Review of "A benchmark deep learning dataset for the classification of supraglacial lake drainage mechanism across the central-west Greenland Ice Sheet"

Rines et al. (2026), ESSD-2026-406

This is a novel and important dataset that the glaciology community could benefit from and it is within ESSD's scope; however, there are some major concerns that could be addressed by the authors.

The authors present a dataset of Sentinel-2 image time series for 1679 supraglacial lakes in the central-west (CW) Greenland drainage basin, covering the 2018 ($n = 679$) and 2019 ($n = 1000$) melt seasons. Each lake is stored as one NetCDF file on a daily 153-day grid at 10 m resolution. Every file holds six reflectance bands, a cloud mask, the static lake outline of Dunmire et al. (2021), a new dynamic water mask computed from NDWI, the scalar water-area series `p_water` of Dunmire et al. (2025), and a drainage mechanism label assigned by a human expert together with a soft-probability vector. Lakes are classified by drainage mechanism rather than drainage rate, into five classes: no drainage, hydrofracture, moulin drainage, lateral drainage and crevasse drainage. A ConvLSTM baseline classifier is also trained and evaluated under two splits, one combining both years and one training on 2019 and testing on 2018.

General comments

1. It is not clear if the scalar `p_water` stream is needed as part of the input data. Since the Dunmire et al. inventory covers only 2018 and 2019, the baseline cannot be run on any other melt seasons. I would suggest the authors run an ablation with `p_water` removed, to see how much it contributes to the model's performance. If it does contribute, the $\mathrm{NDWI}_{\mathrm{ice}}$ mask the authors create in Sect. 2.3 could be tested as an input channel in its place, which would also remove the dependence on an external dataset.

   Also, the static `lake_boundary` mask is loaded as a fourth input channel and broadcast across all 153 timesteps, but is then not consumed by the convolutional encoder (Lines 512–514). What does including it achieve?

2. Lakes form in the same basins year after year, so many of the 679 CW2018 lakes are probably the same places as CW2019 lakes. The combined split pools everything randomly (Line 455), which means the same lake can end up in both training and test, with the same location, the same lake boundary raster and nearly the same background imagery. The cross-year split has the same issue, since those lakes appear in the 2019 training set and again in the 2018 test set. I would ask for the number of lakes that appear in both years, and for a split that keeps repeated lakes and geographically close lakes on the same side rather than dividing them between training and test.

3. I am not sure the comparison against Dunmire et al. (2025) in Sect. 6, Lines 610–621, works as an external evaluation. The inventory is not an independent dataset here. It defines the lake population, provides the static boundaries, and supplies the `p_water` series that the baseline uses as one of its two inputs. So, the comparison is between a model trained partly on Dunmire's own product and Dunmire's classifier, scored against labels that only one of the two methods has seen.

Specific comments

Introduction. Line 92 says cooler and warmer melt seasons, but no reference period is given.

Sect. 2.1.

- The Dunmire et al. (2021) versus (2025) citations are inconsistent. Lines 134–135 say the scalar time series is from Dunmire (2021) and the boundary masks are from Dunmire (2025), but it is the opposite in Lines 174–177. Lines 147–148 cite Dunmire (2025) for both.
- Figure 1(a) would be more informative if the outlines of all drainage basins were shown, rather than the CW basin alone. The lakes in the other basins are a little bit difficult to distinguish from exposed rock at this scale and I think a color with more contrast would help.
- Lines 158–161 state that six of the thirteen Sentinel-2 bands are used, but only five are then listed.

Sect. 2.3. Lines 299–300: please rename NDWI to $\mathrm{NDWI}_{\mathrm{ice}}$.

Sect. 2.4.2. Were there lakes that drained more than once during a melt season, for example draining and then refilling and draining again? The taxonomy says the classes are mutually exclusive "at the level of their dominant drainage mechanism" (Line 344), but "dominant" is not defined and the labeling protocol gives no rule for these cases. If a lake drains laterally in June and then by hydrofracture in August, which label does it receive, and how many lakes were affected?

Sect. 2.5. Line 450 states that MD (16.3%) is the smallest class in 2019, but it is ND (13.7%). The comparison "approximately double the size of the smallest class" needs rewording after changing the smallest class in 2019.

Minor and technical corrections

- Line 155 "Source et al., 2022" and reference list: "Source, M. O., McFarland, M., . . . ", please correct the author list.
- Lines 301 and 846: "Kang Yang and Smith, L. C." should be "Yang, K. and Smith, L. C."
- Lines 410–411: `HD` should be `ND` in both probability vectors.
- Line 499: "released a part of a companion paper" → "as part of".
- Fig. 4: the caption text appears out of order in the PDF.