# Snow Avalanche Susceptibility Mapping Using Machine Learning Algorithms

**Remote Sensing and GIS Application to Cryosphere**

**GNR618**

**Course Instructor: Prof. Gulab Singh**

By

**Chandra Prabha R (22D0294)**

**Uday Gohil (20D070030)**

**Umang (200070085)**

**Saketh.K.S (210260025)**

**Atharva Deshpande (210260012)**

**Gourav (210260017)**

**Divyanshu (21D100005)**

**Prajyot (21B080022)**

**Indian Institute of Technology Bombay**

# Table of Contents

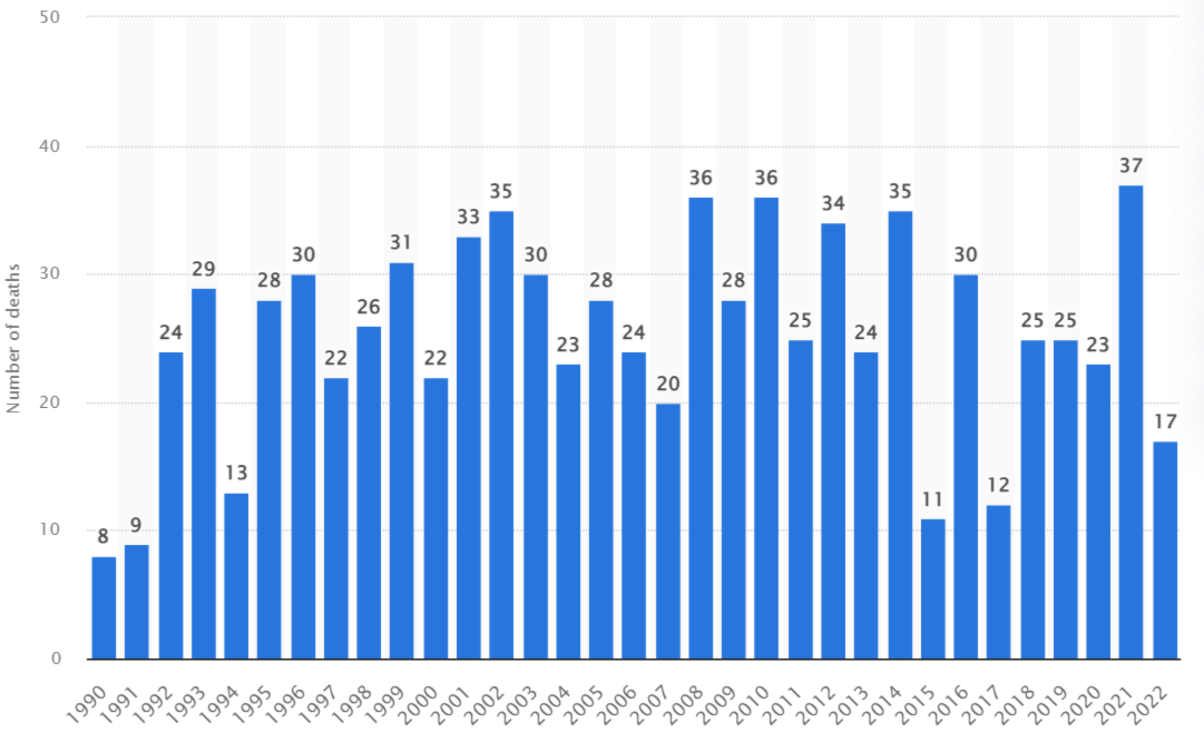
# Chapter 1

# Introduction

## 1.1 Introduction

In mountainous and frigid places, snow avalanches are among the frequently observed natural hazards that influence human lives, the economy, infrastructure, vegetation, and geomorphology. A snow

A quickly moving mass of snow on steep slopes is what an avalanche is. The spatial likelihood of an avalanche occurring is known as snow avalanche susceptibility. Avalanche susceptibility evaluation is the first and most crucial stage of hazard and risk assessment for disaster management and mitigation. Each year avalanches kill more than 150 people worldwide and around 30 people in US (shown in the graph). In 90% of avalanche accidents, the victim or someone in the victim's party causes the snow slide.

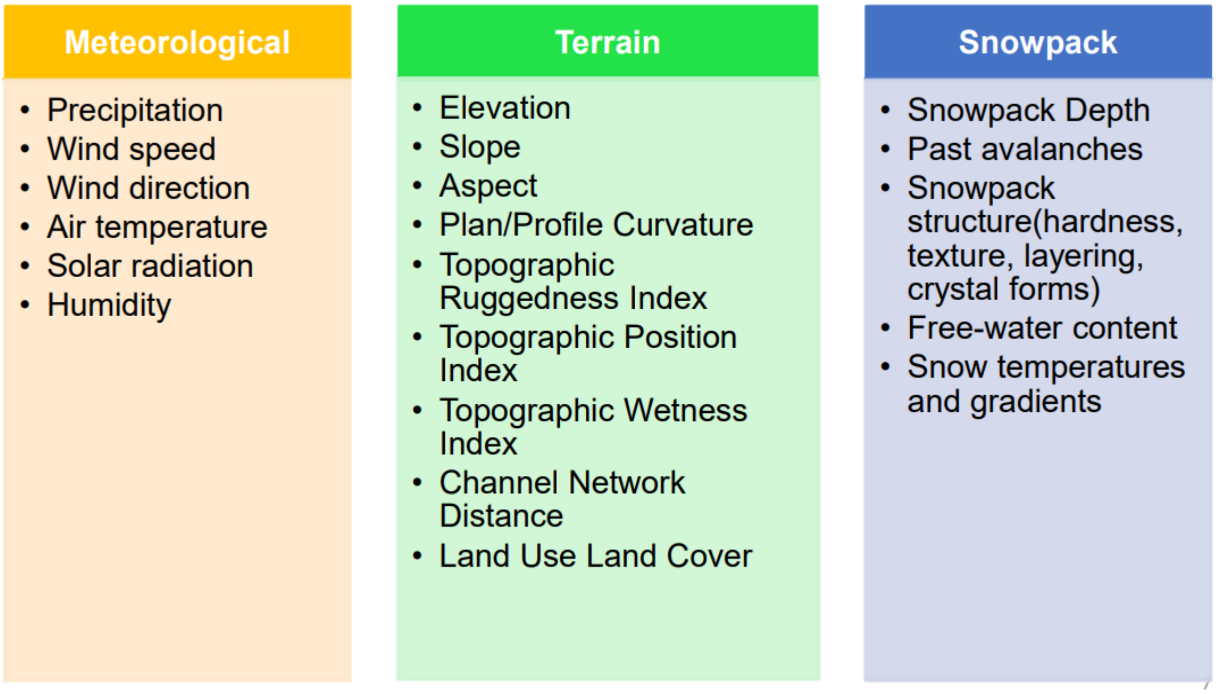



Geographic Information Systems (GIS) and remote sensing (RS) are used to model, map, visualise, and monitor vulnerable zones in order to reduce risks associated with snow avalanches. Compared to GIS and remote sensing-based methodologies, field-based research is constrained by the high-risk exposure. It can be time-consuming due to the instability of the snow mass and unfavourable weather conditions. Yet, GIS and RS are essential and affordable tools for avalanche assessments.

For avalanche susceptibility mapping (ASM), researchers have used a variety of expert-based techniques, including fuzzy frequency ratio (FR), analytical hierarchical process (AHP), etc. The ability of data-driven machine learning (ML) applications to learn, predict, and improve based on past hazard events without human intervention, as well as their capacity to identify trends and patterns and deal with multidimensional and multi-source data like conditioning and triggering factors, have made them extremely effective for natural hazard assessments in recent years. Although there are several ML applications for floods and landslides, the limitations in creating an inventory have prevented us from fully understanding how ASM works.

For the Karaj Watershed in Iran, Mosavi et al. (2020) developed an ensemble machine learning model, the random subspace functional tree (RSFT), and compared the model results with those from other ML techniques like logistic regression (LR), logistic model tree (LMT), alternating decision tree (ADT), and functional trees (FT).
Support Vector Machine (SVM) was used by Tiwari et al. (2021) to forecast avalanche susceptibility using 4 distinct kernel techniques. According to Akay (2021), the Random Forest (RF) is suitable for ASM. Rahmati et al. (2019a) revealed that the RF method outperformed other ML techniques for producing avalanche susceptibility maps (ASMs) at two separate sites.

Twelve conditioning factors, including elevation, slope, plan curvature, profile curvature, aspect, topographic position index (TPI), topographic ruggedness index (TRI), topographic wetness index (TWI), land use/land cover, lithology, distance to road, and distance to the river, were used in this study to produce the ASM of Davos (Switzerland). The avalanche inventory was created by Hafner et al. in prior work (2021a) as vector data (polygons), and it was made available for the current study's use as the training data for the supervised ML algorithms discussed above. In the following Sections, the datasets, methods, and ASM results are presented in detail and discussed accordingly.




*Schaerer, P. and McClung, D. (2006) The Avalanche Handbook. 3rd Edition, Mountaineers Books*

The digital elevation model (DEM) employed in the study (swissALTI3D) was freely provided by Swisstopo (2021), and was downsampled here to 10 m spatial resolution for computational reasons. The slope, plan curvature, profile curvature, aspect, TPI, TRI, and TWI were calculated from the 10 m DEM using the SAGA GIS software (Conrad et al., 2015).

## 1.2 Objectives

a. To train and test different machine learning algorithms for identifying avalanche susceptible zones.

b. To identify avalanche susceptible zones using the trained models in the Indian Himalayan region.

# Chapter 2
# Study Area

## 2.1 Study Area

The study focuses on the snow-covered mountainous region of Davos, Switzerland. The region is well know for tourism and skiing resorts. It is also affected by disasters such as avalanches during the winter.

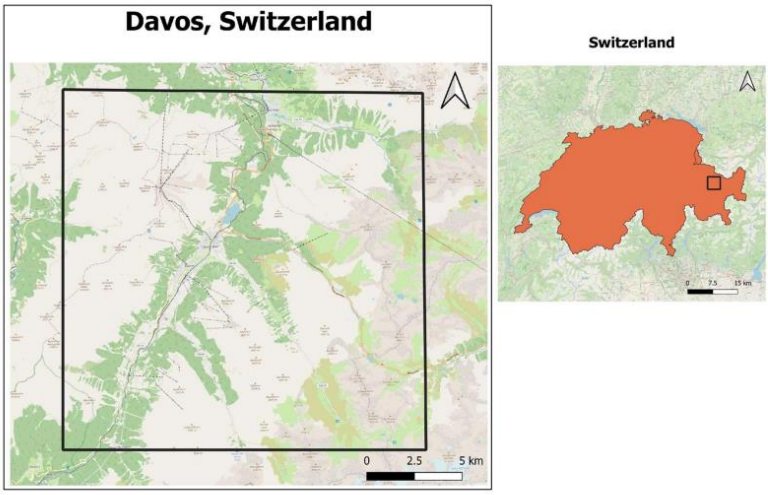
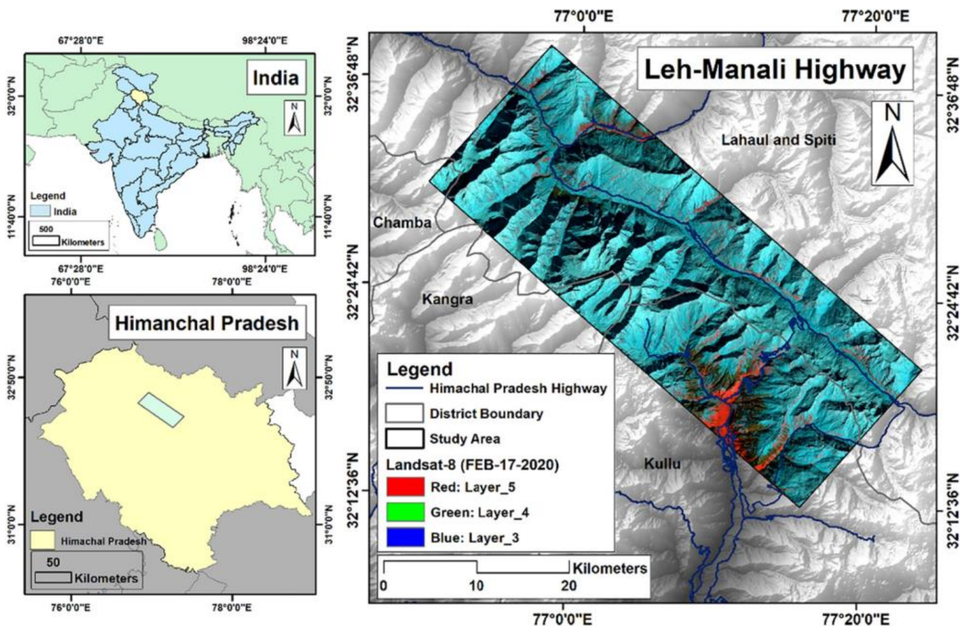



Fig 2.1: Study Area

Another study area was a part of the  Leh Manali Highway in the Indian Himalayan region. This region is annually affected by avalanches which hinders the transportation routes and the livelihood of the local community.



# Chapter 3
# Datasets used

## 3.1 SRTM DEM

The Shuttle Radar Topography Mission (SRTM) is an international research effort that obtained digital elevation models on a near-global scale from 56°S to 60°N,[2]: 4820 to generate the most complete high-resolution digital topographic database of Earth prior to the release of the ASTER GDEM in 2009. SRTM consisted of a specially modified radar system that flew on board the Space Shuttle Endeavour during the 11-day STS-99 mission in February 2000. The radar system was based on the older Spaceborne Imaging Radar-C/X-band Synthetic Aperture Radar (SIR-C/X-SAR), previously used on the Shuttle in 1994. To acquire topographic data, the SRTM payload was outfitted with two radar antennas.[2] One antenna was located in the Shuttle's payload bay, the other – a critical change from the SIR-C/X-SAR, allowing single-pass interferometry – on the end of a 60-meter (200-foot) mast that extended from the payload bay once the Shuttle was in space.[2] The technique employed is known as interferometric synthetic aperture radar. Intermap Technologies was the prime contractor for processing the interferometric synthetic aperture radar data.

## 3.2 Landsat Data

The Landsat Program is a series of Earth-observing satellite missions jointly managed by NASA and the U.S. Geological Survey.

Landsat satellites have the optimal ground resolution and spectral bands to efficiently track land use and to document land change due to climate change, urbanization, drought, wildfire, biomass changes (carbon assessments), and a host of other natural and human-caused changes.

The Landsat Program's continuous archive (1972-present) provides essential land change data and trending information not otherwise available. Landsat represents the world's longest continuously acquired collection of space-based moderate-resolution land remote sensing data. Landsat is an essential capability that enables the U.S. Department of the Interior to wisely manage Federal lands. People around the world are using Landsat data for research, business, education, and other activities.

## 3.3 Avalanche Inventory

Avalanche occurrence zones were identified and demarcated from SPOT 6 high resolution images.

| Sl.No. | Data | Spatial Resolution | Source |
|---|---|---|---|
| 1 | SRTM DEM | 30m | USGS Earth Explorer |
| 2 | Landsat | 30m | USGS Earth Explorer |
| 3 | Avalanche Inventory | Vector Data(Polygons) | Hafner, E. and Bühler, Y 2018 & 2019 |

# Chapter 4
# Methodology

There were 10 data types along with the Avalanche Inventory Data that we gave as Input Data. The conditioning factors were used as model input and divided into two categories as numerical and categorical. The LULC data was considered a categorical feature and was handles through one hot encoding method of converting it to numerical.

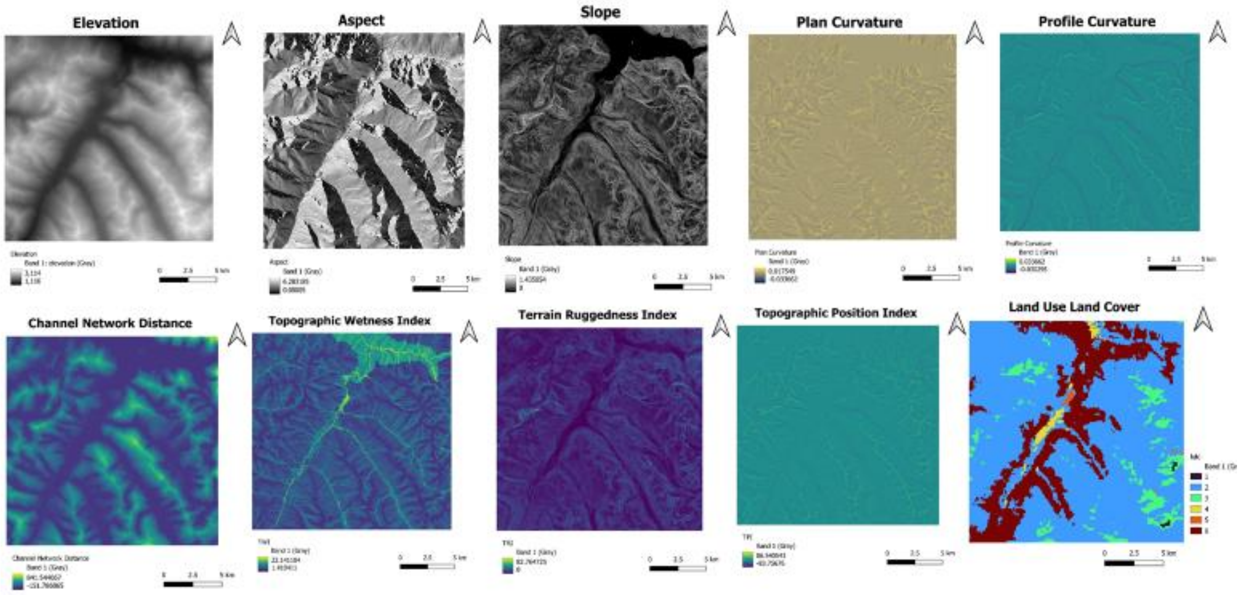
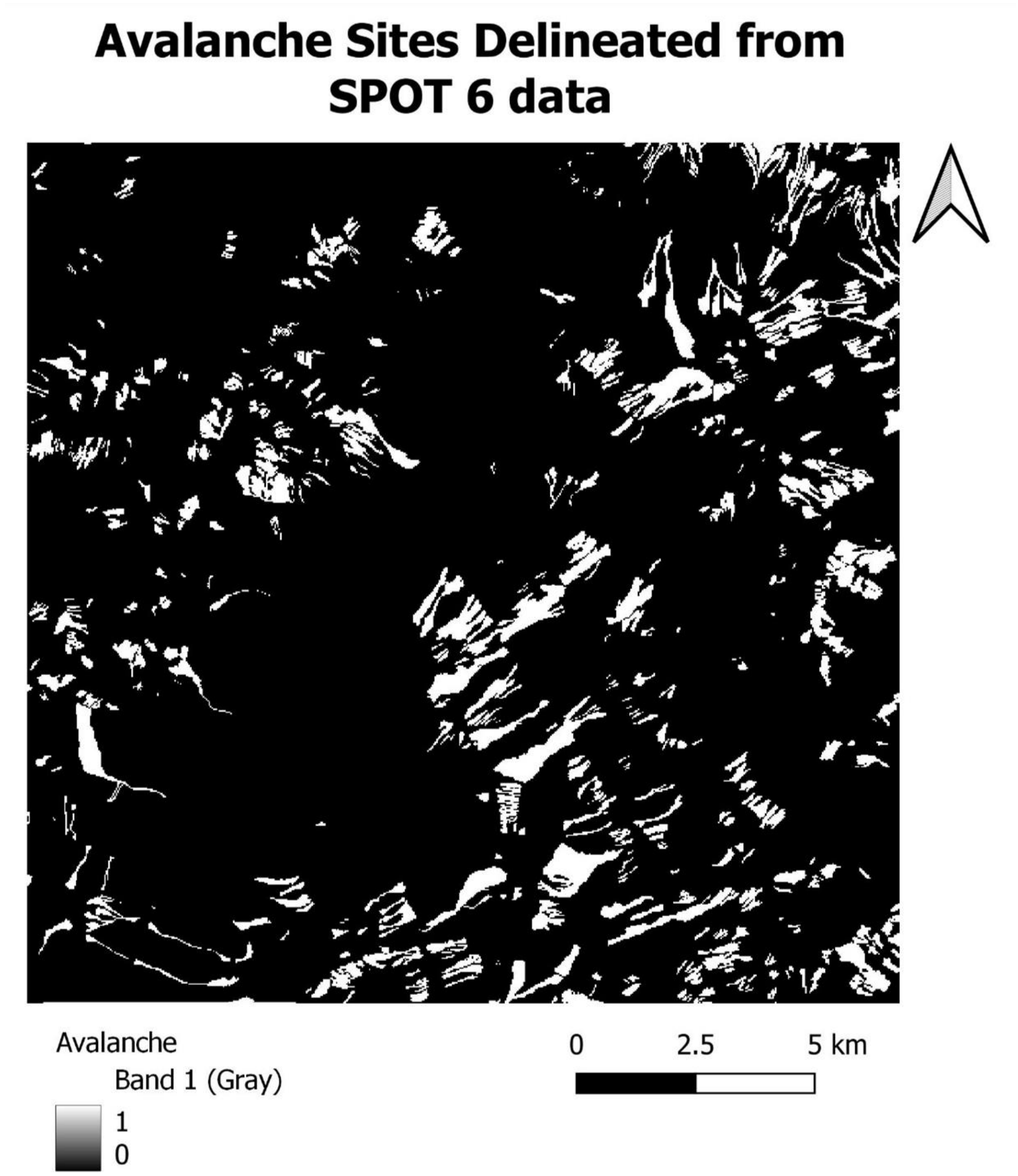



Fig 4.1: Input Features

The avalanche inventory map was rasterized prior to model training and the avalanche and non-avalanche pixels were labelled as True (1) and False (0), respectively.




The pixels employed as non-avalanche class in the training dataset were randomly selected with an equal number to avalanche class pixels (a ratio of 1:1 for avalanche: non-avalanche).

In the second step, the dataset was randomly split as training (70%) and validation (30%) datasets for avalanche prediction. Open source scikit learn library was used for processing the methods in a Python environment.

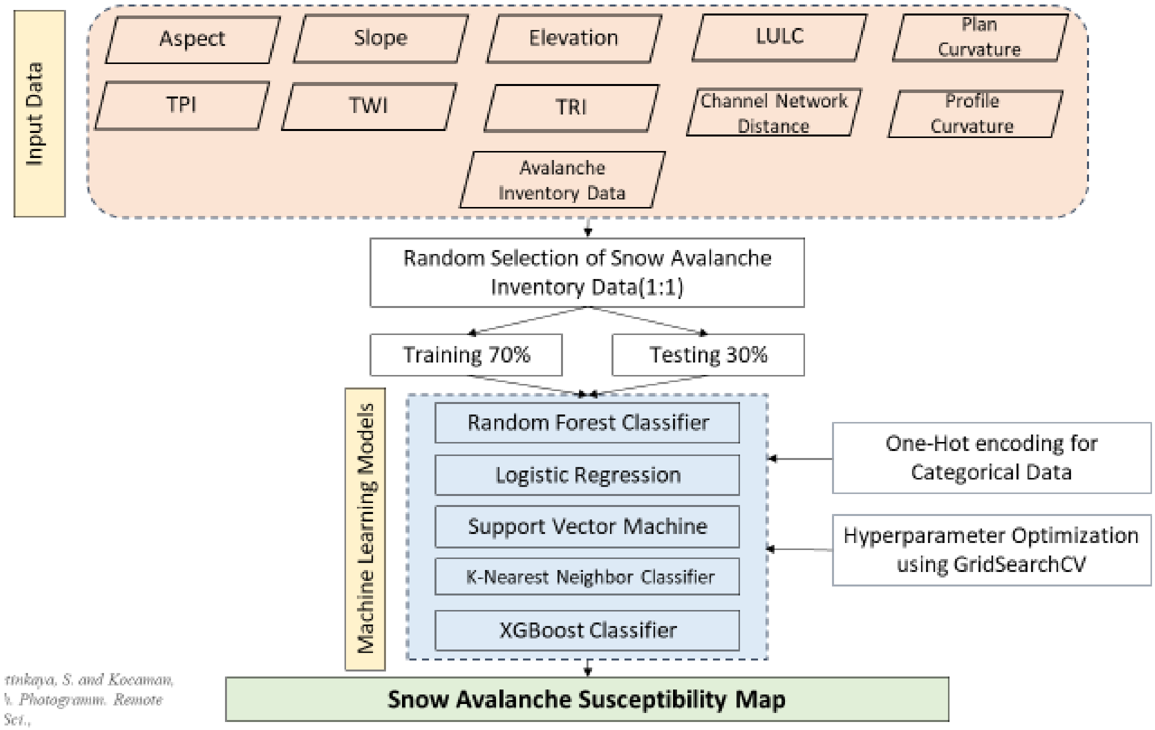



Fig 4.2: Methodology Flowchart

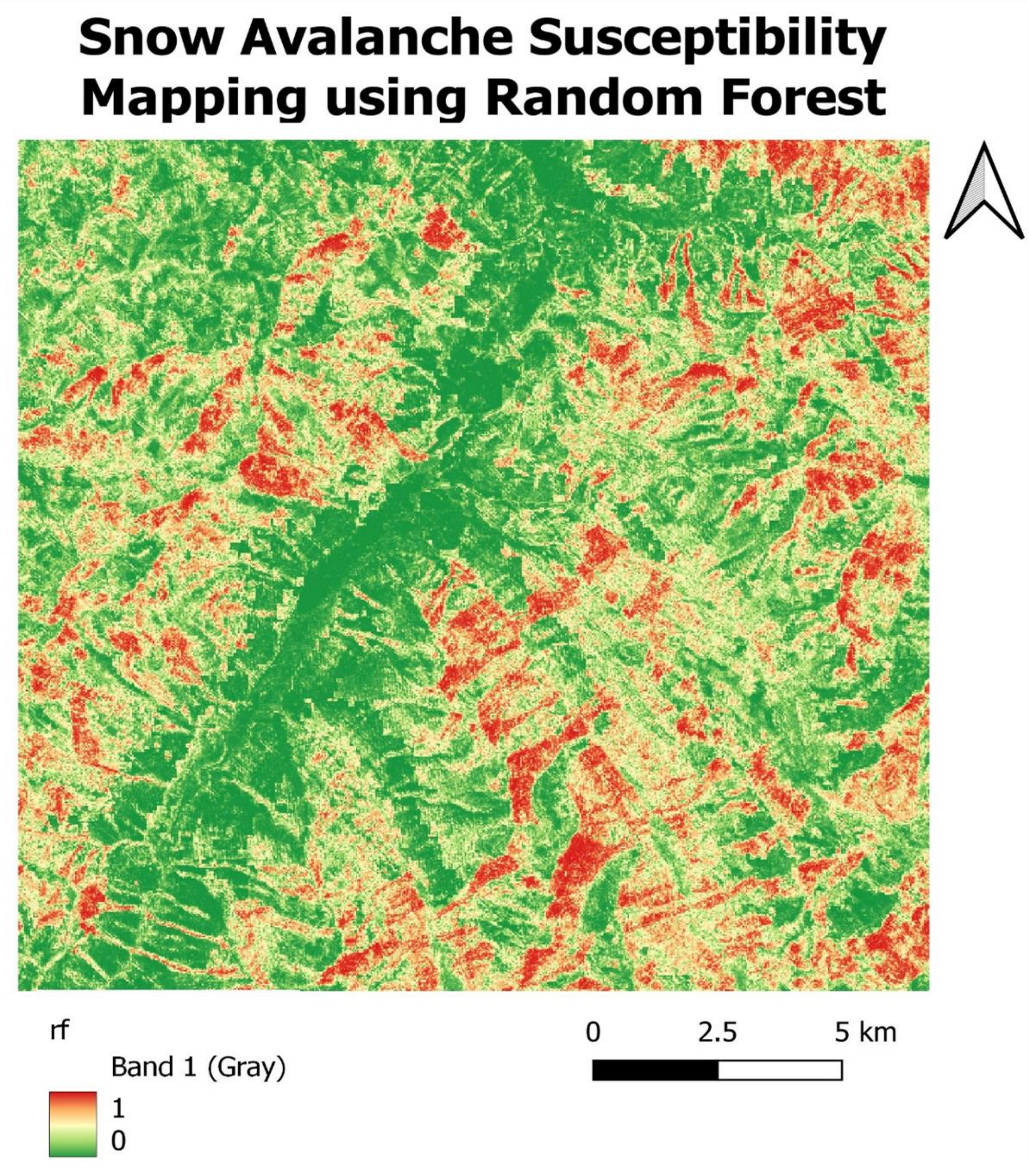
Five machine learning algorithms were implemented and tested. The performance of the algorithms are improved through Hyperparameter Optimization using GridSearchCV.

# Chapter 5

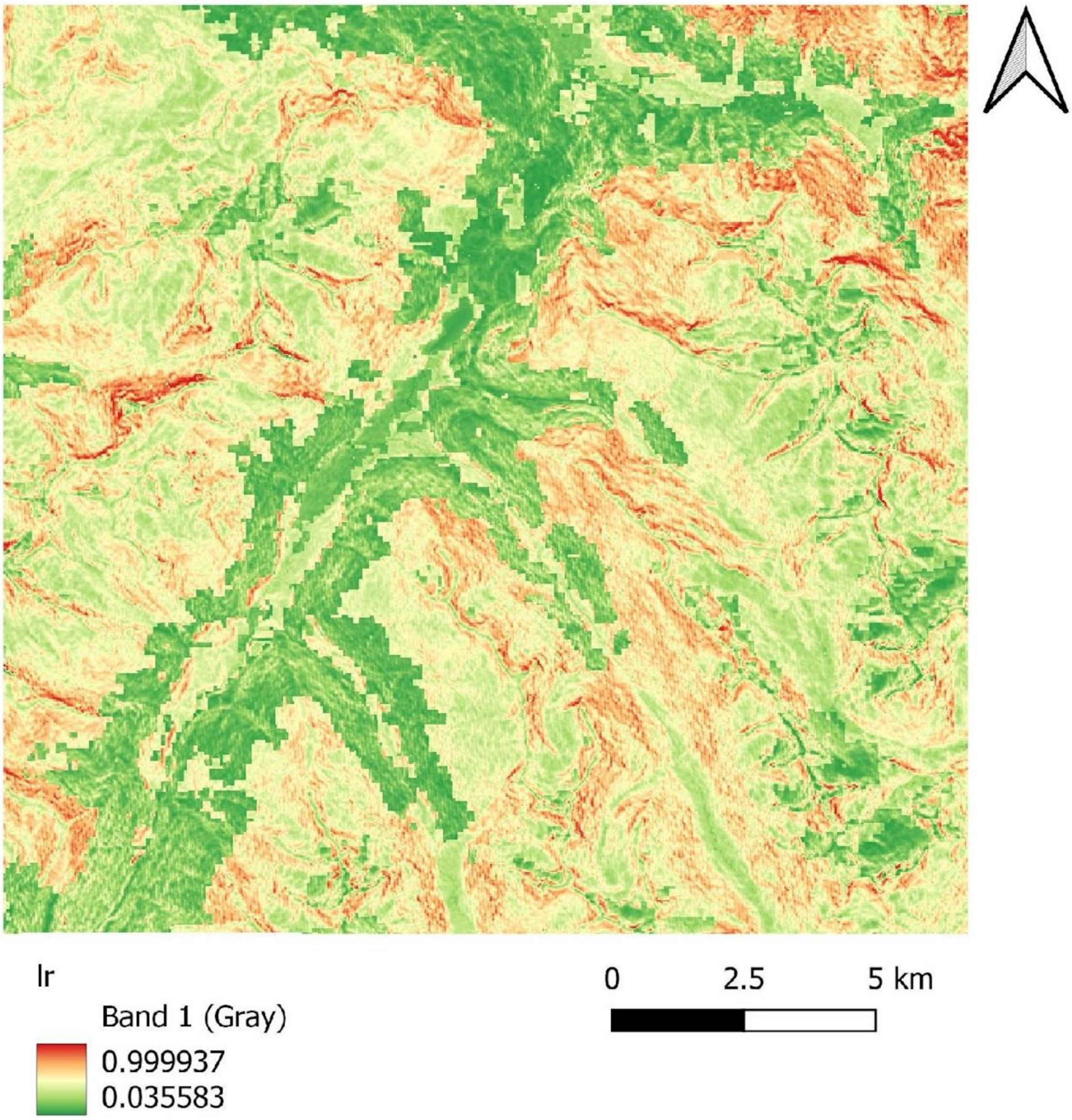
# Results and Discussion

## 5.1 Results and Discussion

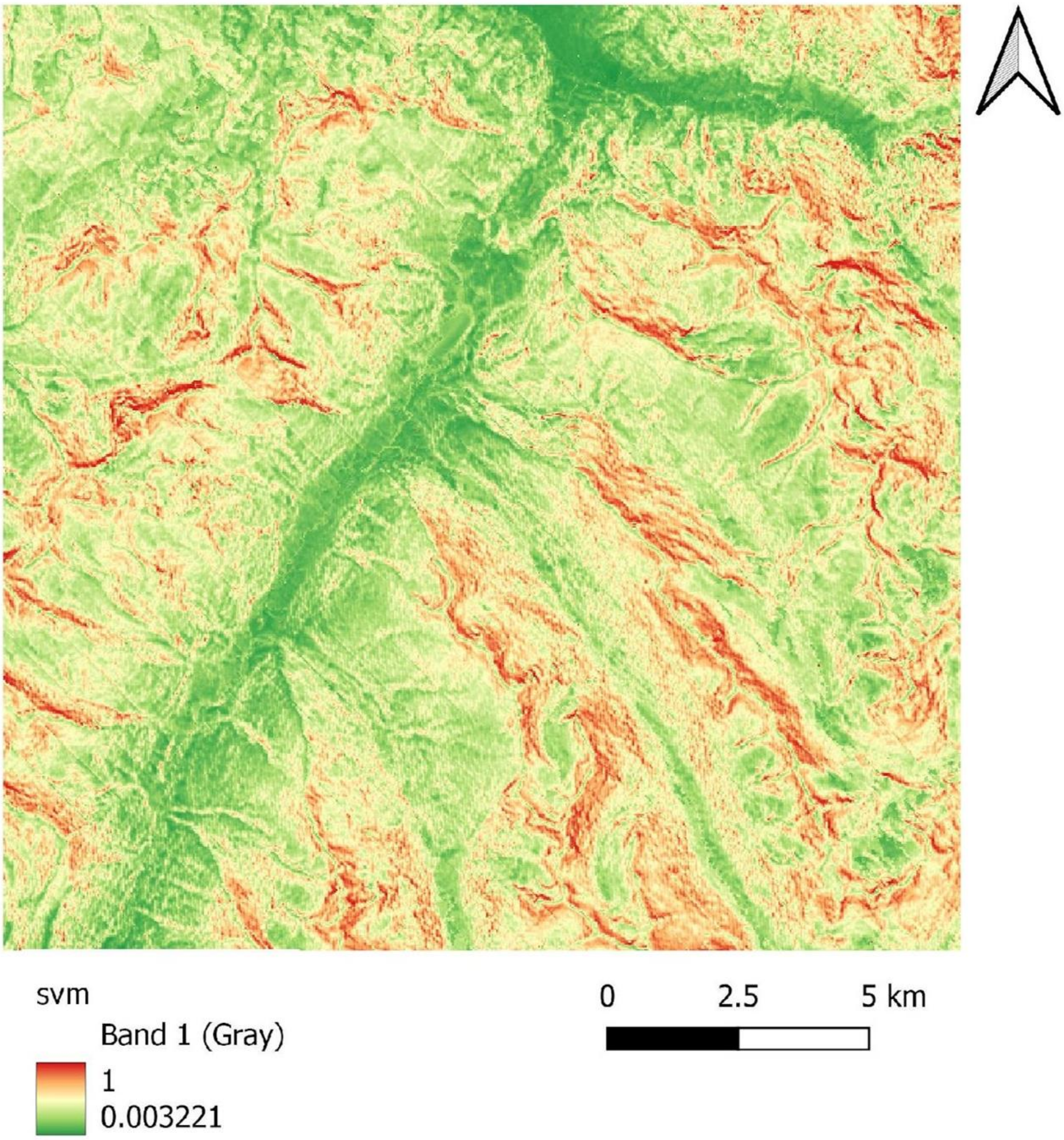
The five machine learning algorithms were trained and implemented to predict the avalanche susceptible zones in Davos region of Switzerland. The susceptibility maps are as follows,

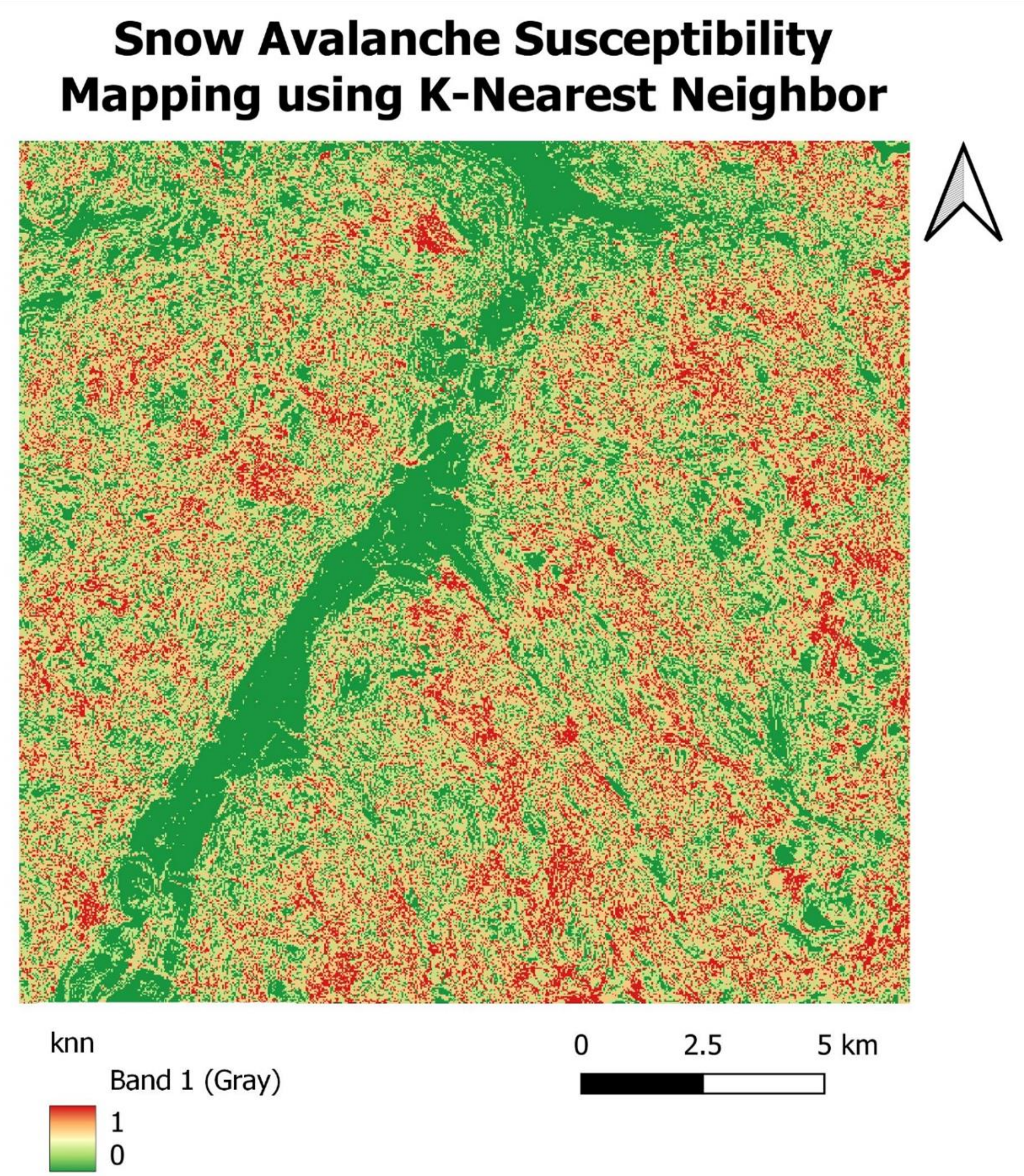


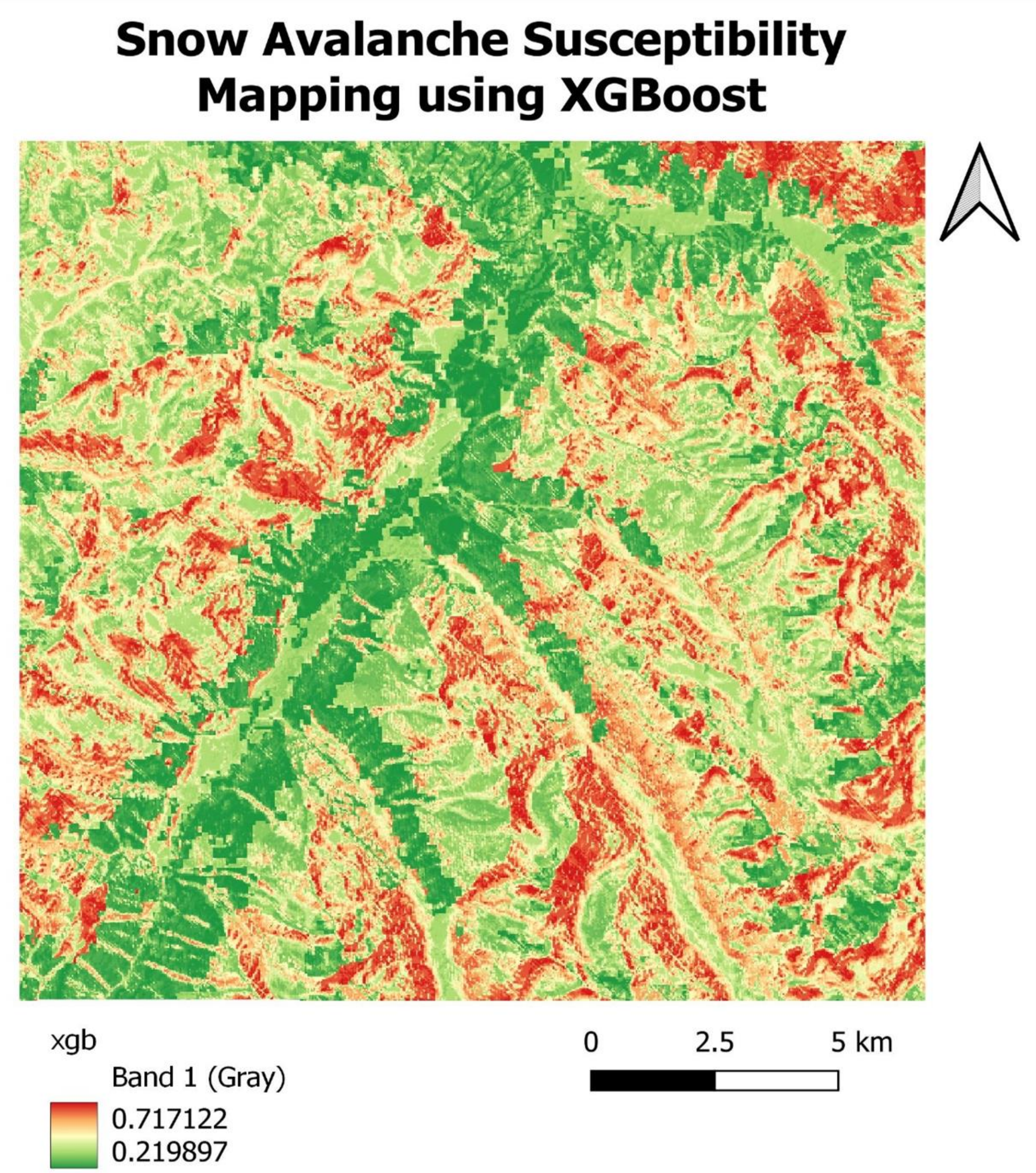
# Snow Avalanche Susceptibility Mapping using Linear Regression



# Snow Avalanche Susceptibility Mapping using SVM




Snow Avalanche Susceptibility
Mapping using K-Nearest Neighbor
knn
Band 1 (Gray)
1
0
0
2.5
5 km


Snow Avalanche Susceptibility
Mapping using XGBoost
xgb
Band 1 (Gray)
0.717122
0.219897
0
2.5
5 km

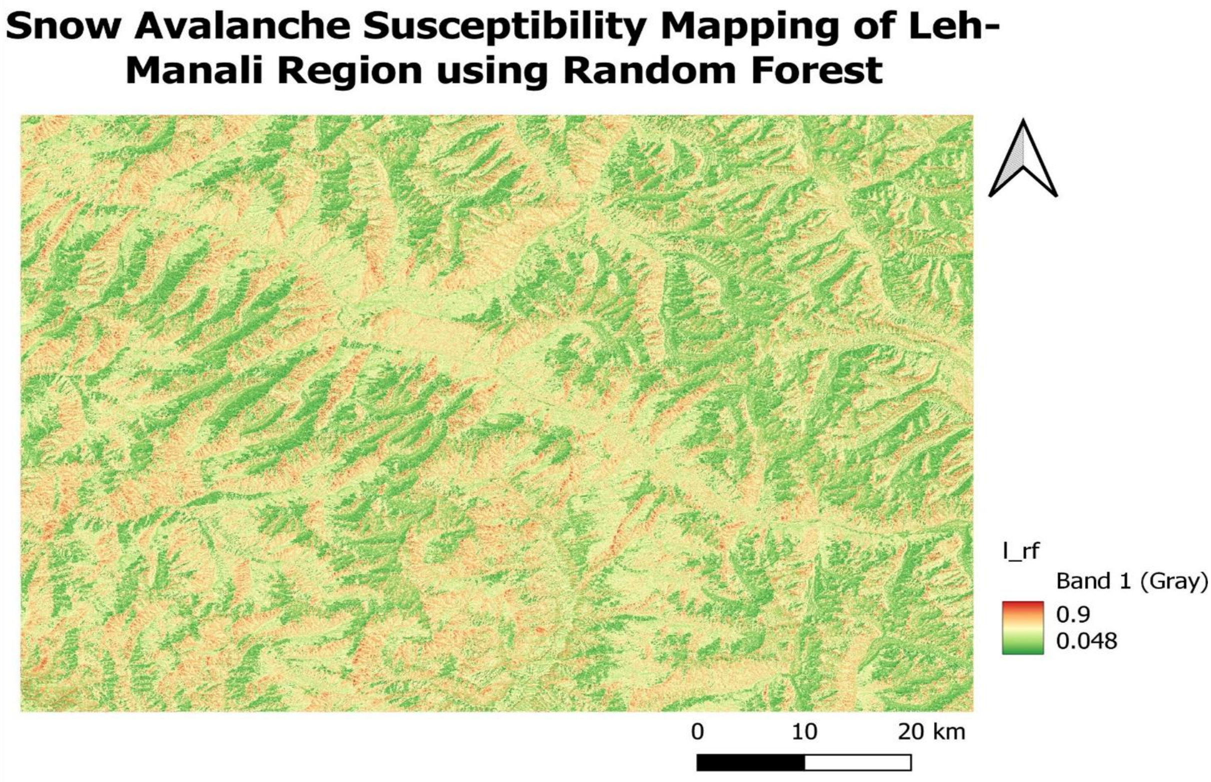
The accuracy metrics of the implemented algorithms are stated below,

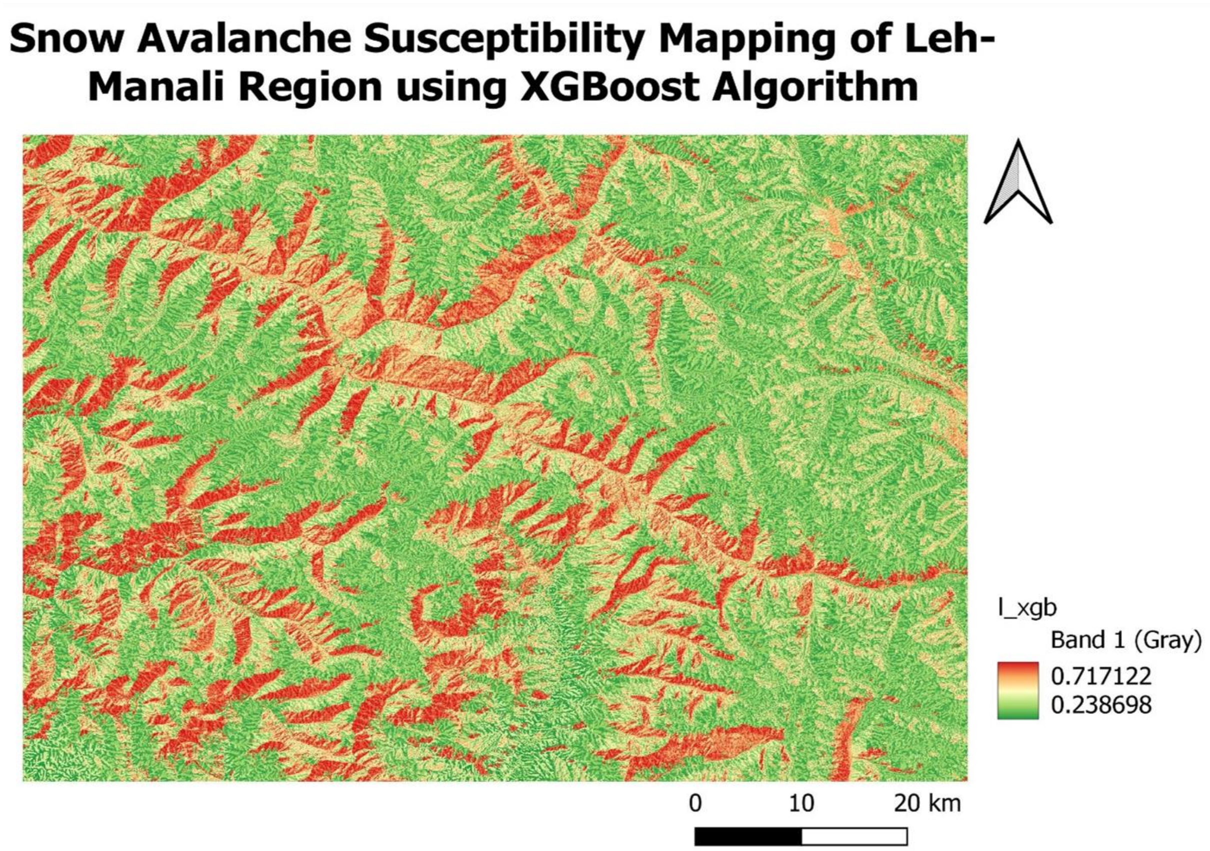
| Executed ML algorithms and their accuracy metrics | | | |
|---|---|---|---|
| **Sl.No** | **Machine Learning Models** | **Accuracy** | **ROC Score** |
| 1 | Random Forest Classifier | 0.788 | 0.787 |
| 2 | Logistic Regression | 0.671 | 0.671 |
| 3 | Support Vector Machine | 0.678 | 0.678 |
| 4 | K-Nearest Neighbor Classifier | 0.652 | 0.652 |
| 5 | XGBoost Classifier | 0.729 | 0.728 |

Of the five algorithms Random forest gave the best results and from visual interpretation also it was observed that it has identified the susceptible regions more precisely. XGBoost also gave good results. K-Nearest Neighbour Classifier comparatively did not give good results. The remaining two algorithms did not perform as good as random forest or XGBoost but was satisfactory.

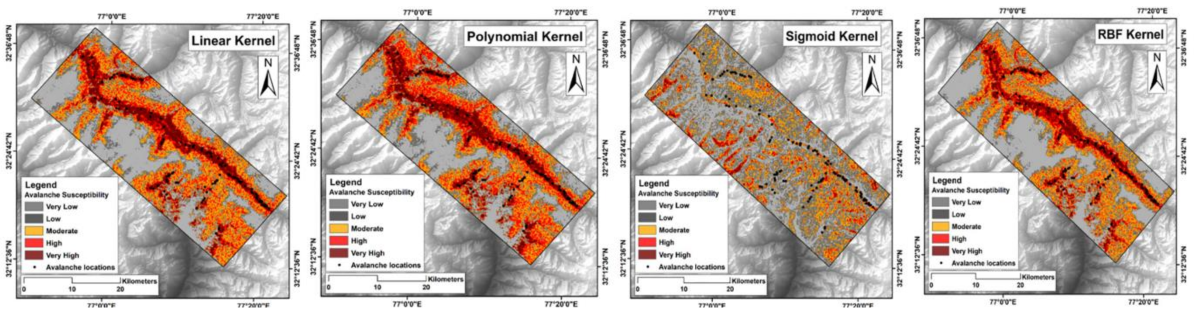
## Avalanche Susceptibility Mapping in the Indian Himalayan Region

The model trained for Switzerland region which performed well was implemented in the Indian Himalayan region due to lack of access to avalanche data in the Indian region. The results are,






For validation, avalanche susceptibility maps from other studies were used to visually check the results.




Avalanche Susceptibility map by Anuj Tiwari et al(2021), Sci. Total Environ.

## 5.2 Conclusion

A snow avalanche is a frequently observed natural hazard threatening lives and properties in mountainous and cold regions. The ASMs can be used as a basemap or initial data by researchers, designers, and decision-makers for regional land use planning, site selection, and avalanche prevention and mitigation purposes. In the present study, five ML models were employed for snow ASM with 10 conditioning factors in Davos, Switzerland. The training data (Hafner et al., 2021) was produced in a previous study for two avalanche periods and provided by the SLF.

Of the 5 ML algorithms used Random Forest and XGBoost gave results with better accuracy. The model trained and validated for Switzerland region was implemented for Himalayan region for which the results matched with other available susceptibility maps. Comparatively XGBoost algorithm gave better results in the Himalayan region. Performance of the algorithms can be improved by using training data from multiple study area or using site specific training data and by classifying the avalanche into run off zone and deposit zones. In addition to the static input variables used other dynamic variables can be included.