National College of Ireland

Leveraging Machine Learning and GANs for Parkinson disease detection

MSc Research Project
Data Analytics

Calista Gonsalves
Student ID: 22186077

School of Computing
National College of Ireland

Supervisor: Paul Stynes, Musfira Jilani and Mark Cudden

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Calista Clifford Gonsalves		
Student ID:	22186077		
Programme:	Msc Data Analytics	**Year:**	2023-2024
Module:	Research Project		
Supervisor:	Paul Stynes, Musfira Jilani and Mark Cudden		
Submission Due Date:	12-08-2024		
Project Title:	Leveraging Machine Learning and GANs for Parkinson disease detection		
Word Count:	6904	**Page Count:**	22

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Calista Clifford Gonsalves

Date: 12-08-2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Leveraging Machine Learning and GANs for Parkinson disease detection

Calista Gonsalves
22186077

Abstract

Parkinson's Disease (PD) is a progressive disorder that affects the nervous system and the parts of the body controlled by nerves. Early-stage detection of PD using spiral and wave images can significantly improve patient outcomes. Many previous researchers studied classification of PD using various sources, but they faced a few limitations like reduced dataset size. Developing a machine learning framework that processes a large and varied dataset, maintains computational efficiency, and achieves high accuracy in PD detection can be a challenge. This research proposes a machine learning framework to improve the early detection of PD. The proposed framework uses the combined images generated using Custom GAN (Generative Adversarial Network) and the original dataset with both hybrid (ResNet50 and InceptionV3 with KNN classifier) and standalone pre-trained CNN models(ResNet50, InceptionV3). The hybrid model consists of a pretrained CNN model along with a machine learning classifier. This framework makes use of a handwritten Parkinson’s disease dataset (original dataset) that comprises of 1632 images each of Parkinson and healthy, for Custom GAN image generation, pre-trained CNN model training and testing. Data pre-processing and transfer learning techniques are applied to two pre-trained CNN models, namely ResNet50 and InceptionV3. Each of these models is evaluated both individually and in combination with KNN classifier. Results of these models are presented in this paper based on accuracy, sensitivity, specificity, F1 score, Cohen’s Kappa and precision. This research shows how this framework can prove useful for patients by detecting PD at an early stage.

1 Introduction

Parkinson's disease is a brain disorder that causes trouble with movement and balance, making it hard for a person to control their movements and perform everyday tasks as the disease progresses. Early detection of PD is important as it helps to manage symptoms in a better way and can also help in starting treatment sooner. Parkinson’s disease doubled to over 6 million between the years 2015-2019 and is expected to double again to over 12 million by 2040 (Dorsey *et al.*, 2018). Due to this increase in the number of cases, it is important to identify Parkinson’s disease at its early stage so that necessary precautions can be taken. Traditionally, the common scale that is used to detect PD is Unified Parkinson's Disease Rating Scale (UPDRS) and even its modified version Movement Disorder Society-UPDRS (MDS-UPDRS), but these have their own limitations (Hendricks and Khasawneh, 2021), hence there is a clear need to look for alternative methodologies/solutions.

The aim of this research is to facilitate the early diagnosis of PD in patients by identifying the condition using spiral and wave images. The major contribution of this research is a novel machine learning framework which includes image generation using Custom GAN, along with hybrid and standalone pretrained CNN models. The main reason for exploring this topic is because diagnosing PD with the help of spiral and wave images is cheap for patients and hence anyone with Parkinson's symptoms can easily get tested.

Most of the research that has been done in this field, has been limited by small dataset, such as the study of Chakraborty *et al*. (2020), which was trained on the data of 55 patients. Hence, there is a clear need to explore methodologies that utilize varied and diverse data. By doing so, the model can be more reliable as it will be able to perform better on unseen data, thereby enabling to see patterns that might not be apparent in smaller datasets. This paper aims to address limitations from previous studies and improve upon them.

The proposed methodology aims to bridge the gap found in past papers by using a GAN combined dataset, that is a combination of Custom GAN generated images and Parkinson's disease dataset (Original dataset). Due to high computational requirements, only 11% of the Custom GAN generated saved images were used, which is 3736 generated images and 3264 images of the Parkinson's image dataset (Original dataset), which makes a total of 7000 images. In the hybrid models, the pretrained models (ResNet50, InceptionV3) are utilized as a deep learning model to extract features, which will be used by KNN classifier to identify PD. By using advanced machine learning like Custom GAN and pretrained CNN models, detailed patterns can be discovered that can help in the future work of medical field and to improve treatments related to PD. Additionally, this research can help in improving the machine learning that is being used in medical diagnosis and can lead to similar methods to be used for diagnosing other brain/medical conditions.

The research is based on the research question:

To what extent can Custom GAN-generated images combined with original images, using standalone CNN models (ResNet50, InceptionV3) and hybrid models (ResNet50, InceptionV3 with KNN classifier) improve early detection of Parkinson's disease?

The research question tries to find out if the images generated using Custom GAN when combined with original images helps in detecting PD, when used with pretrained CNN models. The research also investigates into which pretrained CNN model will help in early detection of PD and whether or not it is effective. By correctly identifying PD from hand drawn images like spiral and wave images, this research can help doctors to start treatment little early and also benefit patients by effectively managing symptoms.

Towards addressing the research question, the study investigates the following research objectives:

1. To assess the effectiveness of Custom GAN generated images along with pretrained CNN models in extracting relevant features from spiral and wave images of hand-drawings to identify PD.
2. To compare the performance of the proposed framework with existing methods in the literature and to highlight improvements and limitations of the proposed work.
3. To investigate the effectiveness of hybrid models (ResNet50, InceptionV3 with KNN classifier) and standalone models in the detection of Parkinson's disease.

4. To assess the computational efficiency of the proposed framework in practical settings.

The main contribution of this research is generating images using Custom-GAN which is inspired by Super-Resolution GAN (SRGAN) also known for generating high-resolution images, but it makes use of complex architecture. SRGAN makes use of residual blocks to enhance image quality using deep network architectures (Ledig *et al.*, 2017). In contrast, the Custom-GAN improves image quality by using a simpler architecture, thereby omitting the residual blocks used in SRGAN. The images generated using Custom-GAN are combined with the Parkinson’s disease dataset (original dataset) for model training and testing of standalone CNN models and hybrid model which is useful for detecting PD at an early stage. The standalone models include ResNet50 and InceptionV3, while the hybrid models combine each of these pre-trained CNN model with KNN classifier.

This paper discusses deep learning models on GAN combined data and its ability to classify PD. The deep learning models and GANs used are discussed in Section 2 Related work. The research methodology is discussed in Section 3. The design specification in section 4. Section 5 discusses the Implementation section and Section 6 discusses Evaluation. Section 7 concludes the research and discusses future work.

2 Related Work

Machine Learning has enhanced the early detection of Parkinson's disease. PD affects around 1% of population above the age of 60 and is a common disorder (Anila and Gera, 2022).

2.1 Deep learning models explored for classification of Parkinson Disease

Islam and Akter (2020) explored the detection of PD using hand drawn images like spiral and wave. The proposed work utilized Histogram of Oriented Gradients (HOG) for feature extraction, which performed well along with Gradient boosting achieving an accuracy of 86.6% and KNN with an accuracy of 89.3%. But the study even emphasized that by increasing the data and using deep learning techniques the solution can be improvised, which will be done in the current research. Kumar and Bansal (2023) focus on using spiral and wave drawings for early detection of PD. The data is augmented by rotating from 204 images to 3264. Modified MobileNetV2 is used which achieves an impressive accuracy of 97.7% even though fewer parameters are used as compared to other models. But one of the main limitations is that it does not focus/discuss other evaluation metrics in details.
Das *et al.* (2021) focus on early detection of PD on 2 different datasets by analysing hand drawn images like spiral, waves, cubes and triangles drawn by the patients. Three different approaches are used, like using various pretrained models like VGG19, ResNet50, MobileNet-v2, InceptionV3, Xception and Inception-Resnet-v2, applying transfer learning with fine-tuning and using two shallow CNNs. VGG19, ResNet50, MobileNet-v2 achieve an accuracy of 91.6% and 100% for first and second dataset respectively. The limitation for this paper is that the size is limited for both the datasets. Similarly, Rajinikanth, Yassine and Bukhari (2024) aimed at detecting PD using wave sketches. Various methods are implemented like image

preprocessing, data augmentation, two fold training and deep feature selection with 50% drop out rate and binary classification. Various CNN models like MobileNet, SqueezeNet, ShuffleNetV1 and NASNet were tested out of which the fused features of MobileNet along with KNN classifier offered 100% detection accuracy.

In Canturk and Gunay (2024), the main aim is to identify PD from scalogram images which are created from speech signals. Various pre trained CNN models like AlexNet, ResNet50 and GoogleNet, along with hybrid system and deep fusion feature method (NasNet and DenseNet) are implemented and verified for 'u', 'o', 'a' phonation. Deep fusion feature along with KNN classifier performed well, having all scores of 0.95 and validated with 10-fold cross validation. The main limitation is that the rate of data sampling can affect the work of classifier, as the rate increases the computational complexity. Anila and Gera (2022) focus on detecting PD by analyzing vocal signals of healthy and PD patients. Noise was removed using median filtering and wavelet transforms. 1D audio signals were converted to 2D time frequency diagram. ResNet50 was used to classify the 2D diagrams into PD and non-PD, the model achieved an accuracy of 86.86% and F1 score of 0.89. The limitation that can be observed here is that the dataset is considerably small, consisting recordings of just 20 individuals.

Yang *et al.* (2022) identified people with PD using gait data, which are then processed into 3d images. Techniques like SMOTE, data augmentation, early stopping and a new loss function called focal loss helps the novel model PD-ResNet to discard abnormal samples, achieving accuracy of 95.51% and accuracy for early PD and PD with different severity levels is 92.03%. The future work focuses on expanding the dataset and keeping it diverse.

Quan *et al*. (2019) make use of data obtained from PPMI databases, to classify DaTscan SPECT images as PD or normal. InceptionV3 was used as a base model with a binary classifier block on top of it and 10-fold cross validation was performed to evaluate the model’s performance. The model achieved an accuracy of 98.48%, but it was even highlighted that the high accuracy was due to the small size of data and class imbalance. Similarly, Kurmi *et al.* (2022) suggested a method that outperforms other state of the art methods in detecting PD using DaTscan images, by using a Fuzzy Fusion logic-based ensemble on VGG16, ResNet50, Inception-V3, and Xception using DaTscan images, the results show that ensemble method performs better than the individual models by achieving an accuracy of 98.45%. A GUI-based software tool was also developed in detecting the disease in real time. The limitation highlighted here was the number of misclassifications which should be reduced.

Vaidya *et al.* (2024) make use of the 204 handwritten images of spiral and wave and are augmented to 3264 to classify PD. Models like Xception and EfficientNetB2 were used on the augmented dataset, where, EfficientNetB2 performed the best, resulting in an accuracy of 96.4% .A flask based web application is created using EfficientNetB2 where users can upload images to find out if or not they have PD. This study highlighted the need of a better approach that can balance sensitivity and specificity. Shaban (2020) used the same dataset of 204 images, which is preprocessed and augmented to 800 images to classify PD. Fine-tuned VGG19 model is used on preprocessed data and is validated using 4 and 10 fold cross-validation technique, achieving an accuracy of 88% and 89% on wave and spiral pattern, this study achieved high performance when compared to tuned AlexNet. Image Rotation within data augmentation was highlighted to reduce overfitting. Morales-Castro *et al*. (2022) also used similar 204 hand-drawn images for detection of PD. ResNet50 and Oriented Gradient Histogram were used as feature extractors, along with classifiers like SVM, KNN, Logistic Regression, out of which ResNet50 along with SVM classifier worked well achieving an accuracy of 90%. However, one possible limitation is that the entire study was conducted on 204 images which is very small.

From the above section, it can be concluded that papers that made use of ResNet50 and InceptionV3 achieved good accuracy but highlighted limited dataset as a disadvantage, which will overcome in the current research by making use of Custom GAN. Also, it can be clearly seen that KNN acts as a good classifier. Hence, the current research makes use of KNN as a classifier in hybrid models.

2.2 Data augmentation using GAN explored in the field of Parkinson Disease Detection

Zanini and Colombini (2020) performed two new data augmentation, namely DCGANs and Style Transfer for augmenting PD's EMG (Electromyography) signals. The main findings of this study are that DCGAN with discriminator CNN pipelines can simulate EMG tremor behavior which can help in the development of reducing tremors. By using Style Transfer, the tremor patterns were transferred to different datasets and protocols. It even highlights the use of these methods for expanding the patient's dataset.

Xu *et al.* (2020) proposed a new ML model for augmenting voice samples called Spectrogram Deep Convolutional Generative Adversarial Network (S-DCGAN) to distinguish between PD patients and healthy individuals. ResNet50 with GAP layer is used to extract features and classify them on Sakar dataset. The S-DCGAN-ResNet50 hybrid model achieved an accuracy of 91.25. The results show that the accuracy improves by using augmented samples, as the model proposed performs well even on adding small samples of images, it also highlights the feasibility and effectiveness of GAN for augmentation of PD dataset.

Peppes *et al*. (2023) make use of a PD symptom called freezing of gait, here FoGGAN is used where generated data that is almost identical to the original one is used for classification of PD. The DNN classifier was tested on original, generated and mixed dataset, with an accuracy of 90%,92% and 90% respectively. The study also highlights the use of GAN for data augmentation to address the issue of data limitations and to improve model performance as a promising strategy in real world applications.

Dzotsenidze *et al.* (2022) compare four different GAN models (StyleGAN2-ADA + LeCam, StyleGAN3, and ProjectedGAN) with the traditional data augmentation methods for Parkinson’s disease diagnostics using spiral images. Various pre-trained CNN methods are used for classification, and the GAN generated images outperform the traditional augmentation methods. Projected GAN performed well with ResNet50 and Xception achieving a sensitivity of 96%. The future work highlights to explore more GAN architectures.

From the above section, it can be understood that GAN is preferred for data augmentation and it provides better results than the standard data augmentation methods, hence GAN is used in this project to address the limitation of small dataset found in previous papers. Custom GAN will be incorporated, which is inspired by SRGAN, as it is known to produce high resolution clear images (Liu and Chen, 2021).

3 Research Methodology

The research methodology consists of 4 steps mainly Data Gathering and Generation, Data Preprocessing, Data Modelling and Testing and Evaluation as shown in figure 1. In this research these 4 steps are carried out, to achieve the desired outcome of classification of PD.

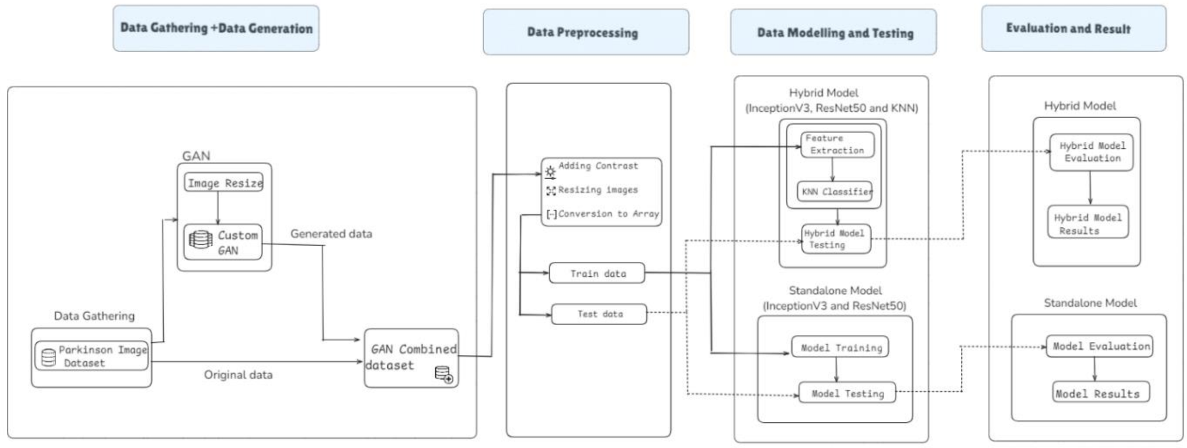

Figure 1: Research Methodology

3.1 Data Gathering and Data Generation

The first step is a combination of **Data Gathering and Data Generation,**
Data Gathering: The data is collected from Kaggle[1]. The Parkinsons Image dataset (original data) is split into folders, namely Parkinson and Healthy. Each folder comprising of 1632 spiral and wave images. As discussed in the above section, since limited and varied data is highlighted in many research papers, it is important to generate more data.

Data Generation: In the Literature review, under GAN section (Section 2.2) it was discussed how GANs outperform traditional augmentation methods, hence another section named Data Generation is proposed which will generate more data using Custom GAN. The generator and discriminator function makes use of simple convolution network to produce high quality images, also Adam optimizers are used with different learning rates. Loss functions like BCE loss is used for adversarial loss and L1 loss is used for the content loss. Custom GAN runs for 20 epoch for each Parkinson and Healthy with a batch size of 8. Custom GAN generates 1632 images for each epoch, but out of 20 epochs, images for just last 10 epochs are saved because of high computational requirements, also the images generated by the last epochs are sharper (An *et al.*, 2020). So, the Custom GAN generated saved images consists of 16320 images (1632*10 epochs), each for Parkinson and Healthy. Figure 2 shows the output of Custom GAN generated images of Parkinson and Healthy as well as the original images.
The first step combines the original dataset with the Custom GAN generated saved images. Due to high computational requirements, only 11% of the Custom GAN generated saved images were used, that is 11% of 16320 which equals to 1868 Custom GAN images, each of

[1] https://www.kaggle.com/datasets/banilkumar20phd7071/handwritten-parkinsons-disease-augmented-data

Parkinson and Healthy. Hence, 3736 Custom GAN images (that is 1868 Custom GAN images each for Parkinson and Healthy) and 3264 original images(1632 images each for Parkinson and Healthy from the original dataset) were combined and used. The breakdown of the images can be seen as follows:

- **Custom GAN Images**: 1868 images each for both Parkinson and Healthy classes were used, summing up to 3736 Custom GAN generated images.
- **Original Dataset Images**: 1632 images each for both Parkinson and Healthy classes from the original dataset were used, summing up to 3264 original images.

In total, the **GAN combined dataset** comprises of **7,000 images**:

- 1868 Custom GAN images + 1632 original images for Parkinson = 3500 images for Parkinson
- 1868 Custom GAN images + 1632 original images for Healthy = 3500 images for Healthy

The distribution of the GAN combined images can be seen in figure 3. For all the further methods in this research, the GAN combined dataset with 7000 images were used.

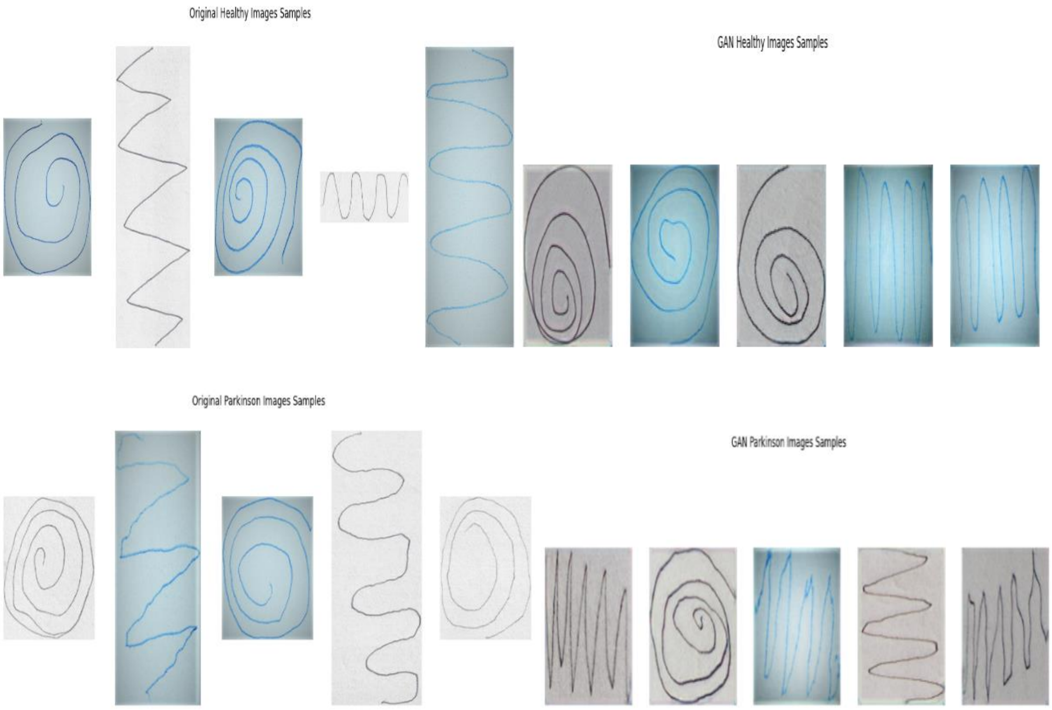

Figure 2: Original (left) and Custom GAN generated images (right)

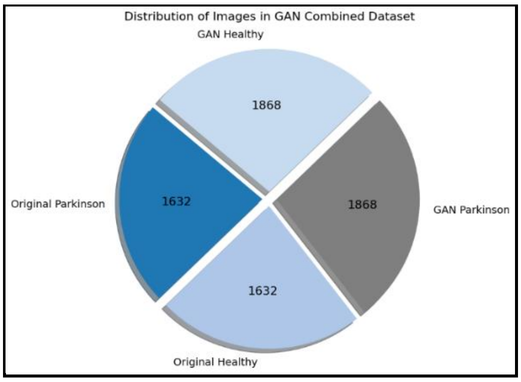

Figure 3: Distribution of images in GAN combined data

3.2 Data Preprocessing

The second step is **Data Preprocessing,** in order to make the GAN combined dataset compatible for modelling, the data needs to be preprocessed in order to ensure that all the images are in one format. The first step in this stage is to ensure that all the images are in RGB format. Based on every model the images are resized, like for ResNet50 – KNN hybrid model the images are resized to (128,128), InceptionV3-KNN hybrid model and Inception V3 (Meena, Mohbey and Kumar, 2023) takes an input of size (224,224), ResNet50 images are also resized to (224,224) to match the input size of the model (Zahisham, Lee and Lim, 2020). Once the images are resized, the images are enhanced by increasing the brightness by 20%, figure 4 shows the images before and after enhancement. After applying all the above steps, the images are normalized by scaling the pixel values in the range of [0,1]. The GAN combined images are now preprocessed and are ready for modelling section. Before passing it to the modelling section, the data is split into train and test. Histogram equalizer was implemented to enhance the images, but the quality of the image degraded after applying it, hence the histogram equalizer method was removed, figure 5 shows the output of histogram equalizer.

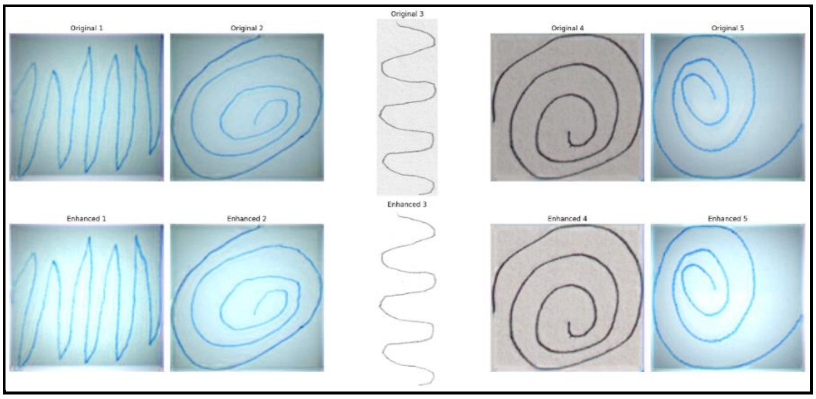

Figure 4: Images before and after increasing the brightness by 20%

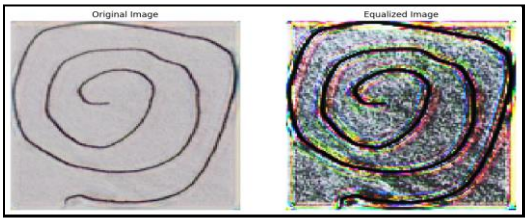

Figure 5: Effects of applying Histogram equalizer

3.3 Data Modelling

The third step is **Data Modelling** which involves model training and model testing on GAN combined dataset. The models are trained on 5600 images and tested on 1400 images. In this research, there are a total of 4 models namely, 2 standalone models: ResNet50 and InceptionV3 and 2 hybrid models that is ResNet50 with KNN classifier and InceptionV3 with KNN classifier. ResNet50 and InceptionV3 are imported from TensorFlow library. For the hybrid models, the pretrained CNN models are used as a feature extractor, and the features extracted are passed to KNN classifier. Hyperparameter tuning is performed on the KNN classifier to

find the optimal value of the number of neighbors (n_neighbors), the weight function used in prediction (weights), and the distance metric (metric) . Five-fold cross validation is implemented to find the best parameters for the KNN classifier.
For the standalone pretrained CNN models like the ResNet50 and InceptionV3, the models are loaded with pretrained weights but the top fully connected layers are removed. The last 20 layers are unfrozen and additional layers are added. The model is compiled with the Adam optimizer, a learning rate, and loss function. The model is trained for 10 epochs with a batch size of 16. Once the hybrid models and standalone CNN models are trained on the training dataset, the model is tested on the test data (1400 images).

3.4 Evaluation and Results

The quality of the Custom GAN generated image is verified using SSIM (Structural Similarity Index Measure) to ensure if the generated images are of good quality. Once the model is trained, and tested on test dataset, different evaluation metrics are checked like Accuracy, Sensitivity, Specificity, Precision, F1 score and Cohen's Kappa. The four models were compared using the metrics mentioned above. The results that are obtained are compared with the research objectives to understand various insights.

4 Design Specification

4.1 Custom GAN Architecture

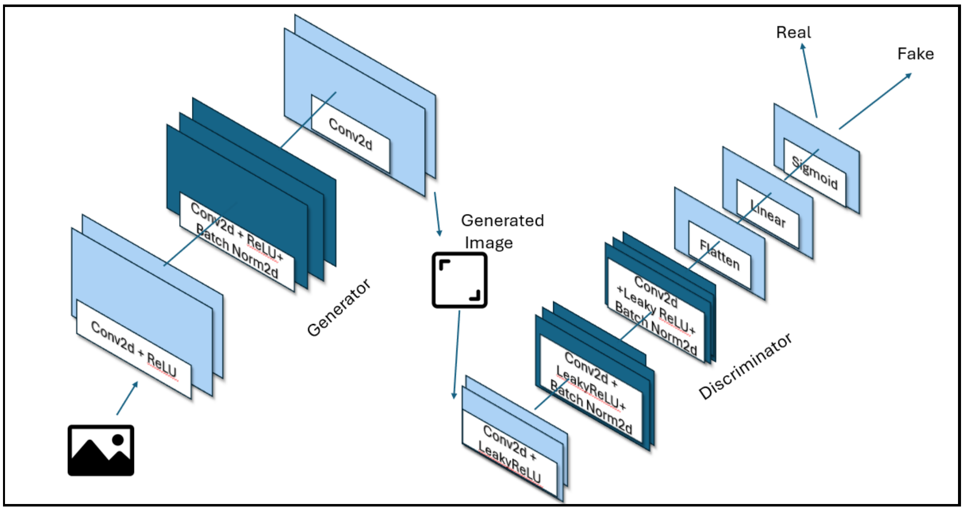

Figure 6: Custom GAN Architecture

The Custom GAN makes use of a generator and discriminator, which works together to generate high quality images. The original dataset is preprocessed by resizing into 128*128 pixels and is then passed to the generator. In the generator section, the first layer uses a 9x9 convolution with 64 filters which helps to capture large features and details. The ReLU activation function introduces non-linearity which helps in learning complex patterns. The second convolutional layer uses 3x3 convolution with 64 filters which helps in refining features. Batch Normalization Layer normalizes the output of the previous layer to stabilize and speed up the training and helps in producing high quality images by maintaining consistent feature distributions, followed by ReLU Activation Layer which introduces non-linearity. The final layer uses 3x3 convolution and reduces the number of filters back to 3 which helps the output image to maintain the high quality details learnt in the previous layers. Similarly, the discriminator uses 3x3 convolution to extract low level features which is followed by the LeakyReLU activation layer. The second convolutional layer uses 3x3 convolution with 64 filters and stride of 2 for downsampling, which helps to focus on high level features. Similarly, batch normalization layer is used to stabilize training, followed by second LeakyReLU activation to introduce non-linearity. The convolutional layer with 3x3 convolution with 128 filters and stride of 2, helps to focus on more abstract features, followed by batch normalization and LeakyReLU activation. The flattening layer prepares the feature maps for final classification layer. The linear layer produces a single value, then the sigmoid layer produces a probability score between 0 and 1, indicating if the image is real or fake. In short, the generator uses large initial convolutions, with ReLU activations, and batch normalization to generate good quality images whereas discriminator uses LeakyReLU activations, strided convolutions for downsampling, and batch normalization for discriminating between real and fake. The generator is trained using content loss (L1 loss) which measures the difference between the generated image and original image (Wu, Xu and Hall, 2017). The discriminator is trained using BCE loss (Adversarial loss) which evaluates how well can it distinguish between real and generated images. This feedback is important because it guides the generator to produce images that appear more real. This combination of losses, like content loss for image accuracy and BCE loss for realistic images help in generating good quality images. Also, the training process involves updating the generator and discriminator with techniques like label smoothing and learning rate decay which helps in stabilizing the process and improving performance.

The use of convolutional layers in generator followed by ReLU activations and Batch Normalization is slightly similar to the SRGAN's approach (Ledig *et al.*, 2017) to generate high resolution images, also the discriminator has a similar structure with convolutional layers which reflects the SRGAN's design to learn and provide feedback to the Generator's discriminator. This combination of L1 loss (which ensure that generated images and original images match in terms of pixel level accuracy) and BCE loss (which encourages generator to create real images thereby fooling the discriminator) helps in generating realistic and high quality images which mirrors the approach that is used in SRGAN to achieve super resolution images

4.2 ResNet50 with KNN classifier

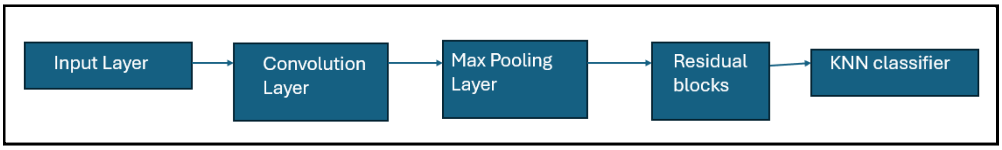

Figure 7: ResNet50 + KNN architecture

Here, the GAN combined image dataset (Custom GAN + Original dataset) will be used which is first preprocessed (resizing to 128*128, normalizing and enhancing brightness as discussed in the methodology section). In this, ResNet50[2] is utilized as a feature extractor where Convolutional layers capture image features, Max Pooling layers reduces the spatial dimension of the features. Residual blocks further captures complex features, which are then flattened by converting to 1D vector. This feature vector will then be used as an input to KNN classifier for classification. The KNN classifier is tuned using a grid search and validated using 5 fold cross validation to ensure good performance.

4.3 InceptionV3 with KNN classifier

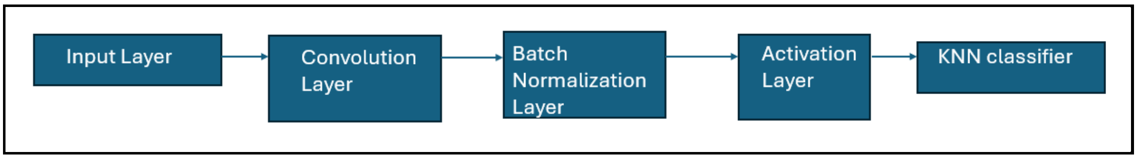

Figure 8: InceptionV3 + KNN architecture

In this, the GAN combined Image dataset (where all the preprocessing steps are same as the above section, except that the images are resized to 224*224) is fed to the model. The convolutional layers, batch normalization, and activation Layers extract high level features from the images and are then flattened to a 1D feature vector. This vector is then used by KNN classifier which is tuned using grid search and validated using 5-fold cross validation to ensure good performance for classification of PD.

4.4 InceptionV3

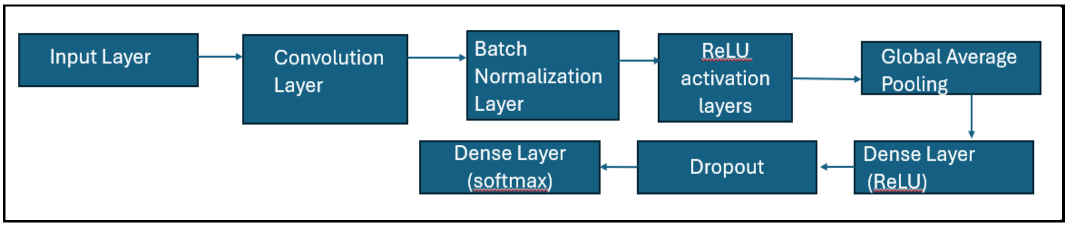

Figure 9: InceptionV3 architecture

The GAN combined dataset (same preprocessing steps except resizing to 224*224 pixels) is fed into InceptionV3 which is pretrained on ImageNet, which consists of Convolutional Layers, batch normalization, and ReLU activation to extract and process high level features. Last 20 layers of the model was unfroze which allows the weights of the layers to be updated(Cheng and Chiang, 2023). Then the global average pooling layer reduces these features into 1D vector, which are then passed through a fully connected dense layer with 256 units and ReLU activation, then followed by dropout layer with rate of 0.5 to reduce overfitting (Liang and Liu, 2015) and then a final dense output layer with softmax activation to classify the images into Parkinson or Healthy.

---

[2] https://wandb.ai/mostafaibrahim17/ml-articles/reports/The-Basics-of-ResNet50---Vmlldzo2NDkwNDE2#:~:text=Residual%20blocks%3A%20Residual%20blocks%20serve,the%20smooth%20flow%20of%20information.

4.5 ResNet50

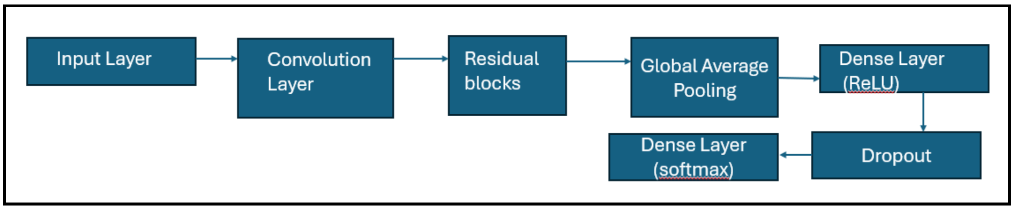

Figure 10: ResNet50 architecture

The GAN combined Image dataset (same preprocessing steps except resizing to 224*224 pixels) is fed into ResNet50 model which is pretrained on ImageNet which extracts high level features with the help of convolutional layers and residual blocks. The code unfreezes the last 20 layers of the model thereby allowing these layers to be fine-tuned. Then followed by a global average pooling layer that converts the features into 1D vector, this vector is then passed through a fully connected dense layer with 256 units and ReLU activation, then followed by a dropout layer with a dropout rate of 0.5 to reduce overfitting. The final dense layer with softmax activation then helps in the classification of PD.

5 Implementation

5.1 Setup

The original dataset is stored on drive. In order to generate GAN images, it is important to use a device that has sufficient computational power due to heavy processing. The generated images were also stored on drive. Generating images and performing classification on such a large dataset (7000 images) requires high computation, hence Google Colab is used with runtime as TPUv2 which provides 334 GB of RAM, which ensures smooth execution and handling of the computational demands. For this research, the entire code is written in Python.

5.2 Custom GAN

The code uses torch library and checks if GPU is available and sets the device accordingly. The dataset is resized to 128*128 pixels and is converted to tensors to prepare the images for input to the model. Subsets of the dataset are created like, batch size of 8 images. The Generator and discriminator are defined with convolutional layers as mentioned in the above section. The training function trains the generator and discriminator for 20 epochs. Initially, for the adam optimizer, the learning rate parameter of 0.0001 is used and is then decreased by a factor of 10 after 10th epoch. Another parameter of Adam optimizer, betas which is set to (0.5,0.999) is used which is the exponential decay rates, also the losses are computed and generated images after the 10th epoch (as discussed in the methodology section) are saved. The generated images are evaluated using SSIM metric. Only 11% of saved generated images (as discussed in methodology) are selected which are then combined with the original data to get a GAN combined dataset (7000 images).

5.3 Model Implementation

The libraries exported are TensorFlow, Keras and sklearn. The GAN combined dataset which contains of 7000 images has equal data that is 3500 Parkinson and 3500 Healthy. This dataset is then used for modelling. The 4 models used here are ResNet50, InceptionV3 and ResNet50, InceptionV3 with KNN classifier. After the preprocessing steps are performed which are discussed in the methodology section the GAN combined data is splitted in a ratio of 80:20 where 1400 images are used for testing and 5600 for training. The standalone CNN models like ResNet50 and InceptionV3 is used for the classification of PD which is compiled with Adam optimizer and trained for 10 epochs, with a batch size of 16 to manage memory efficiently, and a learning rate of 1e-4, and uses sparse categorical cross-entropy as the loss function here. The hybrid models use ResNet50 and InceptionV3 for feature extraction and then those features are sent to KNN classifier, where hyperparameter tuning is performed on parameters as shown in Table 1, using 5-fold cross validation. Using the best parameters obtained using hyperparameter tuning, the model is trained and tested. For both the standalone CNN models and hybrid model, the models are trained and evaluated to determine the best performing model for classifying PD.

Parameters	**Values**
Number of Neighbors	(3,5)
Weights	Uniform or Distance
Metrics (distance)	Euclidean or Manhattan

Table 1: Hyperparameters for KNN

6 Evaluation

This section helps to evaluate the results achieved for the detection of Parkinson Disease, also it helps in understanding how significant the results are. The research is evaluated on the following metrics.

For **Custom GAN**,

Structural Similarity Index Measure (SSIM): SSIM is a metric that is used for evaluating the quality of images, basically it checks the similarity between two images. In the context of GAN, SSIM evaluates how well the generated images preserve structural similarity as compared to the original images. The SSIM quality map is used to calculate the SSIM index for one entire image, this map shows that SSIM is better at predicting the perceived image quality compared to other metrics (Rehman *et al.*, 2012). Once the images are generated using Custom GAN, the quality of the image is evaluated using SSIM. The output of the SSIM for the Custom GAN generated saved images (as discussed in the methodology section, only images after 10^{th} epoch were saved) when compared to the original images are shown below.

Class	**SSIM**
Parkinson	0.60
Healthy	0.63

Table 2: SSIM Scores

Since the Parkinson and Healthy folder have a mix of spiral and wave images, the SSIM score of 0.6 shows that the images have good structural similarity. Also, it shows that the generated images have high level of detail and quality which is important for classifying PD. Therefore, the strong SSIM performance justifies the use of Custom GAN-generated images in this project, as it ensures that the images are sufficiently accurate and reliable for subsequent analysis and classification tasks.

For **Classification, the models are evaluated on the following metrics**,

1. Accuracy: The proportion of correct predictions out of all the predictions made.
2. Precision: The ratio of correct positive predictions out of all positive predictions.
3. Recall (Sensitivity): The percentage of actual positives that are correctly identified.
4. Specificity: The ratio of true negative to the actual negatives.
5. Cohen's Kappa: A measure of how much better the agreement between two raters is compared to chance.
6. F1 Score: The harmonic mean of precision and recall

6.1 Experiment 1: Use of GAN combined dataset (Custom GAN +Original data) on Standalone pretrained CNN models

6.1.1 ResNet50

In this research, when GAN combined data is applied on ResNet50, the confusion matrix is achieved as shown in figure 11. The model achieves an accuracy of 96.6%, and the ability to correctly identify Parkinson’s patients are 93.32%, while the ability to correctly identify healthy individuals is 99.85%. The overall accuracy of positive predictions is 96.8%. The Kappa's score of 93.27% measures the agreement between the model's predictions and actual labels. The overall balance between precision and recall is 96.63%.

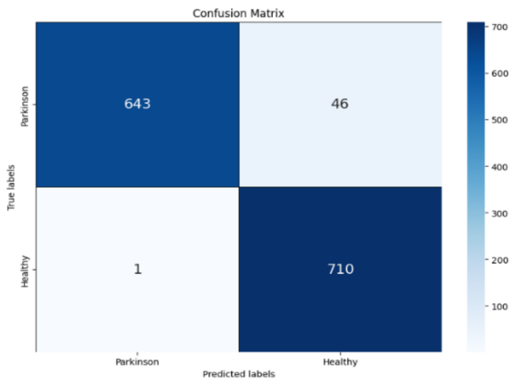

Figure 11: ResNet50 - Confusion Matrix

Metrics	Values
Accuracy	96.64
Sensitivity	93.32
Specificity	99.85
Overall Precision	96.8
Overall F1 Score	96.63
Cohen’s Kappa	93.27

Table 3: ResNet50 metrics

Overall, the model performs well by demonstrating high accuracy, precision and a good balance of precision and recall

6.1.2 Inception V3

In this research, when GAN combined data is applied on InceptionV3, the confusion matrix is achieved shown in figure 12. The model showcases good performance with an accuracy of 98.71%. A high sensitivity of 99.85% shows that the model can correctly identify patients with PD, while high specificity of 97.60% tells that the model is effective at identifying healthy individuals. Overall Precision shows that when the model predicts PD it is correct 98% of the time, Kappa's score of 97.42% shows a high level of agreement between model prediction and actual labels. F1 score of 98.71% shows well balanced performance across precision and recall.

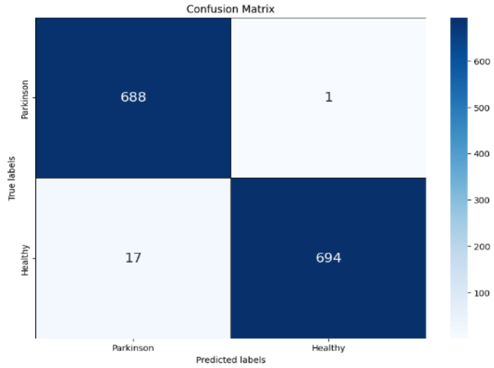

Figure 12: InceptionV3 - Confusion Matrix

Metrics	Values
Accuracy	98.71
Sensitivity	99.85
Specificity	97.60
Overall Precision	98.72
Overall F1 Score	98.71
Cohen’s Kappa	97.42

Table 4: InceptionV3 metrics

Overall, the InceptionV3 model is highly reliable for diagnosing Parkinson's disease showing strong performance across all the metrics

6.2 Experiment 2: Use of GAN combined data (Custom GAN +Original data) on hybrid models

6.2.1 ResNet50 with KNN classifier

The GAN combined data is used on the hybrid model (ResNet50 with KNN classifier), where KNN is tuned to find the best hyperparameters. The best hyperparameters found were ‘Manhattan’, ‘3’ and ‘distance’ for distance metric, neighbors and weights respectively.
When the features extracted using ResNet50 were passed to KNN classifier, the following confusion matrix is obtained as shown in figure 13.

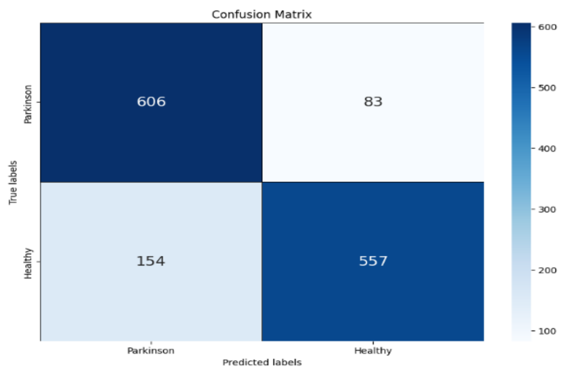

Figure 13: ResNet50 +KNN - Confusion Matrix

Metrics	Values
Accuracy	83.07
Sensitivity	87.95
Specificity	78.34
Overall Precision	83.38
Overall F1 Score	83.05
Cohen’s Kappa	66.18

Table 5: ResNet50 +KNN metrics

The model gives an accuracy of 83.07%, Sensitivity of 87.95% shows that model performs moderately to correctly identify Parkinson, while Specificity of 78.34% shows the model's ability to correctly identify Healthy cases. Overall Precision is 83%, showing the proportion of true positive predictions among all of the positive predictions. F1 score of 83.05% reflects balanced performance across both classes. Kappa's score of 66.18% shows below moderate agreement between model's predictions and the true labels. Overall, the model is not very effective in diagnosing PD.

6.2.2 InceptionV3 with KNN classifier

The GAN combined data is used on the hybrid model (InceptionV3 with KNN classifier), where KNN is tuned to find the best hyperparameters. The best hyperparameters found were 'Manhattan', '3' and 'distance' for distance metric, neighbors and weights respectively.
When the features extracted using InceptionV3 were passed to KNN classifier, the following confusion matrix is obtained as shown in figure 14.

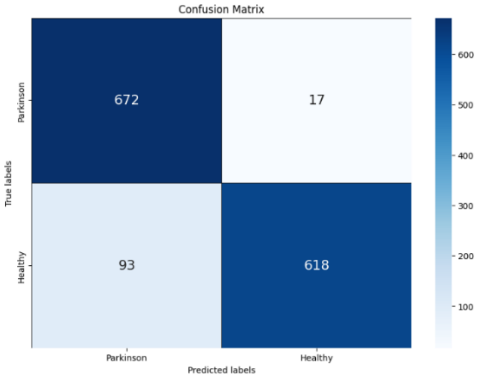

Fig 14: InceptionV3 +KNN - Confusion Matrix

Metrics	Values
Accuracy	92.14
Sensitivity	97.53
Specificity	86.91
Overall Precision	92.5
Overall F1 Score	92.13
Cohen's Kappa	84.3

Table 6: InceptionV3 +KNN metrics

The model gets an accuracy of 92.14%, showing correct predictions. Sensitivity of 97.53% shows that the model is effective in correctly identifying PD, while specificity of 86.91% shows that the model performs well at identifying healthy individuals. Precision of 92% shows a high proportion of true positive predictions among all positive predictions. F1 score of 92.13% shows a balanced performance. Kappa score of 84.3% shows moderate agreement between the model's predictions and the true labels. This model performs well at identifying PD

6.3 Comparison of developed models

The aim of this experiment is to validate how effective is Custom GAN generated images on ResNet50, InceptionV3 and ResNet50, InceptionV3 with KNN classifier. These four models helped in the classification of PD. Table 7 shows the metrics achieved by all 4 models.

	Resnet50+KNN	InceptionV3+KNN	InceptionV3	Resnet50
Accuracy	0.83	0.92	**0.98**	0.96
Sensitivity	0.88	0.97	**0.99**	0.93
Specificity	0.78	0.87	0.97	**0.99**
Precision	0.83	0.92	**0.98**	0.96
F1 Score	0.83	0.92	**0.98**	0.96
Cohen's Kappa	0.66	0.84	**0.97**	0.93

Table 7: Comparison of all model metrics

The above table 7 summarises the model's performance on GAN combined dataset. From the above table it can be inferred that InceptionV3 model performs well across all the metrics except specificity although there's just a difference of 0.02 when compared with ResNet50. This even shows that Custom GAN generated images proved useful in detecting PD with good metric scores for InceptionV3. It can even be seen that both the standalone models performed well when compared to hybrid models. ResNet50 with KNN performed poorly amongst all the models, although ResNet50 standalone gave good results but the addition of KNN classifier did not improve the performance. Whereas InceptionV3 performed well standalone and even with KNN classifier. Overall, it can be concluded that Custom GAN generated images proved beneficial for PD detection.

6.4 Comparison of developed model with existing models

Author Name	Original Dataset Count	Augmented dataset Count	Model	Accuracy
Vaidya *et al.* (2024)	204	3264	EfficientNetB2	96.4%
Kumar and Bansal (2023)	204	3264	MobileNetV2	97.70%
Shaban (2020)	204	800	VGG 19	88%
Morales-Castro *et al* (2022)	204	-	ResNet50 + SVM	90%
This Study	3264	7000	InceptionV3	**98.71%**

Table 8: Comparison with previous studies

In all the previous studies mentioned in Table 8, the original dataset was taken from Kaggle[3] and the augmentation method used were traditional augmentation methods like flipping, rotating, etc. But in this study by incorporating the Custom GAN images, the size of the images almost doubled to the size when compared to the previous studies. The model used in this study is InceptionV3, which outperformed all the previous studies by achieving an accuracy of 98.71%. Also, the current study even outperforms the base paper by Kumar and Bansal (2023) by generating more data and producing better results. Hence in this study, it can be concluded that by increasing the dataset size through Custom GAN can improve model performance and can enhance the accuracy of Parkinson's disease detection.

7 Conclusion and Future Work

The aim of this research was to achieve early detection of PD which could benefit the patients. This research proposes a machine learning framework that combines Custom GAN data with Original data to perform classification on pre-trained CNN models. This study successfully addressed the research question, "To what extent can Custom GAN-generated images combined with original images using standalone CNN models (ResNet50, InceptionV3) and hybrid models (ResNet50, InceptionV3 with KNN classifier) improve early detection of

[3] https://www.kaggle.com/datasets/kmader/parkinsons-drawings/data

Parkinson's Disease?”. This study shows that incorporating the images generated by Custom GAN proved successful in the detection of PD, and Inception V3 outperformed by achieving good performance metrics. This study was successful in achieving all the research objective by assessing the quality of Custom GAN images using SSIM, checking the effectiveness of Custom GAN images by achieving good performance metrics, comparing previous works with the current study to find out that the current study outperformed and comparison of standalone models with hybrid models where standalone models performed well, but one limitation can be highlighted, which is that, this research is computationally intensive which can be difficult in practical settings, unless there are enough computational resources. Also, KNN proved as a good classifier to InceptionV3 but not to ResNet50. When this study was compared to other studies it even showed that increasing the dataset size can improve the reliability of the model in accurately identifying PD.
This research can potentially enhance the early detection of Parkinson Disease. Given the high computational requirements needed for this study, future research can focus on optimizing Custom GAN or trying a different type of GAN. This research was conducted on a large data but the choice of classifier for the hybrid model should be reconsidered which can lead to better performance. Also, by integrating different type of handwritten images like, cubes, triangles, etc. can enrich the dataset. Developing user friendly interfaces can help in real world implementation and can be easier for patients to use it.

References

An, D., Guo, Y., Zhang, M., Qi, X., Lei, N. and Gu, X. (2020) ‘AE-OT-GAN: Training GANs from data specific latent distribution’, *Computer Vision – ECCV 2020. Lecture Notes in Computer Science*, vol 12371, pp. 548-564. Springer. doi: 10.1007/978-3-030-58574-7_33.

Anila, M. and Gera, P. (2022) ‘Parkinson’s disease identification using deep neural network with RESNET50’, *International Journal of Advanced Computer Science and Applications*, 13(11), pp. 488-493. doi: 10.14569/IJACSA.2022.0131156.

Cantürk, İ. and Günay, O. (2024) ‘Investigation of scalograms with a deep feature fusion approach for detection of Parkinson’s disease’, *Cognitive Computation*, 16, pp. 1198–1209. doi: 10.1007/s12559-024-10254-8.

Chakraborty, S., Aich, S., Sim, J.-S., Han, E., Park, J. and Kim, H.-C., (2020) ‘Parkinson’s Disease Detection from Spiral and Wave Drawings using Convolutional Neural Networks: A Multistage Classifier approach’, *International Conference on Advanced Communications Technology (ICACT).* Phoenix Park, Korea (South) pp.298-303. doi: 10.23919/ICACT48636.2020.9061497.

Cheng, P.-C. and Chiang, H.-H. K. (2023) ‘Diagnosis of salivary gland tumors using transfer learning with fine-tuning and gradual unfreezing’, *Diagnostics*, 13(21), 3333. doi: 10.3390/diagnostics13213333

Das, A., Das, H. S., Choudhury, A., Neog, A. and Mazumdar, S. (2021) ‘Detection of Parkinson’s disease from hand-drawn images using deep transfer learning’, *Intelligent*

Learning for Computer Vision. CIS 2020. Lecture Notes on Data Engineering and Communications Technologies, vol 61, pp. 67-84. doi: 10.1007/978-981-33-4582-9_6

Dorsey, E. R., Sherer, T., Okun, M. S. and Bloem, B. R. (2018) 'The emerging evidence of the Parkinson pandemic', *Journal of Parkinson's Disease*, 8(s1), pp. S3-S8. doi: 10.3233/JPD-181474

Dzotsenidze, E., Valla, E., Nõmm, S., Medijainen, K., Taba, P. and Toomela, A. (2022) 'Generative adversarial networks as a data augmentation tool for CNN-based Parkinson's disease diagnostics' *IFAC-PapersOnLine,* 55(29), pp. 108-113, doi: 10.1016/j.ifacol.2022.10.240

Islam, F. A. and Akter, L. (2020) 'Early identification of Parkinson's disease from hand-drawn images using histogram of oriented gradients and machine learning techniques', in *2020 Emerging Technology in Computing, Communication and Electronics (ETCCE),* Bangladesh, 21-22 December 2020, pp. 1-6. doi: 10.1109/ETCCE51779.2020.9350870.

Hendricks, R.M. and Khasawneh, M.T., (2021) 'An Investigation into the use and meaning of Parkinson's Disease clinical scale scores', *Parkinson's Disease*, p.1765220. doi: 10.1155/2021/1765220

Kumar, B. A. and Bansal, M. (2023) 'A transfer learning approach with MobileNetV2 for Parkinson's disease detection using hand-drawings', in *Proceedings of the 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. Delhi, India, 6-8 July 2023, pp.1-8. doi: 10.1109/ICCCNT56998.2023.10307641

Kurmi, A., Biswas, S., Sen, S., Sinitca, A., Kaplun, D. and Sarkar, R. (2022) 'An ensemble of CNN models for Parkinson's disease detection using DaTscan images', *Diagnostics,* 12(5), 1173, doi: 10.3390/diagnostics12051173

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z. and Shi, W. (2017) 'Photo-realistic single image super-resolution using a generative adversarial network', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 4681-4690. doi: 10.48550/arXiv.1609.04802

Liang, J. and Liu, R. (2015) 'Stacked denoising autoencoder and dropout together to prevent overfitting in deep neural network', in *2015 8th International Congress on Image and Signal Processing (CISP)*. Shenyang, China, 14-16 October 2015, pp. 697-701. doi: 10.1109/CISP.2015.7407967

Liu, B. and Chen, J. (2021) 'A super resolution algorithm based on attention mechanism and SRGAN network', *IEEE Access*, 9, pp. 139138-139142. doi: 10.1109/ACCESS.2021.3100069

Meena, G., Mohbey, K. K. and Kumar, S. (2023) 'Sentiment analysis on images using convolutional neural networks based Inception-V3 transfer learning approach', *International Journal of Information Management Data Insights*, 3(1), 100174. doi: 10.1016/j.jjimei.2023.100174

Morales-Castro, J. C., Ruiz-Pinales, J., Lozano-García, J. M. and Guzmán-Cabrera, R. (2022) 'Use of image processing for the detection of Parkinson's disease', *Journal of Physiotherapy and Medical Technology,* 6(16), pp. 27-32. doi: 10.35429/JP.2022.16.6.27.32

Peppes, N., Tsakanikas, P., Daskalakis, E., Alexakis, T., Adamopoulou, E. and Demestichas, K. (2023) ‘FoGGAN: Generating realistic Parkinson’s disease freezing of gait data using GANs’, *Sensors,* 23(19), 8158. doi: 10.3390/s23198158

Quan, J., Xu, L., Xu, R., Tong, T. and Su, J. (2019) ‘DaTscan SPECT image classification for Parkinson’s disease’, *arXiv preprint.* doi: 10.48550/arXiv.1909.04142

Rajinikanth, V., Yassine, S., Bukhari, S. A. (2024) ‘Hand-sketchs based Parkinson's disease screening using lightweight deep-learning with two-fold training and fused optimal features’, *International Journal of Mathematics, Statistics, and Computer Science,* 2(2), pp. 9-18. doi: 10.59543/ijmscs.v2i.7821.

Rehman, A., Rostami, M., Wang, Z., Brunet, D. and Vrscay, E. R. (2012) ‘SSIM-inspired image restoration using sparse representation’, *EURASIP Journal on Advances in Signal Processing,* 2012, 16. doi: 10.1186/1687-6180-2012-16

Shaban, M. (2020) ‘Deep convolutional neural network for Parkinson’s disease based handwriting screening’, in *2020 IEEE 17th International Symposium on Biomedical Imaging Workshops (ISBI Workshops).* Iowa City, IA, USA, 4 April 2020, pp. 1-4, doi: 10.1109/ISBIWorkshops50223.2020.9153407.

Vaidya, O. V., Salvi, O. S., Mahajan, V. and Paraye, M. (2024) ‘An innovative deep-learning approach for identifying Parkinson’s disease through spiral and wave sketch patterns’ in *2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT).* Gharuan, India, 2-3 May 2024, pp. 480-487. doi: 10.1109/InCACCT61598.2024.10551124.

Wu, X., Xu, K., and Hall, P. (2017) ‘A survey of image synthesis and editing with generative adversarial networks’, *Tsinghua Science and Technology,* 22(6), pp. 660-674. doi: 10.23919/TST.2017.8195348.

Xu, Z- J., Wang, R.-F., Wang, J. and Yu, D.-H. (2020) ‘Parkinson’s disease detection based on spectrogram-deep convolutional generative adversarial network sample augmentation’, *IEEE Access,* 8, pp. 206888-206895. doi: 10.1109/ACCESS.2020.3037775

Yang, X., Ye, Q., Cai, G., Wang, Y., and Cai, G. (2022) ‘PD-ResNet for classification of Parkinson’s disease from gait’, *IEEE Journal of Translational Engineering in Health and Medicine,* 10, pp. 1-11. doi: 10.1109/JTEHM.2022.3180933

Zahisham, Z., Lee, C. P. and Lim, K. M. (2020) ‘Food recognition with ResNet-50’, in *IEEE 2nd International Conference on Artificial Intelligence in Engineering and Technology (IICAIET).* Kota Kinabalu, Malaysia, 26-27 September 2020, pp. 1-5. doi: 10.1109/IICAIET49801.2020.9257825.

Zanini, R. A. and Colombini, E. L. (2020) ‘Parkinson’s disease EMG data augmentation and simulation with DCGANs and style transfer’, *Sensors,* 20(9), 2605. doi: 10.3390/s20092605