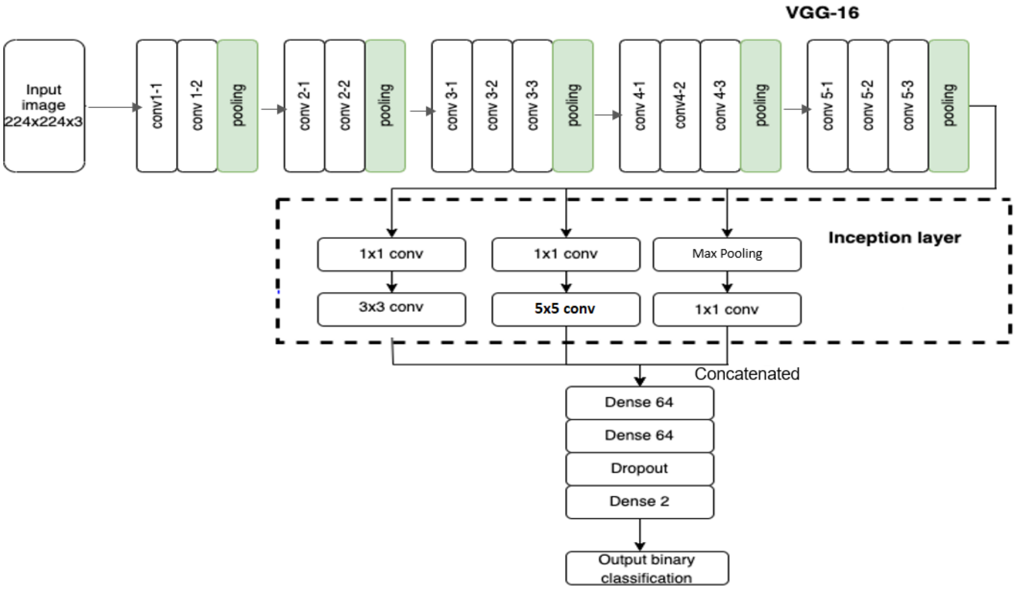

Figure 2: Inception-VGG16 architecture

Net [12] and a proposed novel model, named Inception-VGG16. Each model offers unique characteristics and strengths, which are briefly described below:

- **Vision Transformer** (ViT) is a model that applies the transformer architecture. Instead of scanning the whole image at once, ViT partitions the image into small, equal-sized pieces and then examines each piece through the transformer layers. Pre-trained on a large dataset, ViT has shown promising results in image classification tasks.

- **EfficientNet** is a convolutional neural network (CNN) architecture that optimizes both accuracy and computational efficiency. Introduced by Google researchers, it systematically scales the dimensions of depth, width, and image resolution through a compound scaling method. This balanced scaling results in higher performance levels with reduced computational need, making EfficientNet a powerful model for a wide range of image-processing tasks

- **VGG-16** is another CNN model with a deep architecture consisting of 16 layers, including 13 convolutional layers followed by 3 fully connected layers. It employs small (3x3) convolution filters throughout images, allowing it to capture fine details from images effectively. Despite its straightforward structure, VGG-16 has achieved remarkable success in image recognition tasks and is widely used as a feature extractor in various computer vision applications.

- **Inception-VGG16**, proposed in this work, is a hybrid model combining VGG16 and Inception V3 [13]. The Inception layer applies multiple convolutional filters (1x1, 3x3, 5x5) and max-pooling in parallel to capture features at different scales. In this model, outputs from the final pooling layer of VGG16 are fed into an Inception layer with five Conv2D layers (2x512, 2x128, 1x64), followed by dense layers for classification. This structure enhances the model's ability to recognize diverse objects and shapes. The architecture of the model is shown in Figure 2.

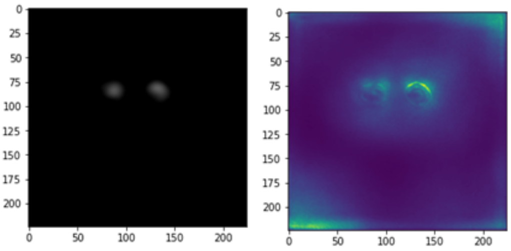

Figure 3: Saliency mapping of Inception-VGG16 model

3 Experimentation and Results

In the experiment, the models were trained and evaluated on both the imbalanced dataset (1,157 PD, 285 non-PD) and the augmented dataset (996 PD, 960 non-PD). The data was split into training (70%), validation (20%),

Table 1: Model's Performance Evaluation with Precision and Recall

Model	Accuracy without Generated Images	Precision without Generated Images	Recall without Generated Images	Accuracy with Generated Images	Precision with Generated Images	Recall with Generated Images
Vision Transformer	77%	61%	78%	98%	99%	97%
EfficientNet	77%	60%	77%	90%	90%	89%
VGG-16	76%	59%	76%	87%	89%	87%
Inception VGG16	77%	60%	77%	96%	97%	96%

and testing (10%) sets to ensure a balanced performance assessment.

For Vision Transformer (ViT), the DeiT Tiny model [14] was chosen for its computational efficiency, with pre-trained weights fine-tuned using the dataset. VGG-16 and EfficientNet-B7 models were also fine-tuned using pre-trained weights from ImageNet. EfficientNet's classification layer was also replaced by a Sigmoid Linear Unit (SiLU) activation function [15] for better performance. The proposed Inception-VGG16 model, a hybrid of VGG-16 and InceptionV3, incorporated an Inception layer to enhance feature extraction. Adam and AdamW optimizers with a learning rate of 0.001 were used for training. Results demonstrate significant improvement with the augmented dataset, with ViT achieving the highest test accuracy of 98%, followed by Inception-VGG16 at 95%. Table 1 shows a comprehensive evaluation of the performance of the models on both balanced and imbalanced datasets. Additionally, Saliency maps [16] were also created by computing the gradients of the model's output with respect to the input image. These gradients highlight which pixels in the image are most influential in the model's prediction. In Figure 3, the image on the left represents the original image, while the right image is the corresponding saliency map showing how the model interprets regions of interest. The heatmap can help validate the model by ensuring it focuses on medically relevant areas in the images.

Discussion

The results highlight the effectiveness of balancing datasets and the comparative performance of different models. When trained on imbalanced datasets, all models, including ViT, EfficientNet, VGG16, and the proposed Inception-VGG16, struggled with poor performance for the minority class (non-pd class), as reflected in low precision, recall, and accuracy. However, training on a balanced dataset significantly improved the performance of all the models. While ViT achieved the highest accuracy on the balanced dataset (98%), the proposed Inception-VGG16 model achieved the second-best result with an overall accuracy of 97% with a weighted precision of 97% and a weighted recall of 96%. The proposed Inception-VGG16 model offers better interpretability than ViT due to its convolutional architecture, which captures localized, pixel-level features and enables clear visualization of regions influencing predictions. This is crucial for identifying subtle patterns in sensitive applications like medical imaging. In contrast, ViT's global self-attention mechanisms can make its decision-making process less transparent and harder to interpret for high-stakes use cases. In addition, ViT's transformer-based architecture, though highly accurate, is computationally intensive and requires substantial resources, making it less practical for resource-constrained environments.

Limitations

This study has limitations due to the conversion of 3D SPECT images into 2D by manually selecting the 42nd slice. While this simplifies computational requirements, it may not always capture the most diagnostic slice. Critical patterns or features might be more visible in other slices, and relying solely on a single slice introduces subjectivity and risks omitting key information necessary for accurate classification.

Another significant concern is potential data leakage. If multiple samples from the same patient are included in the training, validation, and test sets, the model may learn patient-specific features rather than generalizable patterns. This could lead to inflated performance metrics and misrepresent the model's real-world applicability. Patient-level data splitting is essential to ensure unbiased evaluation.

Moreover, the model was only tested on SPECT images, limiting its generalizability to other imaging modalities such as MRI, CT, or PET scans. Each modality presents unique challenges, and performance in one does not guarantee similar results in others. Expanding the study to include multimodal datasets or using 3D image data could provide more robust and comprehensive insights into the model's applicability across clinical settings.

4 Conclusion

This work demonstrates that augmenting the dataset with synthetic images can significantly enhance the performance of pre-trained models. Among the models evaluated, Vision Transformer (ViT) outperformed the other models in classifying SPECT images to detect Parkinson's disease. The proposed Inception-VGG16 model, while not surpassing ViT, showed promising results, achieving the second-highest accuracy compared to VGG16 and EfficientNet. Additionally, saliency mapping could provide valuable insights by explaining the model's decision-making process. Future work would involve testing it on larger and more diverse datasets to further validate the robustness and reliability of the proposed Inception-VGG16 model.

References

[1] G. DeMaagd and A. Philip, "Parkinson's disease and its management: part 1: disease entity, risk factors, pathophysiology, clinical presentation, and diagnosis," *Pharmacy and therapeutics*, vol. 40, no. 8, p. 504, 2015.

[2] C. E. Clarke, "Parkinson's disease," *Bmj*, vol. 335, no. 7617, pp. 441–445, 2007.

[3] G. Pagano, F. Niccolini, and M. Politis, "Imaging in parkinson's disease," *Clinical Medicine*, vol. 16, no. 4, pp. 371–375, 2016.

[4] L. Wang, Q. Zhang, H. Li, and H. Zhang, "Spect molecular imaging in parkinson's disease," *BioMed Research International*, vol. 2012, no. 1, p. 412486, 2012.

[5] C. R. Pereira, D. R. Pereira, S. A. Weber, C. Hook, V. H. C. De Albuquerque, and J. P. Papa, "A survey on computer-assisted parkinson's disease diagnosis," *Artificial intelligence in medicine*, vol. 95, pp. 48–63, 2019.

[6] S. Sivaranjini and C. Sujatha, "Deep learning based diagnosis of parkinson's disease using convolutional neural network," *Multimedia tools and applications*, vol. 79, no. 21, pp. 15 467–15 479, 2020.

[7] M. Xu, S. Yoon, A. Fuentes, and D. S. Park, "A comprehensive survey of image augmentation techniques for deep learning," *Pattern Recognition*, vol. 137, p. 109347, 2023.

[8] T. Karras, M. Aittala, S. Laine, E. Härkönen, J. Hellsten, J. Lehtinen, and T. Aila, "Alias-free generative adversarial networks," *Advances in neural information processing systems*, vol. 34, pp. 852–863, 2021.

[9] A. Van Looveren, J. Klaise, G. Vacanti, O. Cobb, A. Scillitoe, R. Samoilescu, and A. Athorne, "Alibi detect: Algorithms for outlier, adversarial and drift detection," 2019. [Online]. Available: https://github.com/SeldonIO/alibi-detect

[10] A. Dosovitskiy, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[11] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[12] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.

[13] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[14] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jegou, "Training data-efficient image transformers and distillation through attention," in *Proceedings of the 38th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 139. PMLR, 18–24 Jul 2021, pp. 10 347–10 357.

[15] S. Elfwing, E. Uchibe, and K. Doya, "Sigmoid-weighted linear units for neural network function approximation in reinforcement learning," *Neural Networks*, vol. 107, pp. 3–11, 2018.

[16] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," *arXiv preprint arXiv:1312.6034*, 2013.