Model-Centric vs Data-Centric Deep Learning Approaches for Landslide Detection

Hejar Shahabi[1], Omid Ghorbanzadeh[2]

[1]Centre Eau Terre Environnement, Institut National de la Recherche Scientifique (INRS), Quebec City, QC G1K 9A9, Quebec, Canada
[2]Institute of Advanced Research in Artificial Intelligence (IARAI), Vienna, Austria

Abstract
The implementation of deep learning (DL) models has significantly improved the accuracy and automation of remote sensing (RS) image classification tasks, such as landslide detection. The reason is that DL models have independent feature learning and strong computing capabilities and have attracted continuous attention in modifications and enhancements through numerous model-centric efforts. In practice, however, the impact of the quality of training samples on classification performance is usually ignored. This study uses a model-centric approach in which a U-Net network is regarded as a baseline model. A ResNet-34 model is used to optimize the baseline model, and the optimized model is further enhanced by adding an attention mechanism. However, in the data-centric approach, the baseline model is only trained based on the enhanced training samples. Our data-centric approach increased the F1-score by over 13 percentage points, which is the same increase as the most sophisticated and complex model-centric approach.

Keywords
Attention mechanism, deep learning, object-based image analysis (OBIA), landslide extraction

1. Introduction

As remote sensing (RS) imagery has become the basis of data collection across various fields, such as agriculture, environment, and disaster risk management, critical information can be extracted from such multi-temporal and multi-resolution images through image classification, object detection, and time series analysis [1]. However, one of the most critical aspects of such data processing is selecting the appropriate method to use. For many years, the remote sensing community has used artificial neural networks (ANN) such as Multi-Perceptron Layer (MLP) as a conventional method of image classification [2]. Until recently, however, conventional machine learning models like support vector machines and ensemble classifiers such as random forests nearly replaced ANN models for tasks like image classification and change detection because they can handle data with high dimensions and provide acceptable performance even with limited labeled data [2]. Recent developments in computer vision and graphics processing units (GPUs) have led to a rise in the popularity of deep neural networks within the RS community for different tasks, as they have generated robust results in image classification and image segmentation [3]. Several image classification models have been developed, such as AlexNet, VGG net, GoogleNet, ResNet, and DenseNet [4].

In RS tasks such as image classification, however, the goal usually is to label every pixel within an image, and DL semantic segmentation techniques such as Fully Convolutional Networks (FCN) are used to achieve that goal [5, 6]. The U-Net [7] algorithm, which utilizes encoder-decoder architectures to improve FCN, is widely used by the RS community for image segmentation and object detection, although it was initially designed for medical image segmentation. Other region-based models are used to detect and segment objects, including Faster R-CNN and Mask R-CNN, successfully applied to landslide inventory mapping [8]. These segmentation models have also become more sophisticated and advanced by incorporating concepts like attention mechanisms and or incorporating backbone models and weights such as the residual networks (ResNets) [9]. The attention mechanisms focus on certain features or regions while overlooking others, such the way of working of human vision.

Some current studies incorporated other approaches with DL models to increase the transferability [10] and also achieve higher accuracy in RS classification tasks and in landslide detection. Ghorbanzadeh*et al.* [11] did a model-centric strategy by synchronizing the heat map resulting in a ResU-Net network by knowledge-based object-based image analysis (OBIA) for landslide detection. Their experiences have done based on the satellite Sentinel-2 imagery. Their result evaluation indicated that integrating OBIA with U-net resulted in an F1 score value nearly 8% higher than the baseline ResU-Net model for the landslide detection task. In another study, Dong*et al.* [12], improved the U-Net's ability for landslide detection by adding a multi-scale feature-fusion module,

CDCEO 2022: 2nd Workshop on Complex Data Challenges in Earth Observation, July 25, 2022, Vienna, Austria
hejar.shahabi@inrs.ca (H. Shahabi);
omid.ghorbanzadeh@iarai.ac.at (O. Ghorbanzadeh)
0000-0002-3275-8436 (H. Shahabi); 0000-0002-9664-8770 (O. Ghorbanzadeh)

a residual attention network, and a data-dependent up sampling method. Their enhanced network, named L-Unet could increase the F1 score by more than 3%. While Ghorbanzadeh*et al.* [13] followed a data-centric strategy by preparing six training data sets based on optical data and different topographic information to evaluate the performance of a deep-learning convolution neural network (CNN) for landslide detection, a study area in Nepal. Their most remarkable improvement increased the mIOU by more than 17 percentage points. Yang*et al.* [14] have done a training samples enhancement and developed a background-enhancement technique that could support distinguishing landslides and similar background features for training the Mask R-CNN model. The F1 score was significantly higher (22.38%) than the one obtained using only satellite images as input data in their experiment.

As a result of a literature review on applying DL to RS applications, including landslide detection, the focus is mainly on model-centric approaches such as adapting, comparing architectures, and developing advanced models. At the same time, input data plays a less important role in model performance. In the model-centric approach, the input data is the same, while the main effort is focused on code and developing experimental research to improve the model performance. This involves selecting the best model architecture and training process from various possibilities [15]. On the other hand, in the data-centric approach, the goal is to systematically alter, synthesize, and improve datasets to increase the model's accuracy with a fixed architecture [16]. To our best knowledge, no study has comprehensively compared or discussed model-centric vs. data-centric approaches in landslide detection purposes. Since most DL studies for RS tasks are model-centric, in this experimental case study, we aim to compare and evaluate the performance of model-centric and data-centric approaches in landslide detection using Sentinel-2 imagery, ALOS elevation data U-net segmentation method.

2. Study area and data set

A magnitude 6.6 earthquake struck Eastern Iburi, Hokkaido, Japan, on September 6, 2018. There were extensive damages caused by this incident, including power cuts, damage to transmission and distribution networks, and damage to the Tomato-Atsuma Power Station, which supplies electricity to Hokkaido Island. There were 41 fatalities in all, 36 of which were caused by landslides triggered by the earthquake. Typhoon Jebi brought torrential rains to the region just a day before earthquake, which made hills unstable and prone to landslides. This caused nearly 5600 landslides in the area. The result was a significant number of shallow landslides as well as a

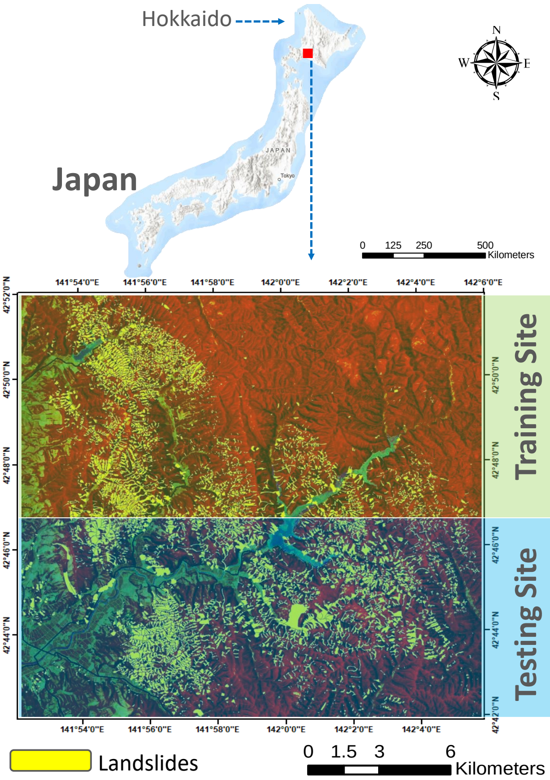

Figure 1: The location of the study area.

few planar and spoon-type deep landslides. A landslide inventory map was generated and updated by the Geographical Survey Institute (GSI) of Japan using aerial orthophotos, very high-resolution aerial images, and a 10m resolution digital elevation model (DEM) [9].

As part of this study, an inventory map (around 4950 landslides in an area of 43 km2) in shape file format provided by GIS in ESRI was acquired and used as ground truth for further analysis. Landslides were detected using Sentinel-2 multispectral imagery in this case. We applied atmospheric corrections to images using Sen2Cor [17], a SNAP plugin. We generated slope layers using the 12-m ALOS DEM, which is an important data set for mapping landslides [3]. Sentinel 2 images range in spatial resolution from 10 to 60 meters, so we selected only bands 2-4 and 8, which have a 10-meter resolution and excluded other bands. The generated slope layer was re sampled to 10-m and stacked with the Sentinel bands. Figure 1 shows the study area and landslide inventory.

3. Methodology

3.1. Model-centric approach

3.1.1. U-Net

As stated in the introduction section, U-Net was initially introduced for the segmentation of biomedical images Ronnenberger*et al.* [7]. Because of its robust performance, it has been used for a wide range of image segmentation problems, including remote sensing image classification and object detection [9]. In U-Net Architecture, encoder and decoder are the two main components. By applying convolution, activation functions, and pooling operations, the encoder learns how to abstractly represent the input image. Pooling reduces computational cost in this part, but spatial data is lost [7]. Through operations such as transpose convolution model, the decoder part attempts to restore the original size of the abstracted representation. Concatenating the output of transpose convolution with the skip connection feature map in the encoder part is the skip-connection feature map at the same level as the output of transpose convolution. While in the encoder part the number of feature channels is doubled at each down-sampling, in the decoder part it is diminished by half until in the final layer a convolution (1x1) is used (in this case with sigmoid activation) to map the channel patterns to a given number of classes. The architecture used by Ghorbanzadeh*et al.* [9] will be implemented for this study.

3.1.2. Residual network (ResNet)

ResNet serves as a backbone for many computer vision tasks, including remote sensing image classification and segmentation [9]. In 2015, it won the ImageNet challenge with extremely deep neural networks (more than 150 layers). ResNet's success is mainly due to its novel architecture that introduced skip connections for the first time, which add the output from the previous layer to the layer ahead. This alternative shortcut path allows gradient to flow through and prevents the problem of vanishing gradient. The baseline model of the U-Net design is used in this research, but with ResNet-34 acting as a backbone.

3.1.3. Attention mechanism

Bahdanau*et al.* [18] introduced the attention mechanism to enhance the performance of the encoder-decoder models for machine translation system. Later its variants were used in other application including the RS applications. Through a weighted combination of encoded input data, the decoder has access to the most valuable parts of the input sequence, thus, the most relevant parts will be associated with the highest weights. A detailed description of the attention mechanism in CNN-based models for the RS applications is provided by [1]. In this case, an attention mechanism is added to the U-Net model with the ResNet 34 backbone for landslide detection.

3.2. Data-centric approach

In this approach, the U-Net architecture applied by [9], which was introduced as our baseline model is used without any model enhancement. However, Data enhancement will be done by synthesizing and augmenting training samples. Therefore, along with the 10-m Sentinel 2 images and slope layer, normalized difference vegetation index (NDVI) will be generated and fed into the model as well. This index is helpful for discriminating some of landslides that removed the surface vegetation from the background objects [14, 13]. Moreover, using the OBIA concept and multi-resolution segmentation (MRS), which is a bottom-up segmentation technique based on the pairwise region-merging approach the size of the generated object is controlled by the scale factor [19]. In order to avoid errors such as over-segmentation and under-segmentation an index called as object fitness index (OFI) introduced by [20] is applied to guaranty the quality of objects. The mean values of objects for each image band, including NDVI and slope, will be calculated, and then exported in a raster format to be stacked with another dataset as input to the U-Net model.

4. Experimental results

In this study, a tile size of 128 × 128 without any overlap was used as an input to all applied models. The accuracy of the model was also validated by selecting 30% of training data sets at random. As our task is binary classification, we applied Sigmoid as the activation function in the last layers and rectified linear activation function (ReLU) in the earlier layers. All models were set at 100 epochs, but a function was defined to save the model on the epoch number to ensure minimum losses. Our DL models were all implemented in Python using TensorFlow API and Keras library, for segmentation part we used eCognition software. As previously noted, data augmentation was only applied to a data-centric approach.

Each model was evaluated based on standard accuracy assessment metrics of the precision, recall and F1-score. A loss and F1-score of 0.14 and 0.88 were achieved during the training of the conventional U-Net model while, these values were 0.26 and 0.68 for validation data. A total of almost 4 million parameters were trained in this model. For test area, the trained model was used to detect landslides; the accuracy assessment results showed precision, recall, and F1-score values of 0.76, 0.48, and

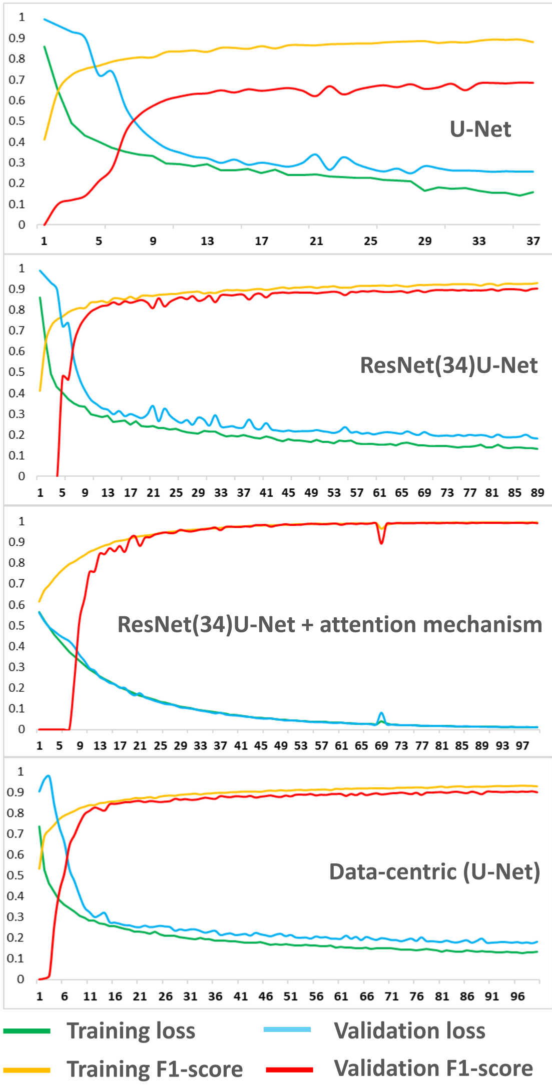

Figure 2: Training metrics graphs for model-centric approaches and data-centric approach.

0.59, respectively. The same U-Net was trained in the following, but with ResNet 34 as a backbone model. The total number of parameters for this model was 23 million, but only 1.8 million parameters were trained, and for the rest pre-trained weights were applied. Training Loss and F1-score values for U-Net with ResNet 34 backbone were 0.13 and 0.93, respectively, and for the validation data set, the scores were 0.18 and 0.90. Although the model's accuracy on training data and validation is quite close, its performance in the test area did not provide much higher accuracy. Precision, recall, and F1-score for U-Net with ResNet backbone were 0.67, 0.72, and 0.70, respectively. Finally, the most complex version of U-Net that includes both ResNet backbone and attention mechanism with 10.5 million parameters was trained. For training, loss and F1-score values were 0.05 and 0.95, and for validation were 0.08 and 91, respectively. However, like U-Net with ResNet backbone model, the performance of the model in the test area even with adding attention mechanism to the model was not significant, and values 0.88, 0.62, 0.72 were achieved for precision, recall, and F1-score, accordingly. It provided the best precision score among other models.

For the data-centric approach, only the base line U-Net model with the same architecture was used as for the first scenario in the model-centric approach. However, the input data has been modified. The OBIA features were stacked with other images, and then data augmentation techniques such as flipping (horizontally and vertically) and rotating (90, 180, 270 degrees) were performed. This resulted in 10395 image patches being fed to the network instead of 2079 image patches. In addition, data augmentation was not applied to validation data. The U-Net model achieved values of 0.13 for loss and 0.93 for F1-score during training. With validation data, however, the values were 0.18 and 0.90. The trained model was used to predict test data, resulting in precision, recall, and F1-score of 0.71, 0.73, and 0.72. The figures 2 and 3 are respectively depicting the training curve of a model and the landslide prediction map for a model-centric and a data-centric approach to landslide prediction.

5. Discussions

In the model-centric approach, using the same labeled data while varying architecture and parameters resulted in varying accuracies in training and validation phases, general U-Net models performed relatively poorly (underfitting error) in the training process compared to others. U-Net based on ResNet Backbone, however, performs better during training, while the best performance is achieved with U-Net based on ResNet and attention mechanism. Loss values for both training and validation data indicate that by increasing models' parameters higher accuracy during training can be achieved. But getting such high accuracy throughout training can be a sign of overfitting. Therefore, prediction results by such a model are evaluated with inventory data. Consequently, excepting the general U-Net provided the lowest accuracy with a recall of 0.48 and an F1-score of 0.60. which means it U-Net was able to detect only 48% of landslides. Furthermore, U-Net with ResNet backbone provided much better performance compared to the general U-Net model, with a recall of 0.72 and F1-score of 072. Finally in model-centric, the best performance according

Table 1
Quantitative evaluation of models.

Model	Tr-loss	Va-loss	Precision	Recall	F1-score
U-Net	0.14	0.26	0.76	0.48	0.59
ResNet(34)U-Net	0.13	0.18	0.67	0.72	0.7
ResNet(34)U-Net + attention mechanism	0.05	0.08	0.88	0.62	0.72
Data-centric (U-Net)	0.13	0.17	0.71	0.73	0.72

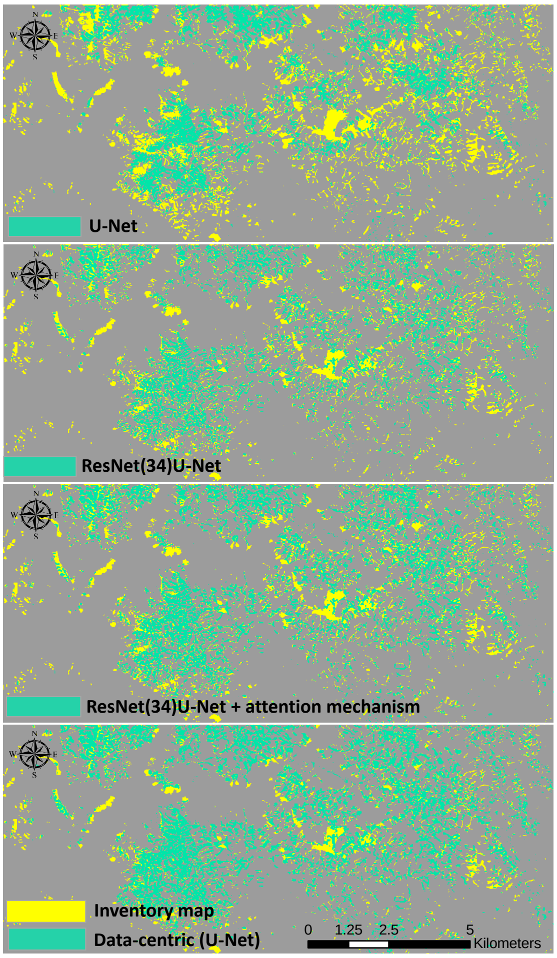

Figure 3: Landslide detection results based on the model-centric approaches and data-centric approach.

to F1-score achieved by the U-Net model based on ResNet backbone and attention mechanism, the score achieved was 0.73 although the recall value was 0.62 it provided the highest precision of 0.88. In Data-Centric, the conventional U-Net with the same architecture was used as the fixed model, while data went through argumentation. Training the model using synthesized and augmented data provided a great performance on both training data and validation. And evaluating predicted landslide with inventory map indicated the F1-score of 0.72 while great consistency between other metrics such as precision with 0.71 and recall of 0.73. This experiment clearly shows that by generating/synthesizing data and argumentation available data higher accuracy can be achieved even with simple model architecture. For example, the performance and accuracy of the data-centric U-Net model were quite similar to an advanced U-Net model with a ResNet backbone and attention mechanism, and in terms of F1-score, the difference in both models' performance was 1%.

6. Conclusions

In this study, the goal was to compare two approaches namely model-centric and data-centric in DL for remote sensing application of landslide detection. According to our accuracy assessment result we conclude that the accuracy of landslide detection can be improved by optimizing network structures or training data set to a certain extent. We showed that the process of enhancing sample sets in the data-centric and may adding additional information is an optimization on the data level, which is applicable any DL models and the common ones like the U-Net model. A direction worth pursuing is how we can enhance the landslide detection accuracy of the DL results by modifying the training samples before or in the feature learning step. We developed a data-centric approach that includes different measurements, and we compared the results with those obtained from complex network structures to represent the potential capabilities of data optimization. The application of popular segmentation models like FCN, SegNet, Deeplab, and ASPP, also the impact of the data-centric approach on the model transferability to new areas is the focus of our next work.

Acknowledgments

This research was funded by the Institute of Advanced Research in Artificial Intelligence (IARAI) GmbH, Vienna, Austria.

References

[1] S. Ghaffarian, J. Valente, M. Van Der Voort, B. Tekinerdogan, Effect of attention mechanism in deep learning-based remote sensing image processing: A systematic literature review, Remote Sensing 13 (2021) 2965.

[2] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, B. A. Johnson, Deep learning in remote sensing applications: A meta-analysis and review, ISPRS journal of photogrammetry and remote sensing 152 (2019) 166–177.

[3] O. Ghorbanzadeh, Y. Xu, P. Ghamis, M. Kopp, D. Kreil, Landslide4sense: Reference benchmark data and deep learning models for landslide detection, arXiv preprint arXiv:2206.00515 (2022).

[4] Z. Ma, G. Mei, Deep learning for geological hazards analysis: Data, models, applications, and opportunities, Earth-Science Reviews 223 (2021) 103858.

[5] E. Maggiori, Y. Tarabalka, G. Charpiat, P. Alliez, Convolutional neural networks for large-scale remote-sensing image classification, IEEE Transactions on geoscience and remote sensing 55 (2016) 645–657.

[6] Y. Xu, P. Ghamisi, Region-growing fully convolutional networks for hyperspectral image classification with point-level supervision (2021).

[7] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer, 2015, pp. 234–241.

[8] S. L. Ullo, A. Mohan, A. Sebastianelli, S. E. Ahamed, B. Kumar, R. Dwivedi, G. R. Sinha, A new mask r-cnn-based method for improved landslide detection, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 14 (2021) 3799–3810.

[9] O. Ghorbanzadeh, A. Crivellari, P. Ghamisi, H. Shahabi, T. Blaschke, A comprehensive transferability evaluation of u-net and resu-net for landslide detection from sentinel-2 data (case study areas from taiwan, china, and japan), Scientific Reports 11 (2021) 1–20.

[10] M. Zhang, H. Singh, L. Chok, R. Chunara, Segmenting across places: The need for fair transfer learning with satellite imagery, arXiv preprint arXiv:2204.04358 (2022).

[11] O. Ghorbanzadeh, H. Shahabi, A. Crivellari, S. Homayouni, T. Blaschke, P. Ghamisi, Landslide detection using deep learning and object-based image analysis, Landslides (2022) 1–11.

[12] Z. Dong, S. An, J. Zhang, J. Yu, J. Li, D. Xu, L-unet: A landslide extraction model using multi-scale feature fusion and attention mechanism, Remote Sensing 14 (2022). URL: https://www.mdpi.com/2072-4292/14/11/2552. doi:10.3390/rs14112552.

[13] O. Ghorbanzadeh, S. R. Meena, H. S. S. Abadi, S. T. Piralilou, L. Zhiyong, T. Blaschke, Landslide mapping using two main deep-learning convolution neural network streams combined by the dempster–shafer model, IEEE Journal of selected topics in applied earth observations and remote sensing 14 (2020) 452–463.

[14] R. Yang, F. Zhang, J. Xia, C. Wu, Landslide extraction using mask r-cnn with background-enhancement method, Remote Sensing 14 (2022) 2206.

[15] L. J. Miranda, Towards data-centric machine learning: a short review, ljvmiranda921. github. io (2021).

[16] I. Pan, L. R. Mason, O. K. Matar, Data-centric engineering: integrating simulation, machine learning and statistics. challenges and opportunities, Chemical Engineering Science 249 (2022) 117271.

[17] M. Main-Knorn, B. Pflug, J. Louis, V. Debaecker, U. Müller-Wilm, F. Gascon, Sen2cor for sentinel-2, in: Image and Signal Processing for Remote Sensing XXIII, volume 10427, International Society for Optics and Photonics, 2017, p. 1042704.

[18] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, arXiv preprint arXiv:1409.0473 (2014).

[19] T. Blaschke, Object based image analysis for remote sensing, ISPRS journal of photogrammetry and remote sensing 65 (2010) 2–16.

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