**List of changes:**

* We have rechecked the paper to enhance and improve the English and fix any grammatical issues, and then updated the manuscript.
* We have updated the state-of-the-art methods in section 1 by extending our comparisons with more and different recent methods such as transformer-based methods. (we’ve added 3 more studies)
* We have merged figure 1 and 3 into one figure.
* In Section 2.1, we updated the manuscript by adding more detailed information regarding the resulting distribution of samples. We have also reduced redundant text in this section.
* In Section 2.3, we provide further details regarding the feature extractor layers.
* In Section 3, we updated this section in the manuscript to reduce the heavy text about hyperparameter selection and Optuna. We have cited the necessary tools and frameworks utilized in our study, including Optuna for hyperparameter optimization, TensorFlow for model implementation, Adam optimizer for training and scikit learn. These citations provide a foundation for the reproducibility and transparency of our research. We clarified and commented on the results of table 2 and the reason of having lower results in AD vs MCI and MCI vs CN task compared to the AD vs CN task.
* In Section 4, we have updated Table 4 to include comparisons with two additional studies (References 5 and 25) and reordered the paragraphs to ensure consistency throughout the manuscript. We expanded the discussion to provide more detailed comparisons with the other works previously mentioned in the related work section, highlighting how our approach differentiates from these studies. Additionally, we recognized the importance of feature interpretability, identifying it as a limitation of our current study. We outlined our intention to address this in future work to enhance the study's clinical relevance, along with working on the multiclass task.