PERTURBATION-BASED DEEP LEARNING REVEALS DYNAMIC SPATIOTEMPORAL SHIFTS IN PEDIATRIC BRAIN DEVELOPMENT FROM CT IMAGING

FUBUKI SAWA[1], NAOYA TAKASHIMA[1], SAYA ANDO[2], NAOMI YAGI[1], KUMIKO ANDO[3], REIICHI ISHIKURA[3], SYOJI KOBASHI[1]

[1]Graduate School of Engineering, University of Hyogo, Hyogo, Japan
[2] Hyogo Prefectural Amagasaki General Medical Center; Hyogo, Japan
[3] Kobe City Medical Center General Hospital; Hyogo Japan
E-MAIL:kobashi@eng.u-hyogo.ac.jp

Abstract:

Understanding pediatric brain maturation is essential for detecting neurodevelopmental disorders. However, quantitative spatiotemporal analysis remains limited in clinical practice, particularly using computed tomography (CT). This study proposes a perturbation-based deep learning framework that estimates brain age from pediatric CT scans while revealing region-wise contributions across developmental stages. A 3D ResNet architecture is trained on pediatric CT images to perform age regression and to enable region-wise interpretability through perturbation analysis. Anatomically guided perturbation analysis is conducted, consisting of lobe-wise masking to evaluate major regions. The proposed method was validated using head CT data from 201 pediatric patients aged 0 to 47 months with five-fold cross-validation stratified by age. The model achieved high accuracy with an RMSE of 5.205 months, an MAE of 4.007 months, and a Pearson correlation coefficient (r) of 0.925. The analysis reveals a dynamic posterior to anterior shift, with occipital lobe dominance in early infancy, emerging parietal lobe contributions in mid-infancy, and increasing frontal and temporal lobe importance in later stages. To our knowledge, this is the first CT-based study to investigate perturbation-driven region-wise interpretability in pediatric brain age estimation. The proposed framework offers a clinically meaningful approach for early developmental assessment.

Keywords:
Pediatric CT, Brain Age Estimation, Deep Learning, Perturbation Analysis, Explainable AI (XAI)

1. Introduction

The human brain undergoes complex maturation during early development, progressing across both space and time. Accurately characterizing this process is essential for identifying deviations from typical growth trajectories and for facilitating early detection of neurodevelopmental disorders [1, 2]. Brain regions do not mature uniformly; for example, the occipital lobe generally develops earlier than frontal areas [3]. Therefore, spatiotemporal characterization of brain maturation holds considerable clinical importance. While magnetic resonance imaging (MRI) has been the dominant modality for studying brain development, computed tomography (CT) remains widely used in pediatric settings due to its speed, accessibility, and suitability for emergency care [4]. However, CT's relatively poor soft-tissue contrast makes it difficult to apply conventional MRI-based developmental markers such as cortical thickness or gray matter volume [5]. As a result, CT-based evaluation of brain maturation remains limited and largely qualitative.

Brain age estimation (BAE) provides a quantitative index of developmental status by predicting brain age based on imaging data. Most BAE studies have relied on MRI and predefined structural features such as cortical thickness or gray matter volume, which are input into traditional machine learning models like support vector regression or ensemble methods [6, 7]. These approaches, however, are difficult to adapt to CT due to poor soft-tissue contrast. Moreover, they often lack interpretability, making it unclear which regions contribute to predictions. This is particularly problematic in infants and young children, whose brains are still undergoing myelination and present even less structural contrast.

Recent advances in deep learning, especially convolutional neural networks (CNNs), have enabled direct learning from imaging data without relying on handcrafted features. This makes CNNs suitable for analyzing CT scans. A few studies have applied CNNs to predict brain age from CT and demonstrated promising accuracy. For example, Morita et al. [8, 9]. used an AlexNet-based CNN to estimate brain age from CT, achieving a root mean square error (RMSE) of 6.26 months, and showing that deep learning can detect subtle developmental signals beyond the reach of

visual inspection. However, these models remain black boxes: they lack interpretability and provide little insight into the anatomical basis of predictions. Without understanding what the model relies on, clinicians are unlikely to trust or adopt such tools. Interpretability is therefore critical to ensure biological plausibility and clinical acceptance in pediatric applications.

To address this limitation, we propose a perturbation-based deep learning framework that enables both accurate prediction and region-wise interpretability in pediatric CT-based BAE. The model employs a ResNet34 architecture [10] trained on 3D head CT scans from children aged 0 to 47 months. To evaluate regional contributions, we implement a perturbation analysis: lobe-wise masking using anatomical masks. By stratifying outcomes across developmental stages, the method reveals spatiotemporal shifts in region importance, aligning with known maturation sequences. This anatomically grounded and quantitatively interpretable framework may facilitate earlier identification of developmental delays or pathological conditions and support more targeted risk assessments or intervention strategies.

Unlike previous BAE studies that rely on MRI-derived features or gradient-based visualization methods such as Gradient-weighted Class Activation Mapping (Grad-CAM) [11], our framework provides direct, region-specific interpretability tailored to CT data. Grad-CAM produces intuitive heatmaps but lacks spatial precision and quantitative assessment of regional contributions. In contrast, our perturbation-based approach systematically masks anatomically defined brain regions and directly measures their effect on predictions. This strategy (lobe-wise perturbation) enables quantitative comparisons across age groups and revealing patterns aligned with developmental processes.

While perturbation-based interpretability methods have been previously explored [12], many of them operate by masking or corrupting arbitrary image patches in 2D classification settings, often disregarding anatomical boundaries. This reduces spatial specificity, introduces biologically implausible alterations, and is poorly suited to volumetric medical data or regression tasks. In contrast, our framework applies anatomically guided masking to 3D pediatric head CT and evaluates its impact on age regression, allowing region-wise contributions to be interpreted in clinically and developmentally meaningful terms. Importantly, we quantify these contributions using changes in absolute prediction error, providing a direct and interpretable measure of each region's influence on developmental estimation.

To the best of our knowledge, this is the first study to systematically analyze region-wise contributions to BAE using pediatric CT data. By integrating perturbation-based interpretability with age-stratified analysis, our framework bridges the gap between predictive modeling and clinical relevance. It not only enhances model transparency but also provides developmental insights that are anatomically meaningful and quantitatively robust. These findings lay the groundwork for future CT-based screening tools to support early detection and personalized care in pediatric neurodevelopment.

2. Subjects and Materials

This study was approved by the Institutional Review Board (IRB) of Kobe City Medical Center General Hospital, with the requirement for informed consent waived due to its retrospective nature. A total of 201 pediatric patients (age range: 0 to 47 months) who underwent non-contrast head CT were included. All scans were reviewed by board-certified radiologists to confirm the absence of structural abnormalities that could affect global brain maturation, such as hydrocephalus or mass lesions. Each CT scan consisted of 200 to 300 axial slices with an in-plane resolution of 512 × 512 pixels. The voxel size ranged from 0.3 mm to 0.6 mm in-plane, with a consistent slice interval of 0.625 mm. Scans were acquired using standard pediatric protocols, with acquisition parameters adjusted by technicians based on patient age and head size. Voxel values were recorded in Hounsfield units (HU). All images were anonymized prior to analysis.

3. Proposed Method

3.1. Overall Workflow

To achieve both accurate BAE and region-wise interpretability, we propose a perturbation-based deep learning framework tailored for 3D pediatric head CT. While conventional perturbation approaches often rely on non-anatomical masking in 2D classification contexts, our method employs anatomically guided masking in 3D regression tasks, allowing biologically coherent analysis of brain maturation. This workflow consists of input standardization, model training using ResNet34, anatomical lobe segmentation, region-wise perturbation analysis, and statistical evaluation across developmental stages. An overview of the framework is shown in Figure 1.

3.2. Standardization of Pediatric Head CT

To ensure spatial and intensity consistency across subjects, all head CT images undergo a standardized preprocessing pipeline based on procedures established in

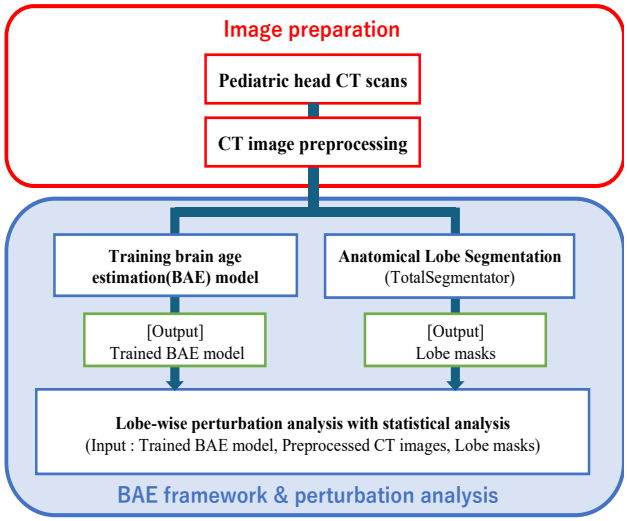


FIGURE 1. Overview of the framework in this study.

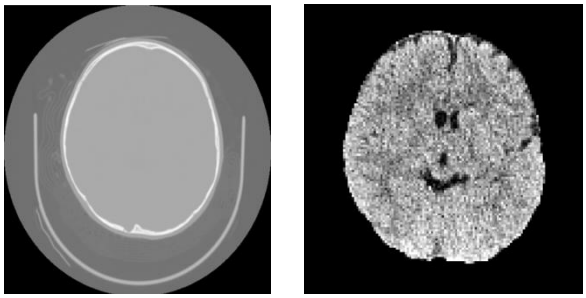
previous studies [8]. Non-brain tissues are removed through skull stripping, and spatial orientation is normalized using principal component analysis (PCA), aligning the brain's principal axes with the coordinate system. Voxel intensities are linearly rescaled using percentile-based normalization (P-tile method), applied only to the brain region after skull stripping, in order to reduce inter-scan variability while preserving anatomical contrast. Each scan is then resampled to isotropic 0.488 mm spacing and cropped to a fixed volume of 128 × 128 × 128 voxels centered on the brain. These procedures ensure that the input to the deep learning model is anatomically aligned and numerically stable across the dataset. An example of a head CT before and after preprocessing is shown in Figure 2.

3.3. Brain Age Prediction Model

A 3D ResNet34 architecture is adopted to perform brain age regression from preprocessed CT volumes. Chronological age, measured in months, is normalized to the range [0, 1] and used as the regression target. The model is trained to minimize the mean squared error (MSE) between the predicted and actual ages.

The use of a 3D residual network enables effective learning of developmental morphology from volumetric CT data, which presents greater challenges than MRI due to its limited soft-tissue contrast. This configuration captures subtle age-related structural patterns from normalized whole-brain CT scans and serves as a compact yet expressive baseline for subsequent region-wise interpretability analysis.

3.4. Anatomical Interpretability Analysis



(a) Original non-contrast axial CT image. (b) Preprocessed CT image with anatomical standardization.

FIGURE 2. Examples of a pediatric CT before and after preprocessing (subject age: 43 months).

To enable anatomical interpretability of brain age prediction, this study proposes a perturbation-based masking framework that quantifies the contribution of individual brain regions. This approach systematically modifies the input CT volume and measures its impact on model output. The goal is to estimate the spatial contribution of anatomical structures to the predicted brain age. Two levels of analysis are implemented: lobe-wise masking based on anatomically defined regions.

To identify anatomically defined regions in head CT scans, individual brain structures are automatically segmented using TotalSegmentator [13-15]. This tool extracts anatomical regions, including the frontal, temporal, parietal, and occipital lobes, and generates corresponding binary masks in which each voxel is labeled as either belonging or not belonging to a given region. The four lobes serve as regions of interest (ROIs) for masking analysis.

To evaluate each lobe's contribution to brain age prediction, voxels within the target lobe are replaced with uniform random noise in the range [0, 1]. This unstructured noise is used instead of zero-masking to mitigate potential bias caused by intensity normalization. The modified image is then input into the trained regression model, and the change in absolute prediction error is calculated as follows:

$$\Delta AE_{lobe} = |\hat{y}_{masked} - y| - |\hat{y}_{whole} - y| \quad (1)$$

where $\hat{y}_{masked}$ is the predicted age from the lobe-masked input, $\hat{y}_{whole}$ is the prediction from the unmasked input, y is the subject's chronological age. A higher ΔAE_{lobe} indicates that masking the given lobe leads to a greater prediction error, suggesting a greater influence of that region on the model's output.

3.5. Developmental Trends in Regional Importance

In addition to anatomical perturbation masking, this study proposes a group-level analysis framework to investigate developmental trends in regional brain

importance for age prediction. Subjects are stratified into four groups based on chronological age: 0-11 months, 12-23 months, 24-35 months, and 36-47 months. The division into four non-overlapping age groups is chosen for simplicity and to reflect major developmental milestones. However, the framework also flexibly accommodates alternative grouping strategies, such as overlapping bins or continuous sliding windows, enabling finer resolution analyses of developmental trends.

To further enhance the quantitative interpretability of the framework, we propose incorporating statistical validation of regional contributions through paired comparisons. For each subject, the absolute prediction error is calculated both from the original whole-brain input and from the input in which a specific lobe (frontal, temporal, parietal, or occipital) is masked. Paired comparisons are then performed between these two conditions to evaluate the impact of masking. The pairing structure directly compares each subject's prediction error between the whole-brain and lobe-masked conditions, enabling isolation of regional contributions while controlling for inter-subject variability.

Specifically, for each lobe and each age group, the distribution of the change in absolute prediction error (ΔAE_{lobe}) across subjects is first assessed for normality using the Shapiro-Wilk test. If the normality assumption is satisfied, a paired t-test is applied; otherwise, a Wilcoxon signed-rank test is used. In addition, effect sizes for each lobe and age group are calculated using Cohen's d to evaluate the magnitude of the observed differences. This process tests whether ΔAE_{lobe} values are significantly different from zero, indicating that masking the given lobe consistently affects age prediction performance.

4. Experimental Results

4.1. Experimental Setup

All experiments were conducted on a workstation running Ubuntu 20.04, equipped with an NVIDIA GPU (model NVIDIA RTX A6000). The primary programming environment was Python 3.10.16. Deep learning models were implemented using PyTorch Lightning 2.5.0.post0 and Torchvision 0.21.0. Image processing was performed using SimpleITK 2.4.1 and NumPy 1.26.4. Statistical analyses were conducted with SciPy 1.15.2 and scikit-learn 1.6.1. Anatomical lobe segmentation was performed using TotalSegmentator 2.7.0.

All head CT scans were preprocessed through skull stripping, spatial normalization using PCA, intensity normalization with the P-tile method, and resampling to an isotropic voxel size of 0.468 mm. Each scan was cropped to a standardized volume of 128 × 128 × 128 voxels centered on the brain.

TABLE 1. Impact of lobe-wise masking on brain age prediction accuracy across age groups. An asterisk (*) indicates statistical significance (p-value < 0.05). Symbols †, ‡, § indicate effect size magnitudes (Cohen's d), categorized as small ($0.2 \leq |d| < 0.5$), medium ($0.5 \leq |d| < 0.8$), and large ($|d| \geq 0.8$), respectively.

Age group	Mask type	ΔAE_{lobe}	p-values	Cohen's d
0-11 months	Frontal Lobe	-0.207†	0.233	-0.211
	Parietal Lobe	-0.281	0.357	-0.163
	Temporal Lobe	-0.049	0.838	-0.036
	Occipital Lobe	-0.007	0.957	-0.009
12-23 months	Frontal Lobe	-0.575†	0.058	-0.256
	Parietal Lobe	-0.203	0.717	-0.048
	Temporal Lobe	-0.045	0.916	-0.014
	Occipital Lobe	0.588*,†	0.009	0.361
24-35 months	Frontal Lobe	0.142	0.763	0.041
	Parietal Lobe	2.443*,‡	$p < 0.001$	0.552
	Temporal Lobe	1.143†	0.074	0.246
	Occipital Lobe	0.291†	0.133	0.206
36-47 months	Frontal Lobe	4.151*,§	$p < 0.001$	1.705
	Parietal Lobe	7.724*,§	$p < 0.001$	2.304
	Temporal Lobe	5.555*,§	$p < 0.001$	1.750
	Occipital Lobe	-0.318†	0.067	-0.249

The ResNet34 architecture was trained using stratified five-fold cross-validation based on subjects' chronological age, enabling balanced age distribution across folds. Chronological age was normalized to the range [0, 1] and used as the regression target. Training employed the AdamW optimizer with a learning rate of 0.0001 and MSE loss (MSELoss). The model was trained for 75 epochs with a batch size of 4. Random seeds were fixed at 42 to ensure reproducibility.

4.2. Brain Age Prediction Accuracy

The brain age prediction model achieved robust performance across the stratified test five-fold cross-validation. The RMSE averaged across folds was 5.205 months, the MAE was 4.007 months, and the Pearson correlation coefficient (r) was 0.925. These results demonstrate that pediatric brain age can be accurately estimated from standardized head CT images despite inherent limitations in soft-tissue contrast.

4.3. Lobe-wise Perturbation and Statistical Analysis

To quantify the contribution of major anatomical lobes to brain age prediction, lobe-wise perturbation analysis was performed as described in Section III. Each of the four lobes (frontal, temporal, parietal, and occipital) was independently masked, and the change in absolute prediction error (ΔAE_{lobe}) was calculated relative to the unmasked condition.

Subjects were stratified into four chronological age groups: 0-11 months, 12-23 months, 24-35 months, and 36-47 months, as described in Section III. Statistical significance of lobe contributions was evaluated within each group, and effect sizes were calculated.

The analysis revealed dynamic shifts in the importance of different lobes across developmental stages. In the youngest group (0-11 months), masking did not significantly affect prediction accuracy for any lobe. This result may reflect that, during early infancy, developmental features relevant to brain age prediction are highly distributed and subtle, making it difficult to isolate lobe-specific contributions using CT imaging.

At 12-23 months, masking the occipital lobe significantly increased prediction error (ΔAE_{lobe} = 0.588 months, p = 0.009, d = 0.361), indicating its dominant role during early infancy. A notable shift emerged at 24-35 months: while only the parietal lobe showed a significant effect (ΔAE_{lobe} = 2.443 months, p < 0.001, d = 0.552), the temporal lobe also tended to affect predictions (ΔAE_{lobe} = 1.143 months, p = 0.074, d = 0.246).

In the 36-47 month group, masking the frontal (ΔAE_{lobe} = 4.151 months, p < 0.001, d = 1.705), parietal (ΔAE_{lobe} = 7.724 months, p < 0.001, d = 2.304), and temporal lobes (ΔAE_{lobe} = 5.555 months, p < 0.001, d = 1.750) all significantly increased the prediction error, indicating their dominant contributions. In contrast, masking the occipital lobe slightly reduced the error (ΔAE_{lobe} = -0.318 months, p = 0.067, d = -0.249), suggesting reduced relevance at this stage.

These findings, summarized in Table 1, indicate a progressive anterior shift in the regions critical for brain age prediction, consistent with known neurodevelopmental trajectories from primary sensory maturation toward association cortex development.

5. Discussion

This study investigates how predictive brain regions evolve during early brain development based on pediatric CT, revealing a dynamic posterior-to-anterior shift in regional importance. Our perturbation-based deep learning framework not only achieved high accuracy in BAE but also provided region-specific insights across different developmental stages.

Transitions in critical brain regions observed in this study align with established neurodevelopmental processes [16]. The early prominence of the occipital lobe may indicate the rapid maturation of primary visual pathways and early myelination patterns. As development progresses, the increasing importance of frontal and parietal association areas around the age of three (36-47 months) corresponds to the maturation of language, memory, and higher-order cognitive functions. The strong predictive contributions from multiple associative regions may reflect the emergence of increasingly complex neural networks. Conversely, the diminishing impact of the occipital lobe at later stages may result from its early maturation plateau, making it less informative relative to still-developing regions, or may reflect limitations inherent to the model or CT imaging modality.

Our interpretability framework evaluates regional contributions based on changes in absolute prediction error rather than raw prediction differences. This choice is critical for ensuring that evaluation accurately reflects the deterioration of predictive performance relative to true chronological age. Simple prediction differences can be misleading, as masking certain regions may coincidentally bring predictions closer to the true value without genuinely improving model reliability. By focusing on changes in absolute error, our method reliably isolates regions whose integrity is essential for maintaining prediction accuracy. This design is particularly important for regression-based tasks like BAE, where deviations from ground truth carry clinical significance.

Unlike conventional perturbation-based interpretability methods like Occlusion Sensitivity [17] which evaluate changes in predicted values without direct reference to the ground truth, our framework assesses regional contributions based on changes in absolute prediction error. Conventional approaches are often sufficient for classification tasks but are less suited for regression problems like BAE, where accurate measurement of deviation from true age is critical. By focusing on absolute error shifts, our method mitigates global prediction bias and more reliably identifies regions essential for maintaining predictive accuracy.

Compared to previous work by Morita et al. [9], who used AlexNet-based CNN to predict brain age from pediatric CT with an RMSE of 6.26 months but without regional interpretability, our framework achieved a lower RMSE of 5.205 months while simultaneously providing spatiotemporal insights into predictive regions. These results demonstrate that our approach improves both predictive accuracy and transparency by revealing anatomically meaningful developmental patterns.

This study demonstrates that interpretable deep learning models can extract developmental insights even from structurally limited modalities like pediatric CT. By providing both accurate BAE and region-specific interpretability, our framework bridges a critical gap between predictive performance and clinical applicability. The ability to assess developmental trajectories from widely available CT data holds promise for supporting the early identification of atypical brain maturation, particularly in settings where

access to MRI is limited. Furthermore, by offering anatomically grounded explanations for model predictions, our approach contributes to enhancing the transparency and trustworthiness of medical AI systems in pediatric neuroimaging.

This study has several limitations. The analysis was based on a single-center study without external validation, and the sample size was modest. The region-masking approach may overlook nonlinear or interaction effects, and segmentation accuracy could affect results. CT's limited soft tissue contrast may also constrain feature extraction. Future work will validate the model in larger, more diverse cohorts and explore improved interpretability methods.

6. Conclusion

This study proposed a perturbation-based deep learning framework for BAE from pediatric CT images, achieving high accuracy while revealing dynamic spatiotemporal shifts in predictive brain regions. Lobe-wise analysis demonstrated early dominance of the occipital lobe, emerging contributions from the parietal lobe around two years of age, and increasing importance of the frontal, parietal, and temporal lobes in later childhood. By quantifying regional contributions through changes in absolute prediction error, the framework provides interpretable insights that align with known neurodevelopmental processes. These results highlight the potential of CT-based deep learning for transparent and clinically meaningful assessment of early brain development.

References

[1] E. L. Dennis and P. M. Thompson, "Typical and atypical brain development: a review of neuroimaging studies," Dialogues Clin Neurosci, vol. 15, no. 3, pp. 359–384, 2013.

[2] N. Gogtay *et al.*, "Dynamic mapping of human cortical development during childhood through early adulthood," Proceedings of the National Academy of Sciences, vol. 101, no. 21, pp. 8174–8179, May 2004.

[3] H. T. Chugani, M. E. Phelps, and J. C. Mazziotta, "Positron emission tomography study of human brain functional development," Ann Neurol, vol. 22, no. 4, pp. 487–497, Oct. 1987.

[4] J. Broder, L. A. Fordham, and D. M. Warshauer, "Increasing utilization of computed tomography in the pediatric emergency department, 2000-2006," Emerg Radiol, vol. 14, no. 4, pp. 227–232, Sep. 2007.

[5] A. M. Winkler *et al*, "Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies," Neuroimage, vol. 53, no. 3, pp. 1135–1146, Nov. 2010.

[6] Z. Lao *et al*, "Morphological classification of brains via high-dimensional shape transformations and machine learning methods," NeuroImage, vol. 21, no. 1, pp. 46–57, Jan. 2004.

[7] M. Tanveer *et al*, "Deep learning for brain age estimation: A systematic review," Information Fusion, vol. 96, pp. 130–143, Aug. 2023.

[8] R. Morita *et al*, "Pediatric brain CT image segmentation methods for effective age prediction models," World Automation Congress (WAC) 2022, pp. 408–412, Oct. 2022.

[9] R. Morita *et al*, "Quantification of pediatric brain development with X-ray CT images using 3D-CNN," Joint 12th International Conference on Soft Computing and Intelligent Systems and 23rd International Symposium on Advanced Intelligent Systems (SCIS&ISIS) 2022, pp. 1–3, Jan. 2022.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," Computer Vision and Pattern Recognition, arXiv:1512.03385, Dec. 2015.

[11] R. R. Selvaraju *et al*, "Grad-CAM: visual explanations from deep networks via gradient-based localization," Int J Comput Vis, vol. 128, no. 2, pp. 336–359, Feb. 2020.

[12] M. Ivanovs, R. Kadikis, and K. Ozols, "Perturbation-based methods for explaining deep neural networks: A survey," Pattern Recognition Letters, vol. 150, pp. 228–234, Oct. 2021.

[13] J. Wasserthal *et al*, "TotalSegmentator: robust segmentation of 104 anatomic structures in CT images," Radiology: Artificial Intelligence, vol. 5, no. 5, 2023.

[14] F. Isensee *et al*, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," Nature Methods, vol. 18, pp. 203-211, 2021.

[15] J. C. Cai *et al*, "Fully automated segmentation of head CT neuroanatomy using deep learning," Radiology: Artificial Intelligence, vol. 2, no. 5, Sep. 2020.

[16] J. E. Richards *et al*, "Brain development in infants (Chapter 4)," The Cambridge Handbook of Infant Development Brain, Behavior, and Cultural Context , pp. 94 - 127, 2020.

[17] I. Kakogeorgiou and K. Karantzalos, "Evaluating explainable artificial intelligence methods for multi-label deep learning classification tasks in remote sensing," International Journal of Applied Earth Observation and Geoinformation, vol. 103, p. 102520, Dec. 2021.