



OPEN Automated mapping of glacial lakes in Himachal Pradesh using multi source remote sensing data and machine learning

Bhawna Pathak¹, Ankit Singh¹, Reet Kamal Tiwari² & Dericks P. Shukla¹✉

The Himalayas are undergoing active climate-induced changes, resulting in glacial lake formation and expansion. Glacial lakes are important reservoirs of freshwater but also run the risk of glacial lake outburst floods (GLOFs). As sentinels of climate change, dynamic monitoring of glacial lakes is necessary. This research introduces an automated method for mapping glacial lakes in Himachal Pradesh based on multi-source remote sensing data and a random forest (RF) classifier. The model was tested under various scenarios using spectral bands and remote sensing indices extracted from Sentinel-2 and Planet images. The combination of Sentinel-1 SAR, Sentinel-2 MSI, and SRTM DEM data resulted in a classification accuracy of 93.69%, which increased to 94.44% with the addition of high-resolution Planet images. Although the method was effective in identifying glacial lakes, it faced difficulties in distinguishing glaciers from supraglacial lakes. Postprocessing methods were used to enhance the results. Model performance was evaluated using statistical measures, such as recall, precision, F1-score, and overall accuracy. The RF classifier performed well robustly, identifying its reliability in glacial lake mapping even being a machine learning method.

Keywords Glacial lakes, Random forest, Planet data, Automated detection, Northwest Himalaya

Glaciers are crucial parts of the Earth's cryosphere, acting as enormous reservoirs of freshwater and they play a fundamental role in governing global climate and sea level change. Glaciers hold approximately 69% of the freshwater resources in the world, supporting ecosystems, agriculture, and human communities, especially in mountainous areas such as the Himalayas, Andes, and Alps¹. The stability and integrity of these natural features are significantly affected by changes happening in the climatic scenarios. The accelerated rate of global warming has greatly impacted glacier dynamics and initiated global-scale retreat and thinning of glaciers². Increased temperatures have accelerated mass loss, with research showing that glaciers lost more than 9 trillion tons of ice between the 1960s to 2016³. This accelerated deglaciation not only contributed to sea level rise but also destabilized glacial ecosystems, creating problems for higher elevation areas. One of the most notable effects of this change is the formation and expansion of glacial lakes, which are water bodies present in depressions created by receding glaciers and are usually dammed by moraines or glacier ice⁴.

Glacial lakes are usually harmless but pose considerable dangers when they expand and become unstable and sometimes causing glacial lake outburst floods (GLOFs). GLOFs are among the most devastating floods that are unexpected and disastrous outflow of water due to the collapse of a natural dam/ barrier such as a moraine or ice wall⁵. They are highly dangerous to communities and infrastructure in downstream areas as well as ecosystems, sending torrential volumes of water, sediment, and debris within a short span of time. In addition to GLOFs, these lakes can contribute to landslides, avalanches, and erosion, increasing the vulnerability of mountainous regions⁶. Historically, GLOF events—such as the 1941 Huaraz disaster in Peru or the 1985 Dig Tsho flood in Nepal—have claimed lives, destroyed houses, and disrupted livelihoods, displaying its highly destructive nature^{7,8}. The impact of GLOF is complex, and involves loss of life, displacement, and economic destruction. Glacier communities, which usually rely on glacial meltwater for agriculture and hydroelectric power are at a risk of facing extinction as GLOFs ravage houses, roads, and fields⁹. In the Himalayas alone, millions of residents occupy possible flood paths that lack early warning and adaptive infrastructure¹⁰.

¹Dexter Lab, School of Civil and Environmental Engineering, Indian Institute of Technology (IIT) Mandi, Mandi, Himachal Pradesh 175005, India. ²Indian Institute of Technology, Ropar, Punjab 140001, India. ✉email: dericks@iitmandi.ac.in

As global warming continues to intensify, these events are predicted to become more frequent and intense, fuelled by the accelerating expansion of glacial lakes in High Mountain Asia (HMA), where satellite images show that thousands of new lakes have appeared over recent decades¹¹. Satellite data reveal that the volume, area and count of glacial lakes have increased significantly, especially in high-mountain Asia, where warming is greater than the global average^{12–15}. This rapid development of glacial lakes is an urgent issue at present, as the process of deglaciation has gained momentum and erratic weather conditions lead to their formation. These glacial lakes can be classified/ identified based on spectral indices such as the NDWI (Normalised Difference Water Index); NDSI (Normalised Difference Snow Index) and other contributing satellite features by applying supervised machine learning, support vector machines, and random forests techniques^{14,16}.

Deep learning, particularly CNN, has enhanced the extraction of features and classification even under difficult scenarios such as hilly terrain conditions, clouds, shadows, or mixed pixels^{17–19}. For example, Kaushik et al. 2022²⁰ crafted GLNet, a deep convolutional neural network trained for glacial lake mapping using multisource remote sensing images such as multispectral, thermal, microwave, and Digital Elevation Model (DEM) inputs. They have obtained high accuracy (0.98) with good spatial transferability, as the method is fully automatic. Wangchuk et al. 2020²¹ also addressed the problems of small lake sizes, cloud cover, seasonal snow, and changing turbidity by combining Sentinel-1 Synthetic Aperture Radar, Sentinel-2 Multi-Spectral Instrument, and DEM data in a random forest classifier model. The completely automated Python library “GLakeMap” that maps glacial lakes in alpine areas across geographic and climatic differences. Mustafa et al. 2024²² combined Sentinel-1 radar backscatter parameters, Sentinel-2 spectral indices, and topographic parameters to train their ML models, ANNs, SVMs, and random forest. ANN achieved the highest accuracy at 95% and CNNs were able to identify glacial lakes with minimal human intervention²². LSTM networks and RNNs have also been applied to model temporal variations in hazards and enhance forecasting²³. Nevertheless, high-altitude terrain restricts training data quality and quantity, and extensive validation is needed to account for regional heterogeneity. Addressing these challenges will involve concerted efforts to augment datasets, increase algorithm resilience, and combine ML with conventional techniques.

The improvements in machine learning algorithms have gone a long way in the analysis and observation of glacial lakes beyond their remote and dynamic nature. Thus, the present study focuses on developing an automated and robust multi-level methodology for accurate detection of glacial lakes in Himachal Pradesh utilising high-resolution multisource remote sensing data and machine learning techniques. With the growth and destabilization of glacial lakes, understanding and minimizing their threats is imperative for scientists, policymakers, and affected populations.

Study area

This work is carried out in the state of Himachal Pradesh covering an area of 55,673 km² that falls in the Western Himalaya. The state shares its eastern border with China; its northern and north-western borders are Jammu and Kashmir; Uttarakhand is to the southeast; and Punjab is to the southwest. The elevation of the state varies from ~ 250 to 7026 m above mean sea level, creating a diverse topography and climate. The major rivers are the Beas, Chenab, Indus, Ravi, and Satluj, including their tributaries that support local ecosystems and livelihoods. There are approximately 2100 glaciers, spanning an area of ~ 3799 km², covering approximately 6.8% of the state's area²⁴. The longest glacier in the state is Bara Shigri Glacier, which is approximately 26 km long²⁵. The Indian Summer Monsoon (ISM) and western disturbances influence the region's snowfall, with maximum snowfall occurring from December to March^{26,27}.

Glacial lake inventory

Studies indicate that these glaciers have been retreating since the end of the Little Ice Age, leading to the expansion of glacial lakes in previously glaciated areas^{28,29}. Earlier inventories mapped 958 glacial lakes larger than 500 m² during 2011–2013²⁹. A more recent study reported that the number of glacial lakes in the Satluj River basin of the Himachal Himalaya nearly doubled, increasing from 562 in 2019 to 1048 in 2023³⁰. We created a glacial lake inventory for Himachal Pradesh through manual digitization using Google Earth Pro, which provides high-resolution images from Maxar and Airbus. We mapped 651 glacial lakes in Himachal Pradesh for the year 2017 and 1130 lakes for the year 2022. This rapid growth highlights the urgent need for automated approaches to map glacial lakes and generate frequently updated inventories. To address this need, we selected 95 lakes from the 2022 inventory for our study area, shown by the blue inset box in Fig. 1b, and used them as the response variable for training and testing our model.

Methodology

In this study, we used a combination of spectral, radar, and topographic variables as predictors and a glacial lake inventory as the response variable for lake classification. We extracted the predictor variables from Sentinel 1 (10 m), Sentinel 2 (10 m), SRTM DEM (30 m) and PlanetScope (3 m) of 2022 and 2023. We used September 2023 satellite image because cloud-free data for 2022 were unavailable for the entire study area. The glacial lake inventory, prepared for 2022, served as the response variable, allowing us to train the model on past lake distributions (2022) and apply it to detect lake presence in the following year (2023). A summary of satellite datasets that are used in this work including their spatial resolution, temporal resolution and swath width are given in Table 1. Although we focused on detection for a single time step, we designed the methodology to remain scalable and adaptable for multi-temporal applications in future studies.

For Level I classification, we calculated spectral indices including the Normalized Difference Snow Index (NDSI), Normalized Difference Water Index (NDWI) using blue and green bands, Normalized Difference Vegetation Index (NDVI), and Normalized Difference Glacier Index (NDGI) from Sentinel-2 data. We used

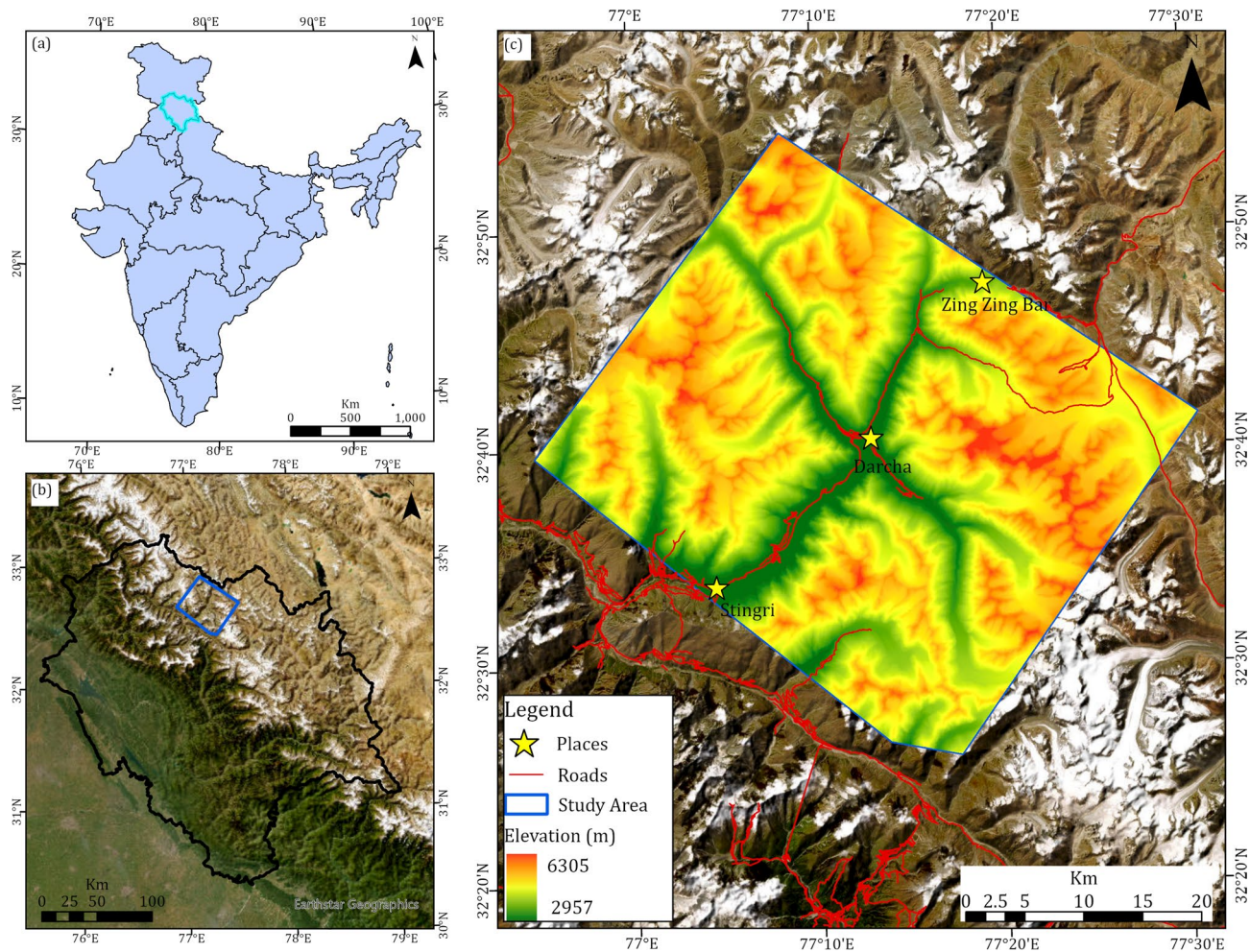


Fig. 1. Location map of the study area. The map illustrates the study area situated in Himachal Pradesh, India, marked by the blue inset box.

Satellite	Sensor	Spatial resolution (m)	Revisit period	Swath width (km)	Bands used
Sentinel-2	MSI	10 (NIR, Green, Blue and Red) 20 (SWIR)	10 days (5 days with combined constellation)	290	NIR, Green, Blue, Red, SWIR
PlanetScope	Dove-C / Dove-R / SuperDove	3	Daily at nadir	25	NIR, Red, Green
Sentinel-1	C-SAR	10 (Interferometric) / 20–40 (Standard)	6–12 days (depending on mode and region)	250	C-band (SAR)
SRTM	C-band SAR	~30 (Public) / ~90 (Global)	One-time mission (2000)	225	C-band (SAR)

Table 1. Summary of satellite datasets used in the study along with their key characteristics.

radar backscatter data (VV and VH) from Sentinel-1 and derived topographic variables such as elevation, aspect, and slope from the SRTM DEM. For Level II classification, we used NIR band, NDVI, and NDGI calculated from high-resolution PlanetScope data, along with the same radar and topographic variables. All predictor variable layers were resampled to a common resolution of 10 m using bilinear interpolation to maintain spatial uniformity and ensure accurate pixel alignment within image stack used for training.

We used a glacial lake inventory as the response variable to train and test the Random Forest model as shown in Fig. 2. We evaluated the model's performance using metrics such as AUC-ROC, recall, precision, accuracy, and F1-score. We applied an automated post-processing workflow and validated the results using high-resolution Planet images.

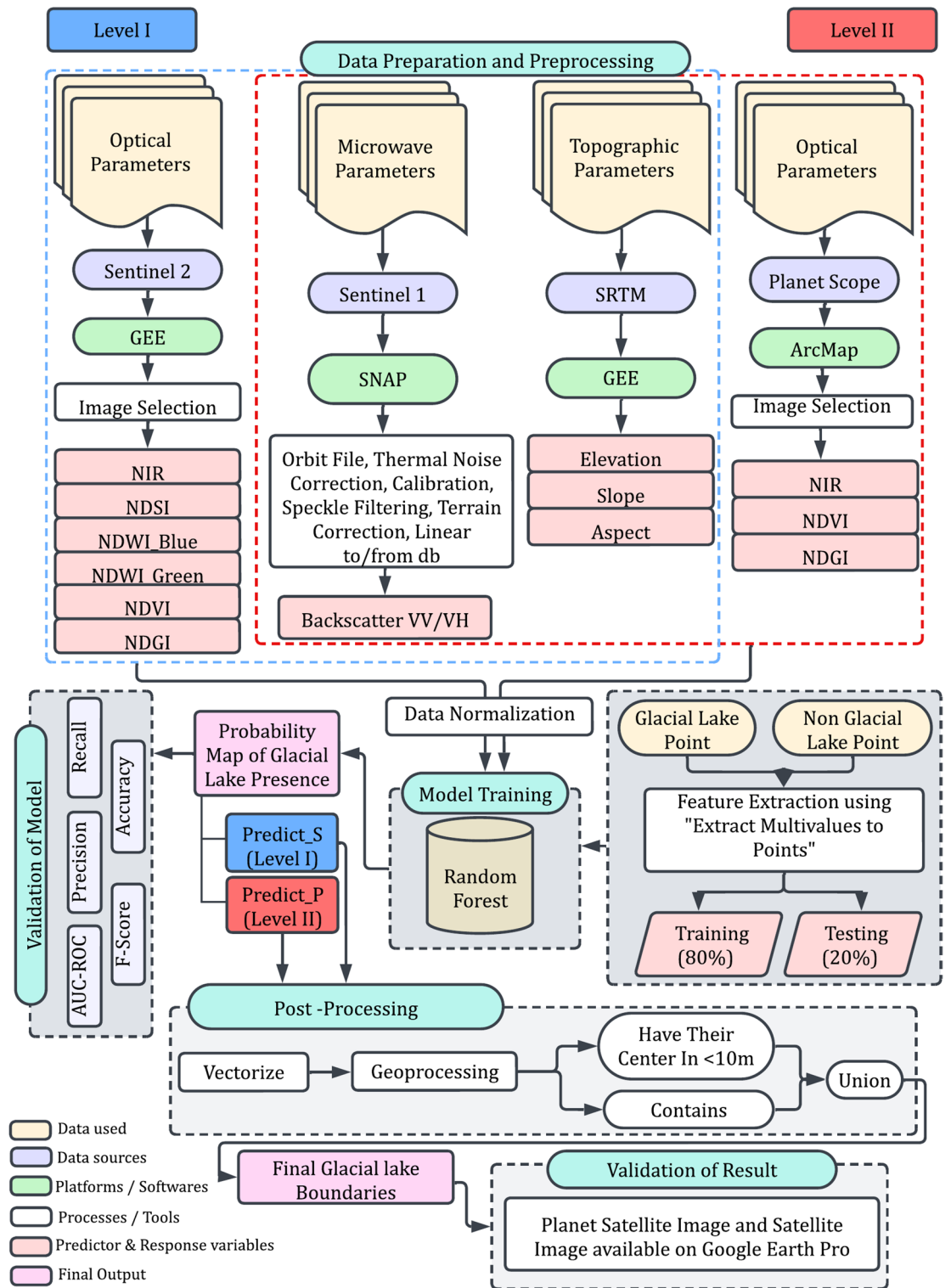


Fig. 2. Workflow chart illustrates the methodology, encompassing data acquisition, preprocessing, classification, analysis, and validation steps.

Level I

For Level I, we used SAR and optical satellite data of 2023 from Sentinel-1, and Sentinel 2 MSI respectively and the SRTM DEM. Optical images assist in identifying spectral contrasts between glacial lakes and their environments and were used to estimate several spectral indices. The NIR band from Sentinel-2 helps in the water-land discrimination because water intensely absorbs NIR radiation³¹. NDSI discriminates ice and snow

from water and land to provide correct lake boundary delineation in glaciated zones. The Sentinel-2 blue, and green bands were utilized to derive two NDWIs namely:

$$NDWI_{Blue} = \frac{B2 - B8}{B2 + B8} \quad (1)$$

$$NDWI_{Green} = \frac{B3 - B8}{B3 + B8} \quad (2)$$

where B2 is blue band, B3 is green band and B8 represent NIR band of Sentinel 2 satellite data. Glacial lakes that reflect higher in B2 band are enhanced more appropriately with the use of NDWI_blue (Eq. 1), whereas NDWI_green is appropriate for glacial lakes reflecting higher in B3 band (Eq. 2)²¹. Thus, using both the indices, glacial lakes can be identified accurately. NDSI discriminates ice and snow from water and land, enabling accurate delineation of glacial lake boundaries in glaciated zones. Along with NDSI, we used NDVI and NDGI for glacial lake identification. NDVI helps differentiate water bodies, which typically show low or negative values, from vegetated areas that reflect strongly in the near-infrared region. NDGI enhances the detection of snow-covered and glaciated areas by emphasizing the contrast between ice and surrounding terrain³². NDSI, NDGI and NDVI were calculated using Eqs. (3), (4) and (5) respectively using Sentinel 2 data.

$$NDVI = \frac{B8 - B4}{B8 + B4} \quad (3)$$

$$NDSI = \frac{B3 - B10}{B3 + B10} \quad (4)$$

$$NDGI = \frac{B3 - B4}{B3 + B4} \quad (5)$$

where B3 is the green band, B4 is the red band and B8 represent the NIR band and B10 is the SWIR band of Sentinel 2.

As, SAR data can penetrate cloud cover, which tends to interfere with optical images in high-altitude regions³³, we used VV and VH backscatter from Sentinel-1 data after applying the pre-processing steps outlined in the methodology flowchart (Fig. 2). The pre-processing framework included an orbit file application, thermal noise removal, radiometric calibration, speckle filtering, terrain correction, and final conversion from linear scale to decibels (dB)^{22,34,35}.

We used spectral indices NDSI, NDWI (green and blue), NDVI, and NDGI; spectral bands such as NIR from both Sentinel-2 and Planet; radar backscatter (VV and VH) from Sentinel-1; and topographic variables including elevation, aspect, and slope as predictor variables. Out of 11, 9 input layers along with the training and testing points are shown in Fig. 3. We resampled all eleven layers to a common spatial resolution of 10 m using bilinear interpolation, normalized them, stacked them as input for the RF model and further processing. This resampling step is important to ensure spatial uniformity across all input layers and proper alignment within the image stack used for model training. This approach allows for easy integration of all datasets within a unified framework suitable for pixel-based classification tasks.

Level II

For Level II, in addition to Sentinel-1 SAR data, SRTM DEM, we used PlanetScope optical images from 2023. NIR, Red and Green bands from PlanetScope images were used to calculate the NDVI and NDGI using Eqs. (3) and (5). Like Level I, we used Sentinel-1 backscatter VV/VH, after pre-processing it. We created an ensemble dataset consisting of training point values extracted from Planet's NDVI, NDGI, and NIR, along with values from SRTM's elevation, aspect, and slope, as well as Sentinel-1's VV and VH backscatter, and Sentinel-2's NDSI, NDWI_Green, and NDWI_Blue. We then executed this ensemble dataset on the integrated image stack used in Level I.

Training and testing data preparation

We mapped 95 glacial lakes in the study area for the year 2022 and generated 950 data points within the glacial lake polygons to obtain a robust dataset for training. We also generated 950 non-glacial lake points within the study area to ensure equal representation of both classes. However, we found severe misclassification in the first iteration, where the model incorrectly classified streams as glacial lakes. To address this, we generated additional 500 points particularly within the streams to help the model better differentiate between lakes and streams. Thus, by more accurately representing stream features, this approach improved the model's accuracy and reduced the false positives in glacial lake detection.

We resampled all predictor variables to a spatial resolution of 10 m and tested for multicollinearity to ensure consistency in the data. We extracted the values of the predictor variables at 950 glacial lake points and 1450 non-glacial lake points using the 'Extract Multi-Value to Points' tool in ArcGIS Pro. We then constructed a binary dataset by assigning a value of 1 to glacial lake points and 0 to non-glacial lake points. Finally, we split the dataset into training and testing sets, using 80% for training and reserving the remaining 20% for testing.

Random forest

Random Forest (RF) is an ensemble machine learning algorithm that has been very commonly employed for classification and characterization problems. It works by building multiple decision trees (DTs)³⁶, each of which

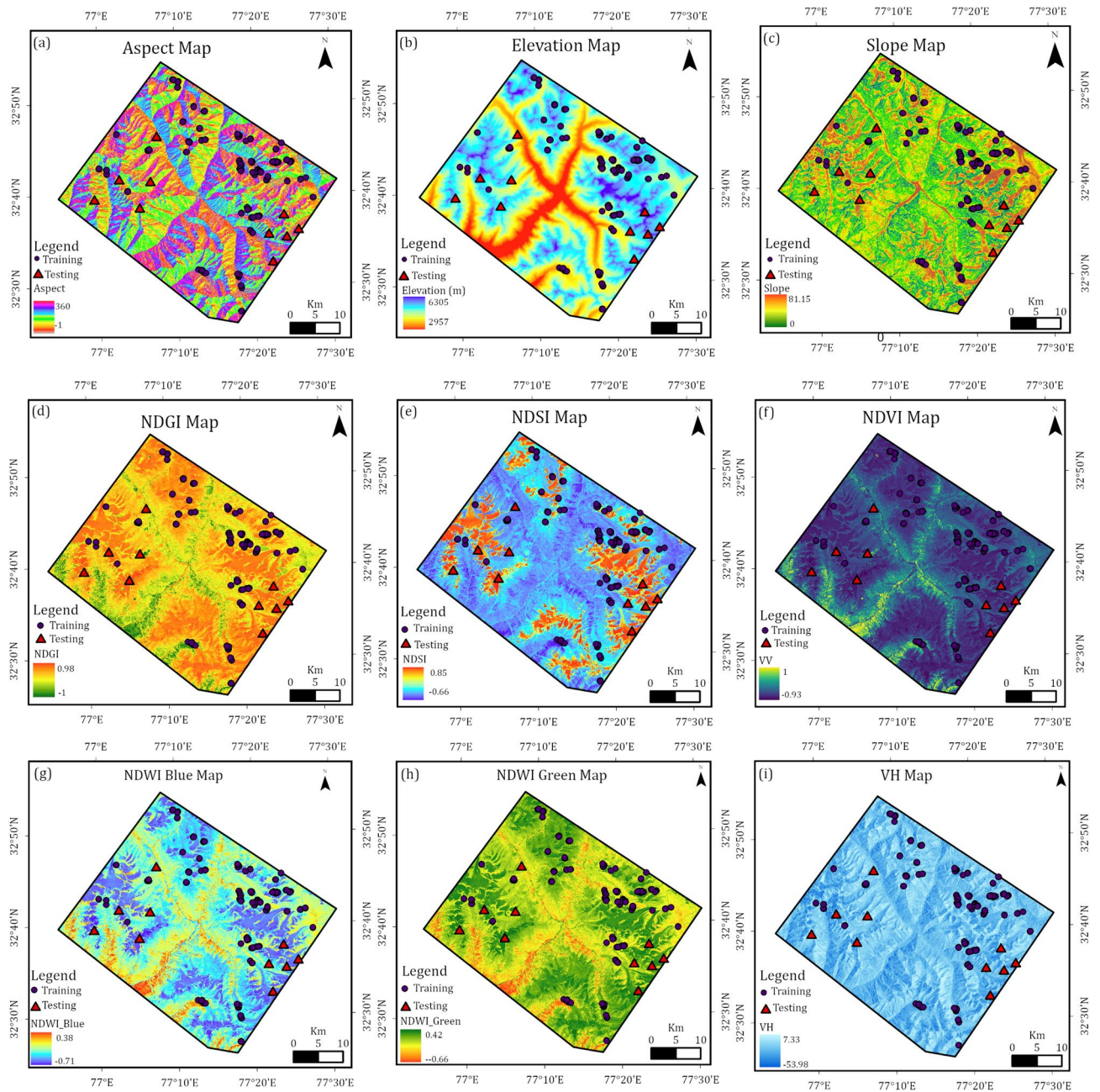


Fig. 3. Spatial distribution maps of various predictor variables used in the study. **a** Aspect Map, **b** Elevation Map, **c** Slope Map, **d** NDGI Map, **e** NDSI Map, **f** NDVI Map, **g** NDWI Blue Map, **h** NDWI Green Map, and **i** VH Map. The maps display training and testing points marked as triangles and circles, respectively, over the study area, with corresponding color gradients representing the range of values for each variable.

is based on different subsets of the training set and features by a process referred to as bagging (bootstrap aggregating). By performing an ensemble, overfitting is avoided and generalization is improved, unlike a single DT, which tends to be biased and prone to overfitting³⁷. In RF, a single tree is constructed based on a bootstrap sample of the training data, and the remaining data are utilized to calculate the out-of-bag (OOB) error, which aids in model performance evaluation. The performance of the algorithm relies on several important hyperparameters, such as the number of trees (ntree), the number of features to consider at each split (mtry), and the node size, which determines the trees' depth^{38,39}. Trees are usually incremented until the OOB error converges. By combining predictions of many decision trees, RF improves classification accuracy and provides stable predictions for glacial lake identification.

Accuracy assessment

The performance of the glacial lake detection model (RF) was assessed using various statistical metrics, such as accuracy, precision, recall, F1-score and the receiver operating characteristic area under the curve (ROC-AUC). Accuracy is the ratio of correct predictions to all predictions and is beneficial when the class distribution is balanced (Eq. 11). However, ROC-AUC is preferable when dealing with class imbalance, as it calculates the model's power of class discrimination through the true positive rate versus the false positive rate or specificity (Eq. 7) versus the 1-sensitivity (Eq. 6) plot.

Precision measures the ratio of correctly identified positive instances among all predicted positives (Eq. 8), whereas recall captures the model's capacity to recognize actual positives (Eq. 9). The F1-score, which is a harmonic mean of recall and precision, provides a balance between the two measures (Eq. 10), such that adequate evaluation is achieved, particularly where false positives or false negatives must be minimized. The performance metrics were calculated as follows:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (6)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (7)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (8)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (9)$$

$$F - \text{score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + TN + FP} \quad (11)$$

where TP is True positive which in our case means pixels which were correctly identified as glacial lakes. TN stands for True Negative, which represents pixels that are not glacial lakes and were not classified as such. FP is False Positive, it denotes pixel incorrectly identified as glacial lakes by the model, while FN is False Negative and it refers to glacial lake pixels that were missed by the model.

Results

Multicollinearity test of predictor variables

Multicollinearity refers to the correlation among predictor variables, which can create redundancy and affect model performance⁴⁰. To assess this, we conducted a Pearson correlation test to evaluate potential multicollinearity between the predictor variables (Fig. 4). In Fig. 4, we can see that the NDGI and NDVI have a high positive correlation ($r = 0.94$), which indicates possible redundancy. However, feature contribution analysis with Random Forest (RF) and SHapley Additive exPlanations (SHAP) values revealed that the NDGI contributes much more than the NDVI does in predictive modelling. Similarly, NDWI_Green and NDWI_Blue have an almost perfect correlation ($r = 0.98$), suggesting that they both carry the same spectral information. Although they are highly correlated, SHAP-based and RF feature importance showed that NDWI_Green makes a valuable contribution to model prediction, justifying its inclusion in the predictor variables. Additionally, the NDVI and NDWI_Blue have very strong negative correlations ($r = -0.85$). Although high correlation values may signal possible redundancy, our investigation shows that correlation alone is not adequate for selecting variables. Even though some predictor variables are highly correlated, the Random Forest and SHAP analyses confirm that each contributes uniquely to improving model performance. Therefore, in our approach, we held on these variables by their predictive significance and not by dropping them solely on the basis of correlation cut-offs.

Independence and importance of predictor variables

Machine learning and deep learning techniques successfully replicate the nonlinearity among variables; nonetheless, they are black-box models and do not, by default, provide insight into the contribution of each factor toward the end prediction. To tackle this challenge, the feature importance of predictor variables was assessed through SHAP values (Fig. 5), which not only quantify the contribution of each variable but also show the direction and magnitude of their influence on the classification outcome. For more insight, the overall variable ranking from the RF feature importance analysis is shown in Supplementary Fig. S1, which complements the SHAP results by illustrating the relative contribution of features based on their role.

The SHAP values derived for the study area show that the NDSI is the most significant feature, as positive SHAP values confirm that high values of the NDSI make classification of a glacial lake more likely. NDWI_Green also shows a considerable contribution, where high values (red) contribute positively to lake classification and low values (blue) decrease the probability of the presence of a lake. This conclusion is consistent with earlier research supporting the effectiveness of spectral water indices in the detection of water bodies⁴¹. Terrain characteristics such as slope and elevation are also significant, with gentle slopes and lower elevations promoting the formation of lakes, as also observed in high elevation lake distributions²².

NIR and NDWI_Blue are moderately significant, indicating their sensitivity to surface reflectance and water content. Aspect also emerged as an important factor as it may impact lake formation patterns. Radar-based

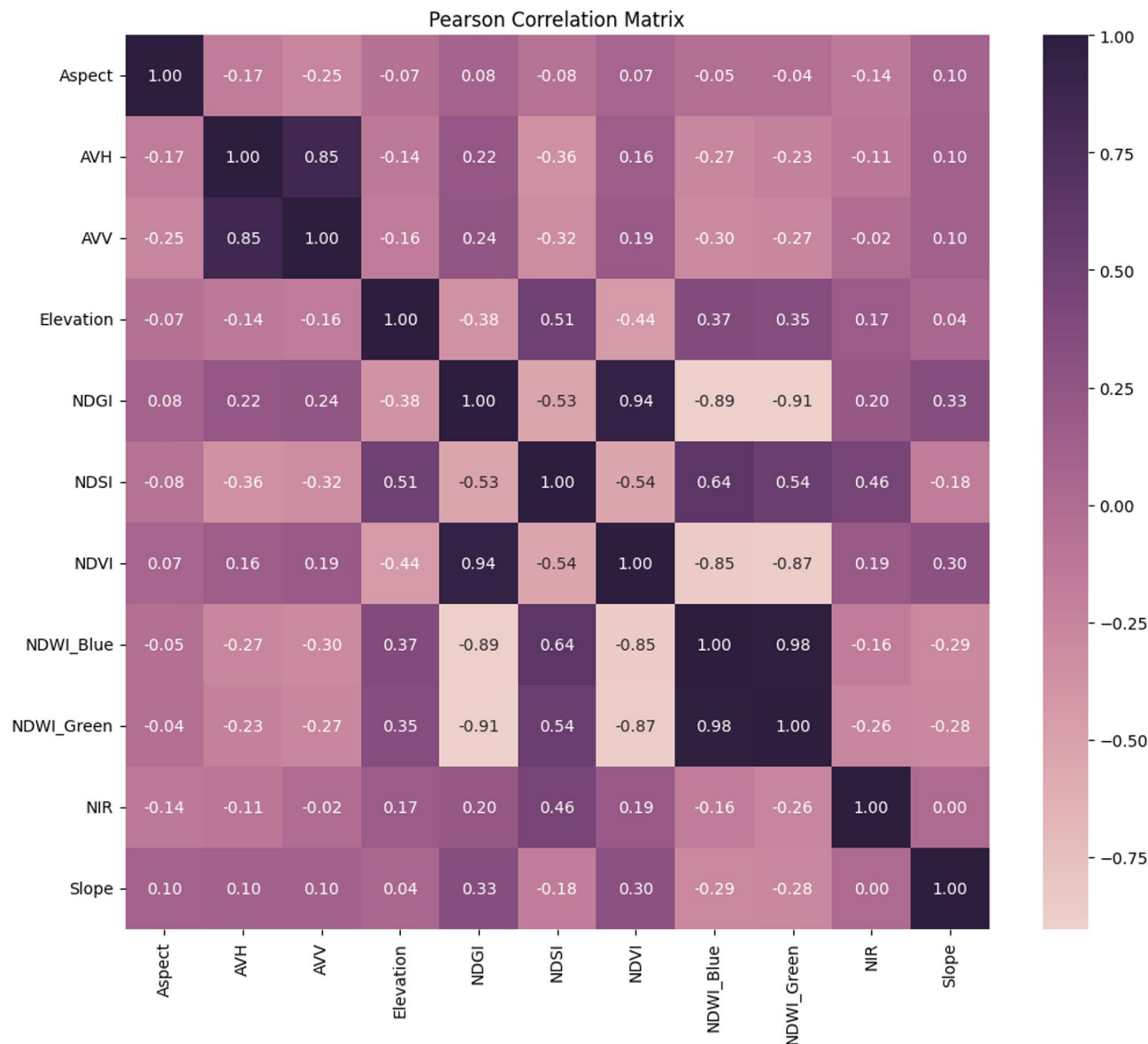


Fig. 4. Pearson Correlation Matrix illustrating the multicollinearity among various response variables used in the study. The color scale represents the strength and direction of the correlation, ranging from -1 (strong negative correlation) to 1 (strong positive correlation).

features such as ascending VV and VH backscatter are of less importance but are useful in providing information on surface roughness and water presence, as indicated by the findings of Shen et al. 2022⁴². The insights from SHAP and RF confirms the relative importance of the variables used in the study. The detailed ranking of variables from the RF model is provided in the Supplementary Fig. S1 and are discussed in detail the supplementary note.

Classification results

Level I

The performance analysis of the Random Forest (RF) model for level I classification yielded promising results (Fig. 6). The model had a total accuracy of 93.69% on the test dataset, which means that it was quite successful in separating glacial lakes from other terrains. Moreover, the statistical measures of class 1 (glacial lakes) also validate the reliability of the model, as the precision is 0.93, the recall is 0.89, and the F 1-score is 0.91. These outcomes emphasize the excellence of the model in accurately predicting glacial lakes with a perfect balance between recall and precision. The validation was conducted visually using the high-resolution Planet Scope satellite data.

Despite the high accuracy, there was some degree of misclassification in the form of false positive and false negative pixels. These classification mistakes are mainly due to spectral similarities of glacial lakes with adjacent features, including wet ice pixels, shadows, and frozen glacial lakes. This similarity makes it difficult to correctly separate water bodies from adjacent terrain.

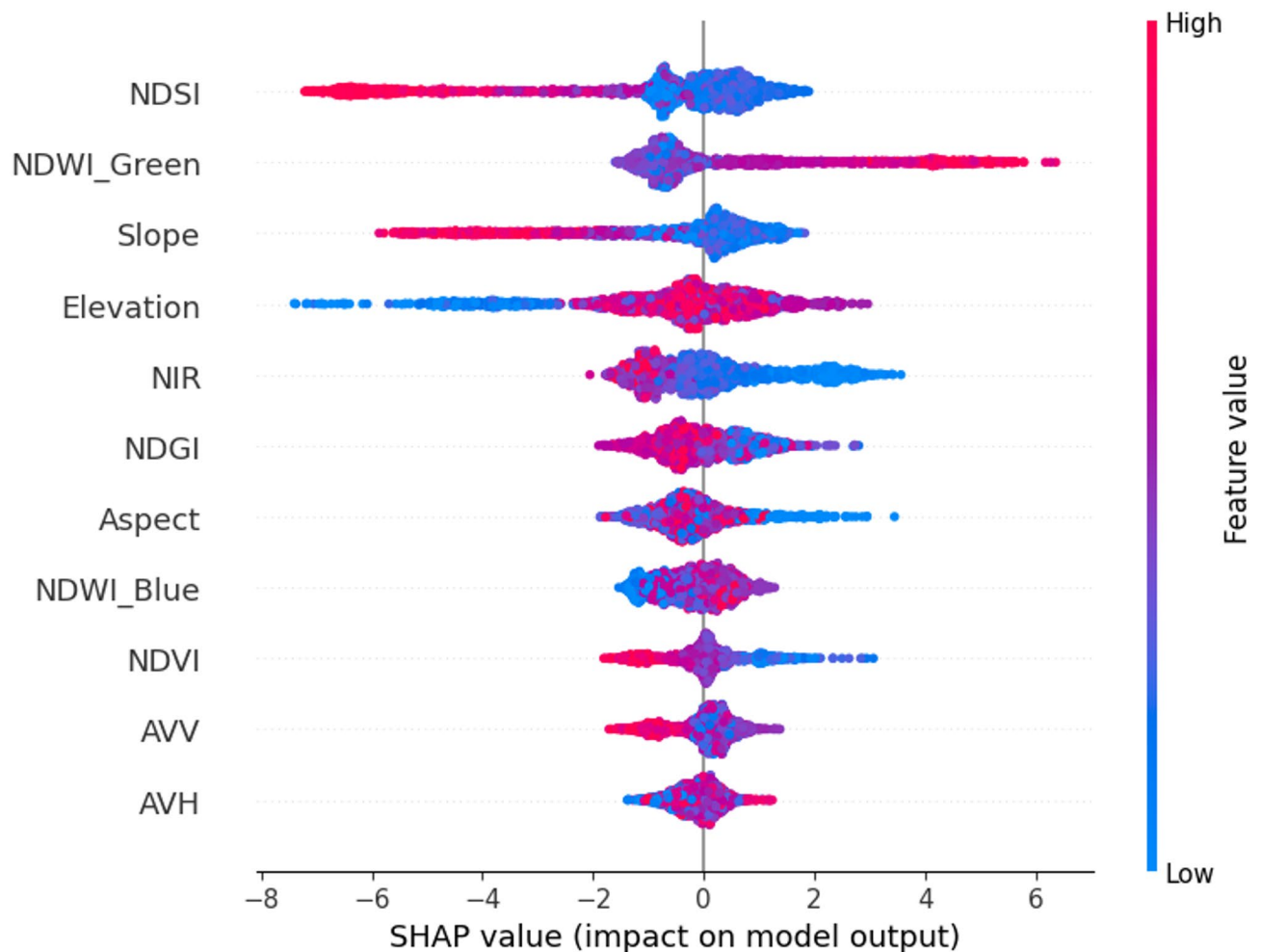


Fig. 5. The plot illustrates the impact of individual features on the model's prediction using SHAP values. Features are ranked by their importance, with NDSI, NDWI_Green, and Slope having the most significant influence.

Level II

Addressing the above-mentioned misclassification errors, we integrated the NIR band and other remote sensing indices, such as the NDVI and NDGI extracted from Planet data (Level II). Multi-value extraction was performed on training points using Planet's NIR, NDVI, and NDGI, and these points were then merged with the original training dataset. This ensemble dataset was subsequently executed on an integrated SRTM + Sentinel-2 + Sentinel-1 image stack at a 10-meter resolution. By utilizing this upgraded methodology, we observed an increase in the classification performance (Fig. 7), with a test accuracy being 94.44%, precision of 0.94, and recall of 0.92 for class 1 (glacial lakes). In addition, the number of false negatives decreased from 27 in Level I to 20 in Level II, while false positives decreased from 15 in Level I to 13 in Level II. This slight reduction in misclassification indicates that combining multiple data sources enhanced the model's ability to distinguish glacial lakes from spectrally similar neighbouring features.

Post-processing

The Random Forest model outputs class probabilities by aggregating predictions from all decision trees in the ensemble. For each pixel, the model calculates the probability of belonging to the "glacial lake" or "non-glacial lake" class based on the proportion of trees that vote for each class. This allows classification to be better represented by the confidence level as well as predicted label, as reflected in the probability score. Such probabilistic output supports both label assignment and uncertainty assessment in classification. Further, for refining the results, post-processing operations were performed to improve the classification results obtained from the RF model. A probability threshold value of 0.75 was used, and pixels with values above 0.75 were kept as glacial lake pixels to minimize misclassification. We vectorized the RF classification outcomes using this threshold of 0.75.

Geoprocessing techniques were employed to enhance the classification. Two classification outputs were considered: one derived from Level I (Predict_S) and the other from the ensembled Level II (Predict_P). A geoprocessing workflow was executed in Predict_S to ensure that the categorized lake pixels in Predict_S were located within 10 m of Predict_P. Subsequently, Predict_S was restricted to encompass Predict_P. The final output

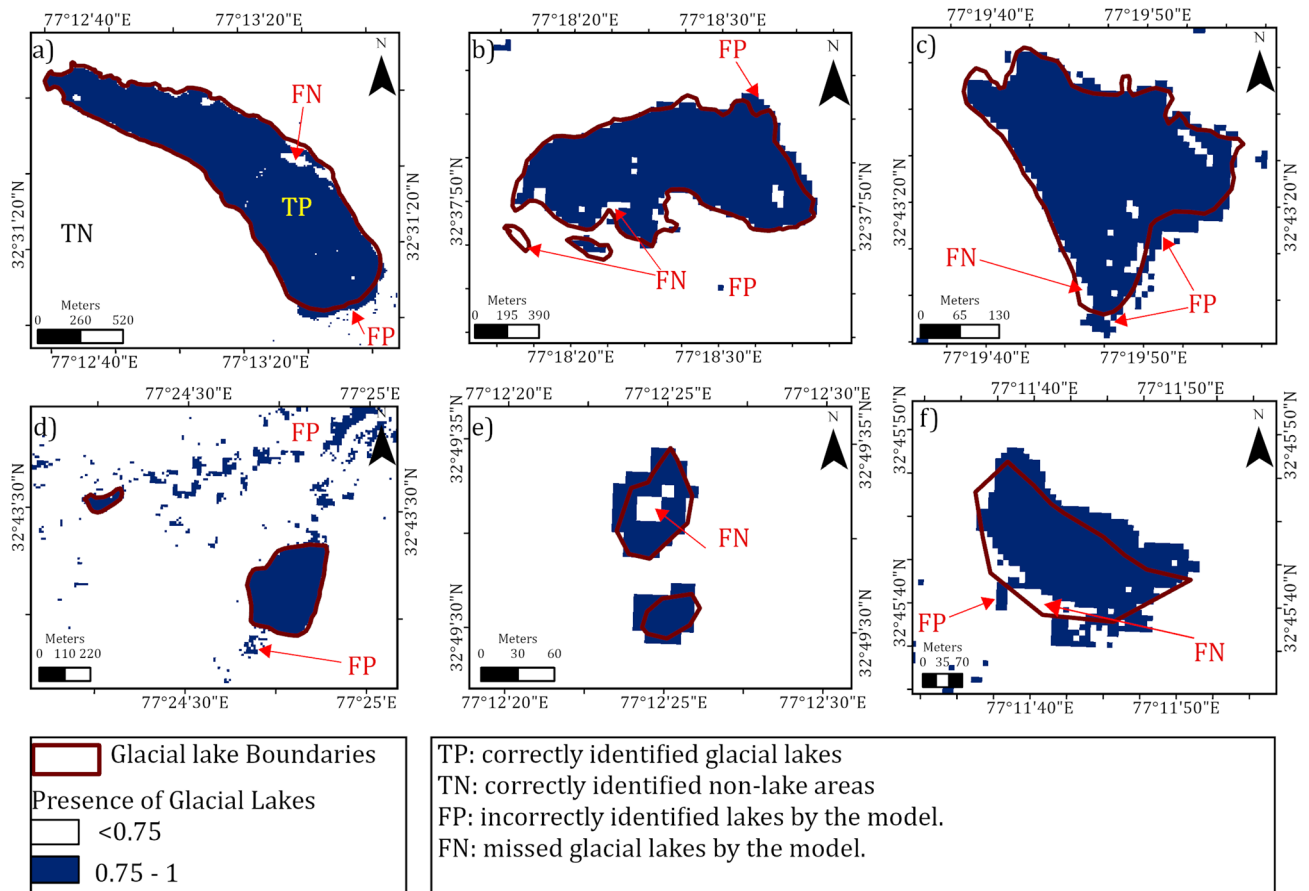


Fig. 6. Classification result using RF for Level I (Predict_S). The figure shows classified raster results obtained using Sentinel-2 optical data and related indices (NDWI, NDSI, NDVI, NDGI). Areas with a probability of glacial lake presence >0.75 are shown in blue, while areas with ≤ 0.75 are depicted in white. Red boundaries indicate actual glacial lake boundaries used for validation.

was subsequently combined using the union operation before being dissolved and smoothed to produce the final shapefile of the RF model- mapped glacial lakes. This strategy facilitated the refinement of lake boundaries and improved the accuracy of the final map.

Accuracy assessment and validation

The RF model has excellent classification accuracy, with 93.69% for Level I (Predict_S), thus reaffirming its efficiency in classifying glacial lakes. The model's AUC-ROC score of 0.984 (Fig. 8) also indicates its high discrimination ability between two classes. For class 1 (glacial lakes), the model performed with a precision of 0.93, recall of 0.89, and F1-score of 0.91, thus maintaining a well-balanced performance between precision and recall. The macro average values of precision, recall, and the F1-score are 0.94, 0.93, and 0.93, respectively, whereas the weighted average is consistently at 0.94. These outcomes affirm the strength of the RF model for both classes.

Similarly, in the case of the Predict_P dataset (Level II), the AUC-ROC score was 0.983 (Fig. 9), verifying the effectiveness of the model in distinguishing classes. The precision, recall, and F1-score for class 1 are 0.94, 0.92, and 0.93, respectively, highlighting its efficiency in classification. The macro and weighted averages increased to 0.95, 0.94 and 0.95, validating both improvement and consistency of the performance. The test accuracy was 94.44%, indicating the high classification ability of the model. Table 2 summarizes the precision, recall, and F1-score metrics for Predict_S and Predict_P, highlighting the classification performance for both the glacial lake and non-glacial lake classes. Moreover, the counts of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) were calculated to increase insights into the performance of the model. These are shown in Table 3, where for Level I (Predict_S) the model recorded 213 TP, 411 TN, 15 FP, and 27 FN, while for Level II (Predict_P) TP increased to 220, FN reduced to 20, and FP dropped to 13 with TN rising slightly to 413.

Additionally, to validate the final results visually, we used high resolution PlanetScope images (3 m) as these images offer enhanced visibility of surface details. The results are depicted in Fig. 10 that shows model derived glacial lake boundaries (red outlines) overlaid on PlanetScope images after applying post-processing steps to the model output. Lakes are precisely identified and delineated in panels Fig. 10a, b, d, h, i, j, k and l, demonstrating the method's ability to capture boundaries with high spatial precision. Few smaller lakes that were not detected

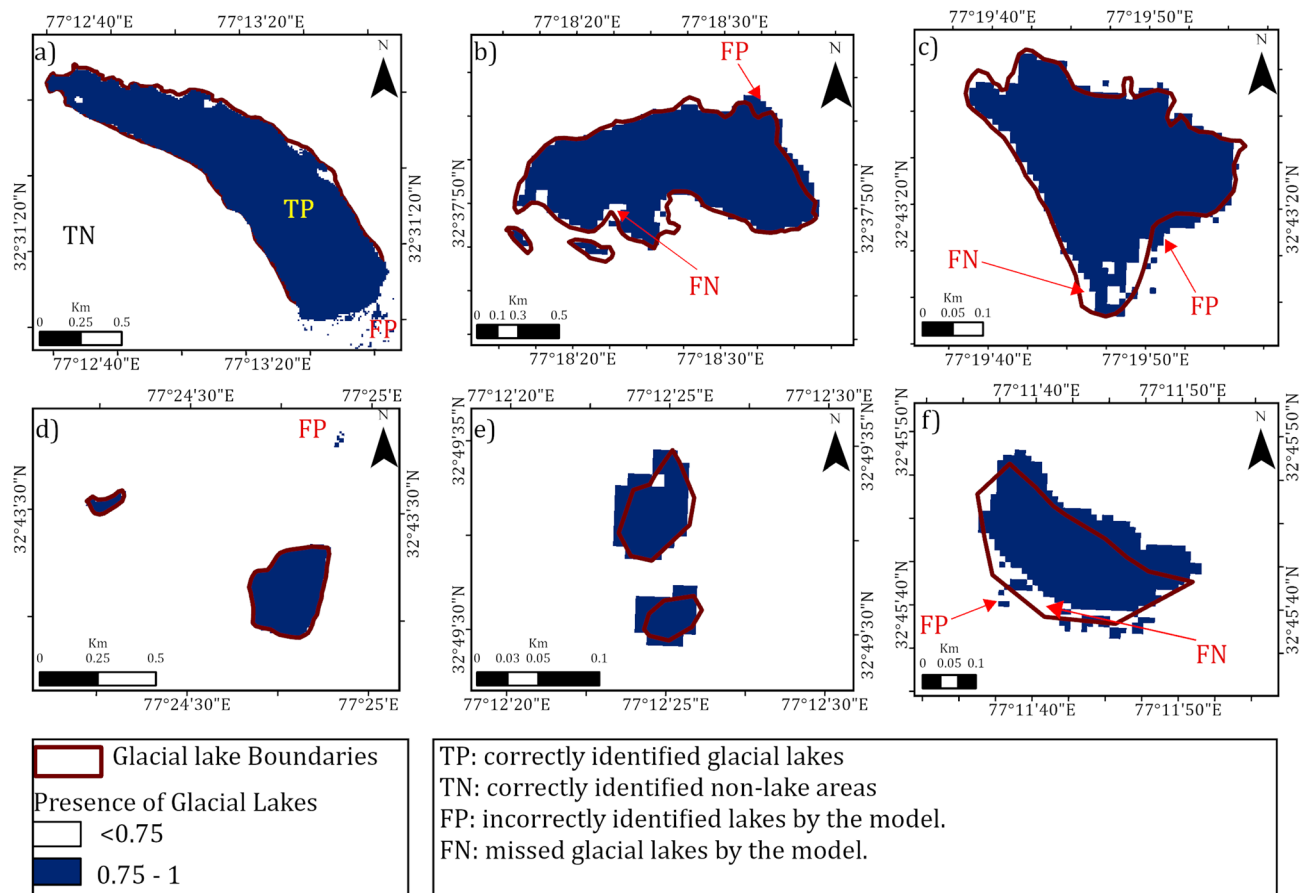


Fig. 7. Classification result using RF for Level II (Predict_P). The figure shows classified raster results obtained using indices derived from Planet data for training and applied to Sentinel-2 bands stack for model execution. Areas with a probability of glacial lake presence >0.75 are shown in blue, while areas with ≤ 0.75 are depicted in white. Red boundaries indicate actual glacial lake boundaries used for validation.

are highlighted in yellow box in Fig. 10f and g, while panels Fig. 10c and e show partial omissions, where portions of lakes were missed. Such errors occurred infrequently and were mainly due to factors such as small lake size, shadow interference, and spectral confusion with snow or debris-covered ice. This visual inspection supports the quantitative evaluation by confirming that the proposed RF approach performs reliably in mapping glacial lakes, with only minimal misclassifications.

Discussion

Proposed method for glacial lake mapping

In the last hundred years, the Himalayas have experienced rising temperatures well above the global average, leading to faster recession of glaciers and growth in glacial lakes^{11,16}. Glacial lakes are valuable sources of fresh water for consumption, hydropower, and agriculture, but when such lakes grow at a fast rate and are structurally unstable, there are considerable risks. Since the formation of glacial lakes are dynamic, frequently updated lake inventories are needed to observe and manage these hazards properly. The accelerated growth of glacial lakes in a warming climate imposes a requirement for accurate and scalable automated mapping solution.

This study provides an operational framework for the detection of glacial lakes in high-altitude regions that addresses critical needs in hazard assessment, monitoring of environmental changes, and resource management, especially in rapidly evolving glacial environments. The method proposed in this study addresses this by outperforming traditional manual and semi-automated techniques. Manual digitization⁴³ is accurate but time-consuming and not possible for large areas and frequent updates. Semi-automatic techniques⁴⁴ usually rely on thresholding and involve substantial human intervention and post-correction that restricts their consistency in diverse terrain. In contrast, our fully automated approach enables reproducible and accurate detection and delineation of glacial lakes in the remote Himalayan region by integrating multi-source remote sensing data as predictor variables within a Random Forest based machine learning framework.

Although the importance of NDWI-green and NIR have been well established^{21,22}, the present study highlights the role of the NDWI-blue as an auxiliary index to delineate lake boundaries with greater accuracy under complex terrain conditions. SAR data are particularly valuable because they can be acquired under all weather conditions, including cloud cover. Consistent with the findings of previous studies^{42,45}, our results

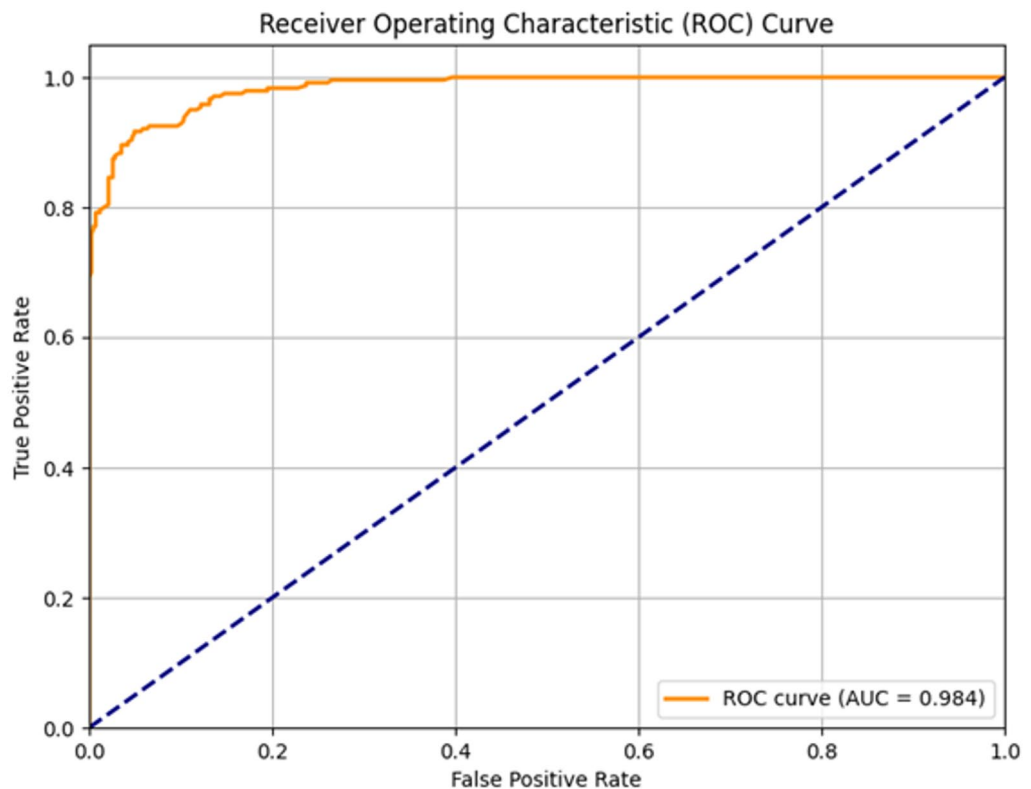


Fig. 8. AUC-ROC Curve for Level I classification. The ROC curve displays the performance of the Random Forest model used for glacial lake detection, with an Area Under the Curve (AUC) of 0.984.

demonstrate that integrating optical and radar data provides synergistic benefits for improved classification performance. The high importance of topographic parameters, particularly elevation and slope, are consistent with the findings of Mustafa et al. 2024²². The RF classifier shows a good performance in classifying glacial lakes. The Sentinel-based classification (Predict_S) model shows an accuracy of 93.69%, precision and recall of 0.93 and 0.89 respectively for glacial lakes. Similarly, the Planet-based classification (Predict_P) model has an overall accuracy of 94.44%. For glacial lakes, the precision and recall are 0.94 and 0.92, respectively, confirming the reliability of the model.

Furthermore, this study utilizes a probability threshold of 0.75 to reduce false positives in the classification output. In high-altitude regions, glacial lakes can resemble features which includes streams, shadows, and supraglacial meltwater which can lead to misclassification. Using a high probability threshold, like 0.75, helps keep only the most confident predictions. This reduces false positives and improves boundary accuracy, especially in complex glacial areas. However, this fixed threshold has a limitation where small or partially shadowed lakes with lower confidence scores may be left out, which leads to false negatives and an underestimation of lake pixel identification. While this approach increases precision, it limits sensitivity.

Although deep learning models like GLNet, investigated by Kaushik et al. 2022¹⁷, perform admirably because of their sophisticated structure, Random Forest has turned out to be an ideal substitute. As a machine learning model, RF yields high accuracy but at a more user-friendly level that may not require familiarity with deep learning⁴⁶. This renders it a practical and effective option for a wide variety of applications.

Comparison with existing literature

As compared with the current literature, especially with the GLNet model introduced by Kaushik et al. 2022²⁰ and the GLaKeMap approach presented by Wangchuk et al. 2020²¹, our Random Forest-based approach shows comparable or better performance in most of the important metrics. For example, GLNet reported F1-scores between 0.70 and 0.91 for different test sites. While GLNet reported correctly mapping both large (>1 km²) and small (< 0.5 km²) glacial lakes without human interference, their F1-score fell to 0.70 at a few locations. Our algorithm, however, produced a stable F1-score of 0.91 (Sentinel) and 0.93 (PlanetScope ensemble), showing excellent performance in detecting even smaller glacial lakes over rugged Himalayan landscapes. Visual verifications ensure that our approach consistently identified smaller lakes that are frequently missed or mislabeled in other methods. A detailed table comparing the proposed methodology with the existing methods is provided in the supplementary material (Supplementary Table 1). The computational requirement was not explicitly stated in their paper²⁰, it is important to note that GLNet is based on a Convolutional Neural Network (CNN) architecture, which, by general understanding, relies heavily on high-end GPU infrastructure for both

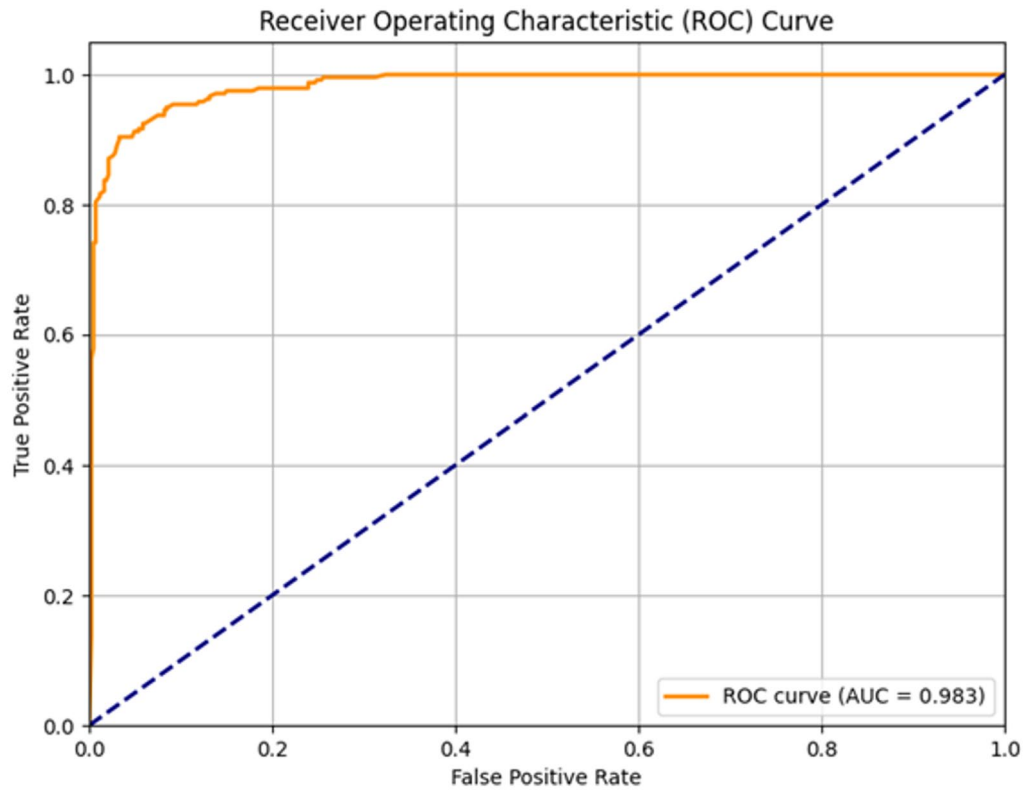


Fig. 9. AUC-ROC Curve for Level II classification based on Planet derived indices. The ROC curve displays the performance of the Random Forest model used, with an Area Under the Curve (AUC) of 0.983.

		Precision	Recall	F1-Score
Level I (Predict_S)	0	0.94	0.97	0.95
	1	0.93	0.89	0.91
	macro avg	0.94	0.93	0.93
	weighted avg	0.94	0.94	0.94
Level II (Predict_P)	0	0.95	0.97	0.96
	1	0.94	0.91	0.93
	macro avg	0.95	0.94	0.95
	weighted avg	0.95	0.95	0.95

Table 2. Model parameter assessment of the models including Precision, recall and F1-score for the two different levels.

	True label	Predicted label	
		0	1
Level I (Predict_S)	0	411	15
	1	27	213
Level II (Predict_P)	0	413	13
	1	20	220

Table 3. Classification performance of the random forest model for two levels, showing true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

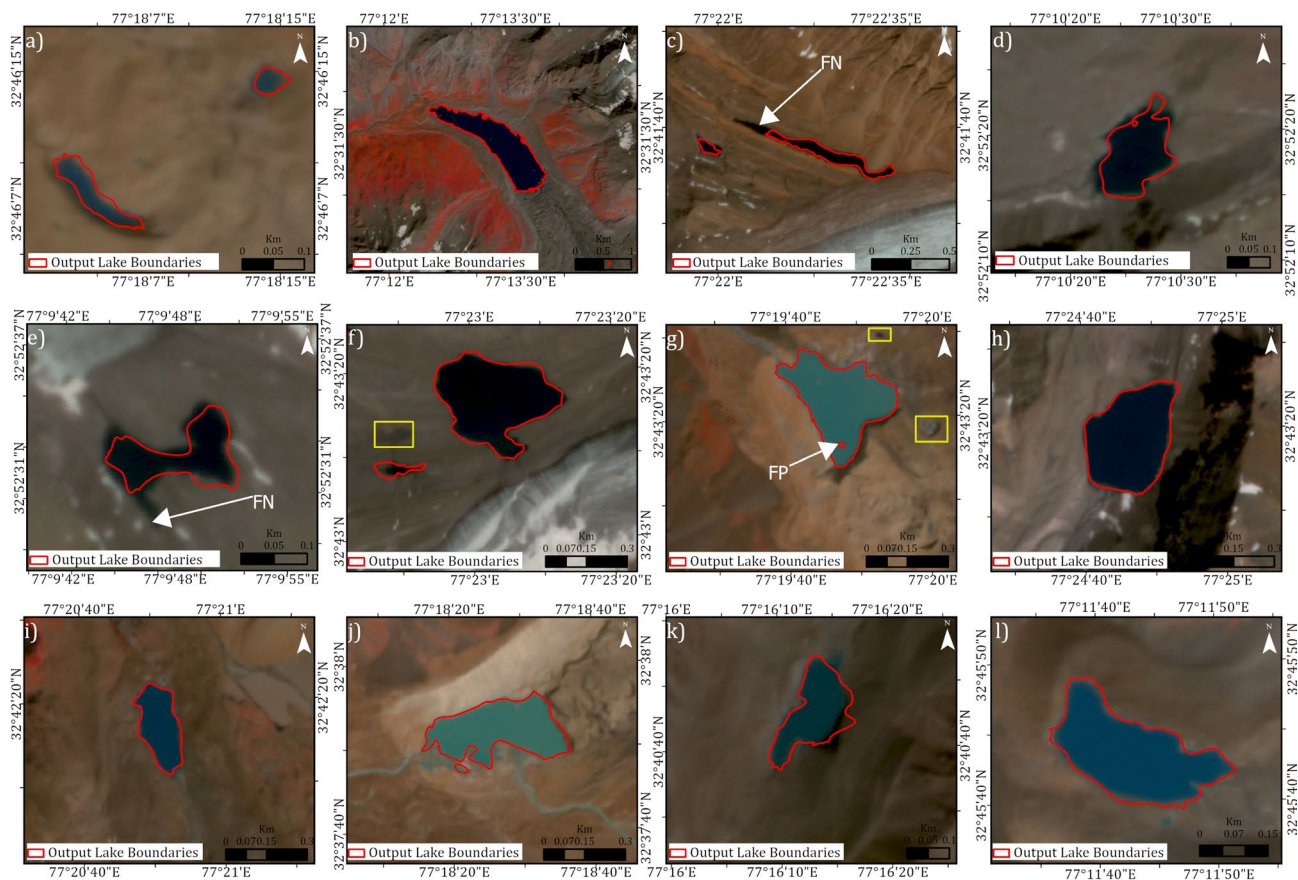


Fig. 10. This figure illustrates the final delineation of glacial lakes, where red boundaries represent the detected lake outlines overlaid on 3 m spatial resolution Planet imagery. The post-processing step significantly improves the accuracy and clarity of lake boundaries by filtering out false positives and enhancing classification precision. Yellow boxes in panels (f) and (g) highlight regions where supraglacial lakes were misclassified due to spectral and textural similarities with nearby wet or saturated glacial surfaces.

training and inference. By contrast, our machine learning based model is computationally light, resource-efficient, and does not require specialized GPU hardware. This makes it more accessible for operational use, especially in scenarios where advanced computational resources are unavailable.

Also, the GLaKeMap technique showed good detection and delineation performance and achieved accuracies of 94.2–96.9%, and mapping accuracies as high as ~97.96%. But their accuracy calculation was done based on comparing the area of manually digitized lakes with that of automatically delineated lakes. This calculation, although simple, is different than the general method of accuracy calculation. As our findings (Fig. 7f) reveal that while automatically extracted lake pixel areas may closely match digitized polygons in size, their spatial position could vary substantially. These positional errors are not accounted for by area-based performance measures and can cause performance scores to be overestimated. Whereas our assessment incorporates pixel-level matching and statistical measures like precision, recall, and F1-score, which provide a more detailed measure of classification performance.

Additionally, our incorporation of multisource data sets such as optical indices (NDWI, NDVI, NDGI), radar backscatter (VV, VH), and topographic variables (slope, elevation, aspect) improved the RF classifier's ability to identify glacial lakes against spectrally similar terrain features like shadows and snow cover. This is especially advantageous in high-altitude, snow-covered, and cloud-ridden environments where single-source data tend to underperform.

Although both GLNet and GLaKeMap contributed meaningfully to glacial lake mapping automation, our approach builds and improves on these with a more flexible, interpretable, and efficient solution. It overcomes limitations of threshold-dependency for image segmentation and area-only accuracy measures and exhibits better performance, specifically in detecting small and spectrally ambiguous lakes. These advancements are critical for operational GLOF risk assessment systems and long-term environmental monitoring in remote Himalayan basins.

Limitation and future works

The results obtained from the classification appear promising after visual validation using high-resolution Planet images. However, we still observed some misclassifications for supra glacial lakes. Despite the use of both

optical and SAR data and rigorous post-processing, detecting supraglacial lakes remains challenging due to their spectral similarity and spatial ambiguity with surrounding glacier surfaces. As seen in Fig. 10f and g (yellow boxes), some of the lakes has not been classified by the model.

Additionally, a fixed probability threshold of 0.75 was applied to reduce false positives. However, this fixed threshold creates a trade-off between precision and recall. Some lakes that are small, partially shadowed or ice-covered may be assigned lower probabilities and thus excluded from the final classification. This might lead to false negatives and slight underestimations of lake areas in specific regions. The threshold significantly improves output quality however future studies could explore adaptive or region-specific thresholds to improve detection accuracy. In future such work can explore adaptive or region-specific thresholds and incorporate temporal information to enhance detection accuracy. Additionally, extending the model for time series analysis could help track lake evolution and support GLOF risk assessments and early-warning systems.

Conclusion

This study contributes to cryospheric remote sensing by presenting a fully automated methodology for glacial lake detection in the Himalayas. It combines remotely sensed data from various sources with Random Forest (RF) classification and SHAP analysis. The model uses the SRTM DEM, Sentinel-1 SAR, Sentinel-2 MSI, and high-resolution PlanetScope data. This setup allows for accurate identification of glacial lakes.

This combination of sensors, particularly the use of PlanetScope for training, is critical for capturing fine-scale glacial features. It also achieves high classification performance across all key metrics, including the AUC-ROC, F1-score, accuracy, precision and recall.

This method offers clear advantages over existing approaches. Unlike GLakeMap, which relies on manually defined thresholds for segmentation, our Random Forest model learns complex decision boundaries directly from the data. This improves adaptability across different terrains and reduces omission errors. Compared with deep learning models such as GLNet, which require large, labelled datasets and substantial computational resources, our model provides comparable classification accuracy, achieving 94.44% overall accuracy at Level II (PlanetScope ensemble), while remaining computationally efficient and accessible to researchers without access to GPU infrastructure. The use of SHAP analysis adds model transparency by highlighting the importance of specific input features that addresses a key issue with black-box models. The false positives from features like streams, shadows, and supraglacial meltwater was minimized by a post-processing step using a probability threshold of 0.75. This filtering retained only high-confidence lake pixels and greatly improved boundary accuracy in complex glacial environments.

Validation against high resolution PlanetScope images confirmed strong agreement with model predictions which supports the reliability of this approach. This method holds significant potential for operational use in hazard assessment such as climate-driven glacial lake expansion and the associated risk of Glacial Lake Outburst Floods (GLOFs). The framework presented here bridges the gap between algorithmic performance and practical usability and contributes to both scientific understanding and real-world glaciological monitoring.

Data availability

All datasets used in this study are publicly available from their respective sources. Sentinel-2 optical imagery and Sentinel-1 SAR data were obtained from the European Space Agency (ESA) and are accessible via Google EarthEngine (GEE) at <https://developers.google.com/earth-engine/datasets>. The Shuttle Radar Topography Mission (SRTM) digital elevation data were also accessed through GEE. PlanetScope high-resolution imagery was used under a research license and is available from Planet Labs (<https://www.planet.com>) upon institutional agreement. Additionally, Google Earth Pro was utilized for high-resolution visual interpretation and geolocation verification. All satellite data (excluding PlanetScope) are freely available for research purposes; access to PlanetScope data may require a license agreement or institutional collaboration. Further the raw datasets can be obtained on reasonable request either from the first author (Ms. Bhawna Pathak, d24091@students.iitmandi.ac.in) or the corresponding author (Dr. Dericks P. Shukla, dericks@iitmandi.ac.in)

Received: 17 May 2025; Accepted: 15 September 2025

Published online: 21 October 2025

References

1. IPCC. *Climate Change 2021: the Physical Science Basis. Contribution of Working Group I To the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, 2021). <https://www.ipcc.ch/report/ar6/wg1/>
2. Pradhan, I. et al. Rock glaciers as proxy for machine learning based debris-covered glacier mapping of Kinnaur District, Himachal Pradesh. *Earth. Surf. Proc. Land.* **49** (11), 3598–3619 (2024).
3. Zemp, M. et al. Global glacier mass changes and their contributions to sea-level rise from 1961 to 2016. *Nature* **568** (7752), 382–386. <https://doi.org/10.1038/s41586-019-1071-0> (2019).
4. Carrivick, J. L. & Tweed, F. S. A global assessment of the societal impacts of glacier outburst floods. *Glob. Planet Change.* **144**, 1–16. <https://doi.org/10.1016/j.gloplacha.2016.07.001> (2016).
5. Richardson, S. D. & Reynolds, J. M. An overview of glacial hazards in the Himalayas. *Quatern. Int.* **65** (66), 31–47 (2000).
6. Clague, J. J. & O'Connor, J. E. Glacier-related outburst floods. In *Encyclopedia of Natural Hazards* (pp. 385–393). Springer. (2015). https://doi.org/10.1007/978-94-007-0264-6_147
7. Carey, M. Living and dying with glaciers: people's historical vulnerability to avalanches and outburst floods in Peru. *Glob. Planet Change.* **47** (2–4), 122–134. <https://doi.org/10.1016/j.gloplacha.2004.10.007> (2005).
8. Harrison, S. et al. Climate change and the global pattern of moraine-dammed glacial lake outburst floods. *Cryosphere* **12** (4), 1195–1209 (2018).
9. Kääb, A., Reynolds, J. M. & Haeberli, W. Glacier and permafrost hazards in high mountains. In *Natural Hazards and Disasters* (225–248). Springer. (2005).
10. Immerzeel, W. W. et al. Climate change will affect the Asian water towers. *Science* **328** (5984), 1382–1385 (2010).

11. Xu, X. et al. Combined use of multi-source satellite imagery and deep learning for automated mapping of glacial lakes in the Bhutan Himalaya. *Sci. Remote Sens.* **10**, 100157 (2024).
12. Zhang, G., Yao, T., Xie, H., Wang, W. & Yang, W. An inventory of glacial lakes in the third pole region and their changes in response to global warming. *Glob. Planet. Change.* **131**, 148–157. <https://doi.org/10.1016/j.gloplacha.2015.05.013> (2015).
13. Bolch, T. et al. The retreat of Himalayan glaciers: climate change impacts and feedbacks. *Nat. Clim. Change.* **9** (12), 993–998 (2019).
14. Tiwari, N., Pradhan, I. P. & Shukla, D. P. Identifying glacier lake dynamics and vulnerable zones for GLOF using AHP. In *AGU fall meeting abstracts* (Vol. 2023, No. 1086, pp. C11E-1086). (2023).
15. Mahanta, K. K., Pradhan, I. P., Gupta, S. K. & Shukla, D. P. Assessing machine learning and statistical methods for rock glacier-based permafrost distribution in Northern Kargil region. *Permafrost Periglac. Process.* **35** (3), 262–277 (2024).
16. Chen, F. et al. Satellite-based remote sensing for monitoring glacial lakes: A review. *Remote Sens. Environ.* **256**, 112322 (2021).
17. Kumar, D. et al. Landslide susceptibility mapping & prediction using support vector machine for Mandakini River Basin, Garhwal Himalaya, India. *Geomorphology* **295**: 115–125. (2017).
18. Pradhan, I. P. et al. Machine learning based high-resolution air temperature modelling from landsat-8, MODIS, and In-Situ measurements with ERA-5 inter-comparison in the data sparse regions of Himachal Pradesh. *Bull. Atmos. Sci. Technol.* **5**, 22. <https://doi.org/10.1007/s42865-024-00085-8> (2024).
19. Pathak, B., Pradhan, P. I., Tiwari, R. K., and Shukla, D. P. Advancing cryospheric monitoring with DEXter air: A satellite-based temperature dataset for the Northwestern Himalaya. (2025). (Accepted).
20. Kaushik, S. et al. Automated mapping of glacial lakes using multisource remote sensing data and deep convolutional neural network. *Int. J. Appl. Earth Obs. Geoinf.* **115**, 103085 (2022).
21. Wangchuk, S. & Bolch, T. Mapping of glacial lakes using Sentinel-1 and Sentinel-2 data and a random forest classifier: strengths and challenges. *Sci. Remote Sens.* **2**, 100008 (2020).
22. Mustafa, H. et al. Integrating multisource data and machine learning for supraglacial lake detection: implications for environmental management and sustainable development goals in high mountainous regions. *J. Environ. Manage.* **370**, 122490 (2024).
23. Hrnjica, B., & Bonacci, O. Lake level prediction using feed forward and recurrent neural networks. *Water Resour. Manage.* **33** (7), 2471–2484 (2019).
24. Chand, P. et al. Shrinking glaciers of the Himachal Himalaya: A critical review. Environmental change in the Himalayan region: Twelve case studies. 89–115 (2019).
25. Chand, P. et al. Reconstructing the pattern of the Bara Shigri glacier fluctuation since the end of the little ice Age, Chandra valley, north-western Himalaya. *Prog. Phys. Geogr.* **41**, 643–675 (2017).
26. Singh, C. & Bharti, V. Rainfall characteristics over the Northwest Himalayan region. Remote Sensing of Northwest Himalayan Ecosystems. Singapore: Springer Singapore, 171–194. (2018).
27. Pradhan, I. P. & Shukla, D. P. Biennial analysis of probable permafrost distribution for Kullu District, north-west Himalaya using Landsat 8 satellite data. *Land. Degrad. Dev.* **35** (1), 360–377 (2024).
28. Bhambri, R. Glacier mapping: a review with special reference to the Indian Himalayas. *Prog. Phys. Geogr.* **33** (5), 672–704 (2009).
29. Bhambri, R. et al. Glacier lake inventory of Himachal Pradesh. *Himal. Geol.* **39** (1), 1–32 (2018).
30. Randhawa, S. S., Thakur, N. and Chauhan, M. Monitoring of glacial lakes/water bodies in Satluj catchment using RS & GIS techniques during the year 2023 (Report No. SCSTE/HPSCCC/Const/SJVNL/2023). State Centre on Climate Change, H.P. Council for Science, Technology & Environment. Submitted to Satluj Jal Vidyut Nigam Ltd. (SJVNL). (2024).
31. Jiang, D. et al. Glacial lake mapping using remote sensing Geo-Foundation model. *Int. J. Appl. Earth Obs. Geoinform.* **136**, 104371. (2025).
32. Das, U. The monitoring Gangotri glacier, using geospatial technique.
33. Mahanta, K. K., Pradhan, I. P., Gupta, S. K., Shukla, D. P. & Gupta, N. Monitoring the spatial and temporal patterns of rock glacier melt induced subsidence using Multi-Temporal Interferometric Synthetic Aperture Radar techniques. *45th COSPAR Scientific Assembly. Held 13–21 July*, **45**, 47. (2024).
34. Niraj, K. C., Gupta, S. K. & Shukla, D. P. Kotrupi landslide deformation study in non-urban area using DInSAR and MTInSAR techniques on Sentinel-1 SAR data. *Adv. Space Res.* **70** (12), 3878–3891 (2022).
35. Mahanta, K. K., Pradhan, I. P., Dhiman, N., Singh, A. & Shukla, D. P. Investigating the first case of permafrost degraded subsidence in Lahaul & Spiti region of Tethyan Himalayas. *Sci. Rep.* **15** (1), 19262 (2025).
36. Singh, A. et al. Ensembled transfer learning approach for error reduction in landslide susceptibility mapping of the data scarce region. *Sci. Rep.* **14** (1), 29060 (2024).
37. Aceña, V. et al. Minimally overfitted learners: A general framework for ensemble learning. *Knowl. Based Syst.* **254**, 109669 (2022).
38. Singh, A. et al. Improving ML-based landslide susceptibility using ensemble method for sample selection: a case study of Kangra district in Himachal Pradesh, India. *Environ. Sci. Pollut. Res.* <https://doi.org/10.1007/s11356-024-34726-4> (2024).
39. Mahanta, K. et al. Assessing machine learning and statistical methods for rock glacier-based permafrost distribution in Northern Kargil region. *Permafrost Periglac. Process.* **35** (3), 262–277 (2024).
40. Gupta, S. K., and Shukla, D. P. Handling data imbalance in machine learning based landslide susceptibility mapping: A case study of Mandakini River Basin, North-Western Himalayas. *Landslides* **20** (5), 933–949 (2023).
41. Zhou, Y. et al. Open surface water mapping algorithms: A comparison of water-related spectral indices and sensors. *Water* **9** (4), 256 (2017).
42. Shen, G. et al. Water body mapping using long time series Sentinel-1 SAR data in Poyang Lake. *Water* **14** (12), 1902 (2022).
43. Zhang, G. et al. An inventory of glacial lakes in the third pole region and their changes in response to global warming. *Glob. Planet. Change.* **131**, 148–157 (2015).
44. Yan, D. et al. Improved landsat-based water and snow indices for extracting lake and snow cover/glacier in the Tibetan plateau. *Water* **12** (5), 1339 (2020).
45. Wu, R. et al. A deep learning method for mapping glacial lakes from the combined use of synthetic-aperture radar and optical satellite images. *Remote Sens.* **12**, 4020 (2020).
46. Ashuli, Adaphro, et al. "Integration of active tectonic index and geomorphic parameters for landslides susceptibility mapping in the Barak River basin." *Natural Hazards* 1–31 (2025).

Acknowledgements

The authors are grateful to IIT Mandi in Kamand, Himachal Pradesh, India, for providing this opportunity and technical tools to conduct this research. Additionally, we would like to thank the Survey of India for providing shapefiles. Authors acknowledge various data sources from where data is downloaded. All data used for this study can be freely downloaded from public websites. The Sentinel 1 SAR data are available at <https://sentinel.esa.int/web/sentinel/missions/sentinel-1/data-products>, the Sentinel 2 MSI data are available at <https://sentinel.esa.int/web/sentinel/sentinel-data-access/sentinel-products/sentinel-2-data-products>. Also, the authors would like to acknowledge the USGS, the Land Processes Distributed Active Archive Center (LP DAAC), Planet Images and Google Earth and SRTM for providing the 30-m DEM to carry out this work.

Author contributions

Bhawna Pathak: Conceptualization, Methodology, Data curation, Analysis, Validation, Writing—original draft. Ankit Singh: Methodology, Analysis, Validation, Writing—review & editing. Dr. Reet Kamal Tewari: Formal analysis, Validation, Writing—review & editing. Dr. Dericks P. Shukla: Conceptualization, Methodology, Software, Formal analysis, Validation, Writing—review & editing.

Declarations

Competing interests

The authors declare that they have no competing interests.

Consent for publication

All authors give their consent to publish this work in whole.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-20434-7>.

Correspondence and requests for materials should be addressed to D.P.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025