Landslide mapping using object-based image analysis and open source tools

Pukar Amatya[1,2,3,*], Dalia Kirschbaum[3], Thomas Stanley[1,2,3] and Hakan Tanyas[1,3,4]

[1] Universities Space Research Association, Columbia, MD, USA
[2] Goddard Earth Sciences Technology and Research, Columbia, MD, USA
[3] Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA
[4]University of Twente, Faculty of Geo-Information Science and Earth Observation (ITC), Enschede, Netherlands

***** Correspondence: pukar.m.amatya@nasa.gov; Tel.: +1-301-614-5956

Abstract

Availability of high-resolution optical imagery and advances in image processing technologies have significantly improved our ability to map landslides. In recent years object-based image analysis (OBIA) has been gaining in popularity for landslide mapping due to its ability to incorporate spectral, textural, morphological and topographical properties. Many studies have been conducted based on commercial software. In this study, we create an open source Semi-Automatic Landslide Detection (SALaD) system utilizing OBIA and machine learning. Configured to run in Linux environment, it uses various open source Python packages and modules. This system was tested in 575 km^2 area along the Pasang Lhamu Highway, Nepal where large numbers of landslides were triggered by the 2015 Gorkha earthquake. Comparison with a manual inventory highlighted that this system was able to detect 70% of the landslide area. The speed and efficiency with which this system was able to detect landslides makes it a viable alternative to manual techniques for landslide mapping over large areas, when establishing approximate landslide locations is of prime importance.

Keywords: Open source, Object-based image analysis, Landslides, RapidEye, Python, Nepal.

1. Introduction

A comprehensive landslide inventory detailing the spatial and temporal characteristics of past landslides is fundamental for formulation of effective landslide hazard and risk management strategies (Corominas et al., 2014; van Westen et al., 2006). Conventionally, landslide inventories are produced by field surveys and/or manual interpretation of aerial photos and satellite images (Guzzetti et al., 2012; Scaioni et al., 2014). This is a time-consuming and subjective process, which can also be limited in space and time (Fan et al., 2019). Availability of high-resolution imagery and development of advanced semi-automatic or automatic processing techniques in recent years has significantly improved our ability to map landslides with less human interaction. Remote sensing classification/detection of landslides can be categorized into two groups: pixel-based and object-based image analysis (OBIA). The pixel-based method uses the spectral information of single pixels and has been successfully used to map landslides (Mondini et al., 2017, 2011; Nichol and Wong, 2005). However, this method suffers from the salt-and-pepper effect due to single pixels being demarcated as landslides (Hölbling et al., 2017). In OBIA, segmentation is used to convert homogenous pixels into objects. Spectral, textural, morphological and topographical characteristics of these objects are utilized and have been found to detect landslides accurately with fewer false positives (Keyport et al., 2018; Moosavi et al., 2014).

OBIA has been used to map landslides around the world (Amatya et al., 2019; Barlow et al., 2006; Blaschke et al., 2014; Comert et al., 2019; Hölbling et al., 2020, 2015, 2012; Lahousse et al., 2011; Lu et al., 2011; Martha et al., 2012, 2011, 2010, 2016; Rau et al., 2014; Sun et al., 2017). Landslide detection using OBIA consists of two important steps, segmentation and classification. The main challenge in implementing a segmentation algorithm is determining the segmentation parameters. The segmentation parameters define the size of the objects. These parameters are often set using iterative trial and error methods and verified visually. Most of the studies utilizing OBIA for landslide detection involves creation of a knowledge-based classification ruleset for classifying segmented objects to landslides. The machine executable ruleset is created based on site-specific characteristics and manual thresholding of landslide diagnostic features (e.g. slope, relief).

With a goal of automating OBIA-based landslide mapping, various machine learning techniques and methods estimating segmentation parameters have been tested. Segmentation parameter estimation techniques such as Estimation of Scale Parameter (ESP) (Drăguţ et al., 2010), Plateau Objective Function (POF) (Martha et al., 2011), ESP-2 (Drăguţ et al., 2014) and Optimum Scale Parameter Selector (OSPS) (Vamsee et al., 2018) have been used. Machine learning methods such as Random Forests (Chen et al., 2017; Stumpf and Kerle, 2011; Tavakkoli Piralilou et al., 2019) and Support Vector Machine (Heleno et al., 2016) have been successfully utilized. Although these improvements have significantly advanced automated/semi-automated landslide detection methods, the accessibility of available algorithms is an issue because most of the above cited studies have relied on commercial software.

Open source tools reduce cost, offers users the flexibility to modify existing codes to best fit their application, and ease dissemination through code sharing services. Open source OBIA system development for broad application is an active field of research. Systems developed by Grippa et al. (2017) and Gonçalves et al. (2019) combine open source software such as GRASS GIS (https://grass.osgeo.org/), SAGA GIS (http://www.saga-gis.org) with R (www.r-project.org). Clewley et al. (2014) developed a Python based OBIA system with a ruleset based classification module. Knevels et al. (2019) combined GRASS GIS, SAGA GIS and R to establish an OBIA-

based workflow for landslide detection using LiDAR. InterIMAGE (Costa et al., 2010) is another OBIA-based open source software used to map landslides (Amato et al., 2019; Hölbling et al., 2017). These studies highlight multiple ways open source systems can be utilized to achieve results following OBIA principles. In this study, we create a Python-based open source Semi-Automatic Landslide Detection (SALaD) system combining OBIA and machine learning methods to eliminate the need for ruleset creation. SALaD was tested in 575 km^2 area along the Pasang Lhamu highway in Nepal where a large number of landsides were induced by the 2015 Gorkha earthquake using RapidEye (Planet Team, 2017) imagery.

2. Study Area

The study area (Fig. 1) covers 625 km^2 at the midpoint of the Pasang Lhamu highway in Nepal. The Pasang Lhamu highway connects Kathmandu, the capital of Nepal, with China and is very important for bilateral trade. The elevation in the area ranges from 604 to 4341 m and slope ranges from 0 to 76°, with mean value of 30°. Forest and agricultural area are the dominant landcover types (Uddin et al., 2015). Geologically, the area is composed of higher Himalayan crystalline rocks mostly schist, quartzite, gneiss, and migmatite and lesser Himalayan Proterozoic phyllite, amphibolite, metasandstone, and schist (Dhital, 2015).The average annual rainfall is about 691 mm (Tavakkoli Piralilou et al., 2019).

Nepal lies in a seismically active zone forming the Himalayan arc as a consequence of the continental collision between the India and Eurasia plate (Ader et al., 2012). Northward underthrusting of India beneath Eurasia has been historically generating large earthquakes in this region (Ambraseys and Douglas, 2004). The 2015 Gorkha earthquake (M_w = 7.8) which occurred on 25 April 2015 and a major aftershock (M_w = 7.2) that occurred on 12 May 2015 are the most recent major earthquakes experienced in this region. The study area was severely affected by landslides during the 2015 Gorkha earthquake (Gnyawali et al., 2019; Xu et al., 2017). The landslides triggered by the Gorkha earthquake are mainly shallow-disrupted with long and narrow runout (Roback et al., 2018).

We created an inventory of landslides triggered by various agents (e.g., earthquake and rainfall). Specifically, we manually mapped 623 landslides using the RapidEye imagery (Fig. 1). We did not differentiate landslide source and depositional areas and delineated them as a part of the same polygon. Landslides ranging in area from 260 to 496,426 m^2 were mapped. The study area was divided into two parts. A 50 km^2 area (yellow polygon in Fig. 1) was used for training the machine-learning model and estimating segmentation parameters. The remaining 575 km^2 area was used for validation of landslides detected by SALaD. 148 landslides were present in the training area and remaining 475 landslides were in the validation area.

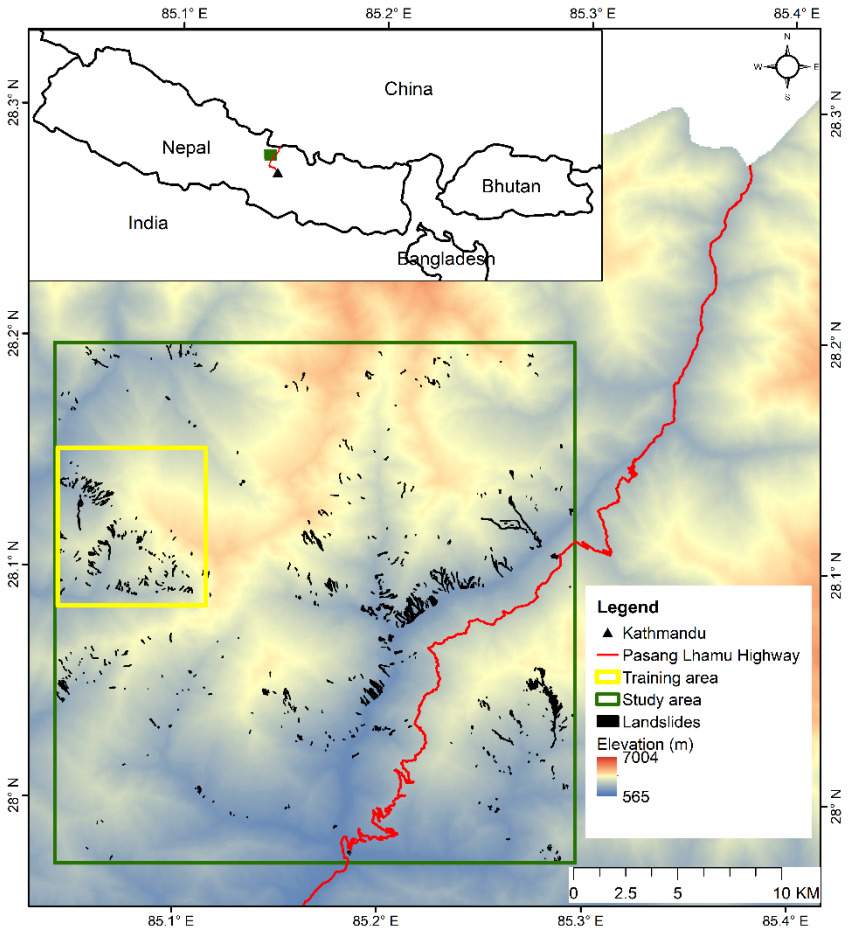

Fig.1 Location of the 625 km^2 study area along the Pasang Lhamu highway with manually mapped landslides. The yellow tile highlights the subset area used for training.

3. Data

We used an optical imagery from RapidEye acquired on 14 December 2015 covering the whole study area for landslide mapping. Level 3A data are available in 25 km × 25 km tiles with a spatial resolution of 5 m. RapidEye has 5 bands (blue, green, red, red edge and near-infrared). Four bands (blue,green,red and near-infrared) with top of atmosphere radiance values were used. The data was available through the Commercial Smallsat Data Acquisition Program (https://earthdata.nasa.gov/esds/csdap). The optical imagery was used in segmentation and creation of spectral and textural landslide diagnostic features. In addition to the optical data, we also utilized the 30 m NASADEM (Crippen et al., 2016), a reprocessed Shuttle Radar Topography Mission (SRTM) data with improve height accuracy and filled missing elevation data. The NASADEM was used to create topographical landslide diagnostic feature.

4. Methods

The SALaD system was developed in a Python environment using OBIA and machine learning methods and was configured to run in a Linux operating system (Fig. 2).This system was deployed in the Advanced Data Analytics Platform (ADAPT) at NASA Goddard Space Flight Center (GSFC)'s NASA Center for Climate Simulation (NCCS) (http://www.nccs.nasa.gov/services/adapt). We used a Linux virtual machine with 10 cores and 50 GB RAM. Details of individual processing steps are described below:

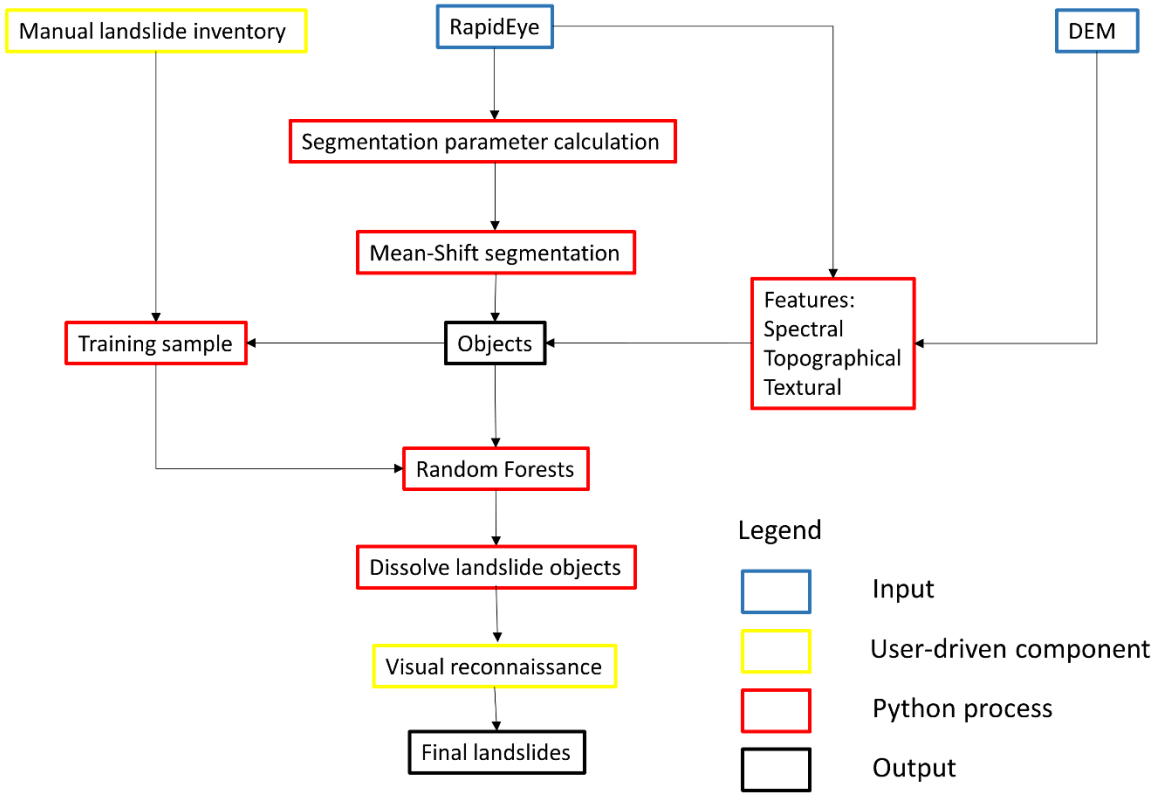

Fig. 2 Flowchart showing each step of the OBIA-based SALaD system. The blue boxes are inputs to the system; yellow boxes are manual steps; red boxes are the Python processes and black boxes are outputs.

4.1 Segmentation

Segmentation is the first and most important step in OBIA. In this study we use Mean-Shift segmentation available in the ORFEO Toolbox (OTB), an open source image processing library develop by CNES, the French Space Agency (Grizonnet et al., 2017). OTB can be accessed in Python through the otbApplication module. The Mean-Shift algorithm, first proposed by Fukunaga and Hostetler (1975), is a non-parametric density gradient estimator which shifts points in the feature space to the local maxima of the density function by effective iterations and can achieve fast convergence. Comaniciu and Meer (2002) extended it to image segmentation. OTB

implements a modified Mean-Shift segmentation algorithm known as Large-Scale Mean-Shift segmentation (Michel et al. 2014). It requires three parameters: the spatial radius (h_s) is the spatial distance between classes; the range radius (h_r) is the spectral difference between classes; and minimum object size is the minimum acceptable spatial size of the class. The mean shift algorithm segments the image by grouping pixels that are closer than h_s spatially and h_r spectrally.

The goal of image segmentation is to produce objects that are internally homogenous and heterogenous from its neighbours (Espindola et al., 2006). We used Plateau Objective Function (POF) proposed by Martha et al. (2011) for segmentation parameter calculation. This method uses the objective function proposed by Espindola et al. (2006). The function aims at maximizing intraobject homogeneity and interobject heterogeneity. Intraobject homogeneity is denoted by variance and interobject heterogeneity is denoted by spatial autocorrelation index Moran's I. The two parameters of the mean shift segmentation, h_s and minimum object size can be determined heuristically based on properties of landslides in an image. h_s can be set based on size of landslides. Larger h_s will significantly increase the computation time. For this study, the h_s was set as 10. The minimum object size can be set according to the smallest area of landslide to be mapped. We set the minimum object size as 10. Based on a single pixel area of 25 m^2, this minimum object size will ensure landslides with an area of 250 m^2 and above will be mapped. Keeping these two parameters constant, the range radius was varied to quantify object homogeneity and heterogeneity. Since this parameter estimation method requires segmenting an image with varying h_r, implementing it over the whole satellite image is very time consuming. The segmentation parameter estimation was carried out in a subset area of size 50 km^2 (yellow polygon in Fig. 1) that contained 148 landslides. The calculated segmentation parameter was then applied over the whole imagery for final segmentation.

Segmentations were carried out with h_s of 10, h_r varying from 2 to 30 with a step size of 2 and minimum object size of 10. For each segmentation output, mean of brightness image (average of blue, green and red pixel values) for each objects were used for calculation of variance and Moran' I. The intraobject homogeneity is calculated using weighted variance:

$$v = \frac{\sum_{i=1}^{n} a_i v_i}{\sum_{i=1}^{n} a_i} \quad (1)$$

where a_i and v_i are the area and intraobject variance of segment i, respectively.

Moran's I index is expressed as

$$I = \frac{n}{S_o} * \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j} z_i z_j}{\sum_{i=1}^{n} {z_i}^2} \quad (2)$$

where z_i is the deviation of the brightness value of object i from its mean $x_i - \bar{X}$, $w_{i,j}$ is the spatial weight between objects i and j, which is one for adjacent regions or zero otherwise, n is the total number of objects, and

S_o is the aggregate of all spatial weights

$$S_o = \sum_{i=1}^{n}\sum_{j=}^{n} w_{i,j} \quad (3)$$

The objective function is defined as

$$F(v, I) = F(v) + F(I) \quad (4)$$

Functions $F(v)$ and $F(I)$ are normalization functions

$$F(x) = \frac{X_{max} - X}{X_{max} - X_{min}} \quad (5)$$

The POF is created using

$$F(plateau) = F(v, I)_{max} - \sigma \quad (6)$$

where $F(v, I)_{max}$ max is the maximum value and σ is the standard deviation of the objective functions calculated from varying h_r.

The POF value demarcates the lower boundary of the plateau in the curve created by plotting h_r values and corresponding objective functions in the *x*- and *y*-axes, respectively. The optimal h_r values are the ones that correspond to objective functions above the plateau. The objects created using the optimal h_r values will have high external and low internal heterogeneity levels (Van Den Eeckhaut et al., 2012). In this study, the h_r value corresponding to the first peak above the plateau was chosen as a suitable value for optimal image segmentation. This h_r value creates smaller objects than subsequent h_r values, and, hence, has the maximum chance of capturing small landslides (Martha et al., 2012).

The Python packages, rasterstats (Perry, 2017) and GeoPandas (Jordahl, 2014) were used for weighted variance calculation. Moran's I was calculated using PySAL (Rey and Anselin, 2010) package. h_s of 10, optimal h_r calculated from the POF and minimum segment size of 10 was used to segment the RapidEye image. The output of this step is a shapefile with segmented objects.

4.2 Object Metrics

Stumpf and Kerle (2011) tested various landslide diagnostic features in different part of the world and found it to be site specific. In order to avoid creation of the excessive features, we used five features that were successful in classifying landslides in similar mountainous environment of Nepal (Amatya et al., 2019) and India (Martha et al., 2010). The spectral features used are normalized difference vegetation index (NDVI) and brightness. Textural features used are the all-direction Grey-level Co-occurrence Matrix (GLCM) mean and homogeneity computed from red band of RapidEye using Haralick's method (Haralick and Shanmugam, 1973). The topographical feature, slope, was calculated from the 30 m NASADEM. Mean of each feature was calculated for all objects and used for landslide detection.

OTB was used to create Haralick texture features. The GDAL (GDAL/OGR Contributors 2019) library was used for creation of spectral features. The RichDEM (Barnes, 2016) and rasterstats Python packages were used for calculation of slope from NASADEM and calculation of object means respectively.

4.3 Machine Learning

In this study we used Random Forests (RF), a non-parametric ensemble learning method for classification or regression based on several decision trees (Breiman, 2001). We chose RF because it is fast to train and has been successfully used for landslide detection (Chen et al., 2018; Ghorbanzadeh et al., 2019; Stumpf and Kerle, 2011; Yu et al., 2018). RF grows multiple decision trees on random subsets of the training data and related features. The final classifications are made by averaging the classification of each individual tree. RF can also estimate the importance of used features for the accuracy of a given classification. We used the RF classifier available in the Scikit-learn package (Pedregosa et al., 2011). We used a RF model with 500 trees.

The RF was trained on the same area used for segmentation parameter estimation (yellow polygon in Fig. 1). Manually mapped landslide polygons in the training area were used to create the training dataset. 148 landslides were present in the training area. The segmented objects that had an overlap of 50% or more with the manual landslides were chosen as landslide objects, and all other objects within the training area were set as non-landslide objects (Stumpf and Kerle, 2011). The training data consisted of 1727 landslide and 15,751 non-landslide objects. Based on this data, a RF classifier was trained, and remaining objects in the validation area were classified.

4.4 Post-processing

Classified landslide objects were dissolved to obtain final landslide areas. A visual check was run to remove some obvious false positives such as roads, riverbanks, houses, and barren areas.

5. Results

5.1 Segmentation

Fig. 3 shows the plot of objective functions for different values of h_r. Using the POF value determined from Eq. (6), the plateau boundary was identified as 1.17. The h_r values with objective functions greater than this plateau represents the optimal parameters. We selected the first peak above the plateau line, h_r = 6 as the optimal parameter. The Mean-Shift segmentation was conducted with h_s as 10, h_r as 6 and mean object size as 10. Fig. 4 shows the results of segmentation using the segmentation parameters. Landslide objects were clearly separated from other objects.

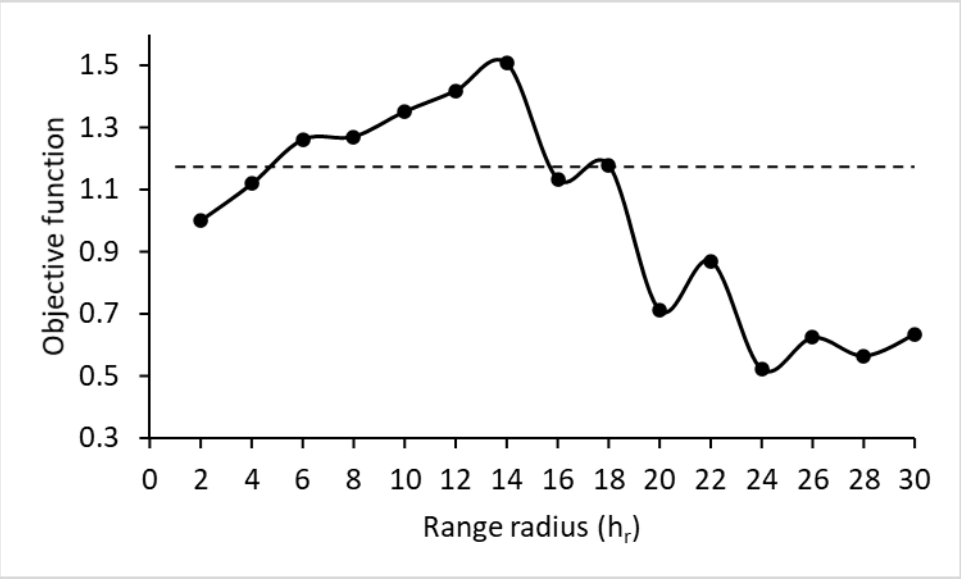

Fig. 3 Objective function curve for the RapidEye image. The horizontal dashed line marks the POF value which demarcates the plateau line. The peaks above the plateau line were used for selecting optimal h_r. Based on first peak above the plateau line h_r of 6 was chosen.

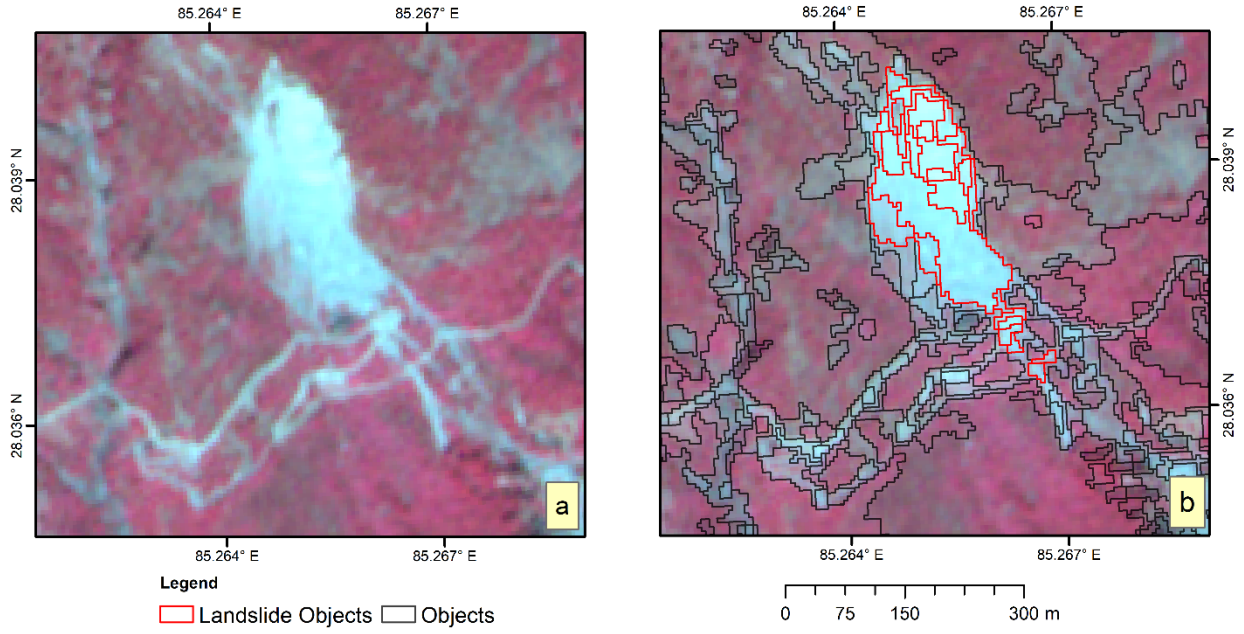

Fig. 4 a. Landslide in the study area; b. Segmented objects using h_s of 10, h_r of 6 and minimum object size of 10.

5.2 Feature importance

Permutation importance (Breiman, 2001) was used to evaluate feature importance (brightness, NDVI, GLCM mean, GLCM homogeneity and slope). The feature importance was calculated using the training dataset. In order to give equal weight to both classes, all landslide objects and an equal number of randomly sampled non-landslide objects were taken. 70% of the sample was used to create the RF model and 30% was held for validation. The importance of a feature is calculated based on the increase in the model's error after permuting the feature. Each feature was permuted 50 times. Fig. 5 shows the feature importance plot for the landslide detection. Brightness is the most important feature for classification of landslides, while slope is the least.

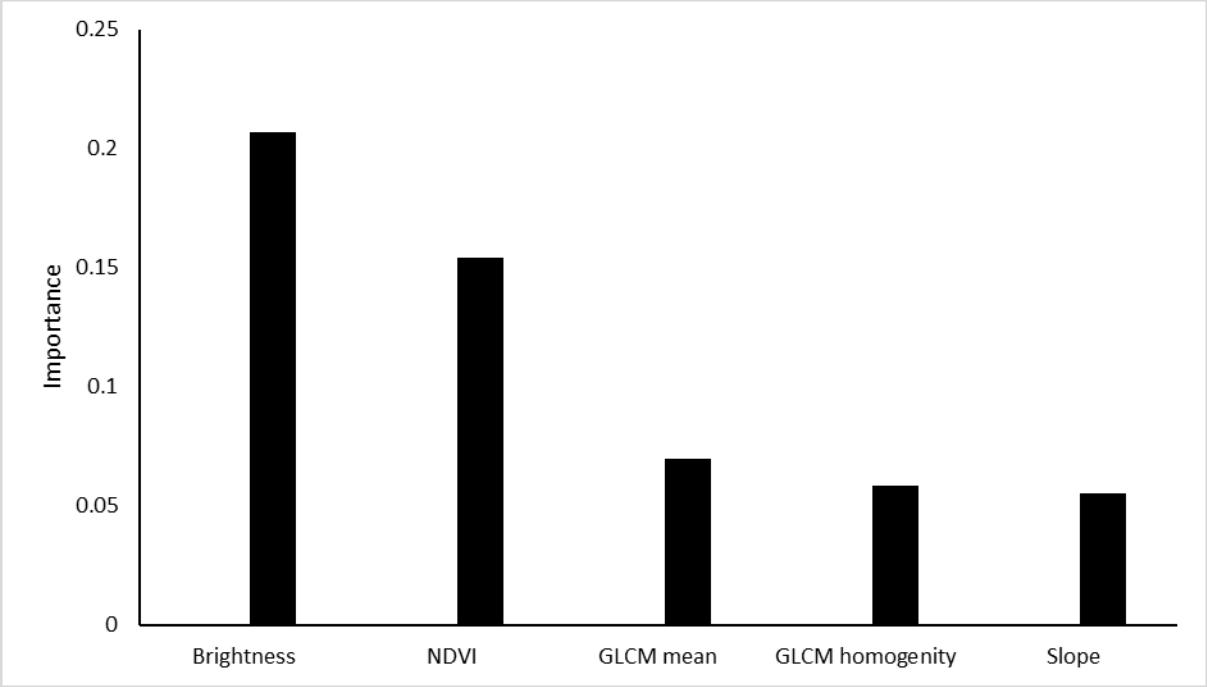

Fig. 5 Landslide detection feature importance plot.

5.3 Landslide validation

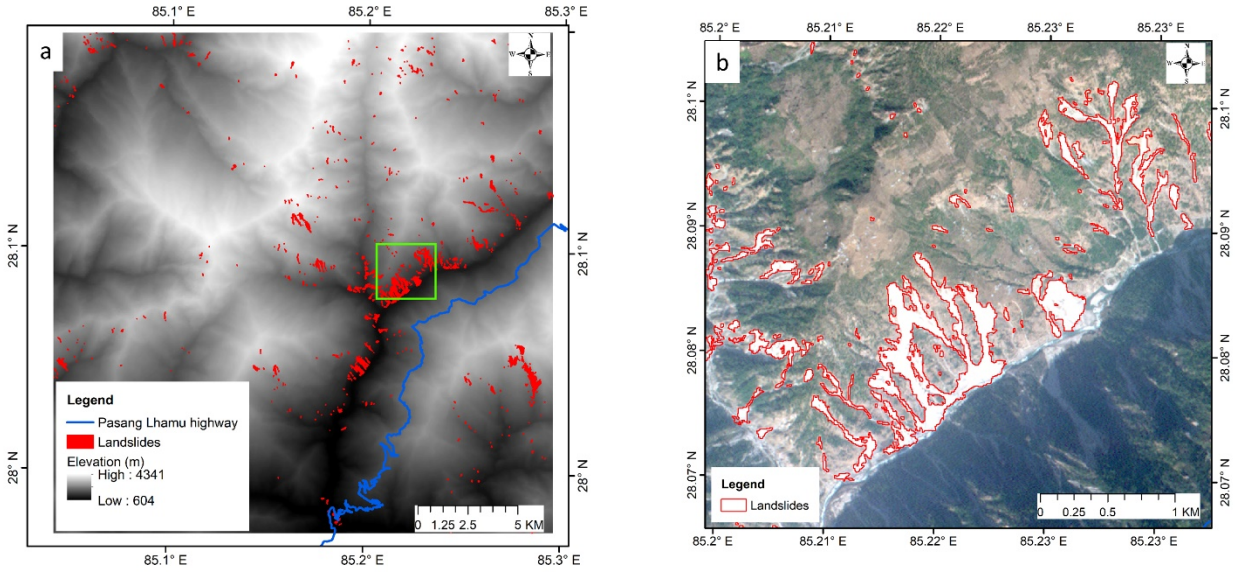

Fig. 6 a. Spatial distribution of SALaD-detected landslides in the study area; b. Zoomed in view of green box in Fig. 6 a, showing SALaD-detected landslides. Landslides within training area are not shown.

Fig. 6 a shows the SALaD-detected landslides for the 575 km^2 validation area. Fig 6 b shows a zoomed in view of landslides. In order to validate SALaD, the detected landslides were compared with 475 manually mapped landslides not used in training. Three metrics based on overlapping area were calculated: true positive (TP), false negative (FN), and false positive (FP). TPs are correctly detected landslides; FPs are detected landslides that have not been mapped manually and FNs are manually mapped landslides not detected by the SALaD. These metrics were not based on the number of landslides because segmentation-derived image objects rarely correspond to single landslides due to over or under-segmentation (Hölbling et al., 2016; Rau et al., 2014). Based on these metrics, the two accuracy indices, producer accuracy (PA) and user accuracy (UA) were calculated as follows:

$$PA = \frac{TP}{TP+FN} * 100 \quad (7)$$

$$UA = \frac{TP}{TP+FP} * 100 \quad (8)$$

The PA denotes how much of the manual inventory was detected. The UA denotes how much of the detected landslides are actual landslides. It was observed that the SALaD was successful in detecting 70% of the area of the manually mapped inventory (Table 1). UA was higher than PA, which suggest an underestimation of TP and an overestimation of FP. When evaluating based on the intersection of any portion of the manual and SALaD-detected landslides, a PA of 91% was obtained. This suggests that most landslides were detected using SALaD, but the areas were not always accurate.

Table 1. Comparison of SALaD and manually mapped landslides based on overlapping landslide area.

True Positive (m^2)	False Positive (m^2)	False Negative (m^2)	Producer Accuracy (%)	User Accuracy (%)
2,687,530	698,151.2	1,127,140	70.45	79.38

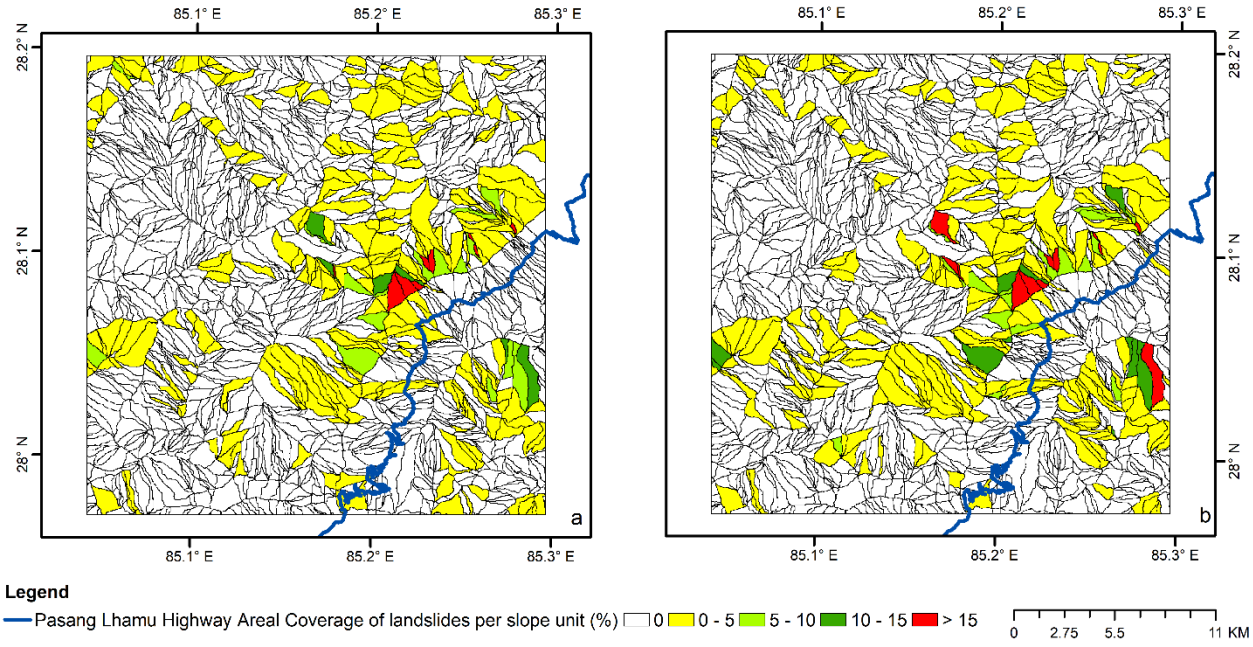

Fig. 7 Areal coverage of landslides mapped: a. SALaD; b. manual. Landslides within training area are not considered.

Additionally, we compared density maps based on areas of landslides detected by the SALaD and manually mapped inventory (Galli et al., 2008; Martha et al., 2012). We use Slope Unit (SU) as a landscape partition method. It is an alternative mapping unit to grid cells and suggested as a well suited terrain subdivision for landslide susceptibility modeling (Carrara, 1988; Guzzetti et al., 2006, 1999). For the delineation of SUs, we use r.slopeunits, an open source software developed by Alvioli et al. (2016). Using this software, SUs subdividing the terrain between streamlines and ridges under the constrain of similar slope and aspect conditions were identified. 1470 SUs were obtained from the 30 m NASADEM. The landslide density (i.e., percentage of SU area intersected by landslides) was calculated by intersecting the landslide polygons of SALaD and manual inventories with the SUs. Fig. 7 a and b shows the SU-based density maps created using SALaD-detected and manually mapped landslides. Although small variations exist, overall, the spatial distribution of landside area coverage is similar in both cases. Comparison of total area of landslides within each SUs for two inventories was also conducted using a scatter plot. The area of landslides within the SUs for SALaD-detected and manually mapped landslides were also in close agreement ($R^2 = 0.96$) (Fig. 8).

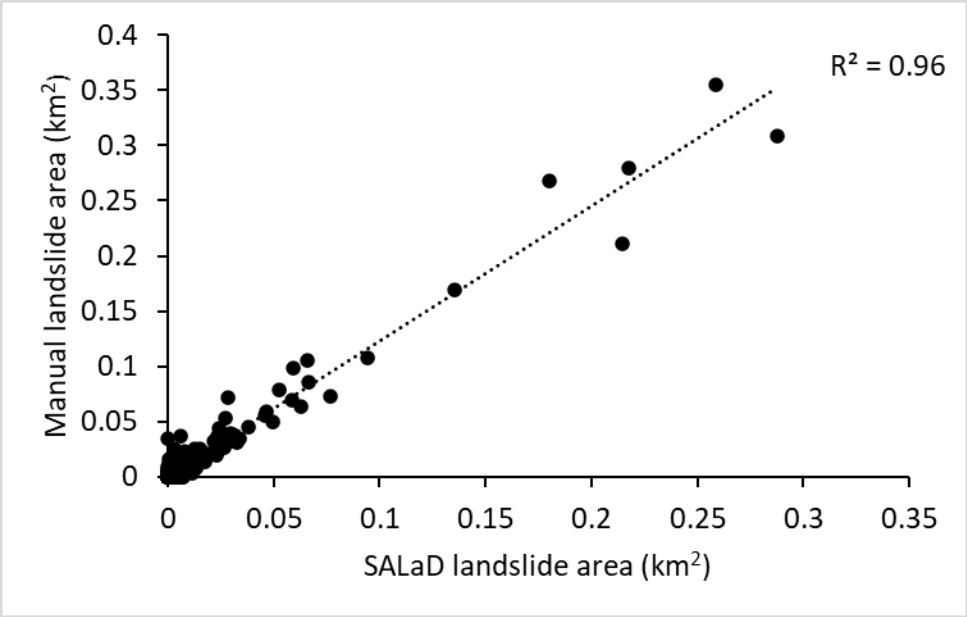

Fig. 8 1:1 comparison of manual and SALaD-detected landslide area under each SU.

6. Discussion

A Python-based open source SALaD system was developed combining OBIA and machine learning. One challenge with OBIA-based detection methods is the calculation of segmentation parameters, which often require significant trial and error as well as visual comparison. To address this issue, we used the POF method to determine segmentation parameters. In Fig. 3, all h_r values corresponding to objectives functions above the plateau are optimal h_r. In order to map small landslides, we chose h_r corresponding to first peak above the plateau. The segmentation did result in some over segmentation (Fig. 4). Although the purpose of segmentation is to create meaningful objects, over segmentation is not necessarily a bad result. Generally, over segmentation is preferred to under segmentation because merging is possible in the final steps (Martha et al., 2010). A user can run segmentation with all suggested h_r values and select the best one. One must note that use of higher h_r values will result in larger objects, which can possibly limit the availability of objects for training purposes. Depending upon the area of landslide to be mapped, increasing the minimum object size will also solve this. The values of h_s and minimum object size are site specific depending upon the size of the landslides whereas h_r is scene specific depending upon the range of the pixel values.

SALaD was tested by mapping landslides in 575 km^2 area along the Pasang Lhamu highway in Nepal. SALaD was able to detect 70% of landslide areas when compared with a manual inventory. The SALaD system detected landslide scarps and large landslides well but failed to detect long narrow runouts (Fig. 9). These types of features lack distinct spectral characteristics as such segmentation was not able to form proper objects. Occlusion due to topography and remaining vegetation led to omission as well. Difficulty mapping flow type landslides for similar reasons has been found in other studies as well (Barlow et al., 2006; Hölbling et al., 2016; Stumpf and Kerle, 2011). Spectral features were the most important feature for landslide mapping (Fig. 5). Some

objects belonging to landslides that lacked distinct brightness and NDVI values were not classified as well. Another source of uncertainty is introduced by the manual mapping procedure, which is the main reference dataset used to train the machine learning model and validate detected landslides. However, the manual mapping process can also be highly subjective depending upon the person doing the mapping and method used (Ardizzone et al., 2002; Carrara et al., 1992; Fookes et al., 1991; Guzzetti et al., 2012; Soeters and Van Westen, 1996). For instance, in an area affected by the 2010 Haiti earthquake, two research groups (Gorum et al., 2013; Harp et al., 2016) examined images having similar spatial resolution and mapped co-seismic landslides using similar visual image-interpretation techniques (Tanyaş et al., 2019). However, in an area of 4.9 km^2 areal coverage, Harp et al. (2016) map 16,379 more landslides than the other research group. Also, within the same areal coverage, 18 km^2 of landslides mapped by Harp et al. (2016) was not identified as landslide by Gorum et al. (2013) mostly because of the difference in their mapping preferences. Errors might have also been introduced during the creation of training data. We implemented an automatic labeling method to generate training data by labeling objects with an overlap of 50% or more with manual landslides as landslide objects and the remaining as non-landslide objects within the training area. This strategy minimizes the overall amount of mislabeling by preventing objects with smaller overlaps from being labeled as landslide objects especially near the boundaries of the manual landslide polygons but there is still a chance of some marginal cases being mislabeled. Other studies have used an overlap percentage ranging from 50 to 75% (Chen et al., 2017; Knevels et al., 2019; Li et al., 2015; Stumpf and Kerle, 2011; Tavakkoli Piralilou et al., 2019). One should check the created training data and select the value accordingly. To avoid automating this step, training data could be created by manually selecting landslide and non-landslide objects from the segmented objects, but this could be very time consuming, depending upon the size and number of objects.

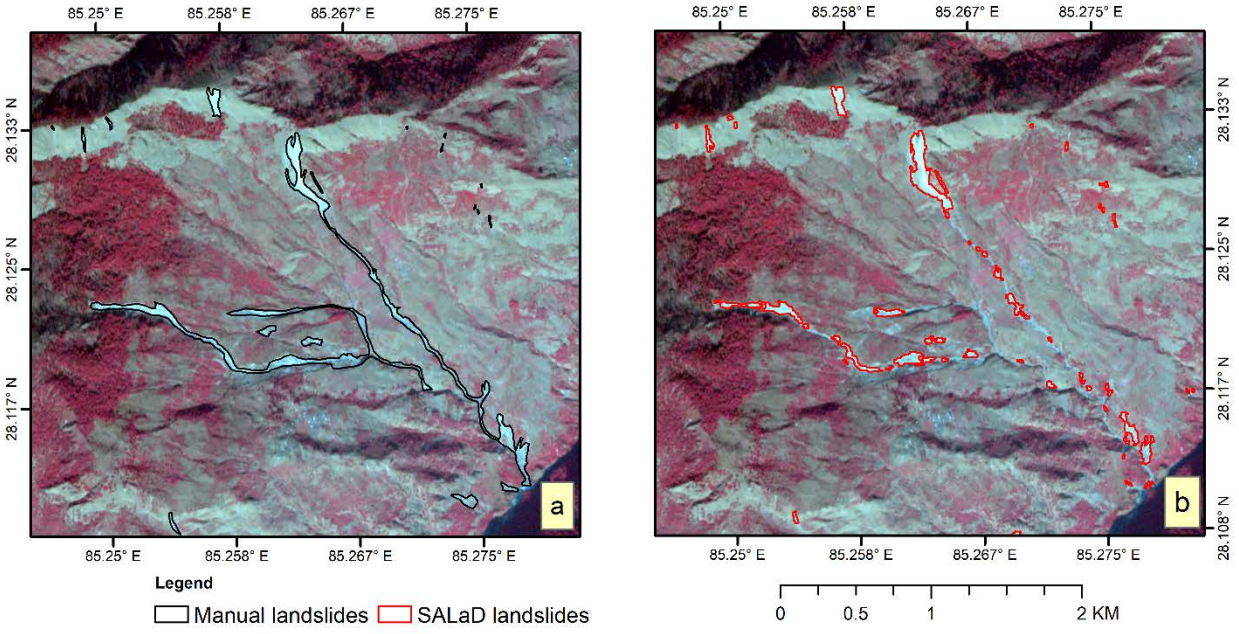

Fig. 9 Insets highlighting landslides mapped: a. manually; b. detected by the SALaD.

Manual mapping of landslides within the study area took us about six hours (one hour in training and five hours in validation area). With one hour of landslide mapping in training area, we were able to complete the processing chain in 45 min. Removal of false positives such as roads, riverbanks, houses, and barren areas added another 15 min. SALaD was able to detect landslides in about an hour. Inventories produced by the SALaD system can provide initial assessment of landslides areas but should not be directly used for local hazard and risk studies without manual corrections. For applications where the approximate density of landsliding must be determined over a large affected area, the SALaD system can produce results much faster than manual mapping. It can also be coupled with slope units to better identify areas where roadcuts and/or infrastructures are likely to be affected by landslides after an earthquake or a storm. Clouds can be a severe limiting factor as SALaD uses optical imagery. Synthetic Aperture Radar (SAR) an active sensor operating in microwave frequency, penetrates clouds and provide day and night coverage. SAR and OBIA has been used to map landslides (Plank et al., 2015). Development of open source method (Esposito et al., 2020) to map landslide using SAR has opened new avenues to provide rapid response products. This is a viable alternative which can be updated in detail once the clouds begin to clear using optical imagery. However, in mountainous environments shadow and layover effects can limit the applicability of SAR.

The processing chain can be easily adopted for landslide mapping in similar areas where landslides are synonymous with change in vegetation using optical imagery. The system configured for this study might not work for every case, so changes needs to be made according to the use case. A challenge with previous studies is that a variety of landslide diagnostic features were used without providing justification on why each of them was used and what value they added to the classification. While five features (NDVI, brightness, slope, GLCM mean and GLCM homogeneity) were effective in the Nepal case, the required features may be different in other study areas. For those instances, additional features can be used and RF can be applied to gauge variable importance and drop unimportant features for final classification. This step has also been shown to improve accuracy (Stumpf and Kerle, 2011) and simplify the processing. Brightness was the most important feature, while slope was the least (Fig 5). Brightness has been successfully utilized as a main diagnostic feature to map landslides (Hölbling et al., 2016; Martha et al., 2012). Rau et al. (2014) found brightness was sometimes more significant than the NDVI when filtering out vegetation areas and bare soil. We used a coarse resolution pre-event NASADEM in our study area. Landslide signatures are not well represented in coarse pre-event DEMs which explains its low importance for landslide classification. Hence, use of a high resolution or post-event DEM will be useful. However, Stumpf and Kerle (2011) found feature importance to be site and segmentation parameter dependent. Hence, this might not be the case everywhere.

The possibility of implementing SALaD in parallel offers an opportunity to efficiently map landslides for large areas. Speed improvements can be made by configuring SALaD in a processing unit based on Graphical Processing Unit. Tavakkoli Piralilou et al. (2019) highlighted that an ensemble of machine learning algorithms offers better accuracy in landslide mapping than a single machine learning algorithm. In a rapid response scenario, the ability to separate event-triggered landslides from existing landslides will be very beneficial for identifying the most significant clusters of landslides and can be used to guide transportation and distribution planning for resources to affected areas. If pre- and post-event images are available, a change detection-based approach will be beneficial for mapping new or event-triggered landslides only. Addition of image co-registration, image normalization steps and use of change detection-based metrics such as

NDVI difference, Principal Component Analysis, Independent Component Analysis etc. (Lu et al., 2019; Martha et al., 2016, 2013; Mondini et al., 2013, 2011) can help achieve this. Advancement of the system in terms of these changes will be carried out in the future. As new and advanced Python packages becomes available, the flexibility of SALaD system enables complete replacement of existing packages or addition into the existing framework. We were also able to configure and run SALaD in a Windows environment.

7. Conclusion

The SALaD system was created by combining various open source Python packages and modules. The OBIA and machine learning-based method was tested mapping landslides in an area along the Pasang Lhamu highway in Nepal. Results show that the SALaD inventory has an accuracy of 70% compared to the manual inventory when looking specifically at landslide area. The SALaD system performs well in terms of detecting landslide initiation areas and large landslides but had trouble detecting narrow landslide runouts lacking distinct spectral characteristics and occluded by topography and remaining vegetation. Therefore, we argue that the proposed method provides a viable alternative to manual techniques for establishing landslide locations within large areas quickly and efficiently. With further manual corrections, the inventories could be significantly improved for more detailed analyses and research-quality assessments. This system can also be configured according to need by removing or adding new components to the current processing chain.

Acknowledgments

This research was funded by the NASA Understanding Changes in High Mountain Asia Program (NNH15ZDA001N-HMA), the NASA Disaster Risk Reduction and Response Program (NNH18ZDA001N-DISASTERS) and the NASA Commercial Smallsat Data Acquisition Program. We would like to thank Mark Carroll, Jian Li and Glenn Tamkin from NASA Center for Climate Simulation (NCCS) for their help in optimizing the code. We would also like to thank authors of the Python packages used in this research.

References

Ader, T., Avouac, J., Liu-Zeng, J., Lyon-Caen, H., Bollinger, L., Galetzka, J., Genrich, J., Thomas, M., Chanard, K., Sapkota, S.N., 2012. Convergence rate across the Nepal Himalaya and interseismic coupling on the Main Himalayan Thrust: Implications for seismic hazard. J. Geophys. Res. Solid Earth 117.

Alvioli, M., Marchesini, I., Reichenbach, P., Rossi, M., Ardizzone, F., Fiorucci, F., Guzzetti, F., 2016. Automatic delineation of geomorphological slope units with r. slopeunits v1. 0 and their optimization for landslide susceptibility modeling. Geosci. Model Dev. 9, 3975.

Amato, G., Eisank, C., Castro-Camilo, D., Lombardo, L., 2019. Accounting for covariate distributions in slope-unit-based landslide susceptibility models. A case study in the alpine

environment. Eng. Geol. 260, 105237.

Amatya, P., Kirschbaum, D., Stanley, T., 2019. Use of Very High-Resolution Optical Data for Landslide Mapping and Susceptibility Analysis along the Karnali Highway, Nepal. Remote Sens. 11, 2284.

Ambraseys, N.N., Douglas, J., 2004. Magnitude calibration of north Indian earthquakes. Geophys. J. Int. 159, 165–206.

Ardizzone, F., Cardinali, M., Carrara, A., Guzzetti, F., Reichenbach, P., 2002. Impact of mapping errors on the reliability of landslide hazard maps.

Barlow, J., Franklin, S., Martin, Y., 2006. High spatial resolution satellite imagery, DEM derivatives, and image segmentation for the detection of mass wasting processes. Photogramm. Eng. Remote Sens. 72, 687–692.

Barnes, R., 2016. RichDEM: Terrain Analysis Software.

Blaschke, T., Feizizadeh, B., Hölbling, D., 2014. Object-Based Image Analysis and Digital Terrain Analysis for Locating Landslides in the Urmia Lake Basin, Iran. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7, 4806–4817.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Carrara, A., 1988. Drainage and divide networks derived from high-fidelity digital terrain models, in: Quantitative Analysis of Mineral and Energy Resources. Springer, pp. 581–597.

Carrara, A., Cardinali, M., Guzzetti, F., 1992. Uncertainty in assessing landslide hazard and risk. ITC J. 172–183.

Chen, F., Yu, B., Li, B., 2018. A practical trial of landslide detection from single-temporal Landsat8 images using contour-based proposals and random forest: a case study of national Nepal. Landslides 15, 453–464.

Chen, T., Trinder, C.J., Niu, R., 2017. Object-Oriented Landslide Mapping Using ZY-3 Satellite Imagery, Random Forest and Mathematical Morphology, for the Three-Gorges Reservoir, China. Remote Sens. . https://doi.org/10.3390/rs9040333

Clewley, D., Bunting, P., Shepherd, J., Gillingham, S., Flood, N., Dymond, J., Lucas, R., Armston, J., Moghaddam, M., 2014. A python-based open source system for geographic object-based image analysis (GEOBIA) utilizing raster attribute tables. Remote Sens. 6, 6111–6135.

Comaniciu, D., Meer, P., 2002. Mean shift: A robust approach toward feature space analysis. IEEE Trans. Pattern Anal. Mach. Intell. 24, 603–619.

Comert, R., Avdan, U., Gorum, T., Nefeslioglu, H.A., 2019. Mapping of shallow landslides with object-based image analysis from unmanned aerial vehicle data. Eng. Geol. 260, 105264.

Contributors, G., 2019. GDAL/OGR Geospatial Data Abstraction software Library.

Corominas, J., van Westen, C., Frattini, P., Cascini, L., Malet, J.-P., Fotopoulou, S., Catani, F., Van Den Eeckhaut, M., Mavrouli, O., Agliardi, F., 2014. Recommendations for the

quantitative analysis of landslide risk. Bull. Eng. Geol. Environ. 73, 209–263.

Costa, G., Feitosa, R.Q., Fonseca, L.M.G., Oliveira, D.A.B., Ferreira, R.S., Castejon, E.F., 2010. Knowledge-based interpretation of remote sensing data with the InterIMAGE system: major characteristics and recent developments. Proc. 3rd GEOBIA.

Crippen, R., Buckley, S., Belz, E., Gurrola, E., Hensley, S., Kobrick, M., Lavalle, M., Martin, J., Neumann, M., Nguyen, Q., 2016. NASADEM global elevation model: methods and progress. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 41, 125–128.

Dhital, M.R., 2015. Geology of the Nepal Himalaya: regional perspective of the classic collided orogen. Springer.

Drăguţ, L., Csillik, O., Eisank, C., Tiede, D., 2014. Automated parameterisation for multi-scale image segmentation on multiple layers. ISPRS J. Photogramm. Remote Sens. 88, 119–127.

Drăguţ, L., Tiede, D., Levick, S.R., 2010. ESP: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. Int. J. Geogr. Inf. Sci. 24, 859–871.

Espindola, G.M., Câmara, G., Reis, I.A., Bins, L.S., Monteiro, A.M., 2006. Parameter selection for region-growing image segmentation algorithms using spatial autocorrelation. Int. J. Remote Sens. 27, 3035–3040.

Esposito, G., Marchesini, I., Mondini, A.C., Reichenbach, P., Rossi, M., Sterlacchini, S., 2020. A spaceborne SAR-based procedure to support the detection of landslides. Nat. Hazards Earth Syst. Sci. 20, 2379–2395. https://doi.org/10.5194/nhess-20-2379-2020

Fan, X., Scaringi, G., Korup, O., West, A.J., van Westen, C.J., Tanyas, H., Hovius, N., Hales, T.C., Jibson, R.W., Allstadt, K.E., 2019. Earthquake-induced chains of geologic hazards: Patterns, mechanisms, and impacts. Rev. Geophys. 57, 421–503.

Fookes, P.G., Dale, S.G., Land, J.M., 1991. Some observations on a comparative aerial photography interpretation of a landslipped area. Q. J. Eng. Geol. Hydrogeol. 24, 249–265.

Fukunaga, K., Hostetler, L., 1975. The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Trans. Inf. theory 21, 32–40.

Galli, M., Ardizzone, F., Cardinali, M., Guzzetti, F., Reichenbach, P., 2008. Comparing landslide inventory maps. Geomorphology 94, 268–289. https://doi.org/https://doi.org/10.1016/j.geomorph.2006.09.023

Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, R.S., Tiede, D., Aryal, J., 2019. Evaluation of Different Machine Learning Methods and Deep-Learning Convolutional Neural Networks for Landslide Detection. Remote Sens. https://doi.org/10.3390/rs11020196

Gnyawali, K.R., Zhang, Y., Wang, G., Miao, L., Pradhan, A.M.S., Adhikari, B.R., Xiao, L., 2019. Mapping the susceptibility of rainfall and earthquake triggered landslides along China–Nepal highways. Bull. Eng. Geol. Environ. 1–15.

Gonçalves, J., Pôças, I., Marcos, B., Mücher, C.A., Honrado, J.P., 2019. SegOptim—A new R

package for optimizing object-based image analyses of high-spatial resolution remotely-sensed data. Int. J. Appl. Earth Obs. Geoinf. 76, 218–230.

Gorum, T., van Westen, C.J., Korup, O., van der Meijde, M., Fan, X., van der Meer, F.D., 2013. Complex rupture mechanism and topography control symmetry of mass-wasting pattern, 2010 Haiti earthquake. Geomorphology 184, 127–138.

Grippa, T., Lennert, M., Beaumont, B., Vanhuysse, S., Stephenne, N., Wolff, E., 2017. An open-source semi-automated processing chain for urban object-based classification. Remote Sens. 9, 358.

Grizonnet, M., Michel, J., Poughon, V., Inglada, J., Savinaud, M., Cresson, R., 2017. Orfeo ToolBox: open source processing of remote sensing images. Open Geospatial Data, Softw. Stand. 2, 1–8.

Guzzetti, F., Carrara, A., Cardinali, M., Reichenbach, P., 1999. Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. Geomorphology 31, 181–216.

Guzzetti, F., Mondini, A.C., Cardinali, M., Fiorucci, F., Santangelo, M., Chang, K.-T., 2012. Landslide inventory maps: New tools for an old problem. Earth-Science Rev. 112, 42–66.

Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M., Galli, M., 2006. Estimating the quality of landslide susceptibility models. Geomorphology 81, 166–184.

Haralick, R.M., Shanmugam, K., 1973. Textural features for image classification. IEEE Trans. Syst. Man. Cybern. SMC-3, 610–621.

Harp, E.L., Jibson, R.W., Schmitt, R.G., 2016. Map of landslides triggered by the January 12, 2010, Haiti earthquake. US Geol. Surv. Sci. Investig. Map 3353, 15.

Heleno, S., Matias, M., Pina, P., Sousa, A.J., 2016. Semiautomated object-based classification of rain-induced landslides with VHR multispectral images on Madeira Island. Nat. Hazards Earth Syst. Sci 16, 1035–1048.

Hölbling, D., Abad, L., Dabiri, Z., Prasicek, G., Tsai, T.-T., Argentin, A.-L., 2020. Mapping and Analyzing the Evolution of the Butangbunasi Landslide using Landsat Time Series with Respect to Heavy Rainfall Events during Typhoons. Appl. Sci. 10, 630.

Hölbling, D., Betts, H., Spiekermann, R., Phillips, C., 2016. Identifying Spatio-Temporal Landslide Hotspots on North Island, New Zealand, by Analyzing Historical and Recent Aerial Photography. Geosci. . https://doi.org/10.3390/geosciences6040048

Hölbling, D., Eisank, C., Albrecht, F., Vecchiotti, F., Friedl, B., Weinke, E., Kociu, A., 2017. Comparing Manual and Semi-Automated Landslide Mapping Based on Optical Satellite Images from Different Sensors. Geosci. . https://doi.org/10.3390/geosciences7020037

Hölbling, D., Friedl, B., Eisank, C., 2015. An object-based approach for semi-automated landslide change detection and attribution of changes to landslide classes in northern Taiwan. Earth Sci. Informatics 8, 327–335.

Hölbling, D., Füreder, P., Antolini, F., Cigna, F., Casagli, N., Lang, S., 2012. A semi-automated

object-based approach for landslide detection validated by persistent scatterer interferometry measures and landslide inventories. Remote Sens. 4, 1310–1336.

Jordahl, K., 2014. GeoPandas: Python tools for geographic data.

Keyport, R.N., Oommen, T., Martha, T.R., Sajinkumar, K.S., Gierke, J.S., 2018. A comparative analysis of pixel-and object-based detection of landslides from very high-resolution images. Int. J. Appl. Earth Obs. Geoinf. 64, 1–11.

Knevels, R., Petschko, H., Leopold, P., Brenning, A., 2019. Geographic Object-Based Image Analysis for Automated Landslide Detection Using Open Source GIS Software. ISPRS Int. J. Geo-Information 8, 551.

Lahousse, T., Chang, K.T., Lin, Y.H., 2011. Landslide mapping with multi-scale object-based image analysis–a case study in the Baichi watershed, Taiwan. Nat. Hazards Earth Syst. Sci. 11, 2715–2726.

Li, X., Cheng, X., Chen, W., Chen, G., Liu, S., 2015. Identification of Forested Landslides Using LiDar Data, Object-based Image Analysis, and Machine Learning Algorithms. Remote Sens. . https://doi.org/10.3390/rs70809705

Lu, P., Qin, Y., Li, Z., Mondini, A.C., Casagli, N., 2019. Landslide mapping from multi-sensor data through improved change detection-based Markov random field. Remote Sens. Environ. 231, 111235.

Lu, P., Stumpf, A., Kerle, N., Casagli, N., 2011. Object-oriented change detection for landslide rapid mapping. IEEE Geosci. Remote Sens. Lett. 8, 701–705.

Martha, T.R., Kamala, P., Jose, J., Vinod Kumar, K., Jai Sankar, G., 2016. Identification of new Landslides from High Resolution Satellite Data Covering a Large Area Using Object-Based Change Detection Methods. J. Indian Soc. Remote Sens. 44, 515–524.

Martha, T.R., Kerle, N., Jetten, V., van Westen, C.J., Kumar, K.V., 2010. Characterising spectral, spatial and morphometric properties of landslides for semi-automatic detection using object-oriented methods. Geomorphology 116, 24–36.

Martha, T.R., Kerle, N., van Westen, C.J., Jetten, V., Kumar, K.V., 2012. Object-oriented analysis of multi-temporal panchromatic images for creation of historical landslide inventories. ISPRS J. Photogramm. Remote Sens. 67, 105–119.

Martha, T.R., Kerle, N., van Westen, C.J., Jetten, V., Kumar, K.V., 2011. Segment optimization and data-driven thresholding for knowledge-based landslide detection by object-based image analysis. IEEE Trans. Geosci. Remote Sens. 49, 4928–4943.

Martha, T.R., van Westen, C.J., Kerle, N., Jetten, V., Kumar, K.V., 2013. Landslide hazard and risk assessment using semi-automatically created landslide inventories. Geomorphology 184, 139–150.

Michel, J., Youssefi, D., Grizonnet, M., 2014. Stable mean-shift algorithm and its application to the segmentation of arbitrarily large remote sensing images. IEEE Trans. Geosci. Remote Sens. 53, 952–964.

Mondini, A.C., Chang, K.-T., Chiang, S.-H., Schlögel, R., Notarnicola, C., Saito, H., 2017. Automatic mapping of event landslides at basin scale in Taiwan using a Montecarlo approach and synthetic land cover fingerprints. Int. J. Appl. earth Obs. Geoinf. 63, 112–121.

Mondini, A.C., Guzzetti, F., Reichenbach, P., Rossi, M., Cardinali, M., Ardizzone, F., 2011. Semi-automatic recognition and mapping of rainfall induced shallow landslides using optical satellite images. Remote Sens. Environ. 115, 1743–1757.

Mondini, A.C., Marchesini, I., Rossi, M., Chang, K.-T., Pasquariello, G., Guzzetti, F., 2013. Bayesian framework for mapping and classifying shallow landslides exploiting remote sensing and topographic data. Geomorphology 201, 135–147.

Moosavi, V., Talebi, A., Shirmohammadi, B., 2014. Producing a landslide inventory map using pixel-based and object-oriented approaches optimized by Taguchi method. Geomorphology 204, 646–656.

Nichol, J., Wong, M.S., 2005. Satellite remote sensing for detailed landslide inventories using change detection and image fusion. Int. J. Remote Sens. 26, 1913–1926.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Perry, M.T., 2017. Rasterstats.

Planet Team, 2017. Planet Application Program Interface: In Space for Life on Earth. San Francisco, CA. [WWW Document]. URL https://api.planet.com

Plank, S., Hölbling, D., Eisank, C., Friedl, B., Martinis, S., Twele, A., 2015. Comparing object-based landslide detection methods based on polarimetric SAR and optical satellite imagery—A case study in Taiwan, in: 7th International Workshop on Science and Applications of SAR Polarimetry and Polarimetric Interferometry, POLinSAR. Frascati, Italy, p. 5.

Rau, J.-Y., Jhan, J.-P., Rau, R.-J., 2014. Semiautomatic object-oriented landslide recognition scheme from multisensor optical imagery and DEM. IEEE Trans. Geosci. Remote Sens. 52, 1336–1349.

Rey, S.J., Anselin, L., 2010. PySAL: A Python library of spatial analytical methods, in: Handbook of Applied Spatial Analysis. Springer, pp. 175–193.

Roback, K., Clark, M.K., West, A.J., Zekkos, D., Li, G., Gallen, S.F., Chamlagain, D., Godt, J.W., 2018. The size, distribution, and mobility of landslides caused by the 2015 Mw7. 8 Gorkha earthquake, Nepal. Geomorphology 301, 121–138.

Scaioni, M., Longoni, L., Melillo, V., Papini, M., 2014. Remote Sensing for Landslide Investigations: An Overview of Recent Achievements and Perspectives. Remote Sens. https://doi.org/10.3390/rs6109600

Soeters, R., Van Westen, C.J., 1996. Slope instability recognition, analysis and zonation: Landslides, investigation and mitigation, Special Report. Transp. Res. Board, Natl. Res. Counc. 247, 129–177.

Stumpf, A., Kerle, N., 2011. Object-oriented mapping of landslides using Random Forests. Remote Sens. Environ. 115, 2564–2577.

Sun, W., Tian, Y., Mu, X., Zhai, J., Gao, P., Zhao, G., 2017. Loess landslide inventory map based on GF-1 satellite imagery. Remote Sens. 9, 314.

Tanyaş, H., van Westen, C.J., Allstadt, K.E., Jibson, R.W., 2019. Factors controlling landslide frequency–area distributions. Earth Surf. Process. landforms 44, 900–917.

Tavakkoli Piralilou, S., Shahabi, H., Jarihani, B., Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S.R., Aryal, J., 2019. Landslide detection using multi-scale image segmentation and different machine learning models in the higher Himalayas. Remote Sens. 11, 2575.

Uddin, K., Shrestha, H.L., Murthy, M.S.R., Bajracharya, B., Shrestha, B., Gilani, H., Pradhan, S., Dangol, B., 2015. Development of 2010 national land cover database for the Nepal. J. Environ. Manage. 148, 82–90.

Vamsee, A.M., Kamala, P., Martha, T.R., Kumar, K.V., Amminedu, E., 2018. A tool assessing optimal multi-scale image segmentation. J. Indian Soc. Remote Sens. 46, 31–41.

Van Den Eeckhaut, M., Kerle, N., Poesen, J., Hervás, J., 2012. Object-oriented identification of forested landslides with derivatives of single pulse LiDAR data. Geomorphology 173, 30–42.

van Westen, C.J., Van Asch, T.W.J., Soeters, R., 2006. Landslide hazard and risk zonation—why is it still so difficult? Bull. Eng. Geol. Environ. 65, 167–184.

Xu, C., Tian, Y., Zhou, B., Ran, H., Lyu, G., 2017. Landslide damage along Araniko highway and Pasang Lhamu highway and regional assessment of landslide hazard related to the Gorkha, Nepal earthquake of 25 April 2015. Geoenvironmental Disasters 4, 14.

Yu, B., Chen, F., Muhammad, S., 2018. Analysis of satellite-derived landslide at Central Nepal from 2011 to 2016. Environ. Earth Sci. 77, 331. https://doi.org/10.1007/s12665-018-7516-1