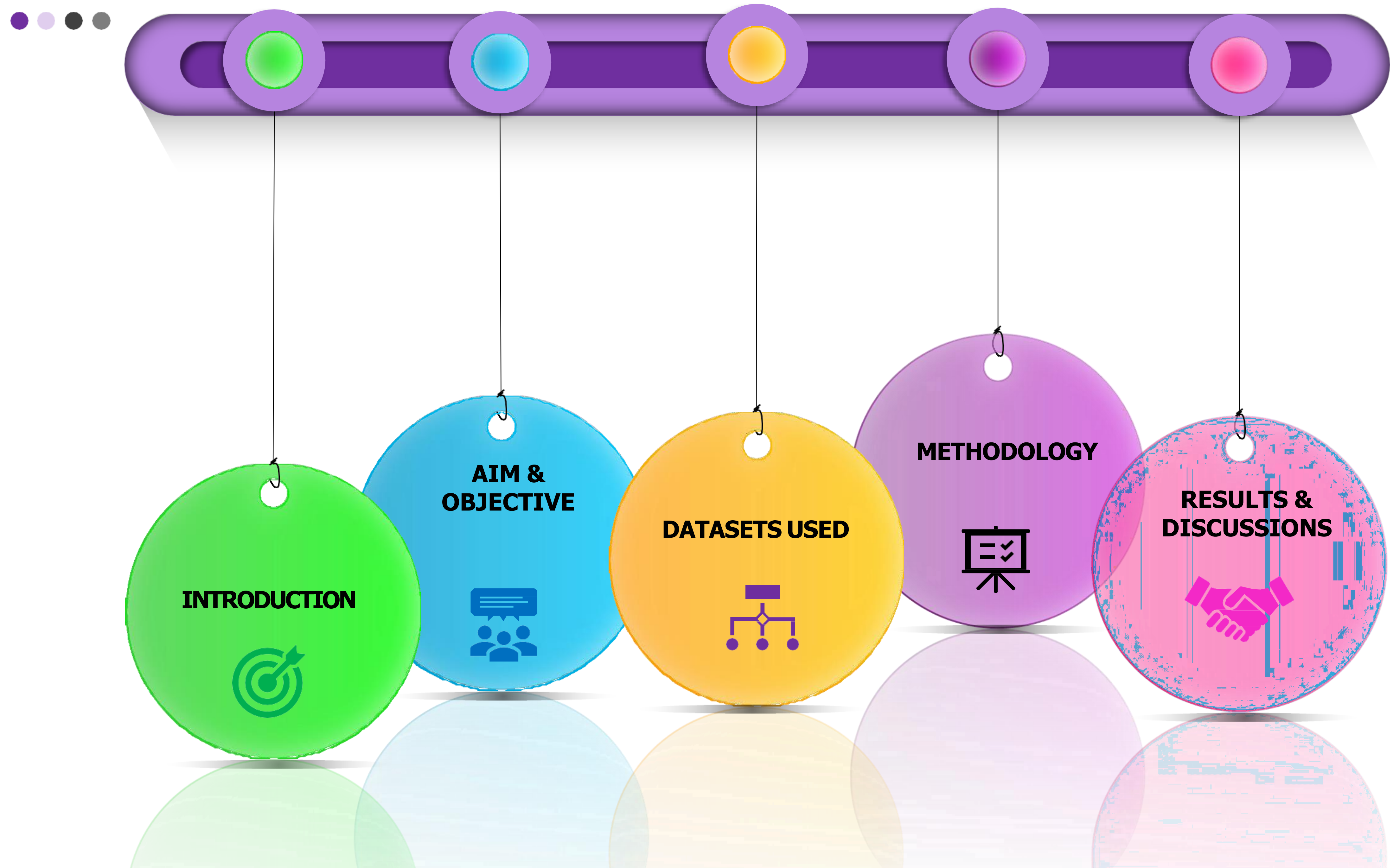
CE 693 – ADVANCED REMOTE SENSING MINI PROJECT

Landslide Susceptibility Mapping using Geospatial and ML Techniques for Dima Hasao and Cachar, Assam, India

By
Meenatchi G
224104007


INTRODUCTION
AIM & OBJECTIVE
DATASETS USED
METHODOLOGY
RESULTS & DISCUSSIONS

AIM & OBJECTIVE

- To identify the landslide-susceptible areas in the Dima Hasao and Cachar regions in Assam, India
- To assess the applicability of Machine Learning algorithms like Random Forest and Support Vector Machine for landslide susceptibility mapping.

Assam: Rains, landslides in Cachar, Dima Hasao affect work on East-West corridor

Ongoing construction of the road has also been affected because of the inclement weather.

NATURAL DISASTERS

Assam's Haflong swept by landslides; activists blame 'development'

The haphazard development of roads and railways without proper study of the soil, has led to disaster, they add

Rain, landslide wreak havoc in Assam; Army, Air Force called in to rescue stranded passengers in Dima Hasao

INTRODUCTION

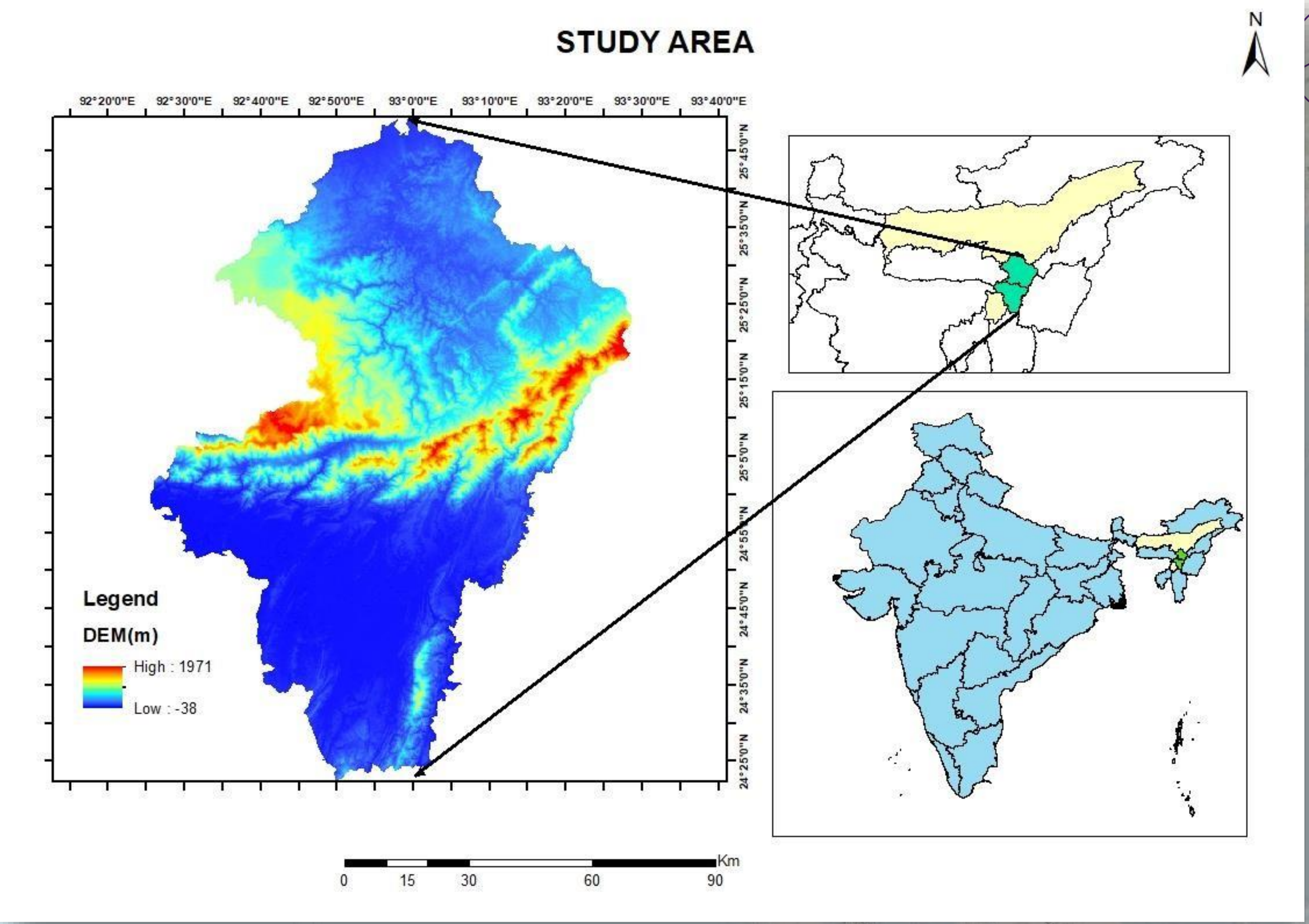
- Guzzetti (2002) defined it as the downslope movement of the mass of rock, debris, or earth materials governed by the typical contribution of different preparatory factors (slope, aspect, elevation, lithology, soil, etc.) that often requires a triggering agent (rainfall, earthquake) before escalating.

- Landslide susceptibility mapping or zonation is the subdivision of the terrain in to zones that have a different likelihood for landslide occurrence. It includes spatial distribution, size, location and displacement of the landslide deposit (Fell et al. 2008; Guzzetti et al. 1999; Varnes 1984).

Related Works

- Pioneering the application of SVM in landslide susceptibility *Yao and Dai, 2006*, *Yao et al., 2008* compared single-class to two-class (binary) SVM in the Hong Kong area. The authors demonstrated how the latter provided better conditions for algorithm training and testing, since it is clearly favorable to know where the landslides exist and where they do not.
- The Study by *Yuan and Zhang (2006)* regarded a specific type of landslide phenomena, the debris flows, by comparing SVM and the fuzzy approach. Since it outperformed the fuzzy method in the testing mode, SVM was considered appropriate and more convenient for this kind of assessment in the area of interest (Yunnan Province, China).
- Machine learning is a branch of artificial intelligence that uses computer algorithms for analysing and predicting information through learning from the training data *(Jordan and Mitchell, 2015)*.
- Literature review shows that various machine learning algorithms have been applied for analysis of landslide susceptibility such as the neuro-fuzzy (*Tien Bui et al., 2012*), *artificial neural network (Chen et al., 2017f; Tien Bui et al., 2016c), decision trees (Hong et al., 2015, Hong et al., 2018b; Tien Bui et al., 2014), support vector machines (Chen et al., 2017b, Chen et al., 2017c; Tien Bui et al., 2017)*, boosted regression trees (Youssef et al., 2015b), naive bayes (Tsangaratos and Ilia, 2016), kernel logistic regression *(Tien Bui et al., 2016c)*, and Random Forests *(Chen et al., 2017d, Chen et al., 2017i; Hong et al., 2016; Trigila et al., 2015)*. However, landslide researchers could not reach a common point about the suitable model for landslide susceptibility study until now.

STUDY AREA

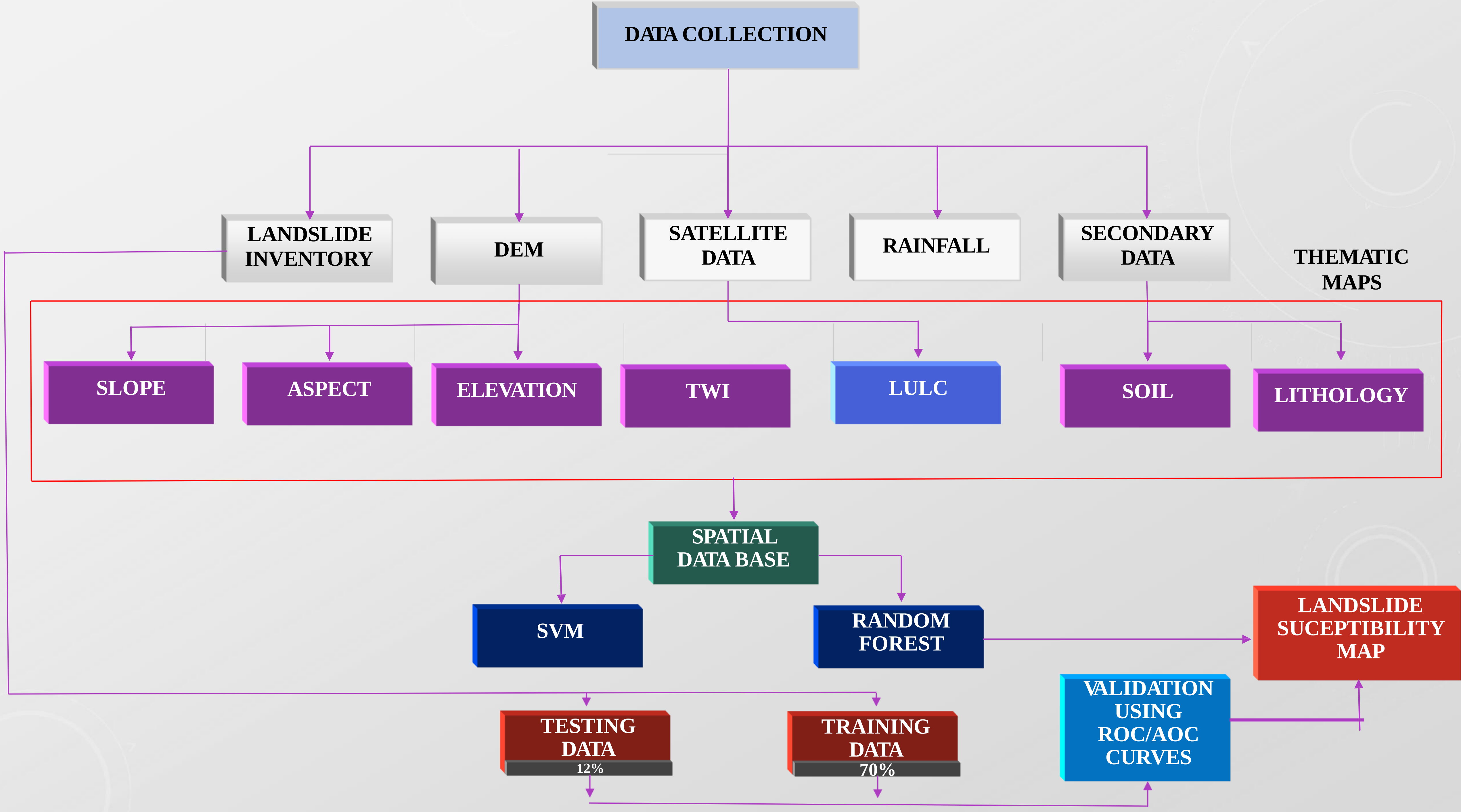


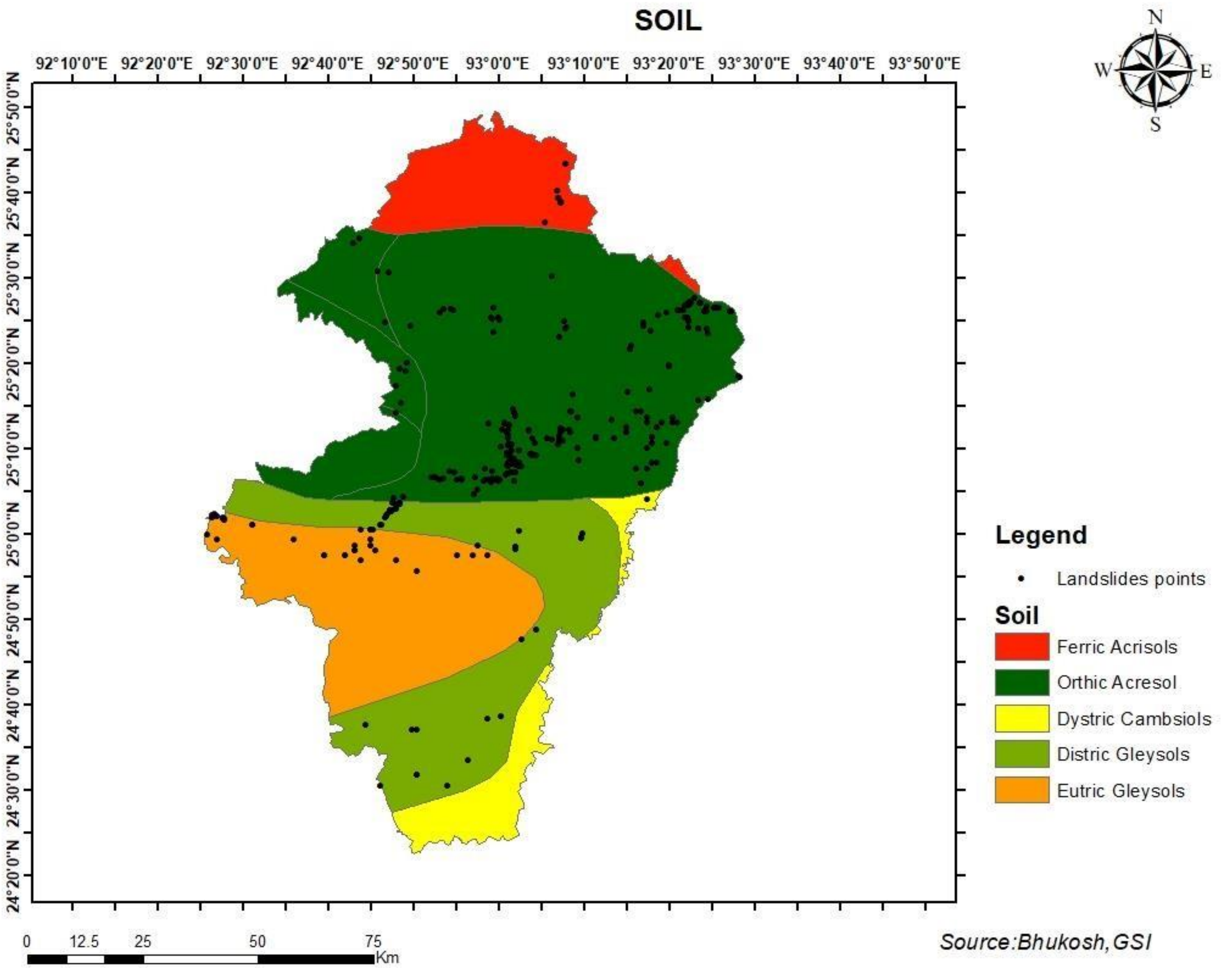
STUDY AREA
N
92°20'0"E
92°30'0"E
92°40'0"E
92°50'0"E
93°0'0"E
93°10'0"E
93°20'0"E
93°30'0"E
93°40'0"E
25°45'0"N
25°35'0"N
25°25'0"N
25°15'0"N
25°5'0"N
24°55'0"N
24°45'0"N
24°35'0"N
24°25'0"N
Legend
DEM(m)
High : 1971
Low : -38
Km
0
15
30
60
90

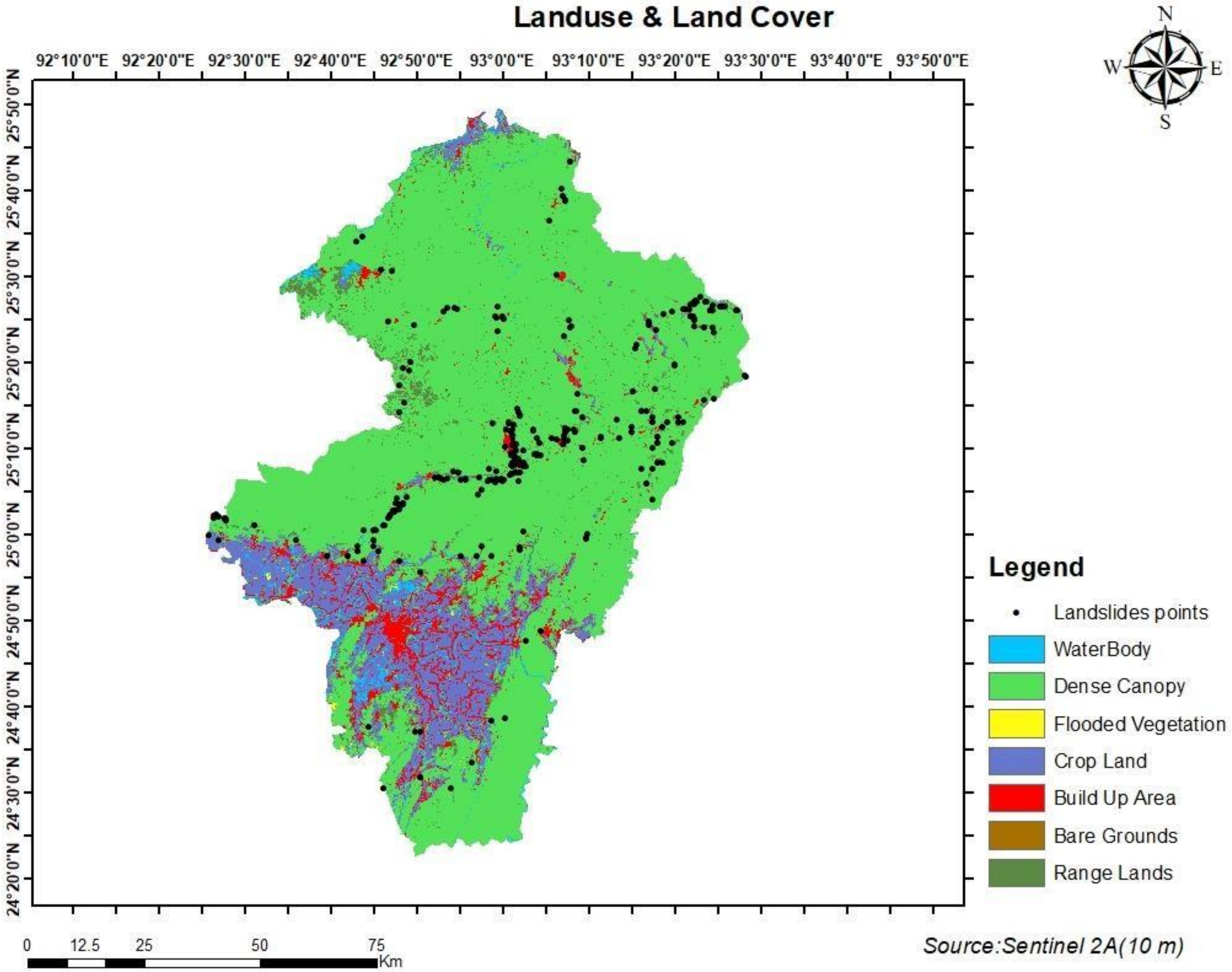
DATASETS USED

Data	Source	Purpose
Landslide Inventory Data and Historical Landslides	Geological Survey of India and satellite imageries	Historical landslides
DEM	SRTM 30 m resolution	Elevation, Slope, Aspect, TWI etc.,
Geological Data	GSI and published data	Soil, Lithology and Geomorphology
Satellite Data – Published	Sentinel 2A 10 m resolution	Landuse and Land cover Mapping
Topographic Wetness Index	CARTOSAT DEM	To calculate the relationship between wetness index and landslide likelihood

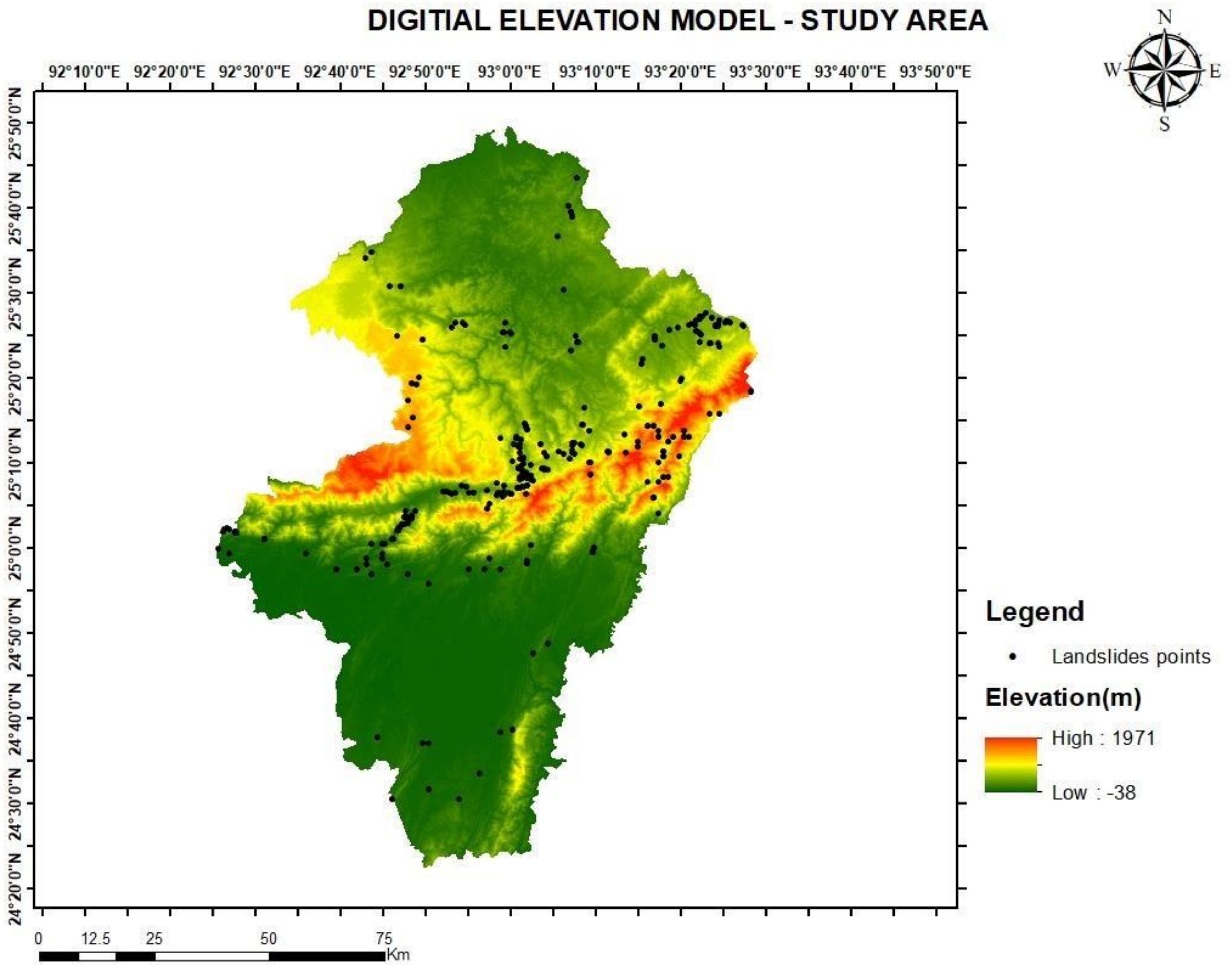
METHODOLOGY

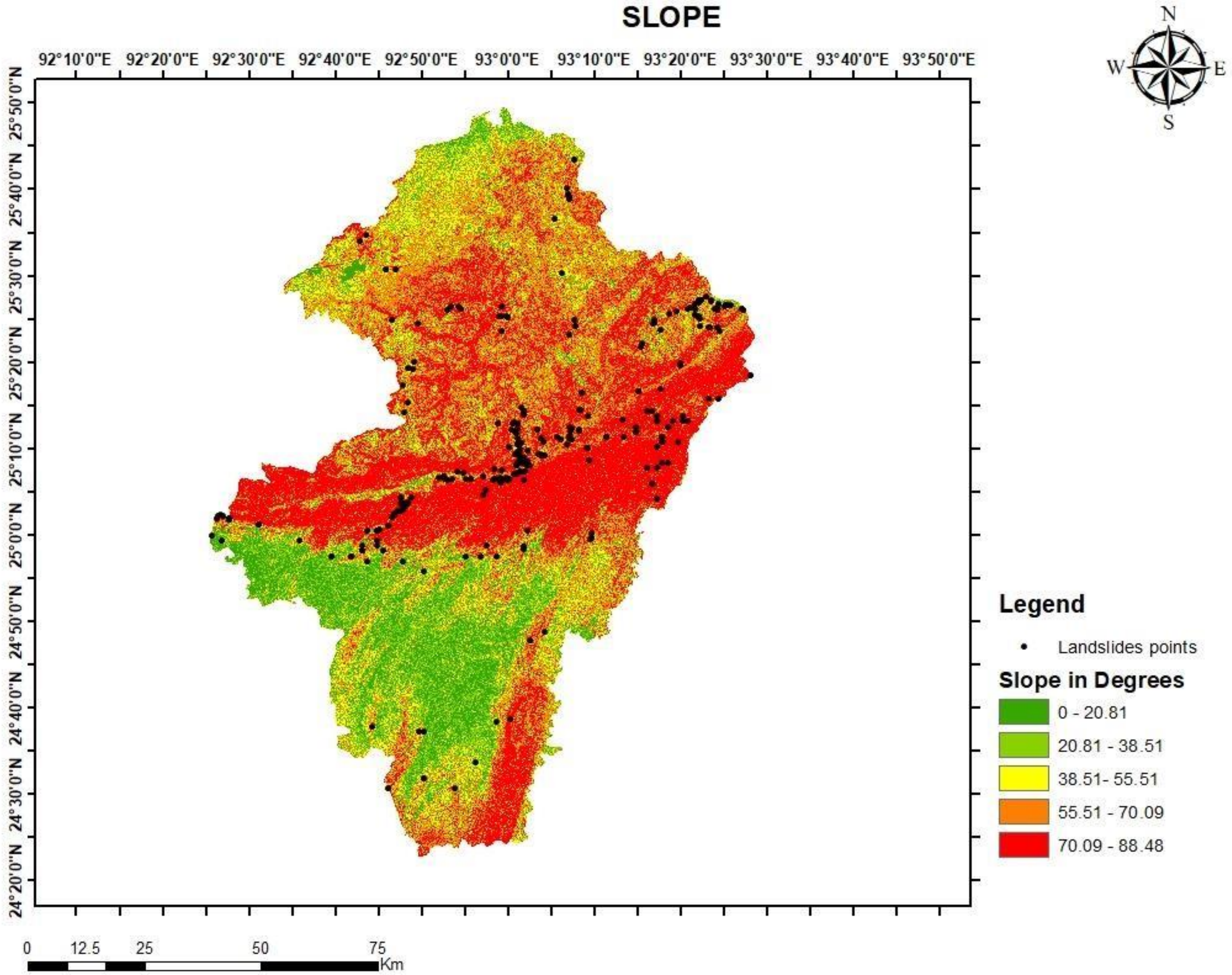

DATA COLLECTION
LANDSLIDE INVENTORY
DEM
SATELLITE DATA
RAINFALL
SECONDARY DATA
THEMATIC MAPS
SLOPE
ASPECT
ELEVATION
TWI
LULC
SOIL
LITHOLOGY
SPATIAL DATA BASE
SVM
RANDOM FOREST
LANDSLIDE SUCEPTIBILITY MAP
VALIDATION USING ROC/AOC CURVES
TESTING DATA
12%
TRAINING DATA
70%

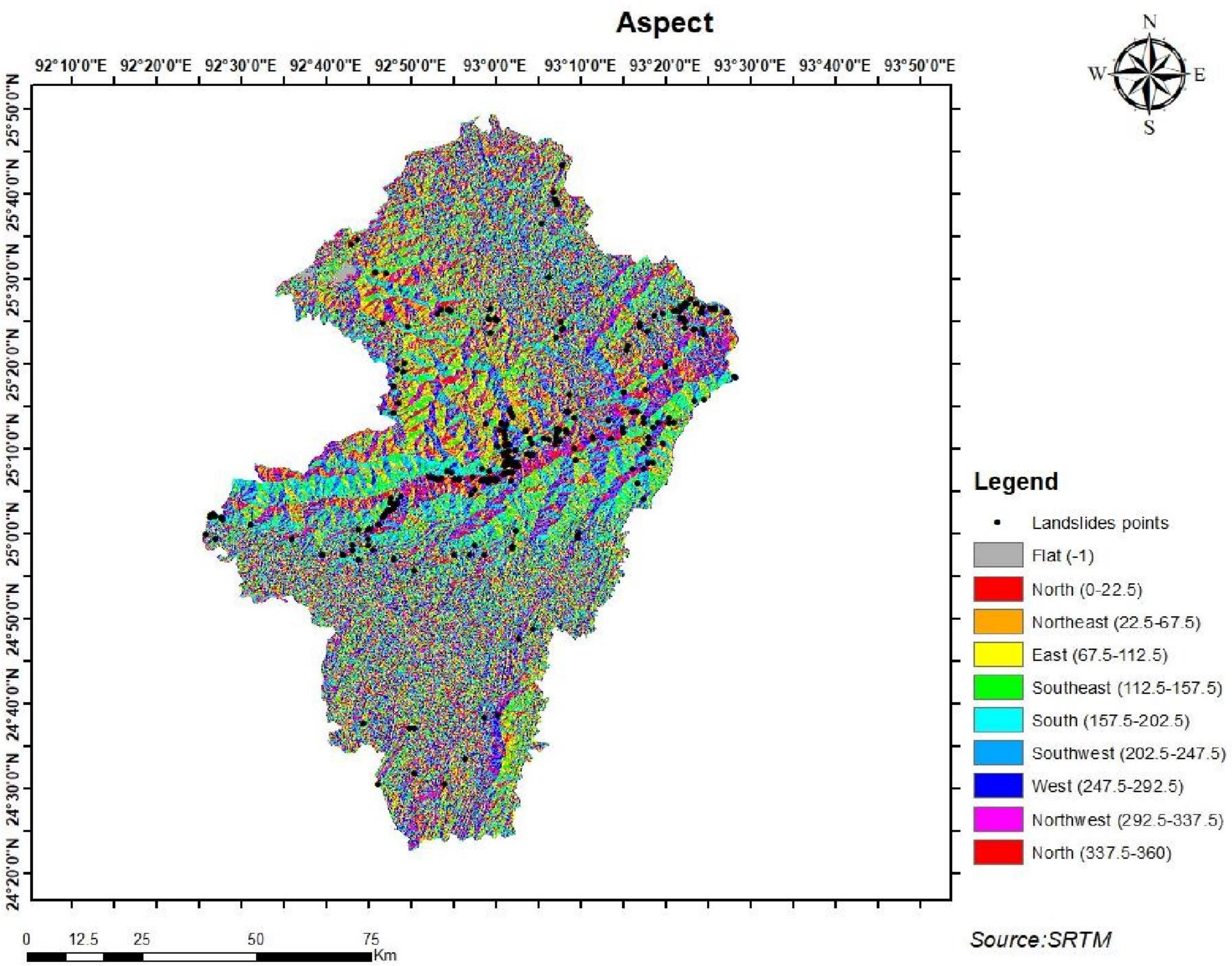

SOIL
92°10'0"E
92°20'0"E
92°30'0"E
92°40'0"E
92°50'0"E
93°0'0"E
93°10'0"E
93°20'0"E
93°30'0"E
93°40'0"E
93°50'0"E
25°50'0"N
25°40'0"N
25°30'0"N
25°20'0"N
25°10'0"N
25°0'0"N
24°50'0"N
24°40'0"N
24°30'0"N
24°20'0"N
N
S
E
W
Legend
Landslides points
Soil
Ferric Acrisols
Orthic Acresol
Dystric Cambsiols
Distric Gleysols
Eutric Gleysols
0
12.5
25
50
75
Km
Source:Bhukosh,GSI

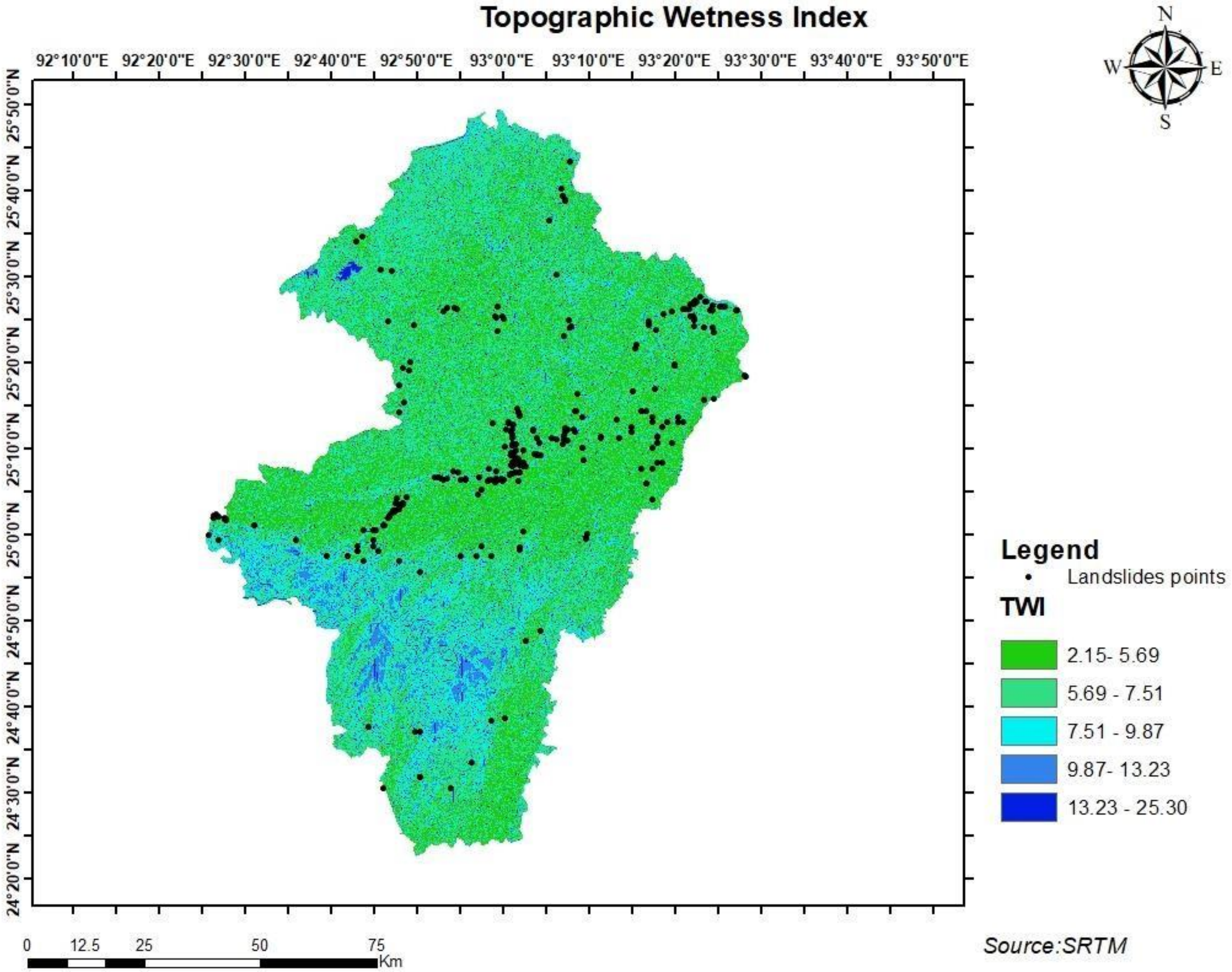


Landuse & Land Cover
92°10'0"E
92°20'0"E
92°30'0"E
92°40'0"E
92°50'0"E
93°0'0"E
93°10'0"E
93°20'0"E
93°30'0"E
93°40'0"E
93°50'0"E
25°50'0"N
25°40'0"N
25°30'0"N
25°20'0"N
25°10'0"N
25°0'0"N
24°50'0"N
24°40'0"N
24°30'0"N
24°20'0"N
N
W
E
S
Legend
Landslides points
WaterBody
Dense Canopy
Flooded Vegetation
Crop Land
Build Up Area
Bare Grounds
Range Lands
0
12.5
25
50
75
Km
Source:Sentinel 2A(10 m)


DIGITIAL ELEVATION MODEL - STUDY AREA
92°10'0"E
92°20'0"E
92°30'0"E
92°40'0"E
92°50'0"E
93°0'0"E
93°10'0"E
93°20'0"E
93°30'0"E
93°40'0"E
93°50'0"E
25°50'0"N
25°40'0"N
25°30'0"N
25°20'0"N
25°10'0"N
25°0'0"N
24°50'0"N
24°40'0"N
24°30'0"N
24°20'0"N
N
W
E
S
Legend
Landslides points
Elevation(m)
High : 1971
Low : -38
0
12.5
25
50
75
Km


SLOPE
92°10'0"E 92°20'0"E 92°30'0"E 92°40'0"E 92°50'0"E 93°0'0"E 93°10'0"E 93°20'0"E 93°30'0"E 93°40'0"E 93°50'0"E
25°50'0"N 25°40'0"N 25°30'0"N 25°20'0"N 25°10'0"N 25°0'0"N 24°50'0"N 24°40'0"N 24°30'0"N 24°20'0"N
N
W
E
S
Legend
Landslides points
Slope in Degrees
0 - 20.81
20.81 - 38.51
38.51- 55.51
55.51 - 70.09
70.09 - 88.48
0 12.5 25 50 75
Km


Aspect
92°10'0"E
92°20'0"E
92°30'0"E
92°40'0"E
92°50'0"E
93°0'0"E
93°10'0"E
93°20'0"E
93°30'0"E
93°40'0"E
93°50'0"E
25°50'0"N
25°40'0"N
25°30'0"N
25°20'0"N
25°10'0"N
25°0'0"N
24°50'0"N
24°40'0"N
24°30'0"N
24°20'0"N
N
S
W
E
Legend
Landslides points
Flat (-1)
North (0-22.5)
Northeast (22.5-67.5)
East (67.5-112.5)
Southeast (112.5-157.5)
South (157.5-202.5)
Southwest (202.5-247.5)
West (247.5-292.5)
Northwest (292.5-337.5)
North (337.5-360)
0
12.5
25
50
75
Km
Source:SRTM



Topographic Wetness Index
92°10'0"E
92°20'0"E
92°30'0"E
92°40'0"E
92°50'0"E
93°0'0"E
93°10'0"E
93°20'0"E
93°30'0"E
93°40'0"E
93°50'0"E
25°50'0"N
25°40'0"N
25°30'0"N
25°20'0"N
25°10'0"N
25°0'0"N
24°50'0"N
24°40'0"N
24°30'0"N
24°20'0"N
N
S
E
W
Legend
Landslides points
TWI
2.15- 5.69
5.69 - 7.51
7.51 - 9.87
9.87- 13.23
13.23 - 25.30
0
12.5
25
50
75
Km
Source:SRTM

RANDOM FOREST ALGORITHM

Random Forest (RF) form of bagged DT-based classifier, which is used to solve complex problems of prediction and multi-classification.

It combines the bagging integrated learning algorithm with the random subspace algorithm, making it a powerful classification tool capable of recognizing large-scale and multivariable data. The implementation processes of the RF algorithm include:

(1) Select n sample features randomly from the total sample;

(2)Generate n trees on n subsets and subsample their features when generating trees;

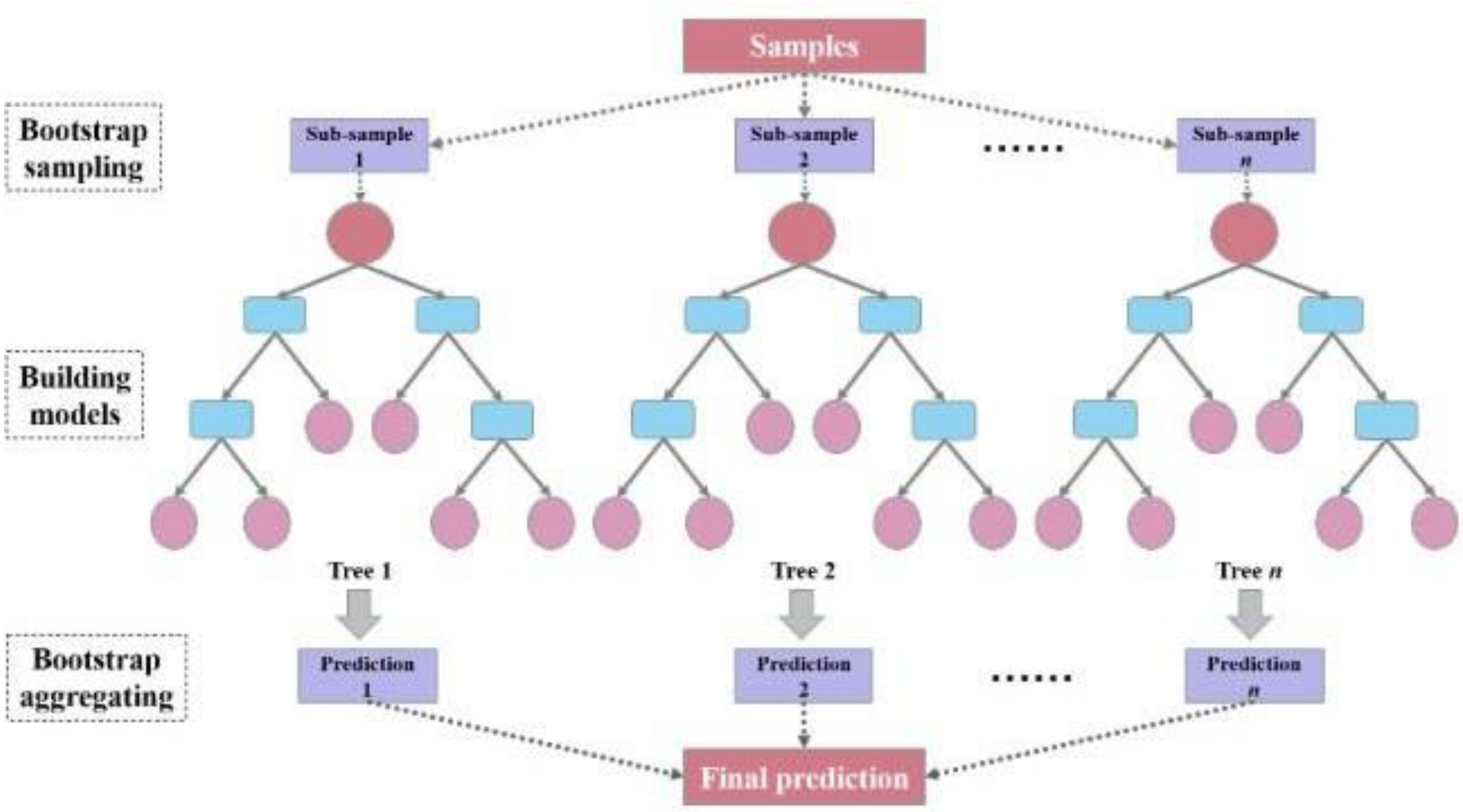
(3) Obtain n predictions by all trees

(4) Calculate the mode or average of the n predictions as the final output.

RANDOM FOREST

- The Random Forest Algorithm is used to prepare a model from training dataset**. Random Forest Classifier** in Python was used to prepare a landslide susceptibility model from the training dataset available.
- In order to distinguish each predictor in the ensemble classifier, a specific number of variables are stochastically selected for generating the necessary nodes in the decision-tree. This construction method enables the RF to further improve the prediction performance through the increase of the difference among the individual classification trees and to avoid over-fitting. Also, the model was used on test data set to check the
- accuracy of the model.

Flowchart of classification using the random forest (RF) algorithm

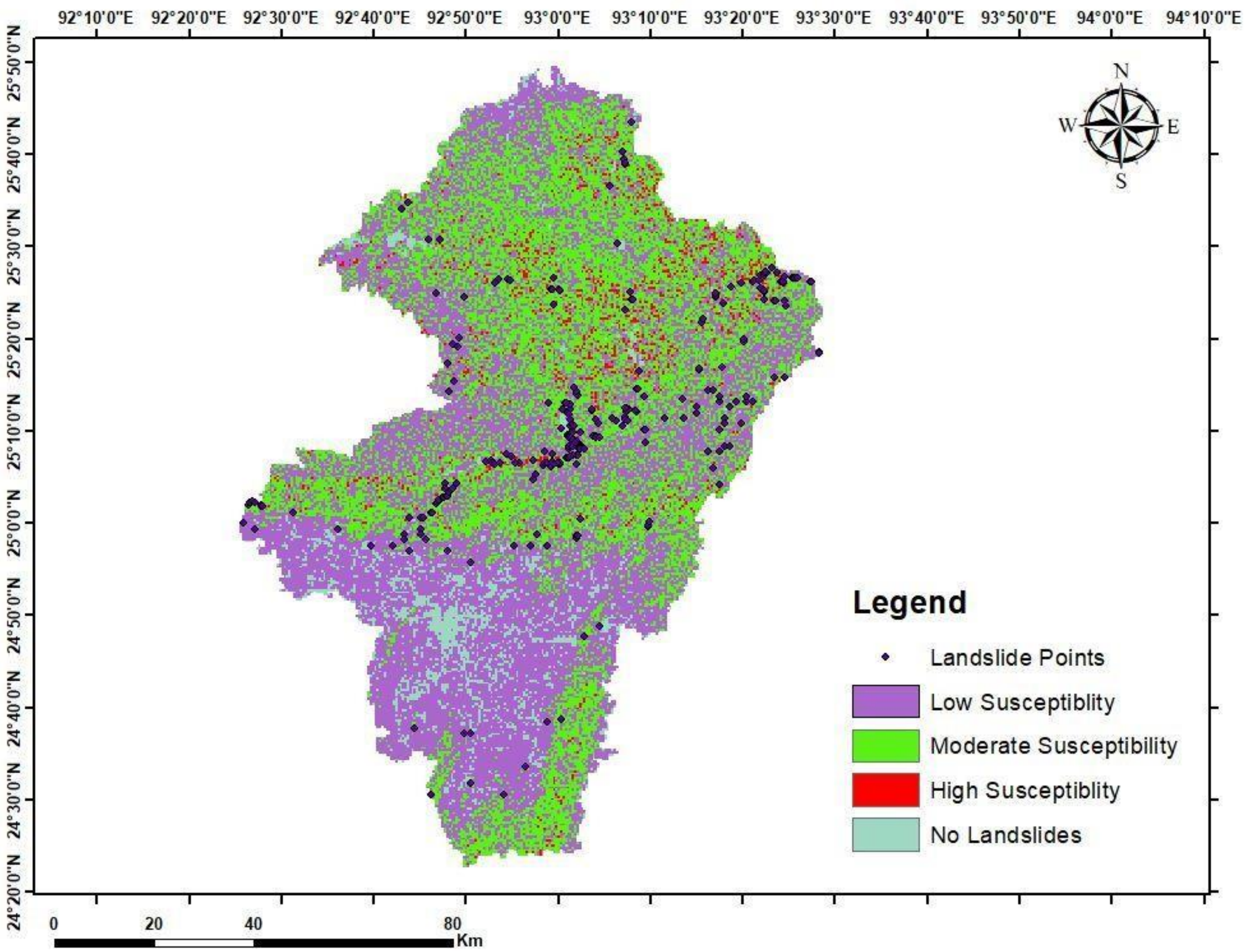


Source: Yuke Huan

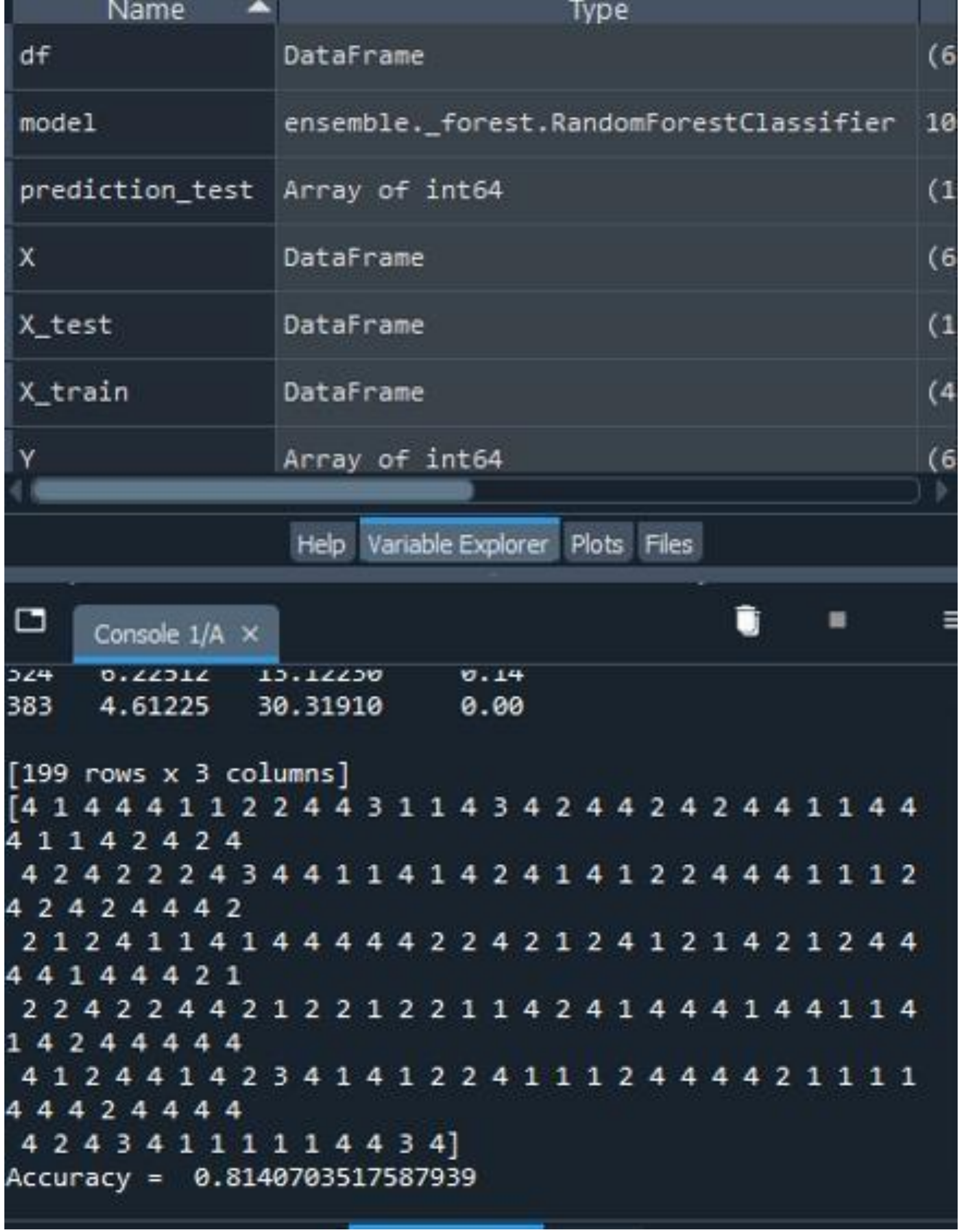
MODEL VALIDATION – TRAINING DATASET

318	13627	4.33969	21.619	0.14	0.25	125	2

Landslide Suscpetibility Map



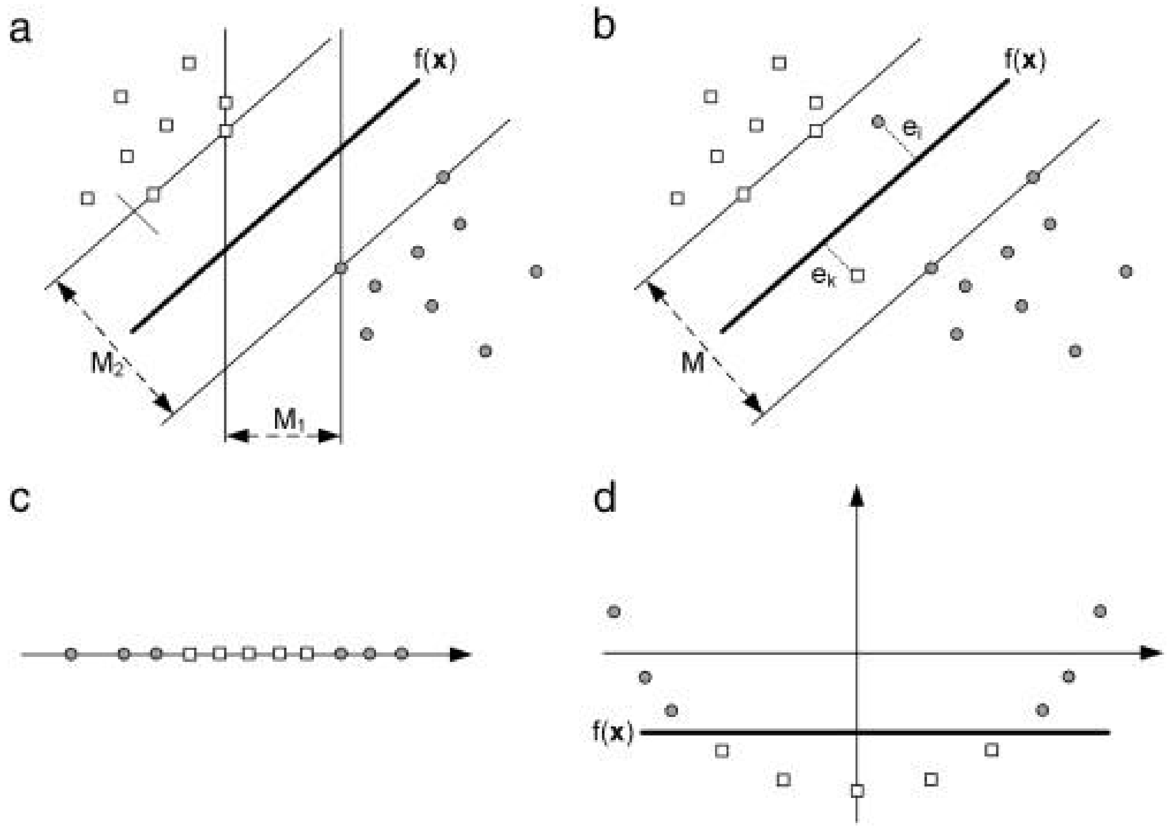
CLASSIFICATION ACCURACY = 81.4%



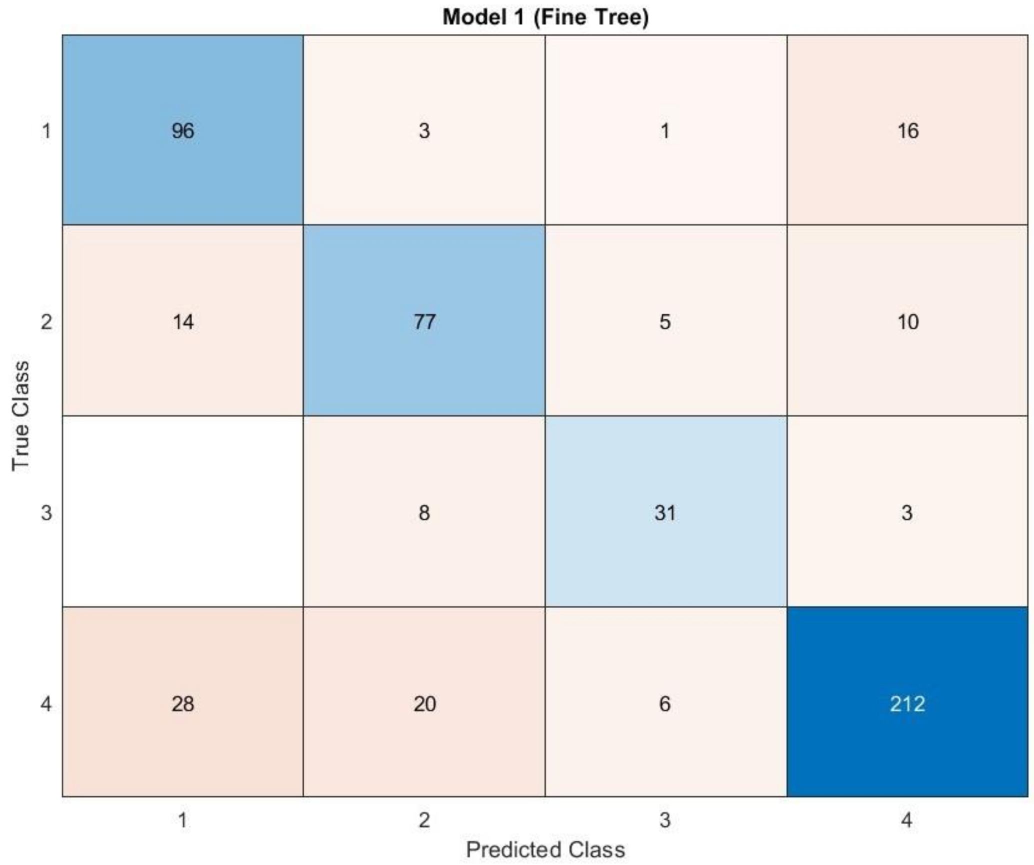
DECISION TREE AND SUPPORT VECTOR MACHINE

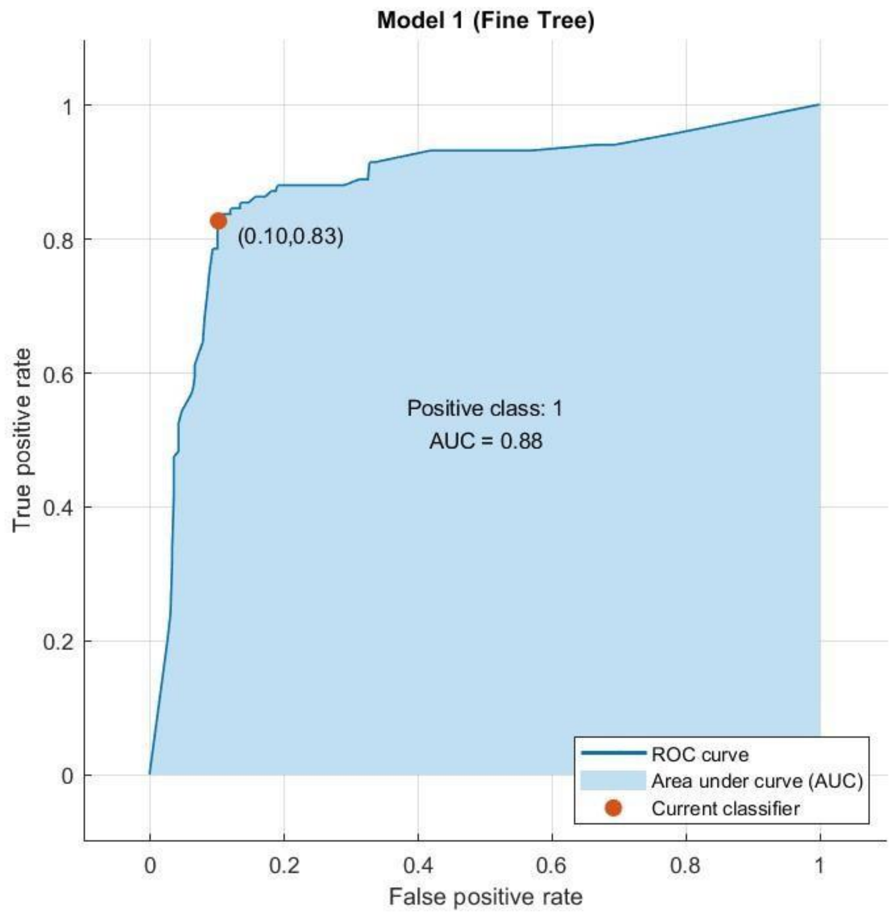
- **The DT model** is essentially a tree involving a set of **decision nodes**, among which the root and each internal node are labelled with a question (Pradhan, 2013).
- The arcs descend from each root node to leaf nodes, where a solution to the associated question is offered.
- **SVM is a binary classifier** (instances could be classified to only one of the two classes), but one can easily transform *n*-class problems into the sequence of *n* (one-versus-all) or $n(n-1)/2$ (one-versus-one) binary classification tasks (Belousov et al., 2002). **The basic variant of the algorithm attempts to generate a separating hyper-plane in the original space of *n* coordinates (x_i parameters in vector x) between the points of two distinct classes**.
- The SVM differs from other separating hyper-plane approaches in the way the hyper-plane is constructed from the training set points (The algorithm seeks a maximum margin of separation between the classes ($M_2 > M_1$)and constructs a classification hyper-plane in the middle of the maximum margin If a point $\mathbf{x} \in R^n$ is above the hyper-plane, it is classified as + 1 otherwise it is − 1.

DECISION TREE AND SUPPORT VECTOR MACHINE

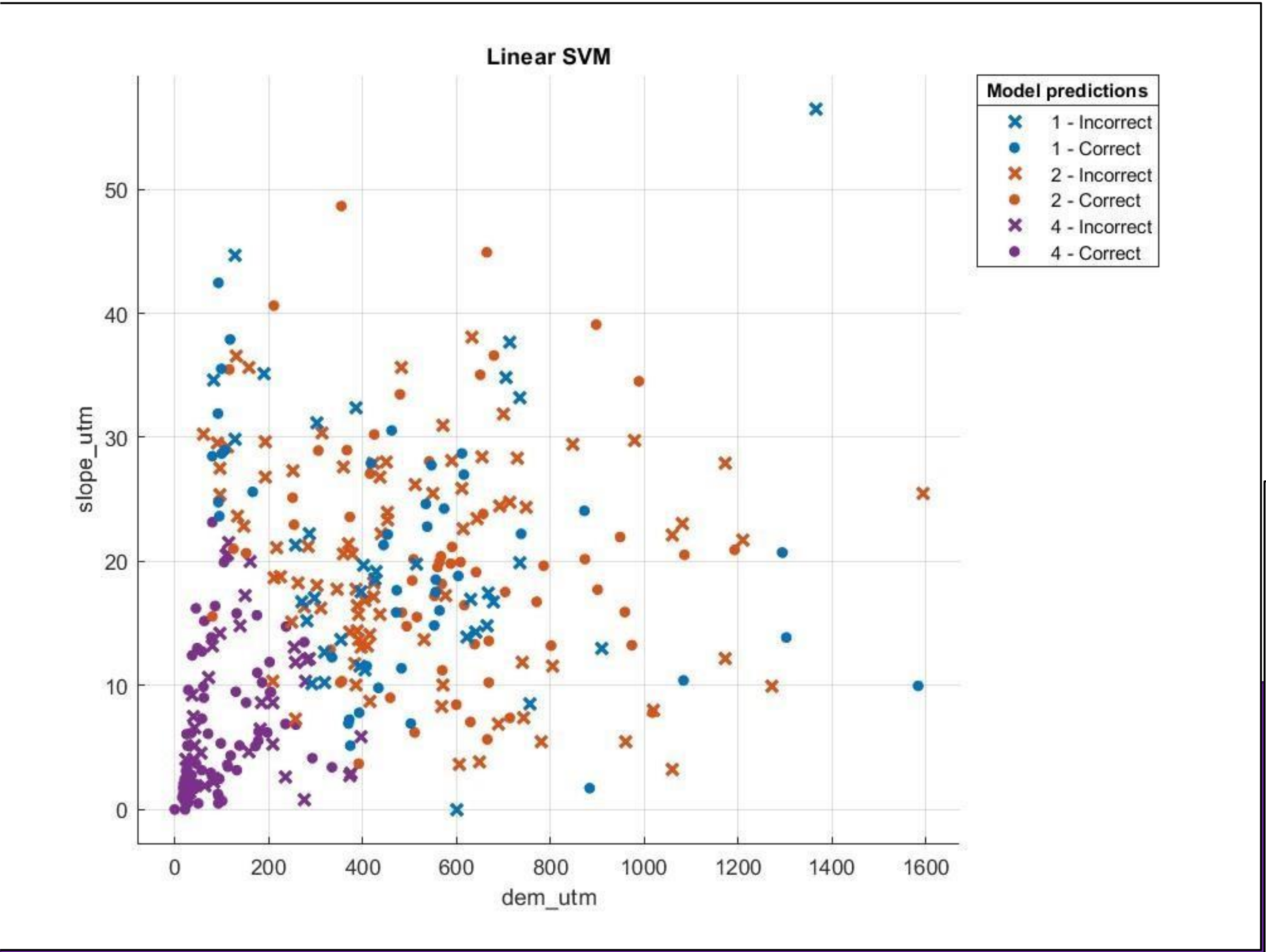


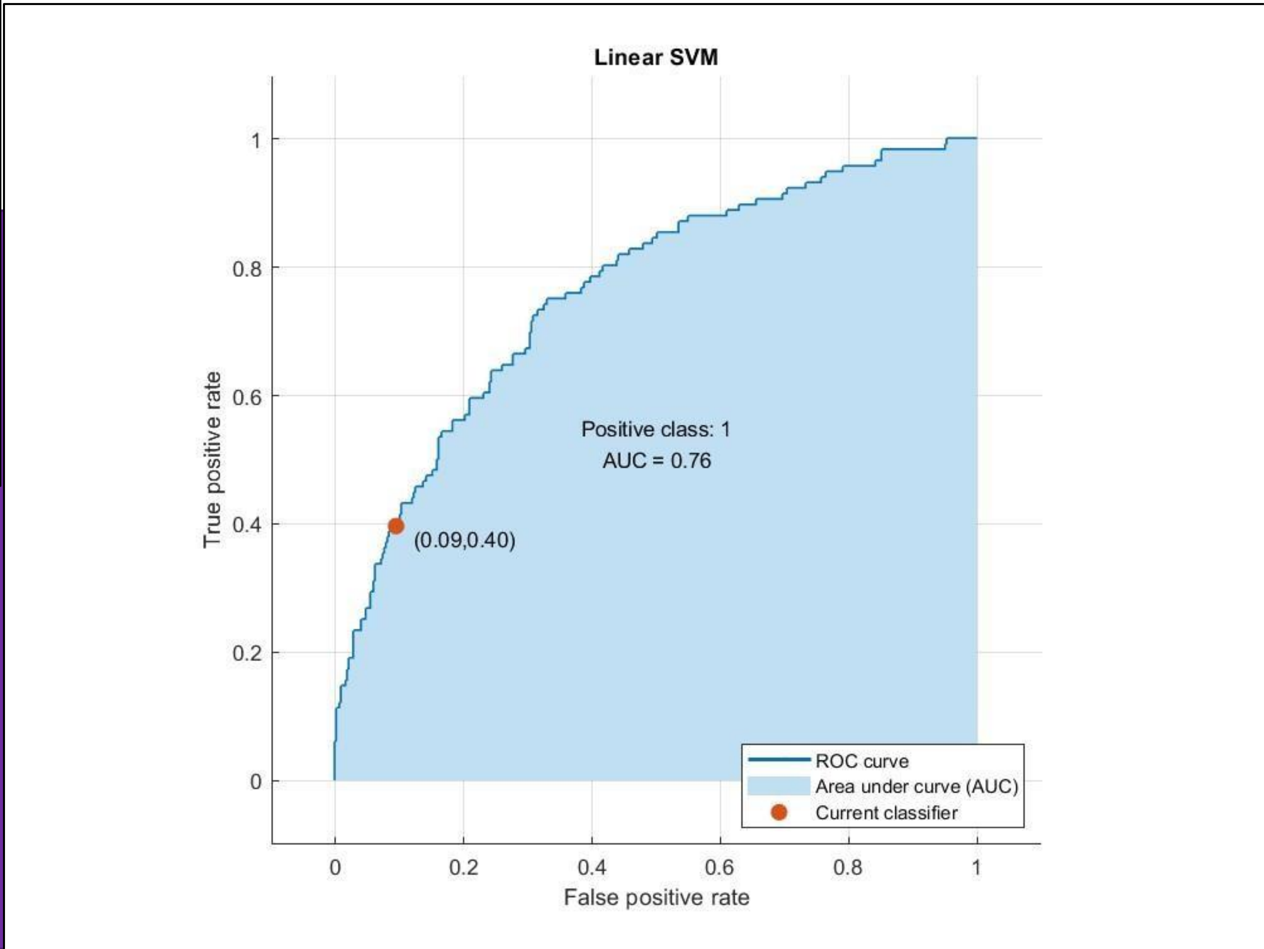
a) Maximum-margin classifier $f(\mathbf{x})$ separating circles from squares in R2; b) Soft-margin classifier allows some points to be misclassified; c) Linearly inseparable case in R1; d) Mapping original input space into feature space of higher dimension (R^2). Classes become linearly separable after the mapping.





DECISION TREE MODEL SUMMARY





SVM MODEL SUMMARY

Accuracy = True Positive + True Negative/Total

= 68.7%

CONCLUSIONS

1. Majority of the landslides were found to occur in the acrisol soils. The gleysols registered a few with cambsiols showing the least number of landslides.
2. Built-up areas, crop lands and bare lands showed low susceptibility to landslides whereas the regions occupied by dense canopies have the majority of the landslides within them.
3. Although elevation didn't have a significant influence on the landslide susceptibility, regions with slopes higher than 38° were found to have be more susceptible.
4. Landslide distribution was seen more prominently in areas with low TWI values.
5. By inspecting the landslide susceptibility map we can see that most of the landslide points are concentrated in the high and moderately susceptibility regions.
6. The accuracy of the Random forest method was found to be 81.4%.
7. SVM, involves a training phase in which the model is trained by a training dataset of associated input and target output values. The trained model is then used to evaluate a separate set of testing data. The overall performance of the SVM model depends on the kernel parameters, such as the regularization parameter (C) and the kernel width (γ). This indicates how the model accurately predicts landslides from input data in the training and test. Among them, the prediction success rate was 69.7%.This shows that the established SVM model had good predictive performance.
8. The confusion matrix provides the following four parameters: the true-positive (TP) and true-negative (TN), referring to the numbers of pixels correctly classified (landslide, non-landslide), and the false-positive (FP) and false-negative (FN), indicating the numbers of pixels classified incorrectly. Once the models were trained and validated, they can be used to estimate the Landslide susceptibility for every pixel in the study area.
9. The final results shows that the RF model has good predictive performance than the SVM and Decision Tree for Landslide Susceptibility Mapping.

REFERENCES

- Chang Z, Du Z, Zhang F et al (2020) Landslide susceptibility prediction based on remote sensing images and GIS: Comparisons of supervised and unsupervised machine learning models
- Chen W, Xie X, Peng J et al (2017) GIS-based landslide susceptibility modelling: a comparative assessment of kernel logistic regression, Naïve-Bayes tree, and alternating decision tree models. Geomatics Nat Hazards Risk 8:950–973.
- de Oliveira GG, Ruiz LFC, Guasselli LA, Haetinger C (2019) Random forest and artificial neural networks in landslide susceptibility modeling: a case study of the Fão River Basin, Southern Brazil. Nat Hazards 99:1049–1073
- Di Napoli M, Carotenuto F, Cevasco A et al (2020) Machine learning ensemble modelling as a tool to improve landslide susceptibility mapping reliability. Landslides 17:1897–1914.
- Das, I., Stein, A., Kerle, N., and Dadhwal, V. K. V. (2012). Landslide susceptibility mapping along road corridors in the Indian Himalayas using Bayesian logistic regression models. *Geomorphology* 179, 116–125. doi:10.1016/j.geomorph.2012.08.004
- Erener, A., and Düzgün, H. S. B. (2011). Landslide susceptibility assessment: what are the effects of mapping unit and mapping method? *Environ. Earth Sci.* 66 (3), 859–877. doi:10.1007/s12665-011-1297-0
- Feizizadeh, B., Roodposhti, M. S., Blaschke, T., and Aryal, J. (2017). Comparing GIS-based support vector machine kernel functions for landslide susceptibility mapping. *Arabian J. Geosciences* 10 (5), 122. doi:10.1007/s12517-017-2918-z
- Xiao, T., Yin, K., Yao, T., and Liu, S. (2019). Spatial prediction of landslide susceptibility using GIS-based statistical and machine learning models in Wanzhou County, Three Gorges Reservoir, China. *Acta Geochim* 38 (5), 654–669. doi:10.1007/s11631-019-00341-1
- Zhao, X., and Chen, W. (2020). GIS-based evaluation of landslide susceptibility models using certainty factors and functional trees-based ensemble techniques. *Appl. Sci.* 10, 16. doi:10.3390/app10010016

THANK YOU