KNOWLEDGE-DRIVEN LANDSLIDE SUSCEPTIBILITY ZONATION USING DEEP LEARNING

Faraz S. Tehrani, Carmen Martinez Barbosa, Amine Aboufirass

ABSTRACT

Landslide susceptibility mapping using statistics methods require access to landslide inventories. Without a reliable landslide inventory the results of landslide susceptibility zonation can be misleading which can lead to over-estimation or under-estimation of landslide susceptibility in the region of interest. However, access to such data is often hampered by lack of proper data collection and management. In this study, we tried to tackle this issue by training Deep Learning algorithms on a reliable landslide susceptibility dataset and eliminate the need for landslide inventories. The results of this study show that the trained DL model could successfully identify zones with high landslide likelihood, which indicate the potential of the trained model as a practical tool for identifying landslide prone areas based on basic topographic features as well as soil and land cover data.

Keywords: Landslide, Deep Learning, ELSUS, DEM, Soil, Landcover

INTRODUCTION

Landslide susceptibility is the likelihood of the landslide occurrence, which is controlled by local terrain conditions. Landslide susceptibility zonation or mapping refers to the segmentation of a region of interest into homogeneous zones or domains with similar landslide susceptibilities. This is typically done using expert judgement, deterministic methods, statistical methods, or a combination of these. For statistical landslide susceptibility mapping, which is the dominant approach in landslide hazard zonation, location (and preferably date) of historical landslides are required. However, this data is typically not readily available for many regions in the world. Hence, the first activity in landslide susceptibility mapping at regional scale is concerned with acquiring the past landslide data and building a landslide inventory, which is a time-consuming process. Therefore, when the record of past landslides is not available, susceptibility mapping should be performed based on the domain knowledge of the landslide experts and or the governing physics of the landslide occurrence. While the former requires the availability of experts, the latter needs detailed information on terrain topography, soil, lithology, hydrology and land cover and land use.

Given the challenges associated with any of the aforementioned methods, in this study we attempted to take advantage of Deep Learning (DL) as a subset of Artificial Intelligence

(AI) and "teach" AI how to classify areas based on their susceptibility to landslides. In order to teach, similar to any educational framework, we need reliable knowledge resources. In this case, we needed credible landslide susceptibility maps based on which we could train the DL model. We chose to use Pan-European landslide susceptibility map: ELSUS Version2 (Wilde et al. 2018), which was developed by landslide experts across Europe based on extensive records of landslides (149,117 landslide events) and the expert knowledge and judgement of its developers.

The DL model is then trained over ELSUS map supplemented by globally available controlling factors. The trained model is aimed to be used elsewhere in the world for initial and quick assessment of landslide susceptibility.

DATA

In order to train the DL model, we need the landslide susceptibility knowledge, which is provided by ELSUS map. We also need globally available datasets on landslide controlling factors including terrain features, shallow lithology and landcover. In this study we restricted ourselves to these limited number of controlling factors (as did ELSUS developers), but for future improvements, we will consider other relevant factors.

ELSUS v2

The ELSUS v2 map provides landslide susceptibility zonation for individual climate-physiographic zones across Europe (Wilde et al. 2018). ELSUS v2 covers a large area of Europe at a high spatial resolution (200×200m). The updated map was prepared using the semi-quantitative method, combining landslide frequency ratios information with a spatial multi-criteria evaluation model of three thematic predictors: slope angle, shallow subsurface lithology and landcover. Figure 1 shows the ELSUS map. ELSUS also include climate data that will have limited use in training the DL model as explained later.

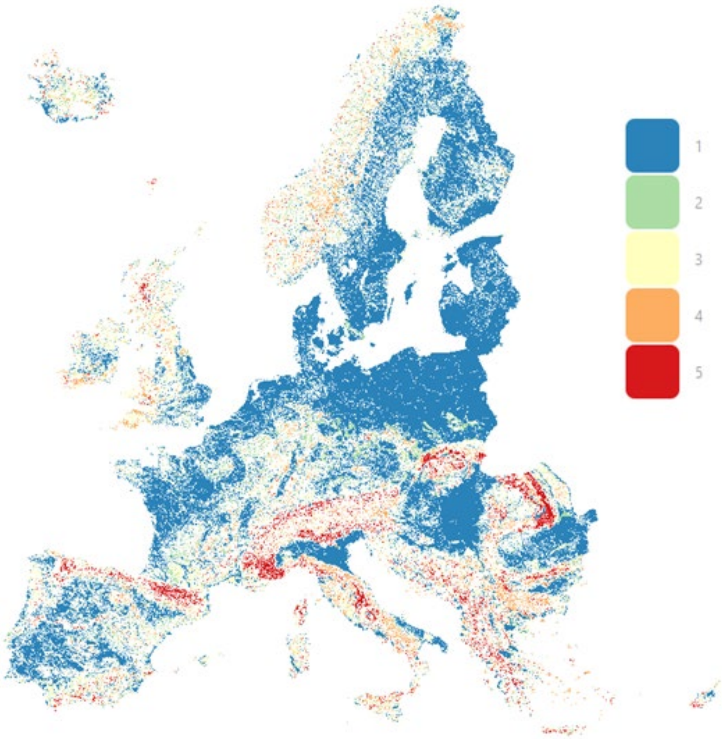

Figure 1. Landslide susceptibility across Europe.

Five landslide classes are defined in ELSUS v2: very low (1), low (2), medium (3), high (4), very high (5). These five classes will serve as the target values (later introduced as *y*) in DL classification. In total ELSUS map consists of 19453 × 20151 cells ~ 391 million cells out of which we selected ~48 million data points to create the dataset for training and testing the DL model.

Digital Elevation Model (DEM)

Digital elevation models (DEMs) are considered as one of the main datasets to extract controlling factors for landslide hazard assessments (van Westen et al. 2008). These three-dimensional representations of the terrain are useful for extracting key topographical and geomorphological parameters including elevation and slope of the ground surface. In this study, we used Multi-Error-Removed Improved-Terrain (MERIT) DEM (Yamazaki et al. 2017) that was developed by removing multiple error components (absolute bias, stripe noise, speckle noise, and tree height bias) from the existing spaceborne DEMs (SRTM3 v2.1 and AW3D-30m v1). It represents the terrain elevations at a 3 arc-sec resolution (~90m at the equator) and covers land areas between 90N-60S. The MERIT map was clipped, over ELSUS map and resampled to the resolution of 200 × 200 m and aligned to be compatible with ELSUS map using open source GIS libraries in python. The terrain features that we extracted from

DEM are Topographic Position Index (TPI) [defined as the difference of every elevation pixel with the average elevation of surrounding pixels], slope, elevation and aspect extracted from the DEM using GDAL command line utilities.

Shallow Lithology

Shallow lithology plays an import role in landslide hazard analysis since the constituent material of slopes can highly affect the hydro-geomechanical response of slopes to landslide triggering factors such as rainfall and earthquake. Therefore, estimating the soil composition of hillslopes can potentially enhance the predictability of landslides. Soil composition was retrieved as raster data from the SoilGrids datasets (Hengl et al. 2014) at 250 m resolution with a global coverage. SoilGrids provides global predictions for standard numeric soil properties such as soil texture fractions at seven standard depths (0, 5, 15, 30, 60, 100 and 200 cm). Soil content was reported as percentage of clay, sand, and silt. The estimated amount of each soil at each depth was normalized by dividing the soil estimated amount by the summation of amount of sand, clay and silt to ensure that all estimations are between 0 and 100 and that the fractions sum up to 100% (Hengl et al. 2014). Among the information available at SoilGrid, the estimated fraction of sand, clay and silt are used in this study. Similar to DEM, the soil map was also clipped, resampled and aligned to the resolution and extent of ELSUS map.

Landcover

The type of landcover in combination with slope material can define the susceptibility to landslides. For instance, the chance of landslide occurrence on barren slopes consisting sand and clay can be higher than vegetated slopes with the same materials, due to the role of vegetation in reinforcing the shallow depths and their contribution to hydrology of slopes (e.g. evapotranspiration, infiltration, rain interception, etc.). Land cover data was obtained from MODIS Land Cover yearly data MCD12Q1 V6. This product provides global land cover types at yearly intervals (2001-present) derived from six different classification schemes (Friedl et al., 2019) [we used three of these six classifications in this study]. The landcover classes provides qualitative (categorical) data such as type of the vegetation for vegetated surfaces. The resolution of land cover map is ~500 m. Similar to other maps, this data was also clipped, resampled and aligned to the extent and resolution of ELSUS map.

METHODOLOGY

The simple idea behind this work is that similar to experts who use their technical knowledge and experience to give meaningful scores to zonate areas based on their landslide susceptibility, machine can also gain knowledge and learn if an area with certain topographic features, soil and landcover can be landslide prone or not. Machine learns that knowledge from experts in this case. Experts developed the susceptibility map of Europe. Machine now ties the susceptibility classes to underlying controlling features and learn the relationships between them. To that end, we need to prepare a sample set that has these controlling factors as predictors or features (*X*), as well as landslide susceptibility classes as target classed (*y*). The goal here is to find a mapping function that maps feature vectors *X* to target vector *y*: $y = f(X)$. This mapping function f (.) is derived through Deep Learning in this study. The reason for choosing DL as the Machine Learning algorithm is the level of complexity in the dataset as well as the quantity of data (millions of data points).

DEEP LEARNING ALGORITHM

Background

DL models are a large set of algorithms that try to mimic the functioning of the human brain, whose structure is built from a large number of neurons that are interconnected. Hence their name, the so-called *Artificial Neural Networks* (ANN). The simplest component of an ANN is a perceptron. A perceptron takes the weighted sum of each feature vector X_i of the dataset and applies a mathematical nonlinear function (e.g. activation function) on each feature. The product of the weighted sum of input features with the activation function is the output of the perceptron. In Figure 2, the resemblance between a neuron and a perceptron is shown.

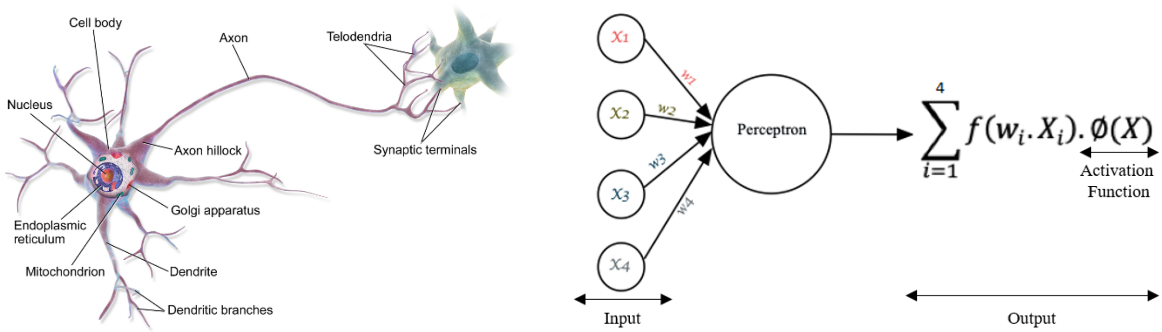

Figure 2: **Left:** Structure of a brain's neuron. **Right:** The simplest ANN known: a perceptron. The dendrites receive a signal which is processed inside a neuron and transferred via the axon to another neuron. In an ANN, the input vectors are received by the perceptron which decides which observations are more important than others by assigning weights to the features. The output can be transferred to another perceptron.

ANNs contain many perceptrons that can be organized in different architectures. ANNs receive as input information the features of the model X_i in batches with a defined size. These input vectors correspond to the *input layer* of the artificial network. This information is received by perceptrons that are located in the ANN's *hidden layers.* In the hidden layers, the ANN generate many combinations of features and learn the hidden relationships present in the data. ANNs can contain several hidden layers whose number depends on the complexity of the problem. As a consequence, the architecture of ANNs depends on the number of hidden layers. Deep networks may contain more than 3 hidden layers. Often, the amount of perceptrons decreases with the increment of hidden layers in the network; although wide ANNs can also be used[1]. Once the Artificial network have learnt the relationships between the data through the hidden layers, an output is created. **In the prediction of landslide susceptibility, the ANN generates 5 outputs, one per landslide susceptibility class**. In Figure 3, we show different ANN's architectures.

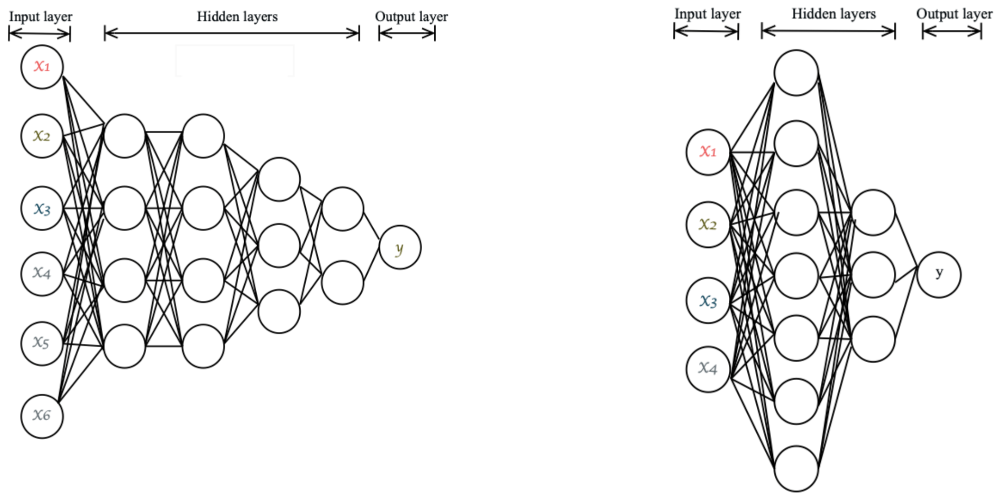

Figure 3: Different ANN architectures assuming a single output. **Left:** Deep neural network. **Right:** Wide and shallow ANN.

During the training an ANN, the input layer of the network receives the feature vectors X_i in batches, weights are generated in the hidden layers and the output of the model is created. These predictions are compared with the true target's observations and the weights are updated such that the loss function is minimized in a process called *back propagation.* The aforementioned steps are repeated until all the observations in the training set are used (e.g. after one epoch). Often, several epochs are needed to obtain a model with good predictive performance. In Table 1, we list the ANN's architectures and model hyperparameters used to predict landslide susceptibility. Here we list the architectures that were explored after

[1] Very wide and shallow networks should be used cautiously, since these models are good at memorization but not so good at generalizing.

optimizing the epochs and the batch size. We found that an increment in the number of epochs does not lead to an improvement in the error metrics. On the other hand, a decrement in the batch size leads to model overfitting (i.e. model does not provide good estimates on unseen data). An increment of the batch size leads to underfitting (i.e. bad predictions even on observations of the train set).

Table 1. ANN's architectures and hyperparameters used in predicting landslide susceptibility. Model 3 and model 4 have in the output layer 5 and 3 perceptrons respectively. See section ***Results and discussion*** for more details.

Model tag	Number of hidden layers	Number of perceptrons per hidden layer	Batch size	Epochs
1	4	28-20-10-6	5000	20
2	4	112-56-28-14	5000	20
3	3	28-28-14	5000	20
4	3	28-28-14	5000	20
5	4	56-28-14-7	5000	20
6	2	28-14	5000	20

Model metrics

To quantify the prediction performance of the DL models, we use the classification accuracy, the confusion matrix and the sensitivity or recall.

Classification Accuracy

Classification accuracy is defined by the ratio of the number of correct predictions to all predictions made. This is the most common evaluation metric; however, it might be quite misleading. It is only suitable when there are an equal number of observations in each class and that all predictions and prediction errors are equally important. During the preprocessing of the data we sampled the observations such that the number of landslide susceptibility classes were balanced.

Confusion matrix

This is one of the most common metrices to evaluate the performance of a model. It shows the quantity of true and false predictions, in terms of the observations of each class. For a binary classification problem, a confusion matrix is defined as shown in Figure 4.

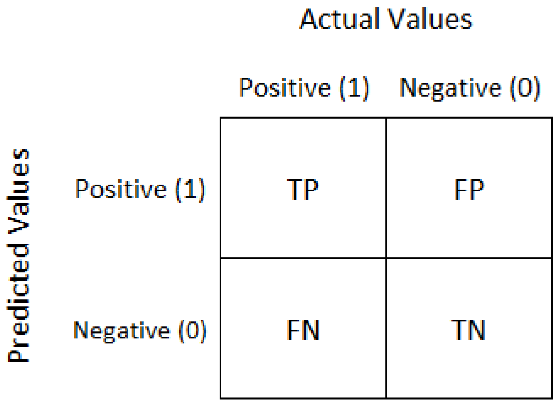

Figure 4: Schema of a confusion matrix in a binary classification problem

In the confusion matrix of binary classification (two classes), we define the following elements:

True Positives (TP): Predicted positive value is correct.

True negatives (TN): Predicted negative value is correct.

False positives (FP): Predicted positive value is wrong.

False negatives (FN): Predicted negative value is wrong.

The confusion matrix can also be normalized. In this case, it shows the *sensibility* or *Recall.*

Recall

Recall refers to the TP rate and it summarizes how well the positive class was predicted. Mathematically, recall is defined as:

$$R = \frac{TP}{TP + FN}$$

Since we are interested in having a DL model capable of making good classifications of how susceptible a region is to landslides (positive cases) or not, we evaluate the quality of the DL models based on this metric.

RESULTS AND DISCUSSION

DL models are known as being ‘black boxes’ due to the difficulty to infer the importance of the input vectors by just looking at the complex relationships generated in deep hidden layers. A combination of statistical techniques and less complex, but yet powerful Machine Learning (ML) models, may indicate the feature vectors that bring more information at predicting the target variable y. In this study we used a combination of ensembles of decision trees; i.e. Random Forest and different boosting algorithms, to obtain the importance of the variables used to predict the landslide susceptibility. We ran Random Forest and Boosting methods during 20 iterations to obtain an estimate (mean and standard deviation) of the importance of each feature vector. The results are shown in Figure 5. As expected, the slope is a very important factor at predicting the landslide susceptibility (average importance around of 0.2); however, features describing the landcover and land use, climate, DEM and lithology (i.e. fraction of sand, clay or silt) also play an important role in the determination of landslide and non-landslide regions. To further test this hypothesis, we train DL models by using three approaches:

Approach 1: Model containing information on climate from ELSUS v2 map + DEM + landcover and lithology. In this scenario we run different ANN architectures, according to those listed in Table 1.

Approach 2: Since the data of ELSUS v2 maps are not available worldwide, we removed from the model the climate features belonging to this data source. Additionally, in this approach we used the ANN architecture from approach 1 that provided the highest recall in all the landslide classes. Note that this model can be applied in all the regions of the world, given the availability of the data.

Approach 3: We trained a DL model where slope is the only feature vector. We aim to find whether there is a significant difference between this model and the one built in approach 2. The ANN architecture used in this approach was 1 perceptron as input layer and 5 perceptrons in the output layer.

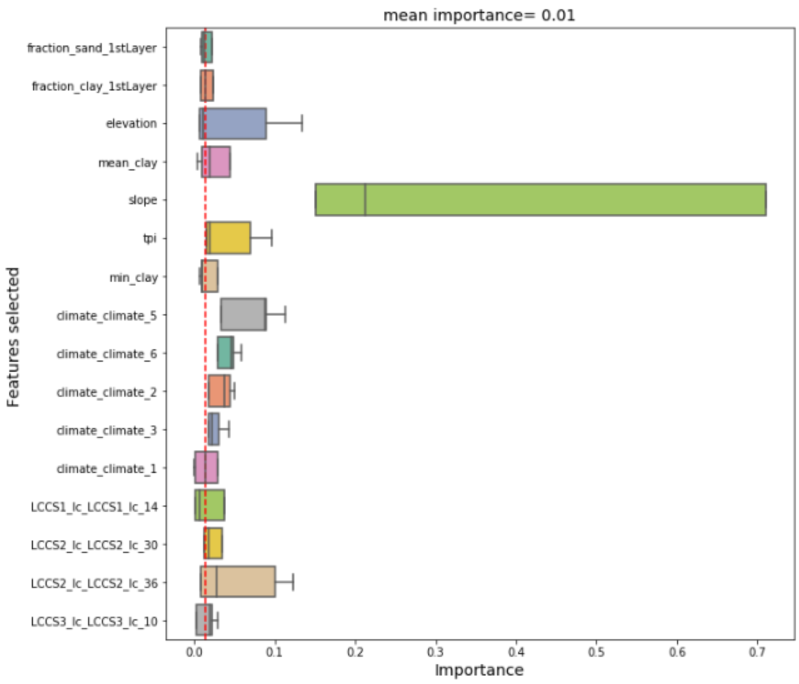

Figure 5: Importance of the input vectors. The feature importance is a score based on how useful an input vector X is at predicting the target variable.

In determining the feature importance, each ensemble of decision tree models provides a measure of feature importance. In Figure 5, we see the distribution (per feature) obtained after 20 iterations. The vertical dashed red line is the average importance of all the input vectors.

To train the DL models, we split the dataset into train, validation and test sets. The train and validation sets are used respectively to build the model and evaluate their performance during the training process. The test set is used as a hold-out dataset, to make a final evaluation of error metrics.

As observed in Figure 1, the European regions prone to landslides are quite less than those insusceptible to landslides. In statistical terms, this is referred to as an imbalanced problem. With the purpose of avoiding biases in the classification made by the DL model, we balanced the classes by over-sampling the minority classes. This process was carried out by picking observations of a landslide class at random with replacement. For this task we made use of the Python's package *imbalance-learn* (Lemaitre at el. 2017). The balance of the classes is carried out only in the train set as part of the data preprocessing. Thus, the number of observations in the train, validation and test sets are respectively: 30,000,000, 6,616,911 and 8,271,138.

Table 2: Model metrics of the DL models built in this study. The highlighted rows indicate the models that best classify the landslides classes.

Approach	**Model tag**	**Recall**					**accuracy**
		1	**2**	**3**	**4**	**5**	
1	1	0.8	0.37	0.41	0.55	0.89	0.60
	2	0.78	0.48	0.39	0.56	0.88	0.61
	3	0.77	0.54	0.4	0.55	0.88	0.62
	4	0.91	-	0.17	-	0.87	0.78
	5	0.81	0.44	0.37	0.59	0.86	0.61
	6	0.76	0.52	0.43	0.54	0.88	0.62
2	3	0.7	0.55	0.25	0.45	0.80	0.54
3	3	0.75	0.29	0.24	0.51	0.71	0.51

The results of the DL models' error metrics are listed in Table 2. In model 4 we show metrics of 3 classes (low, moderate and high), since we merged similar categories (low and very low, high and very high) here. Overall, ANN models are capable of classifying non-landslide and landslide-prone regions; with higher recall for class 5 (very high likelihood). The accuracy of the models is, on the other hand, relatively low due to the similarity between classes (i.e. 1,2 and 4,5). This can be confirmed by looking at the accuracy of model 3 and then comparing it with model 4.

The results listed in Table 2 also suggest that having the slope as unique feature vector does make good predictions of landslide susceptibility; however, the inclusion of other subsoil features is of high importance to improve the estimates. The climate features provided by ELSUS v2 data are also important in making a good landslide classification (see models 3, approach 1 and 2 for comparison); however, ELSUS weather data is not available worldwide, making the DL model developed with this feature, be only applicable in Europe. In future works, we will use global climate features such as mean annual rainfall and temperature to replace climate feature.

In Figure 6, we present the confusion matrix of the best models in the three approaches described (naming description: *A-m* means model *m* from approach *A*). The estimates of the confusion matrix, however, are calculations made over one test set. To get a complete estimate of the error metrics of a DL model, we carried out a 10-fold cross-validation on the whole dataset[2], to obtain the mean and standard deviation of the recall per class.

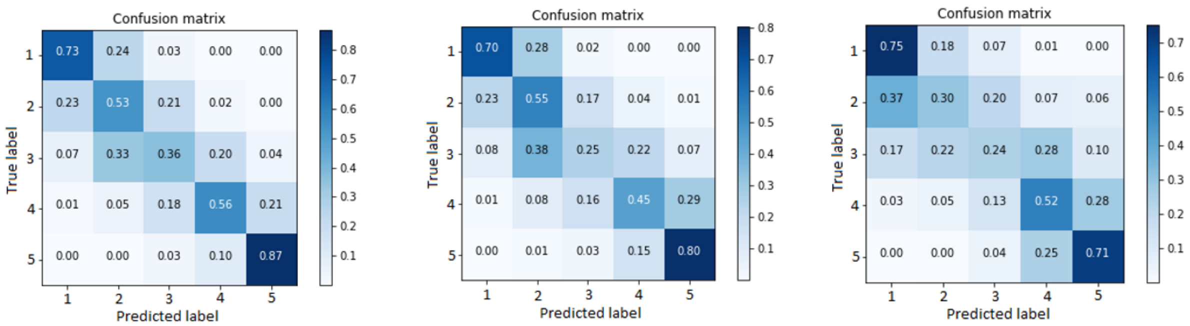

Figure 6: Confusion matrix of: **Left:** Best model with all features (1-3). **Middle:** Model with no Climate from ELSUS v2 data (2-3). **Right:** Model with only slope as input (3-3).

We performed the cross-validation on model 3 in approach 2 (2-3), since this model is aimed to be applicable in all regions in the world where the data of the input vectors are available. The error metrics of this model are:

Recall class 1: 0.70 ± 0.02

Recall class 2: 0.52 ± 0.07

[2] At every iteration of the cross-validation, observations of the dataset are randomly selected to generate the train, validation and test sets. The preprocessing, training and the model evaluation are also carried out such that after the 10 iterations, a mean and std of the error metrics are obtained.

Recall class 3: 0.26 ± 0.03

Recall class 4: 0.46 ± 0.005

Recall class 5: 0.79 ± 0.008

We found that these error metrics are statistically different than those obtained in the model that uses slope as unique Feature. To better visualize the predictions made by this DL model, in Figure 7 we show a comparison between the landslide classes observed in Europe with their corresponding predictions.

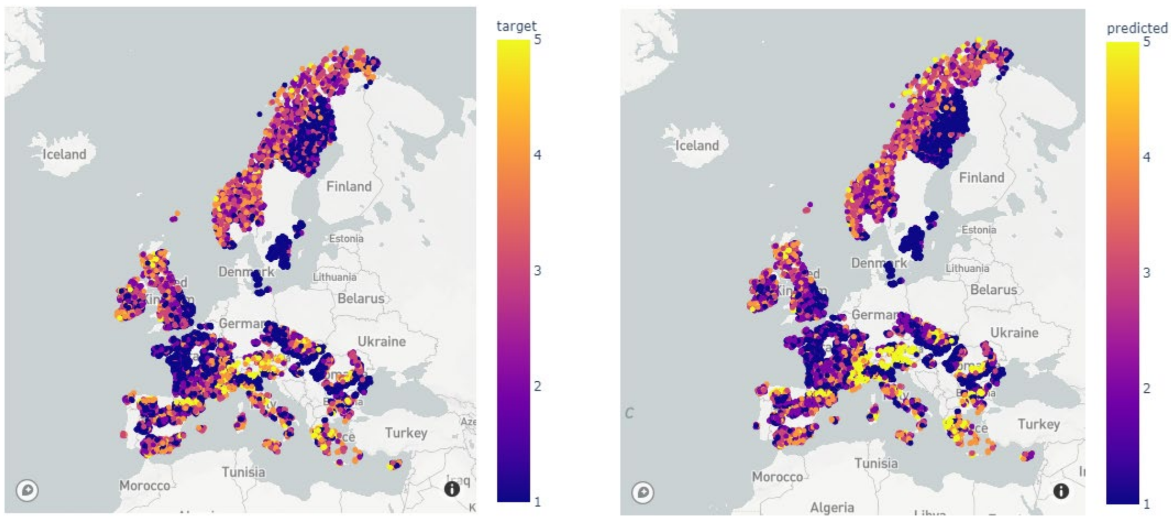

Figure 7: Observations of landslide susceptibility (**left**) and model predictions (**right**) over Europe.

Figure 7 confirms the fact that the model developed in this study is able to detect landslide prone regions. A combination of this model with a landslide forecasting system and run-out models, will be a product that can be broadly used in underdeveloped countries where landslide is a major problem in building and maintaining resilient infrastructures.

CONCLUSION

In this study, we developed a deep learning model and framework for landslide susceptibility mapping based on previously developed susceptibility maps. We used ELSUS version 2 augmented with slope, land cover and soil data to train the DL model. The trained model showed promising results which enable us to use the model for quick landslide susceptibility analysis across the world. A combination of this model with a landslide forecasting system and run-out models, will be a product that can be broadly used in underdeveloped countries where landslide is a major problem in building and maintaining resilient infrastructures.

REFERENCES

Hengl, T., de Jesus, J. M., Heuvelink, G. B., Gonzalez, M. R., Kilibarda, M., Blagotić, A., ... & Guevara, M. A. (2017). SoilGrids250m: Global gridded soil infor-mation based on machine learning. PLoS one, 12(2), e0169748.

van Westen, C. J., Castellanos, E., & Kuria-kose, S. L. (2008). Spatial data for landslide susceptibility, hazard, and vulnerability as-sessment: an overview. Engineering geology, 102(3-4), 112-131.

Wilde, M., Günther, A., Reichenbach, P., Malet, J. P., & Hervás, J. (2018). Pan-European landslide susceptibility mapping: ELSUS Version 2. Journal of Maps, 14(2), 97-104.

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., ... & Bates, P. D. (2017). A high-accuracy map of global terrain elevations. Geophysical Research Letters, 44(11), 5844-5853.