

Dear author,

Thank you for considering publication in the journal and for the opportunity to review your study.

The study explores the development of a DL model to aid in the diagnosis of cerebrospinal fluid (CSF) leak in patients with suspected spontaneous intracranial hypotension (SIH) using brain magnetic resonance imaging (MRI).

Title: The title is clear and informative, providing a concise overview of the paper's content.

Abstract: The abstract offers a good summary of the paper's background, methods, results, and conclusions.

It effectively conveys the purpose of the study and its findings.

Introduction: The introduction effectively summarizes the existing diagnostic methods and highlights the research gap the study aims to address.

My suggestions:

- Although it is common in ML and DL to use the term "validation" to refer to various stages of model evaluation, including internal validation. The term "validate" in developing a DL algorithm may be misleading if the study has limitations, such as a small dataset. The word "validate" often implies external or generalizability validation, which might not be achievable with a small dataset. It would be a good idea to re-write the aim statement.
- Clearly emphasize why a DL Model should be necessary. Explain the potential clinical and patient care benefits of a more accurate method for diagnosing CSF leak.
- Ensure that the introduction flows smoothly from one point to the next and that each paragraph naturally leads to the next. It will make it easier for the reader to follow your argument.

Methodology:

Why was this specific technique and approach chosen? It can help readers understand the reasoning behind your methods.

In the section related to dataset partitioning, there is a need for more clarity regarding the precise method employed to segregate the dataset into distinct training and testing subsets.

Splitting the data into training and testing datasets is fundamental in the ML and DL model development process. It allows you to train the model on one data portion and evaluate its performance on another unseen portion.

While k-fold cross-validation is valuable for assessing model performance and robustness, it is typically used within a more comprehensive data-splitting strategy. The data must be initially divided into training and testing sets, with cross-validation applied to the training set.

This setup ensures that the model is assessed on a truly separate testing dataset, offering a more reliable estimate of its performance in real-world applications.

If the decision to directly partition the dataset into five subsets has been made, I would appreciate a comprehensive explanation elucidating the rationale behind this choice.

Additionally, I am keen to discern the methodology employed for testing the model after its

training. Clarity in this regard would augment understanding of the study's internal validity.

Results: It could be interesting to contextualize the findings by comparing the model's performance with established diagnostic criteria for CSF leak.

The regions highlighted in Figure 3, which significantly influence the model's predictions, are surprising. I'm keen to understand whether the authors can offer insights or explanations to substantiate these intriguing findings. Enhancing this section involves offering a more concise elucidation of the pivotal regions discerned in the occlusion map. Articulate their anatomical significance and explain how the model employs these regions to inform decision-making.

Discussion and Conclusion:

In your statement, you mentioned, "Most notably, the algorithm was trained and validated using data from a single institution, albeit using multiple MRI scanner models and protocols."

The term "validated" may be inaccurate if the model was not tested on unseen data. It is crucial to clarify how the dataset was partitioned into training and testing subsets. Even with the provided explanation, it might be more accurate to refer to this as "internal validity."

There is more potential limitation you need to consider:

- The study has a retrospective design, which introduces the potential for selection bias and data quality issues (scanner type, imaging protocols, and artifacts). Prospective studies are often considered more robust.
- The model's performance should ideally be validated on an independent external dataset to assess its real-world effectiveness. The absence of external validation limits confidence in its performance.

I suggest including a dedicated section for succinct and definitive conclusions. This would encapsulate the study's key findings and their broader implications, enhancing the paper's overall clarity and impact.

These observations can enhance the depth and clarity of your study.