International Journal of Multidisciplinary Research and Growth Evaluation.

Machine learning application in GIS and remote sensing: An overview

Anjeel Upreti
Capital College and Research Center, Kathmandu, Nepal

* Corresponding Author: **Anjeel Upreti**

Article Info

ISSN (online): 2582-7138
Volume: 03
Issue: 04
July-August 2022
Received: 18-07-2022;
Accepted: 12-08-2022
Page No: 546-553
DOI: https://doi.org/10.54660/anfo.2022.3.4.20

Abstract
Machine learning (ML) is a subdivision of artificial intelligence in which the machine learns from machine-readable data and information. It uses data, learns the pattern and predicts the new outcomes. Its popularity is growing because it helps to understand the trend and provides a solution that can be either a model or a product. Applications of ML algorithms have increased drastically in G.I.S. and remote sensing in recent years. It has a broad range of applications, from developing energy-based models to assessing soil liquefaction to creating a relation between air quality and mortality. Here, in this paper, we discuss the most popular supervised ML models (classification and regression) in G.I.S. and remote sensing. The motivation for writing this paper is that ML models produce higher accuracy than traditional parametric classifiers, especially for complex data with many predictor variables. This paper provides a general overview of some popular supervised non-parametric ML models that can be used in most of the G.I.S. and remote sensing based projects. We discuss classification (Naïve Bayes (NB), Support Vector Machine (SVM), Random Forest (RF), Decision Trees (DT)) and regression models (Random Forest (RF), Support Vector Machine (SVM), Linear and Non-Linear) here. Therefore, the article can be a guide to those interested in using ML models in their G.I.S. and remote sensing based projects.

Keywords: Machine Learning; artificial intelligence; pattern; models; classification; regression; GIS; remote sensing

1. Introduction

Machine learning (ML) is a subdivision of artificial intelligence in which the machine learns from machine-readable data and information [2]. It uses data, learns the pattern and predicts the new outcomes [3]. Its popularity is growing because it helps to understand the trend and provides a solution that can be either a model or a product. There are four types of machine learning approaches: supervised, unsupervised, semi-supervised and reinforcement learning [4]. In supervised learning, the labelled training data is provided; in unsupervised learning, unlabeled training data is provided [4]. The semi-supervised learning approach is a hybrid of both supervised and unsupervised learning where mostly labelled information is provided for the training [5]. However, the model is free to figure out the trend in the data on its own. In reinforcement learning, the agent learns from trial and error to make decisions and cope with the interactive environment [4]. A ML project consists of several steps and each steps should be planned carefully (Figure 1).

Applications of machine learning algorithms have increased drastically in G.I.S. and remote sensing in recent years [6]. It has a broad range of applications, from developing energy-based models to assessing soil liquefaction to creating a relation between air quality and mortality [7]. Other examples include qualitative and quantitative evaluation of satellite imagery sensor data for regional and urban scale air quality [8], support vector machine approach for longitudinal dispersion coefficients in natural streams [9], crisis management [10], disaster, linear programming for irrigation scheduling [11], global climate change and weather forecast [12], the status of land cover classification accuracy assessment [13], air pollutants and sources associated with health effects [2], settlement detection [14] features such as roads/highways and ditch segments extraction [8], identify crops' diseases and their yield estimation, building vegetation indices, natural disaster response, and disease outbreak response [15].

In addition, researchers/users are benefitted from the publicly available remote sensing datasets using which they can develop, test and run their ML models for their research [16]. Most of the remote sensing datasets are global and unbiased [17]. This further simplifies the workflow in building accurate ML models in this domain [18]. Furthermore, remote sensing-based research is not halted due to natural disasters or unexpected accidents [16].

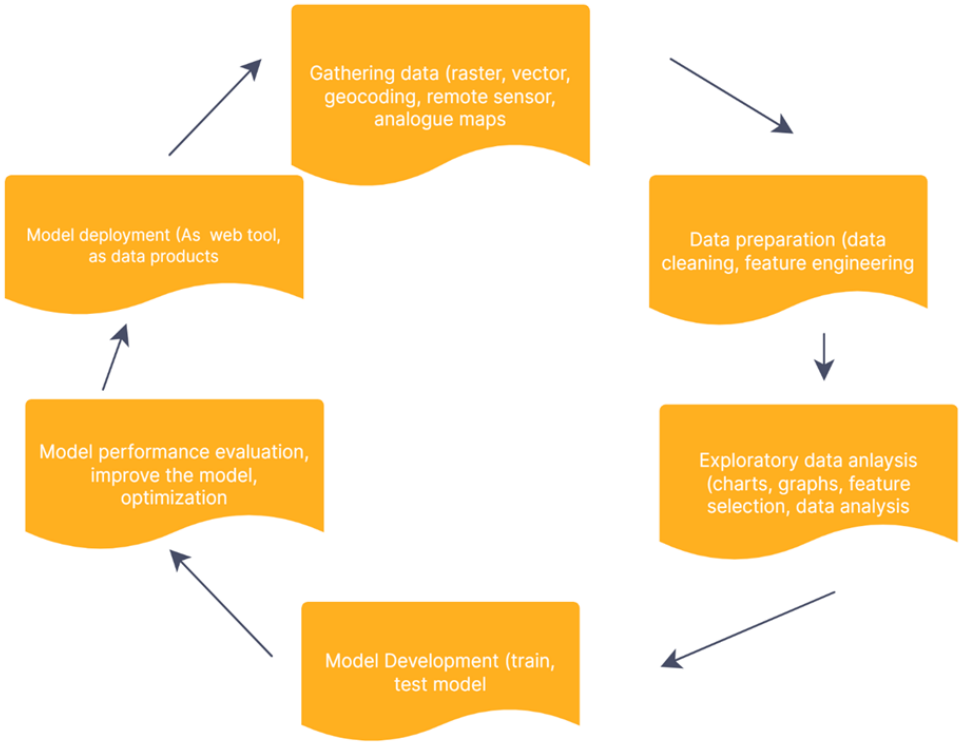

Fig 1: Machine learning workflow

Here, in this paper, we discuss the most popular supervised ML models (classification and regression) in G.I.S. and remote sensing. The motivation for writing this paper is that machine learning models produce higher accuracy than traditional parametric classifiers, especially for complex data with many predictor variables [19]. Therefore, the article can be a guide to those interested in using ML models in their G.I.S. and remote sensing based projects [20]. This paper provides a general overview of 4 supervised non-parametric ML models that can be used in most of the G.I.S. and remote sensing based projects. We discuss classification (Naïve Bayes (NB), Support Vector Machine (SVM), Random Forest (RF), Decision Trees (DT)) and regression models (Random Forest (RF), Support Vector Machine (SVM), Linear, Count and Poisson) here. Binomial and multiclass classification models are more common in G.I.S. and remote sensing-based projects [8]. If the classification has two classes, the classifier is known as binomial; if there are more than two classes, the category is multiclass.

...	Long_	Elev	ElevMin	ElevMax	Bname	Shape_Leng	Shape_Area	HydroID	OutletID	geometry
...	-95.004900	3.857793	1.0	9.0	None	20278.2932	4.692447e+06	300001	100018	POLYGON ((304289.891 3264141.990, 304280.085 3...
...	-95.028273	3.764508	0.0	13.0	None	28593.5702	1.114920e+07	300002	100001	POLYGON ((303858.438 3263161.415, 303838.826 3...
...	-95.006282	3.241213	0.0	9.0	None	14571.3466	3.493518e+06	300003	100002	POLYGON ((304515.423 3262916.271, 304515.423 3...
...	-95.034368	3.574497	0.0	6.0	None	16061.8212	2.994389e+06	300004	100003	POLYGON ((303250.481 3262631.904, 303250.481 3...

Attributes

Fig 2: Data engineering in G.I.S. world

1.1 Supervised machine learning models in G.I.S. and remote sensing

1.1.1 Naïve Bayes Algorithms

These supervised models are the easiest to build, less complex and can be applied to massive datasets [21]. It is fast. However, Naïve Bayes classification cannot be used for continuous numerical values [22]. It ignores noise, hence might lead to inaccurate predictions [23]. There are three types

of Naïve Bayes: Gaussian, Multinomial, Bernoulli. Gaussian assumes the distribution to be normal [24]. Multinomial for discrete counts and Bernoulli for binary outcomes [25].These classifiers are efficient for multiclass predictions [26]. These models can be best utilized in making best management practices models (B.M.P.s), habitat suitability models, weather prediction.

1.1.2. Random Forest Classifier

It's a supervised classification model that can be applied to classification and regression models [27]. It is a collection of decision trees and predicts the results based on the multiple models/sub-models [28]. Therefore it is also known as the ensemble classifier [29]. R.F. works on the bagging principle while making models, which means it makes different models based on the subset of training sample data, and the outcome is based on the majority/average of the sub-models [30]. Multiple studies suggest that the number of trees generally does not significantly impact the resulting R.F. classification accuracy, as long as the number is sufficiently large enough. This is because when the number of trees in the classifier is small, the prediction accuracy increases as additional trees are added. Still, the accuracy tends to plateau with a large number of trees [28]. Some common examples of the projects that can be solved using R.F. algorithm include: land use land cover classification [32], feature extraction such as ditch segments, roads, settlements, or objects of interest, object detection such as tree species, vehicle, species identification such as tigers, elephants, bird species, insects, habitat classification and modelling related projects such as flood-prone/drought [33], core habitat, classify soil types, diseases, weeds, climate and weather-related models and their forecasting [34].

1.1.3. Support vector machine

It's an ML model that can be applied to classification and regression problems [35]. It fits the data based on a distinct line known as a hyperplane [36]. As the model is easy to build and robust to outliers, it is widely used in the G.I.S. and remote sensing domains [37]. Building a support vector ML model requires the use to specify the kernel type [38]. Some popular kernels in remote sensing are polynomial kernels and the radial basis function (RBF) kernel [39]. Classification of satellite based imagery, detection of features like roads, wetlands, grasslands, can be solved using SVM models.

1.1.4. Linear regression

These models are the most popular research models in G.I.S. and remote sensing [40]. Linear regression helps to identify and evaluate the relationship between two or more factors/covariates when we leverage the power of space in our analysis using the distance features, for example, the influence of distance to water in habitat selection [41]. This ML model helps to address the questions like:

- Is there a linear relationship between diameter at breast height and crown diameter of trees?
- What demographic factors contribute to the use of high rates of public transport?
- What factors contribute to the high spread of COVID in geographical regions?
- What is the relation between environmental factors and the cyanobacteria population?
- What variables affect gender-specific leadership?
- What is the relation between climate change and migration?

There are 3 types of linear regression commonly used in GIS and remote sensing based projects. They are Continuous (Gaussian), Logistic and Poisson distribution. The distribution should be normal [42] for the Gaussian distribution linear regression. It is also called continuous because the dependent variable can take a wide range of values such as temperature, rainfall, and tree diameter [8]. If the dependent variable is not normally distributed, we can change it to binary values using reclassify function [43]. Binary is also known as logistic regression models, which builds models with only two outputs:- pass/fail, presence/absence [44]. We use count/Poisson regression models if the dependent variables are the counts/number of occurrences of an event [45].

The dependent variable cannot be negative or decimal values [45]. These models are generally used for species distribution models and understanding event patterns.

1.1.5. Non-linear regression

The regression in which the predictor and response variable has a non-linear relationship is known as non-linear regression [46]. Since most relationships in G.I.S. and remote sensing are non-linear, it is widely used in this sector [47]. Due to its flexibility, a wide variety of models can be built using these models [48]. For example, study the crops and soil processes, study the real estate price and immigration relation, study the relation between diameter and canopy cover.

1.2 Methods to improve the accuracy of the ML models

1.2.1. Feature engineering

Feature engineering is most prevalent in predictive models [49]. It is the process of filtering the most logical and influential variables/covariates in the models from the less important/influential variables, in ML terms, it is known as feature reduction [50]. It requires domain knowledge and understanding of the requirements of the projects [49]. Researchers run exploratory data analyses to observe the relationship between different variables/covariates and extract only the best variables to make an ML model [51].

1.2.2. Boosting

Boosting is a method used in machine learning to reduce errors in predictive data analysis [52]. Data scientists train machine learning software, called machine learning models, on labelled data to make guesses about unlabeled data [53]. A single machine learning model might make prediction errors depending on the accuracy of the training dataset [54]. For example, if a cat-identifying model has been trained only on images of white cats, it may occasionally misidentify a black cat. Boosting tries to overcome this issue by training multiple models sequentially to improve the accuracy of the overall system. Boosting improves machine models' predictive accuracy and performance by converting multiple weak learners into a single robust learning model. Machine learning models can be vulnerable learners or strong learners:

1.2.3. Hyper parameter optimization

Hyperparameter tuning depends on several factors: sample size, classifier/regression models used, and model type [55-57]. It's an additional step to improve the accuracy and performance of the model [58]. For example, selection of the

best polynomial features in linear regression models, number of trees in a random forest, number of layers and neurons in a neural network, maximum depth in decision trees, and learning rate for gradient descent [58]. Some common hyper parameter tuning techniques are grid search, randomized search, Bayesian optimization, sequential model-based optimization, and genetic algorithms [55].

1.3. Overfitting and Underfitting in ML models

Overfitting occurs when the model learns the noise and unwanted details in the learning data, which negatively impacts predicting the new data [59]. Underfitting refers to a model that neither models the training data nor generalizes to new data [60-63]. In comparing classifiers, we emphasize that more than just overall accuracy should be considered; the user's and producer's accuracies for individual classes should also be considered [64]. This is particularly true if the mapping focuses on rare classes (i.e. classes of limited extent in the image data). Rare classes tend to have little effect on the overall accuracy but may nevertheless be vital in determining the usefulness of the classification [65]. However, if it is not feasible to test a variety of classifiers, SVM and R.F. generally appear to be reliable classification methods [65]. Some common approaches to reduce overfitting and underfitting of ML models are to use cross entropy, cross validation, early stopping and regularization approach.

1.3.1 Cross-Entropy and Cross-Validation

Entropy is a measurable physical quality most usually linked with disorder, unpredictability, or uncertainty [66]. The smallest average encoding size per transmission with which a source can efficiently convey a message to a destination without losing any data is defined as entropy [67]. The difference between two probability distributions for a given random variable or set of occurrences is measured by cross-entropy. [68-72]. As a loss function, cross-entropy is extensively employed in ML [73]. Each example has a known class label with a probability of 1.0, whereas all other labels have a probability of 0.0 in classification [74]. In this case, the model determines the probability that a given example corresponds to each class label [75]. Cross-entropy can then be used to calculate the difference between two probability distributions [69].

1.3.2. Cross-Validation

Cross-validation is a technique in which we train our model using the subset of the dataset and then evaluate it using the complementary subset of the dataset [76]. It is useful when there is a limited amount of data available [77]. An example of cross-validation is K-fold cross-validation [78]. The data is divided into K parts, where 1 part is used as a validation dataset and the other remaining as a training dataset [79]. And this process is repeated K times to reduce the biases and produce an effective model [80].

1.3.3. Early Stopping and Regularization

Early stopping and regularization are other techniques used to reduce the overfitting of the data [81]. The early stopping technique stops the training on ML models once the ML model's performance starts dropping and then increasing [82]. The regularization technique can be applied in multiple ways. Their examples are L1, L2, and Dropout regularization [83].

1.4. Model performance calculation

1.4.1. Confusion matrix (Popular for classification models)

A confusion matrix, also known as an error matrix, is used for classification models [84]. These matrices help in evaluating, monitoring and managing models [85]. From these matrices, we can develop metrices like accuracy, precision, recall, specificity, and F1 score [85]. When we create a confusion matrix, positive observation is known as Positives (P) [86], negative observation is known as Negative (N) ([86]), an outcome where the model correctly predicts the positive class is called True Positives (T.P.). In this outcome, the model correctly predicts the negative classes are, known as True Negatives (T.N.). The model incorrectly predicts the positive class when negative, also called a type 1 error are False Positive (F.P.)[87]. An outcome where the model incorrectly predicts the negative class when it is positive also called a type 2 error, is known as a False Negative (F.N.)[87]. We should learn about the accuracy, precision, recall, specificity, F1 score to read and interpret the output of the confusion matrix.

- **Accuracy**

Accuracy can be calculated by using the following formula

$$Accuracy = \frac{\#\ of\ correct\ predictions}{total\ \#\ of\ predictions} = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Precision**

Precision can be calculated by using the following formula

$$Precision = \frac{TP}{TP + FP}$$

- **Recall**

Recall, also known as the sensitivity**,** hit rate**,** or the true positive rate (T.P.R.) [88]. Answers the question, "What proportion of actual positives were identified correctly?"

$$Recall = \frac{TP}{TP + FN}$$

- **Specificity**

Specificity, also known as the true negative rate (TNR), measures the proportion of actual negatives that are correctly identified as such [89]. It is the opposite of recall.

$$Specificity = \frac{TN}{TN + FP}$$

- **F1 score**

The F1 score measures a test's accuracy-it is the harmonic mean of precision and recall [90].

It can have a maximum score of 1 (perfect precision and recall) and a minimum of 0. Overall, it measures the preciseness and robustness of your model [91].

$$F1\ score = \frac{2 * (precision * recall)}{precision + recall} = \frac{2TP}{2TP + FP + FN}$$

Receiver operator characteristic curve

Receiver operator characteristic (R.O.C.) analysis is a quantitative method for determining a binary classification based on a threshold (cut-off) value usually calculated from continuous data [92-94]. Plotting the true positive rate (T.P.R.) against the false positive rate (F.P.R.) at various threshold levels yields the R.O.C. curve[95].Sensitivity, recall, and the chance of detection are all terms used to describe the true-positive rate [96, 97]. The likelihood of a false alarm is also known as the false-positive rate, and it can be computed as (1- specificity). It's also known as a plot of the power as a function of the decision rule's Type I Error (when the performance is calculated from just a sample of the population, it can be thought of as estimators of these quantities). As a result, the R.O.C. curve represents sensitivity or recall as a function of fall-out.

1.4.2. For regression models

Classifying the accuracy of the regression models is a little different than the classification models because, in regression models, we are not only concerned with the model predicting right or wrong but also how accurately the models have predicted the actual value [98]. For example, when we use regression models to forecast the temperature, if the model gives the value as 43 C and the actual value is 43.5 C, the model is better and vice versa [99]. We measure the accuracy of the regression models using explained variance and mean squared error

- **Explained variance**

Explained variance is the amount of variation in the original dataset that our model can explain [98, 100].

- **Mean squared error**

It is the average of the squared differences between the predicted and actual output. R2 coefficient represents the proportion of variance in the outcome that our model can predict based on its features [101].

1.5. Factors to consider while selecting the ML models in GIS and remote sensing based projects

- No rule of thumb
- Experiment with multiple classifiers
- Hyper tuning parameters for the accuracy
- Use random forest classifiers for the weak datasets and Decision tress when simple and fast models are needed
- The default value for the number of trees in R.F. can be 500; for kernel size in SVM, it can be polynomial kernels and radial basis kernels
- Visualize the relationships between the input and predictors to evaluate their relationship and find if there is any band that can help in predicting things better
- Normalize the rare classes/imbalanced datasets
- Computation time also depends on user-defined parameters, classifier chosen, sample size
- If parameters cannot be tuned, R.F. should be used, setting the number of trees to 500 to provide
- Balance the datasets/data normalization. The classes with few samples/rare classes can be affected
- Computational complexities of different ML models which is the amount of resources to run a ML model.

N=number of training examples, m=number of features, n'=number of support vectors, k=number of neighbors, k'= number of trees [102].

Table 1: Computational complexity of discussed ML models

S.N.	Model	Train time complexity	Test time complexity	Space complexity
1	Linear regression	O(n*m^2 + m^3)	O(m)	O(m)
2	Logistic regression	O(n*m)	O(m)	O(m)
3.	Support Vector Machine	O(n^2)	O(n'*m)	O(n*m)
4.	Decision tree	O(n*log(n)*m	O(m)	O(depth of tree)
5.	Random forest	O(k' *n*log(n)*m	O(m*k')	O(k'*depth of tree)
6.	Naïve Bayes	O(n*m)	O(m)	O(c*m)

2. Conclusion

In recent years, ML models are increasingly being used in GIS and remote sensing based projects. ML models helps in solving GIS and remote sensing problems by identifying the underlying patterns, for example satellite based image classification [103], detection of features likes roads, wetlands, grasslands, image segmentation. We discuss few popular ML models and methods of their application in GIS and remote sensing based projects here. Researchers can use this paper as a reference while starting a ML based project. There are other ML models which can be learnt easily after learning above discussed models.

3. Conflict of Interest: None

4. References

1. Lary DJ, Alavi AH, Gandomi AH, Walker ALJGF. Machine learning in geosciences and remote sensing. 2016; 7:3-10.
2. Verma VK, Verma SJMTP. Machine learning applications in healthcare sector: An overview, 2021.
3. Maxwell AE, Warner TA, Fang FJIJoRS. Implementation of machine-learning classification in remote sensing: An applied review. 2018; 39:2784-2817.
4. Sarker IHJSCS. Machine learning: Algorithms, real-world applications and research directions. 2021; 2:1-21.
5. Sidey-Gibbons JA, Sidey-Gibbons CJJB mrm. Machine learning in medicine: a practical introduction. 2019; 19:1-18.
6. Xu C, Jackson SAJGb. Machine learning and complex biological data. 2019; 20:1-4.
7. Greener JG, Kandathil SM, Moffat L. Jones DTJNRMCB. A guide to machine learning for biologists. 2022; 23:40-55.
8. Avand M, Moradi HJJoH. Using machine learning models, remote sensing, and GIS to investigate the effects of changing climates and land uses on flood probability. 2021; 595:125663.
9. Bahari NIS, Ahmad A, Aboobaider BM. Application of support vector machine for classification of multispectral data. In Proceedings of the IOP Conference Series: Earth and Environmental Science, 2014, 012038.
10. Yu XJNH. Disaster prediction model based on support vector machine for regression and improved differential

evolution. 2017; 85:959-976.
11. Sun L, Zhu Z. Linear programming Monte Carlo method based on remote sensing for ecological restoration of degraded ecosystem. In Proceedings of the IOP Conference Series: Earth and Environmental Science, 2019, 012041.
12. Ise T, Oba YJFiR, AI. Forecasting climatic trends using neural networks: an experimental study using global historical data, 2019, 32.
13. Wang J, Bretz M, Dewan MAA, Delavar MAJSoTTE. Machine learning in modelling land-use and land cover-change (LULCC): Current status, challenges and prospects, 2022, 153559.
14. Assarkhaniki Z, Sabri S, Rajabifard AJBED. Using open data to detect the structure and pattern of informal settlements: an outset to support inclusive SDGs' achievement. 2021; 5:497-526.
15. Hossain MT, Zarin T, Sahriar MR, Haque MNJP. Chemistry of the Earth, P.A.B.C. Machine learning based modeling for future prospects of land use land cover change in Gopalganj District, Bangladesh. 2022; 126:103022.
16. Das M. Online prediction of derived remote sensing image time series: An autonomous machine learning approach. In Proceedings of the IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium, 2020, 1496-1499.
17. Palacios Salinas NR, Baratchi M, Rijn JNv, Vollrath A. Automated machine learning for satellite data: integrating remote sensing pre-trained models into AutoML systems. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, 2021, 447-462.
18. Odebiri O, Odindi J, Mutanga OJIJoAEO. Geo information. Basic and deep learning models in remote sensing of soil organic carbon estimation: A brief review. 2021; 102:102389.
19. Das M, Ghosh SK, Chowdary VM, Mitra P, Rijal SJRS. Statistical and Machine Learning Models for Remote Sensing Data Mining-Recent Advancements. 2022; 14:1906.
20. Zerrouki N, Harrou F, Sun Y, Hocini LJISJ. A machine learning-based approach for land cover change detection using remote sensing and radiometric measurements. 2019; 19:5843-5850.
21. Liu R, Chen Y, Wu J, Gao L, Barrett D, Xu T, *et al*. Integrating entropy-based naïve Bayes and GIS for spatial evaluation of flood hazard. 2017; 37:756-773.
22. Sitthi A, Nagai M, Dailey M, Ninsawat SJS. Exploring land use and land cover of geotagged social-sensing images using naive bayes classifier. 2016; 8:921.
23. Tien Bui D, Shahabi H, Shirzadi A, Chapi K, Hoang ND, Pham BT, *et al*. A novel integrated approach of relevance vector machine optimized by imperialist competitive algorithm for spatial modeling of shallow landslides. 2018; 10:1538.
24. Chen H, Hu S, Hua R, Zhao XJEJoAiSP. Improved naive Bayes classification algorithm for traffic risk management, 2021, 1-12.
25. Mitchell TMJML. Logistic regression. 2005; 10:701.
26. El-Magd A, Ahmed SJAJoG. Random forest and naïve Bayes approaches as tools for flash flood hazard susceptibility prediction, South Ras El-Zait, Gulf of Suez Coast, Egypt. 2022; 15:1-12.
27. Belgiu M, Drăguţ LJIjop. sensing, r. Random forest in remote sensing: A review of applications and future directions. 2016; 114:24-31.
28. Pal MJIjors. Random forest classifier for remote sensing classification. 2005; 26:217-222.
29. Piramanayagam S, Schwartzkopf W, Koehler F, Saber E. Classification of remote sensed images using random forests and deep learning framework. In Proceedings of the Image and signal processing for remote sensing XXII, 2016, 205-212.
30. Berhane TM, Lane CR, Wu Q, Autrey BC, Anenkhonov OA, Chepinoga VV, Liu HJRs. Decision-tree, rule-based, and random forest classification of high-resolution multispectral imagery for wetland mapping and inventory. 2018; 10:580.
31. Kulkarni AD, Lowe B. Random forest algorithm for land cover classification, 2016.
32. Prasai R, Schwertner TW, Mainali K, Mathewson H, Kafley H, Thapa S, *et al*. Application of Google earth engine python API and NAIP imagery for land use and land cover classification: A case study in Florida, USA. Ecological Informatics. 2021; 66:101474. https://doi.org/10.1016/j.ecoinf.2021.101474.
33. Prasai R. An open-source web-based tool to perform spatial multicriteria analysis. International Journal of Multidisciplinary Research and Growth Evaluation, 2022a, 297-301. https://doi.org/10.54660/anfo.2022.3.3.19.
34. Chan JCW, Huang C, Defries RJIToG, Sensing R. Enhanced algorithm performance for land cover classification from remotely sensed data using bagging and boosting. 2001; 39:693-695.
35. Chan JCW, Paelinckx DJRSoE. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. 2008; 112:2999-3011.
36. Rodriguez-Galiano VF, Ghimire B, Rogan J, Chica-Olmo M, Rigol-Sanchez JPJIjop, sensing R. An assessment of the effectiveness of a random forest classifier for land-cover classification. 2012; 67:93-104.
37. Mountrakis G, Im J, Ogole CJIJoP, Sensing R. Support vector machines in remote sensing: A review. 2011; 66:247-259.
38. Sheykhmousa M, Mahdianpari M, Ghanbari H, Mohammadimanesh F, Ghamisi P, Homayouni SJIJoSTiAEO, Sensing R. Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review. 2020; 13:6308-6325.
39. Cavallaro G, Willsch D, Willsch M, Michielsen K, Riedel M. Approaching remote sensing image classification with ensembles of support vector machines on the d-wave quantum annealer. In Proceedings of the IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium, 2020, 1973-1976.
40. Waske B, Benediktsson JA, Sveinsson JR. Classifying remote sensing data with support vector machines and imbalanced training data. In Proceedings of the International Workshop on Multiple Classifier Systems, 2009, 375-384.
41. Huang C, Davis L, Townshend JJIJors. An assessment of support vector machines for land cover classification. 2002; 23:725-749.
42. Sudalaimuthu K, Sudalayandi K. Development of linear

regression model to predict ground elevation from satellite elevation–statistical approach. In Proceedings of the AIP Conference Proceedings, 2019, 020031.
43. Mansouri E, Feizi F, Jafari Rad A, Arian MJSE. Remote-sensing data processing with the multivariate regression analysis method for iron mineral resource potential mapping: a case study in the Sarvian area, central Iran. 2018; 9:373-384.
44. Susiluoto J, Spantini A, Haario H, Härkönen T, Marzouk YJGMD. Efficient multi-scale Gaussian process regression for massive remote sensing data with sat GP. 2020; 13:3439-3463.
45. Shi X, Li Y, Zhao QJRS. Flexible hierarchical Gaussian mixture model for high-resolution remote sensing image segmentation. 2020; 12:1219.
46. Ghosh SS, Dey S, Bhogapurapu N, Homayouni S, Bhattacharya A, McNairn HJRS. Gaussian Process Regression Model for Crop Biophysical Parameter Retrieval from Multi-Polarized C-Band SAR Data. 2022; 14:934.
47. Graff CA, Coffield SR, Chen Y, Foufoula-Georgiou E, Randerson JT, Smyth PJIToG, *et al*. Forecasting daily wildfire activity using poisson regression. 2020; 58:4837-4851.
48. Liang D, Frederick DA, Lledo EE, Rosenfield N, Berardi V, Linstead E, *et al*. Examining the utility of nonlinear machine learning approaches versus linear regression for predicting body image outcomes: The US Body Project I. 2022; 41:32-45.
49. Adsuara JE, Pérez-Suay A, Muñoz-Marí J, Mateo-Sanchis A, Piles M, Camps-Valls GJIToG, *et al*. Nonlinear distribution regression for remote sensing applications. 2019; 57:10025-10035.
50. Hsieh WWJapa. Improving predictions by nonlinear regression models from outlying input data, 2020.
51. Paulson NH, Kubal J, Ward L, Saxena S, Lu W, Babinec SJJJoPS. Feature engineering for machine learning enabled early prediction of battery lifetime. 2022; 527:231127.
52. Sarith Divakar M, Sudheep Elayidom M, Rajesh R. Feature Engineering of Remote Sensing Satellite Imagery Using Principal Component Analysis for Efficient Crop Yield Prediction. In Evolutionary Computing and Mobile Sustainable Networks; Springer, 2022, 189-199.
53. Song H, Yang W, Dai S, Yuan HJMB. Engineering. Multi-source remote sensing image classification based on two-channel densely connected convolutional networks. 2020; 17:7353-7378.
54. Schapire REJNe. Classification. The boosting approach to machine learning: An overview. 2003, 149-171.
55. Mayr A, Binder H, Gefeller O, Schmid MJMoiim. The evolution of boosting algorithms. 2014; 53:419-427.
56. Jafarzadeh H, Mahdianpari M, Gill E, Mohammadimanesh F, Homayouni SJRS. Bagging and boosting ensemble classifiers for classification of multispectral, hyperspectral and PolSAR data: a comparative evaluation. 2021; 13:4405.
57. Yang L, Shami AJN. On hyperparameter optimization of machine learning algorithms: Theory and practice. 2020; 415:295-316.
58. Xu S, Zhao Y, Wang M, Shi XJEJoSS. A comparison of machine learning algorithms for mapping soil iron parameters indicative of pedogenic processes by hyperspectral imaging of intact soil profiles. 2022; 73:e13204.
59. Audebert N, Le Saux B, Lefèvre SJIg, magazine RS. Deep learning for classification of hyperspectral data: A comparative review. 2019; 7:159-173.
60. Pannakkong W, Thiwa-Anont K, Singthong K, Parthanadee P, Buddhakulsomsiri JJMPiE. Hyperparameter Tuning of Machine Learning Algorithms Using Response Surface Methodology: A Case Study of ANN, SVM, and DBN, 2022.
61. Gu Y, Wylie BK, Boyte SP, Picotte J, Howard DM, Smith K, *et al*. An optimal sample data usage strategy to minimize overfitting and underfitting effects in regression tree models based on remotely-sensed data. 2016; 8:943.
62. Van der Aalst WM, Rubin V, Verbeek H, van Dongen BF, Kindler E, Günther CWJS, *et al*. Process mining: a two-step approach to balance between underfitting and overfitting. 2010; 9:87-111.
63. Guyon X, Yao JfJJoMA. On the underfitting and overfitting sets of models chosen by order selection criteria. 1999; 70:221-249.
64. Bashir D, Montañez GD, Sehra S, Segura PS, Lauw J. An information-theoretic perspective on overfitting and underfitting. In Proceedings of the Australasian Joint Conference on Artificial Intelligence, 2020, 347-358.
65. Jabbar H, Khan RZJCS. Communication; Devices, I. Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). 2015, 70.
66. Alnaim A, Sun Z, Tong DJRS. Evaluating Machine Learning and Remote Sensing in Monitoring NO2 Emission of Power Plants. 2022; 14:729.
67. Bogner C, Seo B, Rohner D, Reineking BJPo. Classification of rare land cover types: Distinguishing annual and perennial crops in an agricultural catchment in South Korea. 2018; 13:e0190476.
68. Huang EW, Lee WJ, Singh SS, Kumar P, Lee CY, Lam TN, *et al*. Machine-learning and high-throughput studies for high-entropy materials. 2022; 147:100645.
69. Janik RAJapa. Entropy from Machine Learning, 2019.
70. Ruby U, Yendapalli VJIJATCSE. Binary cross entropy with deep learning technique for image classification, 2020, 9.
71. Gordon-Rodriguez E, Loaiza-Ganem G, Pleiss G, Cunningham JP. Uses and abuses of the cross-entropy loss: Case studies in modern deep learning, 2020.
72. Wang ZZ, Goh SHJAG. A maximum entropy method using fractional moments and deep learning for geotechnical reliability analysis. 2022; 17:1147-1166.
73. Pacheco AG, Ali AR, Trappenberg TJapa. Skin cancer detection based on deep learning and entropy to detect outlier samples. 2019.
74. Brochet T, Lapuyade-Lahorgue J, Bougleux S, Salaün M, Ruan SJI. Deep Learning Using Havrda-Charvat Entropy for Classification of Pulmonary Optical Endomicroscopy. 2021; 42:400-406.
75. Juszczuk P, Kozak J, Dziczkowski G, Głowania S, Jach T, Probierz BJE. Real-World Data Difficulty Estimation with the Use of Entropy. 2021; 23:1621.
76. Ho Y, Wookey SJIA. The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling. 2019; 8:4806-4813.
77. Singh VP. Entropy theory and its application in

environmental and water engineering; John Wiley & Sons, 2013.
78. Tougui, I.; Jilbab, A.; El Mhamdi, J.J.H.i.r. Impact of the choice of cross-validation techniques on the results of machine learning-based diagnostic applications. 2021, 27, 189-199.
79. Battula KJIJoE. Technology, A. Research OF machine learning algorithms using K-fold cross validation. 2021; 8:215-218.
80. Learn SJLAHS.-LOSMCVHC. Cross-validation: evaluating estimator performance, 2022.
81. Phinzi K, Abriha D, Szabó SJRS. Classification efficacy using k-fold cross-validation and bootstrapping resampling techniques on the example of mapping complex gully systems. 2021; 13:2980.
82. A Ramezan C, A Warner T, E Maxwell AJRS. Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification. 2019; 11:185.
83. Brownlee JJUhmce.-s.-t.-a.-n.-n.-m. A gentle introduction to early stopping to avoid overtraining neural networks, 2018.
84. Behnke M, Guo SJPM. Comparison of early stopping neural network and random forest for in-situ quality prediction in laser based additive manufacturing. 2021; 53:656-663.
85. Alem A, Kumar SJAAI. Transfer Learning Models for Land Cover and Land Use Classification in Remote Sensing Image. 2022; 36:2014192.
86. Piscini A, Carboni E, Del Frate F, Grainger RGJAiRS. A Neural Network Algorithm to Detect Sulphur Dioxide Using IASI Measurements, 2014.
87. Jeong S, Ko J, Shin T, Yeom JmJSR. Incorporation of machine learning and deep neural network approaches into a remote sensing-integrated crop model for the simulation of rice growth. 2022; 12:1-10.
88. Weaver J, Moore B, Reith A, McKee J, Lunga D. A comparison of machine learning techniques to extract human settlements from high resolution imagery. In Proceedings of the IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, 2018, 6412-6415.
89. Bosman HH, Liotta A, Iacca G, Wörtche HJ. Anomaly detection in sensor systems using lightweight machine learning. In Proceedings of the 2013 IEEE International Conference on Systems, Man, and Cybernetics, 2013, 7-13.
90. Alakus TB, Turkoglu IJC, Solitons; Fractals. Comparison of deep learning approaches to predict COVID-19 infection. 2020; 140:110120.
91. Erickson BJ, Kitamura FJRAI. Magician's corner: 9. Performance metrics for machine learning models, 2021, 3.
92. Chicco D, Jurman GJBG. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. 2020; 21:1-13.
93. Goutte C, Gaussier E. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In Proceedings of the European conference on information retrieval, 2005, 345-359.
94. Carter JV, Pan J, Rai SN, Galandiuk SJS. ROC-ing along: Evaluation and interpretation of receiver operating characteristic curves. 2016; 159:1638-1645.
95. Mbizvo GK, Larner AJJNDM. Receiver operating characteristic plot and area under the curve with binary classifiers: pragmatic analysis of cognitive screening instruments. 2021; 11:353-360.
96. Søreide KJJocp. Receiver-operating characteristic curve analysis in diagnostic, prognostic and predictive biomarker research. 2009; 62:1-5.
97. Brownlee JJMlm. How to use ROC curves and precision-recall curves for classification in Python, 2018, 30.
98. Statnikov A, Aliferis CF, Hardin DP, Guyon I. Gentle Introduction to Support Vector Machines In Biomedicine, A-Volume 2: Case Studies and Benchmarks; World Scientific Publishing Company, 2013.
99. LeDell E, Petersen M, van der Laan MJEjos. Computationally efficient confidence intervals for cross-validated area under the ROC curve estimates. 2015; 9:1583.
100. Goldstein HJEosibs. Heteroscedasticity and complex variation. 2005; 2:790-795.
101. Haack B, Rafter AJGI. Regression estimation techniques with remote sensing: a review and case study. 2010; 25:71-82.
102. Estrella EH, Gilerson A, Foster R, Groetsch PJJoARS. Spectral decomposition of remote sensing reflectance variance due to the spatial variability from ocean color and high-resolution satellite sensors. 2021; 15:024522.
103. Prasai R. Earth engine application to retrieve long-term terrestrial and aquatic time series of satellite reflectance data. International Journal of Multidisciplinary Research and Growth Evaluation, 2022b, 165-171. https://doi.org/10.54660/anfo.2022.3.3.11.