

Dear Reviewer 3,

We thank the reviewer for the positive feedback regarding the novelty and importance of our work. We also appreciate the reviewers comments and suggestions. We have taken into account each of these comments (black text, below), and provide our detailed responses to each (blue text, below). One overarching piece of feedback from all reviewers was to focus the manuscript more closely on the dataset, and only use the baseline classifier as a demonstration of the dataset, which we have done. Regarding this demonstration, in light of the comments by the reviewer regarding potential data leakage in the previously-reported baseline demonstration of the classifier in the original manuscript, we have opted for a train/val/test split fully within the 2019 meltseason. This serves two purposes: (1) as a basic demonstration of the dataset in a deep learning model as opposed to over-describing the implications of the former cross-year and combined splits and (2) as a step to reduce data leakage from the fact that, as the reviewer points out, lakes tend to appear in the same location year-on-year. We believe that the reviewer's suggestions have improved our manuscript, as detailed below.

Sincerely,
Joshua Rines, on behalf of all coauthors

General comments:

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 `NDWI_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.

Thank you for this keen observation. Out of interest, we ran an ablation with and without `p_water` and report the results in Appendix B. Under the baseline model, the `p_water` stream does not seem to improve the performance of the model. We have noted this, but refrain from further investigation in the name of keeping the manuscript focused on the dataset and not the classifier or its optimization, though we are keen to follow up on this work. In addition, running the classifier with `NDWI_ice` is a future step and saved for future work. Indeed, another consideration for future work is to extend the dataset beyond the 2018 and 2019 years, which would allow us to have `p_water` for any year, and we believe that including it in the dataset now is advantageous, as its replication is a natural component of an extension of the 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?

The `lake_boundary` mask is now consumed by the encoder in the revised baseline model.

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.

We appreciate your point, and recognize the importance of protecting against data leakage during training. In the effort to stay focused on this being a data descriptor, we have scaled back the description of the model, and in doing so have restricted ourselves to a train/val/test split on the 2019 year, leaving the 2018 year untouched. This addresses some of the concerns raised here, such as the same lake appearing in multiple years. We recognize that it does not address the “geographically close lakes” within the same year, such as lakes that might appear in each others’ 512x512 tiles, but presently accept this as a modeling limitation.

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.

We have removed this comparison in favor of a simplified “Technical validation” section, which demonstrates the dataset is structured correctly and directly ingestable into a deep learning model.

Specific comments:

(Introduction)

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

We have updated the text to be clear about what we mean by “cooler and warmer”, following Dunmire’s work (L44).

(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.

The static boundary polygons are from Dunmire et al. 2021 and the scalar p_water series are from Dunmire et al. 2025, both distributed with the 2025 inventory. We have corrected the reversed attribution in Section 2 and made the distinction clear and consistent in Section 2.1, Figure 2 caption, and the data availability statement. Thank you.

(Sect. 2.1)

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.

Thank you for the suggestion. We have outlined the drainage basins and delineated the lakes in the other basins more clearly against the exposed rock, as suggested, with a different color contrast.

(Sect. 2.1)

Lines 158–161 state that six of the thirteen Sentinel-2 bands are used, but only five are then listed.

Thank you. This has been rectified by listing B12 as well (L148).

(Sect. 2.3)

Lines 299–300: please rename NDWI to NDWI_ice.

Done.

(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?

Thanks for the question. Yes, interestingly enough, there are some lakes which drain multiple times during a melt season, in both the ways you describe. To address your latter point, we have added a line to the labeling protocol (L564) as well as defining what we mean by "dominant" and adding this exact example to L329 for clarity. For your former question, there are some lakes that appear to drain via hydrofracture multiple times, and we have taken an interest in these phenomena for future study.

(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.

We thank the reviewer for bringing this to our attention. Indeed the smallest class in 2019 is ND (13.7%), not MD (16.3%). Table 1 and Fig 1 both correctly report these numbers, so the error was confined to that sentence. The paragraph has been removed during the restructuring of Section 2, as the split definitions that followed it now sit in the "Technical validation" section, a

section title that was suggested by Reviewer 2. We have restored the class-distribution table. LD is the most prevalent class in both melt seasons at roughly three times the size of the smallest class in 2018 (MD) and two and half times the size of the smallest class in 2019 (ND).

Minor comments:

Line 155 "Source et al., 2022" and reference list: "Source, M. O., McFarland, M., . . . ", please correct the author list.

Done. Thank you for catching this.

Lines 301 and 846: "Kang Yang and Smith, L. C." should be "Yang, K. and Smith, L. C."

Done. Thank you.

Lines 410–411: HD should be ND in both probability vectors.

Thank you for catching this typo. We have fixed it.

Line 499: "released a part of a companion paper" → "as part of".

Thank you. We have removed this sentence entirely as it was deemed unnecessary, as flagged by other reviewers.

Fig. 4: the caption text appears out of order in the PDF.

We have moved the figures to their proper place in the PDF, but are unsure what the reviewer is referring to regarding the caption text being out of order. If further clarification is provided we will be happy to make further adjustments.