Referee 2 comments and replies

Dear Referee 2,

We greatly appreciate the time and effort that you have dedicating to evaluating our work. We have found the feedback insightful and have carefully considered all the comments, which have helped us significantly improve the clarity, structure and quality of our paper.

Reviewer comments/suggestions are in *black*; our responses are in red.

Morgan and colleagues present a workflow for extracting glacial lake delineations using an automated OBIA/ML classification method. The paper provides a thorough overview of classification and segmentation parameter optimisation, feature selection, computational requirements, and a comparison of using 1 vs. 5 Landsat images for training the classifier. This workflow is a welcome contribution in a field increasingly dominated by deep learning approaches. By demonstrating that a more computationally efficient and accessible method can still achieve competitive accuracy, the authors make a convincing case that there remains an important place for simpler classification methods.

One of the main findings that a single training image can achieve accurate classification of a time series is intriguing and very exciting, but it does raise a methodological concern that is worth addressing. As described in Section 2, the same image was used to collect both training and testing datasets. The resulting accuracies were then compared against those from a classifier trained and tested on five images (if I understood correctly). In this case, I would assume the higher accuracy of the single-image method may simply reflect reduced dataset variety rather than superior classification performance.

I wish the authors would acknowledge this issue and discuss its potential effect on the results.

This is a valid point that was also raised by Referee 1. To address this, we have re-done this portion of the analysis, selecting four images for validation (one each from LT05, LE07, LC08, and LC09) that were not used to train the classifier. We were unable to select another LT04 image on account of the lack of suitable alternatives for the study area; however, because of the similarities between LT04 and LT05 we do not feel this will impact the conclusions drawn. The results of this portion are summarised in the Tables below and will be included in the updated manuscript.

Table 1. Images used to train the Classifier for the multi-sensor comparison.

Landsat satellite	Landsat ID	Path / Row	Cloud Cover (%)
LT04	LT04_L1TP_075090_19901230_20200915_02_T1	75 / 90	9
LT05	LT05_L1TP_075090_20070204_20200831_02_T1	75 / 90	2
LE07	LE07_L1TP_075090_20030201_20200916_02_T1	75 / 90	5
LC08	LC08_L1TP_075090_20220213_20220222_02_T1	75 / 90	5
LC09	LC09_L1TP_075090_20230224_20230308_02_T1	75 / 90	4.6

Table 2. Images used to validate the Classifier for the multi-sensor comparison.

Landsat satellite	Landsat ID	Path / Row	Cloud Cover (%)
LT04	LT04_L1TP_075090_19901230_20200915_02_T1	75 / 90	9
LT05	LT05_L1TP_075090_20070204_20200831_02_T1	75 / 90	10
LE07	LE07_L1TP_075090_20030201_20200916_02_T1	75 / 90	9
LC08	LC08_L1TP_075090_20170215_20200905_02_T1	75 / 90	5.9
LC09	LC09_L1TP_075090_20241227_20241228_02_T1	75 / 90	11.5

Table 3. Validation metrics for the classifier trained on the single image OBIA classifier.

Satellite ID	Overall accuracy (%)	$\hat{\kappa}$ statistic	PA Water (%)	UA Water (%)	$\hat{K}$ Water
Multi-class classification					
LC09-2024-12-27	95.5	0.94	99.0	99.8	0.99
LC08-2017-02-15	93.5	0.92	93.5	93.0	0.91
LT05-2003-12-10	97.0	0.95	97.1	97.5	0.97
LE07-2001-01-10	93.4	0.91	93.9	93.2	0.92
LT04-1990-12-30	93.3	0.91	93.3	93.3	0.91
Binary classification					
LC09-2024-12-27	98.5	0.98	99.8	97.9	0.97
LC08-2017-02-15	98.1	0.97	98.6	98.0	0.97
LT05-2003-12-10	98.5	0.98	99.7	98.3	0.98
LE07-2001-01-10	98.9	0.98	98.6	99.6	0.98
LT04-1990-12-30	98.4	0.98	98.7	98.9	0.98

Table 4. Validation metrics for the classifier trained on individual OBIA classifiers per Landsat sensor.

Satellite ID	Overall accuracy (%)	$\hat{\kappa}$ statistic	PA Water (%)	UA Water (%)	$\hat{K}$ Water
Multi-class OBIA classification					
LC09-2024-12-27	95.4	0.94	99.0	99.8	0.99
LC08-2017-02-15	94.0	0.92	98.0	98.5	0.97
LT05-2003-12-10	95.6	0.94	99.8	99.7	0.99
LE07-2001-01-10	96.4	0.95	99.7	98.6	0.98
LT04-1990-12-30	94.6	0.93	99.1	98.0	0.97
Binary OBIA classification					
LC09-2024-12-27	98.5	0.98	99.9	97.8	0.97
LC08-2017-02-15	98.3	0.97	99.8	98.6	0.97
LT05-2003-12-10	98.1	0.97	98.6	96.8	0.96
LE07-2001-01-10	98.2	0.97	99.8	97.8	0.97
LT04-1990-12-30	98.2	0.97	100	95.8	0.93

Table 5. Validation metrics for the classifier trained on multiple images (one per Landsat sensor) to a singular OBIA classifier.

Satellite ID	Overall accuracy (%)	$\hat{\kappa}$ statistic	PA Water (%)	UA Water (%)	$\hat{K}$ Water
Multi-class OBIA classification					
LC09-2024-12-27	92.2	0.89	96.5	97.5	0.95
LC08-2017-02-15	91.2	0.88	99.4	97.5	0.95
LT05-2003-12-10	88.3	0.84	87.0	94.0	0.90
LE07-2001-01-10	87.2	0.83	86.1	94.1	0.89
LT04-1990-12-30	84.9	0.8	84.4	92.2	0.9
Binary OBIA classification					
LC09-2024-12-27	98.3	0.97	99.8	98.8	0.97
LC08-2017-02-15	98.3	0.97	98.0	99.8	0.99
LT05-2003-12-10	98.7	0.98	96.7	98.8	0.98
LE07-2001-01-10	97.6	0.97	96.7	98.8	0.97
LT04-1990-12-30	96.6	0.95	95.6	95.9	0.95

Another unexplored opportunity is applying the classifier to a longer or more frequent time series, or to a new region. With the imagery archive available in GEE, extending the study in this way could be relatively straightforward and would show how well the classifier generalises beyond the study area.

To address this, we have exported the classifier as an asset available through GEE. We have also applied the trained classifier to a single image in Patagonia (LC08-2021-03-03). The results of this can be seen below.

Table 6. Validation metrics for the Patagonia image classified using the classifier trained on an image of New Zealand.

Satellite ID	Overall accuracy (%)	$\hat{\kappa}$ statistic	PA Water (%)	UA Water (%)	$\hat{K}$ Water
Multi-class OBIA classification					
LC08-2021-03-03	92.8	0.91	98.9	97	0.96
Binary OBIA classification					
LC09-2023-03-03	98.8	0.97	97.9	100	1

Line 36-37. Missing the word "Index" in Normalized Difference Snow.

Done.

Line 80-84 I would add the reference where the study is first mentioned.

Done.

Line 115 Add space in 23km for consistency.

Done.

Line 138 Add space in timeseries.

Done.

Line 139-140 Using the same image for both training and testing could risk data leakage and the two datasets being too similar.

We thank the reviewer for highlighting this point. Please see the above corrections (**Table 1**-**Table 5**) that address this comment.

Line 184 Classification and segmentation experiments sound vague to me. Maybe consider switching to "hyperparameter tuning" or "parameter optimisation" as they would fit your description in the paragraph.

We have gone with "parameter optimisation" or similar language throughout the manuscript, in response to this and feedback from Referee 1.

Line 199-201 I found this difficult to understand. Maybe simplify to something like: In this study, OBIA was used for both multi-class classification of all feature types and binary classification of water versus non-water features.

Thank you for the suggestion – we have modified this sentence to read "In this study, OBIA was used for both multi-class classification of all feature types and binary classification of water against non-water features."

Line 202 I would consider switching the title to a clearer "feature selection".

We agree with this and have corrected this by changing the title to "The effect of feature selection on OBIA classification".

Line 208-210 Longer explanation of commonly known NDVI not necessary, especially when other indices didn't have one.

Rewritten as follows: "The MNDWI, NDVI, hillshade and DEM slope were added as input features for the classifier."

Line 229 I don't think it is necessary to explain QGIS for this journal's audience.

Sentence/explanation removed.

Line 352-354 If I understood correctly, all LS images cover the same area, which makes your results highly region-specific. How can you be confident that the OBIA classifier is applicable across the Southern Alps if it has not been tested in other areas? I think this is a missed opportunity to evaluate the generalisability of your classifier.

As detailed above, we have now evaluated the classifier's performance on an image over Patagonia to help address this concern.

Line 359-366 Since you used the same image for both training and testing, the higher accuracy in the single-image method is expected. For the five-image classifier, was the data collected from all five images for both training and testing? If so, the slightly greater variety introduced might explain the difference.

As detailed above, we have revised this to ensure our validation images are separate from the training images.

Line 432 10m > 10 m for consistency.

Done.

Line 468 The word "Classifier" is capitalized, whereas all other instances using lowercase.

Adopted "classifier" throughout for consistency.

Line 518-519 Did you apply the single-image classifier to the whole collection, and if so, what were the results? Also, does the 'whole collection' refer to the full time series for your specific study area of a wider selection of areas?

The goal of this study was to evaluate the impact of varying classification parameters on the classification result; in a second study, we have applied the classifier to the whole collection of Landsat images over the Southern Alps as part of a larger investigation into the glaciological aspects of glacial lake changes in the study region. Here, "the whole collection" refers to all available Landsat 4, 5, 7, 8, and 9 imageries spanning the entire operational lifetime of these sensors across our study region.

Line 602 I had some confusion in trying to locate the supporting scripts and main scripts. Maybe doing some explanation on how access these would be beneficial for other users.

I agree with this comment and with a similar comment by another referee I will ensure that I do a video tutorial online that details how to access the data within GitHub and then another video tutorial on applying our approach in Google Earth Engine. We will also export the trained classifier in Google Earth Engine and provide an in-depth explanation on how you would be able to apply this yourself.