

Changes after review

Reviewer 1

- **R1.1 – Hyperband clarification:**

Clarified that validation loss is used for configuration selection in Hyperband; noted that no explicit patience mechanism is used. This is now explained in the Methods section.

- **R1.2 – Ensembling motivation and rationale:**

Added explanation and motivation for the ensembling strategy, including a reference in the ensembling subsection of the Methods (Section 2.3).

- **R1.4 – Baseline computation time:**

Improved the explanation of baseline computational cost.

- **R1.7 – Number of dataset samples in Table 1:**

Added number of samples (Ns) for each dataset in Table 1.

Reviewer 2

- **R2.1a – Literature review:**

Added additional references and expanded the discussion in the Introduction section to better reflect recent developments in AutoML for medical image analysis.

- **R2.1b – Choice of DenseNet:**

Added a reference supporting the use of DenseNet for medical image classification.

- **R2.2 – Loss function and learning rate:**

Clarified the motivation for using cross-entropy loss; also added the motivation for excluding the learning rate from the search space and cited relevant literature supporting this design choice

- **R2.3 – Tumor-location-dependent cropping:**

Specified that tumor-location-dependent cropping was applied only to training data, not to testing data.

- **R2.5 – DenseNet training:**

Clarified that DenseNet-121 was trained from scratch.

Reviewer 3

- **R3.5 – Fingerprinting and nnU-Net reference:**

Explained how the fingerprinting concept relates to the current setting and what was inspired by nnU-Net.

- **R3.8 – Learning rate not included in the search space:**

Added a motivation for excluding the learning rate from the search space and cited relevant literature supporting this design choice.

- **R3.9 – Hyperband citation:**

Moved the citation for Hyperband to the first occurrence in the text, ensuring it is properly referenced where it is first mentioned.