INDONESIAN JOURNAL ON GEOSCIENCE
Geological Agency
Ministry of Energy and Mineral Resources

Journal homepage: http://ijog.geologi.esdm.go.id
ISSN 2355-9314, e-ISSN 2355-9306

Hyperparameter Tuning on Machine Learning-Based Landslide Susceptibility Mapping (Case study: Palu City and Its Surrounding areas)

Sukristiyanti Sukristiyanti[1], Pamela Pamela[2], Moch Hilmi Zaenal Putra[1], Yukni Arifianti[1], Andri Fachrur Rozie[3], Hilda Lestiana[4], Tri Muji Susantoro[5], Sumaryono[2], Yohandi Kristiawan[2], and Iqbal Eras Putra[2,6]

[1]Research Center for Geological Disaster, National Research and Innovation Agency (BRIN), Indonesia
[2] Center of Volcanology and Geological Disaster Mitigation, Geological Agency, Indonesia
[3]Research Center for Data and Information Sciences, BRIN, Indonesia
[4]Research Center for Geological Resources, BRIN, Indonesia
[5]Research Center for Geoinformatics, BRIN, Cibinong, Indonesia
[6]Magister Program in Regional and City Planning, School of Architecture, Planning and Policy Development, ITB, Indonesia

Corresponding author: sukristiyanti @brin.go.id
Manuscript received: December, 01, 2023; revised: January, 19, 2024;
approved: February, 10, 2025; available online: February, 27, 2025

Abstract - Landslide susceptibility mapping (LSM) produces a zonation map of landslide susceptibility levels, representing the future probability of landslides. It is necessary to give a guideline regarding spatial planning. A machine learning method was used, namely a random forest (RF) algorithm to map landslide susceptibility in Python. The case study is Palu City and its surrounding areas, which were attacked by a big earthquake on September 28th, 2018. Some earlier LSM studies did not discuss hyperparameter tuning, and several others did not mention the training accuracy. Therefore, this study is to find out whether the fast model without hyperparameter tuning and frequently overfitting, can well produce landslide susceptibility maps. The research questions were addressed by comparing two landslide susceptibility maps built with and without hyperparameter tuning using receiver operating characteristics (ROC) and landslide density (LD) analyses. This study shows that the area under the curve (AUC) of the landslide susceptibility map from the fast RF model without hyperparameter tuning is as high as the AUC from the tuned model map. It also happened in both landslide density (LD) maps, and there is no anomaly in the fast model map. Nevertheless, there are strange appearances in the fast model map. Therefore, hyperparameter tuning to obtain the optimal model with no overfitting is mandatory to predict landslide susceptibility spatially.

Keywords: landslide susceptibility mapping (LSM), random forest, machine learning, hyperparameter tuning, Python

How to cite this article:
Sukristiyanti, S., Pamela, P., Putra, M.H.Z., Arifianti, Y., Rozie, A.F., Lestiana, H., Susantoro, T.M., Sumaryono, Kristiawan, Y., and Putra, I.E., 2025. Hyperparameter Tuning on Machine Learning-Based Landslide Susceptibility Mapping (Case study: Palu City and Its Surrounding areas). *Indonesian Journal on Geoscience*, 12 (1), p.43-53. DOI: 10.17014/ijog.12.1.43-53

Introduction

Landslide as a natural disaster needs a lot of attention since its effect can destroy infrastructure, kill people, and cause many more losses that come with it (Harsa *et al.*, 2023). A landslide susceptibility map indicates the landslide-prone areas, which can be used by planners or policy-makers, scientists, and the community for the landslide-disaster risk reduction effort. Machine learning-based landslide susceptibility mapping (LSM) methods have recently become popular (Kavzoglu *et al.*, 2019), since information technology is growing fast. It makes machine learning modeling can be run lightly and easily nowadays. The abundance of geospatial data such as satellite images, remotely sensed data, and historical landslide data in the last decade supports it. This mapping method is suitable on a medium scale (Westen *et al*., 2006) for quite large areas, like in this studied area (Barredo *et al*., 2000).

The location of the study is Palu City and its surrounding areas. The lithology composition of this area mainly consists of Holocene alluvial deposits in the centre, characterized by a plain area. This area is encompassed by two hills on the right and left, composed of Pleistocene Pakuli Formation, a Miocene granite intrusive rock exposed in the eastern part of the studied area. The Upper Miocene Lariang Formation, underlain by metamorphic rocks of the Latimojong Formation, dominates the western part (Sukamto, 1973). The studied area is located along the Palu Koro Fault Valley, flanked by two hills on the right and left sides. Palu City faced a big 7.5 Richter Scale earthquake on September 28th, 2018, which caused 4,340 fatalities (Arifianti *et al*., 2023). The shakes triggered the landslides in the hills arround the city (Figure 1).

The characteristics of Palu acted as landslide causative factors were compiled into a geodatabase. The causative factors consist of twelve layers, *i.e.,* aspect, elevation, slope gradient, lithology, peak ground acceleration (PGA), land use, distance to roads, distance to rivers, distance to lineaments, road density, river density, and lineament density. This study utilized the causative factors as independent types with the successfully inventoried landslides to do the LSM. The output of LSM is a zonation map of landslide susceptibility levels, which depicts the probability of landslides in the future (Tyagi *et al.*, 2022).

The machine learning method is a novel method to conduct LSM (Reichenbach *et al.*, 2018) belonging to data-driven models, which tend to be able to avoid subjectivity compared to other previous methods. Machine learning methods perform more see-through calculations and improve accuracy (Chen *et al*., 2018). Bivariate statistical-based LSM is a data-driven model (Nohani *et al.*, 2019; Arifianti *et al.*, 2020). However, it needs categorical type data for each landslide causative factor, and there is lack of standard to classify each parameter instead of using heuristic classification (Chen *et al*., 2017). In the machine learning, various data types can be run together, even continuous and nominal-type data (Cao *et al*., 2020), and then it can avoid subjectivity. Some continuous data represent geo-environmental values close to actual values, like earth surface altitude, slope gradient, distance to rivers, *etc*.

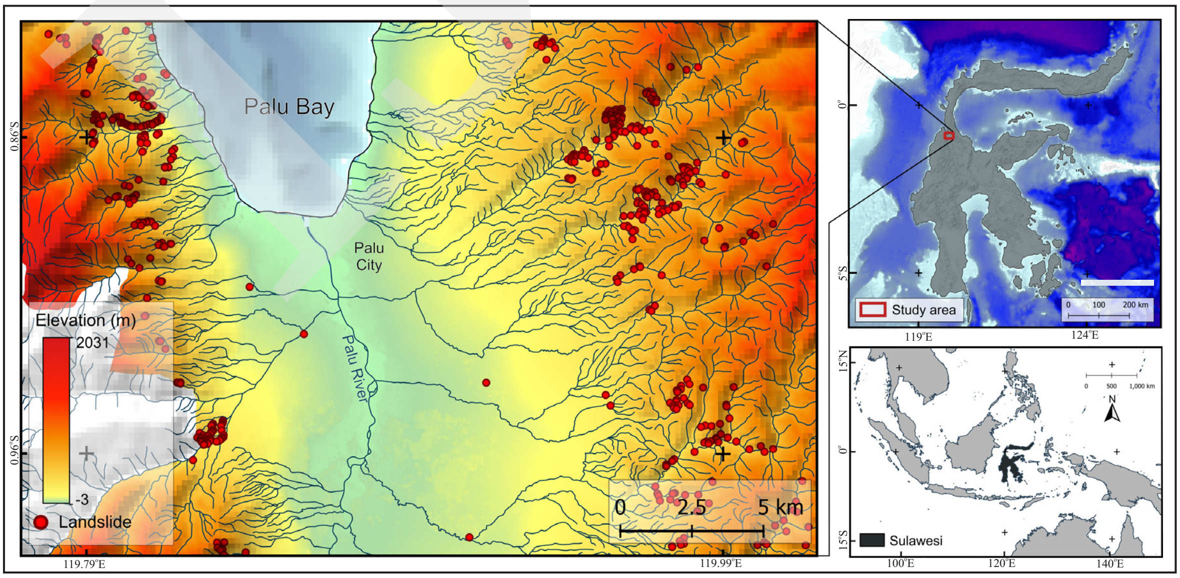

Figure 1. Landslide distribution map in the studied area.

Many prior LSM studies used machine learning models. Some researchers used one (Yao *et al*., 2008; Taalab *et al*., 2018) or more machine learning algorithms (Kalantar *et al*., 2018; El-Magd *et al.*, 2021; Shahzad *et al.*, 2022). Those algorithms are decision trees, random forests (Taalab *et al.*, 2018; El-Magd *et al*., 2021; Shahzad *et al*., 2022), support vector machines (Shahzad *et al*., 2022), naïve Bayes (El-Magd *et al*., 2021), k-nearest neighbour (El-Magd *et al*., 2021), and many more. This LSM study applied random forest (RF), considering previous researchers proved this method superior (Chen *et al*., 2018). It is a nonlinear algorithm suitable for overcoming classification problems in complex phenomena like landslide occurrence (Cao *et al.*, 2020).

The RF model requires hyperparameter tuning to optimize it (Sun *et al.*, 2020). An optimum model has test accuracy lower than training accuracy, and the training accuracy is not equal to 1.0 or overfitting (Muller and Guido, 2016). Overfitting occurs when a model is too complex and too fit to the particularities of the training data (Muller and Guido, 2016). Test accuracy compares the prediction of test data with the actual test data, while training accuracy compares the prediction of the training data with the actual training data. The test data is not involved in the modeling, whereas the training data is used to build a model.

Some previous LSM studies did not mention training accuracy or discuss hyperparameter tuning. Therefore, this paper aims to study whether the fast model without hyperparameter tuning, which frequently overfits, can produce good landslide susceptibility maps. This study answered the research question by looking at two landslide susceptibility maps that were made with and without hyperparameter tuning. The maps were made using receiver operating characteristics (ROC) and landslide density (LD) analysis (Chen et al., 2018).

Methods

Random forest (RF) is an ensemble method that combines many decision trees. The training dataset was built using the bagging (bootstrap aggregating) concept. Bagging produces many new training datasets. Each new training dataset takes a sample from the original dataset with a replacement (bootstrap sample), *i.e.*, it takes a sample, returns it to the original dataset, and takes another sample at random. By sampling with replacement, the same sample can be repeated in each new training dataset, and this causes the new dataset to be different for each tree (Taalab *et al.*, 2018). As a supervised learning method, RF needs a response variable or desired output (Muller and Guido, 2016; Islam *et al*., 2021).

The twelve landslide causative factors were prepared with a cell size of 10 m. TerraSAR X satellite data with a cell size of 7.5 m was used to get all the elements that make up the digital elevation model (DEM), such as aspect, elevation, and slope gradient (Arifianti *et al.*, 2023). Because LSM is a binary classification, it requires two classes for the response variable: landslide location and non-landslide location (He *et al*., 2021). Landslide locations were obtained from landslide inventory, while nonlandslide locations were constructed from sampling on the two lowest classes of bivariate statistical landslide susceptibility map produced by a previous study (Arifianti *et al*., 2023).

The number of landslides is 591 points, and the same number of nonlandslides were generated (Zhang *et al*., 2019). All the thematic layers and the response variables were prepared in the geographic information system (GIS) environment. Three causative factors have nominal type data, *i.e*., lithology, aspect, and land use, while the rest have continuous type data.

The slope gradient is one of the mandatory parameters for landslide susceptibility study (Çellek, 2022). It influences slope stability, where a more significant slope gradient makes the slope likely to collapse if the shear stress exceeds its shear strength. The elevation factor was chosen in this modeling since it may relate to rainfall and vegetation types, affecting landslides. Especially for earthquake-induced landslides, the ground tremor will be higher if the elevation increases. The slope aspect refers to sun exposure, which affects weathering levels and is closely connected to landslide probability.

Meanwhile, lithology influences soil permeability and strength associated with slope gradients. The lithology map is scaled to 1:50,000 obtained from the Centre for Geological Survey of Indonesia. Peak ground acceleration (PGA) is assumed to be an earthquake-leading indicator and is essential as a driving factor for earthquake-induced landslides. Besides, the lineament element was used as an independent factor. Faults and fractures were added in the lineament element, which showed the weak zone regarding lithological characteristics. Lineaments also control the existence of rivers structurally (Ngo *et al*., 2021), so river element was used to contribute to this modeling. Two factors based on rivers used here were the distance to rivers and river density.

Land use may control slope stability; which is enhanced by vegetation regarding mechanical characteristics. Land use is one of the anthropogenic factors where land-use changes affect the slope stability. Regarding anthropogenic factors, this study considered roads as an essential element, so distance to roads and road density layers were utilized in this modeling. The street development alters the slope stability of natural slopes since there is a cut-and-fill activity (Awawdeh *et al*., 2018).

A total of 1,182 samples were used as input in Python; 70 % was used as training data, and the rest for test data (Zhang *et al*., 2019). Each data point has a label, whether landslide or nonlandslide, bringing twelve geo-environment values. Besides, each sample has a geographical location: longitude and latitude. The coordinates for all data points are essential to trace which data points are landslides for training and test data, *etc*. Geographical locations are needed for landslide susceptibility map validation using receiver operating characteristics (ROC) (Pourghasemi *et al*., 2014).

Scikit-learn was applied, a Python package widely used for data science, to implement a random forest (RF) algorithm for landslide susceptibility mapping (LSM). The default values of the model parameters were used to build a fast RF model, as shown in Table 1. Aside from that, hyperparameter tuning was done to create the tuned model. The most common method, *grid search*, was used for this hyperparameter tuning. *Grid search* tries all possible combinations of desired model parameters (Muller and Guido, 2016).

Table 1. Default Random Forest Parameters

No	Parameter	Value
1	*n_estimators*	100
2	*criterion*	*gini*
3	*max_depth*	None
4	*min_samples_split*	2
5	*min_samples_leaf*	1
6	*min_weight_fraction_leaf*	0.0
7	*max_features*	*sqrt*
8	*max_leaf_nodes*	None
9	*min_impurity_decrease*	0.0
10	*bootstrap*	True
11	*oob_score*	False
12	*class_weight*	None
13	*max_samples*	None

In this study, five model parameters were tuned as follows: the number of decision trees (*n_estimators*), maximum depth of all trees (*max_depth*), minimum number of samples to split (*min_samples_split*), minimum number of samples in the leaves (*min_samples_leaf*), and maximum number of independent variables in all trees (*max_features*). Model parameters define the random forest and decision tree structure (Figure 2), where numerous decision trees compound a random forest. Every decision tree in the random forest (RF) is different because of the nature of randomness in it, like the samples, the number of independent variables, and which variables are involved in each decision tree, and so on.

Fast and tuned models were evaluated using the accuracy score in the Python environment. Two kinds of accuracy scores were performed, *i.e.*, training and test accuracy scores. The test accuracy is the model accuracy which compares

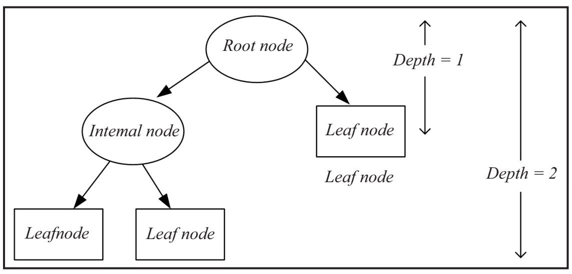

Figure 2. Decision tree diagram.

the prediction of test data with the actual test data. Meanwhile, the training accuracy compares the prediction of the training data with the actual training data. Whether the model is overfitting can be determined by calculating training and test accuracy.

These two models were used to generate spatial prediction results in the form of spatial landslide susceptibility indices, and to classify each index map into a landslide susceptibility map. The index ranges from 0.0 to 1.0, and is categorized into five susceptibility classes: very low, low, moderate, high, and very high. To compare the two resulting maps from the fast model and the tuned ones, they were evaluated by calculating the area under the curve (AUC) of the receiver operating characteristic (ROC) (Pourghasemi *et al.*, 2014) beyond the Python environment. The range of AUC is between 0.0 to 1.0 with details as follows: 0.5 - 0.6 (poor), 0.6 - 0.7 (average), 0.7 - 0.8 (good), 0.8 - 0.9 (very good), and more than 0.9 is excellent (Pourghasemi *et al.*, 2014). Besides, landslide density (LD) analysis was used to compare two landslide susceptibility maps. The previous study used LD analysis to compare the landslide susceptibility maps from three different models (Chen *et al.*, 2018). The LD is well-defined as the percentage of landslides in each class (PL) by the percentage of area in each class (PA). PA is the area of each class divided by the total area of the researched area, and PL is the count of landslides divided by the total number of landslides. Accurate landslide susceptibility maps should display a higher LD for classes of higher susceptibility (Chen *et al.*, 2018). The flow chart of this method can be seen in Figure 3.

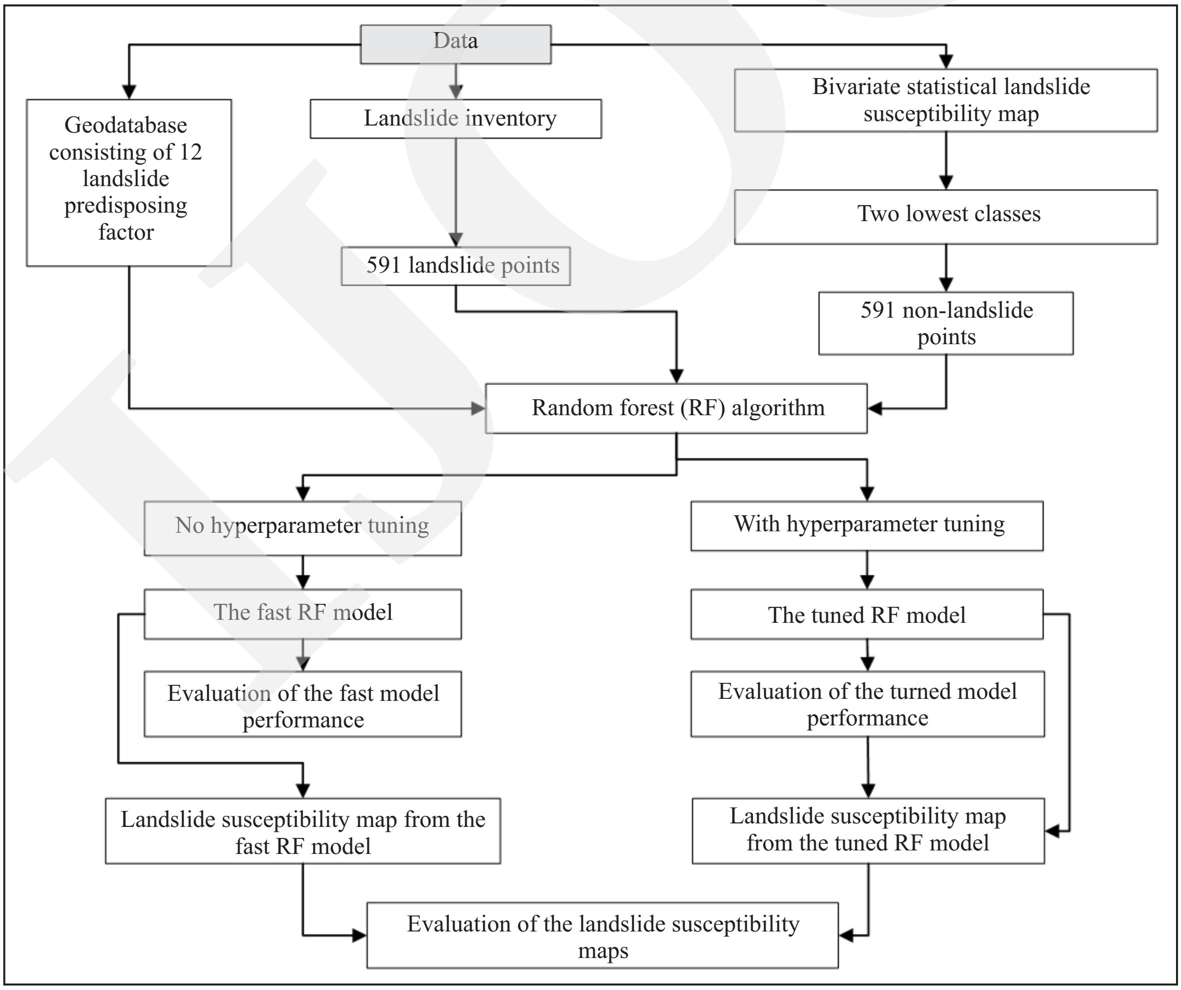

Figure 3. Flow chart of this study.

Result

The twelve processed causative factors, which are superimposed with landslide and nonlandslide points, can be seen in Figure 4.

In this RF modeling, 1,182 data points were used; half are landslides, and the other half are nonlandslides. Some 827 samples (70 %) were used to train the model, and 355 were used to test the model.

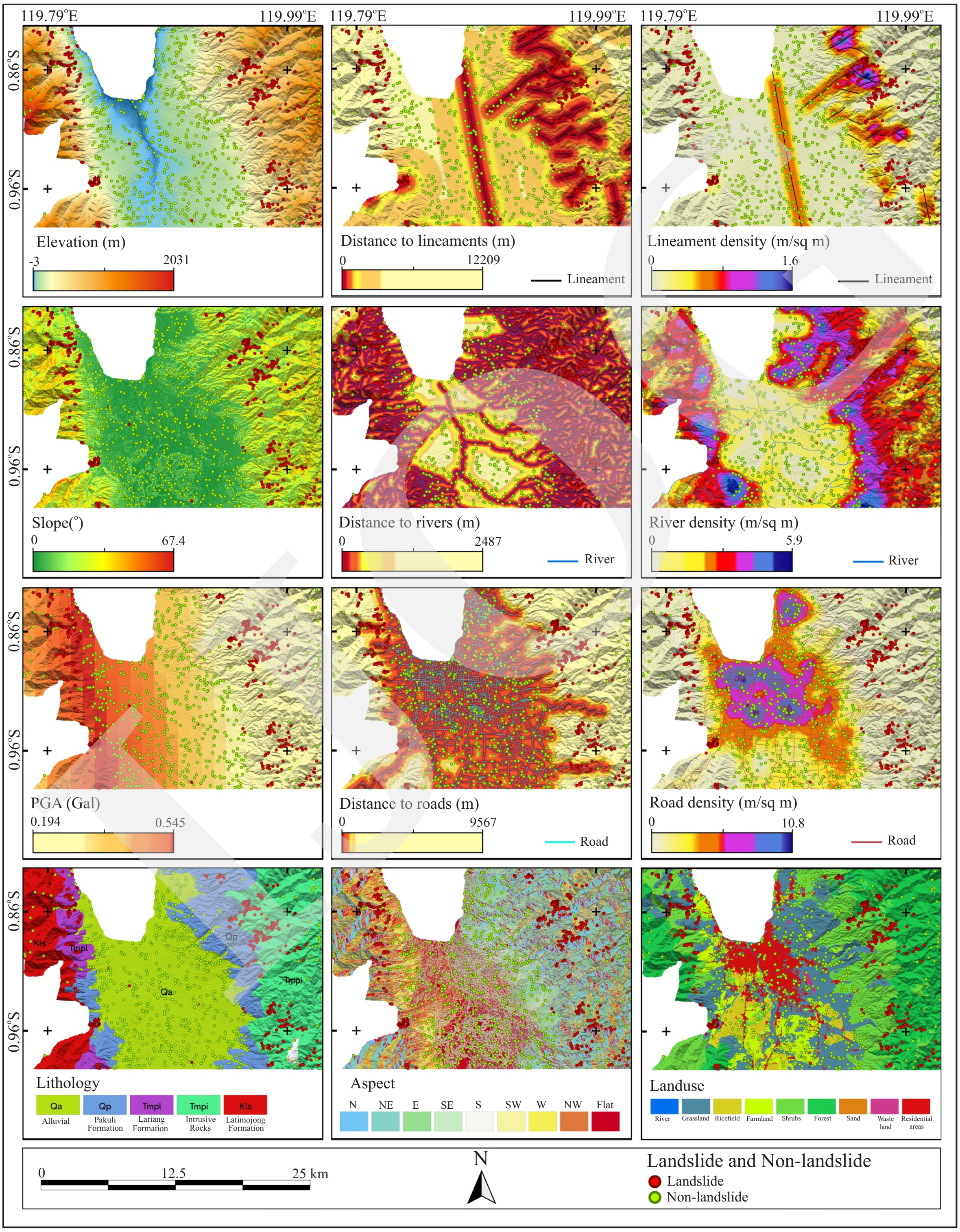

Figure 4. Twelve landslide causative maps were superimposed with landslide and nonlandslide points.

Two landslide susceptibility models were conducted, one with hyperparameter tuning and not on another one. In the fast model, hyperparameter tuning was not employed; the model is simplistic as it solely utilizes the default values, hence conserving time. Hyperparameter tuning was done using the *grid search* method involving five model parameters. The tuned model is the best constructed by the following parameter values: *max_depth* was 7, *max_features* was 3, *min_samples_leaf* was 2, *min_samples_split* was 5, and *n_estimators* was 150. The tuned model consists of 150 decision trees with a default value of 100. All decision trees have seven depths maximum, showing that they are not too long to reach the leaves. The causative factors used in each decision tree maximum is 3 (*max_features*= 3), not much involving most of the variables. The decision trees in the random forest will differ (Muller and Guido, 2016), because the randomness in picking the variables is high, considering that the total causative factors involved are 12.

Both RF, the fast RF and the tuned model, have the same model accuracy of 0.963, but the training accuracy of the fast RF model is 1.0, which shows there is overfitting. Conversely, hyperparameter tuning can avoid overfitting (Table 2). Both models were used to generate landslide susceptibility index maps.

Table 2. Performances of the Random Forest Models

	Training accuracy	Test accuracy
Fast RF model	1.000	0.963
Tuned RF model	0.995	0.963

ROC analysis was performed to obtain the map accuracy identified by their AUCs, which compared the landslides in test data with the landslide susceptibility map. In this analysis, each landslide susceptibility index map was categorized into thirty-two classes using the geometrical interval classification method. The number of landslides and the area were calculated for each class using GIS tools. The values were plotted in a curve in which the x-axis represents the cumulative percentage of the area, and the y-axis represents the cumulative percentage of landslides. The results showed that the AUC of landslide susceptibility maps from the fast RF model is 0.897, and from the tuned model is 0.893 (Figure 5).

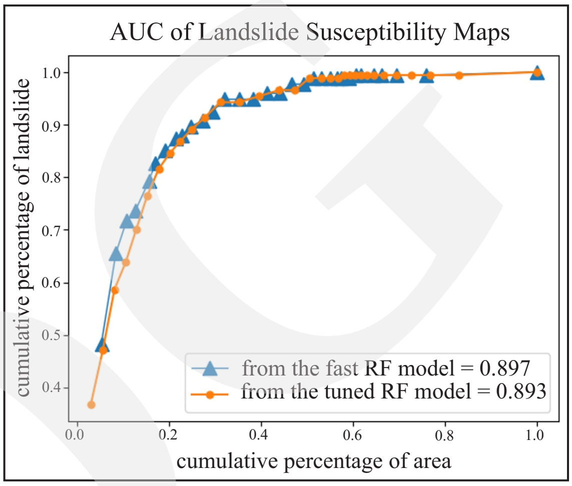

Figure 5. The area under the curve (AUC) represents the quality of landslide susceptibility maps.

Besides, landslide density (LD) was calculated. The previous study used the formula to compare landslide susceptibility maps, generalized from three different machine learning models (Chen *et al*., 2018). In this study, the LD analysis observed the accuracy of every landslide susceptibility map. The landslide susceptibility maps (Figure 6) were classified into five classes using the geometrical interval classification method. In this LD analysis, landslides are the total landslides used for the training and test data. PL means the percentage of landslides in every class, PA means the percentage of the area of each class, and LD is the landslide density of each class (Table 3).

Accurate landslide susceptibility maps should display a higher LD for classes of higher susceptibility (Chen *et al*., 2018), which agrees with LD in the two maps. Figure 6 presents the maps, zooming in on specific sections.

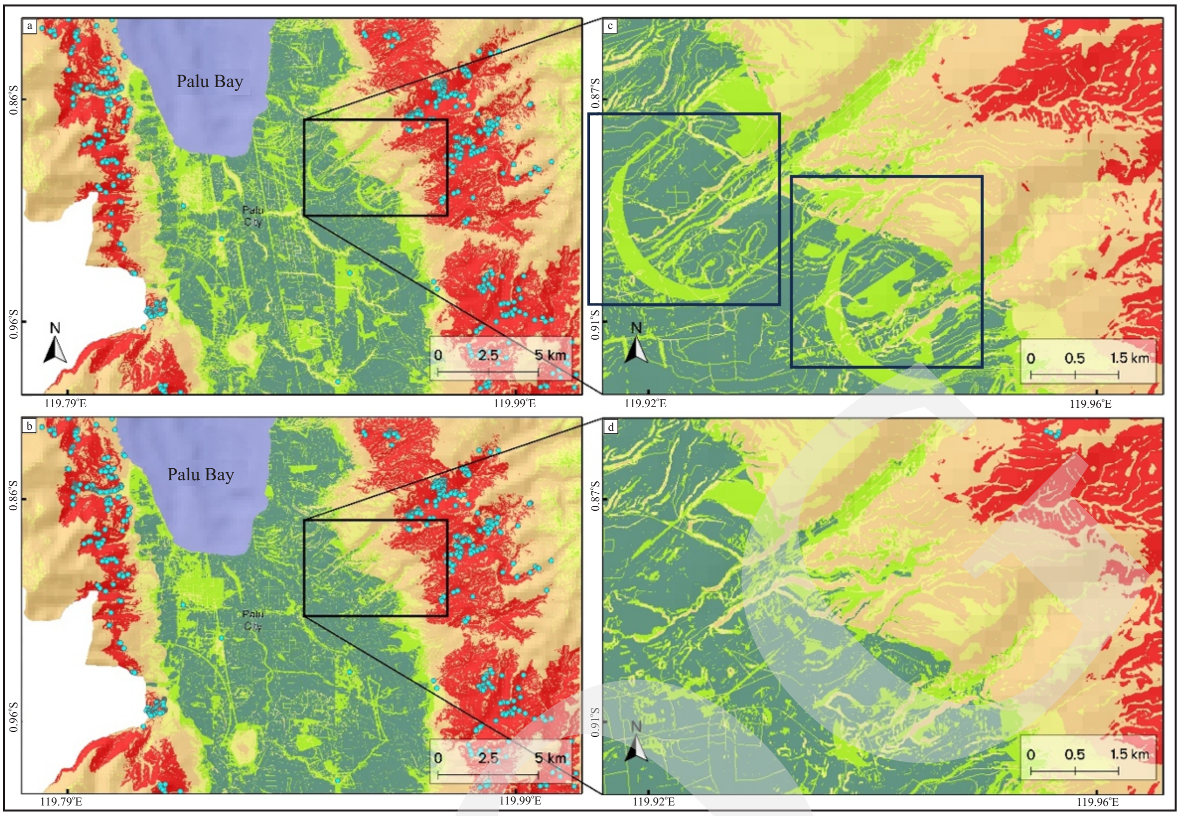

Figure 6. Landslide susceptibility maps using the RF algorithm, from: (a) Fast model, (b) Tuned model, and (c) the Zoom-in of the anomaly appearances.

Table 3. Landslide Density (LD) of Two Landslide Susceptibility Maps

Landslide susceptibility class	The map from the fast RF model			The map from the tuned RF model		
	PL	PA	LD	PL	PA	LD
Very low	0.006	0.240	0.024	0.006	0.241	0.024
Low	0.000	0.130	0.000	0.000	0.126	0.000
Moderate	0.006	0.098	0.059	0.006	0.083	0.069
High	0.132	0.291	0.455	0.241	0.319	0.757
Very high	3.247	0.241	13.452	3.138	0.231	13.582

Discussion

The accuracy of the model or the map is relatively high, > 0.8, while the fast RF model is 1.0, showing overfitting. It happens when a model is too complex and too fit to the particularities of the training data (Muller and Guido, 2016). The overfitting in the fast RF model can show an anomaly in the resulting map (Figure 6). Conversely, hyperparameter tuning can avoid overfitting, and the resulting map looks better.

The very high class on the susceptibility maps has the highest LD (Table 3). Both maps have similar LD, which means that there is no anomaly in the LD of the fast model map. The fast and tuned model maps are shown in Figure 6a and Figure 6b, respectively. The zoomed-in map presented in Figure 6a exhibits unusual features, as illustrated in Figure 6c. Figure 6c displays two crescent shapes, whereas Figure 6d, which provides a zoomed-in view of the tuned model map (Figure 6b), does not exhibit any such anomalies.

The calculation of AUC (Figure 5) and landslide density (LD) (Table 3) for the no-hyperparameter model map did not reveal any anomalies. However, it is evident that hyperparameter tuning is essential until overfitting is eliminated in the model.

A model is the fast model or the tuned model, which was built based on the training data, pairs of input, and target. In this study, the input, or predictor variable, is the landslide causative factor, and the target is landslides and non-landslides. A supervised algorithm, like the random forest (RF), learns the relationship between the input and target to produce a model to predict the new data. The optimal model lies in the optimal range of model complexity, avoiding both overfitting and underfitting (Figure 7).

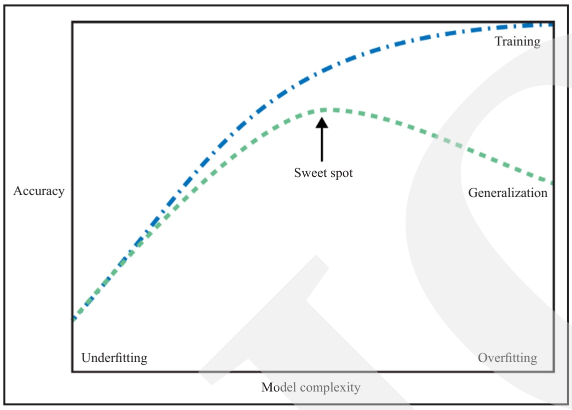

Figure 7. Model complexity trade-off versus training and test accuracy (Muller and Guido, 2016).

The underfitting model shows the model is too simple, and there is a lot of generalization in creating the model. The underfitting model has low accuracy both in training and test accuracy. Conversely, the overfitting model achieves the highest training accuracy, yet it struggles to make accurate predictions. Although it is already a principle in machine learning modeling, several studies have given little consideration. Overfitting generally occurs when a model is created without parameter tuning or when tuning is performed but not to the best of its ability. For example, this research used many model parameters, those are five parameters. Each parameter utilized multiple input values during the grid search process to ensure that the resulting model did not have a training accuracy of 1.0. Therefore, a significant effort in hyperparameter tuning is required to build an optimal random forest model and prevent overfitting.

Conclusions

This study is to find out whether the fast model without hyperparameter tuning, frequently overfitting, can well produce landslide susceptibility maps. The result shows that the area under the curve (AUC) of the landslide susceptibility map from the fast RF model without hyperparameter tuning is as high as the AUC from the tuned model map. It also happened in both map landslide density (LD), and there is no anomaly in the fast model map. Nevertheless, there are strange appearances in the fast model map. It can be concluded that hyperparameter tuning to obtain the optimal model with no overfitting is mandatory to predict the landslide susceptibility spatially. The model accuracy has to be calculated toward training and test data to observe whether there is overfitting. If overfitting happens, it is encouraged to do more hyperparameter tuning until the model has no overfitting.

Acknowledgements

This research was funded through the Programme House of Research Organization for Electronic And Informatics (OREI) of the National Research and Innovation Agency (BRIN). The authors would like to thank the Head of Research Centre for Geological Disaster (BRIN) who has encouraged the authors to carry out this research.

References

Arifianti, Y., Pamela, Agustin, F., and Muslim, D., 2020. Comparative study among bivariate statistical models in landslide susceptibility map, *Indonesian Journal on Geoscience*, 7 (1), p.51-63. DOI:10.17014/IJOG.7.1.51-63.

Arifianti, Y., Pamela, P., Iqbal, P., Sumaryono, S., Omang, A., and Lestiana, H., 2023. Susceptibility Assessment of Earthquake-induced Landslides, Sulawesi Mw 7.5 Earthquake, Indonesia. *Rudarsko-Geološko-Naftni Zbornik*, (*August*), p.43-54. DOI:10.17794/rgn.2023.3.4.

Awawdeh, M.M., El Mughrabi, M.A., and Atallah, M.Y., 2018. Landslide susceptibility mapping using GIS and weighted overlay method: a case study from North Jordan, *Environmental Earth Sciences*, 77 (21), p1-15. DOI:10.1007/s12665-018-7910-8.

Barredo, J.I., Benavides, A., Hervás, J., and Westen, C.J. van, 2000. Comparing heuristic landslide hazard assessment techniques using GIS in the Tirajana basin, Gran Canaria Island, Spain. *International Journal of Applied Earth Observation and Geoinformation*, 2000 (1), p.9-23. DOI:10.1016/s0303-2434(00)85022-9.

Cao, J., Zhang, Z., Du, J., Zhang, L., Song, Y., and Sun, G., 2020. Multi-geohazards susceptibility mapping based on machine learning—a case study in Jiuzhaigou, China, *Natural Hazards*, 102 (3), p.851-871. DOI:10.1007/s11069-020-03927-8.

Çellek, S., 2022. Effect of the slope angle and its classification on landslides, *Himalayan Geology*, 43 (1), p.85-95.

Chen, W., Pourghasemi, H.R., and Naghibi, S.A., 2017. A comparative study of landslide susceptibility maps produced using support vector machine with different kernel functions and entropy data mining models in China. *Bulletin of Engineering Geology and the Environment*, 77 (2), p.647-664. DOI:10.1007/s10064-017-1010-y.

Chen, W., Zhang, S., Li, R., and Shahabi, H., 2018. Performance evaluation of the GIS-based data mining techniques of best-first decision tree, random forest, and naïve Bayes tree for landslide susceptibility modeling, *Science of the Total Environment*, 644, p.1006-1018. DOI:10.1016/j.scitotenv.2018.06.389.

El-Magd, S.A.A., Ali, S.A., and Pham, Q.B., 2021. Spatial modeling and susceptibility zonation of landslides using random forest, naïve bayes and K-nearest neighbor in a complicated terrain. *Earth Science Informatics*, 14 (3), p.1227-1243. DOI:10.1007/s12145-021-00653-y.

Harsa, H., Hidayat, A.M., Mulsandi, A., Suprihadi, B., Kurniawan, R., Habibie, M.N., Hutapea, T.D., Swarinoto, Y.S., Makmur, E.E.S., Fitria, W., Sri Sudewi, R.S., and Praja, A.S., 2023. Machine learning and artificial intelligence models development in rainfall-induced landslide prediction, *IAES International Journal of Artificial Intelligence*, 12 (1), p.262-270. DOI:10.11591/ijai.v12.i1.pp262-270.

He, Q., Wang, M., and Liu, K., 2021. Rapidly assessing earthquake-induced landslide susceptibility on a global scale using random forest, *Geomorphology*, 391, 107889. DOI:10.1016/j.geomorph.2021.107889.

Islam, M.M., Kashem, M.A., and Uddin, J., 2021. Fish survival prediction in an aquatic environment using random forest model, *IAES International Journal of Artificial Intelligence*, 10 (3), p.614-622. DOI:10.11591/ijai.v10.i3.pp614-622.

Kalantar, B., Pradhan, B., Amir Naghibi, S., Motevalli, A., and Mansor, S., 2018. Assessment of the effects of training data selection on the landslide susceptibility mapping: a comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN). *Geomatics, Natural Hazards and Risk*, 9 (1), p.49-69. DOI:10.1080/19475705.2017.1407368.

Kavzoglu, T., Colkesen, I., and Sahin, E.K., 2019. *Landslides: Theory, Practice and Modelling. In*: Pradhan, S.P., Vishal, V., and Singh, T.N., (eds.), 50, p.283-301. DOI:10.1007/978-3-319-77377-3.

Muller, A.C. and Guido, S., 2016. *Introduction to Machine Learning with Python: A Guide for Data Scientists*. *In*: Schanafelt, D., (ed.), O'Reilly Media, Inc, Sebastopol, CA, retrieved from internet: https://www.oreilly.com/library/view/introduction-to-machine/9781449369880/, 392.

Ngo, T.Q., Dam, N.D., Al-Ansari, N., and Amiri, M., Phong, T. Van, Prakash, I., Le, H. Van, Nguyen, H.B.T., and Pham, B.T., 2021. Landslide Susceptibility Mapping Using Single Machine Learning Models: A Case Study from Pithoragarh District, India. *Advances in Civil Engineering*, 2021, p.1-19. DOI:10.1155/2021/9934732.

Nohani, E., Moharrami, M., and Sharafi, S., 2019. Landslide Susceptibility Mapping Using Different GIS-Based Bivariate Models. *Water MDPI*, 11 (1402), p.1-22.

Pourghasemi, H.R., Moradi, H.R., Fatemi Aghda, S.M., Gokceoglu, C., and Pradhan, B., 2014. GIS-based landslide susceptibility mapping with probabilistic likelihood ratio and spatial multi-criteria evaluation models (North of Tehran, Iran). *Arabian Journal of Geosciences*, 7 (5), p.1857-1878. DOI:10.1007/s12517-012-0825-x.

Reichenbach, P., Rossi, M., Malamud, B.D., Mihir, M., and Guzzetti, F., 2018. A review of statistically-based landslide susceptibility models. *Earth-Science Reviews*, 180 *(March)*, p.60-91. DOI:10.1016/j.earscirev.2018.03.001.

Shahzad, N., Ding, X., and Abbas, S., 2022. A Comparative Assessment of Machine Learning Models for Landslide Susceptibility Mapping in the Rugged Terrain of Northern Pakistan. *Applied Sciences (Switzerland)*, 12 (5), p.1-13. DOI:10.3390/app12052280

Sukamto, R., 1973. *Geological Map of the Palu Sheet, Sulawesi, scale 1:250.000*. Geological Research and Development Centre.

Sun, D., Wen, H., Wang, D., and Xu, J., 2020. A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm. *Geomorphology*, 362, 107201. DOI:10.1016/j.geomorph.2020.107201.

Taalab, K., Cheng, T., and Zhang, Y., 2018. Mapping landslide susceptibility and types using Random Forest. *Big Earth Data*, 2 (2), p.159-178. DOI:10.1080/20964471.2018.1472392.

Tyagi, A., Kamal Tiwari, R., and James, N., 2022. A review on spatial, temporal and magnitude prediction of landslide hazard. *Journal of Asian Earth Sciences,* X, (7), 100099. DOI:10.1016/j.jaesx.2022.100099.

Westen, C.J., Asch, T.W.J. van, and Soeters, R., 2006. Landslide hazard and risk zonation - Why is it still so difficult?. *Bulletin of Engineering Geology and the Environment*, 65 (2), p.167-184. DOI:10.1007/s10064-005-0023-0.

Yao, X., Tham, L.G., and Dai, F.C., 2008. Landslide susceptibility mapping based on Support Vector Machine: A case study on natural slopes of Hong Kong, China. *Geomorphology*, 101 (4), p.572-582. DOI:10.1016/j.geomorph.2008.02.011.

Zhang, T., Han, L., Han, J., Li, X., Zhang, H., and Wang, H., 2019. Assessment of landslide susceptibility using integrated ensemble fractal dimension with Kernel logistic regression model. *Entropy*, 21 (2), p.1-23. DOI:10.3390/e21020218.