Evaluating the Effectiveness of Vision Transformers for Pediatric Polymicrogyria Detection

Jason Hao
Computer Systems Lab
Thomas Jefferson High School
for Science & Technology
Alexandria, USA
2025jhao@tjhsst.edu

Shaurya Jain
Computer Systems Lab
Thomas Jefferson High School
for Science & Technology
Alexandria, USA
2025sjain@tjhsst.edu

***Abstract*—Polymicrogyria (PMG) is a rare neurological condition characterized by abnormal cerebral cortex development, leading to excessive small, malformed gyri. With clinical manifestations including seizures, developmental delays, and intellectual difficulties, accurate diagnosis is crucial. This study explores the application of machine learning (ML) techniques, particularly fine-tuned Vision Transformers, to automate the detection of PMG in pediatric neuroimaging data. Utilizing the publicly available PPMR dataset, consisting of 4,517 PMG scan slices and 10,539 normal slices, we developed a scalable and reliable diagnostic tool that significantly enhances the accuracy and efficiency of PMG identification. Our results demonstrate that the fine-tuned BEiT and DeiT models outperform existing benchmarks, achieving high precision and recall while maintaining robustness against unseen data. Our best scores were achieved using the Microsoft BEiT model, having reached an accuracy, recall, precision, and accuracy of 0.999, 0.998, 1.000, and 0.999 respectively, on the entire dataset. This work highlights the potential of ML in clinical neurology, offering a path towards more timely interventions.**

***Index Terms*—polymicrogyria, deep learning, vision transformers, neurology**

I. Introduction

A rare neurological condition known as polymicrogyria (PMG) is defined by aberrant cerebral cortex development in the brain, which results in an excessive number of small, deformed gyri. Among the malformations that could occur in cortical development, 16% are attributed to PMG [1]. Many clinical signs, including seizures, crossed eyes, developmental delays, motor dysfunction, and intellectual difficulties, can result from the illness [2]. Advanced imaging methods, such magnetic resonance imaging (MRI) [3], [4], are usually necessary for PMG diagnosis. Trained radiologists examine these images. But this procedure can be laborious, subjective, and vulnerable to variation across observers, particularly when dealing with nuanced or unusual presentations.

The development of machine learning (ML) in recent years has opened up new possibilities for improving the efficiency and accuracy of diagnostics in a variety of medical fields. ML models can independently identify patterns and anomalies in medical images with a precision that matches—and occasionally surpasses—human skill by utilizing massive datasets and complex techniques. Our goal in this work is to investigate how machine learning methods may be applied to automatically identify polymicrogyria from neuroimaging data. In order to reduce diagnostic delays and improve patient outcomes, our project aims to create a scalable and dependable tool that helps doctors diagnose PMG through advanced image processing and classification algorithms. The remainder of this paper will discuss the relevant literature, the methodologies employed, the performance of the ML models, and their potential implications in clinical practice.

II. Related Work

Several recent studies have focused on the application of machine learning for the detection and diagnosis of neurological conditions through medical imaging, providing a foundation for the current research. A novel method proposed by Zhang et al. [5] introduces a center-based deep contrastive metric learning (cDCM) approach specifically for detecting PMG in pediatric brain MRI scans. Using the PPMR dataset, the authors employed a custom CNN architecture integrating dilated convolutions, squeeze-and-excitation blocks, and multi-level feature fusion. Their model achieved a recall of 92.01% but exhibited lower precision at 55.04%, indicating strong sensitivity but also a need for improvement in reducing false positives in PMG detection.

Related research in traumatic brain injury (TBI) diagnosis has also utilized deep learning to enhance imaging-based assessments. For instance, Phaphuangwittayakul et al. [6] developed a CNN-based framework for detecting hemorrhagic lesions in TBI using both public and private datasets, including the RSNA 2019 Brain Hemorrhage Challenge and PhysioNet datasets. Their approach highlights the importance of utilizing diverse data sources to improve the robustness of machine learning models in clinical applications.

Another study by Ellethy et al. [7] focused on mild traumatic brain injury (mTBI) in pediatric patients, employing a hybrid random forest-artificial neural network (RF-ANN) model. This model ranked clinical and demographic features alongside CT findings to distinguish between mTBI-positive and negative subjects. With an accuracy of 99.74% and precision of 99.25%, the RF-ANN model demonstrated remarkable potential for mTBI diagnosis, showcasing the capacity of artificial neural networks in handling clinical data efficiently.

Building on these approaches, an improved classification model for TBI was proposed by Gan et al. [8], integrating CNNs with recurrent neural networks (RNN) and an embedded squeeze-and-excitation (SE) module. This model, which also incorporated transfer learning to mitigate data insufficiency, achieved 95.9% accuracy in predicting damage at the slice level. The study underscores the significance of hybrid models in addressing local optimization challenges and improving diagnostic accuracy in complex neurological conditions.

Deep learning has further extended to the detection of post-traumatic epilepsy (PTE) biomarkers using electroencephalogram (EEG) data, as demonstrated by Faghihpirayesh et al. [9]. Their study developed a recurrent neural network to detect epileptiform abnormalities (EAs) in TBI patients, achieving an accuracy of 80.78%. This model represents a step toward automated and robust detection of PTE biomarkers, paving the way for enhanced patient monitoring in clinical environments.

These studies collectively demonstrate the growing efficacy of machine learning models in improving the detection and diagnosis of brain disorders, ranging from structural abnormalities like PMG to conditions such as TBI and mTBI. The application of deep learning in these areas is helping to overcome the limitations of traditional diagnostic approaches, and the integration of advanced techniques, such as contrastive learning, transfer learning, and hybrid models, continues to push the boundaries of diagnostic precision. This body of work serves as a strong basis for further research into automating the detection of PMG and related conditions using advanced machine learning methodologies. Attalah et al. also tackled general brain disorder detection, using datasets that included normal embryonic brain, agenesis of the corpus callosum, colpocephaly, mega-cisterna manga, Dandy–Walker malformation, agenesis of the septi pellucidi, cerebellar hypoplasia, and polymicrogyria [10].

III. Methodology

A. Data Acquisition

We utilized the PPMR dataset available on Kaggle created by Zhang et al. [5], which is to our knowledge the only publicly available dataset for pediatric polymicrogyria detection from MRI scans. Papers 12 and 13 performed PMG detection, but their datasets are both unavailable and only a small portion of their dataset contains PMG instances. The PPMR dataset was created by taking slices of 3D brain MRIs from 23 patients. All patients were aged under 18 years old and were from the Children's Hospital of East Ontario. No further information about the patients was provided. In total, the dataset contains 4,517 PMG scan slices and 10,539 normal scan slices.

B. Preprocessing

The dataset consisted of two folders of brain slices organized as controls and PMG study cases. Each of these two folders contain more folders with the brain images stored as JPGs. To make the dataset more suitable for use with the HuggingFace library and its easy-to-use pre-trained vision transformer models, we reorganized the dataset into "train" and "test" folders which each had a "ppmr" and "control" directory, which stored PMG and control data respectively. We then used the HuggingFace ImageFolder object that stores the data in a DatasetDict object. We created two such HuggingFace datasets from the original dataset available on Kaggle. The first was a balanced version of the dataset that would be utilized for training and it was composed of 4,517 control images and 4,517 PMG images, utilizing the maximum number of PMG instances available. We applied an 80-20 train-test split to this dataset for training. The second dataset we created was the original full unbalanced dataset, which would be used to evaluate the models at the end. Before feeding the images into the model for training and testing, all images were resized to 224 x 224 pixels.

C. Model

We fine-tuned three state-of-the-art pretrained vision transformer models for classifying PMG. The first model we tested was the Google ViT pretrained on the ImageNet-21k dataset [11]. The second model we applied was the Facebook DeiT model, which is a more efficient convolution-free vision transformer that has learned from a ResNet-like model through distillation [12]. Finally, we used the Microsoft BEiT model that used masked image modeling as a method to pretrain a vision transformer [13].

D. Training

All three models were trained on a Google Colab L4 GPU instance, ideal for its accelerated computing abilities. All three models were trained with the HuggingFace Trainer object for 8 epochs with a learning rate of 5×10^{-5}, a warm-up ratio of 0.1, and a batch size of 16. We used the default AdamW optimizer and Cross Entropy Loss. As Table 1 shows, all three models took a similar amount of time to train.

TABLE I: Model Names and Training Times

Model Used for Fine-Tuning	**Training Time(seconds)**
google/vit-base-patch16-224-in21k	2727.9148
facebook/deit-base-patch16-224	2833.3485
microsoft/beit-base-patch16-224-pt22k-ft22k	**2666.3419**

IV. Results and Discussion

We measured accuracy, precision, recall, and f1 for each of the models during training and testing. These metrics collectively assess the model's performance in fall detection, with accuracy providing a broad view of overall correctness, and precision and recall demonstrating the model's effectiveness in classifying images as PMG. A summary of the results of training and testing across our two datasets is in Tables 2 and 3.

Accuracy measures the overall correctness of the model's predictions across both PMG and control, given by the following equation:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Precision measures the proportion of correctly predicted PMG cases among all predicted PMG cases, shown by this equation:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

Recall measures the proportion of correctly predicted PMG cases among all actual PMG cases, given by the equation:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

The formula for the F1 score is:

$$\text{F1} = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{4}$$

In Equations (1), (2), and (3), TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively. Although all three models were able to achieve perfect accuracy on the training set for the balanced dataset, it should be noted that the BEiT and DeiT models converged to perfect accuracy slightly faster than the original Google ViT by almost 2 epochs. Both the Deit and BEiT fine-tuned models surpassed 0.99 across all metrics on both train and test, outperforming the original paper by a significant margin. All three models generalized to the data very well as indicated by the fact that they maintained a high level of accuracy even when tested on more than 6000 images that the model did not see during training. The original paper by Zhang et al. [5] achieves a high precision but low recall, indicating that the model identifies most true positives, but is hindered by false positives. In contrast, our paper's models achieve both a high recall and high precision, which would prevent confusion from false positives when applied in the real world. In fact our fine-tuned BEiT model's precision, recall, and accuracy exceed the DilatedCNN + FeatureFusion + SE model proposed by Zhang et al. [5] by approximately 0.45, 0.076, 0.144 respectively, highlighting that our model improves upon previous work by a huge margin. Overall, the main novelty of our work is creating a significant improvement in the detection of PMG from MRIs.

V. Conclusion

A. Summary

This study demonstrates how fine-tuned Vision Transformers can be used to improve deep learning and PMG recognition in MRI images. By automating the analysis of brain malformations, these algorithms offer a more efficient and accurate alternative to traditional diagnostic methods, which are often time-consuming and prone to human error. Our model's results show promise in identifying anomalies related to brain folding patterns. However, to improve the model's resilience and generalizability, more advanced designs and the integration of multimodal imaging data should be investigated in future research. With the potential to yield more timely interventions due to the capacity to detect small structural variations, we hope that this can one day help radiologists with early diagnosis. Ultimately, the application of AI-assisted tools could revolutionize the field of clinical neurology, not only improving lives, but saving them.

B. Limitations

Although our work shows promise, there are a number of limitations that should be considered and room for improvement. A major concern is that the original dataset was created by simply slicing 3D scans of 23 brains, so it is possible that our model needs to be exposed to data coming from more diverse sources before being deployed in the real world. The fact that the data stemmed from only a few sources could also explain the almost perfect performance of our models on the task of PMG classification. A few ways the variety of the data could be improved would be to collect brain scans from other hospitals across different countries and to maintain a balance across patient demographics like gender, ethnicity, and age. More data and training on a more varied dataset is necessary before these models can be implemented in the real world.

Another potential danger is the risk of over-reliance on the model in clinical settings without thorough validation. While our model shows strong performance on the current dataset, there's a possibility that it could misclassify or overlook subtle abnormalities when faced with more complex or noisy real-world data. Such errors could lead to delayed or incorrect diagnoses, especially if clinicians place too much trust in the model's predictions without cross-verifying with other diagnostic tools. Therefore, a rigorous validation process, including testing in real clinical environments and ensuring that clinicians are adequately trained to interpret the model's output, is essential before widespread adoption.

C. Future Work

As mentioned previously, the first and foremost priority would be to train a more robust model across more diverse. The models could also be more helpful to doctors diagnosing PMG if they highlighted specific regions of the brain that show evidence of PMG, which would require new kinds of annotated MRI scans. Our approaches to PMG classification through fine-tuning pretrained vision transformers could also be expanded to a variety of other brain diseases.

In the future we plan to make improvements to our models through better data and add new features, creating an invaluable tool for doctors around the world. ML could enable real-time monitoring systems that dynamically adjust treatments, improving patient care. However, successful implementation will require collaboration between artificial intelligence researchers and medical professionals to ensure models are

TABLE II: Model Evaluation on Balanced Dataset

Model Used for Fine-Tuning	Train/Test	F1 Score	Recall	Precision	Accuracy
google/vit-base-patch16-224-in21k	Train	0.976	0.952	1.000	0.986
google/vit-base-patch16-224-in21k	Test	0.973	0.948	1.000	0.974
facebook/deit-base-patch16-224	Train	0.998	0.997	1.000	0.999
facebook/deit-base-patch16-224	Test	0.998	0.996	1.000	0.998
microsoft/beit-base-patch16-224-pt22k-ft22k	Train	0.999	0.998	1.000	1.000
microsoft/beit-base-patch16-224-pt22k-ft22k	Test	1.000	1.000	1.000	1.000

TABLE III: Model Evaluation on Full Unbalanced Dataset

Model Used for Fine-Tuning	F1 Score	Recall	Precision	Accuracy
google/vit-base-patch16-224-in21k	0.975	0.952	1.000	0.986
microsoft/beit-base-patch16-224-pt22k-ft22k	**0.999**	**0.998**	**1.000**	**0.999**
facebook/deit-base-patch16-224	0.998	0.996	1.000	0.999
DilatedCNN +FeatureFusion +SE by Zhang et al. [5]	Not Provided	0.9201	0.5504	0.846

interpretable, clinically relevant, and ethically deployed, with continuous validation to address potential biases and risks.

D. Source Code

Source code for the project can be found at:

https://github.com/ShauryaJ1/polymicrogyria

References

[1] Kolbjer, S., Martín Muñoz, D. A., Örtqvist, A. K., Pettersson, M., Hammarsjö, A., Anderlid, B. M., & Dahlin, M. (2023). Polymicrogyria: epidemiology, imaging, and clinical aspects in a population-based cohort. *Brain communications*, 5(4), fcad213. https://doi.org/10.1093/braincomms/fcad213

[2] U.S. National Library of Medicine. (n.d.). Polymicrogyria: Medlineplus genetics. MedlinePlus. Retrieved from https://medlineplus.gov/genetics/condition/polymicrogyria/

[3] Leventer, R. J., Jansen, A., Pilz, D. T., Stoodley, N., Marini, C., Dubeau, F., Malone, J., Mitchell, L. A., Mandelstam, S., Scheffer, I. E., Berkovic, S. F., Andermann, F., Andermann, E., Guerrini, R., & Dobyns, W. B. (2010). Clinical and imaging heterogeneity of polymicrogyria: a study of 328 patients. Brain : a journal of neurology, 133(Pt 5), 1415–1427. https://doi.org/10.1093/brain/awq078

[4] Płoński, P., Gradkowski, W., Altarelli, I., Monzalvo, K., van Ermingen-Marbach, M., Grande, M., Heim, S., Marchewka, A., Bogorodzki, P., Ramus, F., & Jednoróg, K. (2017). Multi-parameter machine learning approach to the neuroanatomical basis of developmental dyslexia. Human brain mapping, 38(2), 900–908. https://doi.org/10.1002/hbm.23426

[5] Zhang, L., Abdeen, N., & Lang, J. (2024). A Novel Center-based Deep Contrastive Metric Learning Method for the Detection of Polymicrogyria in Pediatric Brain MRI. ArXiv.org. https://arxiv.org/abs/2211.12565

[6] Phaphuangwittayakul, A., Guo, Y., Ying, F. et al. An optimal deep learning framework for multi-type hemorrhagic lesions detection and quantification in head CT images for traumatic brain injury. Appl Intell 52, 7320–7338 (2022). https://doi.org/10.1007/s10489-021-02782-9

[7] Ellethy, H., Chandra, S. S., & Nasrallah, F. A. (2021). The detection of mild traumatic brain injury in paediatrics using artificial neural networks. Computers in Biology and Medicine, 135, 104614. https://doi.org/10.1016/j.compbiomed.2021.104614

[8] Gan, C., Sun, D., Qin, K., Zhao, H., & Xiao, F. (2020). Improved Traumatic Brain Injury Classification Approach Based on Deep Learning. https://doi.org/10.1145/3431943.3432288

[9] Faghihpirayesh, R., Ruf, S., Rocca, M. L., Garner, R., Vespa, P., Erdogmus, D., & Duncan, D. (2021). Automatic Detection of EEG Epileptiform Abnormalities in Traumatic Brain Injury using Deep Learning. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, 2021, 302–305. https://doi.org/10.1109/EMBC46164.2021.9630242

[10] Attallah, O., Sharkas, M. A., & Gadelkarim, H. (2020). Deep Learning Techniques for Automatic Detection of Embryonic Neurodevelopmental Disorders. Diagnostics, 10(1), 27. https://doi.org/10.3390/diagnostics10010027

[11] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. ArXiv:2010.11929 [Cs]. https://arxiv.org/abs/2010.11929

[12] Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., & Jégou, H. (2021). Training data-efficient image transformers & distillation through attention. ArXiv:2012.12877 [Cs]. https://arxiv.org/abs/2012.12877

[13] Bao, H., Dong, L., & Wei, F. (2021). BEiT: BERT Pre-Training of Image Transformers. ArXiv:2106.08254 [Cs]. https://arxiv.org/abs/2106.08254

Author Background

Jason and Shauyra are both seniors at Thomas Jefferson High School for Science and Technology (TJHSST), where they share a strong passion for artificial intelligence, mathematics, and their applications in solving real-world problems. Their interests lie particularly at the intersection of technology and medicine, with a focus on developing tools that can make a tangible impact on healthcare. As active members of their school's Computer Systems Lab, they have collaborated on a variety of projects aimed at harnessing computational techniques to address complex medical challenges, from creating diagnostic tools to optimizing healthcare delivery systems. Through their work in AI and mathematics, they continue to explore innovative ways to merge their technical knowledge with medical advancements, aiming to contribute to fields like neuromuscular research and medical technology. Their shared enthusiasm for research drives their commitment to using technology to improve lives, making them dedicated contributors to the field.