Measuring the surface area of glaciers in the Himalayas

Term Project, Spatial Statistics and Machine Learning
Jurrian van de Kraats, Loredana Gliga & Sander Rustenburg

1. Introduction

Glaciers form an integral part of Earth's ecosystem, with glacial ice making up roughly ten percent of the planet's total land surface area [1]. They play an important role in many landscapes and are a defining factor in the ecological landscape of their surrounding areas [2]. While most of the planet's glaciers can be found on the vast ice sheets of Greenland and Antarctica, the Himalayas harbor the greatest volumes of ice outside the polar regions, earning the mountain range the nickname 'The Third Pole' [3]. However, the Himalayan glaciers are at risk due to the effects of climate change, resulting in rapid melting [4].

Although this process has been researched extensively, continuous monitoring of this process is necessary to ensure an accurate overview of glacial behaviors and the impact of climate change on the region's ecosystem [3]. This monitoring is commonly conducted through remote sensing [5]. This data can be used to calculate indices that enable glacial mapping and classification [6]. Common indices are the Normalized Difference Snow Index (NDSI), the Normalized Difference Water Index (NDWI) and the Normalized Difference Glacial Index (NDGI) [7] [8]. While these indices have been previously used for the classification of glaciers, there is currently no consensus on which indices are best suited to identify glacial patterns [9]. This makes it difficult to create a uniform algorithm to analyze the Himalayan glaciers on a grander scale.

For this study the following question was researched: *'How do the NDSI, NDWI and NDGI indices compare to each other when used for mapping glacial extent in the Sagarmatha and the Makalu Barun National Park through machine learning?'* Comparing the performance of the three different indices on a simple machine learning algorithm could determine what index fits such an algorithm best. This will enable future research to map the glacial extent in a faster manner over a larger area in the Himalaya.

2. Methodology

Study area

For this study The Sagarmatha and the Makalu Barun National were chosen (Figure 1). These national parks contain some of the world's tallest mountains and well-known Himalayan glaciers. Previous mappings of their glacial extent can help in training the model for ranges outside of this park. Clear weather windows before and after monsoon also provide a good basis for a remote sensing problem.

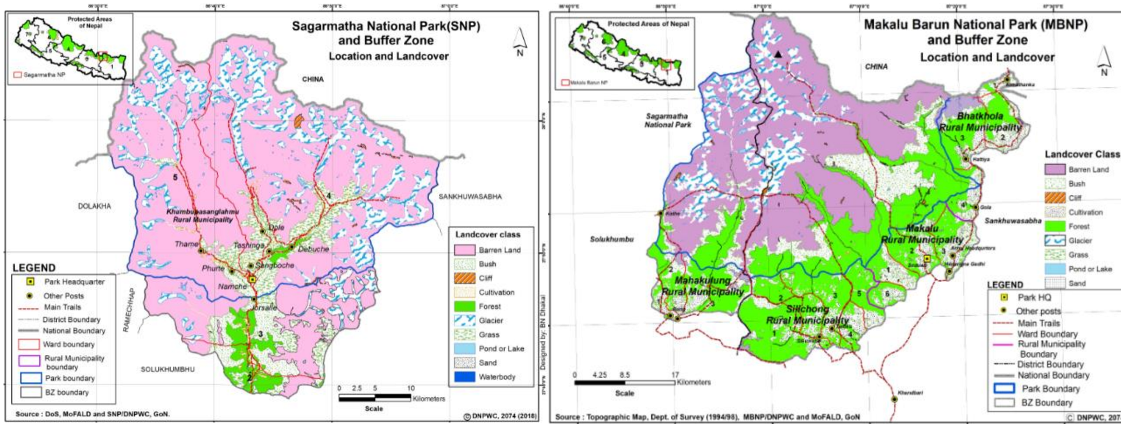


Figure 1: *Sagarmatha National Park (left)* [10] *and Makalu Barun National Park (right)* [11]

Data collection

Figure 2 shows the pipeline used to collect, process, and analyze the data. Data from the Landsat Collection 2 Level-2 satellites was collected through the Planetary Computer [12]. These data ranged from the first of September until the first of November in intervals of three years from 1983 until 2022. The year 2009 was collected to train the model and the year 1999 was collected for validation. In these last two years the glaciers in Sagarmatha National Park and Makalu Barun National Park were extensively mapped [13], providing sufficient data to create classifications.

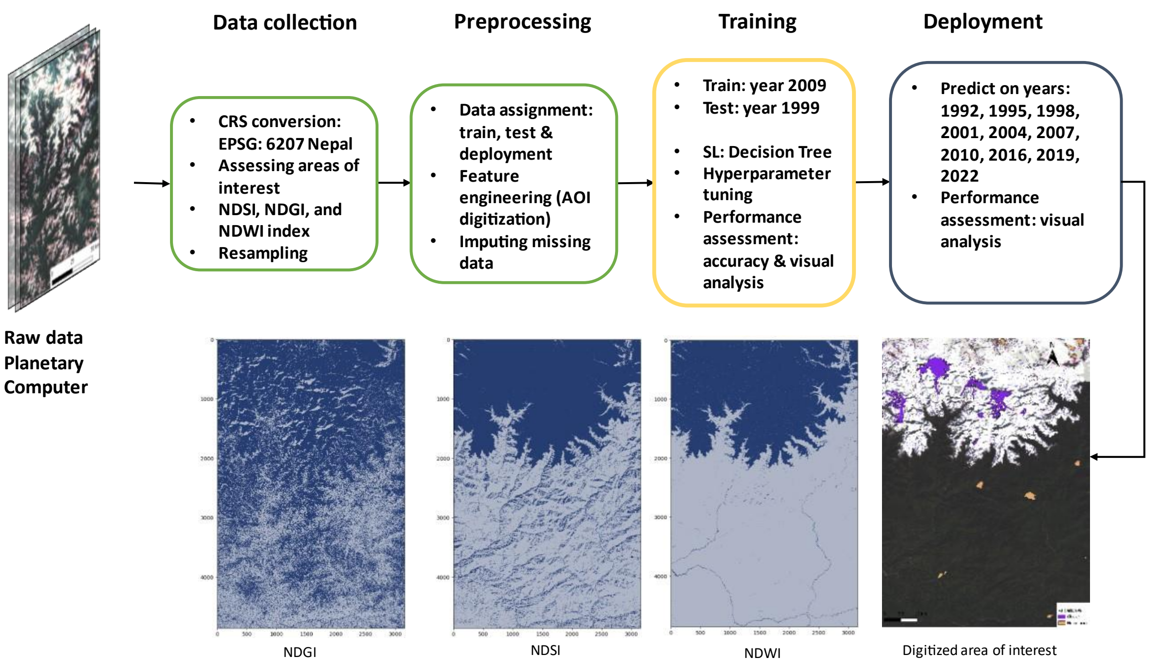


Figure 2: *Project pipeline*

In total 56 measurements were captured with a cloud cover of less than 20 percent (Table 1). The data was converted to the Nepal CRS (EPSG:6207) and clipped to a smaller spatial extent, which covers the two national parks. These measurements were resampled in three datasets consisting of the yearly average NDSI, NDWI and NDGI through the formulas given in Figure 3. Averaging the bands also reduced to noise formed by snowfall and clouds.

Year	**Items**
Training data	
2009	7
Validation data	
1999	2
Deployment Data	
1983	0
1986	0
1989	0
1992	3
1995	5
1998	3
2001	2
2004	6
2007	4
2010	4
2013	5
2016	5
2019	4
2022	6

Table 1: *Measurement moments per interval*

$$NDSI = \frac{Green - SWIR}{Green + SWIR} \quad (1)$$

$$NDGI = \frac{Green - RED}{Green + RED} \quad (2)$$

$$NDWI = \frac{Green - NIR}{Green + NIR} \quad (3)$$

Figure 3: *The bands used to create the NDSI, NDGI, and NDWI indices*

Model creation

For the train and validation data, polygons were mapped from the Ngojumba, Khumbu, Makalu Southeast, Hungu, Merla, Lumsamba, Ripimo Shar, Rolwaling, Barun, Nuptse and Mera Peak glaciers [13]. Non-glacial objects like the parts of the Dnudh Kosi and the Arun rivers, lakes, forests and rock formations were also mapped. A Decision Tree (DT) model was made and hyper tuned using grid search and 5-fold Cross Validation. Parameters used were tree depth, minimum samples per split and minimum samples per leaf [14]. Table 3 shows the best parameters for each index. A DT was chosen because, from the machine learning models with a low computational load, it was assumed to be the better fit to the distribution of the data. The model performance on the validation dataset was assessed though visual inspection and performance measurement statistics.

Decision Tree best parameters

	NDWI	NDSI	NDGI
Tree depth	2	2	4
Min sample per split	2	2	2
Min sample per leaf	1	1	1

Table 3: *The best DT parameters for each index*

Deployment

Missing data in the imagery in the years 2007 until 2019 was imputed through the iterative imputer from Scikit-Learn [15]. The DT model was used to classify glacial and non-glacial objects in the deployment data. The predicted areas were compared between the three indices to identify any differences.

4. Results

Figure 4 shows the visualized results from the different indices on the validation data from 1999. Table 3 shows the DT performance. The NDWI scores best on *accuracy* (.89), *precision (glacial)* (.88) and *recall (non-glacial)* (.53). The NDSI has the highest *precision (non-glacial)* (1).

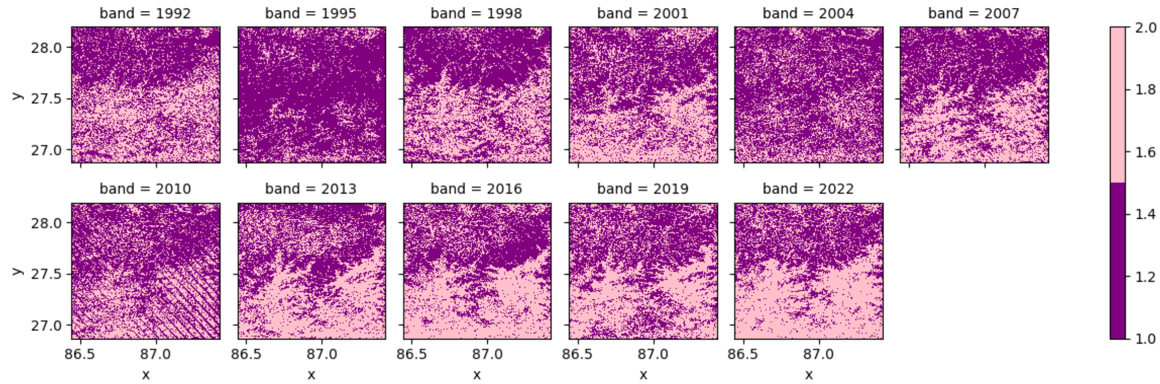
Decision Tree performance

	NDWI	NDSI	NDGI
Accuracy	0.89	0.87	0.81
Precision (glacial)	0.88	0.85	0.85
Precision (non-glacial)	0.98	1	0.61
Recall (glacial)	1	1	0.92
Recall (non-glacial)	0.53	0.43	0.45

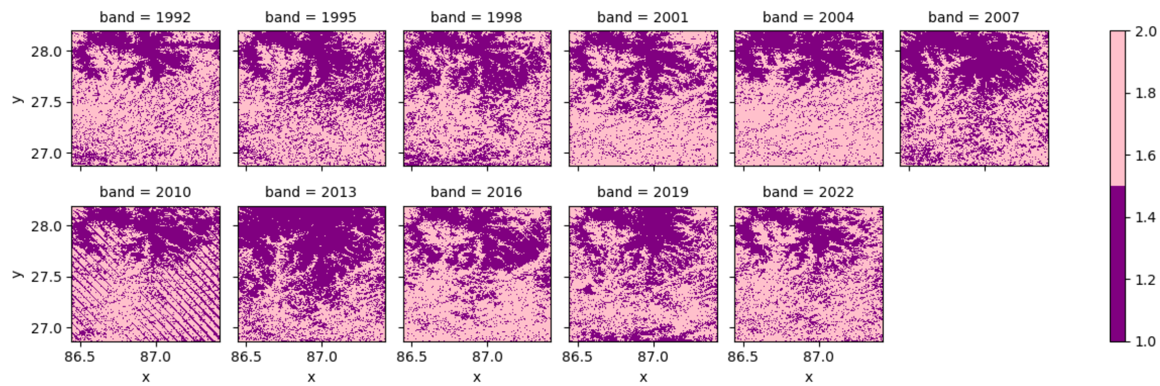
Table 3: Decision Tree performance matrix

Figure 5 shows the visualized results from the different indices over time. In many years the NDGI falsely classifies non-glacial objects as glaciers. The NDSI forms a better fit, but this index seems to have difficulty with cloud cover and heavier snowfall. The NDSI seems to be an adequate index for classification in tree models. However, a stronger model could better assess the nuances between snow, clouds, shadows and glaciers. The NDWI shows clear results over a longer time. It clearly differentiates glacial objects from non-glacial objects, even differentiating glaciers from lakes. The year 2013 indicates that the NDWI does have a harder time with heavy snowfall. Smaller waterbodies like rivers cause some false classifications when they have bigger water density.

NDGI



NDSI



NDWI

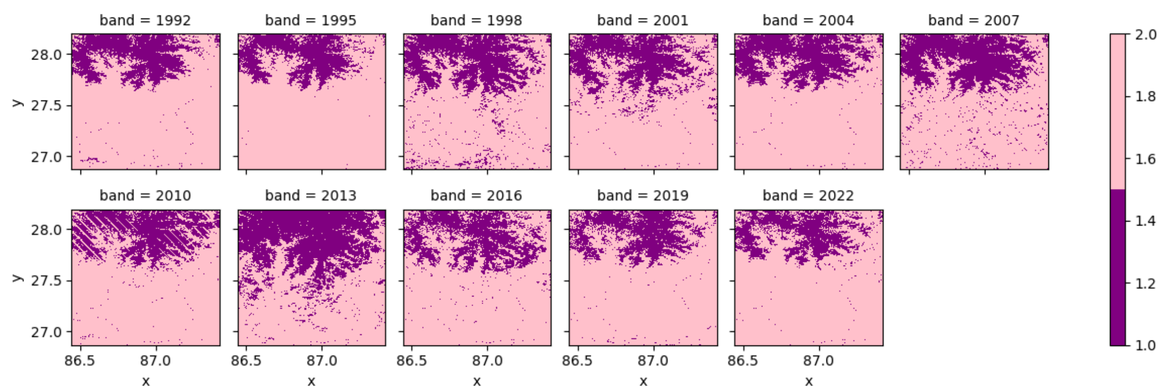


Figure 5: *The NDGI, NDSI and NDWI glacial classifications over a 30-year period (1 is glacial, 2 is non-glacial)*

For assessing the differences in glacial extent, the results of the three indices were plotted as the difference between 1992 and 2022. In Figure 6, the red areas represent the areas that used to be glaciers in 1992 but not in 2009. There is a clear difference between the NDGI, NDSI and NDWI. The NDGI and NDSI have a large decrease in glacial extent, while the NDWI shows a smaller difference. All three results showcase that in the western parts of the study area, the glacial extent has decreased the most. While in the northern parts of the study area there is a small increase in glacial extent.

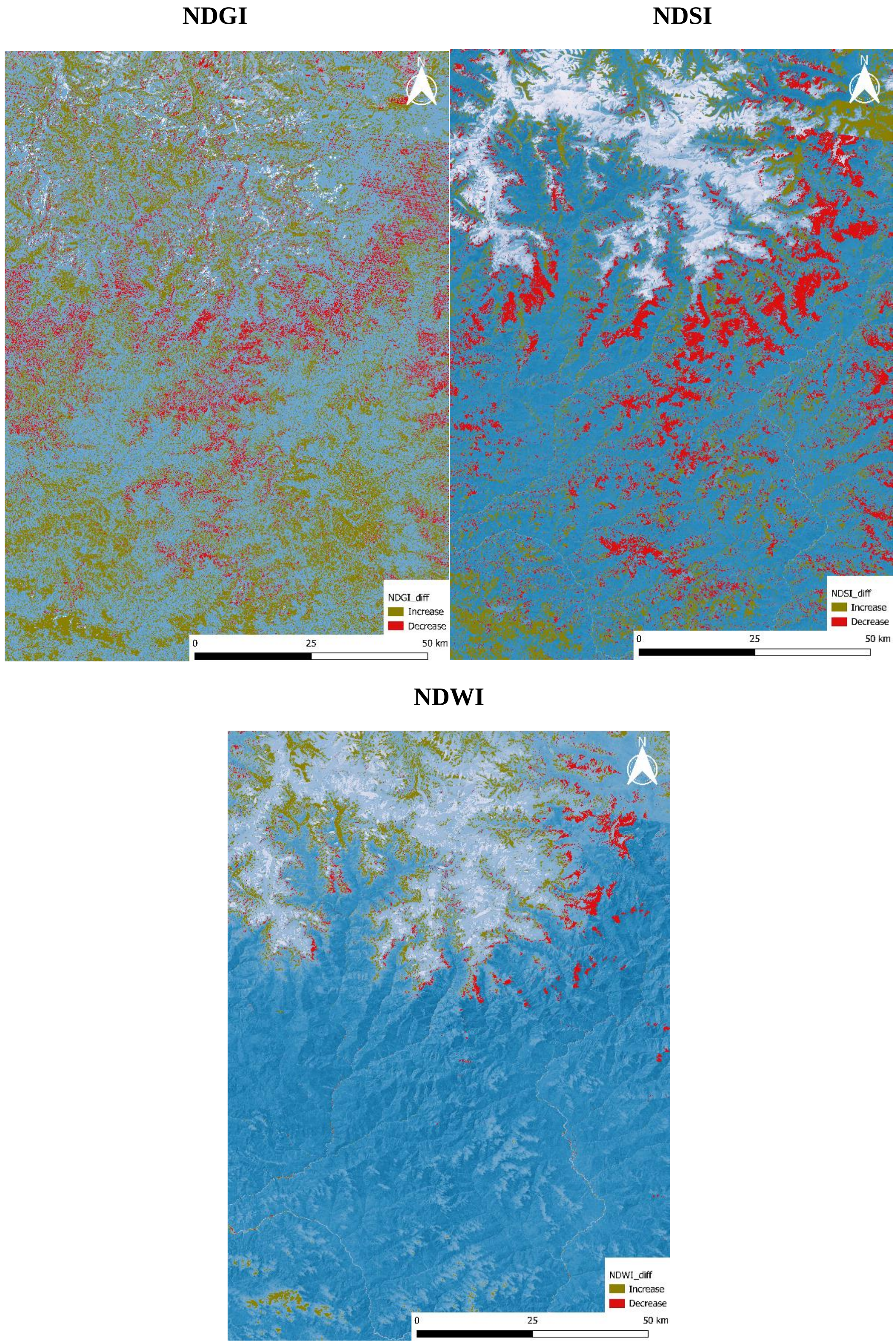


Figure 6: *Difference in glacial extent between 1992 and 2022*

5. Discussion and Conclusion

The NDWI seems to be the best fit for a machine learning problem. We assume that the structure of a tree model has a good fit for the NDWI because it enables a clear differentiating between low water bodies, like forests, snow and small rivers, bigger water bodies like glaciers and large water bodies like lakes.

The results show that the NDGI forms a bad data source when used in a machine learning problem. This is caused by the skewed distribution of this index, giving very high values for ice sheets, but very low values for non-ice objects. This is partly due to the NDGI's nature as a cut off value [16]. A clear cut-off value gives less nuanced results than when used in a classification problem, that relies on its distribution.

Cloud masking techniques could help in creating a cleaner dataset with less noise. This could benefit for example the NDSI model. The NDSI and the NDGI also seem to have problems identifying glacial tongues. These formations of rock and ice differentiate spectrally in a clear way from ice sheets on these indices. Combining the three indices into a, weighted, new index could provide the clear boundaries and the inclusion of glacial tongues from the NDWI with the clear ice definition from the NDGI and NDSI.

Furthermore, some years the NDWI also classifies the rivers in the study area. When these rivers become larger water bodies, they obtain similar spectral values as the glaciers. Because the water from these rivers comes from the glaciers, the NDWI could also be used to classify the difference in melting water over time, which could in turn explain the glacial extent.

Finally, the mapping of the glacial extent shows an interesting development in which some parts of the glaciers in the national parks seem to have increased. In this research we cannot state for certain if this is due to an actual increase, or because of false classifications. A possible recommendation for future research could be to deploy for example the NDWI model on fully indexed glacial polygons and view whether they classify the same increase over time.

In conclusion: the three bands give clearly differing results when used for classification in a tree model. The NDGI forms a bad data source when used in a simple classification algorithm. The NDSI provides adequate results but has a hard time when used with complex data. The NDWI seems to fit the model best and gives the best classifications over time.

6. Bibliography

[1] K. M. Cuffey and W. S. B. Paterson, The Physics of Glaciers, Amsterdam: Elsevier, 2010.

[2] D. Benn and D. J. A. Evans, Glaciers and Glaciation, London: Routledge, 2013.

[3] T. Bolsch, A. Kulkarni, A. Kääb, C. Huggel, F. Paul, J. G. Cogley, H. Frey, J. S. Kargel, K. Fujita, M. Scheel, S. Bajracharya and M. Stoffel, "The State and Fate of Himalayan Glaciers," *Science,* pp. 310-314, 2012.

[4] B. Talukder, R. Matthew, G. W. van Loon, M. J. Bunch, K. W. Hipel and J. Orbinski, "Melting of Himalayan glaciers and planetary health," Elsevier, Amsterdam, 2021.

[5] F. Paul, N. Barrand, S. Baumann, E. Berthier, T. Bolch, K. Casey and H. Frey, "On the accuracy of glacier outlines derived from remote-sensing data," *Annals of Glaciology,* vol. 54, no. 63, pp. 171-182, 2013.

[6] A. Jabbar, A. A. Othman, B. Merkel and S. R. Hasan, "Change detection of glaciers and snow cover and temperature using remote sensing and GIS: A case study of the Upper Indus Basin, Pakistan," Elsevier, Amsterdam, 2020.

[7] J. Florath, S. Keller, R. Abarca-del-Rio, S. Hinz, G. Staub and M. Weinmann, "Glacier Monitoring Based on Multi-Spectral and Multi-Temporal Satellite Data: A Case Study for Classification with Respect to Different Snow and Ice Types," *Remote Sensing,* vol. 13, no. 4, 2022.

[8] V. Bazilova and A. Kääb, "Mapping Area Changes of Glacial Lakes Using Stacks of Optical Satellite Images," *Remote Sens,* vol. 14, no. 23, 2022.

[9] S. Yan, L. Xu, G. Yu, L. Yang, W. Yun, D. Zhu, S. Ye and X. Yao, "Glacier classification from Sentinel-2 imagery using spatial-spectral attention convolutional model," Elsevier, Amsterdam, 2021.

[10] Ministry of Forests and Environment, Nepal, "Sagarmatha National Park," [Online]. Available: https://dnpwc.gov.np/en/conservation-area-detail/72/.

[11] Ministry of Forests and Environment, Nepal, "Makalu Barun National Park," [Online]. Available: https://dnpwc.gov.np/en/conservation-area-detail/75/.

[12] M. McFarland, R. Emanuele, D. Morris and T. Augspurger, "microsoft/PlanetaryComputer: October 2022," Zenodo, 2022.

[13] GLIMS Consortium, "GLIMS Glacier Database, Version 1 [Analysis_IDs 804440--939214, Sakai, Akiko (submitter); Sakai, Akiko (analyst(s))].," Boulder, Colorado, USA: NASA National Snow and Ice Data Center Distributed Active Archive Center, http://dx.doi.org/10.7265/N5V98602, 2005.

[14] Pedregosa et al, "Scikit-learn: Machine Learning in Python , volume 12," in *Journal of Machine Learning Research*, 2011.

[15] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel and P. P., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011.

[16] A. Shukla and I. Ali, "A hierarchical knowledge-based classification for glacier terrain mapping: a case study from Kolahoi Glacier, Kashmir Himalaya," *Cambridge University Press*, vol. 57, pp. 1-10, 2016.