Article

A New Integrated Approach for Landslide Data Balancing and Spatial Prediction Based on Generative Adversarial Networks (GAN)

Husam A. H. Al-Najjar [1], Biswajeet Pradhan [1,2,*], Raju Sarkar [3], Ghassan Beydoun [1] and Abdullah Alamri [4]

[1] Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and IT, University of Technology Sydney, Sydney, NSW 2007, Australia; Husam.AL-NAJJAR@student.uts.edu.au (H.A.H.A.-N.); Ghassan.Beydoun@uts.edu.au (G.B.)

[2] Earth Observation Centre, Institute of Climate Change, Universiti Kebangsaan Malaysia, UKM, Bangi, Selangor 43600, Malaysia

[3] Department of Civil Engineering, Delhi Technological University, Bawana Road, Delhi 110 042, India; rajusarkar@dce.ac.in

[4] Department of Geology and Geophysics, College of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia; amsamri@ksu.edu.sa

* Correspondence: Biswajeet.Pradhan@uts.edu.au or biswajeet24@gmail.com

Abstract: Landslide susceptibility mapping has significantly progressed with improvements in machine learning techniques. However, the inventory/data imbalance (DI) problem remains one of the challenges in this domain. This problem exists as a good quality landslide inventory map, including a complete record of historical data, is difficult or expensive to collect. As such, this can considerably affect one's ability to obtain a sufficient inventory or representative samples. This research developed a new approach based on generative adversarial networks (GAN) to correct imbalanced landslide datasets. The proposed method was tested at Chukha Dzongkhag, Bhutan, one of the most frequent landslide prone areas in the Himalayan region. The proposed approach was then compared with the standard methods such as the synthetic minority oversampling technique (SMOTE), dense imbalanced sampling, and sparse sampling (i.e., producing non-landslide samples as many as landslide samples). The comparisons were based on five machine learning models, including artificial neural networks (ANN), random forests (RF), decision trees (DT), k-nearest neighbours (kNN), and the support vector machine (SVM). The model evaluation was carried out based on overall accuracy (OA), Kappa Index, F1-score, and area under receiver operating characteristic curves (AUROC). The spatial database was established with a total of 269 landslides and 10 conditioning factors, including altitude, slope, aspect, total curvature, slope length, lithology, distance from the road, distance from the stream, topographic wetness index (TWI), and sediment transport index (STI). The findings of this study have shown that both GAN and SMOTE data balancing approaches have helped to improve the accuracy of machine learning models. According to AUROC, the GAN method was able to boost the models by reaching the maximum accuracy of ANN (0.918), RF (0.933), DT (0.927), kNN (0.878), and SVM (0.907) when default parameters used. With the optimum parameters, all models performed best with GAN at their highest accuracy of ANN (0.927), RF (0.943), DT (0.923) and kNN (0.889), except SVM obtained the highest accuracy of (0.906) with SMOTE. Our finding suggests that RF balanced with GAN can provide the most reasonable criterion for landslide prediction. This research indicates that landslide data balancing may substantially affect the predictive capabilities of machine learning models. Therefore, the issue of DI in the spatial prediction of landslides should not be ignored. Future studies could explore other generative models for landslide data balancing. By using state-of-the-art GAN, the proposed model can be considered in the areas where the data are limited or imbalanced.

Keywords: landslide susceptibility; imbalanced dataset; machine learning; generative adversarial network; GIS; remote sensing; Bhutan

1. Introduction

Landslides are a form of natural hazard that pose significant threats to the environment and society [1]. In many parts of the world, landslides frequently occur due to single, high-intensity (e.g., shallow, fast-moving landslides) or prolonged (days or weeks, e.g., slow-moving deep-seated landslides) rainfall events, earthquakes, or human activity [2]. Landslides also occur due to other mechanisms, such as volcanic eruptions, rapid snowmelt, and elevated water levels [3]. Landslides occur in mountainous areas at high frequencies relative to areas with low terrain. The higher the angle of inclination, the more dominant the gravity, which leading to "pulling" material down the slope. Landslides begin to occur when the resisting force exceeds the certain limit depending upon the strength of the material, the frictional properties between the slide material and the rock, or both.

The Himalayan region suffers excessive precipitations which contribute more than 70% of the world's disastrous landslides [4]. Himalayan landslides are often caused by seasonal rains, spawning substantial economic losses and occasionally human lives [5]. A compilation of a fatal landslide database by Froude and Petley [4] recorded 64 fatalities in Bhutan from 2004 to 2017. Still, the real figure of casualties is estimated to be higher. The threat is projected to accelerate as hills are cut and deforested, emphasizing infrastructure development linked to population growth [6].

The spatial prediction of landslides is an essential step in the preparation of risk maps for landslide [7]. In this step, statistical or machine learning models are developed and assessed, taking into consideration a number of landslide conditioning factors (slope, lithology, land use, distance to road, etc.), as well as landslide inventories that generate a spatial probability map of landslide events [8].

Landslide spatial prediction studies have used several types of statistical and machine learning models. The studies are varied and focused on several aspects, such as optimising conditioning factors [9,10], factor normalization methods [11], sampling strategies [12–14] generating training and test samples methods [15], model ensembling [16,17], fine-tuning model's hyperparameters [18], modelling with limited data [7], data imbalance (DI) [19,20], and model assessment [21].

The current study focuses on the DI problem in the spatial prediction of landslides [19]. DI is one of the understudied problems in the spatial prediction of landslides. Landslide dataset contains two classes (positive class or landslide presence and negative class or landslide absence). In most cases, the negative class dominates or overpowers the positive class. In other words, the distribution of observations in the negative and positive classes is unbalanced. The problem is that most typical machine learning models assume/expect balanced class distributions and suffer DI [22]. This leads to the classifier being more biased towards the dominating class; so the algorithms are not able to classify the landslide pixels properly. Thus, it is essential to decrease the imbalance effect in the landslide dataset. The two main approaches for reducing imbalance are oversampling of a minority class and under-sampling of the majority class [23]. Various methods utilised by several scientists to control DI include random oversampling and undersampling [24,25], informed under-sampling [20], and synthetic sampling with data generation [23,26–28].

This study aims to develop and test a new method to correct the class imbalance in landslide datasets using generative adversarial networks (GAN) [29]. The method is based on oversampling landslide presence class with data generation. The method was then compared to other standard methods such as synthetic minority oversampling technique (SMOTE) [30] and tested with several machine learning models, including neural networks (ANN), random forest (RF), decision trees (DT), k-nearest neighbour (kNN), and support vector machine (SVM).

2. Previous Works

DI is a common problem in machine learning with real-world datasets, and spatial prediction of landslides is no exception. Landslide databases are often imbalanced as landslides are described in a network/grid raster spatial data and comprised of a small

number of pixels [31]. This problem affects the prediction performance of the machine learning models by making the model biased towards the dominating class. It can be overcome by dataset balancing [32].

In the oversampling of the minority class, the number of landslide records is artificially increased through sample repetition or data generation [23]. However, in under-sampling the majority class method, it aims to reduce the samples of the majority class by randomly (or synthetically) sampling a smaller subset from the original dataset [8]. Random over-sampling works by selecting a minority class instance randomly and replicating it until the desired balance between classes is reached [26]. Beyond random sampling, synthesised oversampling generates new artificial minority class examples by interpolating among several existing minority class examples that are similar to each other.

Several researchers have investigated the impacts of DI on machine learning models' performance for the spatial prediction of landslides [20,33]. The most common and usual practice is randomly under-sampling of non-landslide data in the training samples. By doing so, the landslide samples will have approximately the same order of magnitude as the non-landslide samples. This method leads to poor prediction ability of the model since the non-landslide sample data are often wasted [34]. Agrawal et al. [33] compared random oversampling and synthesised oversampling (SMOTE, SMOTE-IPF) for landslide prediction. Based on an analysis performed using different machine learning models, such as logistic regression (LR), DT, RF, SVM, and neural networks (NN), they suggested that the synthesised oversampling performed better than that of random oversampling. Braun et al. [35] studied the impacts of data balancing on the performance of machine learning models for spatial prediction of landslides. Their results suggested that data balancing (i.e., by increasing the data density of the positive class) had almost no effect on the overall performance of the DT model using the training data. However, the performance of the model was affected when the test data was used. This indicates that the model had a poor ability to generalise the learned parameters for unknown areas. A hybrid method for spatial prediction of landslides using a solitary landslide-occurring data was proposed by Mutlu and Goz [36]. They used several performance metrics of imbalanced data to evaluate their proposed solution and showed that clustering significantly improved the performance of the models. This was interpreted as landslides occurring in large areas may have different geographical and morphological properties. In another research, Gupta et al. [20] employed two methods for balancing landslide data, EasyEnsemble and BalanceCascade. They showed that models such as NN had significantly been impacted by data balancing. However, the LR model was ill-affected by data balancing and the fisher discriminant analysis (FDA) did not show considerable influence between balanced and imbalanced data. Zhang and Yu [37] also showed that combining adaptive synthetic (ADASYN) sample method and linear discriminant analysis (LDA) for seismic spatial prediction of landslides helps to improve the modelling results affected by the sample imbalance. Other studies have addressed the DI problem at the algorithm level. For example, Song et al. [38] considered landslides' spatial prediction as a DI learning problem. They used an approach that solves the DI problem at the algorithm level using a cost matrix designed to make the negative samples' misclassification cost more considerable than that of the positive samples.

The studies above showed that the DI has a great effect on machine learning models' prediction ability for spatial prediction of landslides. Most of the studies have shown that synthetic sampling (e.g., SMOTE) outperforms the random methods. However, some models gain fewer benefits from data balancing techniques [20]. To the best of the authors' knowledge, generative techniques such as GAN have not been employed in this particular domain. As a result, this paper introduces a new approach to data balancing based on GAN. We think that the new approach would be able to outperform the other techniques due to GAN models' ability to learn abstract representations of the dataset and generate more useful samples compared to random or synthesised samples as other methods generate.

This paper compares SMOTE and GAN for landslide data balancing based on several common/standard machine learning models, such as ANN, RF, DT, kNN, and SVM.

3. Materials and Methods

3.1. Study Area

The case study of this research is Chukha Dzongkhag area (longitudes 89°15′–89°49′ and latitudes 26°44′–27°18′) located in the southwest of Bhutan (Figure 1). The area approximately covers 1879.5 km^2. The area contains steep slope terrain, which makes it highly at risk of slope failures brought by rainfall and associated disasters due to several road cuttings [39].
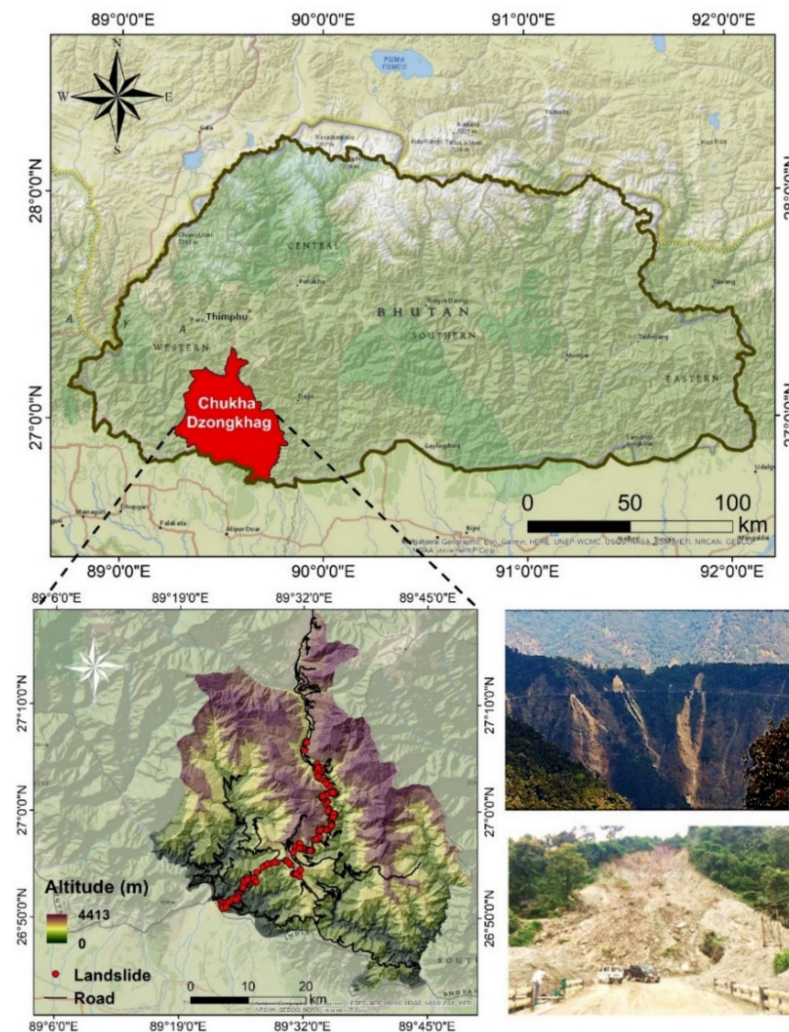
Figure 1. Location of the study area (Chukha Dzongkhag, Bhutan).

Chukha Dzongkhag is located in a region with a subtropical and temperate climate. The area is highly prone to rainfall-induced landslides, exposed to an annual high level of rainfall (maximum of 4000–6000 mm), typically by 80% throughout the southwestern monsoon (from June to September). It records up to 800 mm·day−1 as the most frequent heavy rainfall in the region. Landslides in the study area mostly occur due to intense monsoon rainfall, weak geology, and human-induced activities. The majority of landslides occur along the Phuentsholing–Thimphu highway, a lifeline link that connects the capital Thimphu with neighbouring country India, by road as shown in Figure 1 [40,41]. Landslides often block the roads, thereby causing substantial damage to the economy and infrastructures. The geology of the area is characterised by Himalayan Crystalline Com-

posite [42]. The most disposed zones to slope failures are deeply cracked and weathered rocks of phyllite, slate, and schist that comprise a large quantity of clay [41,43,44].

3.2. Datasets

3.2.1. Geodatabase of Conditioning Factors

In this work, ten landslide conditioning factors were used based on previous works suggestions [8–10] (Figure 2). These include seven geo-morphological factors, such as altitude, slope, aspect, total curvature, slope length, topographic wetness index (TWI), and sediment transport index (STI). It also included three geo-environmental factors such as lithology, distance from the road, and distance from the stream. The details of these factors, including the calculation method, rationale, and the data range within the study area, are provided in Table 1.

Table 1. List of landslide conditioning factors used in this research.

Factor	Calculation Method	Rationale	Data Range (Study Area)
Altitude	Extracted from digital elevation model (DEM) data with 10 m spatial resolution (DEM data from ALOS PALSAR satellite) prepared based on topographical maps.	Areas with high altitude impact loading on the slope, increasing the possibility of landslides.	0–4413 (m)
Slope	$\arctan^{0.5}\left(f_x^2 + f_y^2\right),$ $f_x = \frac{z_3 - z_5 + 2(z_2 - z_6) + z_1 - z_7}{8 \times \Delta x},$ $f_y = \frac{z_7 - z_5 + 2(z_8 - z_4) + z_1 - z_3}{8 \times \Delta y}$	Slope is commonly used in landslide studies. It is an important topographical factor that positively contributes to landslides. In other words, landslides frequency is often high on steep slopes.	0–89 (Degree)
Aspect	$180 - \arctan\left(\frac{f_y}{f_x}\right) + 90\left(\frac{f_x}{|f_x|}\right)$	Aspect is an important factor for landslide studies as it affects daylight, wind, and precipitation exposure. As a result, it also impacts vegetation cover, soil thickness, and water content in the soil.	−1–360 (Degree)
Total curvature	Calculated using the formula for the calculation of total curvature as [45].	Curvature has a critical role in altering the characteristics of landforms. Thus, it is an important factor for landslide modelling. The convex surface instantly drains wetness, while for an extended duration, the concave surface keeps humidity.	−10416–0 (Convex) 0 (Flat) 0–3720 (Concave)
Slope length	Based on the DEM raster, for each cell, the upstream or downstream distance, or weighted distance, along the flow path was calculated.	Slope length is a factor that contributes to the increase of erosive capacity to displace and transport materials downslope.	0–1220 (m)
Lithology	The various lithological units were extracted and mapped from the geology map of the study area based on the Bhutan geological map revised after [46].	Lithological units have different impacts on landslides. For example, weathered and fine rocks are more exposed to slope failure than strong separate/disjointed ones.	[Baxa Formation, Jaishidanda, Tethyan Sedimentary Series, Paro Metasediments, Chhukha, Daling Shumar Gp, Greater Himalayan Series]

Table 1. *Cont.*

| Factor	Calculation Method	Rationale	Data Range (Study Area)
Distance from road	The Euclidean distance was calculated from each pixel within the raster data to the nearest source point (road data).	Shallow to deep excavations, external loads application, and vegetative cover removal are general activities throughout construction alongside main roads/highways.	0–17,377 (m)
Distance from stream	The Euclidean distance was calculated from each pixel within the raster data to the nearest source point (stream data).	The irregular movement system of a hydrological circle and valleys involves eroding and saturating progressions. Consequently, causing slope failure due to a surge in pore water pressure level in zones that neighboring drainage networks.	0–3356 (m)
TWI	$\ln\left(\frac{\alpha}{tan\beta}\right)$	A hydrology factor that combines upslope contributing area and slope which impacts on landslide occurrence. Higher TWI values indicate a low probability of landslide occurrence.	−14–19.5
STI	$\left(\frac{A_s}{22.13}\right)^{0.6} \cdot \left(\frac{sin\beta}{0.0896}\right)^{1.3}$	The amount of sediment transportation through the on-land streams is mainly based on the eroding of catchment evolution concepts and carrying ability limiting deposit flux.	0–2399
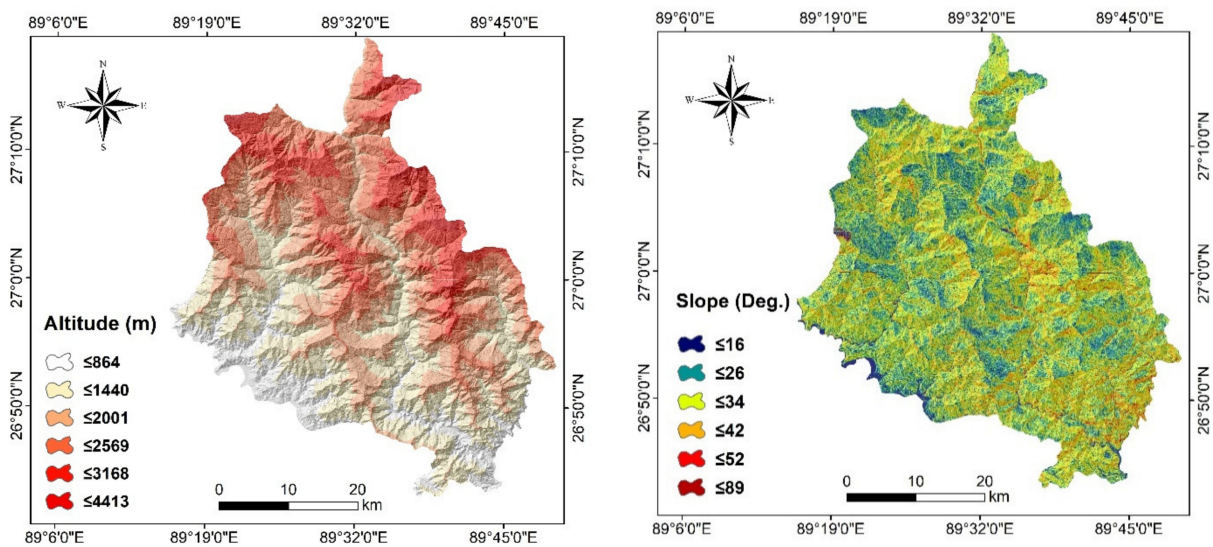
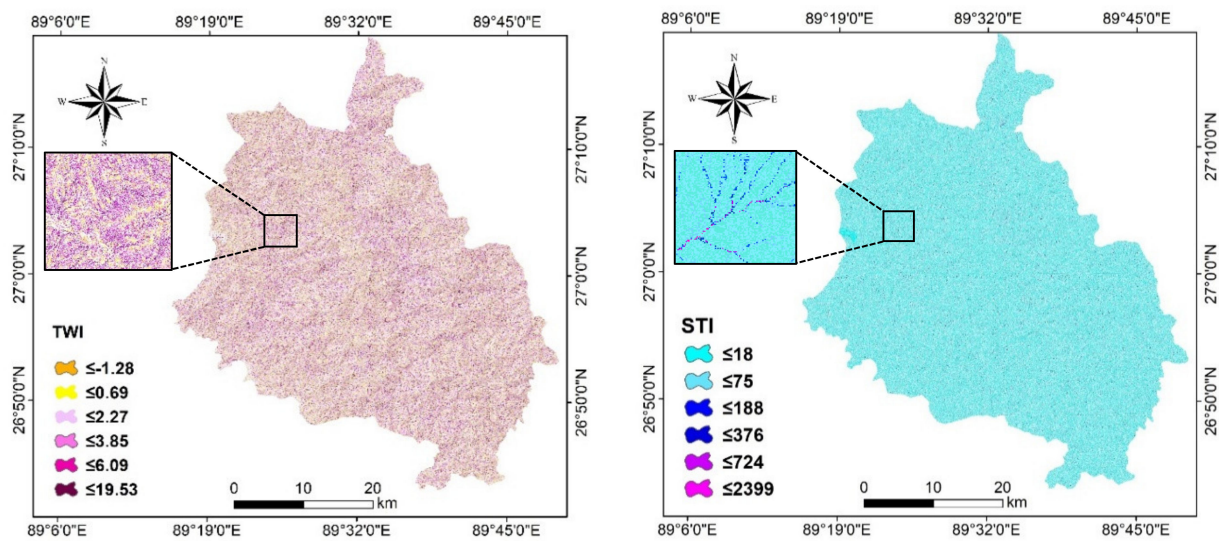**Figure 2.** *Cont.*
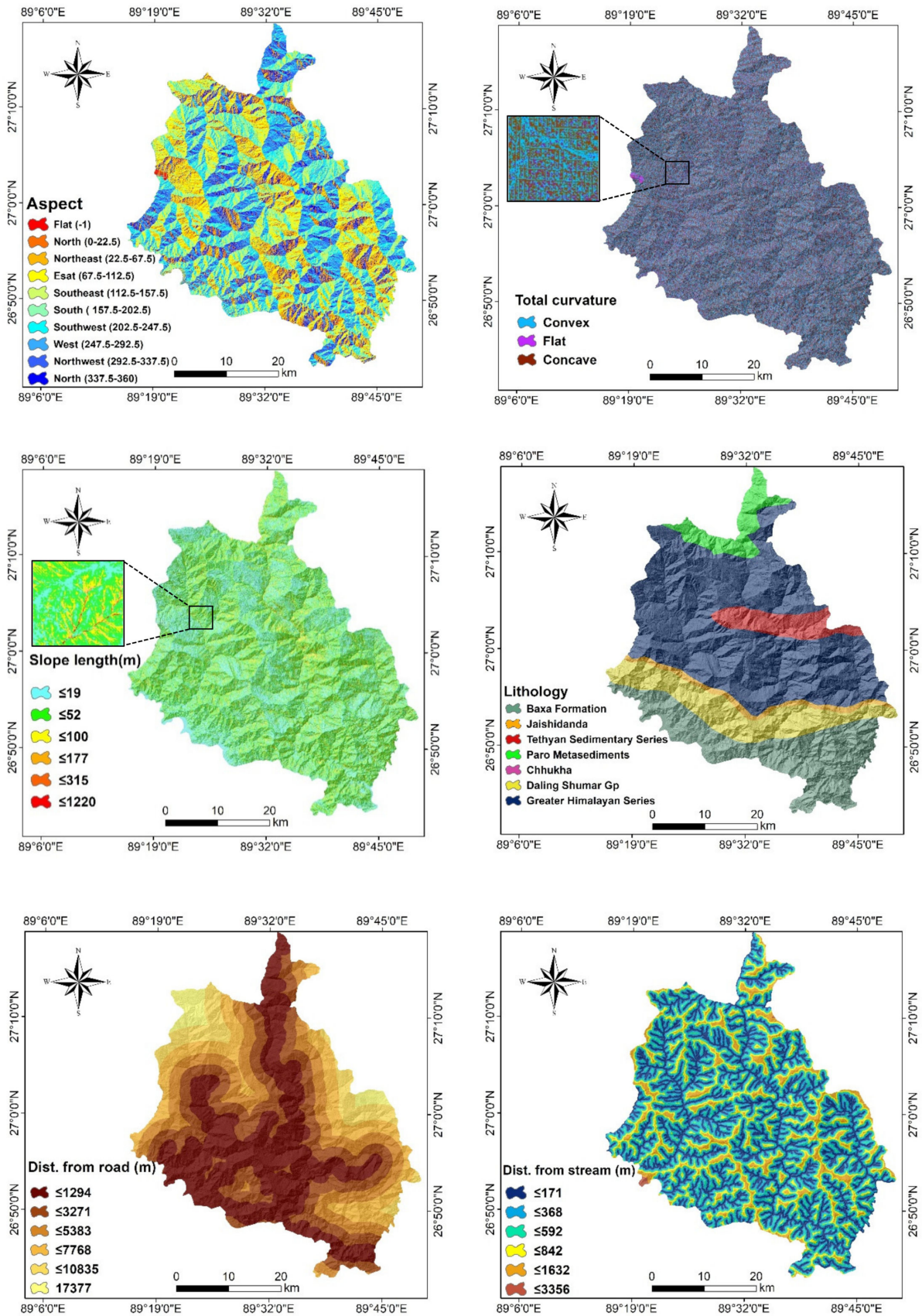
Figure 2. *Cont.*

Figure 2. Landslide conditioning factors.

3.2.2. Landslide Inventory Dataset

Data-driven models always require landslide inventories for training. As a result, the first step in the spatial prediction of landslides with such models is to prepare a landslide inventory database for the investigated area. The database should contain the location and landslide attributes (i.e., type, size, impact) of past events. Nevertheless, in landslide studies, usually obtaining a complete dataset is difficult and expensive. In this research, the database, including 269 landslide inventories, was collected based on the Border Roads Organisation, Govt. of India (http://www.bro.gov.in) project DANTAK framework. The model was constructed as a point-base landslide locations, which is commonly used in machine learning landslide modeling [7,15,16]. We utilised the spatial and location of temporal landslides that took place in a period from 1998 to 2015 in the study area. The landslide inventory was prepared by the Indian Border Roads Organisation, and the most common types of landslides observed were earth slides (shallow landslides). The landslides were mapped as single points.
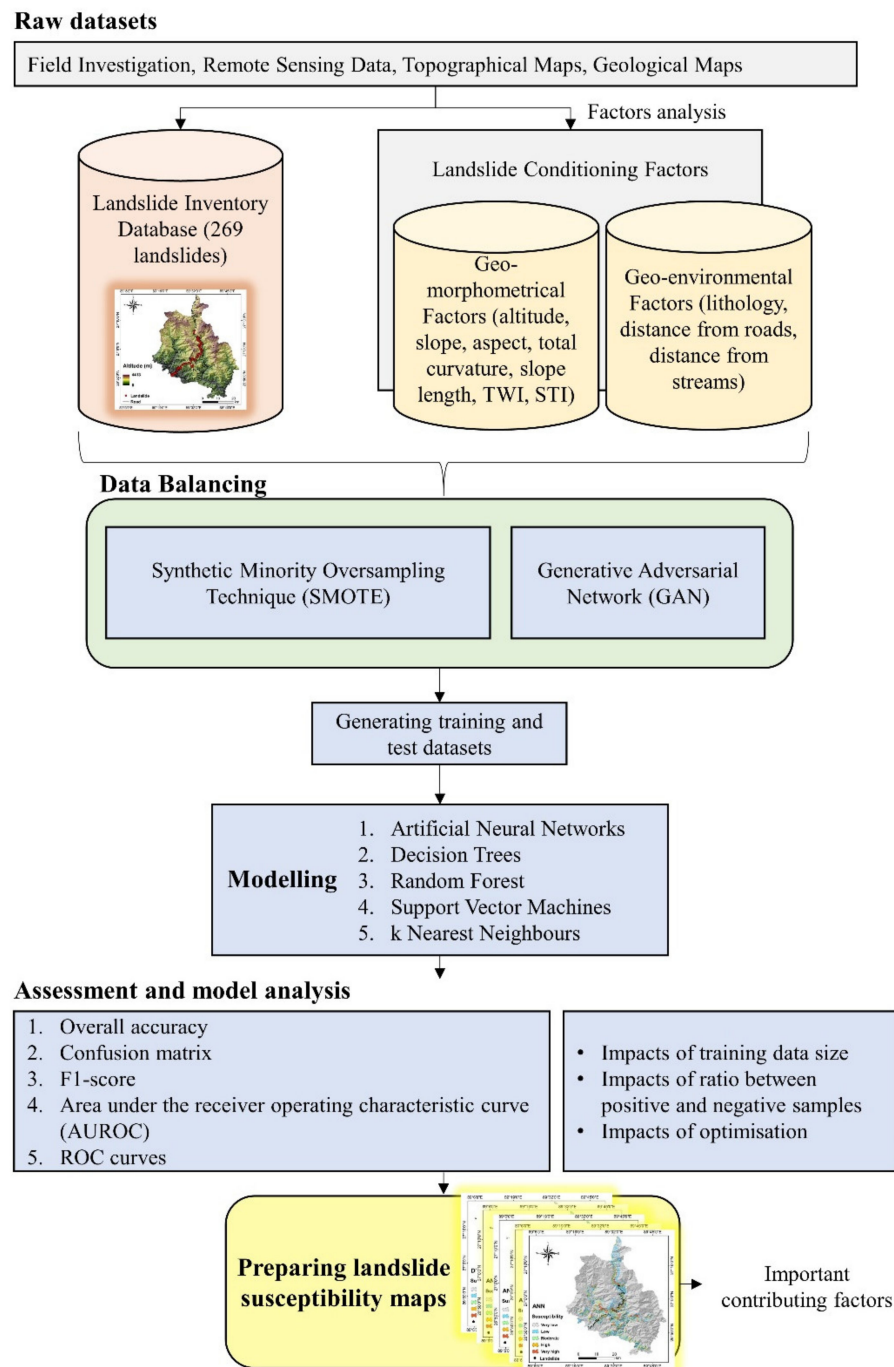
As shown in Figure 1, most of the landslides occurred along the Phuentsholing–Thimphu highway caused by intense or prolonged precipitation. Field data shows that the depths of the landslides vary from several decimetres to a few meters.

3.3. Methodology
3.3.1. Overview

Figure 3 shows the overview of the proposed methodology for spatial prediction of landslides at Bhutan, based on the GAN technique to address the imbalanced landslide inventories problem. After acquiring the inventories, the geospatial databases were prepared for the landslide inventories and landslide conditioning factors. As the non-landslide pixels are more than that of landslide pixels (~3.54%), two data balancing were applied to correct the DI, namely SMOTE and GAN. The corrected/updated dataset was then used to generate the training and test samples according to the (30/70) ratio method used commonly in previous works. Next, five different machine learning models (namely ANN, RF, DT, kNN and SVM) were constructed and trained. Those models were then tested and analysed using performance metrics, such as overall accuracy (OA), confusion matrix, F1-score, and area under receiver operating characteristic curves (AUROC) curves. The model analysis included the impacts of training data size, the ratio between the positive and negative samples, and the model's optimisation on the predictive capability of the models. Once the models were tested and verified, the landslide susceptibility maps of the study area were generated and prepared.

Raw datasets

Field Investigation, Remote Sensing Data, Topographical Maps, Geological Maps

Factors analysis

Landslide Inventory Database (269 landslides)

Landslide Conditioning Factors

Geo-morphometrical Factors (altitude, slope, aspect, total curvature, slope length, TWI, STI)

Geo-environmental Factors (lithology, distance from roads, distance from streams)

Data Balancing

Synthetic Minority Oversampling Technique (SMOTE)

Generative Adversarial Network (GAN)

Generating training and test datasets

Modelling
1. Artificial Neural Networks
2. Decision Trees
3. Random Forest
4. Support Vector Machines
5. k Nearest Neighbours

Assessment and model analysis

1. Overall accuracy
2. Confusion matrix
3. F1-score
4. Area under the receiver operating characteristic curve (AUROC)
5. ROC curves

- Impacts of training data size
- Impacts of ratio between positive and negative samples
- Impacts of optimisation

Preparing landslide susceptibility maps

Important contributing factors

Figure 3. Overview of the proposed methodology for spatial prediction of landslides in Bhutan.

3.3.2. Data Balancing

a.   Synthetic Minority Oversampling Technique (SMOTE)

SMOTE is an approach to correct DI problems [30]. It is an oversampling technique that works based on the interpolation process as the following. First, the k nearest neighbours of all minority samples are determined. Next, artificial minority samples are generated along the lines (i.e., the line lay between the minority samples and their kNN until the dataset reaching a balance. SMOTE is considered a pre-processing step within a modelling pipeline. It allows the application of common machine learning models after correcting DI. However, SMOTE cannot handle noise adequately as the introduced samples might be the result of interpolation between noisy samples. Thus, improvements are required to

clean the data after applying SMOTE. This approach and its extensions have been used for spatial prediction of landslides in studies by Agrawal et al. [33].

b.   Generative Adversarial Network (GAN)

As first presented by Goodfellow et al. [29], GAN is a NN model; the concept of GAN's training is based on adversarial form to generate new data simulating particular distributions. It consists of two main components: a generator (G) and a discriminator (D). The G maps a sample from the data distribution, whereas the D is trained, distinguishing in case the produced sample has a position/place in the data's genuine distribution. Both G and D are regularly learned together based on games concept (Figure 4).
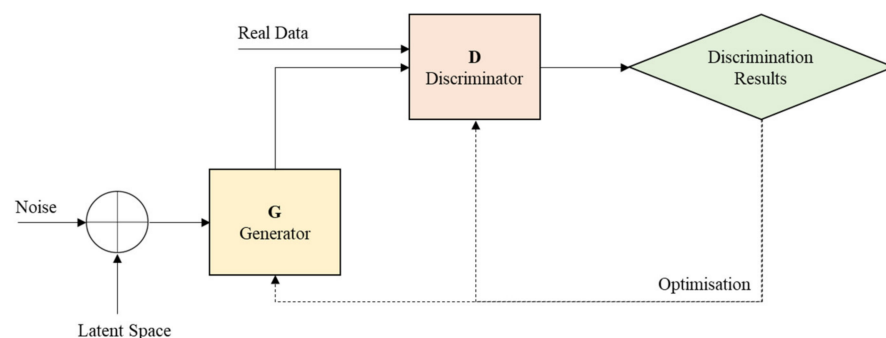
Figure 4. The general architecture of GANs.

A sample from random noise z (for every task) is produced via G to mislead D. The actual samples are then stated via D and the samples generated via G, sorting the samples as actual or fake. Through creating samples that can fool D, the G is rewarded. As a result of making the right classification, D is also rewarded. Both tasks are continuously repeated until a Nash equilibrium is obtained. The replication is then stopped. In particular, let D(s) be the probability that initiates from a genuine/real data rather than the generator. Both G and D act as a minimax match through Equation (1), as introduced by [29].

$$\min_{G} \max_{D} V(D, G) \ = E_{s \sim p_{data}(s)}[\log D(s)] \ + \ E_{z \sim p_z(z)}[\log(1 - D(G(z)))] \tag{1}$$

3.3.3. Machine Learning Models

Five machine learning models were used to evaluate the proposed GAN-based data balancing method for the spatial prediction of landslides. Those models are briefly explained in the following subsections:

Artificial Neural Networks (ANN)

ANN is a learning algorithm composed of neurons, organised in layers that can be used both for classification and regression tasks [47]. The training examples are continually offered to the algorithm, whereas the weights are updated to achieve the favorite objective value. As a benefit in landslide studies, it does not need a straightforward practice to evaluate their favorite objectives [48].

Decision Trees (DT)

DT is a supervised classification algorithm built through recursive data partitioning. In each iteration, the data is split according to a selected attribute's values until data subsets include examples of the same class. At each split in the tree, all input attributes are evaluated for their impact on the predictable attribute. DT is doable without a prior understanding of data distribution and straightforward analysis. Moreover, it is easy to design and explainable for in charge decision-makers [49,50].

Random Forests (RF)

A forest is constructed with several decision trees. Each tree fits a bootstrap sample from the training set with a random subset of features and samples selected at each node [51]. This way, the correlation between trees is minimised. It uses each tree's strongness and their correlations, which makes it less influenced by overfitting issues. Moreover, RF estimates the relative importance of the input variables, which makes it an explainable model compared to other black box models as NN. For landslide studies, mixed variables, (i.e., both categorical and numerical) are more likely. Therefore, RF is ideal for working with such variables [52].

Support Vector Machines (SVM)

SVM is a binary classification algorithm that generates a separating hyperplane (i.e., from the training set points) in the original space of *n* coordinates between the points of two distinct classes [53]. It aims to find a maximum margin of separation between the two classes and constructs a classification hyperplane in the middle of the maximum margin. In the spatial prediction of landslides, the purpose is to differentiate whether pixels are susceptible (1) or not susceptible (-1) [54].

k Nearest Neighbours (kNN)

kNN is a nonparametric classification algorithm that assumes that the classification of an object is most similar to the classes of other samples adjacent to the vector space [32]. The kNN classifier ranks the sample's neighbours among the training examples and uses the k most similar neighbors' class labels to predict the new class. These neighbours' classes are weighted using the similarity by Euclidean distance or the cosine value between the vectors.

3.3.4. Assessment and Model Analysis

a.   Accuracy metrics

The proposed models are assessed through the following accuracy metrics:

i.   Overall accuracy (OA)

OA is a common and primary metric used to measure the general performance of a prediction model. It is the rate of correct classifications using an independent test set or using cross-validation data. In this study, OA was measured using the following Equation (2):

$$\text{OA} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{2}$$

where TP is true positive which means the model predicts positive and it is true, TN is true negative (the model predicts negative and it is true), FP is false positive (the model predicts positive and it is false), and FN is false negative (the model predicts negative and it is false).

ii.   Confusion matrix

A confusion matrix, another statistical method for validating the model, provides the accuracy of the obtained classification [55]. The confusion matrix was calculated by comparing the location and class of each ground truth pixel with the corresponding location and class in the obtained classified image.

iii.   F1-score

The F1-score is the harmonic mean of recall (*r*) and precision (*p*), with a higher score as a better model's function. Precision (*p*) is determined by dividing the true positives (number of landslide pixels) with the total number of pixels classified as a landslide. The sensitivity (*r*) is the degree of true positives acceptably (correctly) predicted and identified.

It can be defined as the number of true positives divided by the overall number of pixels belonging to landslide's class. The F1-score is calculated using the following formula [56]:

$$\text{F1-score} = 2 \times \frac{p \times r}{p + r} \tag{3}$$

$$p = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4}$$

$$r = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5}$$

The problem with the F1-score metric is assuming a 0.5 threshold for selecting which samples are predicted as positive. Changing this threshold would change the performance metrics. AUROC curve is a very common method to solve this problem.

iv.   Area under the receiver operating characteristic curve (AUROC)

The area under ROC is known as AUROC, a quantitative measure summarizing model performance. ROC curves help to recognize the balance between true-positive and false-positive rates. A perfect model has 1.0 AUROC, and 0.5 AUROC indicates random models. The closer the AUROC to 1.0, the higher the model performance.

b.   *Model analysis*

To achieve this goal, three experiments were performed as follows:

1.   Impacts of training data size: The quantity of training samples affects the machine learning models' performance [57]. Thus, we conducted an experiment to understand how each model is impacted by the size of the training dataset. Also, this experiment was performed to understand how data balancing methods work with different sizes of training samples.
2.   Impacts of the ratio between positive and negative samples: The ratio between the positive and negative samples has a great impact on the machine learning models' performance. As a result, this experiment was performed to understand how each model is impacted by the ratio between the landslide and non-landslide samples.
3.   Impacts of model's optimisation: Several hyperparameters for each machine learning model were employed in this research. Those require optimisation to achieve the best performance for the model and also to improve their generalisation in unknown areas. Thus, this experiment was conducted to study the impact of model optimisation on the accuracy of the landslide susceptibility maps.

3.3.5. Preparing Landslide Susceptibility Maps

Based on the models' outputs and analysis step, each cell in the study was assigned a probability of landslide occurrence value ranging from 0 to 1 using the models' predictions: the greater the probability value, the more exposed (susceptible) to landslides [58]. The resulting raster maps were then reclassified according to five categorical classes via ArcGIS 10.7 software: very-low (<0.2), low (0.2–0.4), moderate (0.4–0.6), high (0.6–0.8), and very-high (>0.8). The classification was carried out based on the natural breaks method (the Jenks method of optimisation), which was proven as a reliable classifier in landslide susceptibility mapping [58–60]. This classifier is an iterative process based on data segmentation aims to calculate the best value arrangement in various classes. It has no class bias, and the intra-class deviation is minimum and inter-class deviation is maximum [60].

4. Results

This section describes the results obtained from a number of experiments conducted in this study. It investigates various sampling methods (Table 2) integrated with five machine learning models (ANN, RF, DT, kNN, SVM) for the spatial prediction of landslides at Chukha, Bhutan. Those experiments included: (i) assessing two oversampling techniques i.e., SMOTE and GAN for balancing landslide datasets; (ii) assessing and comparing
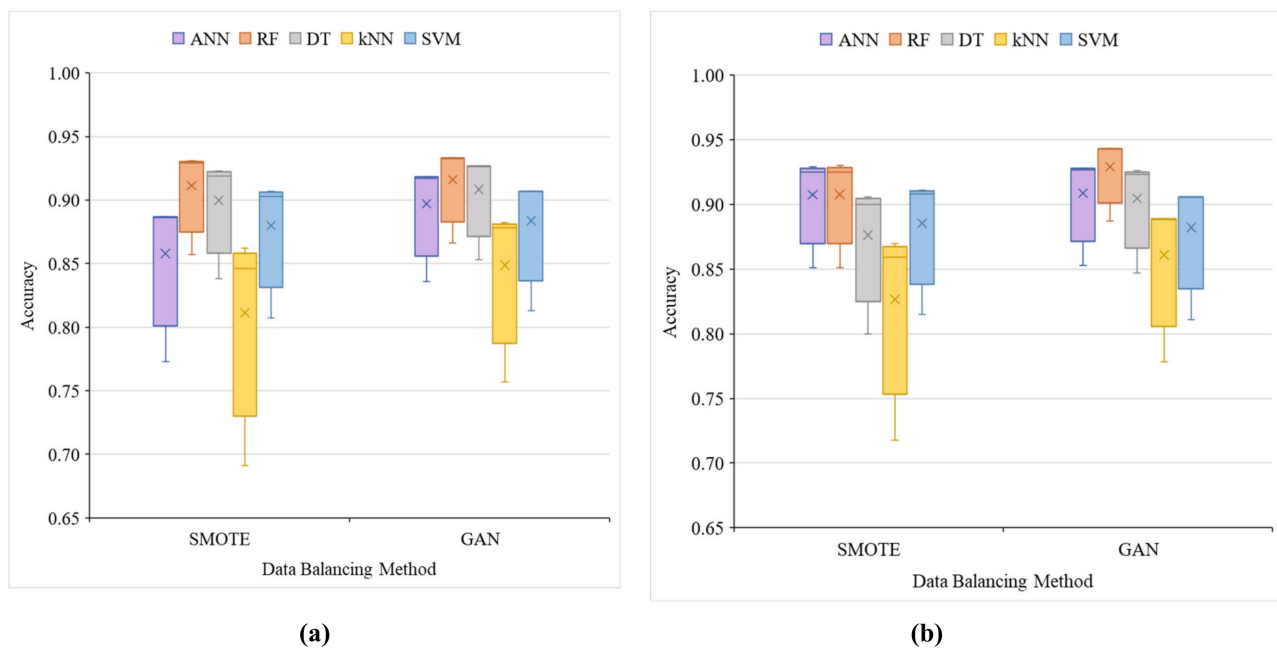
different machine learning models for the susceptibility mapping considering various sampling techniques; and (iii) analysing the proposed models through (1) optimisation of their hyperparameters, (2) studying the effect of training data size, and (3) studying the effect of the ratio between non-landslide and landslide samples in the training datasets. The following subsections describe the main findings obtained from these experiments.

Table 2. The definition of sampling methods used in this research.

Sampling Method	Explanation
Dense sampling	Non-landslide samples have been generated randomly, covering the investigated area with 500 m as a minimum distance between the points. The total number of non-landslide samples was 952 compared to 269 landslide samples. The ratio between the non-landslides to landslide samples is ~3.54.
Sparse sampling	Non-landslide samples have been generated randomly to sparsely cover the study area. The number of non-landslide samples is the same as the landslide samples. The minimum distance allowed between the points was 500 m. The ratio is 1:1 between the non-landslides and landslide samples.
SMOTE	Given the original samples, which contained 952 non-landslide samples and 269 landslide samples, the SMOTE procedure was applied to generate an additional of 683 landslide samples. Thus, the ratio became 1:1 between the non-landslides and landslide samples.
GAN	Given the original samples, which contained 952 non-landslide samples and 269 landslide samples, the GAN model was applied to generate an additional 683 landslide samples. Thus, the ratio became 1:1 between the non-landslides and landslide samples.

4.1. Assessing Data Balancing Methods, i.e., SMOTE and GAN

The two approaches (SMOTE, GAN) were assessed using five different machine learning models, such as ANN, RF, DT, kNN and SVM. The assessment was carried out based on several standard accuracy metrics, including OA, Kappa Index, F1-score, and AUROC. Figure 5 demonstrates the evaluation of the two methods based on the four accuracy metrics and the five models.

(a)   **(b)**

Figure 5. Assessment of SMOTE and GAN data balancing methods using different machine learning models. The accuracy refers to the total combination of four multi-metrics including OA, Kappa Index, F1-score, and AUROC. (**a**) Default parameters, and (**b**) optimal models.

Considering the models with default parameters, they all performed better (any accuracy metric) with the samples generated by the GAN method compared to SMOTE, except the two data balancing methods performed equally for SVM based on F1-score. Based on AUROC, the models i.e., ANN, RF, DT, kNN, and SVM achieved 0.918, 0.933, 0.927, 0.878, and 0.907 with GAN compared to SMOTE which yielded 0.887, 0.929, 0.919, 0.846, and 0.903, respectively. Nevertheless, using the optimal parameters for the models, SVM gained accuracy with SMOTE higher than that from GAN. The remaining models all performed better with GAN. Employing the optimal parameters of the models ANN, RF, DT, kNN, and SVM achieved 0.93, 0.95, 0.923, 0.91, and 0.906 AUROC with GAN compared to SMOTE which yielded 0.925, 0.925, 0.900, 0.859, and 0.908 AUROC, respectively.

Consequently, as it appears from the above assessment, GAN was found as an effective approach that can be used as a data balancing method for spatial prediction of landslides compared with more standard methods such as SMOTE. Although the results from this study were consistent with several machine learning models, some algorithms like SVM (with optimal parameters) seem to gain higher accuracy with the SMOTE method compared to the proposed GAN approach. Additional experiments have been considered to analyse further the two data balancing methods in more detail, explained in the following subsections.

4.2. Assessment and Comparison of used Models

The inventory data used in this research (available for the study area) are highly concentrated along the highway. This issue has created a DI problem since the non-landslide samples generated to cover the study area were more than the number of landslide samples (by ~3.54 times). As mentioned earlier, two data balancing methods (i.e., SMOTE and GAN) were applied to correct the DI. In addition, those techniques were compared to the standard sampling methods, namely dense sampling and sparse sampling. Those methods are undersampling techniques widely used in spatial prediction of landslides studies. Sparse sampling is applied in usual practice, which generates non-landslide samples as much as the landslide samples (equally). No DI issues occur with this sampling method. However, it is not always effective in real-world situations to use such methods as the case in this research. Therefore, it was important to solve the DI problem before the application of machine learning models.

The results are summarized in Appendix A and Figure 6a,b, for (a) default parameters, and (b) optimal parameters. The results based on OA show that the dense sampling is better than the sparse sampling for both default and optimal models, except in the optimal case, DT performed better with the sparse samples. Considering the Kappa index, the two methods, i.e., dense and sparse sampling, both performed differently based on the type of model used. With the default parameters, ANN and SVM performed better with the dense sampling, whereas the remaining models performed best with the sparse sampling.

Utilising the optimal parameters, kNN and SVM achieved the best results with the dense sampling, whereas the remaining models performed best with the sparse sampling. Based on F1-score, sparse sampling was found better than dense sampling for all the models except the default ANN. Moreover, tree models, such as DT and RF with default parameters, performed better with the sparse sampling compared to the other method. DT with the optimal parameters also achieved higher accuracy with the sparse sampling. RF with the optimal parameters worked best with the dense sampling. The remaining models all worked best with the dense sampling for both the default and optimal parameters.

According to any accuracy metric used, both SMOTE and GAN methods achieved higher accuracies than the standard dense and sparse sampling methods. Based on F1-score, on average, the SMOTE and GAN methods improved the accuracy of the models (with default parameters) by 0.104 ($\pm$0.05) and 0.115 ($\pm$0.058), respectively. Regarding the optimal models, the accuracy was improved by 0.051 ($\pm$0.039) and 0.06 ($\pm$0.034) after correcting the DI by the SMOTE and GAN methods, respectively.

(a)

(b)
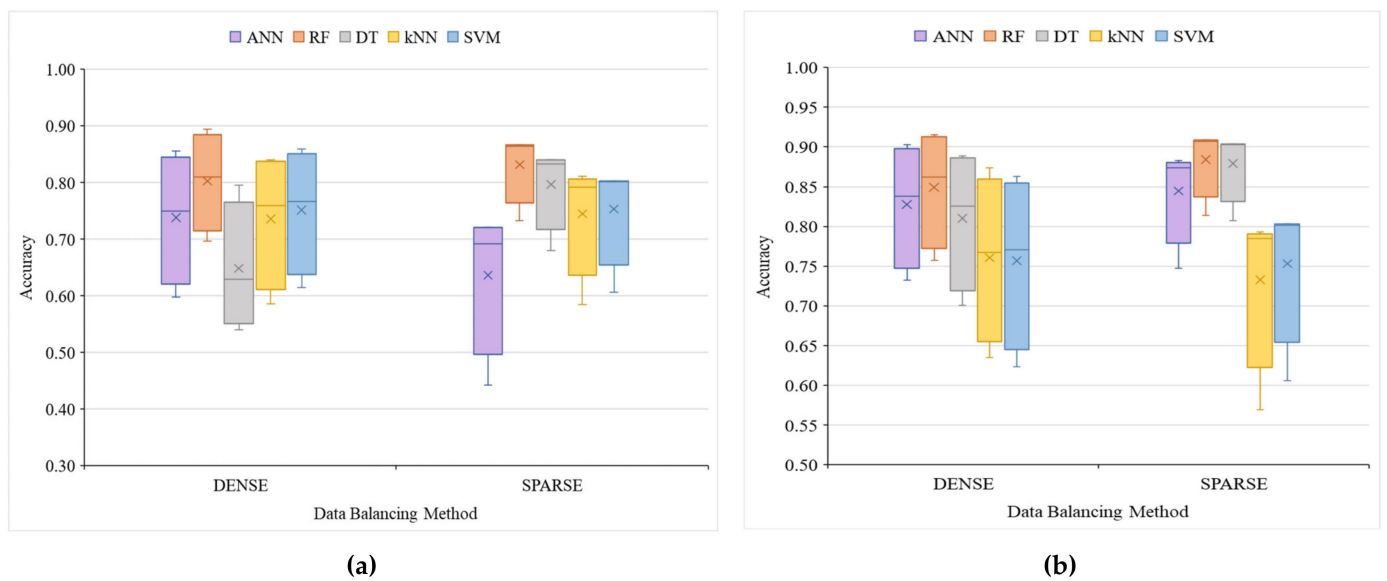
Figure 6. Assessment of DENSE and SPARSE data balancing methods using different machine learning models. The accuracy refers to the total combination of four multi-metrics including OA, Kappa Index, F1-score, and AUROC. (**a**) Default parameters, and (**b**) optimal parameters.

4.3. Model Analysis

4.3.1. Analysis of Optimisation Impact on Machine Learning Models

The performance of the used machine learning models depends on several hyperparameters. This research used both the default values of those hyperparameters provided by the sklearn package and the optimal values found by a randomised search algorithm over a specific search space of values (Table 3). The optimisation was based on the critical hyperparameters for the models. The search space for those hyperparameters was based on common values and a logical minimum and maximum values for each hyperparameter. The randomised search algorithm was run for 20 iterations to find the best values using a cross-validation AUROC accuracy metric. The results of this process are summarised in Table 3. Those best models were used in the experiments (conducted in this study) and compared to the default models.

Table 3. Summary results of optimisation of machine learning models.

Model (Parameters)	Search Space	Best Value	Validation Score (Std)
ANN			
Solver	[LBFGS, SGD, ADAM]	LBFGS	
Max Iteration	(10, 1000, steps:10)	650	
Learning Rate	power(10, range(-3, 1))	0.1	
Type of Learning Rate	[Constant, Invscaling, Adaptive]	Constant	0.882
Hidden Layer Sizes	[(128,), (64,), (32,), (128, 64), (64, 32), (32, 16), (128, 64, 32), (64, 32, 16), (32, 16, 8)]	(32,)	(0.007)
Alpha (L2)	power(10, range(−4, 1))	1.0	
Activation Function	[Identity, Logistic, Tanh, ReLU]	Logistic	
RF			
Number of Estimators	(10, 1000, steps:10)	350	0.910
Max Depth	(2, 15, steps:1)	8	(0.027)
Max Features	[Auto, Sqrt, Log2]	Log2	
DT			0.910
Max Depth	(2, 15, steps:1)	5	(0.024)

Table 3. *Cont.*

Model (Parameters)	Search Space	Best Value	Validation Score (Std)
kNN Nearest Neighbours	(3, 15, steps:1)	6	0.842 (0.008)
SVM C Kernel Function	(1, 1000, steps:10) [Linear, RBF, Sigmoid]	81 RBF	0.854 (0.012)

4.3.2. Analysis of Training Data Size Impact on Machine Learning Models

The size of the training data has an impact on the performance of the machine learning models. In this study, several training data sizes have been tested with the five models used. In addition, different sampling methods, including dense, sparse, SMOTE and GAN, were also tested with the training data sizes. Generally, the results (as presented in Appendix B) show that a training size of 0.7 (test data size = 0.3) can achieve good results. Some models such as RF, DT, and kNN, when combined with SMOTE, showed better accuracies with 0.9 training data size. When combined with different sampling methods, some other models showed the best results, with smaller training data sizes 0.5 or 0.3. However, those results were random, and no consistent pattern was observed in the results. Consequently, this study suggests that a training data size of 0.7 should be used with the test models.

4.3.3. Analysis of Landslide to Non-Landslide Sample Ratio Impact on Machine Learning Models

This research also investigated the effect of the ratio between landslide and non-landslide samples on different machine learning models [57]. The study also analysed how the two data balancing, i.e., SMOTE and GAN, perform with different ratios of landslide/non-landslide samples. In particular, three different experiments have been conducted, including the ratio of 1:1, 1:2, and 1:3 of landslide to non-landslide samples.

First, a balanced dataset (sparse) is created based on the number of landslide samples. The models were tested using this dataset and compared to each other. The results are summarised in Table 4. RF was found to perform the best according to all the four accuracy metrics (OA= 0.914, Kappa= 0.827, F1-score= 0.918, AUROC= 0.914), whereas kNN achieved the lowest accuracies. DT outperformed the remaining models but performed slightly worse than RF. ANN and SVM performed comparably.

Table 4. Assessment of models with a dataset created based on a ratio of 1:1.

	Metric			
	OA	**Kappa**	**F1-Score**	**AUROC**
ANN	0.846	0.691	0.855	0.846
RF	**0.914**	**0.827**	**0.918**	**0.914**
DT	0.901	0.802	0.904	0.901
kNN	0.802	0.605	0.820	0.802
SVM	0.852	0.704	0.852	0.852

Values presented in bold represent the best according to all the four accuracy metrics achieved by RF based on a ratio of 1:1 landslide to non-landslide samples.

Second, an imbalanced dataset was created based on the ratio of 1:2 landslide to non-landslide samples. The five models were tested based on this dataset first. Then, the two data balancing (i.e., SMOTE and GAN) were used to correct the DI. The models were tested again on these datasets and compared with the first data. Table 5 shows the results obtained from this experiment. For all the models tested, data balancing was an important step and improved the models' accuracy. GAN worked best for ANN, RF, DT, and kNN, except for SVM that SMOTE had slightly better performance than GAN (since generated

samples are linear combinations of existing samples). The results of ANN were slightly mixed. The accuracies of ANN obtained with the samples generated by the GAN method. According to OA and AUROC metrics, ANN achieved better results without correcting the DI compared to the SMOTE approach. However, according to Kappa and F1-score, the ANN achieved better results with SMOTE compared to the imbalanced dense samples.

Table 5. Assessment of models with datasets created based on a ratio of 1:2.

Sampling	Model	Metric			
		OA	Kappa	F1-Score	AUROC
Dense	ANN	0.922	0.829	0.889	0.926
	RF	0.909	0.804	0.874	0.917
	DT	0.909	0.799	0.867	0.904
	kNN	0.840	0.636	0.755	0.815
	SVM	0.860	0.685	0.790	0.843
SMOTE	ANN	0.916	0.833	0.920	0.917
	RF	0.941	0.882	0.944	0.941
	DT	0.904	0.808	0.910	0.904
	kNN	0.870	0.740	0.881	0.870
	SVM	**0.932**	**0.864**	**0.933**	**0.932**
GAN	ANN	**0.951**	**0.901**	**0.951**	**0.951**
	RF	**0.960**	**0.920**	**0.961**	**0.960**
	DT	**0.935**	**0.870**	**0.938**	**0.935**
	kNN	**0.895**	**0.790**	**0.895**	**0.895**
	SVM	0.926	0.852	0.927	0.926

Values presented in bold represent the best according to all the four accuracy metrics based on a ratio of 1:2 landslide to non-landslide samples.

Finally, a dataset was created with a 1:3 ratio between the landslides to non-landslide samples. The dataset was used to test the data balancing methods, i.e., SMOTE and GAN, against using the five machine learning models' imbalanced dataset. The results are shown in Table 6. The best results were achieved with the GAN method for models (i.e., ANN, RF, DT, and kNN). In contrast, SVM performed best with the SMOTE method. Thus, the used data balancing methods were found useful in improving the accuracy of the models.

**Table 6.** Assessment of models with a dataset created based on a ratio of 1:3.

Sampling	Model	Metric			
		OA	Kappa	F1-Score	AUROC
Dense	ANN	0.910	0.773	0.834	0.907
	RF	0.932	0.830	0.876	0.942
	DT	0.910	0.770	0.830	0.899
	kNN	0.895	0.715	0.785	0.852
	SVM	0.926	0.813	0.864	0.930
SMOTE	ANN	0.928	0.856	0.932	0.928
	RF	0.944	0.889	0.947	0.944
	DT	0.922	0.843	0.927	0.922
	kNN	0.878	0.757	0.886	0.878
	SVM	**0.938**	**0.876**	**0.939**	**0.938**
GAN	ANN	**0.951**	**0.901**	**0.950**	**0.951**
	RF	**0.951**	**0.901**	**0.951**	**0.951**
	DT	**0.930**	**0.860**	**0.931**	**0.930**
	kNN	**0.905**	**0.810**	**0.905**	**0.905**
	SVM	0.934	0.868	0.935	0.934

Values presented in bold represent the best according to all the four accuracy metrics based on a ratio of 1:3 landslide to non-landslide samples.

4.4. Spatial Prediction of Landslides with Machine Learning Models

Finally, five landslide susceptibility maps were produced for the study area (Chuka Dzongkhag) by ANN, RF, DT, kNN, and SVM models using the corrected datasets created by the proposed GAN approach with optimal parameters (Figure 7). In the maps, the high susceptibility areas are distributed along the Phuentsholing–Thimphu highway. Besides, the majority of the genuine landslides are scattered in the higher susceptibility area, whereas only limited landslides are spread in the lower susceptibility area. According to the maps, the results reveal that the proposed models' predicted susceptibility values are in good agreement with the real landslide distribution. Figure 8 illustrates the AUROC plot for the best models balanced with GAN. These graphs can better interpret the relationship between true positive rate and false-positive rate, where 1.0 indicates the best AUROC and 0.5 indicates the worst AUROC.

The results of variable importance for RF using the corrected datasets created by the proposed GAN are provided in Figure 9. Based on the RF analysis for variable contribution, the most significant conditioning factors were STI, lithology, followed by the distance from the road. While total curvature, TWI, and distance from the stream were in lower ranks in terms of their significant effects on the landslide occurrence in the region. This result indicates a logical outcome, as the study area is located in a mountainous region that is exposed to intense monsoon rainfall, and identified as vastly cracked and weathered rocks with plenty of clay. Additionally, road construction activities such as hill cutting are a common process in such areas, which all could initiate soil instability. Considering STI, it ranked the highest conditioning factor among the others in the study area. It represents the slope failure and deposition processes and directly relates to the water accumulation at the bottom of the catchment and the erosion quantity [61]. As Bhutan is exposed to an annual high level of rainfall (maximum of 4000–6000 mm), usually throughout the southwestern monsoon, it is expected that the greater amount of the flow can carry more sediment that contributes to soil instability/landslides triggering. Another important factor was lithology. Lithological units have different impacts on landslides [62–64]. Generally, weathering prompts slopes to failures. Weathered and fine/cracked rocks are more exposed to slope failure than strong separate/disjointed ones [65]. The most disposed zones to slope failures of the Bhutan region are deeply cracked (results in unstable soils) and weathered rocks of phyllite, slate, and schist that comprise a large quantity of clay. Clay particles control soil strength and shearing mechanism when their proportion is more than 30% [66]. Clay cannot be stabilized and consequently holds the particles together when they are wet [67]. As such, soil erosion frequently occurs during the wet season and this could be one of the important factors causing landslides [67]. Distance from road was another crucial factor in the study area. The majority of landslides occur along Phuentsholing–Thimphu highway, a lifeline link that connects the capital Thimphu with India. Numerous road cuttings, external loads applications, and plants removal are general activities while constructing and maintaining roads and highways, raising the probability of landslides adjacent to such places as the case in the current study.
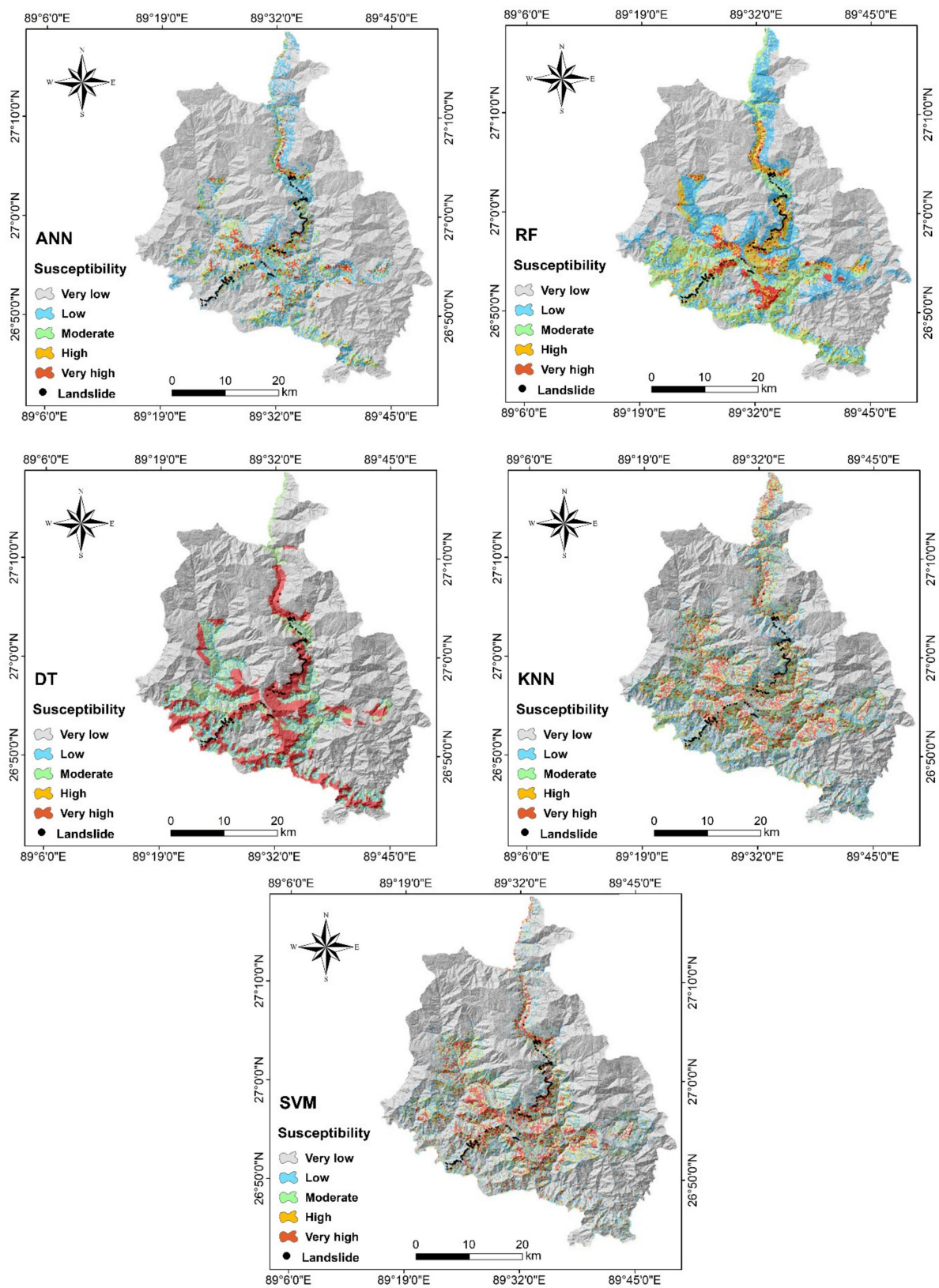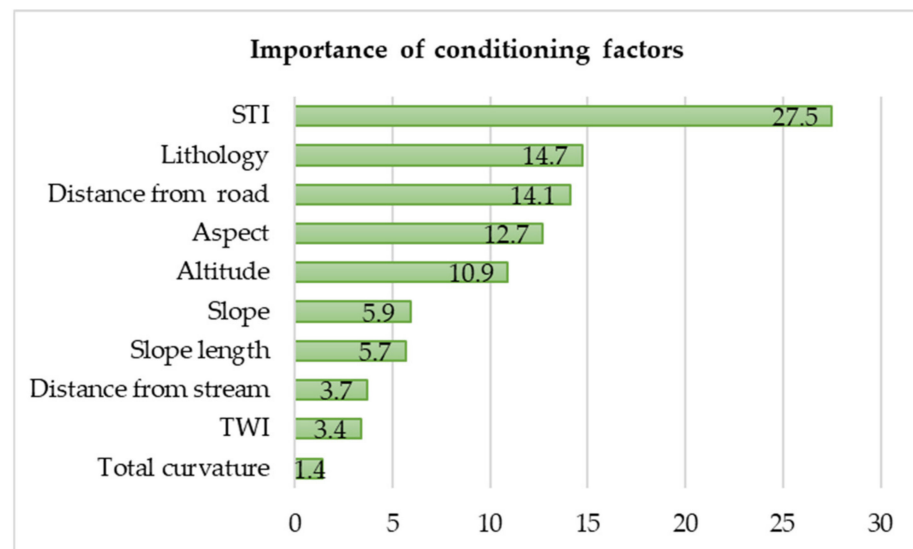
Figure 7. Landslide susceptibility maps produced by ANN, RF, DT, kNN, and SVM models using the corrected datasets created by the proposed GAN.
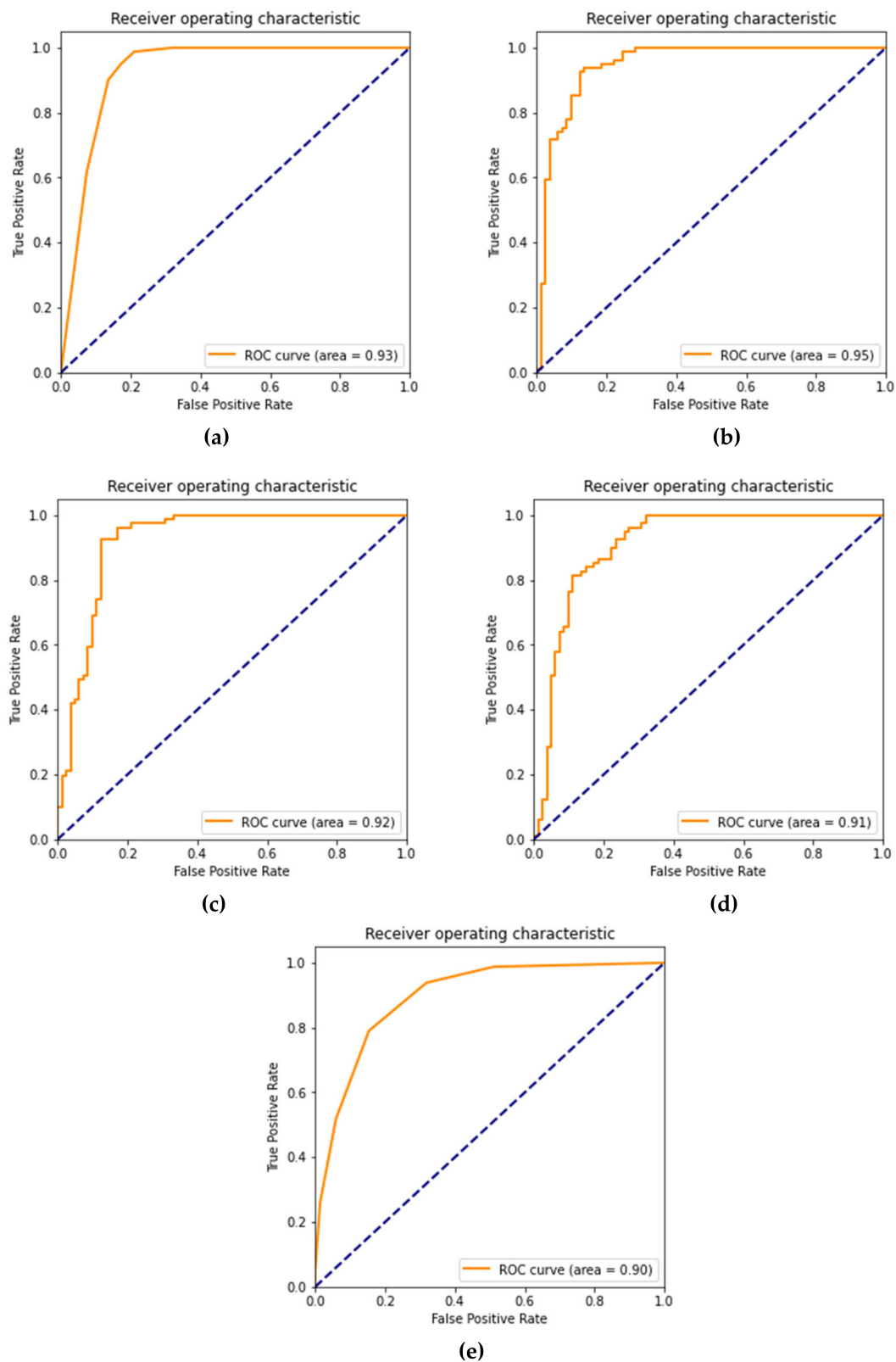
Figure 8. Evaluation of the landslide susceptibility maps (balanced by GAN) based on the AUROC test using: (**a**) ANN, (**b**) RF, (**c**) DT, (**d**) KNN, and (**e**) SVM.

Figure 9. Variable importance analysis derived from RF corrected datasets created by GAN.

5. Discussion

This paper introduced a new method of data balancing based on GAN for improving training in various machine learning models, i.e., ANN, RF, DT, kNN, and SVM. The new method was compared to the standard data balancing methods, such as SMOTE. Those methods were also compared to commonly used methods, such as dense and sparse sampling. When a dataset is not balanced, the predictive models often produce biased results. The models become biased towards the majority class. As a result, developing data balancing methods in spatial prediction of landslides is important.

The proposed data balancing method is based on state-of-the-art generative models, such as GAN. In previous works, GAN models were also applied to create synthesised datasets to improve the performance of machine learning models [18]. Such models learn the distribution of a given training dataset and provide the ability to generate samples that are similar to training samples. In this way, additional training samples can be created to correct imbalanced datasets.

Both the proposed GAN and SMOTE data balancing methods yielded higher accuracy than the imbalanced dataset and sparse sampling. The models with default parameters worked best with the samples generated by the proposed GAN method. Based on AUROC, the models (i.e., ANN, RF, DT, kNN, and SVM) achieved 0.918, 0.933, 0.927, 0.878, and 0.907. The optimal models, i.e., ANN (0.93), RF (0.95), DT (0.92), and kNN (0.91) also achieved the best accuracies with GAN, except SVM yielded the highest accuracy (0.906) with SMOTE. These results indicate the high standing of correcting DI in spatial landslide prediction. Four of the tested models showed higher accuracies compared to SMOTE, both with the default and optimal parameters.

The factor importance resulting from the RF balanced by GAN model suggests that the most important conditioning factors for landslides are STI, lithology, and distance from the road, which shows a direct connection with landslide occurrence. In terms of factor importance, limited studies might have similar results as each area has various geo-environmental variables [68]. In previous studies, distance to rivers, altitude, and lithology were the dominant factors for landslides, while the distance to roads, distance to river and lithology were the main factors [69]. Another study found that distance to fault, elevation and distance to road had the most contribution among 17 conditioning factors [70].

Our finding is quite reasonable, as the Chukha Dzongkhag region is highly impacted by monsoon rainfall, leading to erosion of soil and rocks and triggering extensive mass movements, especially near roads where anthropogenic activities are prevailing.

The factor importance results, however, indicate different weights for the conditioning factors within every algorithm and it is one of the reasons that some differences were observed among various results in some regions in susceptibility maps. This is due to their architecture and the way these models approach a problem in model fittings (learning schemes) in machine learning/data science [71]. Furthermore, we used the Jenks natural break method (The Jenks method of optimisation) to classify the data based on their distribution. Several studies showed that this method is suitable for natural hazard studies such as landslide susceptibility [58,60,68]. The primary benefit of the natural break method is that it has no class bias. Also, the intra-class deviation is the lowest (minimum) and inter-class variation is the highest (maximum) Jenks, [59,60]. In the previous similar analysis by Taalab et al. [19], distance to rivers, TWI, and rainfall were the most important factors, while curvature was the least influential factor. Another study by Xiong et al. [72] showed that the mean altitude, altitude difference, aridity index, and groove gradient have the most significant contribution in their study area. In contrast, seismic intensity and area of moderate soil erosion had the least share. These differences can be attributed to each study area's environmental and geological variation and can be considered as a general limitation of machine learning models.

The statistical validation using ANOVA is applied in order to decide whether two very close values are significantly different from each other [73]. The results in Table 7 show that there are significant differences between different classes and the *p*-value are less than 0.05 for all classes.

Table 7. Statistical Validation of RF-GAN using ANOVA.

Classes	Estimate	Std. Error	Mean	+95%	−95%	*p*-Value	
Very Low	−0.5969524	0.0008443	0.0616	0.0619	0.0613	$< 2e^{-16}$	***
Low	−0.3625884	0.0009018	0.2960	0.2967	0.2952	$< 2e^{-16}$	***
Moderate	−0.1826528	0.0009546	0.4759	0.4768	0.4751	$< 2e^{-16}$	***
High	0.6586289	0.0008322	0.6586	0.6596	0.6575	$< 2e^{-16}$	***
Very High	0.1467942	0.0243412	0.8054	0.8084	0.8024	$1.64e^{-09}$	***

Signif. Codes Symbol: '***' for $p \leq 001$.

This study also analysed the impacts of training data size and landslide/non-landslide samples ratio on the performance of the models. Generally, a training size of 70% can achieve decent results. Some models, such as RF, DT, and kNN when combined with SMOTE, showed better accuracies with 0.9 training data size. With the balanced dataset (ratio 1:1), RF achieved the best results, whereas kNN achieved the lowest accuracies. The analysis based on ratio 1:2 showed that the models i.e., RF, DT, and kNN worked best with GAN. However, SMOTE performed better than GAN for SVM. The results of ANN were slightly mixed. With a ratio of 1:3, the results suggested that GAN was the best method for models, i.e., ANN, RF, DT, and kNN. In contrast, SVM performed best with SMOTE. Although the number of training size of landslides (inventory) using machine learning models have affected the landslide susceptibility maps, the type of landslides could affect the distribution of susceptibility classes, with more reliable results when a multi-temporal inventory is considered. Previous studies highlighted that different rainfall events in an area could affect the estimation of shallow landslide susceptibility in the same region [74].

Thus, the role of different types of landslides can further be investigated using advanced machine learning models [74]. Moreover, scholars stated that limitation of accurate information on the type of landslides (e.g., debris flow, rock fall, etc.) might cause unexpected consequences in the models' comparison [63]. On the other hand, increasing positional inventory-based error is commonly related to the misrepresentations of modeling and validation outcomes [75]. However, complex mutuality exists between the statistical models and inventory-based spatial inaccuracies, which can be reduced by choosing a compatible study design [75].

GAN can generate synthetic data that has the original data distribution [29]. Since these models depend on neural networks, which have several hyperparameters to fine-

tune, the generated data can be controlled and targeted for specific applications. However, SMOTE is less flexible to control. Thus, its performance is hard to improve for some models [30,33]. However, training of GAN models is not an easy task and requires extensive testing of the model's architecture and hyperparameters. Careful analysis is crucial to avoid overfitting when implementing GAN models. Furthermore, as with any deep neural networks, optimisation is another important issue that should be taken into account as GANs are highly sensitive to hyperparameters settings [29,76]. Another possible limitation is that only limited samples may be generated by the generator failing (mode collapse), or the discriminator might get over-successful that it could vanish the generator gradient that nothing can be learned (diminished gradient). In some cases, this difficulty may result in SMOTE outperforming the methods that are based on GAN or related complex models. As a result, GAN can be a better alternative for SMOTE when more data are required, with a similar distribution to the training data.

6. Conclusions

Spatial prediction of landslides is an important step for landslide risk assessment and planning mitigation measures. As a result, many studies have developed models for spatial landslide prediction, and research has progressed significantly in this direction. This study developed a new approach to data balancing based on GAN. This approach was compared to other data balancing methods such as SMOTE as well as other sample techniques, such as dense and sparse sampling. The assessment of these sampling methods was performed based on five machine learning models i.e., ANN, RF, DT, kNN, and SVM, including the default and optimal parameters. Optimisation of the models' hyperparameters was based on a randomised search for 20 iterations using a cross-validation AUROC. This study showed that data balancing is an important step that can significantly impact the performance of machine learning models used for spatial prediction of landslides. Other generative models, such as Gaussian mixture models, autoencoders, and variational autoencoders also should be considered for landslide data balancing.

To sum up, one of the major limitation of machine learning models is that they need plenty of data to perform well. Generally, more data (quality and quantity data) can lead to a better learning scheme and better performance. Nevertheless, data from smaller number of samples (particularly landslides data) can lessen the performance and accuracy of such models. Using state-of-the-art generative models (GAN) can fill the gap and solve such problems. Our finding suggests that RF balanced with GAN can provide the most reasonable criterion for landslide prediction and future aim for comprehensive risk assessment in Bhutan. The approach is especially useful when integrated with RF. It can create a better practice of feature selection due to the nature of RF. It also reduces overfitting in decision trees and works well with both continuous and categorical data.

Our attempt in this study can be considered a preliminary experiment to understand the likelihood of landslides occurring in the region where the landslide data are limited, imbalanced, or scarce.

Author Contributions: Project administration: B.P.; conceptualization: H.A.H.A.-N. and B.P.; methodology: H.A.H.A.-N.; modelling: H.A.H.A.-N.; writing original draft: H.A.H.A.-N.; supervision: B.P., G.B., and R.S.; validation: B.P.; visualization: B.P.; resources: B.P., A.A.; review and editing: B.P., R.S., and A.A.; funding acquisition: B.P. and A.A. All authors have read and agreed to the published version of the manuscript.

List of Abbreviations

DI	data imbalance
GAN	generative adversarial networks
SMOTE	synthetic minority oversampling technique
ANN	artificial neural networks
RF	random forests
DT	decision trees
kNN	k-nearest neighbours
SVM	support vector machine
OA	overall accuracy
AUROC	area under receiver operating characteristic curves
TWI	topographic wetness index
STI	sediment transport index
FDA	fisher discriminant analysis
ADASYN	adaptive synthetic sampling approach for imbalanced learning
LDA	linear discriminant analysis
DEM	digital elevation model
ALOS	advanced land observing satellite
PALSAR	the phased array L-band synthetic aperture radar

Appendix A

Table A1. Assessment of different sampling methods used in this study with five machine learning models (both default and optimised parameters)–based on OA.

Parameters	Model	Sampling Method			
		Dense Samples	Sparse Samples	SMOTE	GAN
Default	ANN	0.856	0.721	0.887	0.918
	RF	0.894	0.866	0.929	0.933
	DT	0.854	0.840	0.919	0.927
	kNN	0.840	0.792	0.846	0.878
	SVM	0.859	0.803	0.903	0.907
Optimal	ANN	0.903	0.874	0.925	0.927
	RF	0.908	0.907	0.925	0.943
	DT	0.889	0.903	0.900	0.923
	kNN	0.874	0.784	0.859	0.889
	SVM	0.863	0.803	0.908	0.906

Table A2. Assessment of different sampling methods used in this study with five machine learning models (both default and optimised parameters)–based on Kappa Index.

Parameters	Model	Sampling Method			
		Dense Samples	Sparse Samples	SMOTE	GAN
Default	ANN	0.597	0.442	0.773	0.836
	RF	0.697	0.732	0.857	0.866
	DT	0.582	0.68	0.838	0.853
	*k*NN	0.585	0.584	0.691	0.757
	SVM	0.614	0.606	0.807	0.813
Optimal	ANN	0.732	0.747	0.851	0.853
	RF	0.757	0.814	0.851	0.887
	DT	0.701	0.807	0.800	0.847
	*k*NN	0.635	0.569	0.718	0.778
	SVM	0.623	0.606	0.815	0.811

Table A3. Assessment of different sampling methods used in this study with five machine learning models (both default and optimised parameters)–based on F1-score.

Parameters	Model	Sampling Method			
		Dense Samples	Sparse Samples	SMOTE	GAN
Default	ANN	0.690	0.661	0.885	0.916
	RF	0.765	0.862	0.931	0.932
	DT	0.676	0.826	0.923	0.926
	*k*NN	0.690	0.811	0.862	0.882
	SVM	0.705	0.800	0.907	0.907
Optimal	ANN	0.794	0.883	0.929	0.928
	RF	0.817	0.909	0.930	0.943
	DT	0.773	0.904	0.906	0.926
	*k*NN	0.716	0.793	0.870	0.888
	SVM	0.712	0.800	0.911	0.906

Table A4. Assessment of different sampling methods used in this study with five machine learning models (both default and optimised parameters)–based on AUROC.

Parameters	Model	Sampling Method			
		Dense Samples	Sparse Samples	SMOTE	GAN
Default	ANN	0.809	0.721	0.887	0.918
	RF	0.855	0.866	0.929	0.933
	DT	0.795	0.840	0.919	0.927
	*k*NN	0.828	0.792	0.846	0.878
	SVM	0.825	0.803	0.903	0.907
Optimal	ANN	0.882	0.874	0.925	0.93
	RF	0.915	0.907	0.925	0.95
	DT	0.878	0.903	0.900	0.923
	*k*NN	0.818	0.785	0.859	0.91
	SVM	0.830	0.803	0.908	0.906

Appendix B

Table A5. Assessment of the models with different sampling methods based on OA.

Sampling	Model	Training Size			
		0.3	0.5	0.7	0.9
Dense	ANN	0.855	0.903	0.918	0.894
	RF	0.892	0.908	0.924	0.902
	DT	0.904	0.889	0.899	0.878
	*k*NN	0.832	0.874	0.872	0.837
	SVM	0.843	0.863	0.880	0.862
Sparse	ANN	0.836	0.874	0.870	0.833
	RF	0.881	0.907	0.914	0.907
	DT	0.889	0.903	0.907	0.907
	*k*NN	0.769	0.784	0.833	0.759
	SVM	0.780	0.803	0.852	0.796
SMOTE	ANN	0.921	0.925	0.935	0.932
	RF	0.920	0.925	0.941	0.948
	DT	0.900	0.900	0.899	0.916
	*k*NN	0.839	0.859	0.874	0.885
	SVM	0.903	0.908	0.913	0.911
GAN	ANN	0.898	0.927	0.937	0.916
	RF	0.939	0.943	0.946	0.942
	DT	0.929	0.924	0.934	0.942
	*k*NN	0.890	0.889	0.897	0.890
	SVM	0.900	0.906	0.900	0.880

Table A6. Assessment of models with different sampling methods based on Kappa Index.

Sampling	Model	Training Size			
		0.3	0.5	0.7	0.9
Dense	ANN	0.579	0.732	0.766	0.703
	RF	0.711	0.757	0.795	0.723
	DT	0.737	0.701	0.723	0.658
	*k*NN	0.494	0.635	0.612	0.512
	SVM	0.568	0.623	0.666	0.612
Sparse	ANN	0.671	0.747	0.741	0.667
	RF	0.761	0.814	0.827	0.815
	DT	0.783	0.807	0.815	0.815
	*k*NN	0.539	0.569	0.667	0.519
	SVM	0.560	0.606	0.704	0.593
SMOTE	ANN	0.842	0.851	0.871	0.864
	RF	0.839	0.851	0.881	0.895
	DT	0.800	0.800	0.797	0.833
	*k*NN	0.677	0.718	0.748	0.770
	SVM	0.806	0.815	0.825	0.822
GAN	ANN	0.796	0.853	0.874	0.832
	RF	0.879	0.887	0.892	0.885
	DT	0.858	0.847	0.867	0.885
	*k*NN	0.780	0.778	0.794	0.780
	SVM	0.799	0.811	0.801	0.759

Table A7. Assessment of the models with different sampling methods based on F1-score.

Sampling	Model	Training Size			
		0.3	0.5	0.7	0.9
Dense	ANN	0.672	0.794	0.819	0.772
	RF	0.781	0.817	0.844	0.786
	DT	0.799	0.773	0.789	0.737
	*k*NN	0.600	0.716	0.693	0.615
	SVM	0.670	0.712	0.744	0.702
Sparse	ANN	0.846	0.883	0.879	0.830
	RF	0.887	0.909	0.916	0.912
	DT	0.897	0.904	0.911	0.909
	*k*NN	0.776	0.793	0.838	0.755
	SVM	0.774	0.800	0.852	0.784
SMOTE	ANN	0.926	0.929	0.938	0.935
	RF	0.924	0.930	0.943	0.950
	DT	0.908	0.906	0.905	0.922
	*k*NN	0.855	0.870	0.884	0.894
	SVM	0.907	0.911	0.916	0.913
GAN	ANN	0.900	0.928	0.937	0.917
	RF	0.940	0.943	0.946	0.942
	DT	0.931	0.927	0.936	0.943
	*k*NN	0.889	0.888	0.896	0.888
	SVM	0.903	0.906	0.901	0.883

Table A8. Assessment of the models with different sampling methods based on AUROC.

Sampling	Model	Training Size			
		0.3	0.5	0.7	0.9
Dense	ANN	0.791	0.882	0.890	0.866
	RF	0.885	0.915	0.929	0.871
	DT	0.891	0.878	0.882	0.842
	*k*NN	0.739	0.818	0.794	0.749
	SVM	0.800	0.830	0.848	0.818
Sparse	ANN	0.836	0.874	0.870	0.833
	RF	0.881	0.907	0.914	0.907
	DT	0.889	0.903	0.907	0.907
	*k*NN	0.769	0.785	0.833	0.759
	SVM	0.780	0.803	0.852	0.796
SMOTE	ANN	0.921	0.925	0.935	0.932
	RF	0.920	0.925	0.941	0.948
	DT	0.900	0.900	0.899	0.917
	*k*NN	0.839	0.859	0.874	0.885
	SVM	0.903	0.908	0.913	0.911
GAN	ANN	0.898	0.927	0.937	0.916
	RF	0.939	0.943	0.946	0.942
	DT	0.929	0.923	0.934	0.942
	*k*NN	0.890	0.889	0.897	0.890
	SVM	0.900	0.906	0.900	0.880

References

1. Turner, A.K. Social and environmental impacts of landslides. *Innov. Infrastruct. Solut.* **2018**, *3*, 1–25. [CrossRef]
2. Sidle, R.C. *Using Weather and Climate Information for Landslide Prevention and Mitigation. Climate and Land Degradation*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 285–307. [CrossRef]
3. Mezaal, M.R.; Pradhan, B.; Sameen, M.I.; Shafri, H.Z.M.; Yusoff, Z.M. Optimized Neural Architecture for Automatic Landslide Detection from High-Resolution Airborne Laser Scanning Data. *Appl. Sci.* **2017**, *7*, 730. [CrossRef]

4.   Froude, M.J.; Petley, D.N. Global fatal landslide occurrence from 2004 to 2016. *Nat. Hazards Earth Syst. Sci.* **2018**, *18*, 2161–2181. [CrossRef]

5.   Dikshit, A.; Sarkar, R.; Pradhan, B.; Jena, R.; Drukpa, D.; Alamri, A.M. Temporal Probability Assessment and Its Use in Landslide Susceptibility Mapping for Eastern Bhutan. *Water* **2020**, *12*, 267. [CrossRef]

6.   United Nations Department of Economic and Social Affairs. World Economic Situation and Prospects 2019. 2019. Available online: https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/WESP2019_BOOK-web.pdf (accessed on 1 September 2021).

7.   Lee, J.-H.; Sameen, M.I.; Pradhan, B.; Park, H.-J. Modeling landslide susceptibility in data-scarce environments using optimized data mining and statistical methods. *Geomorphology* **2018**, *303*, 284–298. [CrossRef]

8.   Yang, Y.; Yang, J.; Xu, C.; Xu, C.; Song, C. Local-scale landslide susceptibility mapping using the B-GeoSVC model. *Landslides* **2019**, *16*, 1301–1312. [CrossRef]

9.   Zhao, X.; Chen, W. Optimization of Computational Intelligence Models for Landslide Susceptibility Evaluation. *Remote Sens.* **2020**, *12*, 2180. [CrossRef]

10.   Kavzoglu, T.; Sahin, E.K.; Colkesen, I. Selecting optimal conditioning factors in shallow translational landslide susceptibility mapping using genetic algorithm. *Eng. Geol.* **2015**, *192*, 101–112. [CrossRef]

11.   Zhang, S.; Ren, W.; Zhang, X.; Liu, H. *Prediction Method of Landslide Disaster in Southern China Based on Multi Attribute Group Decision Making, Proceedings of the 2016 6th International Conference on Machinery, Materials, Environment, Biotechnology and Computer*; Atlantis Press: Amsterdam, The Netherlands, 2016; pp. 2078–2083.

12.   Hussin, H.Y.; Zumpano, V.; Reichenbach, P.; Sterlacchini, S.; Micu, M.; van Westen, C.; Bălteanu, D. Different landslide sampling strategies in a grid-based bi-variate statistical susceptibility model. *Geomorphology* **2016**, *253*, 508–523. [CrossRef]

13.   Lai, J.-S.; Chiang, S.-H.; Tsai, F. Exploring Influence of Sampling Strategies on Event-Based Landslide Susceptibility Modeling. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 397. [CrossRef]

14.   Zhu, A.-X.; Miao, Y.; Liu, J.; Bai, S.; Zeng, C.; Ma, T.; Hong, H. A similarity-based approach to sampling absence data for landslide susceptibility mapping using data-driven methods. *Catena* **2019**, *183*, 104188. [CrossRef]

15.   Pourghasemi, H.R.; Kornejady, A.; Kerle, N.; Shabani, F. Investigating the effects of different landslide positioning techniques, landslide partitioning approaches, and presence-absence balances on landslide susceptibility mapping. *Catena* **2020**, *187*, 104364. [CrossRef]

16.   Chowdhuri, I.; Pal, S.C.; Arabameri, A.; Ngo, P.T.T.; Chakrabortty, R.; Malik, S.; Das, B.; Roy, P. Ensemble approach to develop landslide susceptibility map in landslide dominated Sikkim Himalayan region, India. *Environ. Earth Sci.* **2020**, *79*, 1–28. [CrossRef]

17.   Di Napoli, M.; Carotenuto, F.; Cevasco, A.; Confuorto, P.; Di Martire, D.; Firpo, M.; Pepe, G.; Raso, E.; Calcaterra, D. Machine learning ensemble modelling as a tool to improve landslide susceptibility mapping reliability. *Landslides* **2020**, *17*, 1897–1914. [CrossRef]

18.   Al-Najjar, H.A.; Pradhan, B. Spatial landslide susceptibility assessment using machine learning techniques assisted by additional data created with generative adversarial networks. *Geosci. Front.* **2021**, *12*, 625–637. [CrossRef]

19.   Taalab, K.; Cheng, T.; Zhang, Y. Mapping landslide susceptibility and types using Random Forest. *Big Earth Data* **2018**, *2*, 159–178. [CrossRef]

20.   Gupta, S.K.; Jhunjhunwalla, M.; Bhardwaj, A.; Shukla, D.P. Data imbalance in landslide susceptibility zonation: Under-sampling for class-imbalance learning. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *XLII-3/W11*, 51–57. [CrossRef]

21.   Guzzetti, F.; Reichenbach, P.; Cardinali, M.; Galli, M.; Ardizzone, F. Probabilistic landslide hazard assessment at the basin scale. *Geomorphol.* **2005**, *72*, 272–299. [CrossRef]

22.   Krawczyk, B. Learning from imbalanced data: Open challenges and future directions. *Prog. Artif. Intell.* **2016**, *5*, 221–232. [CrossRef]

23.   He, H.; Garcia, E.A. Learning from Imbalanced Data. *IEEE Trans. Knowl. Data Eng.* **2009**, *21*, 1263–1284. [CrossRef]

24.   Chawla, N.V. Data Mining for Imbalanced Datasets: An Overview. *Data Min. Knowl. Discov. Handb.* **2009**, *2009*, 875–886. [CrossRef]

25.   Galar, M.; Fernandez, A.; Barrenechea, E.; Bustince, H.; Herrera, F. A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches. *IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.)* **2012**, *42*, 463–484. [CrossRef]

26.   Haixiang, G.; Yijing, L.; Shang, J.; Mingyun, G.; Yuanyue, H.; Bing, G. Learning from class-imbalanced data: Review of methods and applications. *Expert Syst. Appl.* **2017**, *73*, 220–239. [CrossRef]

27.   Stumpf, A.; Lachiche, N.; Kerle, N.; Malet, J.-P.; Puissant, A. Adaptive spatial sampling with active random forest for object-oriented landslide mapping. In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012; pp. 87–90. [CrossRef]

28.   Stumpf, A.; Lachiche, N.; Malet, J.-P.; Kerle, N.; Puissant, A. Active Learning in the Spatial Domain for Remote Sensing Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 2492–2507. [CrossRef]

29.   Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the Annual Conference on Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 2672–2680.

30. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]

31. Tsangaratos, P.; Ilia, I. Landslide susceptibility mapping using a modified decision tree classifier in the Xanthi Perfection, Greece. *Landslides* **2016**, *13*, 305–320. [CrossRef]

32. Pyle, D. *Business Modeling and Data Mining*; Morgan Kaufmann Publishers: Burlington, MA, USA, 2003.

33. Agrawal, K.; Baweja, Y.; Dwivedi, D.; Saha, R.; Prasad, P.; Agrawal, S.; Kapoor, S.; Chaturvedi, P.; Mali, N.; Kala, V.U.; et al. A Comparison of Class Imbalance Techniques for Real-World Landslide Predictions. In Proceedings of the 2017 International Conference on Machine Learning and Data Science (MLDS), IEEE, Noida, India, 14–15 December 2017; pp. 1–8.

34. Zhao, L.; Wu, X.; Niu, R.; Wang, Y.; Zhang, K. Using the rotation and random forest models of ensemble learning to predict landslide susceptibility. *Geomat. Nat. Hazards Risk* **2020**, *11*, 1542–1564. [CrossRef]

35. Braun, A.; Garcia-Urquia, E.L.; Lopez, R.M.; Yamagishi, H. Landslide Susceptibility Mapping in Tegucigalpa, Honduras, Using Data Mining Methods. In *Proceedings of the IAEG/AEG Annual Meeting Proceedings, San Francisco, CA, USA, 2018, Volume 1*; Springer: Cham, Switzerland, 2019; pp. 207–215. [CrossRef]

36. Mutlu, A.; Goz, F. SkySlide: A Hybrid Method for Landslide Susceptibility Assessment based on Landslide-Occurring Data Only. *Comput. J.* **2020**, *2020*, bxaa063. [CrossRef]

37. Zhang, S.; Yu, P. Seismic landslide susceptibility assessment based on ADASYN-LDA model. In Proceedings of the IOP Conference Series: Earth and Environmental Science, 5th International Conference on Minerals Source, Geotechnology and Civil Engineering, Guiyang, China, 17–19 April 2020; IOP Publishing: Bristol, UK, 2020; 525, p. 12087. [CrossRef]

38. Song, Y.; Niu, R.; Xu, S.; Ye, R.; Peng, L.; Guo, T.; Li, S.; Chen, T. Landslide Susceptibility Mapping Based on Weighted Gradient Boosting Decision Tree in Wanzhou Section of the Three Gorges Reservoir Area (China). *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 4. [CrossRef]

39. Kuenza, K.; Dorji, Y.; Wangda, D. Landslides in Bhutan. In Proceedings of the SAARC Workshop on Landslide Risk Management in South Asia, Thimphu, Bhutan, 11–12 May 2010; UN Office for Disaster Risk Reduction (UNDRR): Geneva, Switzerland, 2010; pp. 73–80. Available online: https://www.preventionweb.net/files/14793_SAARClandslide.pdf (accessed on 1 August 2021).

40. Cardarilli, M.; Lombardi, M.; Corazza, A. Landslide risk management through spatial analysis and stochastic prediction for territorial resilience evaluation. *Int. J. Saf. Secur. Eng.* **2019**, *9*, 109–120. [CrossRef]

41. Gariano, S.L.; Sarkar, R.; Dikshit, A.; Dorji, K.; Brunetti, M.T.; Peruccacci, S.; Melillo, M. Automatic calculation of rainfall thresholds for landslide occurrence in Chukha Dzongkhag, Bhutan. *Bull. Int. Assoc. Eng. Geol. Environ.* **2018**, *78*, 4325–4332. [CrossRef]

42. Gansser, A. Geology of the Bhutan Himalaya. In *Denkschriften der Schweizerischen Naturforschenden Geselschaft*; Birkhäuser Verlag: Basel, Switzerland, 1983; p. 181.

43. Dikshit, A.; Sarkar, R.; Pradhan, B.; Acharya, S.; Dorji, K. Estimating Rainfall Thresholds for Landslide Occurrence in the Bhutan Himalayas. *Water* **2019**, *11*, 1616. [CrossRef]

44. Sarkar, R.; Dorji, K. Determination of the Probabilities of Landslide Events—A Case Study of Bhutan. *Hydrology* **2019**, *6*, 52. [CrossRef]

45. Gallant, J.C.; Wilson, J.P. Primary topographic attributes. In *Terrain Analysis: Principles and Applications*; Wilson, J.P., Gallant, J.C., Eds.; Wiley: New York, NY, USA, 2000; pp. 51–85.

46. Greenwood, L.V.; Argles, T.W.; Parrish, R.R.; Harris, N.B.W.; Warren, C. The geology and tectonics of central Bhutan. *J. Geol. Soc.* **2016**, *173*, 352–369. [CrossRef]

47. Giudici, P. Data Mining Model Comparison. In *Data Mining and Knowledge Discovery Handbook*; Springer: Boston, MA, USA. [CrossRef]

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

49. Saito, H.; Nakayama, D.; Matsuyama, H. Comparison of landslide susceptibility based on a decision-tree model and actual landslide occurrence: The Akaishi Mountains, Japan. *Geomorphology* **2009**, *109*, 108–121. [CrossRef]

50. Yeon, Y.-K.; Han, J.-G.; Ryu, K.H. Landslide susceptibility mapping in Injae, Korea, using a decision tree. *Eng. Geol.* **2010**, *116*, 274–283. [CrossRef]

51. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

52. Schratz, P.; Muenchow, J.; Iturritxa, E.; Richter, J.; Brenning, A. Performance evaluation and hyperparameter tuning of statistical and machine-learning models using spatial data. *arXiv* **2018**, arXiv:1803.11266.

53. Vapnik, V.N. Constructing learning algorithms. In *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 1995; pp. 119–166. [CrossRef]

54. Marjanovic, M.; Kovačević, M.; Bajat, B.; Vozenilek, V. Landslide susceptibility assessment using SVM machine learning algorithm. *Eng. Geol.* **2011**, *123*, 225–234. [CrossRef]

55. Sdao, F.; Lioi, D.S.; Pascale, S.; Caniani, D.; Mancini, I.M. Landslide susceptibility assessment by using a neuro-fuzzy model: A case study in the Rupestrian heritage rich area of Matera. *Nat. Hazards Earth Syst. Sci.* **2013**, *13*, 395–407. [CrossRef]

56. Dang, V.-H.; Hoang, N.-D.; Nguyen, L.-M.-D.; Bui, D.T.; Samui, P. A Novel GIS-Based Random Forest Machine Algorithm for the Spatial Prediction of Shallow Landslide Susceptibility. *Forests* **2020**, *11*, 118. [CrossRef]

57. Dikshit, A.; Pradhan, B.; Alamri, A.M. Pathways and challenges of the application of artificial intelligence to geohazards modelling. *Gondwana Res.* **2020**. [CrossRef]

58. Ayalew, L.; Yamagishi, H. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology* **2005**, *65*, 15–31. [CrossRef]

59. Jenks, G.F. The data model concept in statistical mapping. *Int. Yearb. Cartogr.* **1967**, *7*, 186–190.

60. Arabameri, A.; Saha, S.; Roy, J.; Chen, W.; Blaschke, T.; Bui, D.T. Landslide Susceptibility Evaluation and Management Using Different Machine Learning Methods in The Gallicash River Watershed, Iran. *Remote Sens.* **2020**, *12*, 475. [CrossRef]

61. Zuba, J.A.; Magirl, C.S.; Czuba, C.R.; Grossman, E.E.; Curran, C.A.; Gendaszek, A.S.; Dinicola, R.S. *Comparability of Suspended-Sediment Concentration and Total Suspended Solids DataSediment Load from Major Rivers into Puget Sound and its Adjacent Waters*; USGS Fact Sheet; US Geological Survey: Tacoma, WA, USA, 2011; pp. 2011–3083.

62. Peruccacci, S.; Brunetti, M.T.; Luciani, S.; Vennari, C.; Guzzetti, F. Lithological and seasonal control on rainfall thresholds for the possible initiation of landslides in central Italy. *Geomorphology* **2012**, *139*, 79–90. [CrossRef]

63. Reichenbach, P.; Rossi, M.; Malamud, B.D.; Mihir, M.; Guzzetti, F. A review of statistically-based landslide susceptibility models. *Earth-Sci. Rev.* **2018**, *180*, 60–91. [CrossRef]

64. El Jazouli, A.; Barakat, A.; Khellouk, R. Geotechnical studies for Landslide susceptibility in the high basin of the Oum Er Rbia river (Morocco). *Geol. Ecol. Landscapes* **2020**, *23*, 1–8. [CrossRef]

65. Viles, H.A. Linking weathering and rock slope instability: Non-linear perspectives. *Earth Surf. Process. Landf.* **2013**, *38*, 62–70. [CrossRef]

66. Frydman, S.; Talesnick, M.; Geffen, S.; Shvarzman, A. Landslides and residual strength in marl profiles in Israel. *Eng. Geol.* **2007**, *89*, 36–46. [CrossRef]

67. Heshmati, M.; Arifin, A.; Shamshuddin, J.; Majid, N.; Ghaituri, M. Factors affecting landslides occurrence in agro-ecological zones in the Merek catchment, Iran. *J. Arid. Environ.* **2011**, *75*, 1072–1082. [CrossRef]

68. Merghadi, A.; Yunus, A.P.; Dou, J.; Whiteley, J.; ThaiPham, B.; Bui, D.T.; Avtar, R.; Abderrahmane, B. Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth-Sci. Rev.* **2020**, *207*, 103225. [CrossRef]

69. Zhou, C.; Yin, K.; Cao, Y.; Ahmed, B.; Li, Y.; Catani, F.; Pourghasemi, H.R. Landslide susceptibility modeling applying machine learning methods: A case study from Longju in the Three Gorges Reservoir area, China. *Comput. Geosci.* **2018**, *112*, 23–37. [CrossRef]

70. Nhu, V.-H.; Mohammadi, A.; Shahabi, H.; Bin Ahmad, B.; Al-Ansari, N.; Shirzadi, A.; Clague, J.J.; Jaafari, A.; Chen, W.; Nguyen, H. Landslide Susceptibility Mapping Using Machine Learning Algorithms and Remote Sensing Data in a Tropical Environment. *Int. J. Environ. Res. Public Heal.* **2020**, *17*, 4933. [CrossRef]

71. Singh, A.; Thakur, N.; Sharma, A. A review of supervised machine learning algorithms. 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 16–18 March 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1310–1315.

72. Xiong, K.; Adhikari, B.R.; Stamatopoulos, C.A.; Zhan, Y.; Wu, S.; Dong, Z.; Di, B. Comparison of Different Machine Learning Methods for Debris Flow Susceptibility Mapping: A Case Study in the Sichuan Province, China. *Remote Sens.* **2020**, *12*, 295. [CrossRef]

73. Zhang, Y.; Ge, T.; Tian, W.; Liou, Y.-A. Debris Flow Susceptibility Mapping Using Machine-Learning Techniques in Shigatse Area, China. *Remote Sens.* **2019**, *11*, 2801. [CrossRef]

74. Bordoni, M.; Galanti, Y.; Bartelletti, C.; Persichillo, M.G.; Barsanti, M.; Giannecchini, R.; Avanzi, G.D.; Cevasco, A.; Brandolini, P.; Galve, J.P.; et al. The influence of the inventory on the determination of the rainfall-induced shallow landslides susceptibility using generalized additive models. *Catena* **2020**, *193*, 104630. [CrossRef]

75. Steger, S.; Brenning, A.; Bell, R.; Glade, T. The propagation of inventory-based positional errors into statistical landslide susceptibility models. *Nat. Hazards Earth Syst. Sci.* **2016**, *16*, 2729–2745. [CrossRef]

76. Pan, Z.; Yu, W.; Yi, X.; Khan, A.; Yuan, F.; Zheng, Y. Recent Progress on Generative Adversarial Networks (GANs): A Survey. *IEEE Access* **2019**, *7*, 36322–36333. [CrossRef]