Pediatric Brain Age Prediction from Structural & Diffusion MRI Data

Research Project Final Report

Daksh Patel

Department of Pediatrics, University of Calgary

Supervised By Dr. Matthias Wilms

December 11, 2023

1. Abstract

Pediatric brain age prediction is a developing field that has significant implications for the early detection of neurodevelopmental disorders. The idea is to predict a person's biological brain age based on neuroimaging data and to understand the difference between their predicted biological and true chronological age. This is significant because these conditions are associated with deviations from typical developmental patterns, which may be detectable by a larger gap between predicted biological and true chronological age. This research project is being conducted by the Department of Pediatrics at the University of Calgary, under the supervision of Dr. Matthias Wilms. To predict the brain age of pediatric/adolescent subjects aged 5-22, we are using machine learning techniques and both T1-weighted and diffusion magnetic resonance imaging (MRI) data of the brain. To improve accuracy, we have developed a Convolutional Neural Network (CNN) based model that incorporates both T1-weighted and diffusion map inputs. We have adapted the Simple Fully Convolutional Neural Network (SFCN)[1] to accommodate multi-channel inputs, including four diffusion tensor imaging (DTI) maps. Our model development involves exploring both macrostructural (T1-weighted) and microstructural (DTI) neurodevelopmental differences. The study focuses on the intersection of machine learning and pediatric neuroimaging, and it has the potential for early diagnosis of conditions such as autism spectrum disorder and attention deficit hyperactivity disorder (ADHD). As we delve into the discussion and results, we will explore the nuances of model variations and their diagnostic implications.

2. Introduction

This study delves into the intricate field of predicting the biological age of pediatric brains, a concept with profound implications for understanding neurodevelopmental disorders. The notion of predicting brain age involves assessing the differences between an individual's biological age and chronological age, providing valuable insights into neurodevelopment. Neuroimaging, particularly through techniques like MRI, has become crucial in unraveling the mysteries of brain development. T1-weighted MRI images (T1w) are a foundational tool in this context, providing detailed structural information and capturing the macrostructural features of the brain, serving as a cornerstone for understanding brain anatomy.

Central to our investigation are Diffusion Tensor Imaging (DTI) maps[2], a powerful MRI technique that delves into the microstructural intricacies of brain tissue. Each DTI map measures a different property, collectively offering, for example, insights into the organization of white matter tracts of the human brain.

Four DTI parameter maps are utilized in our analysis:
Fractional Anisotropy (FA): Reflects the directionality of water diffusion, offering insights into microstructural changes related to white matter integrity.
Mean Diffusivity (MD): Represents the average diffusion rate of water molecules, indicating the overall diffusion characteristics within the tissue.
Axial Diffusivity (AD): Characterizes diffusion along the principal axis, contributing to the understanding of axonal integrity and water diffusion patterns.

Radial Diffusivity (RD): Describes diffusion perpendicular to the principal axis, providing additional information on myelin integrity and water diffusion dynamics.

The significance of comprehending pediatric brain age cannot be overstated, particularly in the realm of neurodevelopmental disorders. Disorders such as autism spectrum disorder and ADHD exhibit unique patterns of brain development[3]. For example, ADHD has been linked to a delayed maturation of the brain, a topic to which our study contributes in furthering understanding[4]. Identifying these atypical developmental trajectories early on could transform the way we diagnose and intervene in such cases.

This report explores the application of machine learning in the healthcare industry, focusing on using a Convolutional Neural Network (CNN) to utilize T1-weighted structural data and DTI maps for brain age prediction at both macro- and microstructural levels. This innovative approach enables us to accurately predict pediatric brain age and to potentially identify neurodevelopmental differences that may be linked to conditions affecting young individuals. Our methodology and findings showcase the potential for early detection and intervention in pediatric neurodevelopmental disorders. This research marks an intersection imaging technologies and machine learning, contributing to the broader discourse on pediatric brain age prediction and its implications for healthcare and developmental science.

3. Literature Review

The field of neuroimaging has been greatly influenced by the emergence of machine learning techniques, particularly deep learning[5]. These advancements have led to significant breakthroughs in disease prediction from medical imaging data, with deep learning even surpassing human performance in certain scenarios. In the realm of neuroimaging, deep learning is being tested for various applications, including brain age prediction and modeling[6], sex classification[7], disease prediction[8], and brain lesion segmentation[9]. However, challenges remain, particularly in the domain of 3D neuroimaging data, where memory demands exceed those of 2D images, necessitating innovative approaches such as down sampling, patch-based inputs, or 2D slices.

One application within neuroimaging is predicting chronological age based on structural brain MRI data. This task shares common challenges with other neuroimaging applications and serves as a testbed for developing and testing deep learning algorithms. The predicted age, commonly referred to as "brain age," holds potential clinical and biological relevance.

Various techniques, including regularized linear regression, support vector machines, Gaussian process regression[10], and deep learning methods, have been utilized to develop brain age prediction models. While significant progress has been made in this field, challenges remain, especially when working with small datasets. In these instances, simpler models may perform just as well as deep learning models. The question of whether more complex deep learning models provide significant advantages over simpler ones remains unanswered, and answering this question is essential for optimizing brain age prediction models for structural MRI data. The hypothesis of our project revolves around the potential advancement in accurately predicting brain age through the combination of macrostructural (T1-weighted) and microstructural

(Diffusion Tensor Imaging - DTI) analyses. While T1-weighted images alone have shown promise in predicting brain age, the incorporation of information from DTI maps is a speculative but potentially impactful approach that adds complexity and depth to the predictive model. DTI measures the diffusion of water molecules in brain tissues, providing insights into microstructural properties such as fractional anisotropy, mean diffusivity, axial diffusivity, and radial diffusivity. These microstructural features offer a hypothetical yet promising avenue for a more comprehensive understanding of the brain's intricate development, capturing nuances that may be overlooked by focusing solely on macrostructural features.

Accurately predicting brain age is a complex task that requires careful consideration of several challenges. One of the main challenges is the need for large sample sizes to effectively fit deep learning models, which contrasts with the limited sample sizes of neuroimaging datasets compared to vast natural image datasets[11]. This discrepancy can lead to overfitting, which limits the ability to learn image features effectively. The challenge becomes even more intense when dealing with 3D multi-channel data, as the training process becomes time-consuming and requires substantial computational resources. Despite these challenges, it is crucial to advance model architectures and strategies to address the complexities posed by neuroimaging datasets, ensuring robust and accurate brain age predictions.

The fusion of T1-weighted and DTI data offers a powerful approach that leverages both macro and microstructural information, potentially improving the precision of brain age predictions. This integration presents opportunities for detecting subtle neurodevelopmental variations in pediatric populations, particularly those associated with conditions such as autism spectrum disorder and ADHD. Although predicting brain age remains a complex undertaking, the combination of T1-weighted and DTI data needs to be explored for potential advancements in the field, paving the way for sophisticated analyses and more reliable predictive models.

4. Methodology

4.1 Dataset

The dataset is obtained from the Healthy Brain Network study[12]. With a sample size of over 2000 participants, aged 5-22 years, this dataset is uniquely suited for our research objectives. It comprises both healthy and diagnosed individuals. For each individual participant, the dataset includes a T1-weighted (T1w) image and four diffusion maps, totaling 86GBs. The T1w image, in nii (NIfTI) format, has a dimension of 193 x 229 x 193 with voxel spacing of 1 x 1 x 1, while each diffusion map has dimension of 108 x 129 x 108 with voxel spacing of 1.8 x 1.8 x 1.8. Notably, we employ the T1w image with segmented outer skull to reduce the input size during model training and focus the learning on crucial parts of the image. This strategic choice not only reduces the input size during model training but also directs the learning towards important regions of the image. There are no missing images for any of the participants, ensuring a comprehensive dataset.

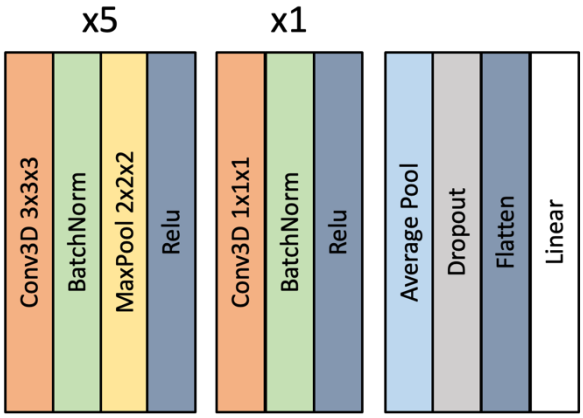
The preprocessing phase, a crucial step in ensuring standardized and reliable input for our deep learning model, is conducted using the MONAI framework[13]. The data for each participant undergoes various transformations, such as converting them to tensors, ensuring consistent voxel

spacing, resizing, padding, normalizing intensity, and concatenating channels. These transformations are essential to prepare the data for effective machine learning model training. Converting the data to tensors allows for compatibility with deep learning frameworks like PyTorch, providing a suitable format for the model to process. Ensuring consistent voxel spacing, resizing, and padding help standardize the dimensions across all images, enabling the model to learn features consistently. Normalizing intensity enhances the comparability of pixel values, contributing to the model's ability to discern relevant patterns. Concatenating channels combines the T1w image and diffusion maps into a cohesive multi-channel input, enriching the dataset with diverse information. These transformations collectively enhance the dataset's quality, facilitating optimal model performance during subsequent training and analysis phases.

4.2 Model Architecture

Our study relies on a Convolutional Neural Network (CNN) architecture for predicting pediatric brain age. More specifically, we extend the popular Simple Fully Convolutional Neural Network (SFCN) model[1]. This model, which is both lightweight and robust, is originally designed for classification and draws inspiration from the fully convolutional network (FCN) and VGG net[14]. However, in our application, we repurpose it for regression, focusing on predicting continuous values for brain age. This regression approach involves training the CNN using the actual age as a reference, allowing the model to learn and improve its predictions by minimizing the difference between its predicted age and the real age of subjects.

SFCN (Fig 1) is based on the FCN architecture and is tailored for predicting brain age using 3D minimally preprocessed T1w brain images. It consists of a sequence of basic blocks (Conv-Batch Norm-Activation) x N-Pooling. To adapt SFCN for our task, we made several modifications, including expanding it to accommodate five channels, including four DTI maps alongside T1-weighted images. Convolutional layers were adjusted accordingly with corrected channel numbers in each layer specified as [5, 32, 64, 128, 256, 256, 64].



(Fig1: Simple Fully Convolutional Neural Network Model)

The SFCN model consists of seven sequential blocks, each designed for hierarchical feature extraction. The first five blocks include 3D convolutional layers, batch normalization, max pooling, and ReLU activation, gradually reducing spatial dimensions. The sixth block introduces a $1 \times 1 \times 1$ 3D convolutional layer, batch normalization, and ReLU activation. The seventh block

integrates average pooling, dropout with a rate of 0.2, a flatten operation, and a fully connected layer with a linear output for predicting brain age.

The architecture follows an iterative process of feature extraction, increased nonlinearity, and mapping of generated features to predict brain age. The removal of fully connected layers reduces parameter count and enhances versatility for different input sizes. The integration of multi-channel DTI maps in addition to T1-weighted input data, allows our model to potentially capture both macrostructural and microstructural features for a more comprehensive and accurate brain age prediction.

4.3 Model Training

The process of training our model involves the division of the dataset into three sets: training, validation, and testing. The training set constitutes 75% of the data, the validation set is 10%, and the testing set is 15%. This strategic division ensures the evaluation of the model's generalization capabilities on previously unseen data, providing a robust assessment framework. Utilizing only healthy or typical children for training establishes a baseline for the model, and when predicting for a neurodevelopmental disorder, we anticipate observing differences in actual age compared to predicted age.

We set the learning rate to 0.001, facilitating a balance between convergence speed and precision during model training. To enhance efficiency in data loading and preparation, we employ 8 workers, allowing parallel processing of the dataset. Additionally, we use a batch size of 10 for both training and testing phases, ensuring an optimal balance between computational resources and model accuracy. Our training spans 100 epochs. To manage computational load, we incorporate gradient accumulation, allowing updates every two steps. To gauge model progress and prevent overfitting, we evaluate the model's performance every 5 epochs.

For optimizing our model, we employ the Mean Squared Error Loss function during training, measuring the difference between predicted and actual values. The Adam optimizer with an initial learning rate of 0.001 propels our model's convergence efficiency. To further refine the learning process, we incorporate a multi-step learning rate scheduler with milestones at epochs 15 and 80, adjusting the learning rate by a factor of 0.1 at these points. This dynamic adaptation promotes sustained model learning and fine-tuning throughout the training process.

5. Results

The trained models exhibit varying degrees of performance across different configurations. Table 1 shows summary of the Mean Absolute Error (MAE) values for models trained on healthy/typical children:

Model No.	Best Epoch	Input Channels	Model Modifications	Train MAE	Validation MAE	Test MAE
1	53	T1w + DTI	Base SFCN	0.7068	1.3053	1.2281
2	55	DTI	Base SFCN	0.6279	1.1681	1.2561
3	79	T1w	Base SFCN	0.6054	1.4130	1.6758

4	81	T1w + DTI	Added more convolution blocks	1.1394	1.7106	1.2423
5	96	DTI	Added more convolution blocks	0.7447	1.3374	1.2793
6	80	T1w	Added more convolution blocks	1.8609	2.0670	1.9249

(Table 1: Model Training and Performance)

The training process was guided by achieving the lowest validation Mean Absolute Error (MAE), determining the optimal epoch to stop training. Here are the key findings from our model training:

- Model 1, utilizing both DTI and T1w with the base SFCN architecture, achieved the lowest test MAE at 1.2281 during the 53rd epoch. Model 2 (DTI) at the 55th epoch exhibited a test MAE of 1.2561, reinforcing the effectiveness of the base SFCN architecture for accurate brain age estimation.

- While Model 3 (T1w) achieved a low training MAE, it showed increased test MAE, indicating potential overfitting. This highlights the importance of balancing training performance with generalization to new data.

- Among the models incorporating additional convolution blocks, Model 4 (DTI + T1w) demonstrated competitive performance with a test MAE of 1.2423 during the 81st epoch, surpassing Model 5 (DTI). This suggests that enhanced model complexity, in the form of added convolution blocks, did not improve model precision in estimating brain age.

- Models 6 (T1w) displayed the higher test MAE value, suggesting that the introduction of more convolution blocks significantly impacted brain age estimation for these configurations.

The overall assessment suggests that the models incorporating both T1w and DTI inputs, namely Model 1 (T1w + DTI) and Model 4 (T1w + DTI), potentially outperformed other configurations in terms of test MAE. This hints at the potential benefits of incorporating microstructural information (DTI) with macrostructural information (T1w) in models.

5.1 Model Selection for Final Analysis

We have identified Models 1, 2, and 3 for further detailed analysis due to their better performance with the base SFCN architecture compared to their counterparts trained with added convolution blocks in estimating brain age. The selection of these models ensures a standardized configuration, allowing us to specifically assess the impact of incorporating DTI maps on brain age prediction. This approach aligns with our project's primary objective of understanding the contribution of microstructural information, represented by DTI maps.

5.2 Statistical Analysis

5.2.1 Accuracy on Healthy Test Cases (T-test between MAEs)

We conduct paired t-tests to assess the accuracy of the models on healthy test cases, focusing on the mean absolute error (MAE). The null hypothesis (H0) posits that there is no significant difference between the model MAEs for healthy participants, while the alternative hypothesis (H1) suggests the presence of a significant difference.

For the t-test between Model 1 and Model 2, the obtained p-value is 0.6703. This result indicates that there is no significant difference in MAE between these two models when predicting the ages of healthy participants. Both models, incorporating T1w and DTI inputs, exhibit comparable performance in terms of accuracy.

However, when comparing Model 2 and Model 3, the p-value is 0.0023. This suggests a significant difference in MAE between the two models when predicting the ages of healthy participants. Model 3, which relies solely on T1w inputs, demonstrates a notable divergence in performance compared to Model 2, emphasizing the impact of input types on predictive accuracy.

The most substantial difference is observed in the t-test between Model 1 and Model 3, yielding a p-value of 0.0006. This signifies a significant discrepancy in MAE between the two models for healthy participants. Model 1, incorporating both T1w and DTI inputs, outperforms Model 3 in accuracy. The results underscore the advantage of integrating DTI maps for improved predictive performance, especially when dealing with healthy test cases.

The statistical analysis reveals nuanced differences in the accuracy of models when predicting ages for healthy participants. The inclusion of DTI inputs appears to contribute positively to model performance, as demonstrated by the superior accuracy of Model 1 compared to models relying solely on T1w inputs. These findings provide valuable insights into the impact of input types on the predictive capabilities of the models in the context of healthy test cases.

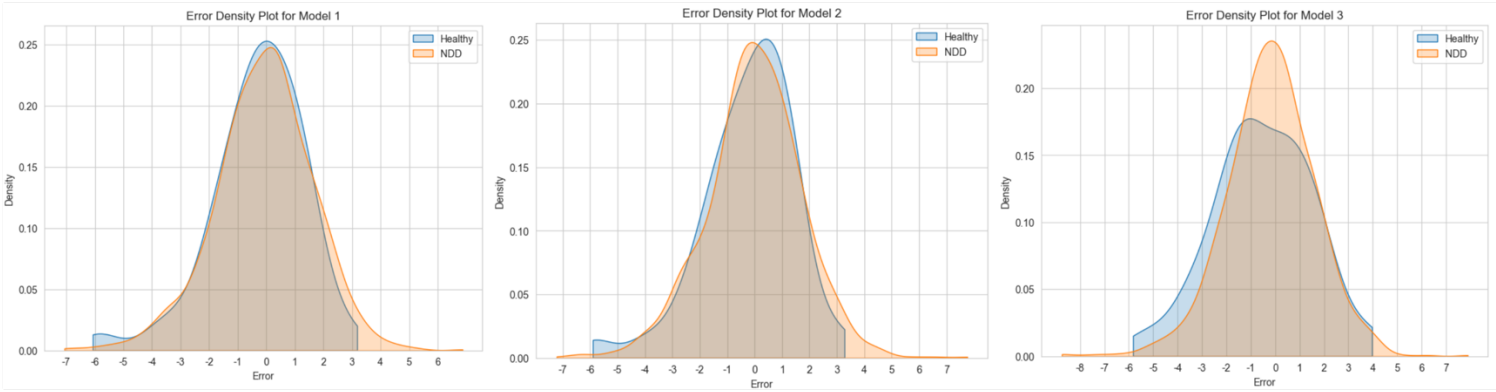
5.2.2 Comparison between healthy predictions and diagnosed predictions.

In addressing the inherent challenge of varying age distributions between healthy and neurodevelopmental disorder diagnosed individuals (specifically ADHD and Autism), we employ a strategic approach to ensure a fair evaluation of our models. The diversity in age ranges among these two groups can introduce biases and complicate the interpretation of model performance. To mitigate this, we have devised a solution by categorizing both healthy and neurodevelopmental disorder participants into specific age groups of equal size. This tailored approach allows for a more balanced comparison, removing potential disparities in population size across different age brackets. By implementing this solution, we aim to enhance the reliability and interpretability of our evaluations, gaining clearer insights into the model's effectiveness in predicting ages for these distinct groups.

a. Healthy vs NDD (No Age Matching)

The Welsh's t-test assessed the significance of the model errors (calculated by subtracting predicted age by actual age) in predicting ages for these distinct groups. The p-values for Model 1, Model 2, and Model 3 were 0.2724, 0.2915, and 0.1729, respectively. These non-significant p-values, consistently observed among the models, suggest that, without age matching, there is a uniformity in their performance when predicting ages for both healthy and neurodevelopmental disorder participants. However, the p-values indicate a lack of robust differentiation between the two populations, raising concerns about the model's ability to distinguish healthy and diagnosed cases effectively.

The Kernel Density Estimation (KDE) analysis (Fig 2) further elucidates the distribution and variability of prediction errors for each model. For Model 1, the KDE curve interpolates the best between healthy prediction error and NDD prediction error. Model 2 exhibits good interpolation as well, but the NDD error is centered around -1, indicating an underprediction of age for NDD. Model 3's KDE curves reveal distinct characteristics. The curve representing prediction errors for NDD participants is narrow, signifying concentrated errors within a smaller range. In contrast, the curve for healthy participants is flatter, suggesting that errors are spread out over a broader range.



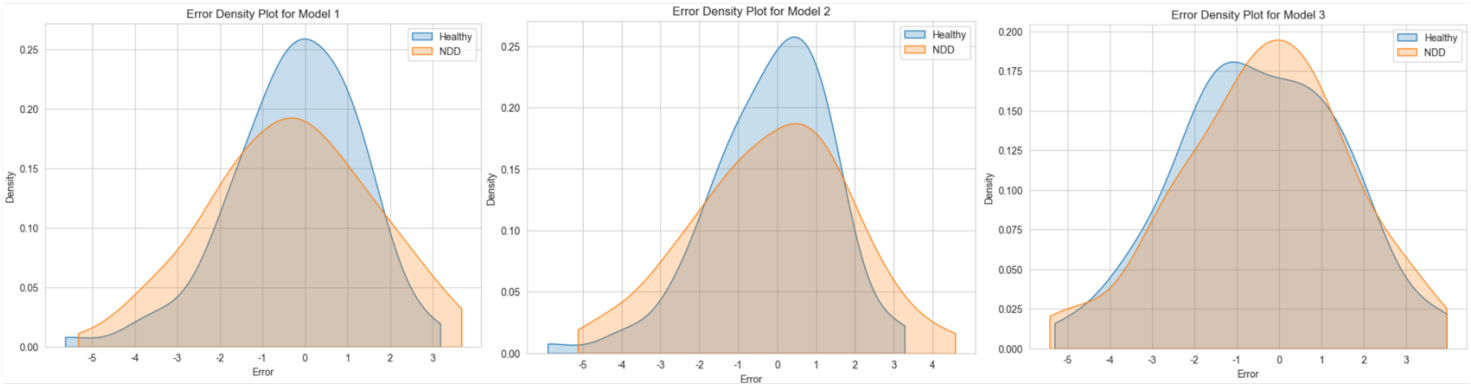
(Fig 2: Difference in prediction error between Healthy and NDD)

b. Healthy vs NDD (With Age Matching)

After matching the ages, we examined the difference in error density between the healthy and diagnosed populations. The subsequent Welsh's t-tests yielded p-values of 0.6654, 0.9914, and 0.7772 for Model 1, Model 2, and Model 3, respectively. Even with age matching, the models demonstrate non-significant differences in predicting ages for healthy and neurodevelopmental disorder participants. This outcome raises concerns about the model's efficacy in accurately distinguishing between these populations, as the lack of significant differences implies a potential limitation in their ability to capture the nuances associated with neurodevelopmental conditions.

The KDE analysis (Fig 3) provides additional insights into the error density distribution. For Model 1 and 2, healthy prediction error density curves are narrower and centered around 0, indicating smaller errors compared to NDD predictions. However, Model 1's NDD prediction errors are centered just below 0, suggesting underprediction, while Model 2's NDD prediction errors are centered slightly above 0, suggesting overestimation. Model

3 exhibits the best overlap between healthy prediction error and NDD prediction error, emphasizing its potential for capturing similar error patterns across both populations.

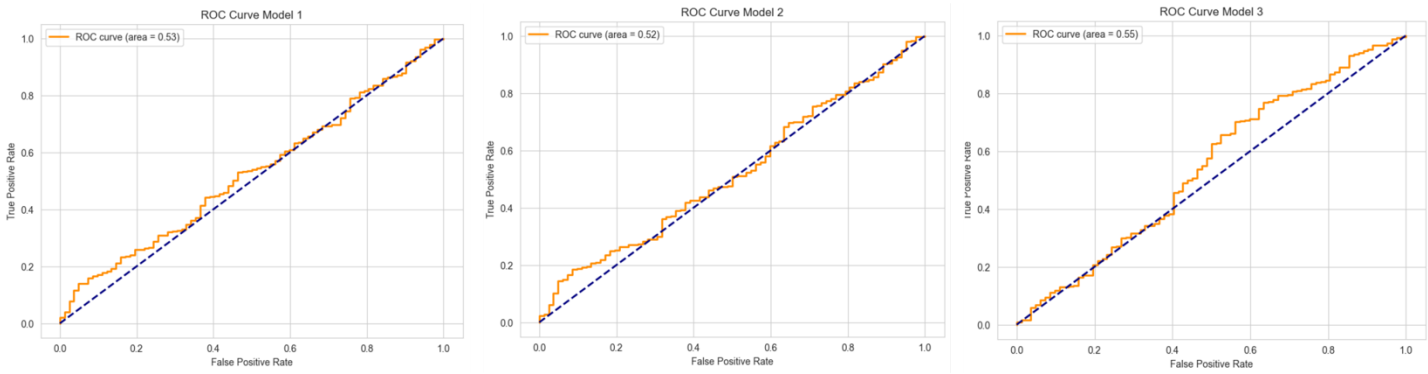


(Fig 3: Difference in prediction error between Healthy and NDD with matched ages)

The consistent model performance across both populations, coupled with the non-significant differences in errors, raises a cautionary note. The models, as reflected in the statistical analyses, may face challenges in effectively differentiating between healthy and diagnosed participants, potentially limiting their clinical utility in neurodevelopmental analysis.

5.2.3 ROC Curve analysis

Determining an optimal threshold for absolute error between the predicted biological age and the true chronological age of a subject is pivotal in assessing the model's predictive capabilities, especially when distinguishing between healthy and NDD predictions. The Receiver Operating Characteristic (ROC) curve, used for this analysis, offers insights into the trade-off between sensitivity and specificity.



(Fig 4: ROC Curve with area under the curve)

Despite considering various thresholds, it is essential to note that the ROC curve areas (Fig 4) for all models (Model 1: 0.53, Model 2: 0.52, Model 3: 0.55) are relatively close to 0.5. An area of 0.5 indicates that the model's ability to discriminate between healthy and NDD predictions is comparable to random chance. This observation, derived from the ROC curve analysis based on age predictions, suggests that the models may face challenges in reliably distinguishing

between the two classes solely based on absolute error. It is important to emphasize that this analysis evaluates the performance of age predictions with different age gap thresholds. This finding prompts a careful consideration of other evaluation metrics and a nuanced interpretation of the model's overall performance in the pediatric brain age prediction task. Further analysis and context-specific considerations are crucial to comprehensively assess the practical utility of these models in clinical applications.

6. Discussion

6.1 Takeaways

The results of our pediatric brain age prediction models offer valuable insights into their performance and potential applications. Notably, Models 1, 2, and 3, utilizing the base SFCN architecture with DTI input, demonstrated consistent performance and were selected for further analysis. The evaluation, based on mean absolute error (MAE) and statistical tests, revealed nuanced differences in their accuracy when predicting the ages. Importantly, the integration of DTI maps appears to positively contribute to the model's performance, particularly in distinguishing healthy and neurodevelopmentally disordered cases.

Achieving our primary objective, the comparison between models with DTI and T1-weighted inputs, unequivocally suggests that DTI aids in better predicting brain age in typically developing subjects than T1-weighted alone. This positive outcome aligns with our initial expectations and emphasizes the significance of incorporating microstructural information for more accurate brain age estimation.

However, when assessing the model's ability to differently predict ages for typically developing kids and a specific neurodevelopmental disorder (ADHD and Autism), the results indicate limitations. The observed non-significant differences in errors between the healthy and diagnosed populations raise concerns about the model's efficacy in accurately distinguishing between these groups. While the current models may not fulfill this specific objective, there remains optimism that they could potentially be effective for other neurodevelopmental disorders. Future work should explore and tailor models for different NDDs to enhance their clinical utility.

These takeaways highlight the achievements and limitations of our models, providing valuable guidance for future research and refinement. The positive impact of DTI in enhancing brain age prediction for typically developing subjects emphasizes the potential of our approach. Simultaneously, the recognition of challenges in predicting age for diagnosed participants encourage a more targeted and nuanced exploration in future studies.

6.2 Challenges

The prediction of pediatric brain age presented several challenges that warrant careful consideration. The dataset's size, especially for healthy participants, posed potential constraints on the model's generalization capabilities. The analysis of diagnoses, with inconsistent notation and severity levels, introduced complexity in drawing conclusive

insights. Training 3D multi-channel models, despite leveraging the University's Arc cluster, demanded meticulous planning and execution, highlighting resource demands in pioneering research.

Comparing models trained solely on healthy participants with those including both healthy and diagnosed subjects posed another challenge. The stark difference in dataset sizes raised questions about model robustness and potential biases. Variations in predictive performance among different models highlighted the delicate balance between complexity and effectiveness, seen in the nuanced impacts of architectural enhancements.

Despite promising MAE results, interpretability challenges persisted. The ROC curve analysis, despite being informative, indicated challenges in distinguishing between healthy and NDD predictions based solely on absolute error. This prompts a careful consideration of other evaluation metrics and a nuanced interpretation of the model's overall performance in the pediatric brain age prediction task.

6.3 Future Work

To address the challenges identified in our study and advance the field of pediatric brain age prediction, future work should focus on several key aspects. Strengthening the model's foundation involves enhancing the dataset through collaboration or integration, specifically addressing size discrepancies between healthy and diagnosed subjects. Refining the diagnosis annotation process with standardized notation protocols contributes to more robust analyses.

Optimizing computational resources via efficient model architectures or advanced hardware accelerates the exploration of model variations. Meticulously analyzing model predictions on diagnosed participants and focusing on diagnostic precision and nuanced differences in neurodevelopmental conditions are crucial steps. The challenges faced underscore the importance of a paradigm shift towards the diagnostic potential of models trained on healthy participants. A targeted approach, considering the intricacies of a healthy brain, is essential for accurate predictions in diverse pediatric neurological profiles.

Incorporating interpretability measures, such as saliency maps, offers a promising avenue for research. This analytical approach unveils specific brain regions contributing significantly to age prediction, bridging the gap between model predictions and clinical interpretability[15]. Future work should also delve into refining the models to effectively differentiate between healthy and diagnosed participants, addressing the challenges highlighted in our analysis.

The challenges encountered in developing and evaluating pediatric brain age prediction models provide a roadmap for future research. Addressing these challenges through dataset enrichment, model refinement, and interpretability advancements will align the project with the evolving landscape of pediatric neuroimaging. This iterative process will contribute to the field's growth, ultimately enhancing our understanding of brain development and age prediction in diverse pediatric populations.

7. Conclusion

In conclusion, our comprehensive analysis of pediatric brain age prediction models reveals valuable insights into their performance and challenges. Models 1, 2, and 3, trained exclusively on healthy participants, exhibit comparable Mean Absolute Error (MAE) values on healthy test cases. Despite incorporating diffusion tensor imaging (DTI) maps alongside T1-weighted images, the models face challenges in achieving robust discrimination. The statistical analyses, encompassing paired t-tests, Kernel Density Estimation (KDE) plots, and Receiver Operating Characteristic (ROC) curve areas, collectively indicate a nuanced landscape. The absence of significant differences in MAE between healthy and NDD predictions, coupled with ROC curve areas close to 0.5, suggests that the models may not be reliably used to identify a neurodevelopmental disorder.

While these findings shed light on the complexities of pediatric brain age prediction, it is essential to reflect on our primary goal. Our objective was to determine whether the inclusion of DTI maps contributes significantly to brain age prediction. The analysis reveals that despite achieving better model performance when trained on DTI alongside T1-weighted images, it does not consistently enhance the model's ability to predict brain age for neurodevelopmental disorder cases. This nuanced understanding contributes to the ongoing discourse on the role of DTI in pediatric brain age prediction and prompts further exploration.

8. References

1. Peng, H., Gong, W., Beckmann, C. F., Vedaldi, A., & Smith, S. M. (2021). Accurate brain age prediction with lightweight deep neural networks. Medical image analysis, 68, 101871.

2. O'Donnell, L. J., & Westin, C. F. (2011). An introduction to diffusion tensor image analysis. Neurosurgery Clinics, 22(2), 185-196.

3. Baecker, L., Garcia-Dias, R., Vieira, S., Scarpazza, C., & Mechelli, A. (2021). Machine learning for brain age prediction: Introduction to methods and clinical applications. EBioMedicine, 72.

4. Berger, I., Slobodin, O., Aboud, M., Melamed, J., & Cassuto, H. (2013). Maturational delay in ADHD: evidence from CPT. Frontiers in Human Neuroscience, 7, 691.

5. Davatzikos, C. (2019). Machine learning in neuroimaging: Progress and challenges. Neuroimage, 197, 652.

6. Cole, J. H., Poudel, R. P., Tsagkrasoulis, D., Caan, M. W., Steves, C., Spector, T. D., & Montana, G. (2017). Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. NeuroImage, 163, 115-124.

7. Arslan, S., Ktena, S. I., Glocker, B., & Rueckert, D. (2018). Graph saliency maps through spectral convolutional networks: Application to sex classification with brain connectivity. In Graphs in Biomedical Image Analysis and Integrating Medical Imaging and Non-Imaging

Modalities: Second International Workshop, GRAIL 2018 and First International Workshop, Beyond MIC 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 2 (pp. 3-13). Springer International Publishing.

8. Liu, M., Zhang, J., Adeli, E., & Shen, D. (2018). Landmark-based deep multi-instance learning for brain disease diagnosis. Medical image analysis, 43, 157-168.

9. Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., ... & Glocker, B. (2017). Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Medical image analysis, 36, 61-78.

10. Franke, K., & Gaser, C. (2019, July 9). Ten years of brainage as a neuroimaging biomarker of brain aging: What insights have we gained?. Frontiers. https://www.frontiersin.org/articles/10.3389/fneur.2019.00789/full

11. Raghu, M., Zhang, C., Kleinberg, J., & Bengio, S. (2019). Transfusion: Understanding transfer learning for medical imaging. Advances in neural information processing systems, 32.

12. Richie-Halford, A., Cieslak, M., Ai, L., Caffarra, S., Covitz, S., Franco, A. R., ... & Rokem, A. (2022). An analysis-ready and quality controlled resource for pediatric brain white-matter research. Scientific data, 9(1), 616.

13. Project Monai. Project MONAI - MONAI 1.3.0 Documentation. (n.d.). https://docs.monai.io/en/stable/

14. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

15. Liu, Z., Adeli, E., Pohl, K. M., & Zhao, Q. (2021). Going beyond saliency maps: Training deep models to interpret deep models. In Information Processing in Medical Imaging: 27th International Conference, IPMI 2021, Virtual Event, June 28–June 30, 2021, Proceedings 27 (pp. 71-82). Springer International Publishing.