RC2 Comments

Dear authors,

your study presents a promising approach to the integration of UAV-derived multispectral data with various machine learning classifiers for vegetation mapping. However, several significant revisions are necessary before your manuscript can be considered for publication, it requires significant changes before it can be resubmitted for further review as a new manuscript. Below, you can find my general comments.

Your published dataset of the previous paper (*https://www.mdpi.com/2072-4292/16/5/840*) was partially used as "ground truth" for a different spectral data in the analysis. While the reuse of your data could be understandable, this manuscript does not introduce substantial new techniques or research novelty. This is highlighted also from the obvious findings, like that the "*spectrally more distinct vegetation types with lower spectral and structural variation showed the greatest proportion of correct classifications in the confusion matrices of the three classifiers*", where the best results are obtained with *Fuscospora* dense forest and sparsely vegetated scree.

We understand this concern and will clearly state how the current analysis differs from the prior study. This manuscript shifts focus from object-based classification to spectral index-based ML classification. Importantly, we analyse feature importance across three classifiers and investigate model performance differences for fine-scale vegetation classes—a novel contribution within this landscape context. We will clarify this distinction and cite our previous work explicitly in the introduction and methods

I recommend a substantial revision of the abstract, as it currently lacks key information regarding the validation or ground-truth data used, the characteristics and spatial extent of the study area, as well as the data collection date. The structure and order of the content in the abstract need to be completely revised. It is widely acknowledged that very few treelines worldwide are entirely unaffected by human influence. This should be acknowledged in the introduction, with consideration given to how such anthropogenic factors—alongside climate—should be incorporated into the modelling framework.

We agree and will revise the abstract to include the 4 ha area, 2018/2019 flight season, use of prior field-based segmentation for ground truth, and summary accuracy/Kappa values. This is an important point, and we will incorporate this into the introduction by noting that while the study area is remote, historical land use (e.g., grazing) or invasive species pressure may shape vegetation dynamics. We will also discuss how such factors could be integrated into future modelling.

Although it is evident that previously published data were used as ground-truth, the manuscript does not clearly explain how these data were incorporated into the training and/or validation phases. While 600 ground control points are mentioned—points that appear to be notably unbalanced in terms of vegetation cover—it remains unclear how exactly these points were utilized in the analysis. Moreover, it is unclear which steps are part of the previous work and which are new (lines 185-196 and lines 212-225). I recommend that the authors revise this section to enhance clarity and prevent any confusion for the reader. Although reference is made to a previously published paper, it would be advisable to clearly specify in the M&M section the spectral bands available from the multispectral sensor mounted on the UAV.

In a revision, we would explicitly delineate which steps were reused (segmentation, field data acquisition) and which were novel (index derivation, ML classifier training). The 600 field GPS points were used both to validate the original segmentation and to label training polygons for ML classification in this study. This distinction will be made clearer in the methods section.

In your previous study, you classified two alpine treeline ecotones in the Canterbury region of New Zealand's South Island with similar vegetation type. Why not use one site for training and one for model testing and/or validation?

We appreciate the suggestion to use one site for training and another for validation. However, the two ecotones—Craigieburn and Lewis Pass—differ substantially in climatic regime and vegetation composition, despite their geographic proximity. Craigieburn is more arid, with cold winters and well-defined snow cover, while Lewis Pass is shaped by higher summer rainfall and supports different dominant vegetation types. These environmental and floristic differences introduce confounding effects that would compromise model transferability without further ecological normalisation. For this reason, we chose to focus on a single ecotone with full field validation and consistent ecological conditions to evaluate classifier performance.

The multivariate analysis could be reduced/deleted to allow more space for the machine learning classifiers, especially since the PCA does not provide much informative value, with 95% of the variance explained by PC1, and possible strong autocorrelations between many of the spectral indices used. In paragraph 3.2, a very expected result of correct classification for the *Fuscospora* forest is shown, and other class-wise percentages are briefly commented on. The sentence "*Classification confidence for the remaining vegetation types was mostly low (< 60%)*" could be further investigated, as the challenges might be to distinguish between similar vegetation types. The most significant part of the paper appears to be the discussion of different algorithms applies (Figure 6 and lines 351–358); however, the manuscript lacks a discussion on how the different approaches or algorithms could be integrated, especially given their distinct and complementary behaviors.

We agree that the PCA provides limited explanatory value in its current form, given the dominance of the first component and the intercorrelation of indices. In a revision, we would condense or relocate this analysis to supplementary material to focus more clearly on the machine learning results. Regarding class-wise performance, we acknowledge that lower accuracy in some vegetation types likely reflects overlapping spectral responses in structurally heterogeneous classes such as scrub and mat-forming vegetation. We will expand our discussion of these results and propose that future efforts could incorporate ensemble models or classifier stacking to leverage the complementary strengths observed across SVM, RF, and XGBoost, as their differing feature priorities suggest integration could improve robustness.

Finally, some aspects mentioned in the discussion seem off-topic: e.g. “a landscape-scale classification of the subalpine can support monitoring the impact of invasive herbivores on these ecosystems, as their grazing pressure threatens both vegetation dynamics and the region’s carbon sequestration potential”, especially considering that the spatial extent of this work cannot be considered sufficient for a 'landscape-scale' study. I suggest modifying the discussion and conclusion sections after a thorough revision of the article.

We acknowledge the reviewer’s concern regarding the spatial extent of the study and agree that care must be taken in using the term “landscape-scale.” However, we consider it justified here given the full altitudinal coverage of the subalpine belt at the site, the high spatial resolution of the UAV data, and the limited accessibility of these environments. Our use of the term reflects ecological representativeness rather than absolute area. Alternatively, we could use a more specific term like ‘ecotone-scale’. That said, we agree that references to herbivore impacts and carbon dynamics are beyond the scope of this analysis and will revise the discussion to limit conclusions to findings directly supported by the data, while moving broader applications to future work.