Responses to Reviewer #3

The authors used ultra-high resolution for vegetation classification and change detection in alpine treeline ecotones, which holds certain scientific value, particularly by focusing on the identification of the krummholz category, a relatively under-researched area. The methods are reasonable and have practical application value in treeline studies. Overall, in the previous round of major revisions, the authors effectively addressed the reviewers' comments with high quality. I recommend accepting the paper after minor revisions.

Specifically, I have the following comments and suggestions:

The authors should address the generalizability of their methodology. This study was conducted in a not very large area, and it is unclear whether the classification process and input feature combination can be applied to larger or other regions. At a larger scale or for remote sensing of alpine treelines in other areas, some concerns may arise, such as whether cloud-free, ultra-high-resolution data can be obtained for most remote mountainous areas. Additionally, the selection of time periods may vary depending on the dominant species in the ecotone. This study area dominated by Juniperus and Abies species, but for widely distributed treeline species like Larix and Betula, the rationale for using autumn data may be questioned. While the authors may not need to add new validation process in this paper, it is recommended to briefly address these points in the discussion.

Response: Thank you for this insightful comment. Specifically, the proposed classification framework combining WorldView-2 ultra-high-resolution imagery and U-Net/RF models can be adapted to other alpine treeline ecotones with cloud-free satellite data available. We also acknowledged that accessibility of imagery and species-specific phenology (e.g., Larix, Betula) may influence the optimal observation period. However, it's worth noting that neither genus, Larix nor Betula, is found in Taiwan's alpine treeline ecotone. We believe this is what makes Taiwan's alpine treeline ecotone unique.

What is the specific purpose of placing the research area indicated by the red marker in the lower-left corner of Figure 1? Why not zoom in on the central part of the map? Alternatively, the label "Mt. Xue main peak" could be reduced in size to minimize unnecessary obstruction.

Response: Thank you for the comments. We have modified Figure 1.

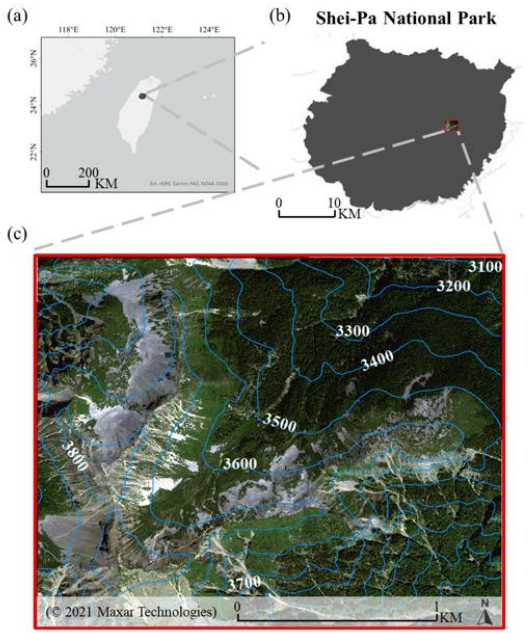

Figure 1. Study area. (a) Geographic location of Shei-Pa National Park in north-central Taiwan. (b) Treeline ecotone study area located in the Xue Mountain glacial cirques within Shei-Pa National Park. (c) WorldView-2 image showing the research area with topographic contours.

In L225, the statement "All classes achieved F1-scores above 0.6" seems somewhat redundant, as a F1-score of 0.6 is not a particularly strong benchmark. Moreover, based on Figure 4, it is clear that most F1-scores are above 0.7, with only one around 0.6. It is recommended to remove this sentence or replace it with the overall average F1-score.

Response: Thank you for the valuable suggestion. The sentence regarding "F1-scores above 0.6" has been deleted, and the section 3.1 and 3.2 has been rewritten.

L251, write out the full name of ATE as "alpine treeline ecotone." "ATE" itself is not a widely used abbreviation.

Response: Thank you for the comment. We have revised the manuscript.

L269-271, The ecological significance reflected in the results can be moved to the discussion section, with relevant citations added to confirm the value of this minor classification accuracy improvement for ecological applications.

Response: Thank you for the comment. The related ecological interpretation has been moved to the Discussion section 4.3 and supported with additional references (L364-368).

> "Although the numerical improvement in overall accuracy appears modest, such enhancement is ecologically meaningful. Even slight gains in classification precision can improve the detection of subtle land cover transitions, particularly the identification of forest expansion boundaries in alpine treeline ecotones. These improvements strengthen the ecological interpretation of spatial change dynamics and provide a more reliable foundation for long-term monitoring (e.g., Bader et al., 2021; Wang et al., 2022)."

The readability of Figure 5 is poor, and the key points are not clear. It is recommended to enlarge the bar chart and highlight the values and ranking of the factors indicating their relative importance. The curve for cumulative model interpretability is not a highlight and does not need to be emphasized. It would be sufficient to label the factors corresponding to the 95% threshold only.

Response: Thank you for the comment. We have modified Figure 5, and the updated version is now presented as Figure 6.

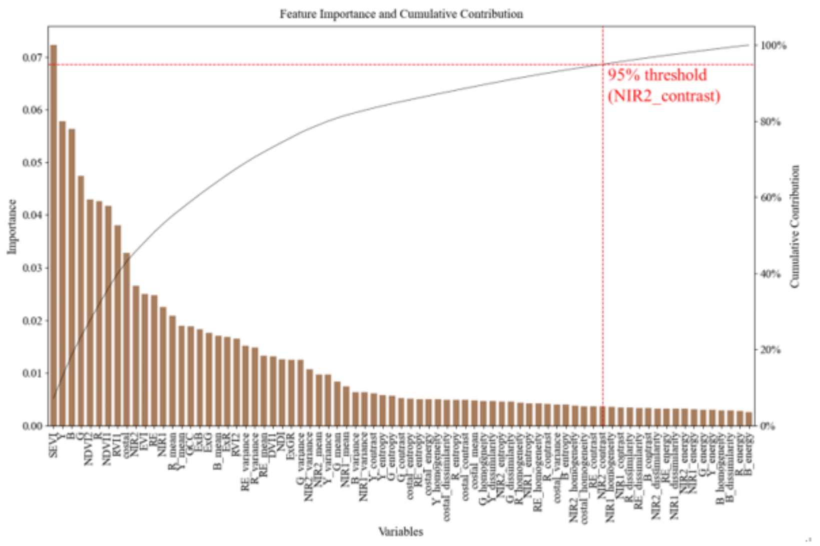

Figure 6. Feature importance ranking derived from the Random Forest model. Features are ranked based on their contribution to classification accuracy, with the top-ranked features including SEVI, Y (yellow), B (blue), G (green), and NDVI2. Most of the top features are spectral bands and vegetation indices, while texture features rank lower.

L291, "expanded by 0.105 km² and was reduced by 0.004 km²": Using km² as the unit makes the values appear insignificant. If the authors intend to convey a significant trend of forest expansion, it is recommended to use hectares instead. Moreover, the unit in Table 8 is also hectares, so it is suggested to standardize the area unit throughout the paper (including the corresponding expressions in the abstract and the other sections).

Response: Thank you for the comment. We have revised the manuscript to standardize the area unit throughout the paper.

In Figures 7 & 8, the "field survey" icon color is not very prominent. Recommended to change the color or add a black border to make it stand out more. Additionally, in Figure 8, it would be better to zoom in, as the pink triangle is hard to find now.

Response: Thank you for the comment. We have modified Figure 7 & 8.

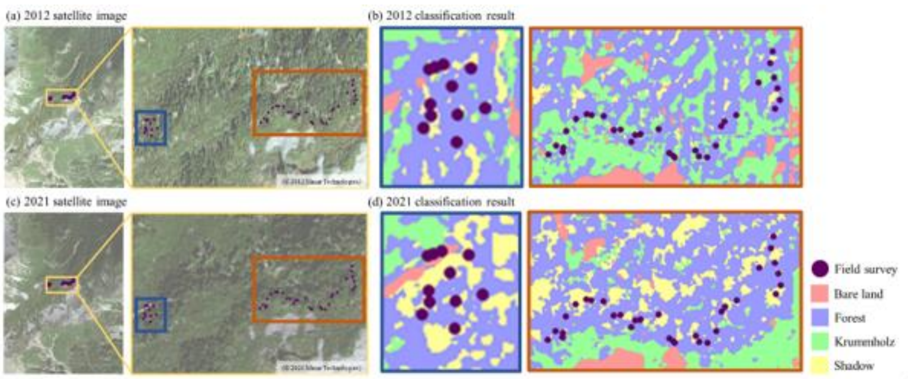

Figure 7 Comparison of satellite imagery and classification results from 2012 and 2021. Panels (a) and (c) show high-resolution satellite images for 2012 and 2021, respectively. Colored boxes in these images indicate the enlarged areas shown in (b) and (d). Panels (b) and (d) present the classification results of the corresponding enlarged regions using a U-Net model trained with 61 selected features. Triangles mark field survey locations.

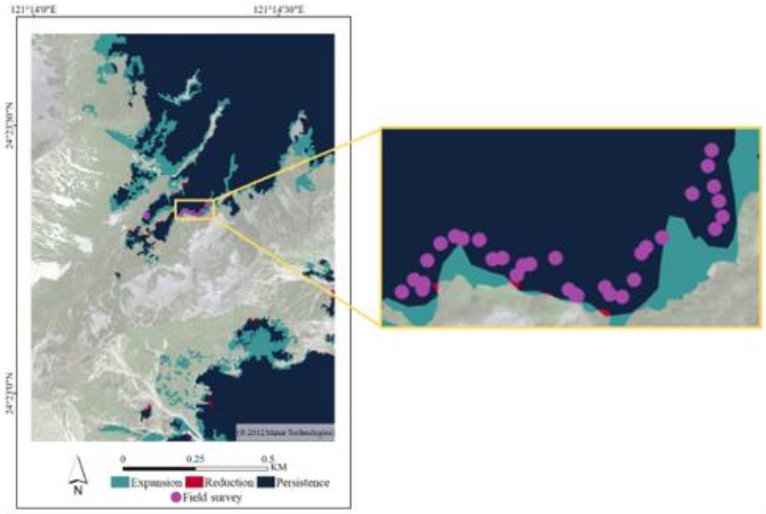

Figure 8. The spatial distribution of ATE area changes from 2012 to 2021. ATE expansion is marked in dark cyan, reduction is marked in dark red, persistence is marked in dark blue, and field survey point in purple.

L364, similarly, provide the full name of "ATE" here.

Response: Thank you for the comment. We have revised the manuscript.

Responses to Reviewer #4

L33, Add reference

Response: Thank you for the comment. We revised the manuscript and added the reference as follows (L33):

> "The island contains more than 200 mountains exceeding 3,000 meters in elevation (Kuo et al., 2022)."

L43-44, Only wildfires? Thre are other natural disturbances potentially influencing such an issue. You should also mention the relevance eof cascading effets between disturbances in this context.

Response: Thank you for the comment. We revised the manuscript and added the reference as follows (L41-45):

> "However, these shifts are also influenced by other drivers, including land-use history, altered disturbance regimes (e.g., fire, landslide, windthrows), herbivory pressure, and species-specific physiological traits. Moreover, cascading effects among these disturbances can further amplify ecological responses and accelerate treeline dynamics (Wang et al., 2016; Johnson et al., 2017; Du et al., 2018; Mohapatra et al., 2019, Stritih et al., 2024; Lu et al., 2025)."

L45-60, This section should be moved to the Discussion, where is generally useful to compare the study with previous researches, specifically focusing on the novelty of your study respect to the available literature.

Response (45-73): Thank you for the comment. Following the reviewer's advice, we have moved the paragraph summarizing previous treeline remote sensing studies from the Introduction to the Discussion and expanded it into a new subsection titled "4.1 Comparison with previous alpine treeline ecotone remote sensing studies." This section now highlights the novelty of our study, emphasizing the integration of ultra-high-resolution WorldView-2 imagery and U-Net for detecting krummholz and forest transitions in a subtropical alpine environment. We revised the manuscript as follows (L307-323):

> "Recent advancements in remote sensing technology have enabled extensive studies on alpine treelines using imagery at various spatial resolutions (Garbarino et al., 2023). For example, Xu et al. (2020) employed Landsat (30 m) data to assess treeline–climate relationships in China, reporting an upward shift of ~50 m per 1°C increase in temperature. At medium to high resolution, Rösch et al. (2022) achieved over 90% classification accuracy for Pinus mugo in the Alps using PlanetScope (3 m) and Sentinel-2 (10 m) data, emphasizing the value of multi-source data fusion. At very high resolution, Terskaia et al. (2020) combined aerial orthophotos (1–2 m) and WorldView-2 imagery (0.5 m) to quantify shrub and tree encroachment in Alaska, detecting substantial vegetation transitions over six decades.
>
> Building on prior work, fine-scale mapping of alpine treeline ecotones (ATEs) remains difficult because transitional vegetation is spatially heterogeneous, often includes stunted or shrubby forms such as krummholz, and exhibits subtle spectral/structural gradients at meter scales (e.g., Bader et al., 2021; Nguyen et al., 2022). Our study uses ultra-high-resolution WorldView-2 imagery (0.4 m) and machine learning workflows to detect fine-scale transitions within the ATE (~400 ha) in Taiwan. Concretely, we

show that integrating spectral bands, vegetation indices, and texture (GLCM) features at sub-meter resolution enables reliable separation of krummholz from closed-canopy forest—an underrepresented class distinction in many alpine studies (cf. Korznikov et al., 2021; Nguyen et al., 2022). This demonstrates the novelty and practical value of combining modern machine-learning segmentation with ultra-high-resolution imagery to fine-scale analyze the alpine treeline ecotone (ATE) in subtropical mountain environments. Related recent work similarly highlights the need for meter-scale approaches to capture ATE patterns and dynamics (Zou et al., 2022; Carrieri et al., 2024)."

L61-73, as above. Here, it should be better to simply adreess the relevance of R, SVM models, as well as U-Net models in this context, of course with references to same recent papers, therefore moving to the aims of your research and respective innovative aspect.

Response: Thank you for the comment. We revised the paragraph to better highlight the study's aims and innovations as follows (L46-67):

"Machine learning is increasingly being combined with high-resolution remote sensing to enhance land-cover and forest-type classification. Among the numerous algorithms, each model has its own strengths. Random forests (RF) and support vector machines (SVM) have gained widespread use due to their robustness and effectiveness in processing multispectral data with limited training samples (Belgiu and Drăguţ, 2016; Jombo et al., 2020; Jackson and Adam, 2021). RF, in particular, exhibits strong interpretability and stability in heterogeneous environments. In contrast, deep learning models such as U-Net demonstrate superior ability to capture both spectral and spatial information, achieving high segmentation accuracy in complex landscapes (Ronneberger et al., 2015; Freudenberg et al., 2019; Wagner et al., 2019). Recent comparative studies further demonstrate that RF and SVM remain reliable and interpretable choices for multispectral classification when training data is limited or imbalanced. At the same time, U-Net and other convolutional neural network (CNN) architectures generally provide superior spatial accuracy and boundary delineation in high-resolution or well-labeled datasets. Furthermore, transferability analysis shows that U-Net models generally have better generalization capabilities in large or heterogeneous regions, while RFs tend to perform more consistently in small sample or sparsely labeled scenarios (Boston et al., 2022; Ge et al., 2021; Nigar et al., 2024).

In Taiwan, many alpine forest studies have been conducted through field surveys using an ecological approach at relatively small spatial scales, focusing on flowering phenology and growth assessment for specific tree species (Chiu et al., 2022; Liao et al., 2023a; Kudo et al., 2024). In recent years, Chung et al. (2021) used Landsat 8 imagery combined with support vector machine (SVM) classification to examine timberline dynamics on Taiwan's highest peak, Yushan, revealing the influence of temperature on timberline shifts. The Xue Mountain, Taiwan's second-highest peak, has also been the subject of long-term ecological monitoring (Chung et al., 2021; Liao et al., 2023b). However, extensive targeting alpine treeline ecotone (ATE) dynamics remains lacking. This study provides the first comprehensive analysis of changes in the ATE landscape in Taiwan's Xue Mountain glacial cirque region. It uses ultra-high-resolution WorldView-2 satellite imagery with Random Forest (RF) and U-Net models. The aim is to quantify spatiotemporal changes between 2012 and 2021."

L88, As done with ATE, explain the term with respective reference.

Response: In this study, krummholz, representing stunted, shrub-like trees typically found at high elevations near the treeline and shaped by wind or snow pressure (Liao et al., 2023a). The definition of krummholz was provided in (L78-79 and L191). Relevant explanation and details can be found in Liao et al., 2023a.

L91-95, Fig. 1: the boundaries of the study area are not clearly shown in the Figure. Is not neither clear the variation in elevations from the DTM.

Please, enlarge the RGB map, show boundaries of the study area, remove the DTM simply defining min and max elevation in the next, and clearly indicate the number of pictures with sequential numbers / letters (a,b,c...1,2,3...), also updating the text and the caption in accordance.

Response: Thank you for the comment. We have modified Figure 1 and revised the manuscript.

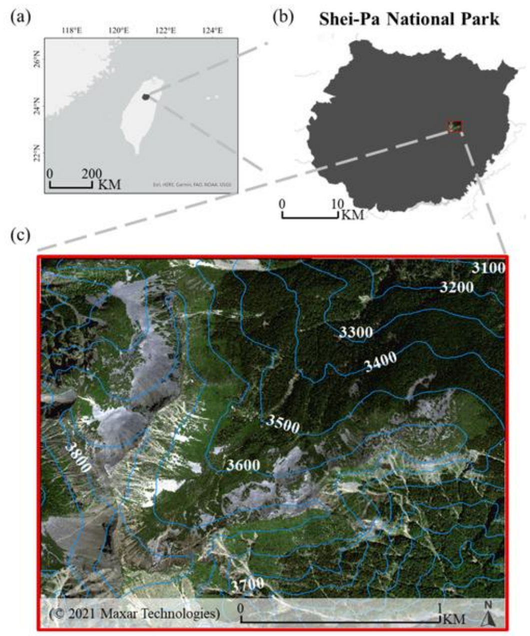

Figure 1. Study area. (a) Geographic location of Shei-Pa National Park in north-central Taiwan. (b) Treeline ecotone study area located in the Xue Mountain glacial cirques within Shei-Pa National Park. (c) WorldView-2 image showing the research area with topographic contours.

L122, Restate better.

Response: Thank you for the comment. The sentence has been revised for clarity as follows (L114):

"Both images were captured in the autumn season when vegetation had entered its dormant phase."

L123, GNSS is the more correct term.

Response: Thank you for the correction. The term "GPS" has been replaced with "Global Navigation Satellite System (GNSS)" throughout the manuscript to ensure technical accuracy. (L115)

L126, Remove the column, not useful.

Response: Thank you for the comment. The "Data quantization (Bits)" column in Table 1 has been removed as suggested.

L128-129, Add reference.

Response: Thank you for the comment. We revised the manuscript and added the reference as follows (L125-126):

> "The reflectance spectrum of plant leaves can reflect their internal physiological status, such as chlorophyll content, water content, intercellular spaces, and cell walls (Croft et al., 2014; Xu et al., 2023; Neuwirthová et al., 2024; Špundová et al., 2024)."

L136, not use the future, but restate like "in this study, 11 vegetation indices was used". Please, update it in accordance alongside the entire manuscript.

Response: Thank you for the comment. We revised the manuscript as follows (L134):

> "In this study, 11 vegetation indices were used, as summarized in Table 2."

L141-144, I suggest to restate these sentences in a better way. It is not clear wahat is the point: relevance to extract ground object (why?), or relevance of texture (?)...please explain better.

Response: Thank you for the comment. We revised the manuscript as follows (138-141):

> "With improvements in satellite image resolution, a single ground object may consist of multiple pixels, making spatial information increasingly important for image interpretation (Wang et al., 2015). Texture features describe the spatial arrangement and structural patterns of objects within an image, providing complementary information to spectral reflectance. This allows for better discrimination of land cover types with similar spectral characteristics."

L146-151, Move to Discussion. Here, simply state your methodology based on previosu researches, but the comparison with other works as suggegsted before should be moved to the final part (Discussion) of the manuscript (of course, underlining the novelty of your research in this regard).

Response: Thank you for the suggestion. The purpose of citing Guo et al. (2020) and Sibiya et al. (2021) was to reference their parameter settings rather than to perform a comparative analysis. To clarify this, the paragraph in the Methods section has been revised to emphasize the methodological reference only, as follows (143-146):

> "Following the parameter settings suggested by previous studies (Guo et al., 2020; Sibiya et al., 2021), texture features were extracted to enhance spatial information for classification. In this study, a moving window size of 7 × 7 was applied based on their findings, which provided an effective balance between detail and noise in texture analysis."

L163, Add reference.

Response: Thank you for the comment. We revised the manuscript and added the reference as follows (L157):

> "The process began by randomly sampling the data to create training datasets. After each sampling, the selected data points were returned to the dataset for the next round of sampling (bootstrap sampling) (Breiman, 2001)."

L168, was.

L168, Better explain for readers.

Response (L168): Thank you for the comment. We revised the manuscript as follows (L162-164):

> "The final classification result was determined by aggregating the predictions of all decision trees through a majority voting approach, which means that each tree casts one "vote" for a class label, and the class receiving the most votes becomes the final prediction."

L169, Add reference.

Response: Thank you for the comment. We revised the manuscript and added the reference as follows (L165):

> "To evaluate the importance of each feature, the Random Forest model uses the Gini Index (Breiman, 2001)"

L169, What "node" stands for should be explained, or alternatively add a reference.

Response: Thank you for the comments. We revised the manuscript as follows (L165-166):

> "which measures the impurity of a node. A node represents a point in the tree where the dataset is split based on a feature, with each node divided using the best split among a random subset of explanatory variables (Breiman, 2001)."

L173, Please explain what "feature importance" stands for in this context.

Response: Thank you for the comments. We revised the manuscript as follows (L171-172):

> "Feature importance quantifies how much each variable reduces node impurity and contributes to improving classification accuracy across all trees in the forest (Belgiu and Drăguţ, 2016; Breiman, 2001; Chen et al., 2023)."

L178, Restate better, since is not clear how this sentence is linked (logically) with the previosu and the next ones.

Response: Thank you for the comments. We revised the manuscript as follows (L174-175):

> "Ronneberger et al. (2015) proposed the original U-Net model, which was devolved from the fully convolutional network (FCN) and was initially designed for biomedical image segmentation."

L179, Add reference.

Response: Thank you for the comment. We revised the manuscript and added the reference as follows (L176):

> "The U-Net model consists of a contracting path (downsampling) and an expanding path (upsampling) (Ronneberger et al., 2015)."

L202, Add reference to support this decision.

Response: Thank you for the comment. We revised the manuscript and added the reference as follows (L198):

> "The dataset split was performed at the patch level, to avoid spatial autocorrelation and data leakage (Roberts et al., 2017)."

L202, Parameters are not explained.

Response: Thank you for your comments. The parameters used in the Kappa coefficient (Po and Pe) were already described in detail in Lines 213–214, where their definitions and corresponding formulas are presented. Since this information is already included, no additional changes were required.

L222, features.

L225-226, So, how such a misclassification influenced your results? You should discuss how to resolve such an issue, and show data supporting the fact that this didn't influenced obtained outcomes and models performances, since misclassification and errors between classes in both segmentation and next classification steps is a crucial point to be taken into account in similar works.

To support the proposed metodology and overall your results, you need to better consider such aspects, specifically in this case when you state that effective and evident misclassification occurred in your data-elaboration process.

L228-229, You need to show results supporting your statement.

Response (L222, 225-226, 228-229): Thank you for the valuable comment. We have revised Section 3.1 to discuss the potential influence of misclassification and to demonstrate, through performance metrics and field validation, that such errors did not significantly affect results or model performance. We revised in Lines 225-248, Section 3.1, and Discussion Lines 338-345, Section 4.2, as follows:

> "This study employed Random Forest (RF) and U-Net models with four feature combinations to examine land cover classes in Taiwan's Xue Mountain glacial cirques in the alpine treeline ecotone (ATE) region. Four land cover classes —bare land, forest, krummholz, and shadow —were investigated using feature combinations of spectral bands (8 features), vegetation indices (13 features), and texture features (56 features). The classification results of the RF and U-Net models with four feature combinations were compared in detail (Fig. 5 and Table 5). In general, the RF model demonstrated stable, robust classification performance across various feature dimensions. Specifically, the average F1-score of the RF model ranged from 0.823 to 0.839, the overall accuracy (OA) ranged from 0.817 to 0.830, and the Kappa coefficient ranged from 0.751 to 0.768 (Table 5). Among all classes, shadow and bare land achieved the highest F1-scores, both exceeding 0.85, while forest and krummholz maintained moderate but stable accuracy, ranging from 0.75 to 0.83. Additionally, the combination 4 yielded the highest F-1 score in forest and krummholz classes, indicating that the RF model improved when vegetation indices and texture features were combined with spectral information.

Furthermore, the U-Net model exhibited a marked improvement after incorporating more features. The F1-score for the forest class increased significantly from 0.609 for feature combination 1 (spectral bands only) to 0.828 for combination 4 (spectral, vegetation indices, and texture features). Likewise, the F1-score for krummholz improved from 0.696 to 0.778. Bare land and shadow also maintained high accuracy above 0.82 across all combinations. The U-Net's overall performance metrics (F1-score of 0.840, OA of 0.838, and Kappa of 0.778 in combination 4) surpassed those of RF, indicating that the U-Net model benefited substantially from integrating spectral, vegetation, and texture information.

Overall, the results showed that incorporating vegetation indices and texture features improved classification performance, particularly for vegetation classes in the U-Net model. Based on the higher F1-score in combination 2 than combination 3, it implied that vegetation indices contributed more than texture features. However, the highest F1-score was obtained in combination 4, indicating a complementary effect from vegetation indices and texture features. Additionally, the consistency between the classified ATE and field-observed forest–krummholz transitions further confirmed the classification's reliability. Overall, both models maintained stable performance across different feature combinations, supporting the robustness of the proposed approach."

"Despite the overall satisfactory classification performance, some confusion between forest and krummholz was observed due to their similar canopy structures and spectral reflectance. This misclassification occurred mainly along the transition between dense forest to stunted krummholz. However, this issue had only a limited influence on the overall outcomes. Field survey validation confirmed that the classified treeline boundaries were consistent with the observed forest–krummholz transitions in situ, and both RF and U-Net models maintained high accuracies (OA > 0.83, Kappa > 0.76). Therefore, the local confusion slightly affected boundary precision but did not alter the overall trend of the alpine treeline ecotone. To further minimize this effect in future work, incorporating structural features, such as LiDAR-derived canopy height models, could improve discrimination between forest and krummholz and enhance classification reliability."

L250, I think that time-consuming analysis is out of scope here. I suggest to remove this, simply highligjting the relevance to find out th best methodoogy able to reduce the time necessary for data elaboration. Infact, you should also consider the computation capabilities of different work stations, for example, as well as the possibility to improve performances (in terms of time-consuming) of data processing with various strategies (specifically concerning modeling parameters, programming languages, etc...)

Response: The time-consuming analysis (previous Section 3.2) has been shortened and integrated into Section 3.1 to improve conciseness of manuscript (L256-261).

"Based on the RF and U-Net model results, a further feature importance analysis was conducted to assess individual features in combination 4, comprising 77 features, including spectral bands, vegetation indices, and texture features. The feature importance analysis results revealed that the cumulative contribution achieved 95% interpretability with 61 features. Additionally, the OA and Kappa values improved slightly to 0.842 and 0.784, respectively. Moreover, computation time was reduced by 14.3% due to fewer features (Table 6). According to the feature ranking results, spectral bands and vegetation indices ranked higher than texture features, with SEVI, Y, B, G, and NDVI2 identified as the top five features (Fig. 6)."

L254, GNSS. Update in all the manuscript.

Response: The term "GPS" has been replaced with "Global Navigation Satellite System (GNSS)" throughout the manuscript to ensure technical accuracy.

L273, Increase the size of X-axis label. Label of the Y-ais is missing. Increase also the size of the title and numbers along the Y-axis as well.

Response: The figure was redrawn to improve readability. (L266 Figure6)

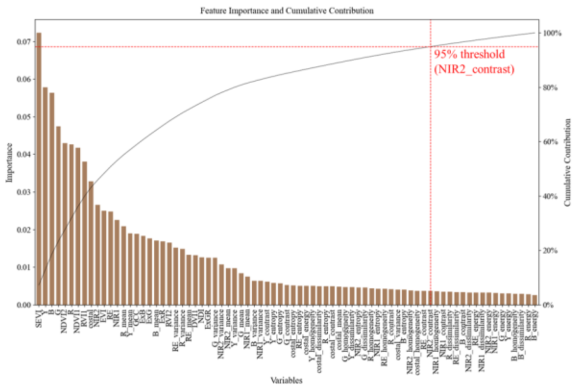

Figure 6. Feature importance ranking derived from the Random Forest model. Features are ranked based on their contribution to classification accuracy, with the top-ranked features including SEVI, Y (yellow), B (blue), G (green), and NDVI2. Most of the top features are spectral bands and vegetation indices, while texture features rank lower.

L276-277, See comment at line 250.

Response: We moved Table 6 to match the updated manuscript in L262. The time-consuming analysis (previous Section 3.2) has been shortened and integrated into Section 3.1 to improve conciseness of manuscript (L256-261).

L280, The figure is not clear for many reasons:
- what the black boxes stand for is not stated, neither in the text
- If results are similar using 77 and 61 features, it should be enought to simply tell it in the text, maybe the figure is not so necessary ? Please, justify your choice to include Fig.6, in accordance with the aims of your research.

Response: We have removed Figure 6.

L285, This was not stated before, but only here. You have to better explain the scope of ground truth validation data at the beginning of the manuscript, not at the end.
In addition, you say that trees' positions were collected along an elevation gradient, but no more details are explained in this regard (e.g., what elevation gradient? why you proceed in this way?...) these aspect need to be addressed at the beginning of the paper.

Response: To avoid misunderstanding, we rephrased the paragraph and explained the field observation in sections 2.3 and 3.2.

L288-289, Of course, but you could try to reduce the presence of shadow in data elaboration process, to improve classification results. This is not considered in the text: I suggest to better explain, in the methodology explanation, how to potentially reduce the presence of shadow adjusting models parameters, and the relevance of such an aspect.

See also comment for table 7.

Response: Thank you for your comments. In the methodology section 2.3, we have explained that we applied histogram matching technique to ensure the radiometric consistency across the two images (L113-115).

L293, No mention of this in the methodology section. Please edit.

Response: Thank you for your comments. We have added it to section 2.6.5. as follows (L216-222):

> "The bootstrap resampling method was a nonparametric approach used to estimate the variability and confidence intervals (CIs) of a statistic by repeatedly resampling with replacement from the original dataset. It enabled robust inference without assuming a specific data distribution (Efron and Tibshirani, 1993). The percentile method was commonly used, in which the 2.5th and 97.5th percentiles of the bootstrap distribution defined the 95% CI (Davison and Hinkley, 1997). To ensure stable and reliable estimates, between 1000 and 10,000 bootstrap iterations were generally recommended (Davison and Hinkley, 1997), with at least 5000 replicates providing sufficient accuracy for most applications (Carpenter and Bithell, 2000)."

L299-302, Results as shown here are not persuasive and acceptable: the tree line detection by classification outcomes is not clear in your figure: I suggest to enlarge the size of the image, add a basemap on the background, adjust transparency and remove the visualization of the shadow class, to better visualize tree line detection from classification procedure. Here it just seems a mosaic of classes with no clear evidence of tree line detection, despite the presence of ground truth point (in addition: add simply dots for ground points).

In addition: looking at RGB images, treeline is not so evident. Please choose a more clear area where tree line is evident from RGB images and therefore by the classification outcomes, considering the comment above.

Response: Thank you for the comment. We have modified Figure 7.

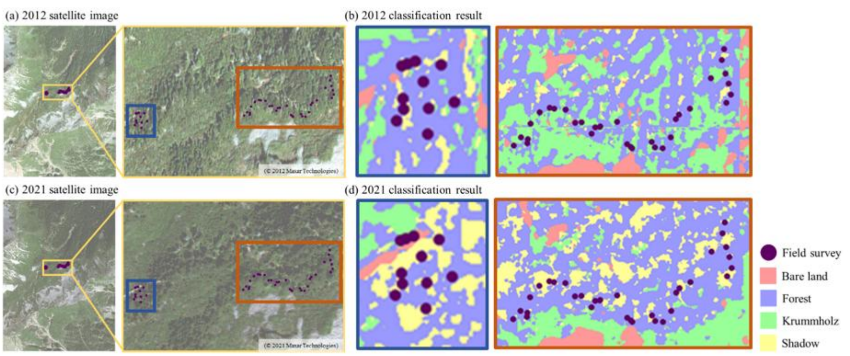

Figure 7 Comparison of satellite imagery and classification results from 2012 and 2021. Panels (a) and (c) show high-resolution satellite images for 2012 and 2021, respectively. Colored boxes in these images indicate the enlarged areas shown in (b) and (d). Panels (b) and (d) present the classification results of the corresponding enlarged regions using a U-Net model trained with 61 selected features. Triangles mark field survey locations.

L303, The reasons of increasing presence of shadows are not stated, neither how to reduce their presence in the classification outcomes). Please consider this issue in the text, adding accurate discussion of such aspects looking at the aims of your research.
See also comment at line 288

L304, This should depend on classification accuracy and data preparation. Consider this issue about shadows looking at the previous comments.

Response: Thank you for your comments. The presence of shadow may reveal evidence of ecological changes, such as growing trees or a denser canopy. Additionally, the increase in shadow also aligns well with our findings of forest expansion. Furthermore, terrain-related shadows could provide important information about the microenvironment for further analysis of habitat conditions.

L307-308, Some improvement are needed:
- remove the DTM on the background, is not useful and create confusion.
- instead, add a basemap with transparency adjustements
- field investigation points are not visible
- this is not the entire study area, or yes? it is not clear
- change colors with a good palette of omogeneous colors
- increase the size of the figure, moving legend and items on the left side
- reduction class is not visible (?)

Response: Thank you for the comment. We have modified Figure 8.

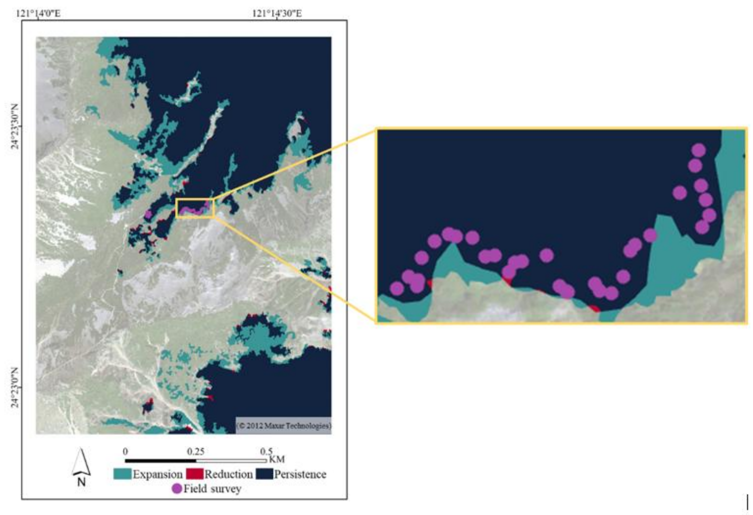

Figure 8. The spatial distribution of ATE area changes from 2012 to 2021. ATE expansion is marked in dark cyan, reduction is marked in dark red, persistence is marked in dark blue, and field survey point in purple.

L317, Update including sections highlighted in the comments above at the beginning of the manuscript.

Response: Thank you for the comment. We carefully edited the manuscript with update subtitle for each section.

L337, Too colloquial, update in a more formal way.

Response: Thank you for the comments. We revised the manuscript as follows (L352):

"In this study, a total of 77 features were derived from the satellite imagery, including 8 spectral bands, 13 vegetation indices, and 56 texture features."

L342, Adjust in accordance.

Response: Thank you for the comments. We revised the manuscript. (L355-357)

"Notably, most of the top-ranked features were spectral or vegetation index variables, whereas texture features contributed less to the classification. The feature selection slightly improved the overall accuracy (+0.4%) and the Kappa coefficient."

L343-344, Use indirect syntax. As above.

Response: Thank you for the comments. We revised the manuscript as follows (L357-360):

"Although OA was used as the primary selection criterion, the analysis also confirmed that the selected features maintained or improved F1-scores for the forest class, the primary focus of detecting treeline changes. It should be noted that optimizing overall accuracy (OA) values may sometimes overlook minority or ecologically important classes."

L363, Conclusions need to be improved: in-depth discussion of the usefulness of your research is missing, as well as concerning aspecs including ecology, changes in habitats and species presence, influence of ATE shift in natural disturbances occurrence and social risks, social impacts of ATE changes over time, changes in forest management practices, climate change and future scenarios...
Please consider all these aspects (and other) and improve the Conclusions of your research.

Response: We reorganized the conclusion section to address ecology aspects related to our findings. We emphasized our contribution and innovation of demonstrating the potential of high-resolution satellite imagery for long-term ecological monitoring.

L362, Use a more accurate term.

Response: Thank you for the comments. We rephrased the sentence as follows. (L385)

"This study investigates changes in the alpine treeline ecotone (ATE) of the Xue Mountain glacial cirques in Taiwan from 2012 to 2021, utilizing WorldView-2 imagery integrated with Random Forest and U-Net models."

The reviewer marked several sentences for deletion directly in the annotated PDF.

Response: We sincerely thank the reviewer for the detailed annotations and constructive suggestions. Most of the sentences marked for deletion have been removed as suggested. In addition, several related sentences in the same sections were also revised to improve clarity, readability, and consistency throughout the manuscript.