

Thank you for the opportunity to review this interesting study. The authors proposed the AutoML-Ens by ensembling six ML algorithms to find the best weights of predictors. Also, they considered different ensemble methods including BMA, MEAN, and so on to indicate the superior performance of the proposed method compared to these ensemble methods. In my opinion, the manuscript is suitable for publication in Geoscientific Model Development (GMD), after the authors have addressed the following comments and questions:

**Major comments:**

- 1- Because neural networks are one of the ML techniques and standardization is critical for this model architecture, I'm curious if the authors addressed it in the workflow.
- 2- According to the authors, the type of problem in this study is classification, as stated in line 159, and they utilized least absolute error to identify the ideal model (Line 118), which is used for regression problems (at least as far as I know). Maybe I misunderstood that, could you help explain this to me?
- 3- The authors address the accuracy of the autoML in section 3.1.3, however they don't specify the classes, and I'm curious about the proportion of classes. Is it an imbalance classification problem since the performance metric is easily skewed toward the major class? If so, how did the authors manage this situation?
- 4- I'm curious if the authors evaluated the predictors' correlation, as it is preferable to supply more informative information rather than a larger number of predictors for a machine learning model.
- 5- Generally speaking, the performance of the developed model is assessed based on benchmark. For example, multi-linear regression and logistic regression methods are used for regression and classification problems as baseline, respectively. I would like to see how well your developed model is compared to the baseline.

**Minor comments:**

- 6- Figure 2 shows 47 flux sites, but the boxplots for mean annual temperature and mean annual precipitation show 44 and 42 Flux sites, respectively. Could you please clarify the differences?
- 7- Could you elaborate the machine learning classifier? It is hard for me to follow this term.