Classifying Alzheimer's Disease, Parkinson's Disease and Control Cases with Transfer Learning and XAI

I. Introduction

Alzheimer's Disease (AD) is a neurological disorder that causes gradual and irreversible damage to cognitive ability and memory, hindering the patient's ability to perform daily tasks. Their conditions only worsen over time, impacting their quality of life [1]. The disease ultimately results in death [2].

Parkinson's Disease (PD) is a progressive disorder that mainly affects movement and motor control, which is due to the loss of brain cells in the Substantia Nigra—the region in the brain responsible for transmitting movement signals [1]. They also face trouble with essential basic skills like speaking, walking, writing, and more [2].

Additionally, AD patients may lose their sense of self as the disease progresses. As patients lose their ability to carry out daily activities like dressing, or cooking, they may feel frustration and helplessness because they feel reliant on others. Furthermore, as the disease progresses, the loss of memories and skills can result in the loss of identity [1]. Similarly, Parkinson's disease can have negative emotional effects on people. The loss of motor functions results in slow movement and difficulty with balance. This makes everyday tasks such as walking downstairs and making breakfast a very challenging task. This can lead to depression and anxiety. Both AD and PD hinder one's quality of life. Early detection and treatment can help "manage symptoms, slow the disease progression, and improve the patient's quality of life" [1]. The "improvement of the quality of life for the impacted individuals largely depends on the timely and precise diagnosis of these illnesses" [2]. Clinical diagnoses can take 3 to 5 years to classify PD [3]. Similarly, clinical diagnoses typically take 5.5 years to reach a final diagnosis of AD [4]. Such a long period allows for symptoms to worsen over time. Furthermore, clinical diagnoses using blood tests to detect AD have an accuracy of 88% to 92% [5]. Similarly, clinical diagnoses can detect PD with an accuracy of 90.3% [6]. While clinical diagnoses are fairly accurate, they take a very long time to make their diagnosis. AI techniques can help reduce the time and increase the precision of diagnosis of neurological diseases, helping improve impacted patients' quality of life.

It is common for medical datasets to be small due to the time and cost of collecting each MRI scan. Transfer Learning can be used to increase the accuracy of the model even with a small dataset [7]. Pre-trained models are machine learning models that have already been trained on a large dataset and can be used to solve new problems utilized in Transfer Learning [7]. Since pre-trained models are originally trained on a large dataset for a different task, it can reach higher accuracy in a short period. However, using transfer learning alone creates a black-box experience, which is a system where the internal workings and decision-making processes are not transparent to the user. Since the model does not illuminate the reason why it made its diagnosis, healthcare workers do not know when to trust AI models. This is where visual eXplainable Artificial Intelligence (XAI) techniques come in. Visual XAI techniques like Grad-CAM, saliency maps, and more highlight features in the MRI scan, elucidating why the AI model made its decision [8]. If the AI model makes a wrong diagnosis, the programmers can view the regions that it highlights and understand why the model made a mistake, which can help them improve the model for the future. This also builds trust between healthcare workers and AI models; since the XAI techniques like Saliency Maps highlight regions on the MRI scan that led it to its diagnosis, healthcare workers know why the model makes its diagnosis. Knowing why the model made its diagnosis can help healthcare workers make their diagnosis with confidence. If the model makes a diagnosis for the right reason, the healthcare workers can trust the model; if the model makes a diagnosis for the wrong reason, the healthcare workers can safely ignore its diagnosis. The usage of XAI helps healthcare workers know when to trust and not to trust the AI model, whether the model provides the right regions or not. Without XAI, the AI model would provide a diagnosis without providing any explanation for why it made its diagnosis. The usage of XAI technique helps increase interpretability and build trust between healthcare workers and AI. Overall, XAI techniques not only allow for easier improvement of the model but also build trust between healthcare workers and AI technology in the medical field. Integrating XAI techniques with Transfer Learning allows for both high accuracy and high interpretability.

II. Literature Review

To help improve the interpretability of AI models, we must view how others utilized Transfer Learning and/or XAI techniques to classify neurological disorders. Dr. Viswan, a lecturer at the University of Technology and Applied Science (UTAS), and colleagues used Transfer Learning with the pre-trained models: ResNet50, ResNet101, InceptionV3, Inception, ResNetV2, and EfficientNetB0 to classify between AD, PD, and control cases. Dr. Viswan also created heatmaps from Grad-CAM, an XAI technique, which highlights regions in the MRI scan that led the model to its diagnosis. Dr. Viswan then calculated Pearson's correlation coefficient, which is a number between –1 and 1 that determines the correlation between the original MRI scan and the heatmap [1].

Dr. Siddiqua, a professor from Nothern University Bangladesh, and colleagues first performed data cleaning techniques such as sharpening and denoising; it used a separate pre-trained denoising model to clean the data. Cleaning the data removes biases and allows the models to train with higher quality images, it improves the model's accuracy in the test dataset. They then used Transfer Learning with pre-trained models: EfficientNetB0, InceptionV3 model, ResNet50, and Xception to classify between AD, PD, and Control Cases [2].

Mr. Mansouri, a lecturer from the National Engineering School of Sfax (ENIS), and colleagues used the pre-trained models: CNN, VGG16, ResNet50, and AlexNet to classify stages of AD: No impairment, Moderate Impairment, Mild Impairment, and Very Mild Impairment. To address the limited number of MRI Scans of AD, it created a synthetic MRI scan using Generative Adversarial Networks (GANs), specifically WGAN-CP. They then trained the models with real and synthetic MRI scans and a model with only real MRI scans. They then used the XAI technique Grad-CAM to extract gradients from the last Convolutional layer from each image to overlay heatmaps onto the image [9].

Dr. Al-Zharani, a professor at Imam Mohammad Ibn Saud Islamic University (IMSIU), and colleagues also used Transfer Learning with the pre-trained models: VGG16, ResNet50, and DenseNet121 to classify stages of Alzheimer's Disease. It performed data augmentation techniques such as horizontal flipping of images, rotation of image by 5°, and shifting the images to increase the number of images in the training dataset. The best pre-trained model was DenseNet121, which had an accuracy of 97.33% [8].

Assistant Professor of Computer Science and Engineering at Rangamati Science and Technology University, Dr. Mahmud and colleagues utilized Transfer learning with the pre-trained models: VGG16, VGG19, DenseNet169, and DenseNet201 to classify stages of Alzheimer's Disease. They used the explainable AI technique (XAI) including saliency maps and Grad-CAM to increase the interpretability of the model. They performed many image-preprocessing techniques such as resizing them to 224 x 244, normalization, removing noise by median filter, and more. Their model received an accuracy of 96% [10].

Mr. Dentamaro, a researcher at the University of Bari who is interested in Machine Learning and AI in healthcare, and his colleagues classified PD with XAI techniques such as Integrated Gradients and Attention Heatmaps for Vision Transformer (ViT) after using Transfer Learning to train the pre-trained models: ResNet and DenseNet. This paper used 3D MRI scans instead of 2D MRI scans. Their DenseNet model performed the best with an accuracy of 96.6% [11].

Dr. B Rama, assistant professor at the University of Kakatiya with over 50 publications, and colleagues used transfer learning to train the two pre-trained models: CNN and VGG-19. They performed data preprocessing steps on the MRI scans. Each model had convolutional layers for feature extraction, pooling layers (e.g., MaxPooling) to reduce spatial dimensions, and dense and dropout layers for classification and regularization. In the end, the CNN model received an accuracy of 98.04%; while the VGG-19 model received 92.04% [12].

There has been research using transfer learning on classifying PD with Control Cases [12], AD with Control Cases [8] and between AD, PD, and Control Cases [2]. There has also been research integrating XAI methods with Transfer Learning to increase both the accuracy and interpretability of the model. Mr. Dentamaro classified PD with Control Cases with the XAI techniques Integrated Gradients and Attention Heatmaps with the pre-trained models: ResNet and DenseNet [11]. Dr. Mahmud classified AD with Control Cases with the XAI techniques: saliency maps and Grad-CAM with the pre-trained models: VGG16, VGG19, DenseNet169, and DenseNet201 [10]. Mr. Mansouri also classified AD with Control Cases with the XAI technique: Grad-CAM. However, he used Transfer Learning with the pre-trained models: CNN, VGG16, ResNet50, and AlexNet. It also created a WGAN-CP model to create synthetic MRI scans to address the limited dataset size, increasing the accuracy [9]. Lastly, Dr. Viswan used the Grad-CAM technique with the pre-trained models: ResNet50, ResNet101, InceptionV3, InceptionResNetV2, and EfficientNetB0 [1].

However, other XAI methods have not been used when classifying between AD and PD. Only the XAI technique Grad-CAM has been used to classify AD and PD with Transfer Learning. Other studies that utilized XAI techniques and Transfer Learning only classify between one neurological disease (AD or PD) with Control cases—not both. Other studies also don't focus on interpretability by using XAI techniques; they use Transfer Learning alone to classify neurological disorders. This led to the research question: How does the integration of XAI, specifically Saliency Maps, and Transfer Learning improve the interpretability and the performance of classification of AD, PD, and Control Cases?

This paper will focus on using Saliency Maps combined with the use of Transfer Learning with pre-trained models. The research aims to build trust between healthcare workers and AI models. XAI techniques provide insights into the features and regions crucial for its decision [9]. This information from the model can be used by clinicians to make their diagnosis of the patients, instilling confidence in healthcare workers. XAI techniques create a collaborative experience between healthcare workers and AI, where humans and AI technology work together to produce accurate and reliable diagnoses, saving lives. Moreover, XAI techniques also help improve the model's accuracy through the debugging process. If the model makes a wrong diagnosis, the programmer can view the Saliency Map to see why the model made the wrong diagnosis. Because the programmer knows where and why the model made the wrong diagnosis, it becomes easier to fix the problem. Overall, the implementation of XAI techniques not only helps improve the performance of the model but also helps create trust between the AI models and healthcare workers. The

collaboration of AI models and healthcare workers helps create fast and reliable diagnoses, helping patients receive treatment as fast as possible, and improving their quality of life.

III. Methods

Acknowledging that there is an existing gap between classifying AD and PD with XAI techniques and Saliency Maps as it has not been done before; there has been various studies classifying AD and PD alone with XAI techniques and Transfer Learning while there has only been one study that classified AD and PD with XAI techniques, specifically they used the GRAD-CAM technique, with Transfer Learning. Since there is no studies that utilized Saliency Maps to classify AD and PD with Transfer Learning, it is an existing gap. The methodology Design-based research was used since it focuses on designing models that can be used in the real world, and it requires iterative improvements to the model.

A. Choosing the Dataset and data preparation

A good dataset must include MRI images of neurological disorders supported by the type of neurological disorder; but, they don't need to be separated by train, validation, and test since the code can separate them. Kaggle, the world's largest data science company, has a dataset with MRI images of Parkinson's, Alzheimer's, and Control Cases. The Kaggle's built-in IDE was used along with its GPU to run the code for the model. The training dataset has 2561 AD MRI images, 3010 Control Case MRI Images, and 906 PD MRI images. The testing dataset has 639 AD MRI images, 662 Control Case MRI Images, and 61 PD MRI images. All of the images provided by Kaggle had the same format: Portable Network Graphics (png). Some of the training data for each class was moved to the testing dataset so that 20% of the data for each class was in the testing class, while the training class had the other 80%. The training dataset AD has 2560 while the testing dataset AD has 640 MRI images. The training dataset PD has 774 while the testing dataset PD has 193 MRI images. The training dataset CONTROL has 2938 while the testing dataset CONTROL has 734 MRI images.

This methodology was approved by the Institutional Review Board (IRB). In this study, I made an AI model to classify MRI scans. I performed several methods employed to improve the model. For example, one method I used is normalization, which adjusts the pixel's RGB value to a predefine range such as [0, 1] or [-1, +1]. This study used the min-max scaler normalization, which normalized each pixel value to a brightness to between 0 and 1, which is better for machines to make calculations because it is more efficient2. Additionally, the dataset for images was already split into training and testing dataset; but the training dataset was further split into training and validation, with 90% of the training dataset and 10% used for the validation. Additionally, after splitting the data in train, validation, and test data, I performed data augmentation techniques, including rotating it clockwise and counter clockwise up to 20 degrees, shifting the image 20% of its width and height, applies random shearing transformation, zooms in up to 20%, and finally flips it horizontally. The data augmentation techniques were only used on the training dataset since validation and test datasets must reflect real world conditions. Data augmentation techniques help increase the number of images in the training dataset; Thus, it can help increase the accuracy of the model since the model has more images to learn from, addressing the small dataset size.

To address the imbalance of images across the class, for instance, there are significantly more AD and CONTROL MRI images than PD images in both training and testing datasets, class weights were used to inform the model of the percentage of AD, PD, and CONTROL images in the dataset. This allows the model to perform better despite the data imbalance of the dataset.

B. Training the model

Transfer Learning techniques were used to train the model. Specifically, the pre-trained models VGG16, ResNet50, InceptionV3, Xception, and EfficientNet models were used. These models were pre-trained on millions on images from ImageNet, which is why when trained to classify neurological disorders they will have a higher accuracy because the model has already been trained on a different task before. The last layer of the model was changed to have 3 output layers: 1 for AD, 1 for PD, and 1 for the control cases. Additionally, the XAI technique Saliency Maps was overlay the weights of the model onto the image, highlighting the most important regions on the MRI scan that led it to its diagnosis. This technique can be useful to help debug the model and help assist healthcare workers see why the model made its decision. Thus, this helps create a collaborative experience between healthcare workers and AI models. Patients begin to trust AI models since healthcare workers have the final decision of the diagnosis.

C. Architecture of VGG-16

VGG is a pre-trained model with 16 layers and is provided below.

- Input layer: The VGG-16 model takes 224 * 224 image input size. The images were from the dataset before placed in the model [13].
- Convolutional layer. The VGG convolutional leverage a minimum receptive field, which had a 3*3 small size kernel that captured left, right, up and down. It has a convolutional stride fixed of 1 pixel, which means that it shifts right/left/up/down by 1 pixel each time it moves [13].
- Hidden Layer: ReLU is usually used in hidden layer. Visual Geometry Graph was not used since it consumes too much memory [13].
- The output layer was changed to have 3 output layers for Alzheimer's Disease, Parkinson's Disease, and Control Cases.

D. Architecture of EfficientNet

Efficient net is also a pre-trained model used to perform transfer learning. Both pre-trained models VGG and EfficientNet were trained on images from ImageNet, which has millions of images from over 1000 classes [10].

- Input layer: The EfficientNet model also takes 224*224 image size for input. The images were resized for the pre-trained model [14].

- Convolutional Layer: This pre-trained model used Mobile inverted Bottleneck Convolution layer, or MBConv, which is a key component of the EfficientNet family of neural network architectures, which is useful when computational resources are limited [14].
- Hidden layers: Similarly, the ReLu was used in the hidden layer [14]
- The output layer was changed to have 3 output layers for Alzheimer's Disease, Parkinson's Disease, and Control Cases.

E. Architecture of InceptionV3

The InceptionV3 model, which belongs to the Inception family of architectures, "is a widely-used pretrained model particularly designed for image - classification and detecting different object as a task. It incorporates inception modules, which are comprised of parallel convolutional layers that possess varying filter sizes. InceptionV3, which has 48 layers total, also introduces factorized convolutions as a means to reduce the computational cost associated with the model. Moreover, it utilizes batch normalization, auxiliary classifiers, and global average pooling. By employing these architectural elements, the model becomes capable of effectively capturing both local and global features" [2].

E. Architecture of ResNet50

ResNet50 is a modified version of the Residual Network (ResNet) design. ResNet "pioneered the notion of residual learning, which involves utilizing skip connections to circumvent one or more layers, thereby facilitating the training of exceedingly deep networks. ResNet50, in particular, encompasses a total of 50 layers and utilizes residual blocks with shortcut connections. The incorporation of skip connections aids in alleviating the vanishing gradient problem, thereby enabling the model to acquire more effective representations" [2].

E. Architecture of Xception

Xception, also known as Extreme Inception, "represents an expansion of the Inception framework, albeit with a distinct methodology for dealing with convolutional operations. Instead of relying on conventional convolutions, Xception employs depth wise separable convolutions, decompose a standard-convolution into depth wise convolution and pointwise convolution. This segregation enhances the efficiency of the model and diminishes the number of parameters. In terms of design, Xception achieves commendable performance across various computer vision tasks while adopting a comparatively simpler approach" [2].

F. Hyperparameters of the models

The same hyperparameters were used for all pre-trained models; the optimizer used was Adam because it is a popular optimizer used and because of its adaptive learning rates. A callback used to increase performance during training reduced the learning rate on a plateau. This method reduces the learning rate when a certain metric is not improving anymore; the metric chosen was validation loss. If the validation loss doesn't decrease after a couple epochs, then the learning rate decreases, which helps continue to increase in performance. Additionally, the loss function used was categorical cross-entropy, which is a loss function used during multi-class AI models.

G. Hypotheses

a. The performance of the InceptionV3 model will perform worse than ResNet50 since InceptionV3 has 48 layers in the model while ResNet50 has 50 layers in the model.
b. The Parkinson's MRI scan will highlight the motor cortex region of the MRI scan since Parkinson's disease affects motor control.
c. The VGG16 model will perform better than Xception since it uses standard convolution instead of depth wise convolution, which has more parameters and is less efficient. But since it has more parameters, it has more control over the model and thus can receive higher accuracy.

H. Measuring the performance of the model

The following metrics are used to measure the performance of the AI model. After training each pre-trained model and implementing Saliency Maps, these metrics were calculated to see how well the model performed.

Accuracy is an important metric to evaluate the overall performance of the model measuring the quality of positives and negative predictions. Equation (1) shows how to calculate accuracy. It is based on the ratio of the correct observations to the total observations [10].

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

Another important metric is the precision value as shown in (2), which measures the quality of a positive prediction. This measures the likelihood that a positive prediction is correct. This is the ratio of the correctly predicted positive observations to the total predicted positive observations [10].

$$PRECISION = \frac{TP}{TP + FP} \quad (2)$$

Another important metric is the recall value as shown in (3). This measures the proportion of true positives out of the amount of positive diagnosis. This is the ratio of the correctly predicted positive observations to all observations in an actual class [10].

$$RECALL = \frac{TP}{TP + FN} \quad (3)$$

Another important metric is F1-score as shown in (4). This can be calculated by the harmonic mean of precision and recall [10]. This is useful when both precision and recall values are equally important for performance.

$$F1 - score = 2 \cdot \frac{PRECISION \cdot RECALL}{PRECISION + RECALL} \quad (4)$$

TP denotes true positive, TN denotes true negative, FP denotes fake positive, and FN denotes false negative.

IV. Results

Table I
Performance of the pre-trained models

Model Name	Accuracy	Precision	Recall	F1-score
VGG16	71.88%	0.7188	0.7188	0.7188
ResNet50	56.25%	0.5282	0.4391	0.4995
InceptionV3	67.04%	0.6728	0.6654	0.6691
Xception	62.84%	0.6280	0.6250	0.6265
EfficientNetB0	45.68%	0	0	N/A

Fig. 1. Performance of each pre-trained model

The pre-trained models had VGG16, ResNet50, InceptionV3, Xception, EfficientNetB0 an accuracy of 82.41%, 44,90%, 67.04%, 62.84%, and 45.68% respectively. These models had a precision value of 0.8049, 0.5282, 0.6728, 0.6280, and 0, respectively. The proposed model for this study is the VGG16 pretrained model since it had the highest accuracy and precision value. This study demonstrates the potential of Explainable AI (XAI) and Deep Transfer Learning by accurately classifying AD and PD with the VGG16 pre-trained model with an accuracy of 82.41%. These findings build trust between AI-based diagnoses and healthcare workers, making it a valuable tool for early detection. The findings highlight the importance of explainability in medical AI, ensuring that healthcare professionals can understand and rely on model predictions.

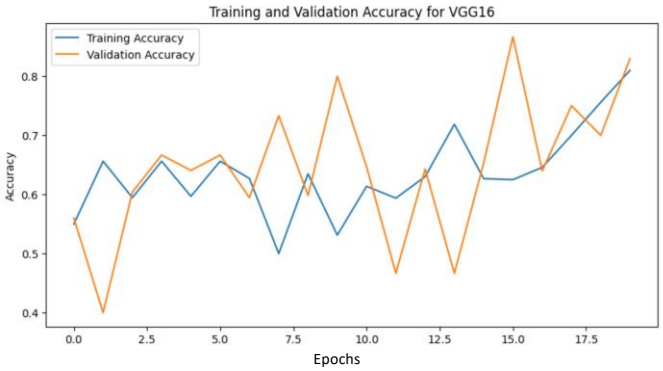

Fig. 2. Training and Validation Accuracy of the VGG16 pre-trained model over 20 epochs.

Both the validation and training accuracy of the model generally increased as the epochs increased. However, the validation accuracy fluctuated a lot between 7 and 16 epochs, each epoch represents the number of times the model goes through the training dataset. At the end of the 20 epochs, the model ended slightly above 80% accuracy in both the training and validation dataset.

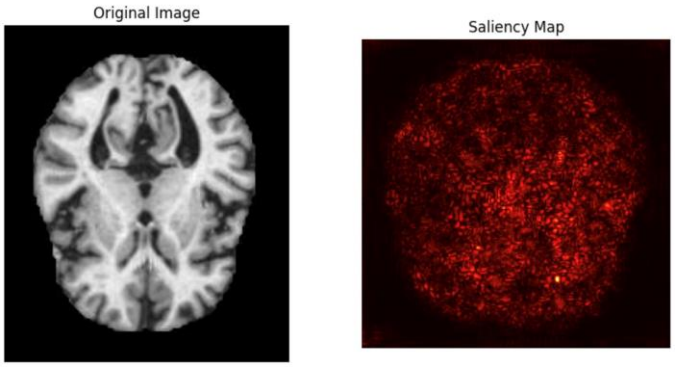

Fig. 3. Original MRI scan of Control Case and the Saliency Map from the Pre-trained Model VGG16

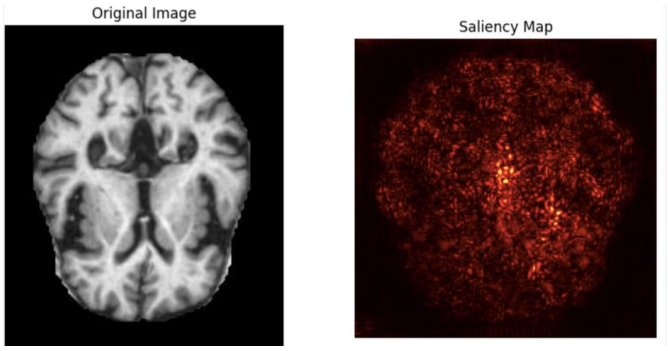

Fig. 4. Original MRI scan of Alzheimer's Disease and the Saliency Map from the Pre-trained Model VGG16

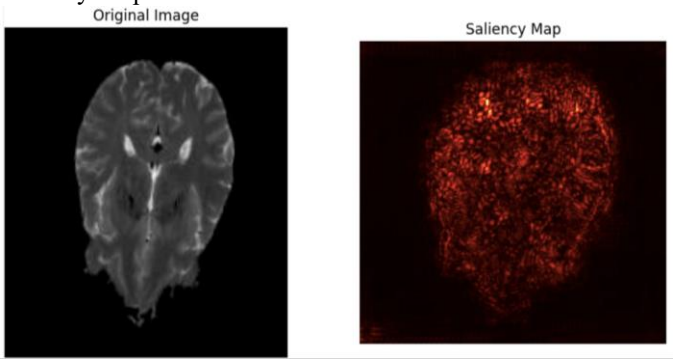

Fig. 5. Original MRI scan of Parkinson's Disease and the Saliency Map from the Pre-trained Model VGG16

The model made a prediction that the MRI scan was a Control Case, which is correct. Looking at the Saliency Map produced by the model (Figure 3), the bottom right region is highlighted in yellow, and this region is also surrounded by brighter red regions. This region is the reason why the VGG16 pre-trained model made its diagnosis.

The model made a prediction that the MRI scan was a Alzheimer's Disease in Figure 4, which is correct. Looking at the Saliency Map produced by the model, the middle region is highlighted in yellow, highlighting that the region the MRI scan made its diagnosis was because of the middle region of the MRI.

The model made a prediction that the MRI scan was a Parkinson's Disease in Figure 5, which is correct. Looking at the Saliency Map produced by the model, the upper region is

highlighted in yellow, illuminating that the region the MRI scan made its diagnosis was because of the upper region of the MRI.

V. Discussion

A. Compared to existing literature

There has been research using transfer learning on classifying PD with Control Cases [12], AD with Control Cases [8] and between AD, PD, and Control Cases [2]. This research lack interpretability compared to this study. There has also been research integrating XAI methods with Transfer Learning to increase both accuracy and interpretability of the model. Mr. Dentamaro classified PD with Control Cases with the XAI techniques Integrated Gradients and Attention Heatmaps with the pre-trained models: ResNet and DenseNet [11]. Dr. Mahmud classified AD with Control Cases with the XAI techniques: saliency maps and Grad-CAM with the pre-trained models: VGG16, VGG19, DenseNet169, and DenseNet201 [10]. Mr. Mansouri also classified AD with Control Cases with the XAI technique: Grad-CAM. However, he used Transfer Learning with the pre-trained models: CNN, VGG16, ResNet50, and AlexNet. It also created a WGAN-CP model to create synthetic MRI scans to address the limited dataset size, increasing the accuracy [9]. While these studies are accurate and have high interpretability, they can only classify one neurological disorder. Lastly, Dr. Viswan used the Grad-CAM technique with the pre-trained models: ResNet50, ResNet101, InceptionV3, InceptionResNetV2, and EfficientNetB0 to classify AD and PD [1]. Its highest performing pre-trained model ResNet50 has accuracy of 98.8%. This study's highest pre-trained model VGG16 had an accuracy of 82.41%. Dr. Viswan used a different XAI technique; he used Grad-CAM while this study utilized Saliency Maps. Grad-CAM utilizes the gradients of the last convolutional layer of a CNN to generate a heatmap while saliency maps can be applied to any type of neural network architecture with gradients from multiple layers. Because this study's proposed model VGG16 performed accurately with an accuracy of 82.41% and a precision of 80.49%, the study helps improve the interpretability of integrating XAI techniques in the classification of AD and PD.

VI. Conclusion

B. Addressing the Hypothesizes

In my first hypothesis, I predicted that ResNet50 would perform better than InceptionV3 since ResNet50 has 50 total layers, which is less than InceptionV3's 48 total layers. ResNet50 and InceptionV3 had an accuracy of 44.90% and 67.04% respectively. I reject my hypothesis because InceptionV3 performed better; InceptionV3 performed better even though it had fewer total layers, which suggests that simplicity in the model architecture caused the model to perform better.

In my second hypothesis, I predicted that the Parkinson Disease's MRI scan would highlight the brain region with the motor cortex since PD is responsible for hindering motor function. I accept this hypothesis since Fig. 5 highlighted the region with the motor cortex.

Finally, in my third hypothesis, I predicted that the pre-trained model VGG16 would perform better than Xception since VGG16 utilizes standard convolution rather than depth wise convolution like Xception. I accept my hypothesis since the VGG16 pre-trained model had an accuracy of 82.41% while the Xception pre-trained model had an accuracy of 62.84%. Even though depth wise convolution is more efficient, it has less parameters, which means it has less control over analyzing the MRI scan, which is why it had less accuracy.

C. Implications

This study improves interpretability of AI in diagnosing neurological disorders with adequate accuracy. The model is able to highlight regions on the MRI scan to make its diagnosis. Since the models without XAI techniques do not illuminate the reason why it made its diagnosis, healthcare workers do not know when to trust AI models. The XAI technique Saliency Maps can highlight features in the MRI scan, elucidating why the AI model made its decision [8]. The high interpretability of the model created by XAI techniques builds trust between the AI model and the physician since physicians know when to trust the model. Physicians can use the AI model to assist them in making fast and accurate decisions, allowing patients to receive immediate treatment. Immediate treatment improves the patient's quality of life by mitigating symptoms of the disease.

D. Limitations

Some limitations face was the coding environment. The model was made in Kaggle, and its GPU was used when compiling the code. Specifically, Kaggle uses a NVIDIA TESLA P100 GPU. Using a higher quality GPU can help improve the accuracy and performance of the model and help decrease the time the model takes to train. Additionally, the model's performance was hindered due to its small dataset size. With either a larger dataset or higher quality GPU, the performance and interpretability of the AI model would be improved.

E. Future Directions

This study made an AI model that classified AD and PD with Transfer Learning and the XAI technique Saliency Maps. In the future, other XAI techniques can be utilized to further increase the interpretability of the model and possibly improve the accuracy of the model. Additionally, instead of using data augmentation techniques, studies can utilize generative adversarial networks (GAN) to generate new MRI scans in each class: AD, PD, Control cases to address the small dataset size, which can increase the performance of the model. Integrating multi-modal data with XAI techniques can also help improve the performance of the model, creating trust with the model. Furthermore, other types of scans instead of MRI scans such as PET, CT scans, or blood samples to diagnosis neurological disorders to improve performance of the model. Additionally, AI can help classify between many more neurological disorders to help improve detection speed and accuracy these diseases.

References

[1] V. Viswan, N. Shaffi, M. Mahmud, K. Subramanian, and F. Hajamohideen, "A comparative study of pretrained deep neural networks for classifying Alzheimer's and Parkinson's disease," 2021 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1334–1339, Dec. 2023, doi: https://doi.org/10.1109/ssci52147.2023.10371843.

[2] A. Siddiqua, A. M. Oni, and M. J. Miah, "A transfer learning approach for neurodegenerative disease classification from brain MRI images: distinguishing Alzheimer's, Parkinson's, and control cases," 2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT), pp. 347–351, May 2024, doi: https://doi.org/10.1109/iceeict62016.2024.10534463.

[3] M. Rossi, S. Perez-Lloret, and M. Merello, "How much time is needed in clinical practice to reach a diagnosis of clinically established Parkinson's disease?," *Parkinsonism & Related Disorders*, vol. 92, pp. 53–58, Nov. 2021, doi: https://doi.org/10.1016/j.parkreldis.2021.10.016.

[4] M. Kvello-Alme, G. Bråthen, L. R. White, and S. B. Sando, "Time to Diagnosis in Young Onset Alzheimer's Disease: A Population-Based Study from Central Norway," *Journal of Alzheimer's Disease*, vol. 82, no. 3, pp. 965–974, Aug. 2021, doi: https://doi.org/10.3233/jad-210090.

[5] S. Reynolds, "Accurate blood test for Alzheimer's disease," *National Institutes of Health (NIH)*, Aug. 13, 2024. https://www.nih.gov/news-events/nih-research-matters/accurate-blood-test-alzheimer-s-disease

[6] S. Virameteekul, T. Revesz, Z. Jaunmuktane, T. T. Warner, and E. De Pablo-Fernández, "Clinical Diagnostic Accuracy of Parkinson's Disease: Where Do We Stand?," *Movement Disorders*, vol. 38, no. 4, Jan. 2023, doi: https://doi.org/10.1002/mds.29317.

[7] Geeksforgeeks, "ML | Introduction to Transfer Learning," *GeeksforGeeks*, Nov. 23, 2019. https://www.geeksforgeeks.org/ml-introduction-to-transfer-learning/

[8] M. Al-Zharani, S. I. Ansarullah, M. S. Al-Eissa, G. M. Dar, R. A. Alqahtani, and S. Alkahtani, "Exploring the efficacy of deep learning techniques in detecting and diagnosing Alzheimer's Disease: a comparative study," *Journal of Disability Research*, vol. 3, no. 6, 2024, doi: https://doi.org/10.57197/jdr-2024-0064.

[9] D. Mansouri, A. Echtioui, R. Khemakhem, and A. B. Hamida, "Explainable AI framework for Alzheimer's diagnosis using convolutional neural networks," 2024 IEEE 7th International Conference on Advanced Technologies, Signal and Image Processing (ATSIP), pp. 93–98, Jul. 2024, doi: . https://doi.org/10.1109/atsip62566.2024.10639037.

[10] T. Mahmud, K. Barua, S. Umme Habiba, N. Sharmen, M. Shahadat Hossain, and K. Andersson, "An explainable AI paradigm for Alzheimer's diagnosis using deep transfer learning," *Diagnostics*, vol. 14, no. 3, pp. 345–345, Feb. 2024, doi: https://doi.org/10.3390/diagnostics14030345.

[11] V. Dentamaro, D. Impedovo, L. Musti, G. Pirlo, and P. Taurisano, "Enhancing early Parkinson's disease detection through multimodal deep learning and explainable AI: insights from the PPMI database," *Scientific Reports*, vol. 14, no. 1, Sep. 2024, doi: https://doi.org/10.1038/s41598-024-70165-4.

[12] B. Rama, P. Praveen, and M. A. Shaik, "Machine learning model to detect Parkinson's Disease using MRI data," 2023 International Conference on Sustainable Communication Networks and Application (ICSCNA), pp. 1516–1521, Nov. 2023, doi: https://doi.org/10.1109/icscna58489.2023.10370527.

[13] K. Muthamil Sudar, P. Nagaraj, S. Nithisaa, R. Aishwarya, M. Aakash, and S. Ishwarya Lakshmi, "Alzheimer's Disease analysis using Explainable Artificial Intelligence (XAI)," 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Apr. 2022, doi: https://doi.org/10.1109/icscds53736.2022.9760858.

[14] S. Knox, A. P. R, and S. K. G, "Detecting Alzheimer's Disease using multi-modal data: an approach combining Transfer Learning and Ensemble Learning," 2023 International Conference on Control, Communication and Computing (ICCC), pp. 1–6, May 2023, doi: https://doi.org/10.1109/iccc57789.2023.10165454.