Using 2D MRI slices to classify Schizophrenia

Youzhi Wang
MSCS student
Columbia University
New York, USA
yw4196@columbia.edu

Roy Michael Madpis
MSDS student
Columbia University
New York, USA
rmm2286@columbia.edu

Yilin Ye
BSOR student
Columbia University
New York, USA
yy3152@columbia.edu

***Abstract*—This paper presents our endeavor to classify schizophrenia using 2D MRI slices instead of entire 3D MRI images. Our methodology involves utilizing a VGG model to generate GradCAMs, identifying the most significant slices in terms of classification contribution, and subsequently training a SWIN Transformer model using these slices. Despite our efforts, constrained by time and resources, we were unable to develop a model capable of successfully classifying schizophrenia based solely on 2D MRI slices. This paper documents our exploration and highlights the challenges encountered in this pursuit. The code for our attempts are on this: GitHub Repository**

***Index Terms*—MRI, deeplearning, SWIN transformer, VGG**

I. Introduction

Schizophrenia (SCZ) is a complex and chronic mental disorder characterized by disturbances in thoughts, perceptions, emotions, and behavior. It typically emerges in late adolescence or early adulthood, affecting approximately 1% of the global population. Symptoms of schizophrenia often manifest in various forms, including hallucinations, delusions, disorganized thinking, and diminished emotional expression. The exact cause of schizophrenia remains elusive, although it is widely believed to involve a combination of genetic, environmental, and neurobiological factors. Neuroimaging studies have revealed structural and functional abnormalities in the brains of individuals with schizophrenia, implicating disruptions in neurotransmitter systems and neural circuitry. Despite advancements in understanding the disorder, diagnosing and treating schizophrenia remain significant challenges, underscoring the importance of ongoing research efforts aimed at elucidating its underlying mechanisms and developing more effective interventions. [11] [12] [13]

II. Background and Related Work

A. Neuroimaging in Schizophrenia Diagnosis

Neuroimaging techniques, particularly magnetic resonance imaging (MRI), have been extensively used to study the structural and functional brain abnormalities associated with SCZ. These studies have revealed alterations in brain volume, connectivity, and activity patterns in patients with SCZ compared to healthy controls [3]. However, the translation of these findings into clinical practice has been limited, partly due to the variability in imaging protocols and the complexity of the data.

B. Deep Learning in Neuroimaging Analysis

In the domain of SCZ diagnosis, deep learning approaches have been employed to analyze magnetic resonance imaging (MRI) and functional MRI (fMRI) data to uncover subtle, yet discriminative features associated with the disorder. For instance, CNNs have been applied to structural MRI data to differentiate between patients with SCZ and healthy controls, achieving notable success in terms of classification accuracy [5] [8]. Moreover, deep learning models have been utilized to explore functional connectivity patterns in fMRI data, providing insights into the aberrant brain networks characteristic of SCZ [7].

C. Swin Transformers for Image Classification

Swin Transformers have shown great promise in the domain of medical image classification, particularly in the analysis of complex medical images such as those found in radiology, pathology, and neuroimaging. Their hierarchical structure with shifted windows enables efficient processing of images at different resolutions, making them adept at capturing fine-grained details in medical scans. This capability is especially crucial in medical image classification tasks where subtle features may be indicative of various pathologies or conditions. Furthermore, Swin Transformers exhibit strong generalization performance, thanks to their self-attention mechanism, which allows them to capture long-range dependencies within the images. This is particularly beneficial in medical imaging, where abnormalities may manifest across large spatial regions [1].

D. Current Challenges and Opportunities

Despite the advancements in deep learning and neuroimaging, several challenges remain in the development of automated SCZ diagnostic tools. One major challenge is the generalization of models across different datasets and imaging sites, as models trained on one dataset may not perform equally well on another. Additionally, interpreting the decisions of deep learning models, known as explainability, is crucial for clinical applications but remains a difficult task. This project's tri-model architecture aims to leverage the strengths of CNNs and Swin Transformers to address these gaps, potentially leading to breakthroughs in the objective diagnosis of schizophrenia.

III. Our Initial Modeling approach

Our proposed methodology involves a multi-step process designed to leverage the strengths of both VGG and SWIN Transformer models for improved performance in schizophrenia detection.

Firstly, we employ a modified VGG model, incorporating squeeze and excitation blocks [10], to process the 3D MRI images. Subsequently, we utilize GradCAMs from the VGG model to identify a subset of 2D slices exhibiting the highest gradient activations. These selected slices are then used to train a pre-trained SWIN Transformer model, which has been originally trained on the ImageNet-1K dataset. [1]

The rationale behind this approach is twofold. Firstly, SWIN Transformers, being trained on extensive natural image datasets, are anticipated to exhibit superior generalizability when applied to medical images like MRI scans. Secondly, by supplying the SWIN Transformer with the most significant MRI image slices, we aim to focus the model's attention on the most pertinent regions, facilitating more effective learning.

This methodology represents a deliberate integration of established techniques from computer vision with state-of-the-art transformer architectures, with the aim of enhancing both the robustness and interpretability of our model for schizophrenia detection.

IV. VGG Model

To replicate the original VGG implementation for extracting GradCAM slices from 3D structural MRI, we strictly adhered to the methodologies outlined in the original study to ensure the authenticity and reliability of our results.

Starting with data transformation, we implemented a series of augmentation techniques exactly as specified in the study. We introduced random blur to simulate slight focus issues commonly seen in MRI scans, applied at a probability of 0.1. To replicate electronic noise inherent in MRI data acquisition, we added random noise with a probability of 0.6. Recognizing the potential for small patient movements during the scan, we incorporated random affine transformations, including translations, rotations, and scaling, applied with a probability of 0.2. Additionally, to mimic the anatomical variations and distortions in brain tissues, random elastic deformations were used, also at a probability of 0.2. To address intensity inhomogeneities in the scans, we added a random bias field effect with a probability of 0.1, and to account for potential patient movements during scanning, we introduced random motion with a probability of 0.05.

The core of our model, the VGG architecture, is designed for deep feature extraction from 3D MRI data. The VGG11 model we adapted consists of multiple convolutional layers, each followed by batch normalization and ReLU activation functions to ensure non-linear processing throughout the network. To enhance the model's ability to focus on relevant features within voluminous medical imaging data, we incorporated SE blocks at strategic points within the network. These blocks adaptively recalibrate channel-wise feature responses by explicitly modeling interdependencies between channels, thus boosting the representational power of the network without significant computational overhead.

Each convolutional layer in our VGG11 model is defined with a kernel size of 3x3x3, maintaining padding to preserve spatial dimensions across layers. Following each set of convolutional operations, max pooling is applied to reduce dimensionality and to abstract higher-level features, which is crucial for managing the model's complexity and computational demands. Batch normalization follows each convolutional operation, helping to accelerate convergence and stabilize training by normalizing the input layers by re-centering and re-scaling.

The squeeze-and-excitation block, introduced after the batch normalization step and before the activation, computes the global spatial information of each channel, captures channel-wise dependencies, and reweights the channel feature responses to enhance informative features while suppressing less useful ones. This process involves a squeezing operation via global average pooling that summarizes each channel, which is then followed by an excitation operation involving two fully-connected layers—a dimensionality-reduction layer and a dimensionality-increasing layer—both of which culminate in a sigmoid activation that scales the feature maps accordingly.

The classifier part of our VGG model transitions from feature extraction to making predictions. It is composed of a sequence of fully-connected layers that decode the deep features extracted by the convolutional base into final output predictions. The dropout layers interspersed between these fully-connected layers help reduce overfitting by randomly disabling a fraction of the neurons during training, thus ensuring that the network generalizes well to new, unseen data.

We set the learning rate at 1e-4, which strikes a balance between efficient learning and stability, avoiding the common pitfalls of rapid convergence or overshooting the target. The model parameters were optimized using the Adam optimizer, known for its effectiveness in managing sparse gradients in noisy datasets like those found in medical imaging. The choice of cross-entropy as the loss function aligns with its prevalent use in binary classification tasks, particularly effective in managing class imbalance, a typical issue in medical diagnostic settings.

Finally, we got an AUC value of 0.81 for the VGG model. we believe we could get better results by exploring other configurations and different training strategies.

V. SWIN Transformer

We commenced our exploration by loading a pre-trained SWIN Transformer model, initially trained on the ImageNet-1K dataset [1], and subsequently fine-tuning it for our specific task. We chose the SWIN transformer pre-trained with ImageNet-1k due to the unavailability of such models pre-trained on brain MRI data.

A variety of configurations were experimented with during the fine-tuning process:

- Training with using the most significant slices from each orientation

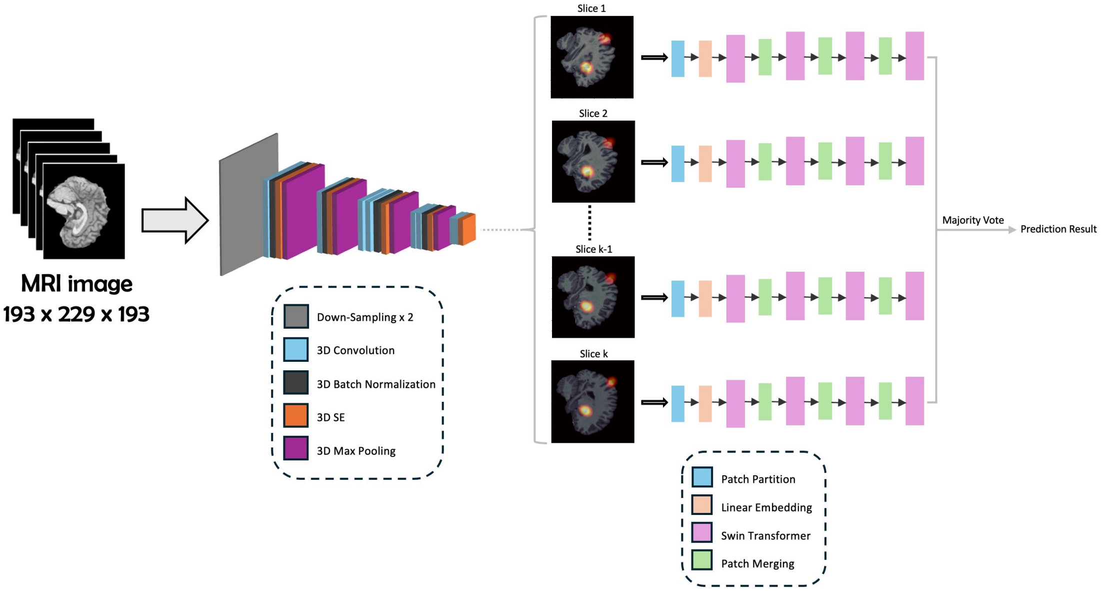


Fig. 1: Proposed model. In this illustration, we can see that the 3D MRI image first goes through a VGG model. Then the most significant slices are extracted using gradCAMs. These 2D slices are then fed into a SWIN transformer model. The final output of the whole model is then decided by a majority vote on the outputs of the SWIN transformer.

- Inputting 3 most significant slices from each direction into 9 channels for the transformer
- Training on the middle 15 slices from each orientation
- Training on the middle 20 slices from each orientation
- Training on the middle 30 slices from each orientation
- Freezing all the transformer layers except the last linear layer
- Unfreezing all the transformer layers
- Tried learning rates of 1e-4, 1e-5, 1e-6
- Tried batch sizes of 1 image to 8 images
- Switching to no pretrained weight
- Using the Swin B transformer model, initially used Swin T model

Despite exhaustive experimentation, none of these configurations yielded a SWIN Transformer model capable of outperforming random guessing in schizophrenia classification.

A significant constraint we encountered was the limited availability of GPU resources and time. Each configuration iteration required several hours to assess, and some attempts were hindered by GPU constraints. However, we remain optimistic that additional configurations may exist that could lead to the development of a viable model. This comprehensive exploration underscores the complexity of fine-tuning transformer models for medical image analysis and highlights the need for further research in this domain.

VI. Discussion

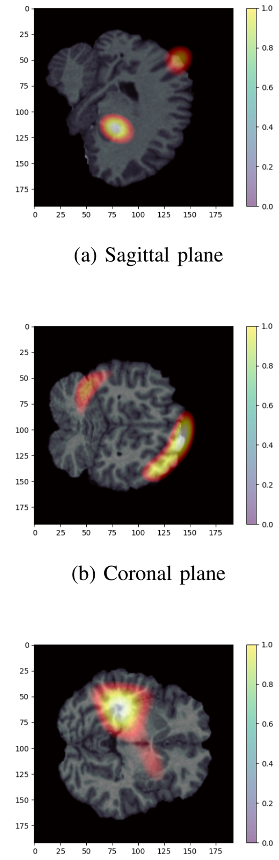
A. *Analysis of Results*

The analysis of the VGG model's performance with an AUC of 0.81 indicates a reasonable level of predictive ability but falls short of being exemplary. This outcome is crucial in the context of using Grad-CAM to identify significant slices for further analysis with the SWIN Transformer model. Although Grad-CAM is innovative in its approach to highlight areas within MRI scans that contribute most to the model's output, its application in this scenario raises several concerns about the model's overall efficacy and the reliability of diagnostic predictions.

Firstly, the method's reliance on Grad-CAM for selecting significant slices introduces the risk of missing vital contextual information. MRI scans encapsulate complex, multidimensional structures, and focusing solely on 2D slices highlighted by Grad-CAM might result in a dataset that does not fully represent all the nuances of schizophrenia. This approach potentially overlooks crucial 3D spatial relationships within the brain. These relationships are often key to understanding comprehensive pathological manifestations of schizophrenia, suggesting that a purely 2D focus could impair the model's ability to grasp the full spectrum of diagnostic features.

Moreover, the choice of slices based on Grad-CAM's activation maps may lead to another layer of complication. If these maps do not accurately reflect the most diagnostically relevant regions due to inherent limitations in the VGG model's understanding of MRI data or the method's own biases, the training data for the SWIN Transformer could lack critical features necessary for robust schizophrenia diagnosis. This misalignment could hinder the model's generalization capability, making it less effective at diagnosing new, unseen cases.

The use of the SWIN Transformer, pretrained on ImageNet, further complicates this issue. The features learned from non-medical images, while diverse, might not translate well to the domain-specific requirements of medical imaging, particularly for complex conditions like schizophrenia. The limited fine-



(a) Sagittal plane

(b) Coronal plane

(c) Axial plane

Fig. 2: gradCAMs from VGG model. These gradCAMs were set to a threshold of 0.9, so only gradCAMs with values above 0.9 are shown to illusrate areas that are most important for the VGG model to give its prediction.

tuning on slices selected via Grad-CAM, which might already be suboptimal, can exacerbate this issue, leading to a model that is not fully equipped to handle the intricacies of MRI-based diagnosis.

This analysis suggests a need for a more holistic approach that considers the 3D structural complexity of the brain and incorporates broader context and multiple perspectives within the MRI data. Enhancing the dataset with more comprehensive and accurately targeted slices, potentially through improved or additional methods beyond Grad-CAM, could be pivotal. Additionally, rethinking the reliance on pre-trained models like the SWIN Transformer, or significantly expanding the fine-tuning process to better accommodate medical imaging contexts, might improve diagnostic accuracy and reliability.

B. Future Works

In future work, we plan to extend beyond the use of SWIN Transformers by exploring a range of advanced neural network architectures that have been pretrained specifically on brain MRIs. This exploration will include Vision Transformers (ViTs) and Medical Transformers (MedT), which may offer enhanced capabilities for capturing the complex patterns inherent in neurological imaging. Additionally, considering the unique demands of medical imaging, we aim to establish a custom pretraining pipeline. This pipeline will utilize a diverse dataset of brain MRIs from various conditions, thus forming a robust foundational model that more accurately reflects the diagnostic features relevant to neurological disorders.

To leverage the strengths of both local feature extraction and global contextual analysis, we propose the development of hybrid End-to-End architectures. These systems will combine Convolutional Neural Networks (CNNs) and transformers in a sequential framework—using CNNs to meticulously extract detailed local features from individual MRI slices, followed by the application of transformers to synthesize and interpret broader contextual relationships across slices. This dual approach promises to enhance the model's ability to interpret complex medical images by providing a more nuanced understanding of both local and systemic anomalies.

Collaboration with domain experts such as neurologists, radiologists, and psychiatrists will be crucial to align the model development with clinical needs and realities. By integrating their expert insights into the model training process, we can ensure that the developed systems focus on clinically relevant features and outcomes, thereby improving the practical utility and reliability of the diagnostic models in real-world medical settings. Through these interdisciplinary partnerships, we aim to refine our models to better meet the diagnostic challenges and enhance patient care in the field of neurology.

VII. Conclusion

Our study contributes to the ongoing efforts in understanding schizophrenia and improving the classification of MRI images associated with the disorder. Despite our initial hypothesis proving unsuccessful, our exploration has laid a foundation for future research endeavors and provides valuable insights and avenues for future investigation. Specifically, our utilization of activation maps from the VGG model presents a promising avenue for further exploration. Future studies may benefit from incorporating more advanced techniques and exploring alternative models for generating activation maps.

In essence, while our current findings may not have met our expectations, they represent a stepping stone towards advancing the field of schizophrenia detection and classification. By building upon our work and exploring innovative methodologies, researchers may uncover novel insights and develop more effective diagnostic tools for this complex disorder.

References

[1] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, “Swin Transformer: Hierarchical Vision Transformer using Shifted Windows,” in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2021, pp. 9992–10002.

[2] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, “A survey on deep learning in medical image analysis,” *Medical Image Analysis*, vol. 42, pp. 60–88, December 2017.

[3] K. H. Karlsgodt, D. Sun, and T. D. Cannon, “Structural and Functional Brain Abnormalities in Schizophrenia,” *Current Directions in Psychological Science*, vol. 19, no. 4, pp. 226–231, August 2010.

[4] Fornito, “Network scaling effects in graph analytic studies of human resting-state fMRI data,” *Frontiers in Systems Neuroscience*, 2010.

[5] M. R. Arbabshirani, S. Plis, J. Sui, and V. D. Calhoun, “Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls,” *NeuroImage*, vol. 145, pp. 137–165, January 2017.

[6] J. Zhang, V. M. Rao, Y. Tian, Y. Yang, N. Acosta, Z. Wan, P.-Y. Lee, C. Zhang, L. S. Kegeles, S. A. Small, and J. Guo, “Detecting schizophrenia with 3D structural brain MRI using deep learning,” *Scientific Reports*, vol. 13, no. 1, pp. 14433, September 2023.

[7] J. Kim, V. D. Calhoun, E. Shim, and J. H. Lee, “Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia,” *NeuroImage*, vol. 124, pp. 127–146, January 2016.

[8] S. M. Plis, D. R. Hjelm, R. Salakhutdinov, E. A. Allen, H. J. Bockholt, J. D. Long, H. J. Johnson, J. S. Paulsen, J. A. Turner, V. D. Calhoun, and J.-B. Poline, “Deep learning for neuroimaging: a validation study,” 2014, www.frontiersin.org.

[9] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, “Grad-CAM: Why did you say that? Visual Explanations from Deep Networks via Gradient-based Localization,” *CoRR*, vol. abs/1610.02391, 2016, http://arxiv.org/abs/1610.02391.

[10] J. Zhang, V. M. Rao, Y. Tian, Y. Yang, N. Acosta, Z. Wan, P.-Y. Lee, C. Zhang, L. S. Kegeles, S. A. Small, and J. Guo, “Detecting schizophrenia with 3D structural brain MRI using deep learning,” *Scientific Reports*, vol. 13, no. 1, pp. 14433, 2023, https://doi.org/10.1038/s41598-023-41359-z.

[11] R. A. McCutcheon, T. Reis Marques, and O. D. Howes, ”Schizophrenia—An Overview,” JAMA Psychiatry, vol. 77, no. 2, pp. 201-210, 2020, https://doi.org/10.1001/jamapsychiatry.2019.3360.

[12] National Institute of Mental Health, ”Schizophrenia,” [Online]. Available: https://www.nimh.nih.gov/health/topics/schizophrenia#:~:text=What%20is%20schizophrenia%3F,for%20their%20family%20and%20friends.. [Accessed: May 8, 2024].

[13] Janssen Pharmaceuticals, Inc., ”Schizophrenia Basics,” [Online]. Available: https://www.janssenschizophreniainjections.com/understanding-schizophrenia/schizophrenia-basics/. [Accessed: May 8, 2024].

Appendix

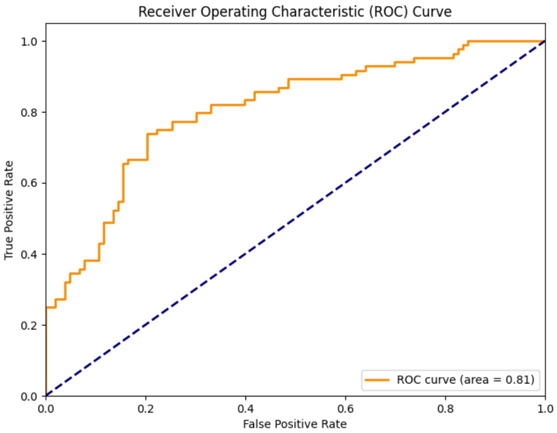


Fig. 3: ROC curves of the VGG model

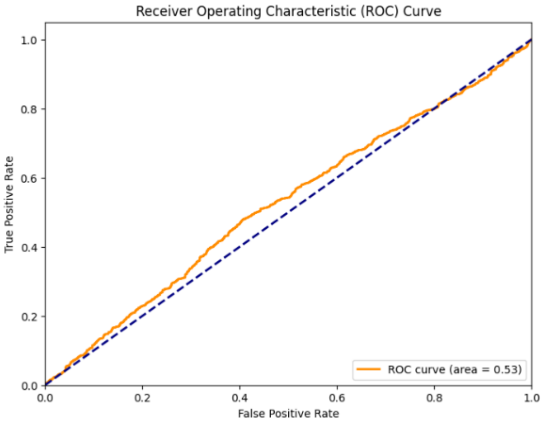


Fig. 4: ROC curves of the SWIN transformer