



Research Article

Semi-automatic classification for rapid delineation of the geohazard-prone areas using Sentinel-2 satellite imagery



Karma Tempa¹  · Komal Raj Aryal² 

Received: 7 December 2021 / Accepted: 25 March 2022

Published online: 09 April 2022

© The Author(s) 2022 [OPEN](#)

Abstract

The study of land use land cover has become increasingly significant with the availability of remote sensing data. The main objective of this study is to delineate geohazard-prone areas using semi-automatic classification technique and Sentinel-2 satellite imagery in Bhutan. An open-source, semi-automatic classification plugin tools in QGIS software enabled efficient and rapid conduct of land cover classification. Band sets 2-8, 8A, and 11-12 are utilized and the virtual colour composites have been used for the clustering and creation of training samples or regions of interest. An iterative self-organizing data analysis technique is used for clustering and the image is classified by a minimum distance algorithm in the unsupervised classification. The Random Forest (RF) classifier is used for the supervised classification. The unsupervised classification shows an overall accuracy of 85.47% (Kappa coefficient = 0.71) and the RF classifier resulted in an accuracy of 92.62% (Kappa coefficient = 0.86). A comparison of the classification shows a higher overall accuracy of the RF classifier with an improvement of 7.15%. The study highlights 35.59% (512,100 m²) of the study area under the geohazard-prone area. The study also overlaid the major landslide polygons to roughly validate the landslide hazards.

Article highlights

- (a) Semi-automatic classification technique was applied to delineate the geohazard-prone area in the heterogeneous region of Bhutan Himalaya.
- (b) Unsupervised and supervised classification technique were used to perform land cover classification using the semi-automatic classification plugin (SCP).
- (c) The Random Forest classifier predicted higher accuracy and the application is rapid and efficient compared to the unsupervised classification.

Keywords Semi-automatic classification · Sentinel-2 · ISODATA · Random Forest · Geohazard · Bhutan

✉ Karma Tempa, karmatempa.cst@rub.edu.bt; Komal Raj Aryal, karyal@ra.ac.ae | ¹Civil Engineering Department, College of Science and Technology, Royal University of Bhutan, 21101 Phuentsholing, Bhutan. ²Faculty of Resilience, Rabdan Academy, Abu Dhabi, United Arab Emirates.



SN Applied Sciences

(2022) 4:141

| <https://doi.org/10.1007/s42452-022-05028-6>

1 Introduction

Geohazard disasters such as landslides, riverine floods, and flash floods accompanied by debris and mudflow are common in the Bhutan Himalayan regions. Landslide disaster incidents predominantly cause land degradation and socio-economic impacts around the world [1]. Similarly, many research studies have identified flood as one of the most destructive and costliest natural disasters [2, 3]. There is undoubtedly a great demand for geohazard studies using the RS technique as it is recognized as a powerful tool for extracting information about the earth's surface with spatiotemporal resolution [4]. Hence, geohazard studies are extensively carried out through remote sensing and GIS techniques globally. GIS-based landslide studies are widely undertaken using environmental factors under different regional settings at different scales for landslide susceptibility mapping [5–9]. In recent times, artificial intelligence (AI)-based machine learning and deep learning have become the state-of-the-art technology in delivering high-level performance, especially in developing landslide susceptibility maps [10–12]. More recently, there has been significant focus on the methodological process to accurately perform the image classification under the domain of AI [13]. The machine learning algorithms (MLAs), a sub-domain of AI is also widely used, through statistical or non-probabilistic approaches in various classification techniques to automatically extract information from RS data [14]. In particular, classification mapping has become an area of interest in light of the availability of medium to high resolution multispectral remote sensing imagery, which has relatively contributed to the delineation of disaster-prone areas [15]. The present study uses freely available Sentinel-2A imagery. Since 2015, Sentinel-2A multispectral Imager (MSI) has been providing 13 spectral bands: four bands at 10 m spatial resolution, six bands at 20 m spatial resolution, and three bands at 60 m spatial resolution. Among these, the spatial resolution of 10 m can be used to improve the spatial resolution of other 20 or 60 m bands [16, 17]. The Sentinel-2 spatial and temporal resolution offers new potential for generating accurate datasets in ready-to-use vector format [18]. RS data contain valuable knowledge on the earth's surface that can be coupled with image classification techniques to extract the LULC information [19]. In addition, with the advent of image processing technologies using GIS, image classification methods have made a significant advancement in the conduct of LULC mapping [20, 21] including assessment and identification of LULC change [22]. Undoubtedly, the LULC maps are some of the most important documents that provide information for various applications in land use policy development, agricultural monitoring, urban

planning, ecosystem services, conservation, and LULC dynamic assessment [23]. It has been also used for an integrated risk and vulnerability assessment in many parts of the world.

Automated or semi-automated methods of remote sensing analysis, such as pixel-based and object-based image analysis, differ in the way they depend on individual pixels or the way homogeneous pixels are converted into objects [16, 24, 25]. The performance of these remote sensing techniques often depends on the geographical characteristics of the study area [26]. The automated approaches are simple [27, 28], and cost-effective without the need for physical interventions [29]. Some of the research studies used the SCP for application in various land cover classification, e.g., [24, 30]. The SCP is a Python tool for downloading and processing remote sensing images [31], which was developed in the QGIS environment [32]. The SCP plugin works semi-automatically allowing the user to download multiple RS products associated with satellite imagery pre-processing and perform unsupervised and supervised classification automatically by setting some parameters in the user interface. Minimum distance, maximum likelihood, and spectral angle mapping classification algorithms are available in the SCP. The SCP plugin uses a traditional pixel-based image classification approach that computes the spectral signatures and compares the random nearest neighbourhood pixels. The unsupervised classification implements the ISODATA or K-means algorithms for clustering of the band sets. RF, a machine learning algorithm is also available to perform supervised classification. RF classifier use a similar but improved method of bootstrapping as bagging with ensembles of tree-type classifiers of potentially high accuracies and are computationally less intensive [33]. Hence, the SCP provides an efficient platform to be used by a wide range of professionals who could benefit from remote sensing analysis.

At present, use of RS data on studies of geohazards by LULC mapping in the Bhutanese Himalayas is rare, and notably an assessment of exposure and vulnerability to geohazards associated with multiple natural hazards is lacking. Prior to this study, relatively few studies on landslide hazard based on causal factors were conducted [34, 35]. Therefore, in the present study, an attempt was made to apply the semi-automatic classification technique to the heterogeneous geographical settings of different land use compositions characterized by active geohazard activities. The geohazards exposed in the study area are mainly the result of landslides and, flood associated activities resulting in debris and mudslides, and subsequent sediment deposition. Therefore, this paper accomplishes the geohazard delineation mapping using the GIS-based SCP plugin and the Sentinel-2 satellite imagery. Application of the SCP first performed rapid land cover classification

using unsupervised and supervised classification methods, followed by an advanced accuracy assessment and validation. A comparison between the classification algorithms and their accuracy is also conducted.

The main features of this study consist of six main sections, beginning with an introduction that provides information on literature research and describes the aim and motivation of the work. In Sect. 2, we introduce the study area located in Phuentsholing, Bhutan. Section 3 provides details on Sentinel-2 satellite imagery and describes the various processes associated with the

SCP plugin, including implementation of the classification methods. In Sect. 4, we present the detailed image classification results and the quantitative comparison of the accuracy assessment results for the unsupervised and the FR classifier. Section 5 discusses the importance of the spectral signatures of the materials (land classes) for LULC classification and the challenges associated with the heterogeneous land composition. Finally, we conclude how the SCP plugin can be used efficiently for image classification in the context of geohazard-prone

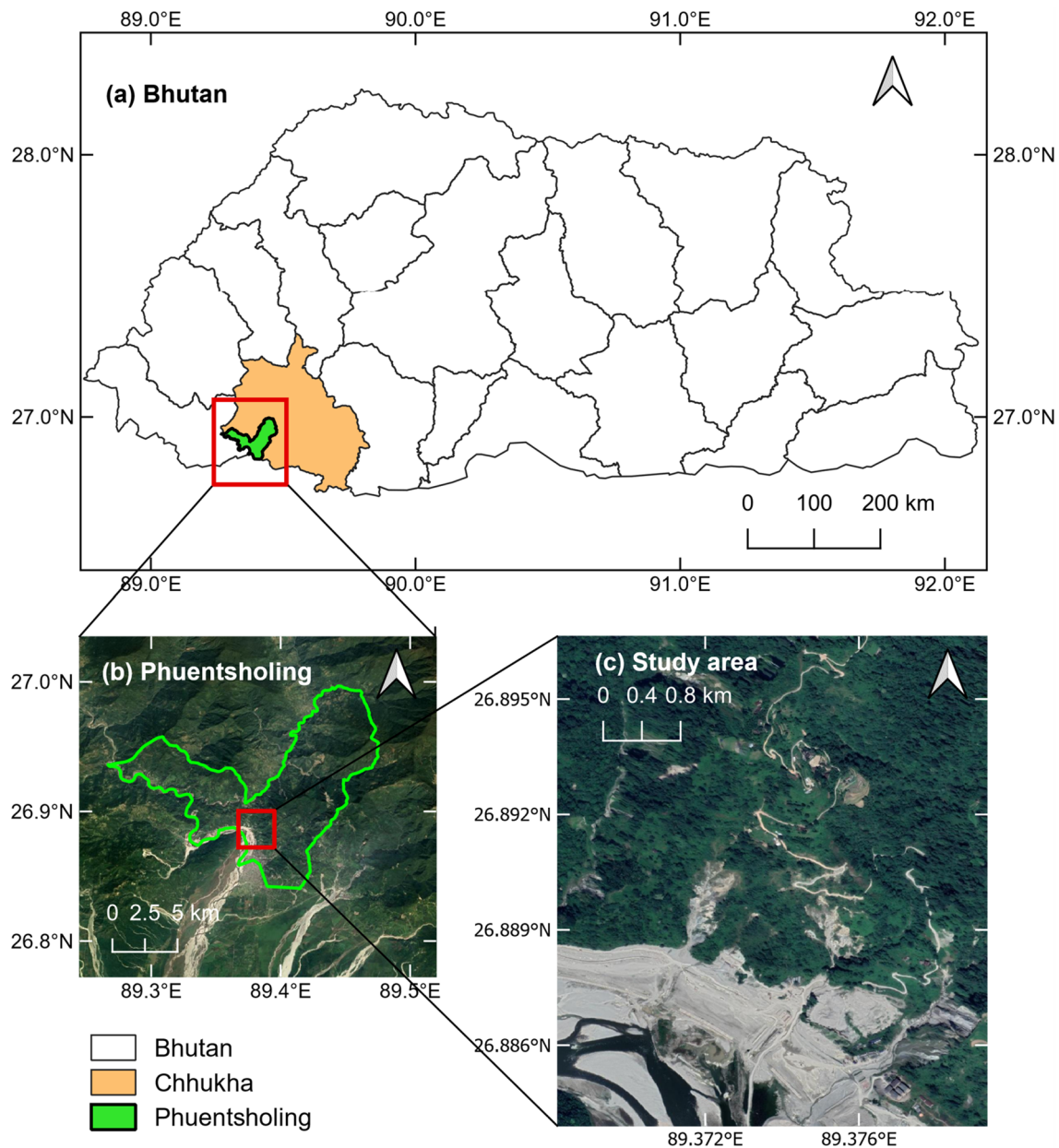


Fig. 1 Geographical location of the study area: **a** Bhutan administrative units, **b** Location of the Phuentsholing area under Chhukha district, **c** Study area location, an extended Phuentsholing township area

area delineation and outlined suggestions for future work.

2 Study area

The study area (Fig. 1) covers an area of 1,438,800 m² in the southern belt of Bhutan (east of Phuentsholing). The area was selected in the current study to complement the application of SCP and predict the outstanding geohazard scenario. The city of Phuentsholing (Fig. 1b) covers an area of 15.6 km² and is located at 26.86° E and 89.39° N. The city has a population of 27,658 and consists of a total of 2263 residential and commercial buildings (<http://www.pcc.bt/index.php/>). The altitude of Phuentsholing ranges between 183 and 2500 m above sea level [35]. Land at downstream of the study area, 2 km from the core city of Phuentsholing (Fig. 1c) has been allocated for the development of the new township of the extended municipality and is important to Phuentsholing, which will serve as one of the largest trading centres in the country. When

complete, the expanded new community will house planned educational, commercial and residential buildings with modern amenities.

Geohazards, particularly landslides and floods accompanied by large debris flows and sediment deposits in the mountainous regions of the Himalayas, are widespread [36]. The majority of the landslides in the Bhutan Himalayas are mainly rainfall-induced phenomena [37] with consequences similar to flooding every monsoon. Geologically, Bhutan Himalayas is characterized by fragile geological settings over complex topographical features [38]. The study area falls within the Phuentsholing Formation under the Buxa Group of the Lesser Himalayas [39]. The proximity of the Main (or Himalayan) Frontal Thrust (MFT) and the Main Boundary Thrust (MBT) and the presence of shear zones in the region reveals a weak geological setting [40]. In addition, the lithological composition of the area consisting of fractured and weathered rocks of grey and dark coloured phyllite strata, slate, and schist that contain high amounts of disintegrated clay indicates the study area to be highly susceptible to the geohazards [35].

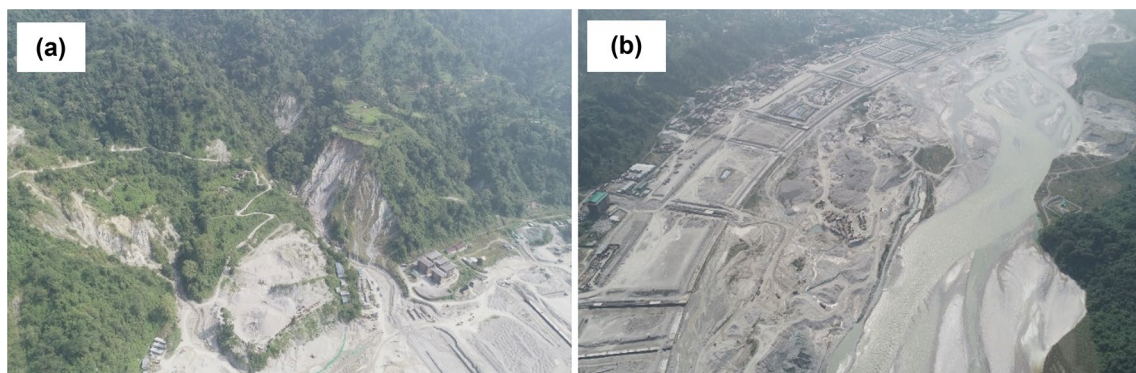
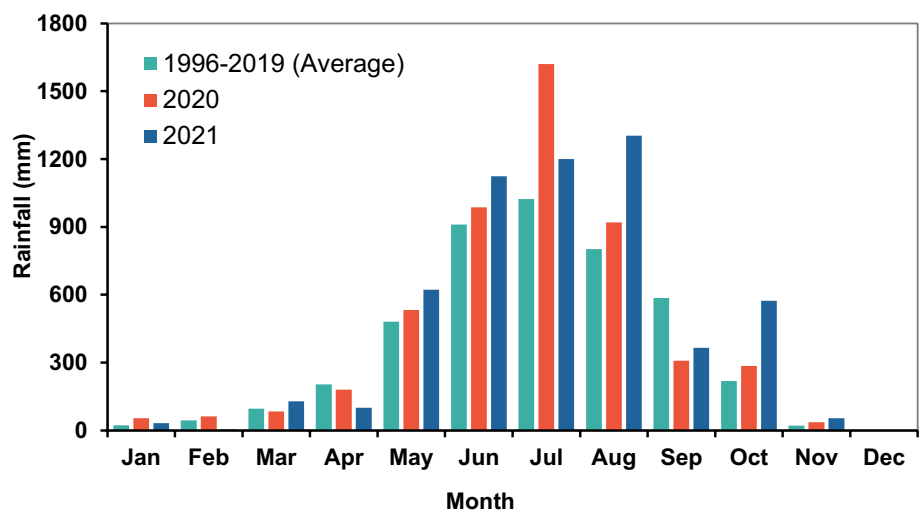


Fig. 2 The aerial view of the geohazard-prone area, **a** Part of upstream of the study area, **b** Extended new township development area

Fig. 3 Phuentsholing monthly rainfall record from 1996–2021



The landslide and flood hazards in the study area have been complex phenomena usually triggered by extreme temporal rainfall events and the landforms have been consequently changing with substantial subsidence of landmass every monsoon [41] (Fig. 2). Recent geohazards in the study area include the 2016 Amochhu flood, which had devastating and enormous social and economic impacts. The landslides, debris flows and flash floods from the upstream tributaries and the sedimentation of the Amochhu River are very frequent and recurring events in each monsoon. According to Bhutan state of climate change 2020 [42], the study area usually receives one of the highest rainfalls. Historical precipitation records from 1996 show an increase in precipitation patterns in recent years (Fig. 3). The average cumulative annual rainfall from 1996 to 2019 is 4406 mm and in the last two years 2020 and 2021 the cumulative annual rainfall was 5070 and 5508.3 mm respectively, with more rainfall recorded in the month between May and August.

3 Materials and methods

3.1 The semi-automatic classification plugin

The open-source, semi-automatic classification plugin (SCP) developed by Congedo [31], primarily provides a means to perform an unsupervised and supervised classification through an image processing technique that aims to classify the land cover by training an algorithm with the samples of the spectral signature of the material (In particular, minimum distance, maximum likelihood, and spectral angle mapping algorithms). The SCP is also enabled with the machine learning algorithm hosted with the Sentinel application platform (SNAP). Creation of training inputs (ROIs) in the form of polygons, manually or by a region growing, and computation of spectral signatures is available with the SCP dock interface. The SCP plugin currently allows downloading of seven satellite datasets and provides a full suite of processing tools that facilitate the pre-processing and post-processing phases for image classification. Basically, the SCP provides an advanced remote sensing techniques for non-remote sensing experts through standard step-by-step procedures that can be applied to urban planning, agriculture, environmental monitoring, etc. In essence, the approach offers fast computation by encoding the context information through clustering and machine learning algorithms. The SCP interface can be sketched as follows to give an overview:

- (i) Platform to download remote sensing images
- (ii) Pre-processing: clipping, conversion to reflectance, and other tools

- (iii) Processing: classification and analysis
- (iv) Postprocessing: refinement of the classification and interpretation of data
- (v) Tools for training input creation (ROIs), computation of spectral signatures, and accuracy assessment.

3.2 Multispectral satellite imagery

The satellite remote-sensing images and GIS datasets are widely used for event-based spatial analysis and multi-temporal assessment of various applications at different regional settings and scales [43]. Sentinel-2 is a multispectral satellite sensor developed by the European Space Agency (ESA) as part of the Copernicus land monitoring services. The Sentinel-2 multispectral instrument (MSI) consists of two satellites that observe the Earth at 10, 20, and 60 m spatial resolutions [20] and the three red-edge bands can capture the strong reflectance of the vegetation in the near-infrared region of the electromagnetic spectrum.

The basic criteria for the selection of satellite imagery for geospatial analysis, in particular for the land cover classification, is that the imagery must contain little or no cloud cover or haze. Sentinel-2A images from October 11, 2021, with less than 0.1% cloud cover, were downloaded from the ESA website (<https://scihub.copernicus.eu>), showing an approximate 20 km radius of the study area. These can be also downloaded with integrated SCP features. The virtual band sets are created from the converted reflectance images, which enabled the visualization of colour composites of the study area. The colour composites of the various band set combinations provided essential information and guidance for designing training samples (ROIs). This technique complements the analysis process and improve the overall performance. Figure 4 shows four different colour composites used during land cover classification post-processing to create training samples. (Table 1).

The unique aspect of the Sentinel level 2A satellite products is that the multispectral bands are atmospherically corrected [45] and can be used directly for pre-processing in SCP. In remote sensing analysis, the spectral reflectance is the basic common parameter that characterizes the spectral reflectance curves of the objects and allows them to be distinguish from each other [46, 47]. Mostly, remote sensing products are available as calibrated digital numbers (DN) and the conversion to radiance or reflectance is necessary. The pre-processing procedure in SCP includes clipping the band sets using geographical coordinates or vector shapefile for the study area and converting them from discrete values in DN to decimal values of reflectance. The Sentinel-2A satellite images has been

Fig. 4 Application of different colour composites of Sentinel-2A images used to perform clustering and create a ROI of different land classes to design training samples: **a** RGB=4-3-2 **b** RGB=3-2-1, **c** RGB=7-3-2, **d** RGB=8-4-3

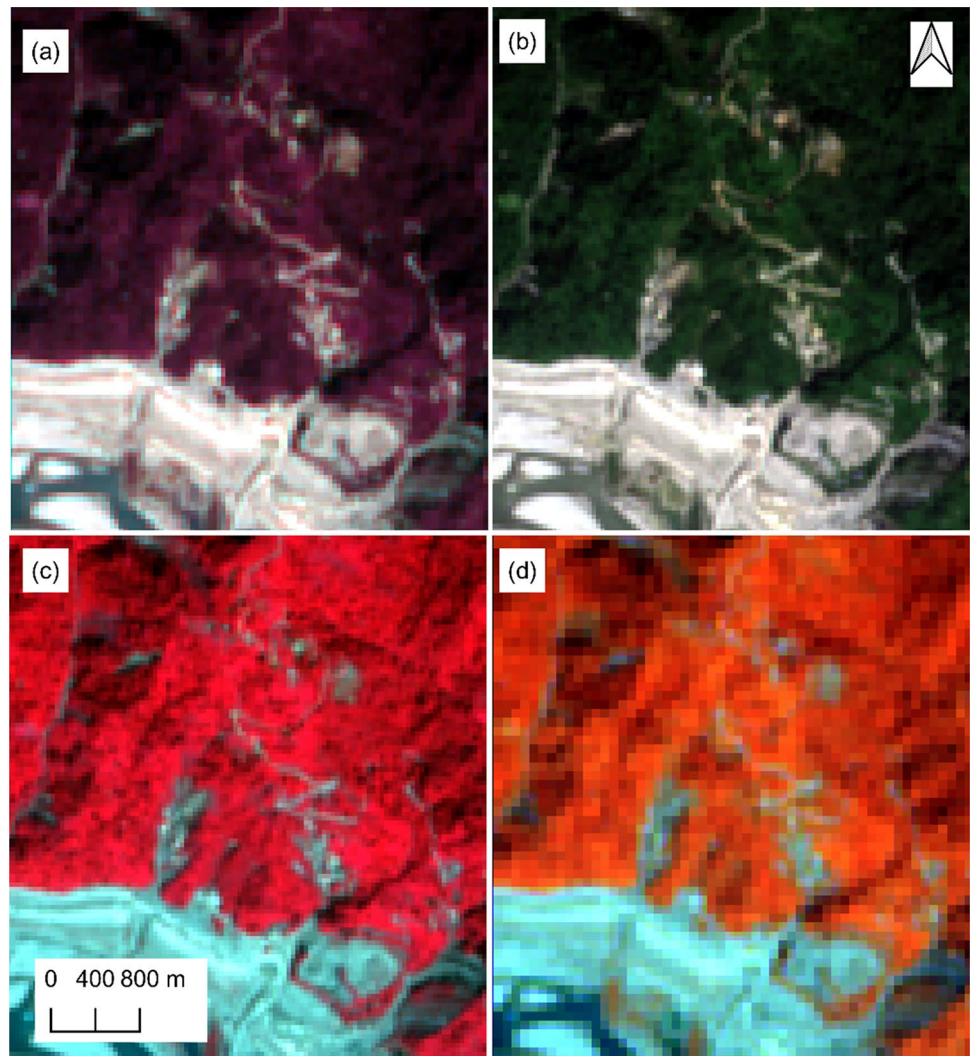


Table 1 Sentinel-2 multispectral bands [44]

Band	Central wave-length (µm)	Resolution (m)	Description
B1	0.443	60	Coastal aerosol
B2	0.490	10	Blue (B)
B3	0.560	10	Green (G)
B4	0.665	10	Red (R)
B5	0.705	20	Near Infrared (NIR)
B6	0.740	20	NIR
B7	0.783	20	NIR
B8	0.842	10	NIR
B8A	0.865	20	NIR
B9	0.945	60	Water vapour
B10	1.375	60	Cirrus
B11	1.610	20	Short Wave Infrared (SWIR)
B12	2.190	20	SWIR

projected onto the Coordinate Reference System (CRS) of EPSG: 32,645-WGS 84/UTM Zone 45 N.

3.3 Unsupervised classification

Unsupervised classification refers to the process of identifying a large number of unknown pixels of the inherent categories from the dataset of the particular image to group into classes (i.e., cluster) without the use of analyst-specified or labeled training samples [48, 49]. Unsupervised image classification relies upon the implementation of various clustering algorithms needed for the categorization of pixels. To device this, the ISODATA algorithm was implemented in SCP using band processing tools. Furthermore, ISODATA introduced an additional feature of merging clusters of similar spectral signatures compared to the k-means method. The algorithm splits clusters with too high variability of spectral signatures based on the standard deviation, since many spectral signatures

(i.e., seed) can be defined in the initial pre-processing phase to capture the number of clusters expected in the image. Hence, the initial parameters provided by the users in SCP implementation are C = number of desired clusters, N_{min} = minimum number of pixels for a cluster, σ_t = maximum standard deviation threshold for splitting and, D_t = distance threshold for merging. The algorithm proceeds with the iterative procedure by calculating Euclidean distance (minimum distance algorithm), and the pixels are assigned to produce clusters according to the most similar spectral signature. With several trials of clustering, where the number of classes = 10, maximum standard deviation = 0.02, minimum class size in pixel = 10, 20 iterations applying minimum distance seed signatures at a distance threshold of 0.01 were used to produce the initial image cluster (Fig. 5a). Then, the image cluster was converted to unique values raster under each cluster using a postprocessing tool and reclassification was carried out to produce a single-band (grey) land cover classification map (Fig. 5b) of the study area.

3.4 The Random Forest classifier

Random Forest is a supervised machine learning algorithm composed of many decision trees. The random forest consists of a large number of individual decision trees that operate as an ensemble. Each tree in the random forest spits out a class prediction and the class with the most

votes become our model's prediction. The technique is used to solve regression and classification and is capable of combining many classifiers to provide a solution to complex problems through ensemble learning [50, 51]. The resulting decision trees (i.e., the forest) generated by the RF algorithm are trained by bagging or bootstrap aggregation using an ensemble meta-algorithm that improves the accuracy of the algorithm [52]. The RF algorithm predicts the decision trees and produces the result, and the prediction takes into account the average or mean of the output of various trees [53]. The accuracy of the result can be increased by increasing the number of trees [50]. Figure 6 shows the generalized implementation of the RF decision tree model.

The reflectance images are uploaded in the SCP, which is a common procedure for executing all algorithms available in the SCP plugin. These band sets are defined as the reference band set to which the wavelength has been assigned by the quick-wavelength setting tool. The RF classifier allows for image classification from the set of bands using the ROI polygons in the training sample input. The ROIs were then used to create classes and macro classes identified by a class ID (i.e., CID) and each ROI was assigned a land cover class. The ROI polygons cover pixels (training data) randomly for each land class, and the number of pixels covered depends largely on the size of each polygon. Hence, the number of maximum training pixels in the SCP was set to 5000 to provision

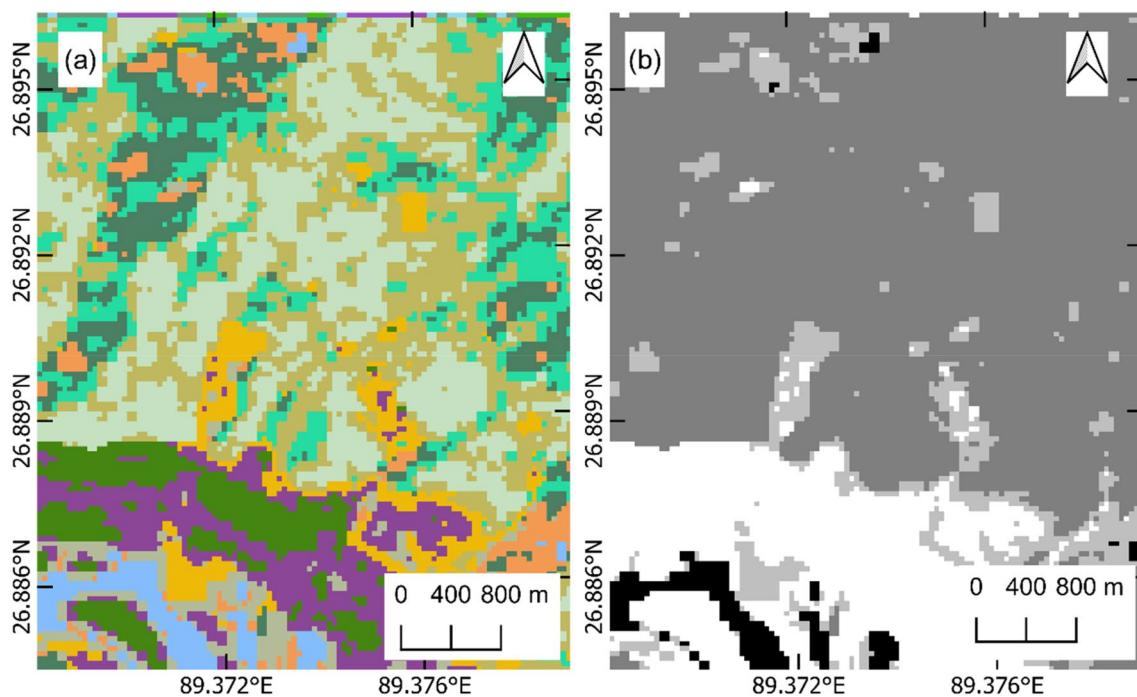


Fig. 5 Preprocessing of Sentinel-2 band sets in SCP: **a** Clustered pixels, **b** Reclassified cluster

Fig. 6 Implementation of generalized RF decision tree models on the training data sets (ROIs)

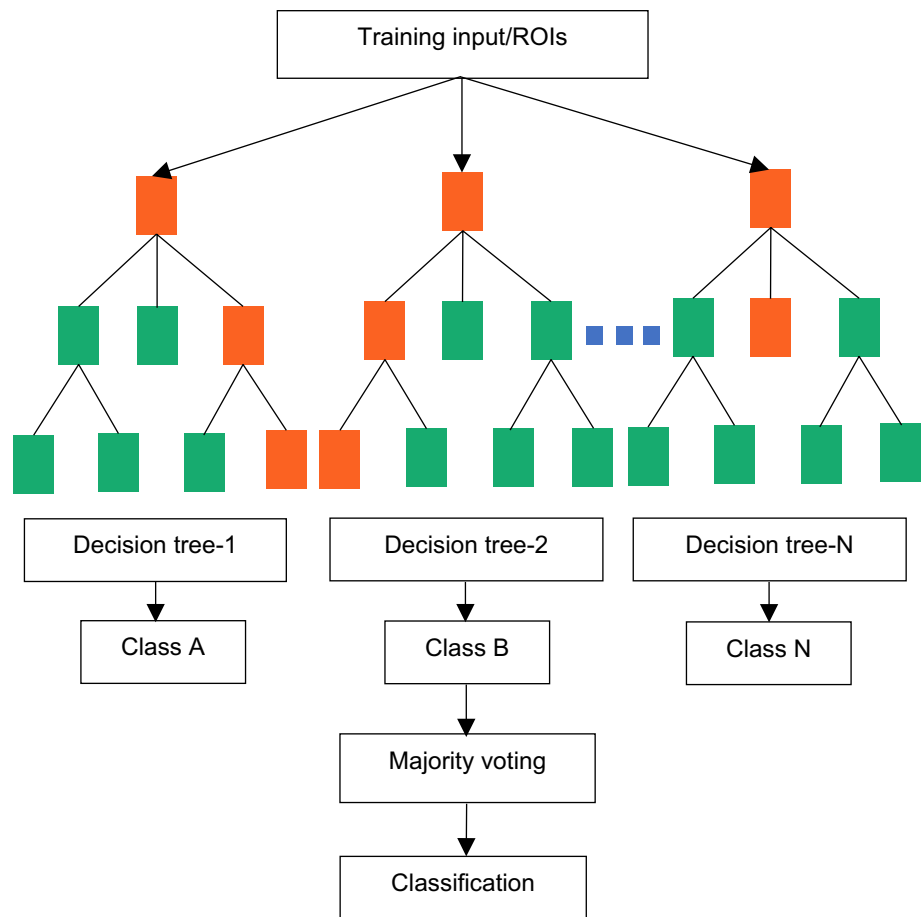


Table 2 Land class coverage for the two algorithms

Method/Class	Land class coverage (%)			
	Water	Vegetation	Landslide	Sediments
Unsupervised	2.65	67.22	10.28	19.84
Supervised	3.69	64.41	9.43	22.47

and accommodate the pixel counts generated inside the ROI polygons. A total of 1863 pixels (training samples) considering all the ROI polygons of different classes was trained in the current RF model. The performance of the RF classifier depends on the number of decision trees and the higher number allows for more accurate models, albeit in exchange for a longer execution time to run the classification. In this study, we employed 100 decision trees to run the RF classifier.

4 Results

4.1 Land cover classification maps

The coverage of the land classes (Table 2) of the study area using the two classification algorithms is shown in Fig. 7a, b. The unsupervised classification of the Sentinel-2 image resulted in 38,200 m² of water cover, 967,200 m² of vegetation, 147,900 m² of the landslide area, and 285,500 m² of sedimental deposits within the study area.

Within the study area, vegetation forms the largest land cover type with 67.22%, and water forms the smallest, with 2.65%. The landslide and the sediment deposit areas consist of 10.28 and 19.84% respectively. Similarly, RF classifier results show 53,100 m² (3.69%) for water, 926,700 m² (12.88%) for vegetation, 135,700 m² (9.43%) as landslide, and 323,300 m² (22.47%) as the area covered by the sediment deposits of riverine flood and debris flow due to flash floods and landslides. The land cover classification map produced from unsupervised classification and RF classifier is presented in Fig. 7.

The land cover classification map produced using the RF classifier is reclassified to map the geohazard-prone

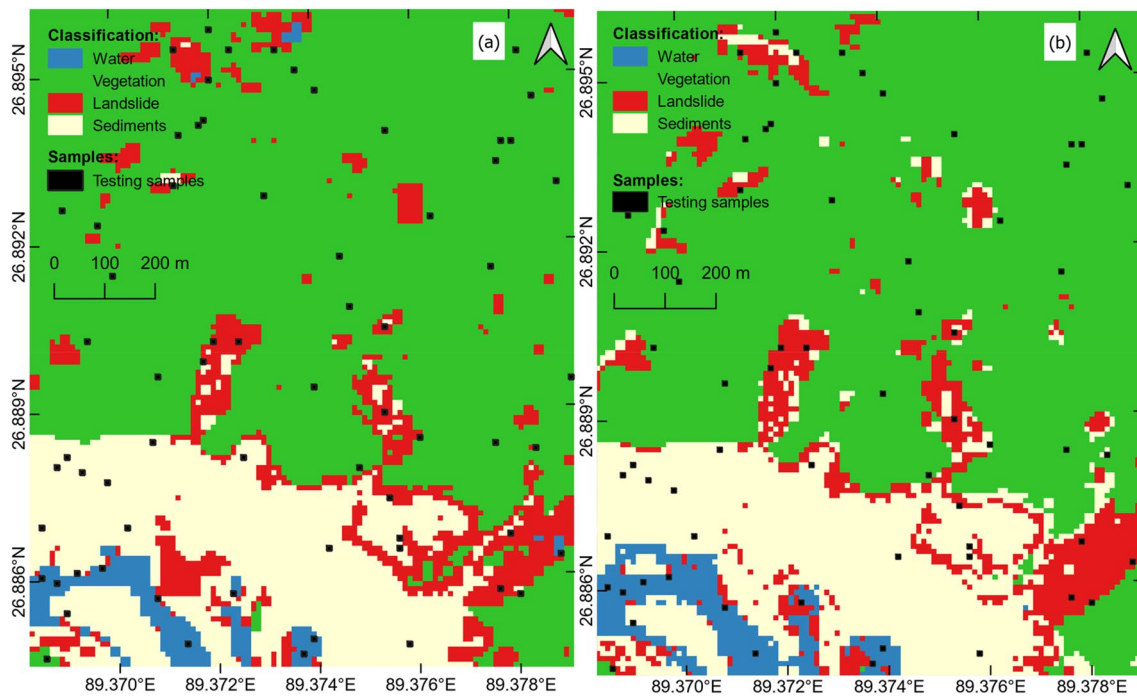


Fig. 7 Land classification map of the study area produced using SCP, **a** Unsupervised classification map using ISODATA, **b** Supervised classification map using RF algorithm. Geohazard attributes in consideration: Amochhu River, sediments/debris, and landslides

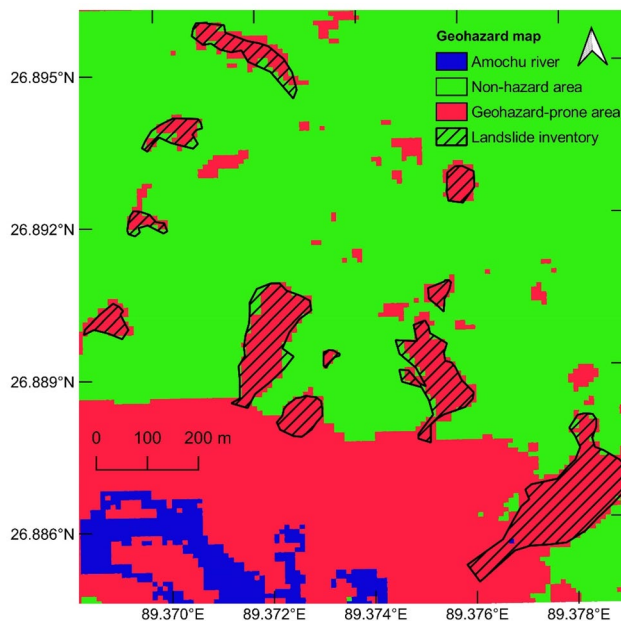


Fig. 8 Delineated geohazard map from land cover classification (FR classifier)

areas (Fig. 8). The geohazard attributes in the map include the Amochhu River (3.69%) and the landslide and flood area (31.90%) and in total, 35.59% of the area is indicative of the geohazard-prone area. The non-hazard area covers 64.41%. Further, we used high-resolution satellite imagery

of Google Earth to create polygons to demonstrate the extent of major landslides. The polygon of the landslides was georeferenced in QGIS software to produce a landslide inventory map of the study area and roughly validate this on the geohazard map.

4.2 Accuracy assessment

Accuracy assessment is necessary to assess the reliability of the results [54]. Consequently, producer (PA) and user accuracies (UA) are computed based on an error matrix resulting from accuracy assessment [55, 56]. In this study, an advance accuracy assessment using SCP functions for both unsupervised classification and RF classifier was performed and the results compared. Using the generated landcover classification map, the next analytical step was to run a post-processing tool to obtain each classifier's classification report, which provided the pixel sum and area coverage of each landcover class (Tables 3 and 4). Accuracy evaluation requires these parameters to calculate the total number of training samples required for the classified image using Eq. 1.

$$N = \left(\sum_{i=1}^c (W_i S_i) / S_o \right)^2, \quad (1)$$

where W_i is the mapped area portion of class i , S_i is the standard deviation of stratum i , S_o is the expected standard deviation of overall accuracy (0.01) and c is the total number of classes. The classification raster must be defined in the band set whose accuracy must be assessed based on the vector or the raster training samples created. In addition, the sample stratification method corresponding to the signature of each class was used to develop a sample design by creating multiple ROIs with basic tool options in SCP. In an accuracy assessment, to stratify the samples,

we conjectured user accuracy and standard deviations of strata such that the user's accuracy is lower and standard deviations S_i is higher for the classes possessing low area proportion and evaluated to arrive at adequate accuracy [54] as shown in Tables 3 and 4. After that, each pixel sample was listed as point coordinates of the respective classes and the total stratified samples were aggregated. The final step involved verifying the training pixel samples by performing photo interpretation to resample the training datasets. In the current study, we employed

Table 3 Classification report and sampling attributes of unsupervised classification

Land class cover	Area (m ²)	Pixel sum	W_i	S_i	$W_i \times S_i$	Weighted	Equal	Mean
Water	38,200	382	0.0266	0.5	0.0133	17.6978	166.3326	92.015
Vegetation	967,200	9672	0.6722	0.2	0.13444	447.2351	166.3326	306.784
Landslides	147,900	1779	0.1028	0.3	0.03084	68.3960	166.3326	117.364
Sediments	285,500	2855	0.1984	0.4	0.07936	132.0016	166.3326	149.167

Table 4 Classification report and sampling attributes for the RF classifier

Land class cover	Area (m ²)	Pixel sum	W_i	S_i	$W_i \times S_i$	Weighted	Equal	Mean
Water	53,100	531	0.0369	0.5	0.01845	25.9991	176.146	101.073
Vegetation	926,700	9267	0.6441	0.2	0.12882	453.8225	176.146	314.984
Landslides	135,700	1357	0.0943	0.3	0.02829	66.4423	176.146	121.294
Sediments	323,300	3233	0.2247	0.4	0.08988	158.3200	176.146	167.233

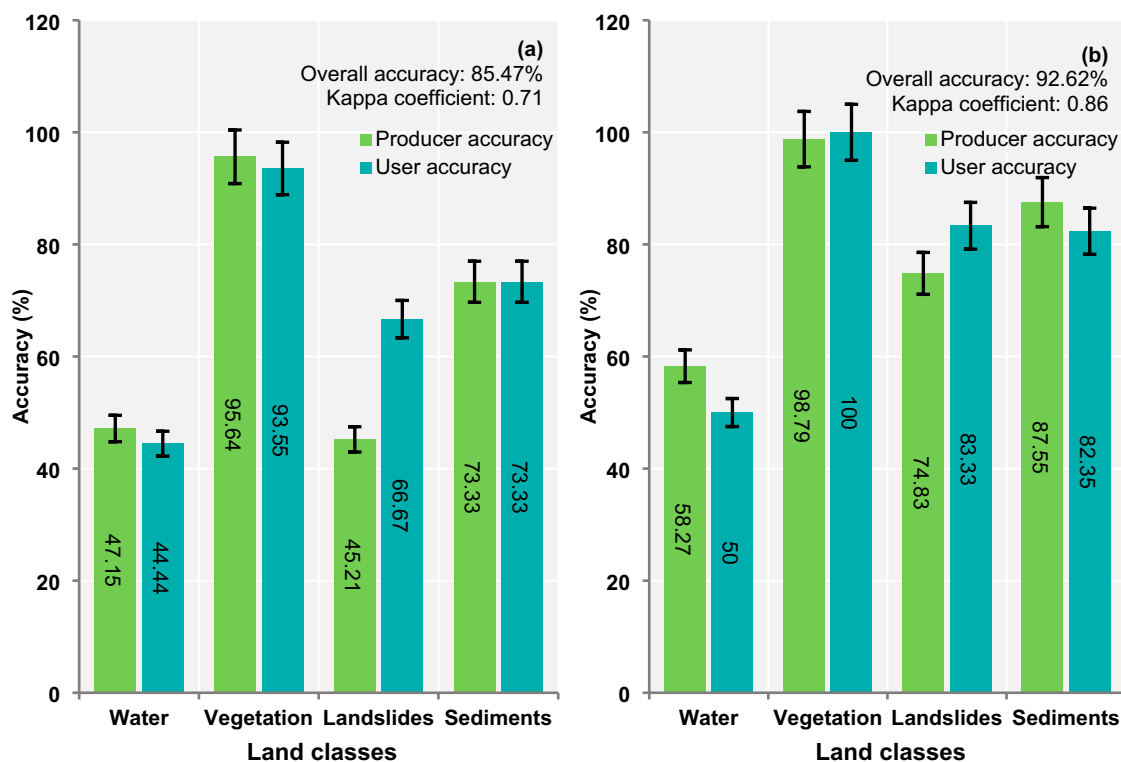


Fig. 9 Accuracy assessment plot, **a** ISODATA, **b** RF classifier

high-resolution Google Earth and virtual composite images to resample the training samples (test samples). Finally, an accuracy assessment between the land cover classification map and the resampled training input was performed. The SCP plugin also provided a first predictive assessment of the classifier model. The current RF decision tree model represents 96.99% of the prediction, with 0.26 root mean square error (RMSE) and 0.02 bias from 1863 training samples. The accuracy assessment of the land cover classification is compared in Fig. 9, between the two algorithms (unsupervised and supervised classification). In general, both the PA and UA for water class resulted in a lower percentage of accuracy compared to other classes. The unsupervised classification for water class shows 47.15% and 44.44% of PA and UA respectively, while the FR classifier indicates slightly higher accuracy of 58.27% and 50% with a difference of accuracy in the range of 11.12–5.56%.

The vegetation land class dominates the classification accuracy with 95.64% PA and 93.55% UA. The level of accuracy is even higher when the RF classifier is employed, with PA at 98.79% and 100% of UA. The improvement in accuracy of the land cover classification is in the range of 3.15–6.45% with the supervised classification using the RF algorithm.

The most essential part of the study is to delineate the geohazard-prone area using the SCP technique and to adequately assess the accuracy of the outcomes. The unsupervised classification indicates poor delineation of geohazard areas compared to the supervised RF classifier, in particular, delineation of landslides and sediment deposit areas. The accuracy results demonstrate 45.21%

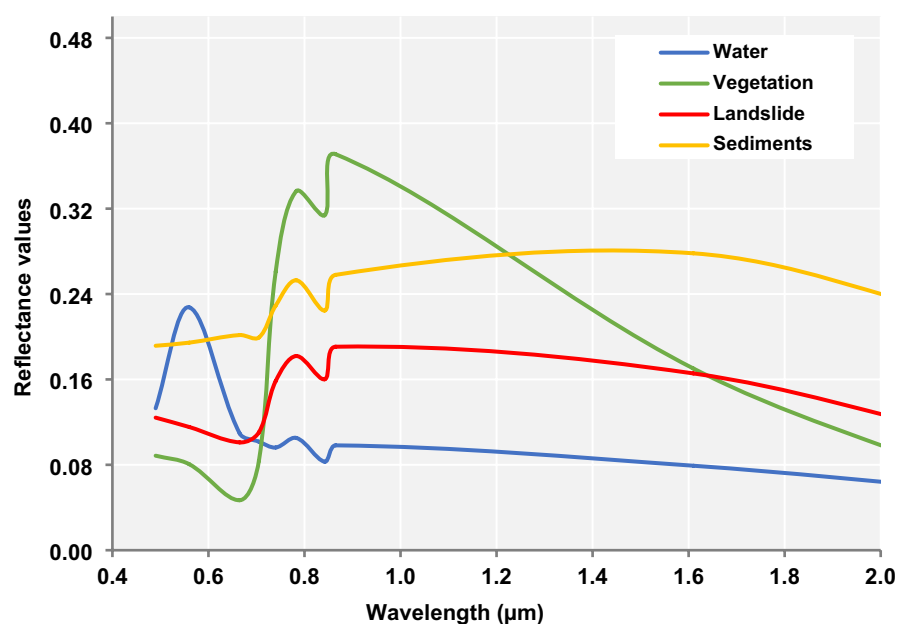
and 66.67% of PA and UA respectively. On the contrary, the application of the FR algorithm shows a substantial improvement in accuracy, of 74.83% and 83.33% in PA and UA respectively. The improvement is in the range of 29.62–16.66%. Similarly, a notable difference in the accuracy of classification for sediment land class can be observed. The PA of 73.33% by unsupervised learning and 87.55% by supervised learning has been achieved indicating an improvement of classification accuracy by 14.22%. For UA, we saw an increase in accuracy by 9.02% resulting from 73.33% to 82.35%, through unsupervised classification and RF classifier respectively.

The diagnosis of the accuracy assessment indicates overall accuracies of 85.47 and 92.62% for unsupervised classification and RF algorithm, respectively. The study shows that the image classification model using RF classifier, a supervised machine-learning algorithm, demonstrates higher accurate classification results which provide a reliable geohazard map of the study area. In the current study, the supervised classification model (RF) is 7.15% superior in terms of overall classification accuracy. According to the results of the statistical error matrix (confusion matrix) of unsupervised and supervised algorithms, a Kappa coefficient of 0.71 and 0.86, respectively, is achieved.

5 Discussion

The current SCP framework of image classification is traditionally a pixel-based classification method that implements a grouping of pixels that are made up of them or

Fig. 10 The spectral signature plot of land classification by RF classifier



categorization within the range of thresholds (or statistically) defined by the type of algorithms used. The signature curve of land classes in the study area shown in Fig. 10 presents the spectral reflectance value distribution corresponding to the wavelength. The spectral characteristics of water, vegetation, and soil are evident in three categories of visible region (0.4–0.7 μm), near-infrared (0.7–1.3 μm), and shortwave infrared region (beyond 1.3 μm).

Although the study area is not very large and is rather a localized region, the resulting accuracy for water class indicated a lower forecast in both methods (47.15 and 58.27%, respectively). This challenge is usually observed when the pixel-wise approaches are applied to heterogeneous regions where the surface features such as colour, structure, and texture play important roles. Usually, this limitation also happens when the object's size is much smaller than the size of the pixel [57]. The relief of the study area indicated a close alignment to such observations since the angle of incidence of the sunlight has a high potential to differentiate the spectral reflectance pattern of similar signatures. Photo interpretation of high-resolution satellite imagery shows that the east-facing slopes in the study area fall under shadows (dark) and have a similar spectral signature to water bodies, and this effect is likely due to the water class resulting in low accuracy compared to others classes. Soil types of different colours and textures show unique reflectance values and similar classification issues can be expected which can lead to over or underestimation, especially when dealing with landslide applications. Often open construction sites, unpaved farm roads, concrete surfaces, and barren land resembles the spectral signature of landslide soils. In such cases, image classification for a larger area may be extremely challenging.

6 Conclusions

The image classification was performed using the SCP plugin and Sentinel-2 imagery in QGIS software to create land cover classification maps. The map produced by the RF classifier, with an overall accuracy of 92.62%, was further reclassified to delineate the geohazard-prone areas of the study area (512,100 m^2). The pre-processing and basic tools in SCP are able to download the RS images, crop them and convert them to reflectance. The SCP plugin is equipped with multiple classification algorithms for image classification including a machine learning algorithm (RF classifier). The presented methodology and semi-automatic classification technique using the SCP plugin show a high potential for performing a land cover classification

in the heterogeneous mountainous region. The SCP plugin enabled quick image classification using clustering and machine learning algorithms, which proved to be less computationally intensive. The RF classifier proved to be more efficient compared to unsupervised classification. Implementing unsupervised classification requires multiple attempts at clustering with appropriate user input parameters related to the algorithm attributes and achieving reasonable accuracy. The accuracy rating indicates higher accuracy of land cover classification using the RF classifier. However, the study also found some anomalies in the accuracy results related to the terrain relief and topography, depicting near spectral signatures for some of the materials (e.g., water class). To address the anomalies, the flexibility to define the parameters of the algorithms (thresholds) in the SCP plugin with an increasing number of macro-classes and subclasses could be a possible solution. With this scope, the potential of the semi-automatic classification technology can be more precisely exploited for any geographical environment.

Acknowledgements The authors would like to gratefully acknowledge generous philanthropic support from the Office of the Vice President, Rabdan Academy, the United Arab Emirates (UAE).

Authors' contributions All authors contributed to the study's conception and design [KT and KRA]. Material preparation and analysis were performed by KT. The first draft of the manuscript was written by KT and all authors edited and commented on previous versions of the manuscript. All authors also read, revised, provided inputs and approved the final manuscript.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability All data generated or analyzed during this study are included in this published article.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/>.

References

1. Ali SA, Parvin F, Vojteková J, Costache R, Linh NTT, Pham QB, Vojtek M, Gigović L, Ahmad A, Ghorbani MA (2021) GIS-based landslide susceptibility modeling: a comparison between fuzzy multi-criteria and machine learning algorithms. *Geosci Front* 12:857–876. <https://doi.org/10.1016/j.gsf.2020.09.004>
2. Dhital YP, Kayastha RB (2013) Frequency analysis, causes and impacts of flooding in the Bagmati River Basin. *Nepal J Flood Risk Manag* 6:253–260. <https://doi.org/10.1111/jfr3.12013>
3. Mu D, Luo P, Lyu J, Zhou M, Huo A, Duan W, Nover D, He B, Zhao X (2021) Impact of temporal rainfall patterns on flash floods in Hue City, Vietnam. *J Flood Risk Manag* 14:1–15. <https://doi.org/10.1111/jfr3.12668>
4. Mohajane M, Essahlaoui A, Oudija F, Hafyani ME, Hmadi AE, Ouali AE, Randazzo G, Teodoro AC (2018) Land use/land cover (LULC) using landsat data series (MSS, TM, ETM+ and OLI) in Azrou Forest, in the Central Middle Atlas of Morocco. *Environments* 5(12):1–16. <https://doi.org/10.3390/environments5120131>
5. Shahabi H, Hashim M (2015) Landslide susceptibility mapping using GIS-based statistical models and remote sensing data in tropical environment. *Sci Rep* 5:1–15. <https://doi.org/10.1038/srep09899>
6. Zhang Z, Yang F, Chen H, Wu Y, Li T, Li W, Wang Q, Liu P (2016) GIS-based landslide susceptibility analysis using frequency ratio and evidential belief function models. *Environ Earth Sci* 75:1–12. <https://doi.org/10.1007/s12665-016-5732-0>
7. Awawdeh MM, ElMughrabi MA, Atallah MY (2018) Landslide susceptibility mapping using GIS and weighted overlay method: a case study from North Jordan. *Environ Earth Sci* 77:1–15. <https://doi.org/10.1007/s12665-018-7910-8>
8. Vojteková J, Vojtek M (2020) Assessment of landslide susceptibility at a local spatial scale applying the multi-criteria analysis and GIS: a case study from Slovakia. *Geomat Nat Hazards Risk* 11:131–148. <https://doi.org/10.1080/19475705.2020.1713233>
9. Senouci R, Taibi NE, Teodoro AC, Duarte L, Mansour H, Meddah RY (2021) GIS-based expert knowledge for landslide susceptibility mapping (LSM): case of mostaganem coast district, west of Algeria. *Sustainability* 13(2):1–21. <https://doi.org/10.3390/su13020630>
10. Ortiz JAV, Martínez-Graña AM (2018) A neural network model applied to landslide susceptibility analysis (Capitanejo, Colombia). *Geomat Nat Hazards Risk* 9:1106–1128. <https://doi.org/10.1080/19475705.2018.1513083>
11. Nhu VH, Shirzadi A, Shahabi H, Chen W, Clague JJ, Geertsema M, Jaafari A, Avand M, Miraki S, Asl DT, Pham BT, Ahmad BB, Lee S (2020) Shallow landslide susceptibility mapping by random forest base classifier and its ensembles in a semi-arid region of Iran. *Forests* 11:1–28. <https://doi.org/10.3390/F11040421>
12. Thi Ngo PT, Panahi M, Khosravi K, Ghorbanzadeh O, Kariminejad N, Cerda A, Lee S (2021) Evaluation of deep learning algorithms for national scale landslide susceptibility mapping of Iran. *Geosci Front* 12:505–519. <https://doi.org/10.1016/j.gsf.2020.06.013>
13. Chen W, Xie X, Wang J, Pradhan B, Hong H, Bui DT, Duan Z, Ma J (2017) A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *CATENA* 151:147–160. <https://doi.org/10.1016/j.catena.2016.11.032>
14. Bachri I, Hakdaoui M, Raji M, Teodoro AC, Benbouziane A (2019) Machine learning algorithms for automatic lithological mapping using remote sensing data: a case study from Souk Arbaa Sahel, Sidi Ifni Inlier, Western Anti-Atlas, Morocco. *ISPRS Int J Geo-Inf* 8(6):1–20. <https://doi.org/10.3390/ijgi8060248>
15. Amatya P, Kirschbaum D, Stanley T, Tanyas H (2021) Landslide mapping using object-based image analysis and open-source tools. *Eng Geol* 282:106000. <https://doi.org/10.1016/j.enggeo.2021.106000>
16. Bouaziz M, Eisold S, Guermazi E (2017) Semiautomatic approach for land cover classification: a remote sensing study for arid climate in southeastern Tunisia. *Euro-Mediterr J Environ Integr* 2:1–7. <https://doi.org/10.1007/s41207-017-0036-7>
17. Wang Q, Blackburn GA, Onojeghuo AO, Dash J, Zhou L, Zhang Y, Atkinson PM (2017) Fusion of Landsat 8 OLI and Sentinel-2 MSI data. *IEEE Trans Geosci Remote Sens* 55(7):3885–3899. <https://doi.org/10.1109/TGRS.2017.2683444>
18. Belgiu M, Csillik O (2018) Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sens Environ* 204:509–523. <https://doi.org/10.1016/j.rse.2017.10.005>
19. Duarte L, Teodoro AC, Cunha M (2019) A semi-automatic approach to derive land cover classification in soil loss models. In *Proceedings SPIE 11156, Earth resources and environmental remote sensing/GIS applications X*, 111560B, 3 October 2019: 1–13. <https://doi.org/10.1117/12.2532935>
20. Abdi AM (2020) Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. *GIScience Remote Sens* 57:1–20. <https://doi.org/10.1080/15481603.2019.1650447>
21. Belenok V, Noszczyk T, Hebyn-Baidy L, Kryachok S (2021) Investigating anthropogenically transformed landscapes with remote sensing. *Remote Sens Appl Soc Environ* 24:100635. <https://doi.org/10.1016/j.rsase.2021.100635>
22. Pasha SV, Reddy CS, Jha CS, Rao PVVP, Dadhwal VK (2016) Assessment of land cover change hotspots in Gulf of Kachchh, India using multi-temporal remote sensing data and GIS. *J Indian Soc Remote Sens* 44(6):905–913. <https://doi.org/10.1007/s12524-016-0562-9>
23. Noi PT, Kappas M (2017) Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors* 18(1):1–20. <https://doi.org/10.3390/s18010018>
24. Furukawa F, Morimoto J, Yoshimura N, Kaneko M (2020) Comparison of conventional change detection methodologies using high-resolution imagery to find forest damage caused by typhoons. *Remote Sens* 12:1–17. <https://doi.org/10.3390/rs12193242>
25. Höhle J (2021) Automated mapping of buildings through classification of DSM-based ortho-images and cartographic enhancement. *Int J Appl Earth Obs Geoinf* 95:102237. <https://doi.org/10.1016/j.jag.2020.102237>
26. Whiteside TG, Boggs GS, Maier SW (2011) Comparing object-based and pixel-based classifications for mapping savannas. *Int J Appl Earth Obs Geoinf* 13(6):884–893. <https://doi.org/10.1016/j.jag.2011.06.008>
27. Xie S, Liu L, Zhang X, Yang J, Chen X, Gao Y (2019) Automatic land-cover mapping using landsat time-series data based on google earth engine. *Remote Sens* 11:1–20. <https://doi.org/10.3390/rs11243023>
28. Gonzalez-ollauri A, Mickovski SB (2021) A Simple GIS-based tool for the detection of landslide-prone zones on a coastal slope in Scotland. *Land* 10:1–15. <https://doi.org/10.3390/land10070685>
29. Gašparović M, Zrinjski M, Gudelj M (2019) Automatic cost-effective method for land cover classification (ALCC). *Comput Environ Urban Syst* 76:1–10. <https://doi.org/10.1016/j.compenvurbsys.2019.03.001>
30. Arekhi M, Goksel C, Sanli BF, Senel G (2019) Comparative evaluation of the spectral and spatial consistency of Sentinel-2 and

- Landsat-8 OLI data for Igneada longos forest. *Int J Geo-Inf* 8:1–13. <https://doi.org/10.3390/ijgi8020056>
31. Congedo L (2021) Semi-automatic classification plugin: a python tool for the download and processing of remote sensing images in QGIS. *J Open Sour Softw Mag* 6:3172. <https://doi.org/10.21105/joss.03172>
 32. QGIS Development Team (2021) QGIS geographic information system. QGIS Association. <https://www.qgis.org>
 33. Gislason PO, Benediktsson JA, Sveinsson JR (2006) Random forests for land cover classification. *Pattern Recognit Lett* 27(4):294–300. <https://doi.org/10.1016/j.patrec.2005.08.011>
 34. Gariano SL, Sarkar R, Dikshit A, Dorji K, Brunetti MT, Peruccacci S, Melillo M (2019) Automatic calculation of rainfall thresholds for landslide occurrence in Chukha Dzongkhag, Bhutan. *Bull Eng Geol Environ* 78(6):4325–4332. <https://doi.org/10.1007/s10064-018-1415-2>
 35. Dikshit A, Sarkar R, Pradhan B, Acharya S, Alamri AM (2020) Spatial landslide risk assessment at Phuentsholing, Bhutan. *Geosciences* 10(4):1–16. <https://doi.org/10.3390/geosciences10040131>
 36. Tempa K, Chettri N, Sarkar R, Saha S, Gurung L, Dendup T, Nirola BS (2021) Geotechnical parameter assessment of sediment deposit: a case study in Pasakha, Bhutan. *Cogent Eng* 8(1):1–21. <https://doi.org/10.1080/23311916.2020.1869366>
 37. Dikshit A, Sarkar R, Pradhan B, Acharya S, Dorji K (2019) Estimating rainfall thresholds for landslide occurrence in the Bhutan Himalayas. *Water* 11:1–12. <https://doi.org/10.3390/w11081616>
 38. Tempa K, Sarkar R, Dikshit A, Pradhan B, Simonelli AL, Acharya S, Alamri AM (2020) Parametric study of local site response for bedrock ground motion to earthquake in Phuentsholing, Bhutan. *Sustainability* 12:1–20. <https://doi.org/10.3390/su12135273>
 39. Tempa K, Chettri N (2021) Comprehension of conventional methods for ultimate bearing capacity of shallow foundation by PLT and SPT in Southern Bhutan. *Civ Eng Archit* 9:375–385. <https://doi.org/10.13189/cea.2021.090210>
 40. Thakur VC (2001) Regional geology and geological evolution of the Himalaya. In: Tianchi L, Chalise SR, Upreti BN (eds) *Landslide hazard mitigation in the Hindu Kush-Himalayas*. <https://doi.org/10.53055/ICIMOD.374>
 41. Tempa K, Peljor K, Wangdi S, Ghalley R, Jamtsho K, Ghalley S, Pradhan P (2021) UAV technique to localize landslide susceptibility and mitigation proposal: a case of Rinchening Goenpa landslide in Bhutan. *Nat Hazards Res*. <https://doi.org/10.1016/j.nhres.2021.09.001> (in press)
 42. NCHM (2021) Bhutan state of the climate 2020. Weather and climate services division, NCHM, Thimphu, Bhutan.
 43. Lai JS, Tsai F (2019) Improving GIS-based landslide susceptibility assessments with multi-temporal remote sensing and machine learning. *Sensors* 19:1–25. <https://doi.org/10.3390/s19173717>
 44. Zhang T, Su J, Xu Z, Luo Y, Li J (2021) Sentinel-2 satellite imagery for urban land cover classification by optimized random forest classifier. *Appl Sci* 11(2):1–17. <https://doi.org/10.3390/app11020543>
 45. Congedo L (2021) Semi-automatic classification plugin documentation Release 7.9.5.1
 46. Pop A, Zoran M, Braescu CL, Necsoiu M, Serban F (1995) Spectral reflectance signification in satellite imagery. In *Proceedings SPIE 2326, Photon Transport in Highly Scattering Tissue*, Lille, France. <https://doi.org/10.1117/12.200851>
 47. Cardoso-Fernandes J, Teodoro AC, Lima A (2018) Potential of Sentinel-2 data in the detection of lithium (Li)-bearing pegmatites: a study case. In *Proceedings SPIE 10790, Earth Resources and Environmental Remote Sensing/GIS Applications IX*, 107900T (9 October 2018). <https://doi.org/10.1117/12.2326285>
 48. Olaode A, Naghdy GA, Todd C, Naghdy G (2014) Unsupervised classification of images: a review. *Int J Image Process* 8:325–342
 49. Teodoro AC, Pais-Barbosa J, Veloso-Gomes F, Taveira-Pinto F (2009) Evaluation of beach hydromorphological behaviour and classification using image classification techniques. *J Coast Res* 56:1607–1611
 50. Gounaridis D, Apostolou A, Koukoulas S (2016) Land cover of Greece, 2010: a semi-automated classification using random forests. *J Maps* 12:1055–1062. <https://doi.org/10.1080/17445647.2015.1123656>
 51. Taalab K, Cheng T, Zhang Y (2018) Mapping landslide susceptibility and types using Random Forest. *Big Earth Data* 2:159–178. <https://doi.org/10.1080/20964471.2018.1472392>
 52. Wang C, Shu Q, Wang X, Guo B, Liu P, Li Q (2019) A random forest classifier based on pixel comparison features for urban LiDAR data. *ISPRS J Photogramm Remote Sens* 148:75–86. <https://doi.org/10.1016/j.isprsjprs.2018.12.009>
 53. Dang VH, Hoang ND, Nguyen LMD, Bui DT, Samui P (2020) A novel GIS-based random forest machine algorithm for the spatial prediction of shallow landslide susceptibility. *Forests* 11:1–20. <https://doi.org/10.3390/f11010118>
 54. Olofsson P, Foody GM, Herold M, Stehman SV, Woodcock CE, Wulder MA (2014) Remote sensing of environment good practices for estimating area and assessing accuracy of land change. *Remote Sens Environ* 148:42–57. <https://doi.org/10.1016/j.rse.2014.02.015>
 55. Congalton RG (1991) A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens Environ* 37(1):35–46. [https://doi.org/10.1016/0034-4257\(91\)90048-B](https://doi.org/10.1016/0034-4257(91)90048-B)
 56. Zhang SL, Chang TC (2015) A study of image classification of remote sensing based on back-propagation neural network with extended delta bar delta. *Math Probl Eng* 2015:1–10. <https://doi.org/10.1155/2015/178598>
 57. Li M, Zang S, Zhang B, Li S, Wu C (2017) A Review of remote sensing image classification techniques: the role of spatio-contextual information. *Eur J Remote Sens* 47:389–411. <https://doi.org/10.5721/EuJRS201447>

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