

# Benchmarking landslide inventory data handling strategies for landslide susceptibility modeling based on different random forest machine learning workflows

Guruh Samodra (✉ [guruh.samodra@ugm.ac.id](mailto:guruh.samodra@ugm.ac.id))

Faculty of Geography Universitas Gadjah Mada <https://orcid.org/0000-0002-4094-8846>

Ferman Setia Nugroho

LAPAN: Lembaga Penerbangan dan Antariksa Nasional

---

## Research Article

**Keywords:** landslide, machine learning, random forest, sampling strategies, spatial CV, hyperparameter tuning, susceptibility

**Posted Date:** April 12th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1441095/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

Machine learning (ML) algorithms are frequently used in landslide susceptibility modeling. Different data handling strategies may generate variations in landslide susceptibility modeling, even when using the same ML algorithm. This research aims to compare the combinations of inventory data handling, cross validation (CV), and hyperparameter tuning strategies to generate landslide susceptibility maps. The results are expected to provide a general strategy for landslide susceptibility modeling using ML techniques. The authors employed eight landslide inventory data handling scenarios to convert a landslide polygon into a landslide point, i.e., the landslide point is located on the toe (minimum height), on the scarp (maximum height), at the center of the landslide, randomly inside the polygon (1 point), randomly inside the polygon (3 points), randomly inside the polygon (5 points), randomly inside the polygon (10 points), and 15 m grid sampling. Random forest models using CV–nonspatial hyperparameter tuning, spatial CV–nonspatial hyperparameter tuning, and spatial CV–forward feature selection–no hyperparameter tuning were applied for each data handling strategy. The combination generated 24 random forest ML workflows, which are applied using a complete inventory of 743 landslides triggered by Tropical Cyclone Cempaka 2017 in Pacitan Regency, Indonesia, and 11 landslide controlling factors. The results show that grid sampling with spatial CV and spatial hyperparameter tuning is favorable because the strategy can minimize overfitting, generate a relatively high-performance predictive model, and reduce the appearance of susceptibility artifacts in the landslide area. Careful data inventory handling, CV, and hyperparameter tuning strategies should be considered in landslide susceptibility modeling to increase the applicability of landslide susceptibility maps in practical application.

## 1. Introduction

Landslides commonly cause numerous casualties and considerable property damage in hilly and mountainous areas, and landslide occurrence increases because of significant population growth in prone areas. Practical landslide disaster risk reduction tool is important for stakeholders in the landslide prone areas. A landslide susceptibility map that can be generated from landslide susceptibility modeling is considered as one of the tools that can be used by municipalities to design risk-reduction-based development (Fell et al., 2008). The most appropriate method for landslide susceptibility modeling that covers large areas on medium to small scales is statistical quantitative approaches (Cascini, 2008; Soeters and Westen, 1996; van Westen et al., 2008), which need a reliable landslide inventory (Blahut et al., 2010).

A landslide inventory is the simplest form of landslide map (Guzzetti et al., 2012; Ngadisih et al., 2017). The landslide inventory generally represents single or multiple events, which describe the locations and outlines of landslides (Chacón et al., 2006). The landslide inventory should cover information about the location, type, volume, activity, date of occurrence, and other characteristics of landslides in the area (Fell et al., 2008). Furthermore, the landslide inventory must identify and illustrate the triggering factors (Godt et al., 2008) as a single point or polygons of the landslides (Parise, 2001). For example, in Indonesia, landslide data are often represented by point or location coordinates, which are available at <https://dibi.bnpb.go.id/>. Several researchers have created a landslide polygon inventory for some areas (Aditian et al., 2018; Ngadisih et al., 2014; Samodra et al., 2020, 2018) for a specific purpose. Landslide inventory maps and landslide controlling factor maps are required for landslide susceptibility modeling.

A susceptibility map shows the likelihood of landslide occurrence in a given location (Corominas and Moya, 2008). Numerous landslide susceptibility models have been recently developed, and the methods are increasingly sophisticated because of the advancement of computer processing and geographic information system (GIS) technologies. The common procedure for landslide susceptibility modeling is to extract information on each landslide controlling factor from the landslide and non-landslide point/area samples (Lovelace et al., 2020). The differences in landslide inventory, i.e., point or polygon, can cause differences in landslide susceptibility data handling in the model. Landslide inventory data handling, i.e., the placement of landslide samples in a landslide polygon and the number of samples, can affect the appearance and performance of a landslide susceptibility map (Abraham et al., 2021; Hussin et al., 2016; Steger et al., 2016a).

Remote sensing and GIS technologies are becoming increasingly sophisticated and integrated, contributing to the advancement of landslide susceptibility modeling. Statistical learning and machine learning (ML) techniques, such as logistic regression, artificial neural network, decision tree, random forest, support vector machine, and their variations, are the most common methods adopted for landslide susceptibility modeling recently (Merghadi et al., 2020; Reichenbach et al., 2018). Among these methods, random forest is ranked as the most promising ML algorithm for spatial prediction (Hengl et al., 2015; Nussbaum et al., 2018; Park and Kim, 2019; Sun et al., 2021; Vaysse and Lagacherie, 2015).

Even though it exhibits high predictive performance, the ML algorithm for spatial prediction is often prone to overfitting (Jaafari et al., 2019; Just et al., 2020; Meyer et al., 2019, 2018). Overfitting occurs when the model performs well for the training data but probably worse for the independent data (Probst et al., 2019). Many researchers rarely considered the effect of overfitting in their ML models (Kim et al., 2017; Lai et al., 2019; Park and Kim, 2019; Taalab et al., 2018). The evaluation of the training data, particularly spatial cross validation (CV) and spatial hyperparameter tuning, is recommended to avoid overfitting in the ML algorithm (Brenning, 2005; Meyer et al., 2019, 2018; Schratz et al., 2019).

Instead of comparing algorithms, our research focuses on comparing the best strategies to handle landslide inventory data for landslide susceptibility modeling. This study focuses on handling landslide inventory with scenarios that often exist in Indonesia. That is, for point-based data handling, the landslide point is placed randomly on the landslide polygon, the landslide point is placed on the scarp (maximum height), the landslide point is placed on the toe (minimum height), and the landslide point is placed at the center (centroid). For area-based/polygon data handling, the landslide points are often placed in all landslide areas based on grid sampling. In this research, a 15 m grid sample was created in each landslide polygon. The authors also created strategies for randomly generating 3x, 5x, and 10x points in a landslide polygon to compensate for the number of points made based on grid sampling. Eight inventory data handling scenarios will be simulated in landslide susceptibility maps using nonspatial CV for performance estimation combined with nonspatial hyperparameter tuning (CV–nonspatial tuning), spatial CV estimation combined with nonspatial hyperparameter tuning (spatial CV–nonspatial tuning), and spatial CV combined with forward feature selection (FFS) estimation without tuning (spatial CV–FFS–no tuning). The comparison of the combinations of inventory data handling, CV, and hyperparameter tuning strategies is expected to provide general strategies for landslide susceptibility modeling using ML techniques.

## 2. Study Area

The study area is located in the Pacitan Regency, East Java Province, Indonesia (Fig. 1a). The area encompasses a total area of 1,390 km<sup>2</sup>, with a relief of 1,226.5 m. Flat morphology is located in the southern part of the area, with a narrow bay shape and an elevation ranging from 0 m to 10 m. Hilly and mountainous morphologies are dominant in the study area, with the slope ranging from 15° to 70°. Steep topography is mainly located in the Grindulu River valley, with strong dissection and a “V” shape because of high weathering, erosion, and landslide.

The geological setting of the area is affected by the formation of arc volcanism during the middle Eocene to the early Oligocene, termination of arc volcanism during the late Oligocene to the early Miocene, growth of carbonate during the middle Miocene, and uplifting followed by denudational processes during the Pliocene to the recent age (Smyth et al., 2008). The Pacitan Regency consists of Dayakan Formation (sandstone and claystone), Mandalika Formation (volcanic breccia and lava tuff, with intercalations of sandstone and siltstone), Watupatok Formation (basaltic pillow lava, sandstone, claystone, and cherts), Arjosari Formation (polymict breccia, sandstone, and conglomerate, with intercalations of volcanic and intrusive rocks), Semilir Formation (tuff, breccia, sandstone, and claystone), Jaten Formation (conglomerate, sandstone, mudstone, lignite, shale, and tuff), Wuni Formation (volcanic breccia, tuff, and sandstone, with intercalations of lignites and limestone), Nampol Formation (sandstone, siltstone, limestone, claystone, and lignite, with intercalations of conglomerate and breccia), Oyo Formation (sandstone, siltstone, limestone, and marl), Wonosari Formation (reef limestone, bedded limestone, sandy limestone, and marl), and Kalipucang Formation (conglomerate, clay, and alluvium) (Fig. 1b) (Samodra et al., 1992; Sampurno and Samodra, 1997). A geological formation map is used as one of the landslide controlling factor maps employed to model landslide susceptibility.

Climate effects, rugged terrains, and geological conditions make the study area generally prone to landslides. The Pacitan Regency was severely affected by Tropical Cyclone (TC) Cempaka that occurred on November 27 to 29, 2017. The cumulative rainfall of 235.97 mm in one and a half day triggered 743 landslides in the Pacitan Regency during TC Cempaka 2017 (Samodra et al., 2020).

Figure 1

## 3. Methods

### 1.1 Landslide inventory and its data handling

The landslide inventory of TC Cempaka 2017 is represented as landslide polygons without separation of scarp, body, and toe. A total of 743 landslides were obtained from on-screen digitizing by comparing pre-event and post-event high-resolution satellite imageries and conducting field surveys. Atrium's Pleiades Pansharpened Multispectral Natural Color Band Imagery was employed for post-event landslide inventory mapping. The spatial resolution of the images is 50 cm with the date of acquisition ranging from March 15 to 30, 2018.

Most statistical learning and ML models employed for landslide susceptibility modeling require point sampling instead of polygons to generate samples. We employed eight landslide inventory data handling scenarios to convert a landslide polygon into a landslide point, i.e., the landslide point is located on the toe (minimum height), on the scarp (maximum height), at the center of the landslide, randomly inside the polygon (1 point), randomly inside the polygon (3 points), randomly inside the polygon (5 points), randomly inside the polygon (10 points), and each 15 m grid (Fig. 2). As a consequence, the number of datasets is different among scenarios. The scenarios of minimum, maximum, center, and random 1 point per landslide (1pt/l) generate the same number

dataset, i.e., 1,486 points consisting of 743 landslide points and 743 non-landslide points. The total number of datasets is 4,458, 7,430, 14,860, and 19,190 for random 3 pts/l, random 5 pts/l, random 10 pts/l, and 15 m grid, respectively. The non-landslide points are created randomly outside the landslide polygon. The number of non-landslide points is equal to the number of landslide points in each scenario. We applied 75% (training) and 25% (testing/success rate) data splitting for each scenario.

Figure 2

## 1.2 Landslide controlling factors

The Pacitan Regency consists of mountainous terrains characterized by unstable geological formations, humid tropical climate associated with heavy rainfall, high seismicity, and intense anthropogenic activities, which can affect landslide occurrence. Landslide susceptibility modeling using ML algorithms employed landslide inventory points and landslide controlling factors as proxies of terrain characteristics to train the model. In this research, the landslide susceptibility map was created at a 1:100,000 scale, which is applicable for information and advisory purposes (Cascini, 2008). Therefore, the selection of the cartographic scale and spatial resolution of the base maps for creating the landslide controlling factor maps was based on the procedure proposed by Hengl, (2006).

The landslide controlling factors maps were derived from the National Digital Elevation Model of Indonesia (DEMNAS) at approximately 10 m spatial resolution, which is available at <https://tanahair.indonesia.go.id/demnas/#/>, and digital topographical map at a 1:25,000 scale, which is available at <https://tanahair.indonesia.go.id/portal-web>. In this research, landslide controlling factors were represented by digital terrain attributes, such as slope, aspect, distance to the river, distance to the road, elevation, land use, plan curvature, profile curvature, slope, stream power index (SPI), terrain wetness index (TWI) (Fig. 3), and geological formation (Fig. 1). The digital terrain attributes were processed from DEMNAS with System for Automated Geoscientific Analysis (SAGA) GIS (Conrad et al., 2015).

The hypothetical significances of the landslide controlling factors to landslide are briefly explained as follows: Elevation represents the local relief and locates the landslide points with maximum and minimum heights. Slope represents the balance between shear strength and shear stress acting on the slope. Aspect is the direction of the slope, which can reflect the differences in the degree of weathering and soil moisture related to solar insolation. Curvature reflects the slope form, which can affect the directions of surface water and subsurface groundwater flows. SPI represents the strength of stream power calculated from the catchment area and the steepness of the slope. The wide area and steep slope generate a high stream power, which means a large amount of water and a high velocity of water flow. TWI reflects the soil water moisture or tendency of the slope to accumulate water (Beven and Kirkby, 1979; Moore et al., 1991; Quinn et al., 1991). Distance to the river as a landslide controlling factor assumes that a shorter distance to the river will lead to more landslides because of the steep slope and erosion. Land use and distance to the road as landslide controlling factors assume that human activities can increase the instability of the slope. Geological formation represents the strength of the material in the study area. The selection of landslide controlling factors in this research was mainly knowledge driven. The FFS method was also employed to quantitatively select the most significant landslide controlling factors and remove irrelevant landslide controlling factors in the model.

Figure 3

## 1.3 Random Forest algorithm

Random forest is a powerful ML algorithm developed by constructing several decision trees through random samplings and combining multiple decision trees for classification and prediction purposes (Breiman, 2001). In this research, the algorithm starts by randomly drawing a number of tree ( $n_{tree}$ ) samples (landslide data point in controlling factor pixels), growing an unpruned classification tree by randomly sampling the number of controlling factors as a candidate at each split ( $m_{try}$ ) and choosing the best split, and predicting new data by aggregating the predictions of  $n_{tree}$  by majority vote (Liaw and Wiener, 2002). Both  $n_{tree}$  and  $m_{try}$  are considered the hyperparameters that need to be tuned to obtain the optimal prediction. The randomForest package (Liaw and Wiener, 2002) accessed via the caret package (Kuhn, 2021) in the R environment for statistical programming (R Core Team, 2020) was employed to model landslide susceptibility in the Pacitan Regency.

## 1.4 Validation strategies, controlling factor selection, and hyperparameter tuning

In this study, resampling techniques, i.e., nonspatial and spatial CV, were used as validation strategies to estimate the model performance/success rate. Nonspatial CV applied the commonly used random 10-fold CV, in which the samples were partitioned randomly into 10 sets/folds of roughly equal size. Models were repeatedly fit using all samples, except for the first fold, which was subsequently used to estimate performance measures. Spatial CV (Brenning, 2005) or leave-location-out (LLO) CV (Meyer et al., 2018) was applied by partitioning 10-fold data spatially. The illustration of nonspatial and spatial CV is shown in Fig. 4. For a detailed discussion of the effects of nonspatial and spatial CV applied in ML, see the works of Meyer et al. (2018) and Meyer et al. (2019). Both nonspatial CV and spatial CV were applied to each scenario of landslide inventory data handling. We also tested spatial CV combined with the FFS method.

FFS was used to remove the landslide controlling factors from the model to obtain the best model performance. The algorithm initially trains the random forest using every combination of two landslide controlling factors and iteratively increases the number of combinations until none of the remaining landslide controlling factors decreases the performance of the current best model. This study employed FFS, which was implemented in the CAST package (Meyer et al., 2018). The variable importance score, which was calculated from the aggregation of the Gini index across the ensemble trees (Kuhn and Johnson, 2013), was applied to interpret the influence of landslide controlling factors on landslides.

Hyperparameter tuning was applied to optimize the performance of random forest by reducing the bias assessment of the model's predictive power (Schratz et al., 2019). The grid search method (Bergstra and Bengio, 2012) implemented in the caret package (Kuhn, 2021) was applied to tune both nonspatial and spatial CV strategies. Therefore, we applied nonspatial CV for performance estimation combined with nonspatial hyperparameter tuning (CV–nonspatial tuning), spatial CV estimation combined with nonspatial hyperparameter tuning (spatial CV–nonspatial tuning), and spatial CV combined with FFS estimation without tuning (spatial CV–FFS–no tuning). The combinations of landslide inventory data handling, CV, and hyperparameter tuning strategies generated 24 random forest ML workflows.

Figure 4

## 1.5 Performance measure

The performance measures for the landslide susceptibility model were the accuracy index, kappa index (Cohen, 1960), and area under the receiver operating characteristic (ROC) curve (AUROC). The accuracy and kappa indices

were applied to 75% of the dataset during CV and hyperparameter tuning, i.e., the measures calculated the average performance over all 10-fold CVs. The AUROC and confusion matrix plot were derived by comparing the models with the remaining 25% of the independent dataset. In this research, the accuracy and kappa indices applied to 75% of the dataset were considered the success rate and the AUROC and confusion matrix plot applied to 25% of the dataset were considered the predictive rate (Chung and Fabbri, 2003).

The accuracy index was calculated using the confusion matrix of the two-class problem as the agreement between observed and predicted classes, as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

1

,

where TP is the true positive, TN is the true negative, FP is the false positive, and FN is the false negative. The kappa index was also calculated using the confusion matrix, as follows:

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

2

,

$$p_o = \frac{TP + TN}{TP + TN + FP + FN}$$

3

,

$$p_e = \frac{(TP + FN)(TP + FP) + (FP + TN)(FN + TN)}{(TP + TN + FP + FN)^2}$$

4

,

where  $p_o$  is the observed accuracy and  $p_e$  is the expected accuracy. The kappa statistic can take a value between -1 and 1. Negative values represent the opposite direction of the truth (Kuhn and Johnson, 2013). The values 0–0.2, 0.21–0.39, 0.4–0.59, 0.6–0.79, 0.8–0.9, and > 0.9 represent none, minimal, weak, moderate, strong, and almost perfect agreement between observed and predicted classes (McHugh, 2012).

The measure considered the apparent predictive accuracy was calculated by comparing the models with the remaining 25% of the independent dataset using the AUROC. The authors called it “apparent” predictive accuracy because of the unavailability of posterior landslide inventory to calculate the true predictive performance. The ROC curve was plotted based on all of the possible true positive rates (TPR; sensitivity) and the corresponding false positive rates (FPR, 1 – specificity). AUROC values close to 100% indicate an effective model for

discriminating two classes, and AUROC values close to 50% indicate an ineffective model representing no discrimination for the two-class problem.

## 4. Result

### 4.1. Best hyperparameter settings

The range of optimal hyperparameter settings of  $m_{try}$  in nonspatial tuning was larger than that in spatial tuning. The value of  $m_{try}$  in nonspatial tuning mainly ranged between 8 and 10, whereas the value of  $m_{try}$  in spatial tuning mainly ranged between 8 and 10 and between 2 and 3. The range of optimal  $n_{tree}$  in nonspatial tuning mainly ranged between 550 and 2,000, whereas the range of optimal  $n_{tree}$  in spatial tuning mainly ranged between 350 and 2,000. The range of optimal  $n_{tree}$  in spatial tuning of more than one sample in a landslide polygon mainly ranged between 1,000 and 2,000. The range of optimal hyperparameter settings of  $n_{tree}$  in one sample in a landslide polygon is smaller than that in more than one sample for both nonspatial and spatial tuning (Table 1).

Table 1  
Database system of the landslide inventory in the app

Data handling	CV–nonspatial tuning		CV LLO–spatial tuning		CV LLO FFS–no tuning	
	$m_{try}$	$n_{tree}$	$m_{try}$	$n_{tree}$	$m_{try}$ (fixed)	$n_{tree}$
Minimum	2	550	2	350	2	Default
Maximum	8	550	8	350	2	Default
Center	9	550	10	600	2	Default
Random 1 pt.	8	250	3	300	2	Default
Random 3 pt.	8	600	2	2,000	2	Default
Random 5 pt.	4	2,000	2	2,000	2	Default
Random 10 pt.	9	600	9	2,000	2	Default
15 m grid	10	1,000	10	1,000	2	Default

Table 1

### 4.2. Success rate and predictive performance

The variation of success rate between the models was relatively low for different validation and hyperparameter tuning strategies. The accuracy and kappa of CV–nonspatial tuning and CV LLO FFS–no tuning were relatively higher than that of CV LLO–spatial tuning. The variation of success rate for different validation and hyperparameter tuning strategies was relatively low, ranging from 0.01 to 0.06 for accuracy and from 0.01 to 0.11 for kappa (Table 2).

Table 2

Success rate and apparent predictive rate of the landslide susceptibility model using different landslide inventory data handling strategies, validation strategies, and hyperparameter tuning

Data handling	CV–nonspatial tuning			CV LLO–spatial tuning			CV LLO FFS–no tuning		
	Accuracy	Kappa	AUROC	Accuracy	Kappa	AUROC	Accuracy	Kappa	AUROC
Minimum	0.71	0.42	0.78	0.70	0.41	0.78	0.71	0.42	0.78
Maximum	0.70	0.40	0.73	0.70	0.40	0.72	0.71	0.42	0.70
Center	0.70	0.39	0.74	0.69	0.38	0.74	0.70	0.41	0.73
Random 1 pt.	0.66	0.31	0.72	0.66	0.32	0.73	0.68	0.36	0.71
Random 3 pt.	0.77	0.54	0.86	0.77	0.54	0.86	0.79	0.59	0.88
Random 5 pt.	0.81	0.63	0.92	0.81	0.63	0.92	0.85	0.70	0.95
Random 10 pt.	0.87	0.73	0.95	0.84	0.69	0.95	0.90	0.80	0.98
15 m grid	0.87	0.74	0.95	0.83	0.66	0.95	0.88	0.77	0.97

The variation of success rate between the models was relatively high for different data handling strategies. The lowest values of accuracy and kappa for random 1 point data handling with CV–nonspatial tuning were 0.66 and 0.31, respectively. The highest values of accuracy and kappa for random 10 point data handling with CV LLO FFS–no tuning were 0.9 and 0.8, respectively. The highest differences of accuracy and kappa for 1 and > 1 handling strategies were 0.22 and 0.43, respectively. Generally, the larger the number of samples generated, the higher accuracy and kappa values.

Table 2

The variation of predictive rate (AUROC) was relatively similar to the variations of accuracy and kappa, which were low for CV and tuning strategies but significant for different data handling strategies (Fig. 5). The AUROC of CV–nonspatial tuning was relatively similar to that of CV LLO–spatial tuning. By contrast, the AUROC of CV LLO FFS–no tuning was slightly lower than both of them and slightly higher than both of them for 1 and > 1 point sampling strategies, respectively. For 1 point sampling strategies, the lowest and highest AUROC were 0.7 and 0.78, respectively. By contrast, for > 1 point sampling strategies, the lowest and highest AUROC were 0.86 and 0.98, respectively. The highest success rate for 1 and > 1 point sampling strategies obtained using the model was calculated by locating the point samples in the toe of the landslide, the minimum height of a landslide polygon, or a random location with a total of 10 samples. The 15 m grid sampling strategy also achieved the highest AUROC similar to the random 10 point data handling strategy, except for CV LLO FFS–no tuning. CV LLO–spatial tuning with 15 m grid sampling was determined to have the highest accuracy using the confusion matrix (Fig. 5b). Notably, the variations of CV and hyperparameter tuning strategies in the random forest were not significantly different from the variation of AUROC. By contrast, the variations of different landslide inventory data handling strategies were significantly different from the variation of AUROC.

Figure 5

## 4.3. Susceptibility map appearance

The 10 classes of final susceptibility maps (Fig. 6) were classified using the equal interval classification method to represent the differences of susceptibility map appearance considering different landslide inventory data handling strategies, validation strategies, and hyperparameter tuning. The susceptibility maps obtained using different validation strategies and hyperparameter tuning did not show significant variation of map appearance. Meanwhile, different landslide inventory data handling strategies showed significant variation of map appearance. The landslide susceptibility map obtained using 1 point sampling in each landslide polygon showed a high proportion of yellow color representing a pixel value close to 0.5 (Figs. 6a to 6l). By contrast, the landslide susceptibility map obtained using > 1 point sampling in each landslide polygon showed a low proportion of yellow color representing a pixel value close to 0.5, which indicates the inability of the classification model to determine whether the pixel is stable or unstable (Reichenbach et al., 2018) and has higher uncertainty than the pixel that has a value close to 0 or 1 (Guzzetti et al., 2006; Rossi et al., 2010; Van Den Eeckhaut et al., 2009). Landslide susceptibility modeling using > 1 point sampling in each landslide polygon has lower uncertainty than that using 1 point sampling. However, landslide susceptibility modeling using random 1, 3, 5, and 10 points showed abrupt changes in map appearance related to landslide controlling factors (i.e., distance to the river and geological formation) in the western part of the study area (Figs. 6j to 6u).

Figure 6

The superimposed landslide polygon map and landslide susceptibility map (Fig. 7) showed that the landslide model obtained using grid sampling (Figs. 7v, 7w, and 7x) successfully classified all landslide polygons with values ranging between 0.8 and 0.99. By contrast, the landslide model obtained using 1 point sampling classified landslide polygons with values ranging between 0.2 and 0.8 (Figs. 7a, 7b, and 7c). The high pixel values of landslide susceptibility calculated from the 1 point sample located in the minimum height were associated with a lower slope and highly corresponded to two landslide controlling factors (i.e., curvature and TWI).

Figure 7

## 4.4. Landslide controlling factors importance

The ranking of landslide controlling factor importance was relatively different for all different landslide inventory data handling strategies, validation strategies, and hyperparameter tuning (Appendix 1). Slope and geological formation were the only controlling factors that were ranked in the Top 6 for the models that did not apply FFS. By contrast, aspect and geological formation were the only controlling factors that were ranked in the Top 5 for the models that applied FFS. Geological formation had the highest variable importance score in all situations, indicating that this landslide controlling factor played an important role in predicting landslide susceptibility within each model. The most consistently lower-ranked controlling factor (variable importance score < 25) was profile curvature.

## 5. Discussion

### 5.1. Success rate, overfitting, and map appearance

The landslide susceptibility map should be generated from a model that can handle overfitting. Overfitting in the landslide susceptibility model leads to a high performance on training data but a low performance on testing data and such problems can be avoided by the spatial CV and hyperparameter tuning strategies (Probst et al., 2019). In this research, we applied nonspatial CV–nonspatial tuning, spatial CV–nonspatial tuning, and spatial CV–FFS–no tuning to detect overfitting in the random forest model. Generally, random CV–nonspatial tuning generated slight overfitting with a higher accuracy and kappa but a lower or the same AUROC than that of spatial CV–nonspatial tuning and spatial CV–FFS–no tuning. However, we cannot conclude that the use of FFS is favorable because it generated a high AUROC for > 1 point sampling but generated a low AUROC for 1 point sampling. Hence, the spatial CV with spatial tuning strategy is more favorable than the nonspatial CV with nonspatial tuning strategy. Meyer et al. (2019) and Schratz et al. (2019) also recommended the use of spatial CV with spatial tuning to minimize overfitting in the model. As the aim of such a model is to predict the landslide occurrence beyond sampling locations, spatial CV is preferable to nonspatial CV (Ploton et al., 2020). Spatial CV generates a low success rate (accuracy and kappa) but a high predictive performance (AUROC). The variation of success rate is also affected by the number of samples used in the model.

The success rate of the random forest algorithm for landslide susceptibility modeling is more sensitive to landslide inventory data handling strategies than CV and hyperparameter tuning strategies. The accuracy, kappa, and AUROC consistently show that more data samples generate a higher success rate. The 1 point data sampling in a landslide polygon is likely to generate a low success rate. The random 1 point sampling generates the lowest success rate, and the 1 point sample located in the lowest elevation or the toe generates the highest performance for 1 point sampling strategies. Locating the 1 point sample in a landslide polygon may have the consequence that the non-landslide sample is located inside the landslide polygon and can affect the landslide susceptibility map appearance.

The landslide susceptibility maps generated from different models and data handling strategies must be carefully analyzed and critically reviewed (Sterlacchini et al., 2011) based on their appearance. The landslide susceptibility map appearance was reviewed based on the geomorphic plausibility check (Steger et al., 2016b) and analyzed based on the superimposed landslide polygon and susceptibility map (Figs. 8a and 8b). The 1 point sample generated more artifacts in the landslide polygon than the > 1 point sample. A high susceptibility value is highly correlated to the location of the point sample, e.g., the 1 point sample located in the minimum height generates a high value on the lower slope, gully, or valley. By contrast, the 1 point sample located in the maximum height generates a high susceptibility value on the upper slope. The use of the 1 point sample failed to predict all landslide polygon areas (Fig. 8a). The use of the > 1 point sample located randomly generates a better result, as more than half of landslides is classified as 0.8 to 1. However, a small artifact still exists inside the landslide polygon (Figs. 7m to 7u). Point samples based on the grid generates better results because all landslide polygons are classified as 0.8 to 1 (Fig. 8b).

In this study, the applicability of landslide susceptibility maps based on the random forest was investigated using various data handling, CV, and hyperparameter tuning strategies. This study shows that different data handling and treatment strategies lead to different results of landslide susceptibility map and performance. Given the success rate, predictive performance, and map appearance, grid sampling using spatial CV and spatial hyperparameter is favorable for landslide susceptibility modeling using random forest. However, the predictive performance was calculated by comparing the model with independent landslide data triggered by TC Cempaka 2017, which is called by the authors as apparent predictive accuracy. The use of posterior landslide data with

different times is required to test the real predictive accuracy of the landslide susceptibility map. Multitemporal landslide inventory enables the generation of susceptibility maps with both spatial and temporal CV, which may not only reduce overfitting but also generate considerably high predictive performance.

Figure 8

## 5.2. Comparing controlling factors importance

The use of feature selection in landslide susceptibility modeling eliminates the irrelevant controlling factors, resulting in simpler and lower-dimensional models but a high success rate (Micheletti et al., 2013). This research shows that the FFS technique slightly decreases the AUROC for 1 point samples but increases the AUROC for > 1 point samples. Moreover, the FFS technique reduces artifacts but its TPR is slightly lower than that of CV LLO–spatial tuning. The FFS computation is also time-consuming compared with the other strategies. The use of FFS and spatial validation is strongly recommended for the model with controlling factors, leading to a strong random performance; however, FFS and spatial validation cannot predict any dataset other than the training dataset (Meyer et al., 2019). Thus, the selection of appropriate and relevant controlling factors is important.

Slope, aspect, and geology were the factors that influenced the landslide events triggered by TC Cempaka 2017. Slope and material strength are well understood to be the landslide controlling factors. However, the mechanism of slope as a landslide controlling factor is poorly understood. Capitani et al. (2013) investigated the influence of slope as a landslide controlling factor in some basins in Italy and determined that slope only works for superficial and clayey deposit. Slope could also reflect the differences in soil moisture and vegetation, which are related to solar insolation and evapotranspiration. Slope also played an important role in landslides triggered by earthquakes. For example, the landslides triggered by the 2018 Hokkaido Earthquake were mostly located on the west-facing slope because of the mechanism of velocity pulse-like ground motion by earthquakes and strength parameters of the slope (Chen et al., 2021). Most landslides triggered by TC Cempaka 2017 were located in the north-facing and east-facing slopes and affected the map appearance, which could be related to the rainfall pattern caused by TC Cempaka 2017 and the degree of weathering.

## 5.3. Landslide data handling and landslide magnitude.

The representation of landslide features in maps leads to different data handling strategies and scales. GIS representation of landslides through discrete-object conceptualization and continuous-field conceptualization generates vector and raster data, respectively. Landslides, as a geomorphological form and process, are scale-specific (Evans, 2003) and are not scale-free. Landslides are often represented as polygons and points on the medium to large and small landslide inventory maps, respectively. The common procedure for landslide susceptibility modeling based on ML is to overlay the landslide inventory points with the landslide controlling factors. The conversion of polygons to points generates a different number of points in each polygon depending on the methods. The use of different numbers of landslide samples generates significantly different results.

For 1 point samples, we created a landslide susceptibility using 743 landslide points and 743 non-landslide points, which we increased 3, 5, and 10 times. Finally, using grid samples, we obtained 19,190 landslide points and 19,190 non-landslide points. The results show that a high number of samples significantly improves the success rate. However, the study conducted by Dou et al. (2020) using landslide data from Japan and Nepal indicated that different sampling strategies do not affect the success rate. They used 10,120 landslides triggered by the 2018 Hokkaido Earthquake and validated using 24,915 points of the 2015 Gorkha Earthquake landslide

inventory (Roback et al., 2018). We argue that landslide points > 10,000 will not affect the success rate, whereas landslide points < 10,000 will affect the success rate. We recommend the use of grid sampling to create landslide susceptibility maps with landslide samples < 10,000 points. Further investigation on the use of real landslide inventory data with different levels of frequency and magnitude in different geomorphological settings is required.

## 6. Conclusion

Different strategies for landslide susceptibility modeling, i.e., inventory data handling, CV, and hyperparameter tuning, can generate landslide susceptibility maps with a high variation. The authors tested 24 random forest ML workflows of landslide susceptibility modeling by varying the inventory data handling, CV, and hyperparameter tuning strategies. The variation of strategies generates different success rate, predictive performance, and map appearance. Grid sampling with CV LLO–spatial tuning (spatial CV–spatial hyperparameter tuning) is favorable because the strategy can minimize overfitting, generate a relatively high-performance predictive model, and reduce the appearance of susceptibility artifacts in the landslide area.

The application of the ML algorithm and its variation in landslide susceptibility modeling is still developing because of the advancement of computer processing and GIS technologies. The use of an advanced ML algorithm should be carefully and critically applied to avoid developing a model that can have a high success rate but does not make a reliable spatial prediction. Careful data inventory handling, CV, and hyperparameter tuning should be considered to increase the applicability of landslide susceptibility maps in practical application. Otherwise, the practical application landslide susceptibility maps cannot be confidently pursued because it would produce misleading results.

## Declarations

### Acknowledgment

The authors thank Maulana Yudinugroho and Ghalih N. Wicaksono for the drone survey in Pacitan Regency.

### Author Contribution

Guruh Samodra: Conceptualization, Methodology, Software, Formal Analysis, Visualization, Writing – Original Draft, Funding acquisition

Ngadisih: Formal Analysis, Writing – Original Draft , Writing – Review & editing

Ferman Setia Nugroho: Resources, Writing – Review & editing

### Compliance with Ethical Standards:

No funding was received for this research

### Conflict of Interest:

All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

## References

1. Abraham MT, Satyam N, Lokesh R, Pradhan B, Alamri A (2021) Factors Affecting Landslide Susceptibility Mapping: Assessing the Influence of Different Machine Learning Approaches, Sampling Strategies and Data Splitting. *L.* 2021, Vol. 10, Page 989 10, 989. <https://doi.org/10.3390/LAND10090989>
2. Aditian A, Kubota T, Shinohara Y (2018) Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of Ambon. *Indonesia Geomorphology* 318:101–111. <https://doi.org/10.1016/j.geomorph.2018.06.006>
3. Bergstra J, Bengio Y (2012) Random Search for Hyper-Parameter Optimization Yoshua Bengio. *J Mach Learn Res* 13:281–305
4. Beven KJ, Kirkby MJ (1979) A physically based, variable contributing area model of basin hydrology. *Hydrol Sci Bull* 24:43–69. <https://doi.org/10.1080/02626667909491834>
5. Blahut J, van Westen CJ, Sterlacchini S (2010) Analysis of landslide inventories for accurate prediction of debris-flow source areas. *Geomorphology* 119:36–51. <https://doi.org/10.1016/J.GEOMORPH.2010.02.017>
6. Breiman L (2001) Random forests. *Mach Learn* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
7. Brenning A (2005) Spatial prediction models for landslide hazards: Review, comparison and evaluation. *Nat Hazards Earth Syst Sci* 5:853–862. <https://doi.org/10.5194/NHESS-5-853-2005>
8. Capitani M, Ribolini A, Bini M (2013) The slope aspect: A predisposing factor for landsliding? *Comptes Rendus Geosci* 345:427–438. <https://doi.org/10.1016/J.CRTE.2013.11.002>
9. Cascini L (2008) Applicability of landslide susceptibility and hazard zoning at different scales. *Eng Geol* 102:164–177. <https://doi.org/10.1016/j.enggeo.2008.03.016>
10. Chacón J, Irigaray C, Fernández T, El Hamdouni R (2006) Engineering geology maps: Landslides and geographical information systems. *Bull Eng Geol Environ.* <https://doi.org/10.1007/s10064-006-0064-z>
11. Chen G, Xia M, Thuy DT, Zhang Y (2021) A possible mechanism of earthquake-induced landslides focusing on pulse-like ground motions. *Landslides* 2021 185 18:1641–1657. <https://doi.org/10.1007/S10346-020-01597-Y>
12. Chung C-JF, Fabbri AG (2003) Validation of Spatial Prediction Models for Landslide Hazard Mapping. *Nat Hazards* 2003 303 30:451–472. <https://doi.org/10.1023/B:NHAZ.0000007172.62651.2B>
13. Cohen J (1960) A Coefficient of Agreement for Nominal Scales. *Educ Psychol Meas* 20:37–46. <https://doi.org/10.1177/001316446002000104>
14. Conrad O, Bechtel B, Bock M, Dietrich H, Fischer E, Gerlitz L, Wehberg J, Wichmann V, Böhner J (2015) System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geosci. Model Dev* 8:1991–2007. <https://doi.org/10.5194/GMD-8-1991-2015>
15. Corominas J, Moya J (2008) A review of assessing landslide frequency for hazard zoning purposes. *Eng Geol* 102:193–213
16. Dou J, Yunus AP, Merghadi A, Shirzadi A, Nguyen H, Hussain Y, Avtar R, Chen Y, Pham BT, Yamagishi H (2020) Different sampling strategies for predicting landslide susceptibilities are deemed less consequential with deep learning. *Sci Total Environ* 720:137320. <https://doi.org/10.1016/J.SCITOTENV.2020.137320>
17. Evans IS (2003) Scale-specific landforms and aspects of the land surface.. In: Evans IS, Dikau R, Tokunaga E, Ohmori H, Hirano M (eds) *Concepts and Modelling in Geomorphology: International Perspectives*. Terrapub, Tokyo, pp 61–84

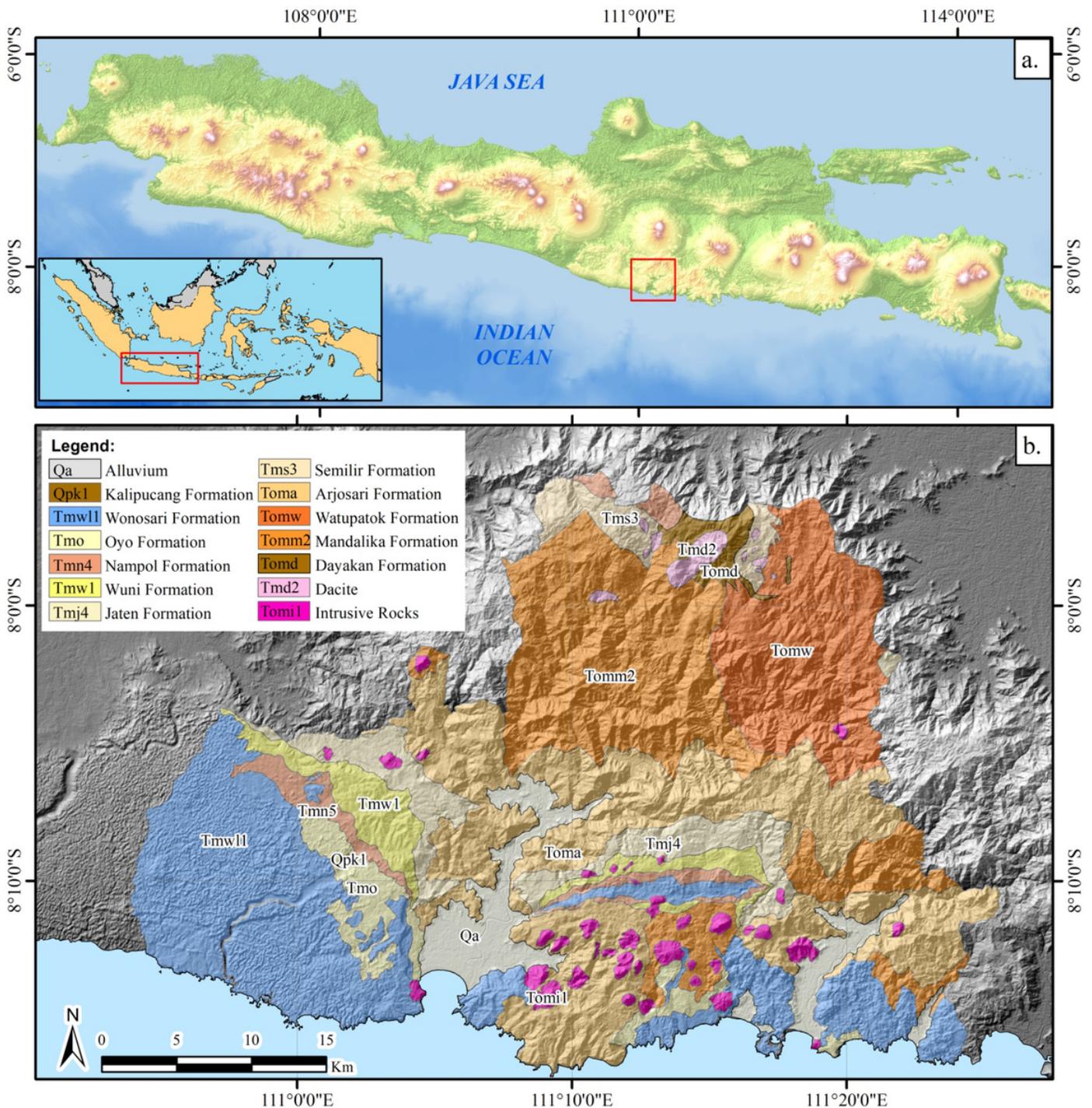
18. Fell R, Corominas J, Bonnard C, Cascini L, Leroi E, Savage WZ (2008) Guidelines for landslide susceptibility, hazard and risk zoning for land-use planning. *Eng Geol* 102:99–111. <https://doi.org/10.1016/j.enggeo.2008.03.014>
19. Godt JW, Baum RL, Savage WZ, Salciarini D, Schulz WH, Harp EL (2008) Transient deterministic shallow landslide modeling: Requirements for susceptibility and hazard assessments in a GIS framework. *Eng Geol* 102:214–226. <https://doi.org/10.1016/j.enggeo.2008.03.019>
20. Guzzetti F, Mondini AC, Cardinali M, Fiorucci F, Santangelo M, Chang KT (2012) Landslide inventory maps: New tools for an old problem. *Earth-Sci Rev* 112:42–66. <https://doi.org/10.1016/j.earscirev.2012.02.001>
21. Guzzetti F, Reichenbach P, Ardizzone F, Cardinali M, Galli M (2006) Estimating the quality of landslide susceptibility models. *Geomorphology* 81:166–184
22. Hengl T (2006) Finding the right pixel size. *Comput Geosci* 32:1283–1298. <https://doi.org/10.1016/j.cageo.2005.11.008>
23. Hengl T, Heuvelink GBM, Kempen B, Leenaars JGB, Walsh MG, Shepherd KD, Sila A, MacMillan RA, de Jesus JM, Tamene L, Tondoh JE (2015) Mapping Soil Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current Predictions. *PLoS ONE* 10:e0125814. <https://doi.org/10.1371/JOURNAL.PONE.0125814>
24. Hussin HY, Zumpano V, Reichenbach P, Sterlacchini S, Micu M, van Westen C, Bălteanu D (2016) Different landslide sampling strategies in a grid-based bi-variate statistical susceptibility model. *Geomorphology* 253:508–523. <https://doi.org/10.1016/J.GEOMORPH.2015.10.030>
25. Jaafari A, Zenner EK, Panahi M, Shahabi H (2019) Hybrid artificial intelligence models based on a neuro-fuzzy system and metaheuristic optimization algorithms for spatial prediction of wildfire probability. *Agric For Meteorol* 266–267:198–207. <https://doi.org/10.1016/J.AGRFORMET.2018.12.015>
26. Just AC, Arfer KB, Rush J, Dorman M, Shtein A, Lyapustin A, Kloog I (2020) Advancing methodologies for applying machine learning and evaluating spatiotemporal models of fine particulate matter (PM<sub>2.5</sub>) using satellite data over large regions. *Atmos Environ* 239:117649. <https://doi.org/10.1016/J.ATMOSENV.2020.117649>
27. Kim J-C, Lee S, Jung H-S, Lee S (2017) Landslide susceptibility mapping using random forest and boosted tree models in Pyeong-Chang, Korea. <https://doi.org/10.1080/10106049.2017.1323964> 33,1000–1015. <https://doi.org/10.1080/10106049.2017.1323964>
28. Kuhn (2021) *Caret: Classification and regression Training*
29. Kuhn M, Johnson K (2013) Applied predictive modeling. *Appl Predict Model* 1–600. <https://doi.org/10.1007/978-1-4614-6849-3>
30. Lai J-S, Chiang S-H, Tsai F (2019) Exploring Influence of Sampling Strategies on Event-Based Landslide Susceptibility Modeling. *ISPRS Int J Geo-Information* 2019 8(8):397. <https://doi.org/10.3390/IJGI8090397>
31. Liaw A, Wiener M (2002) Classification and Regression by randomForest. *R News* 2:18–22
32. Lovelace R, Nowosad J, Münchow J (2020) *Geocomputation with R*, 1st edn. Chapman and Hall/CRC Press
33. McHugh ML (2012) Interrater reliability: the kappa statistic. *Biochem Med* 22:276
34. Merghadi A, Yunus AP, Dou J, Whiteley J, ThaiPham B, Bui DT, Avtar R, Abderrahmane B (2020) Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth Sci Rev* 207:103225. <https://doi.org/10.1016/J.EARSCIREV.2020.103225>

35. Meyer H, Reudenbach C, Hengl T, Katurji M, Nauss T (2018) Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environ Model Softw* 101:1–9. <https://doi.org/10.1016/J.ENVSOFT.2017.12.001>
36. Meyer H, Reudenbach C, Wöllauer S, Nauss T (2019) *Ecol Modell* 411:108815. <https://doi.org/10.1016/J.ECOLMODEL.2019.108815>. Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction
37. Micheletti N, Foresti L, Robert S, Leuenberger M, Pedrazzini A, Jaboyedoff M, Kanevski M (2013) Machine Learning Feature Selection Methods for Landslide Susceptibility Mapping. *Math Geosci* 2013 461(46):33–57. <https://doi.org/10.1007/S11004-013-9511-0>
38. Moore ID, Grayson RB, Ladson AR (1991) Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. *Hydrol Process* 5:3–30. <https://doi.org/10.1002/hyp.3360050103>
39. Ngadisih, Samodra G, Bhandary NP, Yatabe R (2017) Landslide Inventory: Challenge for Landslide Hazard Assessment in Indonesia. *GIS Landslide*. Springer Japan, Tokyo, pp 135–159. [https://doi.org/10.1007/978-4-431-54391-6\\_8](https://doi.org/10.1007/978-4-431-54391-6_8)
40. Ngadisih, Yatabe R, Bhandary NP, Dahal RK (2014) Integration of statistical and heuristic approaches for landslide risk analysis: a case of volcanic mountains in West Java Province, Indonesia. <http://dx.doi.org/10.1080/17499518.2013.826030> 8, 29–47. <https://doi.org/10.1080/17499518.2013.826030>
41. Nussbaum M, Spiess K, Baltensweiler A, Grob U, Keller A, Greiner L, Schaepman ME, Papritz A (2018) Evaluation of digital soil mapping approaches with large sets of environmental covariates. *SOIL* 4:1–22. <https://doi.org/10.5194/SOIL-4-1-2018>
42. Parise M, Part C, Solar (2001) *Terr. Planet. Sci.* 26, 697–703. [https://doi.org/10.1016/S1464-1917\(01\)00069-1](https://doi.org/10.1016/S1464-1917(01)00069-1)
43. Park S, Kim J (2019) Landslide Susceptibility Mapping Based on Random Forest and Boosted Regression Tree Models, and a Comparison of Their Performance. *Appl. Sci.* 2019, Vol. 9, Page 942 9, 942. <https://doi.org/10.3390/APP9050942>
44. Ploton P, Mortier F, Réjou-Méchain M, Barbier N, Picard N, Rossi V, Dormann C, Cornu G, Viennois G, Bayol N, Lyapustin A, Gourlet-Fleury S, Péliissier R (2020) Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nat Commun* 2020 111(11):1–11. <https://doi.org/10.1038/s41467-020-18321-y>
45. Probst P, Wright MN, Boulesteix A-L (2019) Hyperparameters and tuning strategies for random forest. *Wiley Interdiscip Rev Data Min Knowl Discov* 9:e1301. <https://doi.org/10.1002/WIDM.1301>
46. Quinn P, Beven K, Chevallier P, Planchon O (1991) The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. *Hydrol Process* 5:59–79. <https://doi.org/10.1002/hyp.3360050106>
47. Core Team R (2020) R: a Language and Environment for Statistical Computing. [WWW Document]. R Found. Stat. Comput. Vienna, Austria. URL <https://www.r-project.org/>
48. Reichenbach P, Rossi M, Malamud BD, Mihir M, Guzzetti F (2018) A review of statistically-based landslide susceptibility models. *Earth Sci Rev*. <https://doi.org/10.1016/j.earscirev.2018.03.001>

49. Roback K, Clark MK, West AJ, Zekkos D, Li G, Gallen SF, Chamlagain D, Godt JW (2018) The size, distribution, and mobility of landslides caused by the 2015 Mw7.8 Gorkha earthquake. *Nepal Geomorphology* 301:121–138. <https://doi.org/10.1016/J.GEOMORPH.2017.01.030>
50. Rossi M, Guzzetti F, Reichenbach P, Mondini AC, Peruccacci S (2010) Optimal landslide susceptibility zonation based on multiple forecasts. *Geomorphology* 114:129–142. <https://doi.org/10.1016/J.GEOMORPH.2009.06.020>
51. Samodra G, Chen G, Sartohadi J, Kasama K (2018) Generating landslide inventory by participatory mapping: an example in Purwosari Area. *Yogyakarta Java Geomorphology* 306:306–313. <https://doi.org/10.1016/j.geomorph.2015.07.035>
52. Samodra G, Ngadisih N, Malawani MN, Mardiatno D, Cahyadi A, Nugroho FS (2020) Frequency–magnitude of landslides affected by the 27–29 November 2017 Tropical Cyclone Cempaka in Pacitan, East Java. *J Mt Sci* 17:773–786. <https://doi.org/10.1007/s11629-019-5734-y>
53. Samodra H, Gafoer S, Tjokrosapoetro S (1992) Geological Map of the Pacitan Quadrangle, Jawa. Bandung
54. Sampurno, Samodra H (1997) Geological Map of the Ponorogo Quadrangle. Jawa. Bandung
55. Schratz P, Muenchow J, Iturritya E, Richter J, Brenning A (2019) Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. *Ecol Modell* 406:109–120. <https://doi.org/10.1016/J.ECOLMODEL.2019.06.002>
56. Smyth HR, Hall R, Nichols GJ (2008) Cenozoic volcanic arc history of East Java, Indonesia: The stratigraphic record of eruptions on an active continental margin. *Spec Pap Geol Soc Am* 436:199–222. [https://doi.org/10.1130/2008.2436\(10\)](https://doi.org/10.1130/2008.2436(10))
57. Soeters R, van Westen CJ (1996) Slope instability recognition, analysis, and zonation. In: Turner AK, Schuster RL (eds) *Landslides, Investigation and Mitigation* (Transportation Research Board, National Research Council, Special Report; 247). National Academy Press, Washington D.C., pp 129–177
58. Steger S, Brenning A, Bell R, Glade T (2016a) The propagation of inventory-based positional errors into statistical landslide susceptibility models. *Nat Hazards Earth Syst Sci* 16:2729–2745. <https://doi.org/10.5194/NHESS-16-2729-2016>
59. Steger S, Brenning A, Bell R, Petschko H, Glade T (2016b) Exploring discrepancies between quantitative validation results and the geomorphic plausibility of statistical landslide susceptibility maps. *Geomorphology* 262:8–23. <https://doi.org/10.1016/j.geomorph.2016.03.015>
60. Sterlacchini S, Ballabio C, Blahut J, Masetti M, Sorichetta A (2011) Spatial agreement of predicted patterns in landslide susceptibility maps. *Geomorphology* 125:51–61. <https://doi.org/10.1016/J.GEOMORPH.2010.09.004>
61. Sun D, Xu J, Wen H, Wang D (2021) Assessment of landslide susceptibility mapping based on Bayesian hyperparameter optimization: A comparison between logistic regression and random forest. *Eng Geol* 281:105972. <https://doi.org/10.1016/J.ENGCEO.2020.105972>
62. Taalab K, Cheng T, Zhang Y (2018) Mapping landslide susceptibility and types using Random Forest. *Big Earth Data* 2:159–178. <https://doi.org/10.1080/20964471.2018.1472392>
63. Van Den Eeckhaut M, Reichenbach P, Guzzetti F, Rossi M, Poesen J (2009) Combined landslide inventory and susceptibility assessment based on different mapping units: an example from the Flemish Ardennes, Belgium. *Nat Hazards Earth Syst Sci* 9:507–521

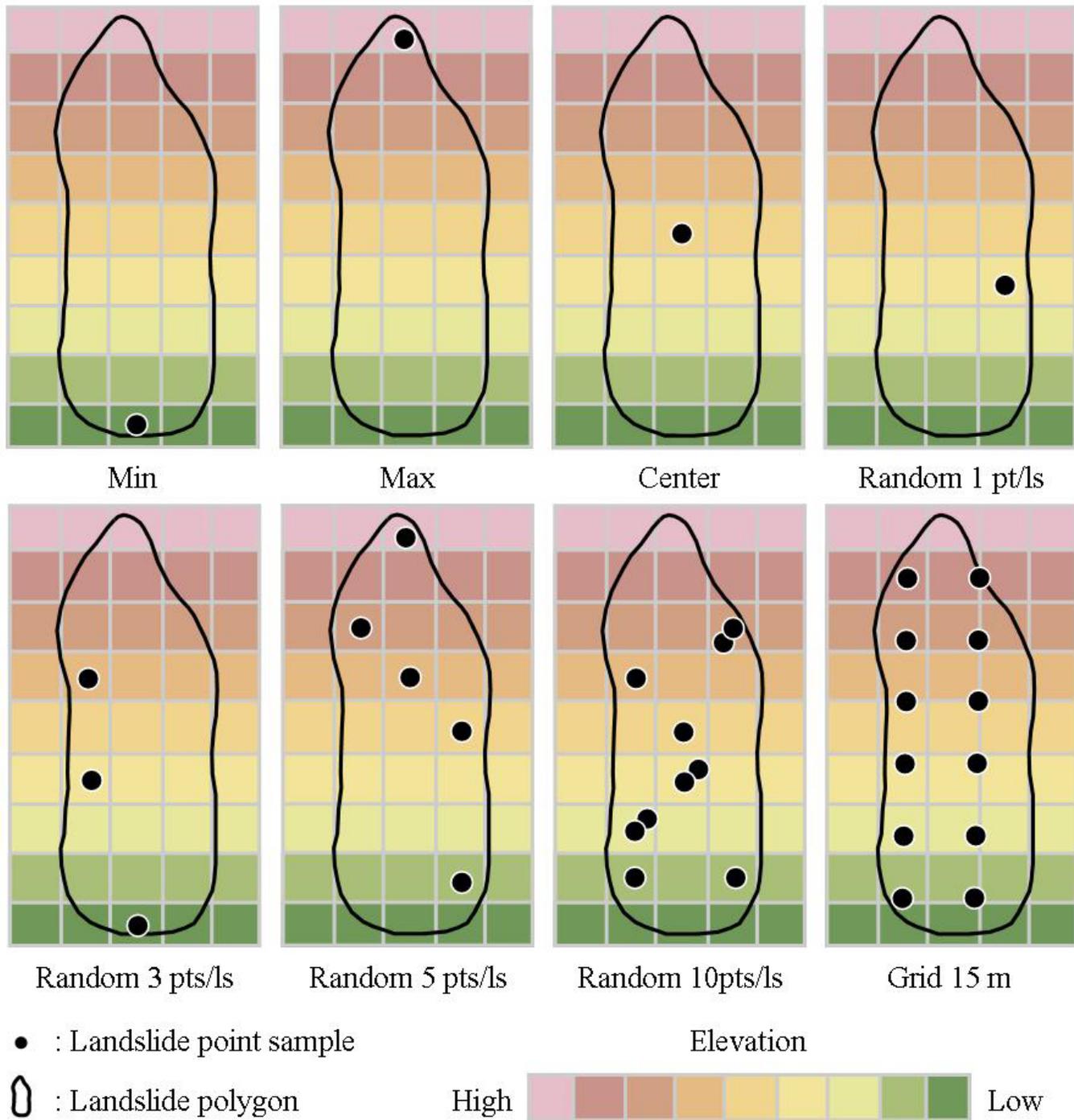
64. van Westen CJ, Castellanos E, Kuriakose SL (2008) Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview. *Eng Geol* 102:112–131.  
<https://doi.org/10.1016/J.ENGEO.2008.03.010>
65. Vaysse K, Lagacherie P (2015) Evaluating Digital Soil Mapping approaches for mapping GlobalSoilMap soil properties from legacy data in Languedoc-Roussillon (France). *Geoderma Reg* 4:20–30.  
<https://doi.org/10.1016/J.GEODRS.2014.11.003>

## Figures



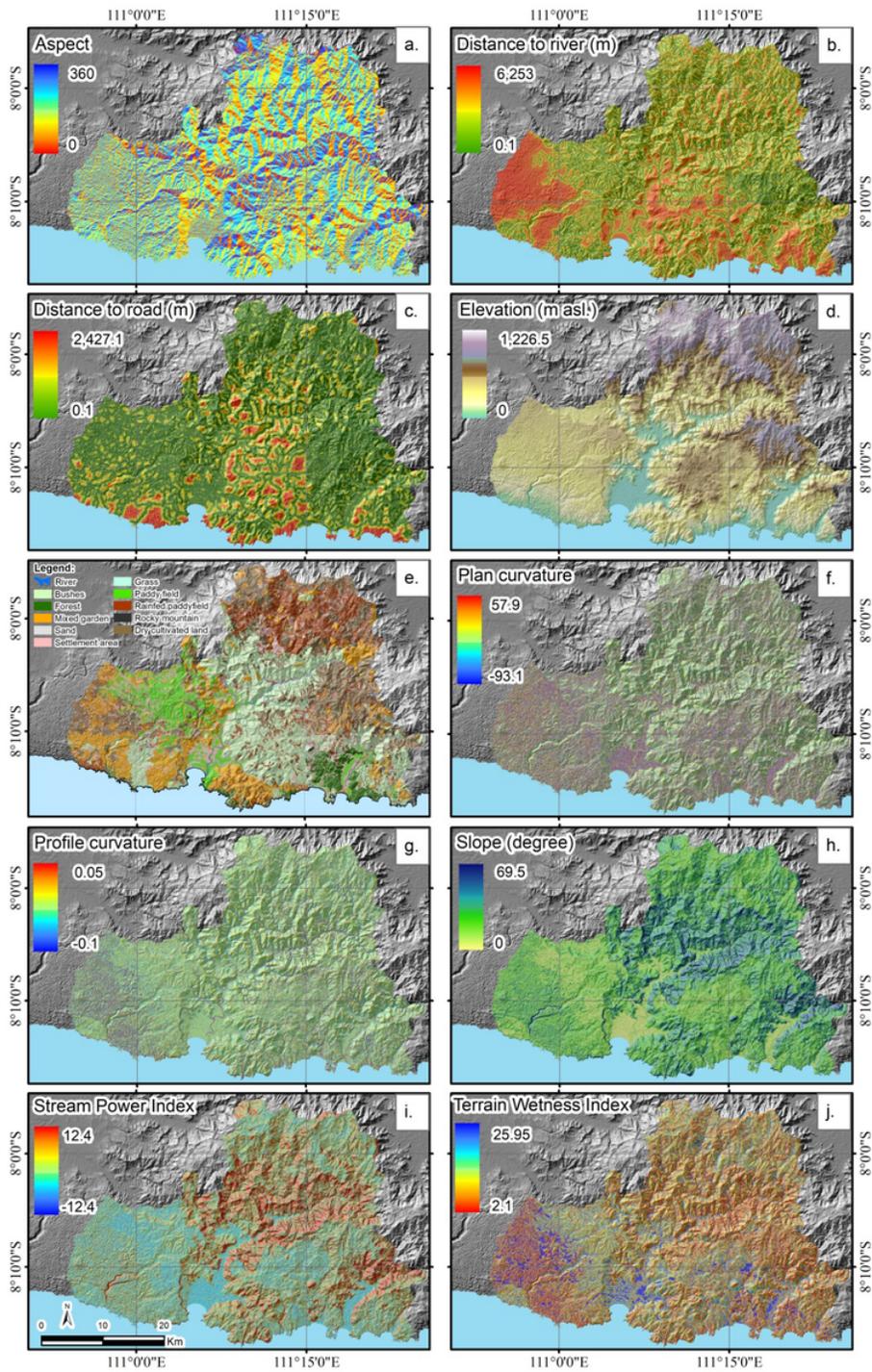
**Figure 1**

Location of the study area location, with the red rectangle denoting the Pacitan area and the red rectangle in the inset denoting the Java Island) (a). Geological formation of the Pacitan Regency (Samodra et al., 1992; Sampurno and Samodra, 1997), which is also used as landslide controlling factor map (b)



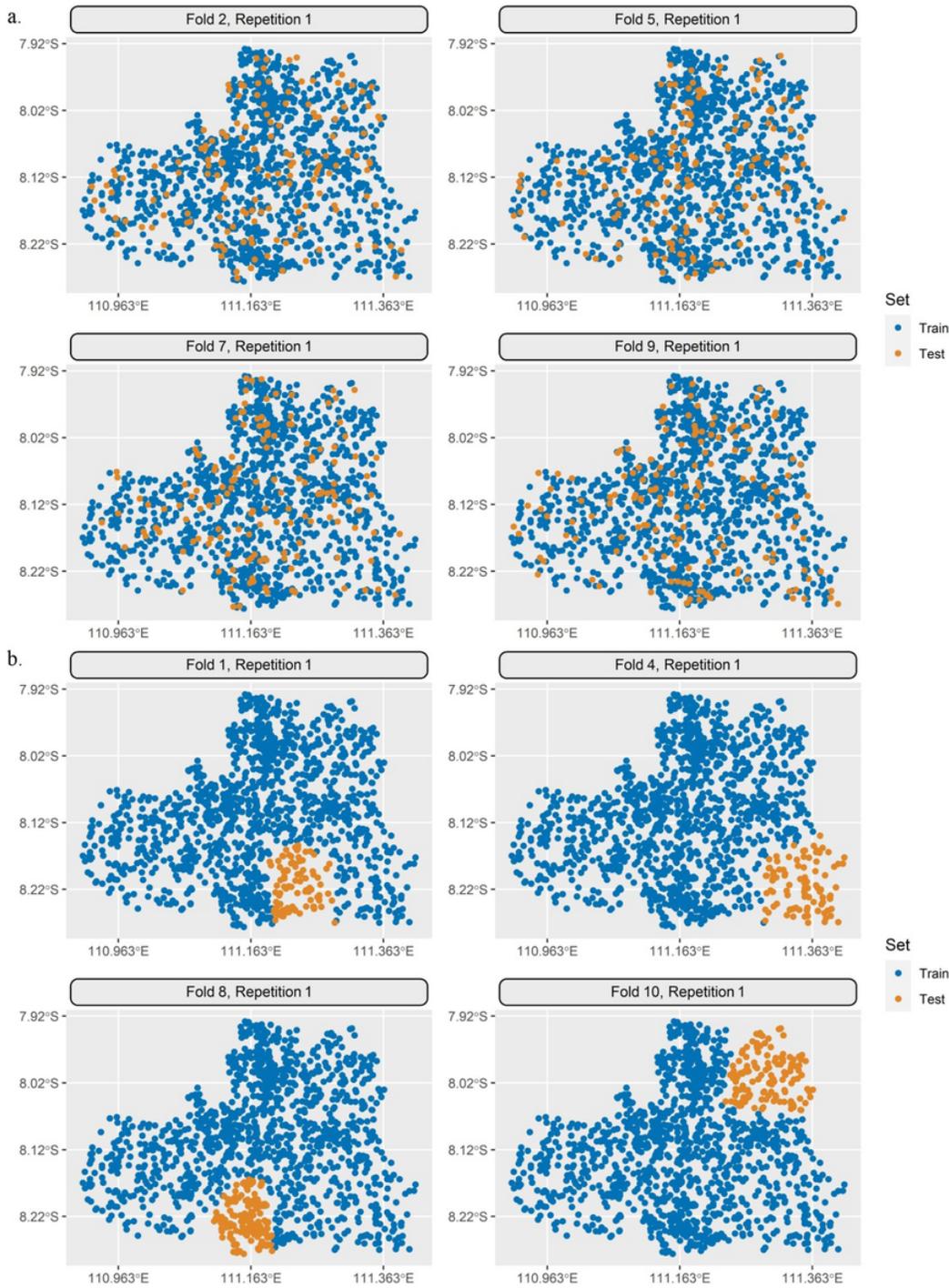
**Figure 2**

Data handling scenarios for converting landslide polygon into point sampling



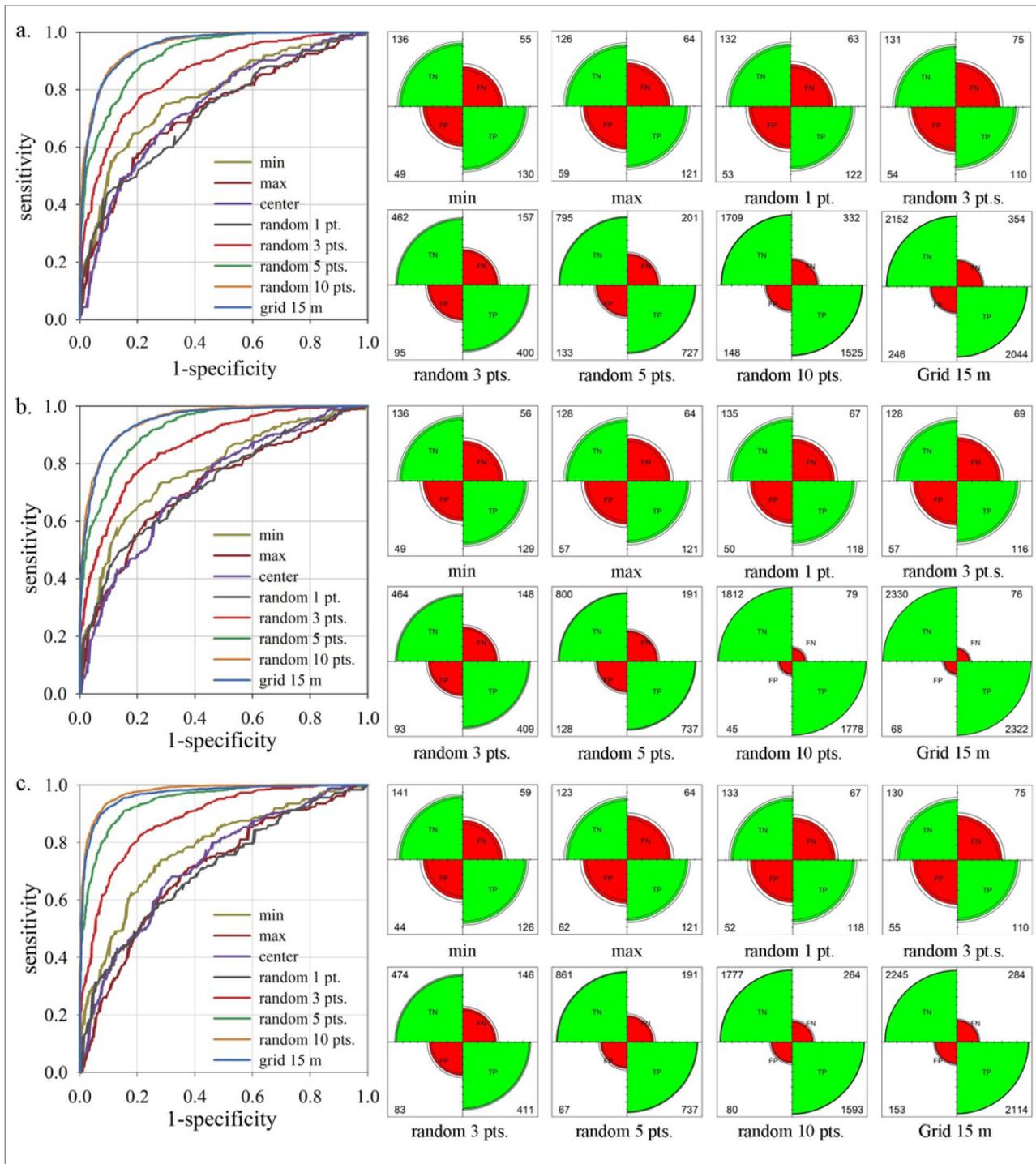
**Figure 3**

Landslide controlling factors: aspect (a), distance to the river (b), distance to the road (c), elevation (d), land use (e), plan curvature (f), profile curvature (g), slope (h), SPI (i), and TWI (j)



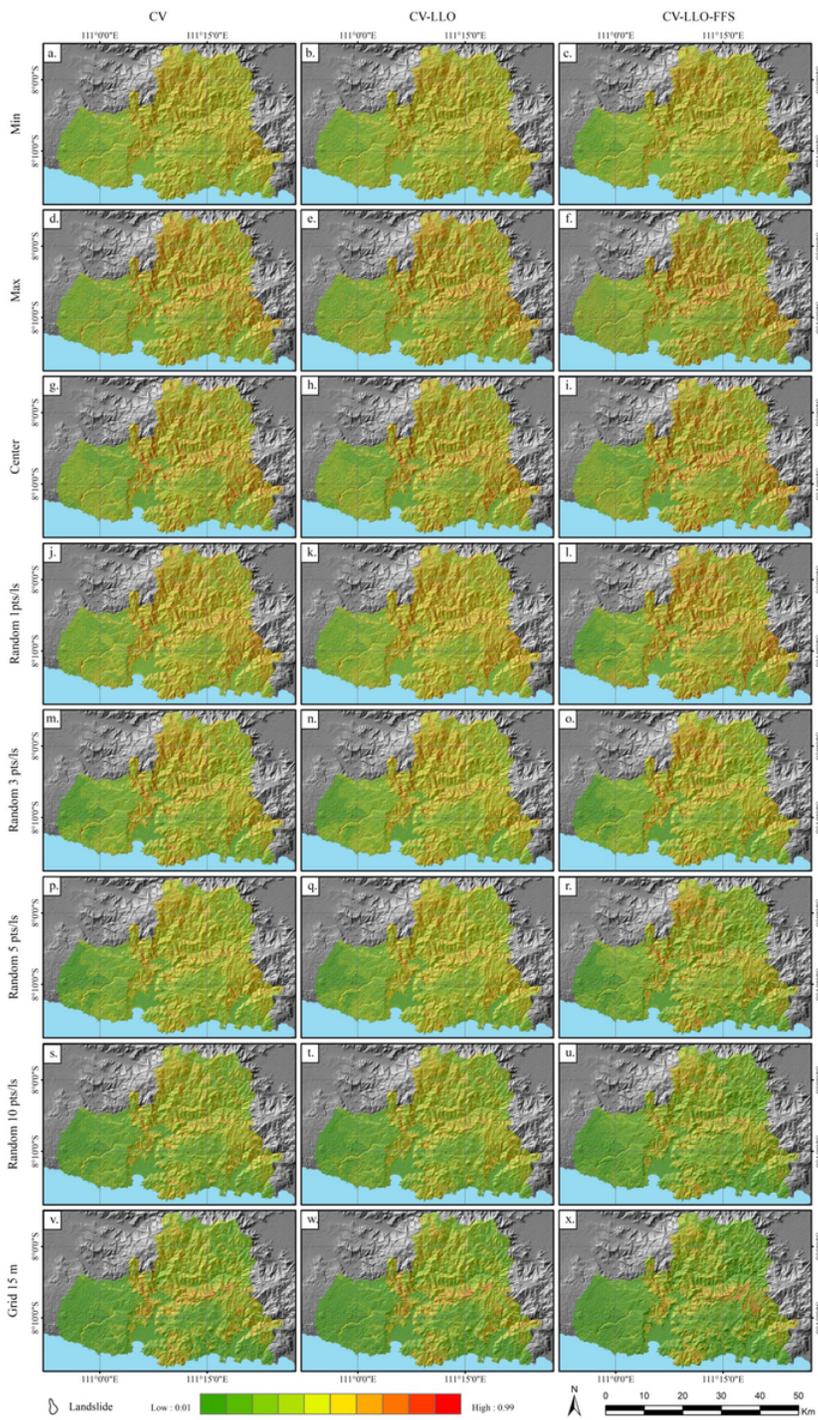
**Figure 4**

Illustrations of nonspatial 10-fold cross validation (CV) showing Folds 2, 5, 7, and 9 (a). Illustrations of spatial 10-fold CV showing Folds 1, 4, 8, and 10 (b)



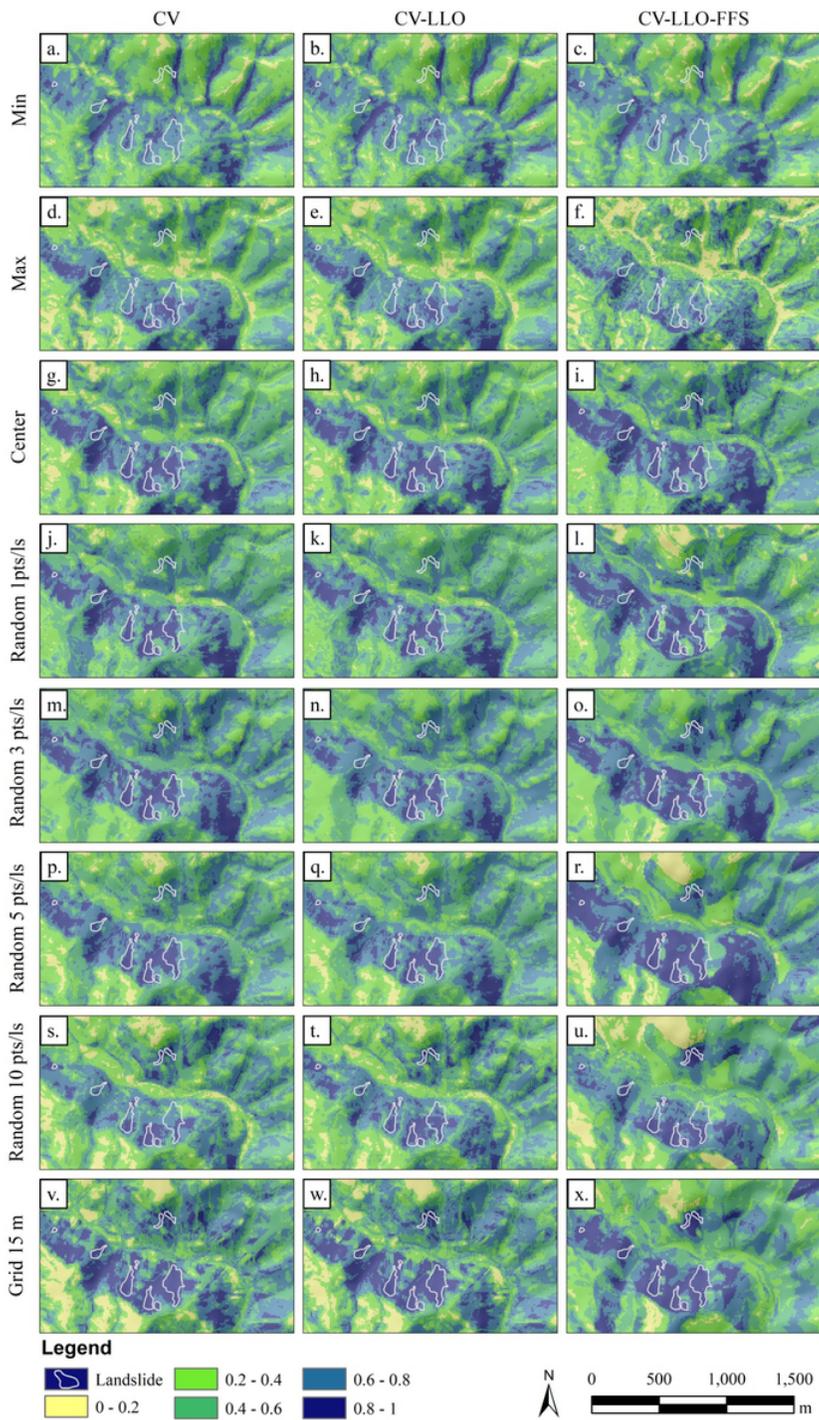
**Figure 5**

ROC and fourfold plot confusion matrix of different CV and hyperparameter tuning strategies: CV–nonspatial tuning (a), CV LLO–spatial tuning (b), and CV LLO FFS–no tuning (c)



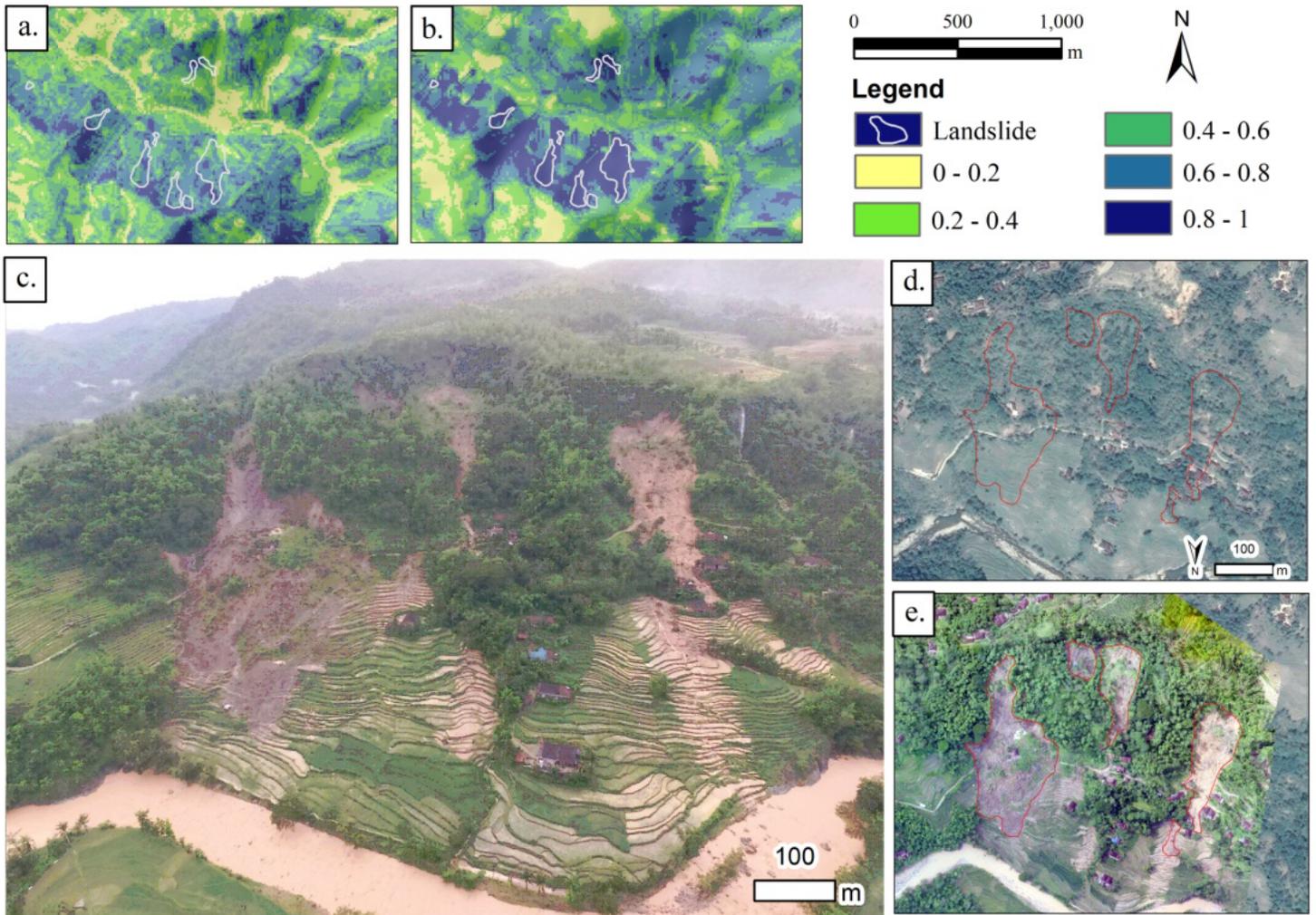
**Figure 6**

Landslide susceptibility maps obtained by the random forest model using different landslide inventory data handling strategies, validation strategies, and hyperparameter tuning



**Figure 7**

Superimposed landslide polygon and landslide susceptibility map (for the location where a detailed survey was conducted) obtained using different landslide inventory data handling strategies, validation strategies, and hyperparameter tuning



**Figure 8**

Maximum CV LLO FFS considered the worst map (a) and grid CV LLO considered the best map (b). Landslide areas in the maps shown in (a) and (b) were obtained from drone survey (c). Images of landslide polygons before TC Cempaka 2017 and (d) landslide polygons after TC Cempaka 2017 (e)

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Appendix.docx](#)