

Comment on **essd-2022-312**

Anonymous Referee #2

Referee comment on "*TreeSatAI Benchmark Archive: a multi-sensor, multi-label dataset for tree species classification in remote sensing*" by Steve Ahlswede et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2022-312-RC2>, 2022

The manuscript presents a dataset for species classification and also answers various questions connected to the used technique/data combinations, which is important and interesting. However, I have some doubts about using Sentinel-1 for this purpose, as it was previously shown in many studies that S-1 only marginally improves accuracy. Secondly, when using only Sentinel-2 from the summer season, it is not surprising that classification accuracy improvement is limited. The majority of existing studies highlight the role of multi-temporal information (spring, autumn) from S-2 in species discrimination. You actually mention that finally in the conclusions. Maybe you can think about improving that process in the future, for example with seasonal metrics from S-2. Still, your work is very valuable.

I have some minor comments:

Could you add what is the study area size/examined forest size?

Figure 4, particularly the a) part is difficult to read. And what is shown in the chart, the number of samples?

Some methods used to assess accuracy are not described (or I couldn't find them). Maybe you could just briefly say something about what is weighted/micro, F1, and mAP score?

Maybe you could put the subchapters in chapter 4 in the same order as the questions (or change the order of questions ;))

In lines 308-311 You refer to a study from Immitzer based on World-View data and I'm

not sure if it makes sense as the difference between S-2 and World-view spatial resolution is huge...