Peer Review File

Article information: https://dx.doi.org/10.21037/tp-24-219

Reviewer A

The topic is interesting however there are major issues with the study. The reasoning for the current study needs to be emphasized and compared with similar studies. Please see further comments below,

Reply: Highly appreciate your positive comments, The below is our detailed modification according to your valuable comments.

Abstract

Comment 1. *Please revise the methods section of the abstract. Be specific of the model utilized.*

Reply 1: Thank you for your valuable feedback on the abstract. We appreciate your suggestion and to revise the methods section including specific details about the model utilized.

Changes in the text: We have modified our text as advised (see Page 1-2, line 21-41).

Comment 2. *What are morphology features? Have you used linear or non-linear SVR model?*

Reply 2: Surface morphometry, including both surface and volumetric features, was also analyzed to construct morphology models. Specifically including mean white matter volume (WMV), mean gray matter volume (GMV), Cerebro-Spinal Fluid (CSF), and total intracranial volume (TIV), as well as surface features include cortical thickness, and this is described in detail in the second paragraph of *the Segmentation and Feature extraction* section of *Method* part.

In this study, the optimal nonlinear support vector model (SVM) with radial basis function was trained according to the data characteristics.

Changes in the text: we added this in revised manuscripts (see Page 1, lines 22,26-37)

Comment 3. *How do you know the selected features are optimal?*

Reply 3: Thank you for your insightful feedback, the question about feature selection is a critical step in radiomics processing. For this research, our treatment is as follows: Firstly, a univariate significance analysis(independent sample t-test or Mann-Whitney U test) is performed based on the distributional characteristics of the data to obtain the features that are significantly different between the two groups; then a LASSO regression analysis is executed, which is a commonly used method to deal with the high-dimensional data, making the coefficients of the unimportant features to be 0, thus ensuring that the important features are selected for further analysis.

The description of the detailed method of feature selection and dimensionality reduction in this study is placed in the section *Feature selection and dimensionality reduction* of *Method* part.

Comment 4. *What was the size of the training and test cohorts? Did you validate the model?*

Reply 4: Thank you for your valuable feedback. In the study, 33 preschool children with corrected TOF and 29 children in the control group (36 in the training cohort and 26 in the testing cohort) were included and were randomly divided on a 6:4 into a training (n = 36) and test (n = 26) cohort.

To validate the effectiveness and generalizability of the model, we employed the following validation methods:(1) Cross-Validation: 5-fold cross-validation was used in machine learning models to assess model stability and 10-fold cross-validation was performed for feature selection using LASSO regression. (2) Independent test set validation (6:4 randomized division): After training the model, we validated it using an independent test set.

(3) We utilized various evaluation metrics to comprehensively assess the model's performance, including accuracy, sensitivity, specificity, PPV (Positive predictive value), NPV (Negative predictive value), and the area under the curve (AUC). These metrics provided a thorough evaluation of the model's performance in different aspects. Through these methods, the model demonstrated in this study would perform good generalization and reliability.

Changes in the text: we added this in revised manuscripts (see Page1, lines17-19; Page6, lines 258-259; Pages10, lines 389-413)

Comment 5. *Monomodal was clearly not well trained, it's memorized the data and not replicable.*

Reply 5: Thank you for your insightful feedback. To improve the robustness and repeatability of the monomodal, the following points are carried out in this study (1) Data splitting and cross-validation (2) adjusted key training parameters (3) multiple models were trained (4) model performances were assessed. In summary, the above methods can improve the robustness as well as the repeatability of the training model.

Comment 6. *How did you compare the AUC values?*

Reply 6: Thank you for your insightful question. We compared the AUC values using the DeLong test, which is a non-parametric approach specifically designed for comparing the areas under correlated ROC curves. This method is widely accepted for its robustness and accuracy in assessing the statistical significance of the difference between two AUCs.

Changes in the text: we added this in revised manuscripts (see Page 2, lines 40-41).

Comment 7. *What significant indicators?*

Reply 7: Significance indicators are those optimal metrics that are operated by feature selection and dimensionality reduction and are significantly different after performing significance analysis and LASSO regression.

Main Text

Comment 1. *Radiomics is not " based on machine learning algorithms,". Please correct the statement.*

Reply 1: Thank you for pointing out the inaccuracy in our description of radiomics. We have corrected the relevant statement.

Changes in the text: We have modified our text as advised (see Page 3, line 93-95).

Comment 2. *Radiomics is a blanket name for the features extracted from medical imaging data.*

Reply 2: Yes, couldn't be more certain, thank you for pointing out the sloppy presentation

instructions on the definition of radiomics, I have made changes to the corresponding statements in the text.
Changes in the text: We have modified our text as advised (see Page 3, lines 93-95).

Comment 3. *Please revise the MRI statement which can cause misleads among the readers.*
Reply 3: Thank you very much for pointing that out. I have made changes to this as follows:
Changes in the text: We have modified our text as advised (see Pages 4, lines 123-127).

Comment 4. *Please revise the acquisition time statement as well. DW-MRI is also sort of functional image that can be acquired much faster using EPI sequence. Therefore, the statement is kind of misleading.*
Reply 4: Thank you very much for the correction.
Changes in the text: We have modified our text as advised (see Pages 4, lines 123-127).

Comment 5. *Please avoid first person tone in the manuscript.*
Reply 5: Thank you for your valuable suggestions, we have read through the entire text again and made changes to the presentation.

Comment 6. *In a quick search, similar studies were shown in Pubmed. Please revise the statement.*
Reply 6. Thank you very much for pointing this out, we have made changes to this.
Changes in the text: We have modified our text as advised (see Page 5, lines 217).

Comment 7. *Did you acquire single average T1W? Which T1w MRI sequence did you use?*
Reply 7. Yes, in this study, we acquired single average T1-weighted (T1W) MRI images. We employed the Magnetization-Prepared Rapid Acquisition with Gradient Echo (MPRAGE) sequence. This sequence is widely used in both clinical and research settings for high-resolution structural imaging, providing clear gray matter and white matter contrast.

Comment 8. *Please briefly describe the preprocessing steps.*
Reply8. Okay, we added a detailed description of the preprocessing section in the original article.
Changes in the text: We added this in revised manuscripts (see Page 8, lines 304-319).

Comment 9. *how did you normalize the MRI data? How did you deal with the relative intensity?*
Reply9. Thank you for your insightful questions. We performed spatial normalization of MRI images using SPM. Intensity bias and noise are reduced by bias field correction. And Z-score normalization is performed for intensity.
Changes in the text: We added this in revised manuscripts (see Pages 8, lines 304-319).

Comment 10. *Please give the list of the extracted features as supplementary.*
Reply10. Thank you for your valuable suggestions, all features have been listed and placed in Table S2 of the supplementary material and mentioned in the text.
Changes in the text: We added this in revised manuscripts (see Page 9, lines351-352).

Comment 11. *For texture features, did you use fixed bin or fix width?*
Reply11. In this study, a fixed-width approach is used, which is able to ensure that each interval covers the same range of intensity values and is more suitable for dealing with the continuous distribution of gray values common in medical imaging. It can ensure the balance of the individual grayscale intervals and thus better capture the subtle texture variations in medical images. This approach also helps to standardize the feature extraction process for different images, ensuring consistent and comparable features. The bin width is set to 25 for this study and the description is added to the methods section in the original article.
Changes in the text: We added this in revised manuscripts (see Pages 9, lines 351).

Comment 12. *How did you deal with the NA values?*
Reply 12. Thanks for your reminder. Fortunately, there are no null values in this study.

Comment 13. *Please be specific about number of radiologists performing segmentation.*
Reply 13. Thanks to your suggestion, this study was conducted by two physicians outlining the area of interest layer by layer. See the manuscript for details.
Changes in the text: We have modified our text as advised (see Page 8, lines 330; Page 9, lines 355-357).

Comment 14. *How did you decide which roi to use? How did you deal with the controversies?*
Reply 14. In this study, the consistency and reproducibility of radiomics features were assessed by the **intra-observer and inter-observer agreement analysis** and when the intraclass correlation coefficient (ICC) was greater than or equal to 0.75, it was included in the study. At the same time, the spatial consistency of image segmentation results to evaluate overlap or IOU was assessed by the Dice coefficient.
Changes in the text: We added this in revised manuscripts (see Page 9, lines355-363).

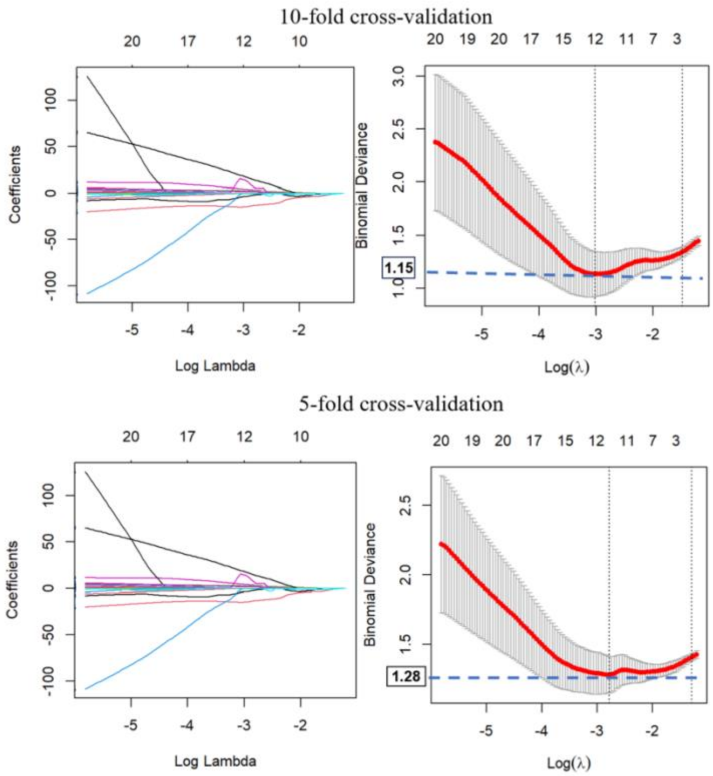
Comment 15. How did you select the hyperparameters?
Reply 15. Thank you for your question regarding the selection of hyperparameters in our radiomics study. The selection of hyperparameters is a critical step to ensure the robustness and generalizability of our model. We adopted the following approach:
(1) Cross-Validation: We employed 5-fold cross-validation to systematically explore a range of hyperparameter values. This approach allowed us to evaluate the performance of our model across different hyperparameter settings and to select those that minimized the cross-validation error while avoiding overfitting. (2) Performance assessment: The final hyperparameters were chosen based on the best performance metrics (accuracy, AUC, sensitivity, specificity, PPV, NPV) obtained during the cross-validation process. (3) Domain Knowledge: Additionally, we leveraged domain knowledge to further refine our hyperparameter selection, ensuring that the choices made were not only results but also clinically reasonable.

Comment 16. *10-fold cross validation may not be the optimal approach for your dataset. You have about 60 samples, 48 for training leaves 5 samples for the validation.*
Reply16: Thank you for your valuable feedback on our manuscript. We appreciate your

insightful comments and understand that the choice of cross-validation can significantly impact the stability and reliability of results, especially with small datasets. Therefore, this study is modified as follows:

Based on the small sample size of this study, it may be appropriate to select a smaller number of folds when performing cross-validation. There are two places in this study where cross-validation is used to improve model stability: (1) Feature selection: this study tries to compare the model performance of ten-fold cross-validation and five-fold cross-validation, and for this dataset, the ten-fold cross-validation obtains smaller variance: Minimum Deviance: 1.132552 (for ten-fold) Minimum Deviance: 1.281921 (for five-fold), as shown in the Fig below. Therefore, considering all the factors, ten-fold cross-validation is finally chosen for feature selection in this study. (2) Machine learning model building. When building the model, in order to improve the stability of the model, according to your suggestion, five-fold cross-validation is used for modeling, and the model still shows good performance. The results are shown in the modified section of the paper. From the results of re-training the machine learning model, the results show that it still exhibits better performance without the risk of overfitting. This is a good indication of the stability of the model.

Changes in the text: We have modified our text as advised (see Page 10, lines394, 408).



Comment 17. *Which filters did you use for the pyradiomics package?*

Reply 17: Thank you for your insightful question regarding the filters used in the pyradiomics package. In our study, we employed the following filters to extract first-order features, texture features, and wavelet features: (1) First-order features: We utilized basic statistical filters to capture first-order statistics from the images. These features include metrics such as mean, variance, skewness, and kurtosis, which describe the intensity distribution within the ROI

(Region of Interest). (2) Texture Features: Gray Level Co-occurrence Matrix (GLCM) Filter: This filter was used to derive texture features by analyzing the spatial relationship between pixel intensities. GLCM features capture information about the texture patterns and their frequency, which is crucial for identifying heterogeneity within the ROI. Gray Level Run Length Matrix (GLRLM) Filter: The GLRLM filter was applied to extract features that quantify the length of consecutive runs of pixels with the same intensity. This helps in capturing texture properties related to the homogeneity and variability of structures within the image. Gray Level Size Zone Matrix (GLSZM) Filter: This filter was used to obtain features that describe the size of homogeneous zones of a specific gray level in the image, providing additional texture information. (3) Wavelet Features: Features extracted from wavelet decomposed images. Typically, eight wavelet sub-bands (LLL, LLH, LHL, LHH, HLL, HLH, HHL, HHH) are used to capture both low and high-frequency information. We ensured that the implementation of these filters followed the guidelines and recommendations provided in the Pyradiomics documentation. (https://pyradiomics.readthedocs.io/en/latest/features.html).

Comment 18. *Which wavelet kernel did you use? How many level did you use as well?*
Reply 18: Thank you for your insightful questions.
(1) We utilized the wavelet features extraction functionality provided by the Pyradiomics package. The specific wavelet kernel employed was the Discrete Wavelet Transform (DWT) with the Daubechies wavelet (db4). This choice is consistent with standard practices in radiomics and complies with IBSI standards, and provides a good balance between localization in the time and frequency domains, which is crucial for accurately capturing the subtle features in brain imaging data.
(2) The analysis involved a three-level wavelet decomposition. This decision was guided by preliminary analyses indicating that three levels were sufficient to capture the relevant textural features while maintaining computational efficiency and avoiding overfitting. The levels were chosen to ensure that we could extract meaningful patterns across different scales of the imaging data.

Comment 19. *What was the percentage of intersection of union for the ROIs?*
Reply19: Thank you for your valuable questions. The IoU is a standard metric for evaluating the accuracy of segmentation, Based on your valuable advice, then the intersection over union (IoU) percentage was used to evaluate the overlap between the ROIs. Calculations were performed using Python with the SimpleITK library. The median IoU for intra-observer ROI overlap was 95%, with an interquartile range (IQR) of 94% to 96%. For inter-observer ROI overlap, the median IoU was 84%, with an IQR of 79% to 85%.
Changes in the text: We added this in revised manuscripts (see Pages 12, lines 460-462).

Comment 20. *Please revise the statement for wavelet in discussion section. How is thresholding utilized in wavelet transform?*
Reply20: Thank you for pointing this out. In this study, we utilized *PyRadiomics* to extract wavelet features from the images. *PyRadiomics* employs wavelet transform to decompose images into different frequency components, allowing the extraction of detailed information. *PyRadiomics* directly extracts features from the wavelet coefficients without explicit

thresholding.
Changes in the text: We have modified our text as advised (see Page 16, lines 756-757).

Comment 21. *Please revise the Footnote section.*
Reply: I appreciate your reminder that the footnote section has been changed in the revised draft.

Reviewer B
In this manuscript, the authors employ machine learning models to identify radiomic and morphometric features from MRIs that identify children with tetralogy of Fallot (TOF, n=33) in comparison to children without TOF (n=29). The performance of individual models and a fusion model with features from both categories were assessed, with the fusion model showing the highest performance. Correlation of specific radiomic features with clinical outcomes and neurodevelopmental scores in children with TOF were assessed. Neurodevelopmental differences are prevalent among people with congenital heart disease including TOF, and identifying potential biomarkers for outcomes is an important need in the field. However, the small sample size limits the generalizability of their findings. Further, the grammar and formatting limited readability. Major revisions to the text of the manuscript are required to improve readability and interpretability of the research.
Reply: Thank you for your detailed and thoughtful review, which is of great help to our improvement. Based on your valuable suggestions, I will carefully and meticulously revise it point by point.

Major consideration:
Comment 1. *The small sample size of this cohort limits the generalizability of the results, particularly given the machine learning method risk of overfitting.*
Reply 1: Thank you for your valuable feedback. The small sample size of our cohort indeed poses a limitation to the generalizability of our results. We acknowledge that with a limited dataset, machine learning models are prone to overfitting.

To avoid this risk, we implemented several strategies:(1) Cross-Validation: We used 5-fold cross-validation to assess the model's performance and ensure that it is not overly fitted to a specific subset of the data. (2) Regularization Techniques: Regularization methods, L1 regularization, were employed to penalize overly complex models and reduce the likelihood of overfitting. (3) Model Simplicity: We opted for simpler machine learning models where suited for high dimensional features, as they are less likely to overfit compared to more complex models.

Moreover, we use more than one machine learning model in our modeling, which also ensures the robustness of the results.

Then, the reason for the small sample size in this study is that the study explores the structural changes in the brain of school-age children with TOF and long-term postoperative neurodevelopment, each child not only needs to collect neurodevelopmental scales, and clinical demographics, but also ensure image quality while undergoing MR scanning, so it is not easy to achieve a large-sample analysis in the short term. This study will continue to follow the long-term neurodevelopment of younger (infants) TOF through school age and beyond. Meanwhile, this study is related to the incidence of TOF.

Further, we perform the re-modeling analysis used 5-fold cross-validation in the revised manuscript, and the results still obtain better performance.

We have described this in the limitations of the study.

Changes in the text: We added this in revised manuscripts (see Pages 18, lines833-835, 842-848).

Comment 2. *Methods describing the statistical correlation assessment between radiomics with clinical and neurodevelopmental were not properly described. Authors should add these methods, including how features were chosen and assessed for correlation with neurodevelopmental assessments.*

Reply2. In this study, the Shapiro-Wilk Normality Test was used to examine the normality distribution of clinical and neurodevelopmental indicators. Then, Pearson correlation or Spearman correlation analyses were performed based on the distributional to further explore the quantitative imaging metrics and neurodevelopmental correlations as well as their underlying clinical mechanisms.

We have added "*Correlations of radiomics features with clinical and neurodevelopment metrics*" and "*Statistical*" sections to the Methods section. analysis" section, which describes the methods used in detail.

Changes in the text: We added this in revised manuscripts (see Page 11, lines 421-429,lines 431-442).

Comment 3. I*n regard to neurodevelopmental correlations, several co-variates known to be important to neurodevelopment are not included in the analysis. These include demographic characteristics for the cohort, as well as clinical characteristics such as gestational age at birth (i.e., whether children were born preterm), parental education and income. Where available, these should be considered in the analysis and if not available is a limitation to the study*

Reply 3: Thank you very much for your insightful comments and for highlighting the importance of considering additional covariates known to influence neurodevelopmental outcomes. We fully acknowledge that demographic characteristics such as age, gender, parental education, and income are critical factors in neurodevelopmental studies. In this study, we excluded preterm infants from the inclusion criteria. Therefore, we have collected and added to this in our revision manuscript.

Changes in the text: We added this in revised manuscripts (see Pages 7, lines 275-283; Page18, lines 849-853).

Minor considerations:

Comment 1. *Grammar and writing quality throughout the manuscript should be improved. Examples of specific instances are listed below but not exhaustive.*

Reply1: Thank you for your thoughtful pointing out, and we will revise it point by point based on your suggestions.

Throughout the manuscript: "morphology" should be replaced with "morphometry" in the context of surface and volume-based analyses.

(It has been revised throughout in the revised manuscript based on your suggestions.)
Line 17 Abstract: “health” should read “healthy”
(It has been revised in the revised manuscript based on your suggestions.)
Line 30: Sentence is not grammatically correct.
(Changes in the text: We have modified our text as advised in the revised manuscript. (see Page 3, lines 46-47).
Line 49: Grammar - “still exist” is incorrect.
(Changes in the text: We have modified our text as advised in the revised manuscript. (see Page 3, lines 78)
Line 75: Grammar - this sentence is incomplete.
(Changes in the text: We have modified our text as advised in the revised manuscript. (see Page 4, lines 119-120)
Lines 78-96: This paragraph is poorly written and should be rephrased to describe studies in the context of importance to this work, as opposed to summarizing individual studies in single sentences.
(Changes in the text: We have modified our text as advised in the revised manuscript. (see Pages 4-5, lines 118-138)
Line 102: Suggestion to provide an explanation of the in-depth quantitative radiomic features, categories, and examples of how they are obtained.
(Changes in the text: We have modified our text as advised in the revised manuscript. (see Pages 5, lines 203-214)

Line 104: Grammar - “my knowledge” should read “our knowledge”
(Changes in the text: We have modified our text as advised in the revised manuscript. (see Pages 5, lines 217)
In general, the introduction is lacking background on Tetralogy of Fallot as a disease, including the type of heart defects it includes, incidence rate in children with CHD, and risk factors/mechanisms for its association with abnormal neurodevelopment. A paragraph on this should be added.
(Changes in the text: We have modified our text as advised in the revised manuscript. (see Pages 2-3, lines 64-76)
Methods:
Line 128: “refusal to participate in the study” is an implied criteria and should be removed.
(Changes in the text: We have modified our text as advised in the revised manuscript. (see Pages 6, lines 252)
Lines 134-135: The exact number of cases used in the training set and validation set should be provided as opposed to the ratio.
(Changes in the text: We have modified our text as advised in the revised manuscript. (see Pages 6, lines 258-259)

Line 139: At which ages were the WPPSI-IV was administered for the TOF group? This information should be added to text and in Table 1.
(Changes in the text: We have modified our text as advised in the revised manuscript. (see Pages 6, lines 265)

Line 148: Provide a definition/context of the McGoon index.
(Changes in the text: We have modified our text as advised in the revised manuscript. (see Pages 7, lines 276-277)

Line 159: Grammar - "respectively" is used incorrectly in this context.
(Changes in the text: We have modified our text as advised in the revised manuscript. (see Pages 7, lines 291)

Line 168: Sentence is grammatically incorrect.
(Changes in the text: We have modified our text as advised in the revised manuscript. (see Pages 8, lines 304)

Line 172: The citation for CAT12 was not included and should be added:
Gaser C, Dahnke R, Thompson PM, Kurth F, Luders E, Alzheimer's Disease Neuroimaging Initiative. A Computational Anatomy Toolbox for the Analysis of Structural MRI Data. bioRxiv. doi: https://doi.org/10.1101/2022.06.11.495736
(Thank you for your thoughtful presentation and references, which have been added to the appropriate places. Changes in the text: We have modified our text as advised in the revised manuscript. (see Pages 8, lines 308)

Line 186: Sentence is grammatically incorrect and missing a verb.
Changes in the text: We have modified our text as advised in the revised manuscript. (see Pages 9, lines 341-342)
Line 187 and line 195: This information is repeated in both paragraphs. It should be condensed for clarity.
(Indeed, the expression is duplicated, and we have deleted line 195.) We have modified our text as advised in the revised manuscript. (see Pages 9, lines 355)

Line 195: How were the ROI defined and chosen by the radiologists? What were the criteria?
Reply: In this study, the consistency and reproducibility of radiomics features were assessed by the intra-observer and inter-observer agreement analysis and when the intraclass correlation coefficient (ICC) was greater than or equal to 0.75, it was included in the study. At the same time, the spatial consistency of image segmentation results was assessed by the Dice coefficient.
Changes in the text: We added this in revised manuscripts (see Pages 9, lines 355-363).

Line 197: Grammar – the word "variability" is missing at the end of the sentence.
Reply: Thank you for your thoughtfulness in pointing out that
Changes in the text: We added this in revised manuscripts (see Pages 9, lines 360-361).

Results:
Table 1: Gender is mislabeled as "1 or 2" and should be changed to M or F.
(Thanks for the thoughtful and detailed suggestions, and have revised Table 1.)
Line 252: "Intra-observer" should not be capitalized.

(Thank you, it has been modified.)
Lines 256-259: Sentence meaning is unclear due to grammatical errors.
(Thanks for pointing this out, changes have been made to make it more logical. We have modified our text as advised in the revised manuscript. (see Pages 12, lines472-480))
Lines 262-263: “Significance” should not be capitalized, and this sentence is poorly written.
Reply: Thank you for pointing this out, the content has been revised and adjusted. We have modified our text as advised in the revised manuscript. (see Pages 12, lines479)
Table 2: The definition of “Image type” should be provided in the caption and text and abbreviations in the table should each be spelled out in the caption.
Reply: We have modified our text as advised in the revised manuscript. Abbreviations are placed below the table. (Page12, line 481)
Lines 276-278: Sentence structure needs to be improved.
We have modified our text as advised in the revised manuscript. (see Pages 13, lines497-509).
Table 3: “The cohorts” column is implied in this study and can be removed from the table. N values for the training and testing sets should be added to the table.
Reply: We have modified the table 3 as advised in the revised manuscript.

In the results, the purpose of the models/features to identify those associated with TOF should be more explicitly stated. How were features chosen to check correlation with neurodevelopmental features?
Reply: Thank you for your well-considered advice. We added this in revised manuscripts (see Pages 13-14, lines 519-638;640-663).
Figure 4: The caption reads “AUC merger” instead of “AUC fusion” and should be fixed for consistency.
Reply: Thank you for your valuable suggestions. Based on your suggestion we have modified it in the figure and presented it uniformly throughout the text.
Figure 5: The font size should be increased for readability. Each abbreviation should be spelled out in the figure caption.
Reply : Thank you for your suggestion, we have made changes to Figure 5.
Lines 292-311: This paragraph is difficult to understand without more information regarding the radiomics features. The specific features should be explained in regard to how they are related to MRI, and abbreviations should be spelled out if not previously defined in the text for the reader. Covariates related to child neurodevelopment should also be described. Methods pertaining to this statistical analysis should be added to the methods section.
Reply: (1) We have modified our text as advised in the revised manuscript. (see 13-14, lines 519-638;640-663). A separate section has been added to the Methods section in order to elaborate on the statistical methods used.
(2) Thank you for your valuable suggestion and we strongly agree on the importance of controlling for covariates, so I further performed multiple linear regression analyses of radiomics characteristics with scale scores and clinical indicators, and the results showed that after correcting for age and gender, there was still a significant negative correlation between Wavelet-HHL- glszm- SALGLE and FSIQ. We summarized the results in Table S3 in the Supplementary Material. The other variables did not show significant correlations after correction, which may be related to the small sample size of our own study. Subsequently, we

will continue to follow up and increase the sample size to increase the stability of the study.

We added this in revised manuscripts (see Pages 15, lines 665-669;Page 18,inles 849-853).

Table S1.

Formatting should be improved, and abbreviations spelled out when possible. Authors should be cautious to use statements implying causation from correlative results.

Reply : Thank you for your careful consideration, we are aware of the need for further consideration and validation of studies with small sample sizes, as well as that the clinical interpretability of radiomics features has always been a challenge, so we have modified Table S1 according to your suggestion, using the conservative term “may” “appeared to”. We hope that this method will be widely studied and more scholars will verify the clinical interpretability and feasibility of such quantitative features.

Discussion:

Line 314: “Age” should not be capitalized in within the sentence.

Reply: Thank you. We have modified our text as advised in the revised manuscript.

Lines 315-17: Grammar - this sentence is difficult to understand and should be rewritten.

Reply: We have modified our text as advised in the revised manuscript. (see Pages 15, lines673-676).

Line 328-336: These results were not described in the text within the results section and should be added (in the results text in addition to the table).

Reply : We added this in revised manuscripts (see Pages 12, lines 470-471;page 11,lines 447-448)

Line 335: References/citations to the specific studies associated with those “similar to previous studies on brain development by this method” should be added.

Reply: This sentence duplicates the expression in the next paragraph(“The prominent features in this study include first-order features and texture features from wavelet transform, which are similar to the previous findings of evaluating neurodevelopment (24,46)”), which has been deleted from the manuscript. (see Pages 16, lines757-759)

Line 335-336: Sentences should not start with “And”. The wording “proves” should be replaced by “suggests” as this is not a definitively proven result given the small sample size of the study.

Reply: We have modified our text as advised in the revised manuscript. This sentence is repeated and deleted. (see Pages 15, line690)

Lines 345-347: This statement is difficult to understand and should be clarified.

Changes in the text: We have modified our text as advised in the revised manuscript. (see Pages 15-16, lines692-757)

Line 352: Spell out the abbreviation OS as it was not previously defined.

Thank you for pointing this out, after consideration this part of the citation has been replaced with an updated study.

Lines 354-363: This sentence should be re-written in the context of the discussion as opposed to summarizing each study separately. Outside of solely describing other studies which used radiomics features to correlate with other diseases, it is unclear the connection between the studies and the results presented in this manuscript.

Reply: Thank you for pointing out. Some of the salient features explored in previous studies on radiomics or neurodevelopmental-related studies are listed here as also containing similar textural features found in this study, as an illustration of the potential that the significantly different textural features found in this study may have in contributing to neurodevelopmental abnormalities

Line 368: It is not clear how morphologic MRI signatures in relation to neurodevelopment were studied in this manuscript. The last paragraph of the results section only described radiomic features which were analyzed with clinical features and neurodevelopmental scores. As neurodevelopmental scores were not included as outcomes in the models, authors should reword this sentence to avoid using "predicting". The authors should clarify how specific features were chosen to study correlation with neurodevelopmental scores.

Reply: Thank you for pointing out. We have modified our text as advised in the revised manuscript.

(1) Thank you very much for your valuable suggestion, we have added the correlation analysis of structural information and presented it in a clear and concise table. (see Pages13-14, lines521-638)

(2) Thanks for your careful reminding, we use "explore" instead of "predict".

Changes in the text: We have modified our text as advised in the revised manuscript. (see Pages 16, lines783)

(3) We first assessed based on univariate analysis that the features with significant differences in morphological and radiomics features in the TOF group compared with the normal group were screened out. Since radiomics are large-throughput high-dimensional features, the features after univariate analysis are selected again by LASSO regression to obtain the most significant and optimal features. Further correlation analysis with neurodevelopmental indicators is performed to quantify the significance of these differential imaging features.

We added methods for obtaining clinical and scale indicators in the revised manuscript. (pages 9-10, lines366-399)