# A Review of Ensemble Learning Algorithms Used in Remote Sensing Applications

Yuzhen Zhang [1,*], Jingjing Liu [1] and Wenjuan Shen [2,3]

1   School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China
2   College of Forestry, Nanjing Forestry University, Nanjing 210037, China
3   Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing 210037, China
*   Correspondence: yzhang@ustb.edu.cn

**Abstract:** Machine learning algorithms are increasingly used in various remote sensing applications due to their ability to identify nonlinear correlations. Ensemble algorithms have been included in many practical applications to improve prediction accuracy. We provide an overview of three widely used ensemble techniques: bagging, boosting, and stacking. We first identify the underlying principles of the algorithms and present an analysis of current literature. We summarize some typical applications of ensemble algorithms, which include predicting crop yield, estimating forest structure parameters, mapping natural hazards, and spatial downscaling of climate parameters and land surface temperature. Finally, we suggest future directions for using ensemble algorithms in practical applications.

**Keywords:** bagging; boosting; stacking; remote sensing applications; ensemble

## 1. Introduction

Remote sensing is widely used in applications such as military reconnaissance, analysis of natural hazards, detection of land use and land cover change, measurement of sea ice distribution, precision agriculture, and estimation and mapping of carbon stocks [1,2]. The primary properties of the target detected by the sensor in these applications are in the form of spatial, spectral, temporal, and polarization signatures [3]. Many models of physical processes have been developed that quantify the data relationships between sensor and target. These models include radiative transfer models, radiosity models, ray-tracing models, dynamic global vegetation models, and land surface microwave emission models, among others [4–9]. However, it is generally difficult to obtain the parametric input required for a physical process model, especially at a large scale. Some studies have coupled a physical process model with a statistical method or used a statistical method with a machine learning algorithm to extract crucial predictions for a specific application. Since the remote sensing signals often provide a nonlinear representation of the target, machine learning and deep learning algorithms have been widely used in remote sensing applications, because they are not based on some underlying assumption regarding the distribution of the data and they have a potent capacity to capture nonlinear correlations [10,11]. Support vector machines (SVM) and random forests (RF) are two well-known nonparametric machine learning algorithms that are used in many studies [12–14].

Some studies have claimed that no single algorithm could outperform every other machine learning algorithm under all situations (sometimes called the no free lunch theorem) [15]. The ensemble learning technique has been developed in response to this claim [16,17]. Ensemble learning techniques create multiple hypotheses and combine them to solve a given problem, in contrast to conventional machine learning techniques that aim to learn a single hypothesis from training data. [18,19]. By combining multiple learners

and taking full advantage of these learners, ensemble algorithms have produced improved results and reduced the overfitting problem, and they possess the flexibility to deal with different tasks [17,20,21]. Bagging, boosting, and stacking are three well established ensemble techniques, although there exist some variants and other ensemble algorithms that have been used in practical applications [17]. Current implementations of these algorithms in R and Python are shown in Table 1.

**Table 1.** Implementation of ensemble learning algorithms using R and Python.

| Algorithm | Implementation with R and Python | |
| --- | --- | --- |
| | R Package | Python Library |
| Bagging | ipred | scikit-learn |
| RF | randomForest/caret | scikit-learn |
| ERT | extraTrees | scikit-learn |
| AdaBoost | Adabag/fastAdaboost | scikit-learn |
| GBM | gbm | scikit-learn |
| XGBoost | xgboost/plyr | xgboost |
| LightGBM | lightgbm | lightgbm |
| CatBoost | catboost | catboost/scikit-learn |
| Stacking | stacks | vecstack/scikit-learn |

We present a general overview of the bagging, boosting, and stacking ensemble techniques used in remote sensing. We first review the principles underlying the three types of ensemble algorithms in Section 2 and describe our literature search for publications about bagging, boosting, and stacking ensembles in Section 3. In Section 4, we focus on several fields of study that widely use ensemble techniques for objectives such as crop yield prediction, estimation of forest structure parameters, mapping natural hazards, and spatial downscaling, as well as other applications. Finally, we examine the issues related to ensemble algorithms and future directions.

## 2. Principles of Ensemble Learning Algorithms

### 2.1. Bagging Algorithms

Bagging is an ensemble technique that combines multiple learners trained on distinct subsamples of the original data. To build a bagging model, we generate multiple datasets by bootstrapping the training data and then develop models based on the individual datasets and make predictions using these models. All the predictions are then combined to produce a representative value such as the mean, median, or majority vote for classification and averaging, for regression, depending on the problem to be solved. Since an individual learner is often sensitive to noise in the training data, bagging, by aggregating multiple results in a single prediction, should provide stable and improved results with decreased variance [22].

RF is a modified version of bagging, in which the classification and regression trees (CART) technique is often used as the individual learner. RF uses subsamples of the original data to build trees and randomly selects a subset of variables to determine each split in the tree [23]. RF excludes approximately 30% of training samples from the modeling due to bootstrapping and random subspace techniques and is often used to compute the out-of-bag (OOB) prediction error.

Many studies have shown that RF outperforms other machine learning and statistical regression algorithms [13]. It has many advantages, some of which are the following. Both classification and regression issues can be resolved using RF. In addition, RF is insensitive to noise or outliers in the training data because it bins them. RF trains rapidly since it incorporates only part of the features for training. RF is easy to use since only one or two hyperparameters (i.e., the number of trees in the forest and the number of predictor variables to be randomly selected at each junction) are required to be tuned. In addition, an intrinsic estimate of the generalization error (the OOB error) is given. RF provides

a measure of variable importance as identified by the Gini index [24]. Despite these advantages, the spatial correlation of nearby observable data is disregarded by RF. Thus, RF kriging (RFK), which couples RF with residuals interpolated by conventional kriging, was proposed and successfully used for forest biomass mapping [25] and prediction of PM2.5 concentrations [26].

Extremely randomized trees (ERT) is also a tree-based ensemble method. ERT differs from RF in the following ways: ERT uses all samples to train the base learner instead of using bootstrap resampling; and ERT randomly selects cutting points instead of calculating local optimal cutting points when splitting nodes, which injects more randomness than RF [22]. The results from different individual trees are therefore more diverse. Some studies have suggested that ERT produces more accurate predictions than RF [27]. ERT has been used in a wide variety of applications such as aboveground biomass estimation [28], prediction of ground level PM2.5 concentrations [29], modeling olive tree phenology [30], retrieval of downward longwave radiation [31], streamflow modeling [32], and estimation of terrestrial latent heat flux [33].

### 2.2. Boosting Algorithms

Boosting algorithms use a forward stagewise process to transform weak learners into strong learners by increasing the weights of training samples that were mistakenly identified or wrongly calculated in a successive iteration. The final output of boosting is obtained by combining the results from all iterations using a weighted vote for classification or a weighted sum for regression [34]. Widely used boosting ensemble algorithms include adaptive boosting (AdaBoost), gradient boosting machine (GBM), extreme gradient boosting (XGBoost), LightGBM, and categorical boosting (CatBoost).

AdaBoost was initially developed to increase the efficiency of binary classifiers. In GBM, all the weak learners of GBM are decision trees. XGBoost is an improved version of GBM that implements parallel preprocessing at the node level, which makes it faster than GBM. XGBoost also introduces a variety of regularization techniques to reduce overfitting [35]. Studies have shown that of XGBoost is excellent in mapping plant species diversity [36], predicting PM2.5 concentrations [37], forest biomass estimation [38], and risk analysis for flash floods [39]. LightGBM has many of XGBoost's advantages, such as parallel training, regularization, and sparse optimization. The major difference between the two lies in the method of constructing trees. LightGBM uses a leafwise split: after the first split, the node with a higher delta loss is selected for the next split [40]. This technique enables LightGBM to easily handle huge amounts of data. A histogram-based method is often adopted to select the best split in LightGBM to reduce the time used in training. However, LightGBM cannot perform well with a small number of samples.

CatBoost is an improved gradient boosting decision tree (GBRT) algorithm and an alternative to XGBoost. As the name suggests, CatBoost can internally handle categorical variables in the data, and is thus suitable for dealing with machine learning tasks involving categorical and heterogeneous data [41,42]. Studies have found that CatBoost is superior to other machine learning algorithms, and it has been used to estimate forest biomass [28,43] and reference evapotranspiration [44].

### 2.3. Stacked Generalization

Stacked generalization, also known as stacking, was proposed by Wolpert in 1992 [45]. It is a heterogeneous learning technique that combines diverse base learners by training a model, unlike the homogeneous bagging and boosting methods which directly aggregate the outputs of several learners to obtain the final prediction. If properly designed, stacking can take full advantage of different base learners and should perform better than an individual base learner, whether using majority voting or weighted averaging [20,46–48].

Generally, stacking consists of several base learners (level 0) and a meta learner (level 1), in which the outputs of base learners serve as inputs of the meta learner. Both the precision and variety of base learners affect the performance of a stacking algorithm [49,50].

Diversity is a measure of the dependence or complementariness among learners [51]. That base learners have high diversity implies that they are skilled in different ways, and thus stacking them could lead to improved results. There are many measures of diversity, such as Q-statistics in classification, and the variance of ensemble predictions around their weighted mean in regression [52]; no one measure has been shown to be the best [51,53,54]. In addition to the diversity and accuracy of base learners, the number of base learners also affects the performance of stacking. More base learners are not always associated with more accurate predictions, but they do require additional memory usage and computing time [55]. Some studies have suggested that three to four base learners stacked together are a suitable choice [56], and some action should be taken to ensure that an optimal subset of base learners is used in stacking.

A two-layer stacking model is often adopted in practical applications, although stacking models with more than two layers are sometimes used to further improve results [57]. A two-layer stacking model is constructed as follows (Figure 1). A k-fold cross-validation is initially used to create k verification datasets on all the base learners. The cross-validated predictions are then used as new training data for the meta learner in the second layer; the number of base learners in the first layer is equal to the number of predictors in the second layer, and the test data are created by adding the averages of the k-fold predictions of the base learners [47,58]. In some classification studies, class probabilities were used instead of predicted classes as input attributes for the meta learner, which provided an effective way of combining base model confidences [20]. The inclusion of engineered original features in stacking has been shown to give better performance [58,59]. There have also been studies that used weighted stacking to improve results [60].
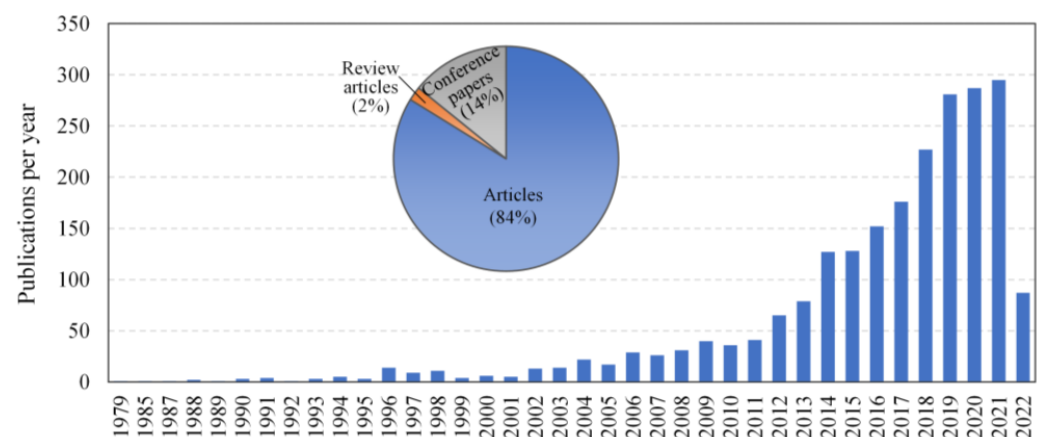


**Figure 1.** Statistics of ensemble literature in the refined areas.

### 3. Literature Search and Analysis

We conducted a literature survey of the ISI Web of Science database using the search term TOPIC ensemble AND TOPIC remote sensing. The search was then refined to include research areas Remote Sensing, Geography, Agriculture, Forestry and Environmental Sciences Ecology and document types Articles, Meeting, and Review articles. This yielded a total of 2247 results that we refer to as ensemble literature.

Research articles accounted for 84% of the records returned, conference papers for 14% and review articles for 2% (Figure 1). The publication years of the refined search records returned ranged from 1979 to 2022. In 2015–2022, more than 100 papers meeting the search criteria were published annually (Figure 1).

We then constructed a co-occurrence network by identifying keywords in the searched ensemble literatures, calculating the frequencies of their co-occurrences, and analyzing the networks to find central words and clusters of themes in the network (Figure 2). The constructed network showed that studies of ensemble learning algorithms were mainly related to classification, and studies of ensemble method use or ensemble models were primarily concerned with RF. In addition, the co-occurrence network suggested that applications concerning crop yield, biomass, rainfall, landslides, and satellite products were connected to ensemble algorithms.
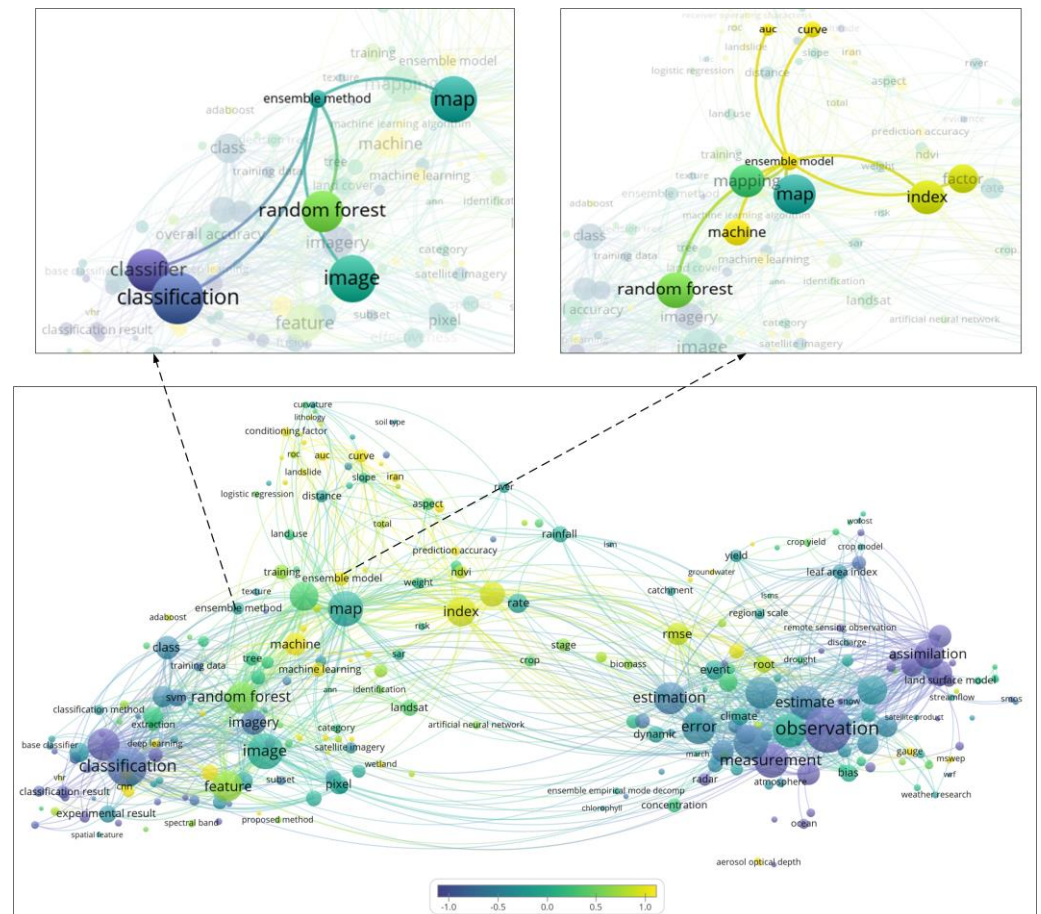


**Figure 2.** Co-occurrence network of keywords in ensemble literature.

## 4. Applications of Ensemble Learning Algorithms

The co-occurrence network (Figure 2) combined with our reading of the ensemble literature shows that typical applications of ensemble learning algorithms were mainly concerned with the following issues: yield prediction, forest structure and biomass estimation, mapping of natural hazards (e.g., land susceptibility to natural disasters), and spatial downscaling of land surface temperature and rainfall (Table 2). RF was the most frequently used ensemble algorithm (30 times) in the 43 applications listed in Table 2, followed by stacking (13 times), while boosting algorithms were less frequently used (Table 3). Only seven studies used XGBoost; GBRT and AdaBoost were used four times. MLP and kNN often served as reference models to evaluate the performance of an ensemble algorithm.

**Table 2.** Summary of reviewed applications that used ensemble learning algorithms.

| Reference | Application | Main Input Datasets or Causative Factors | Algorithms Used | Algorithm with the Best Accuracy |
|---|---|---|---|---|
| [61] | Oil palm yield prediction | Landsat time series imagery | RF, AdaBoost | RF |
| [62] | Flood susceptibility mapping | Slope, elevation, plan curvature, topographic wetness index (TWI), topographic position index, convergence index, stream power index (SPI), distance to stream, drainage density, rainfall, lithology, soil type, land use/land cover (LULC), and normalized difference vegetation index distance (NDVI) | J48 ensemble model, MultiBoosting J48, AdaBoost J48, random subspace J48 | Random subspace J48 |
| [63] | Drought risk assessment | Elevation, slope, distance from the stream, drainage density, temperature, humidity, precipitation, evaporation, soil moisture, soil depth, soil texture, NDVI, LULC, geomorphology, groundwater level, deep tone, agriculture-dependent population, and population density | SVM, RF, SVR, and their ensemble with bagging, boosting and Stacking | SVR-stacking |
| [64] | Downscaling climate variables | Sea surface temperature, air temperature, geopotential height, and sea level pressure | GBRT, SVR | GBRT |
| [65] | Flash flood susceptibility prediction | Altitude, slope, aspect, plan curvature, profile curvature, distance from river, distance from road, land use, lithology, soil depth, rainfall, SPI, and TWI | BRT, RF, PRF, RRF, ERT | ERT |
| [66] | Winter wheat yield prediction | MODIS EVI, climate data, and the subseasonal-to-seasonal (S2S) atmospheric prediction data | MLR, XGBoost, RF, SVR | XGBoost |
| [67] | Estimation of canopy height and growing stock volume | Airborne laser scanner (ALS), phased array type L-band synthetic aperture radar (SAR), and Landsat data | RF | — |
| [68] | Flood susceptibility mapping | Slope, elevation, plan curvature, NDVI, SPI, TWI, lithology, land use, rainfall, stream density, and distance to river | LMT, logistic regression, Bayesian logistic regression, RF, Bagging-LMT | Bagging-LMT |
| [25] | Forest biomass estimation | Advanced land observing satellite 2 L band, and Sentinel-1C band SAR, shuttle radar topography mission (SRTM) digital elevation model (DEM) data, and Sentinel-2 data | RF, RF Kriging | RF Kriging |
| [46] | Spatial interpolation of daily maximum air temperature | LST, NDVI, elevation, slope, aspect, solar radiation, global man-made impervious surface, human built-up and settlement extent, latitude, and longitude | Cokriging, MLR, SVR, RF, stacking, simple average ensemble | Stacking |
| [69] | Individual tree dendrometry | Field data, unmanned aerial vehicle(UAV)-Lidar data | SVR, MLP, RF, XGBoost | SVR |

**Table 2.** *Cont.*

| Reference | Application | Main Input Datasets or Causative Factors | Algorithms Used | Algorithm with the Best Accuracy |
|---|---|---|---|---|
| [70] | Reference evapotranspiration time series forecasting | Maximum and minimum temperature, wind speed at 2 m high, average relative humidity, and the insolation | Three CNN models CNN1, CNN2, CNN3, ensemble-CNN1, ensemble-CNN2, ensemble-CNN3, hybrid ensemble | Ensemble models |
| [71] | Short term electricity consumption forecasting | Electricity consumption | ANN, RF, GBRT, LR, DL, DT, evolutionary algorithms for regression trees, ARMA, ARIMA, stacking, | Stacking |
| [72] | Forest biomass estimation | ALS data and Landsat 8 imagery | ELM, BPNN, RT, RF, SVR, kNN, CNN, MLP, stacking | Stacking |
| [73] | Prediction of eucalyptus stand volume | SPOT-5 raw spectral features, spectral vegetation indices, rainfall data, and stand age | SGB, RF, stepwise MLR | SGB |
| [74] | Alfalfa yield prediction | Narrow-band indices (e.g., simple ratio index, NDVI, chlorophyll absorption ratio index, modified versions of these indices) derived from UAV-based hyperspectral images | RF, SVR, kNN, stacking | Stacking |
| [75] | Wheat grain yield prediction | Vegetation indices derived from UAV-based multispectral images | RF, SVR, GP, RR, stacking | Stacking |
| [76] | Estimation of forest variables | Statistics extracted from LiDAR data | MLR, MLP, SVR, kNN, GP, RT, RF | SVR |
| [77] | Forest biomass estimation | Multi-temporal Sentinel-1 and 2 data-derived variables (vegetation indices, SAR backscatter) | RF, GBM, XGBoost, ensemble model based on weighted averaging | Ensemble model |
| [78] | Land subsidence susceptibility mapping | Elevation, slope, aspect, profile curvature, plan curvature, TWI, distance to road, distance to river, distance to fault, precipitation, land use, lithology, drainage density, and groundwater drawdown | Logistic regression, MLP, AdaBoost and LogitBoost | AdaBoost |
| [79] | Forest disturbance detection | Landsat Thematic Mapper (TM) and Enhanced Thematic Mapper (ETM +) imagery | Eight automated change detection algorithms, stacking | Stacking |
| [80] | Downscaling LST | topographic variables derived from SRTM DEM data, land cover, and surface reflectance in visible red and near-infrared bands | RF, TsHARP | RF |
| [81] | Landslide susceptibility mapping | Altitude, slope, aspect, cross sectional curvature, profile curvature, plan curvature, longitudinal curvature, channel network base, convergence index, distance to fault, distance to river, valley depth, and lithology map | FDA, GLM, GBM, RF, ensemble based on weighted average | Ensemble |

**Table 2.** *Cont.*

| Reference | Application | Main Input Datasets or Causative Factors | Algorithms Used | Algorithm with the Best Accuracy |
|---|---|---|---|---|
| [82] | Estimating wheat yields | MODIS NDVI and climate data time series | RF, Cubist, XGBoost, MLP, SVR, GP, kNN, MARS, average ensemble, bayesian data fusion | SVR |
| [83] | Downscaling soil moisture | Sentinel-1 radar, monthly NDVI, land cover, topography, and surface soil properties. | RF | — |
| [84] | Winter wheat yield prediction | Spectral indices calculated from UAV-based hyperspectral data | SVR, GP, RR, RF, stacking | Stacking |
| [43] | Forest biomass estimation | Landsat spectral variables, vegetation indexes, texture measures, and terrain factors | RF, XGBoost, CatBoost | CatBoost |
| [47] | Landslide susceptibility mapping | Lithology, bedding structure, distance to fault, slope, aspect, plan curvature, profile curvature, elevation, distance to river, and NDVI | DBN, CNN, ResNet, stacking, simple averaging ensemble, weighted averaging ensemble, boosting | Stacking |
| [85] | Spatial prediction of landslide | Slope, aspect, elevation, curvature, lithology, land use, distance to roads, distance to faults, distance to rivers, and rainfall | LSSVM, MADT | MADT |
| [86] | Predicting flood probabilities | Elevation, slope angle, aspect, plan curvature, SPI, TWI, sediment transport index, drainage density, mean annual rainfall, proximity to rivers, proximity to roads, proximity to the coastline, soil texture, geology, land cover, wind speed, and mean sea level | LWLR, random subspace, REPTree, RF, M5P model tree, stacking | sSacking |
| [87] | Flood susceptibility mapping | Elevation, slope, aspect, NDVI, mean monsoonal rainfall, plan curvature, drainage density, population density, land cover, proximity to rivers, proximity to roads, geology, and soil texture | Bayesian regularization back propagation neural network, CART, EBF, weighted average ensemble algorithm | Weighted average ensemble algorithm |
| [88] | Forecasting agricultural commodity prices | Prices of energy commodities, exchange rate, interaction between commodities prices in domestic, and foreign markets | RF, GBM, XGBoost, stacking, MLP, SVR, kNN | XGBoost |
| [89] | Wheat yield prediction | Normalized difference red edge index, temperature, precipitation, relative humidity, sunshine duration, solar radiation, growing degree days, Shannon diversity index of precipitation evenness, abundant and well-distributed rainfall, and days after planting | LR, RR, Lasso, ENR, SVR, kNN, DT, RF, GBDT, MLP, XGBoost | XGBoost |

**Table 2.** *Cont.*

| Reference | Application | Main Input Datasets or Causative Factors | Algorithms Used | Algorithm with the Best Accuracy |
|---|---|---|---|---|
| [90] | Landslide susceptibility mapping | Elevation, slope, slope aspect, general curvature, plan curvature, profile curvature, surface roughness, TWI, SPI, slope length, NDVI, LULC, and distance from roads, rivers, faults and railways | Logistic regression, GBDT, VFI, SVM, DT, neural networks, Naïve bayes, RF, deep learning, majority-based voting ensemble | Majority-based voting ensemble |
| [91] | Spatial downscaling of precipitation data | Enhanced vegetation index, altitude, slope, aspect, latitude, and longitude | RF | — |
| [92] | Downscaling soil moisture | Soil moisture related indices derived from MODIS and a digital elevation model | GBDT | — |
| [93] | Estimating daily reference evapotranspiration | Maximum and minimum air temperature, relative humidity, wind speed at 2 m height, and solar radiation | RF, SVR, MLP, kNN, stacking, blending | Stacking |
| [94] | Downscaling precipitation | North American multi-model ensemble model outputs | Quantile mapping, wavelet SVM, wavelet RF | Wavelet SVM and wavelet RF |
| [95] | Downscaling LST | Landsat 8 and Sentinel-2A images, SRTM data, and daily minimum and maximum air temperatures | multi-factor geographically weighted machine learning (MFGWML), thermal image sharpening (TsHARP), high resolution thermal sharpener for cities | MFGWML |
| [96] | Growing stem volume estimation | Vegetation indices, spectral reflectance variables, backscattering coefficients, and texture features extracted from the Sentinel-1A and Sentinel-2A image datasets | Bagging (CART), Bagging (kNN), Bagging (SVR), Bagging (ANN), AdaBoost (CART), AdaBoost (kNN), AdaBoost (SVR), AdaBoost (ANN), secondary ensemble with an improved weighted average (IWA) | IWA |
| [28] | Forest biomass estimation | Leaf area index, canopy height, net primary production, and tree cover data, climatic data, and topographical data | SVR, MARS, MLP, RF, ERT, SGB, GBRT, CatBoost | CatBoost |
| [58] | Forest biomass estimation | Satellite-derived leaf area index, net primary production, forest canopy height, tree cover data, climate data, and topographical data | CatBoost, GBRT, MLP, MARS, SVR, stacking | Stacking |
| [97] | Mapping fine spatial resolution precipitation | MODIS NDVI, daily land surface temperature, and SRTM DRM data | RF, CART | RF |

**Table 3.** Most-used ensemble learning algorithms.

| Most-Used Ensemble Learning Algorithms | Number of Times Used |
| :---: | :---: |
| RF | 30 |
| SVR | 19 |
| stacking | 13 |
| MLP | 10 |
| KNN | 8 |
| XGBoost | 7 |
| GBRT | 4 |
| Adaboost | 4 |
| CART | 3 |
| CatBoost | 2 |
| ERT | 2 |
| SGB | 2 |

### 4.1. Yield Prediction

Machine learning algorithms have been widely used in crop yield prediction [98]. Van Klompenburg et al. [99] analyzed 50 machine learning papers and 30 deep learning papers concerning crop yield prediction. A neural network was the most frequently used machine learning technique in the machine learning papers; RF was used in 12 studies, and GBRT was used four times. Ruan et al. [89] included eleven statistical and machine learning techniques for predicting field-scale wheat yield and found that the two ensemble learning models RF and XGBoost were most accurate in prediction, with $R^2$ in the range 0.74–0.78 and RMSE in the range 0.78–0.85 t/ha. Cao et al. [66] adopted MLR, XGBoost, RF, and SVR algorithms and three datasets including MODIS EVI, climate data from the climatic research unit, and the subseasonal-to-seasonal (S2S) atmospheric prediction data, to estimate winter wheat yield at the grid level. The results showed that among the four models, XGBoost reached the highest skill with the S2S prediction as inputs, with an $R^2$ of 0.85 and RMSE of 0.78 t/ha. Ang et al. [61] employed the RF and AdaBoost algorithms to estimate oil palm yield prediction from multi-temporal remote sensing data. Results of the study revealed that the RF model (RMSE = 0.384; MSE = 0.148; MAE = 0.147) outperformed the AdaBoost model (RMSE = 0.410; MSE = 0.168; MAE = 0.176). Kamir et al. [82] used nine base learners and two ensemble (average ensemble and Bayesian fusion) methods to estimate wheat yields across the Australian wheat belt from climate data, and satellite image time series. The results showed that SVR with radial basis function merged as the single best learner with the R2 of 0.77 and RMSE of 0.55 t/ha at the pixel level, and the ensemble techniques did not show a significant advantage over the single best model.

Stacking models are also increasingly used in crop yield prediction. Feng et al. [74] predicted alfalfa yield from unmanned aerial vehicle (UAV)-based hyperspectral images using RF, SVR, k-nearest neighbors (kNN) and stacking ensemble algorithms. Comparison of the results indicated that stacking performed the best among all the base learners, with $R^2$ = 0.874. Fei et al. [75] combined four base learners (RF, SVM, Gaussian process (GP) and ridge regression (RR)) to predict grain yield across growth stages and found that stacking improved prediction accuracy in both full and limited irrigation scenarios with respective $R^2$ values of 0.625 and 0.628 at the mid grain filling stage. Li et al. [84] developed four base models (RF, SVM, GP, and RR) as well as stacking model, to predict winter wheat yields from UAV-based hyperspectral image data. They found that SVM outperformed the other learners and, compared with each base model, the stacking model was more accurate.

*4.2. Estimation of Forest Structure Parameters and Biomass*

Forest structure parameters (e.g., forest height) and forest biomass are critical indicators of carbon stocks and are increasingly important in fields related to the carbon cycle and climate change [100]. García-Gutiérrez et al. [76] compared a multiple linear regression (MLR) model, a neural network, SVR, kNN, and RF in estimating forest parameters in the province of Lugo (Galizia, Spain) from lidar data and found that MLR was outperformed by ML algorithms and that SVR with Gaussian kernels outperformed the other algorithms. Corte et al. [69] used SVR, ANN, RF, and XGBoost to estimate tree dendrometry parameters such as volume, height and diameter at breast height from UAV-lidar point clouds. Their results showed that all models were robust, with relative RMSE <29% for volume, <9% for height and <15% for diameter at breast height; SVR performed the best in terms of minimal error rates. SVR outperformed the ensemble algorithms in both studies and gave the most accurate predictions.

Various ensemble algorithms have been developed to estimate forest parameters. For example, Cartus et al. [67] used a two-stage model to derive forest canopy height and growing stock volume (GSV) in Chile from a combination of airborne laser scanner (ALS), PALSAR and Landsat data. They developed an RF model of canopy height and GSV using ALS-derived values that were validated with in situ measurements and then used a second RF model that used multitemporal PALSAR and Landsat data as predictor variables and ALS-based estimates as response variables. At three test sites, the retrieval of canopy height and GSV reached good accuracies with the $R^2$ of 0.70–0.85. Xu et al. [96] developed a two-stage ensemble approach to increase the accuracy of forest GSV estimation. They selected variables using a collinearity test and ran four base learners (CART, kNN, SVR, and ANN) and combined them first using bagging and then using AdaBoost to generate eight ensemble models. The eight ensemble models were then aggregated using a weighted average method in which the weights were determined by the validated relative RMSE values of the eight ensemble models. The experimental results showed that the combined ensemble approach significantly reduced the uncertainty of GSV mapping from the Sentinel-1A and Sentinel-2A data, with relative RMSE values in the range 18.89–21.34%.

Dube et al. [73] used stochastic gradient boosting (SGB) to estimate the stand volume of eucalyptus plantations in South Africa from multisource data and found that SGB accurately predicted stand volume, with $R^2 = 0.78$ and RMSE = 33.16 m$^3$/ha. These results were more accurate than the results given by RF or stepwise regression. Zhang et al. [28] comprehensively assessed the performance of eight algorithms (SVR, MARS, MLP, RF, ERT, SGB, GBRT, and CatBoost) in predicting forest biomass using several remote sensing datasets. Their results indicated that five ensemble algorithms (RF, ERT, SGB, GBRT and CatBoost) produced more accurate predictions than the other three individual algorithms and that CatBoost obtained slightly more accurate results than the other four ensemble algorithms, with $R^2 = 0.72$ and RMSE = 45.63 t/ha. In a subsequent study, they developed a stacking model to combine several accurate base learners to further increase the accuracy of biomass prediction, and the results indicated that the stacking ensemble increased prediction accuracy; in particular, it decreased the biases [58]. Ghosh et al. [77] used a stacked set of ensemble algorithms (RF, GBM, and XGBoost) to predict the aboveground biomass of Indian mangroves from Sentinel-1 and Sentinel-2 time series. The results indicated that stacking increased AGB prediction accuracy with RMSE = 72.864 t/ha and relative RMSE = 11.38%.

Du et al. [72] developed a CNN model using ALS data and Landsat imagery and found that CNN was more accurate than an extreme learning machine (ELM), a backpropagation neural network, a regression tree, RF, SVR, KNN, and other standard machine learning techniques. The stacking algorithm significantly increased prediction accuracy when compared with base models.

### 4.3. Natural Hazards

Natural hazards such as droughts, hurricanes, tornadoes, floods, and landslides can affect human life and property [101]. Accurate prediction or mapping the probabilities of natural disasters is therefore of great importance for human survival. Machine learning algorithms, particularly ensemble algorithms, have high prediction accuracy [85] and so have become increasingly used in identifying areas of long term drought [63], mapping landslide hazards [81,90], assessing the susceptibility of gullies to erosion [102,103] and mapping land susceptible to subsidence [78].

Arabameri et al. [62] combined three meta classifiers (Real AdaBoost, Random Subspace, and MultiBoosting) into a hybrid ensemble framework to predict the likelihood of flooding in the Gorganroud River Basin, Iran. Band et al. [65] used multiple ensemble algorithms (GBM, RF, parallel random forest (PRF), regularized random forest (RRF), and ERT) to quantify the likelihood of flash flooding in Markazi Province, Iran and found that ERT was the most accurate model with an area under curve (AUC) value of 0.82. Chapi et al. [68] combined bagging with logistic model trees (LMT) and developed a bagging–LMT ensemble model to map flood susceptibility. The results showed that in terms of accurately mapping flood susceptibility, the bagging–LMT model performed better than LMT, logistic regression, Bayesian logistic regression, and RF. Hakim et al. [78] compared logistic regression, MLP and two meta ensemble machine learning algorithms (AdaBoost and LogitBoost) in predicting likely subsidence based on a land subsidence inventory map generated from Sentinel-1 synthetic aperture radar (SAR) data. The results showed that AdaBoost gave the greatest prediction accuracy (81.1%), followed by MLP (80%), logistic regression (79.4%), and LogitBoost (79.1%). Kalantar et al. [81] investigated the suitability of flexible discriminant analysis (FDA), generalized logistic models (GLM), GBM, and RF for mapping landslide susceptibility. The test results showed that FDA was similar in prediction accuracy to GLM but was less accurate than GBM, which was in turn less accurate than RF. Rahman et al. [87] compared a Bayesian regularization back propagation neural network (BRBP), CART, an evidence belief function (EBF) and their various combinations in ensemble models to predict flood likelihood in Bangladesh. They found that the ensemble model that combined BRBP, CART, and EBF using weighted averaging was more accurate (AUC > 90%) than other models. In another study, Rahman et al. [86] found that stacking locally weighted linear regression (LWLR) and RF models increased the prediction accuracy of flood susceptibility maps in Bangladesh, with $R^2$ = 0.967–0.999, MAE = 0.022–0.117, RMSE = 0.029–0.148. Sachdeva et al. [90] used a majority voting ensemble technique to predict landslide susceptibility and found that the ensemble model that combined logistic regression, GBDT, and voting feature intervals produced predictions that were close in accuracy to the predictions of widely used machine learning algorithms such as decision trees, SVM, and RF.

### 4.4. Spatial Downscaling

Remote sensing data are often spatially downscaled to obtain fine resolution (FR) data from coarse resolution (CR) remote sensing data. The finer resolution data provide more spatial details and thus bridge the gap between what CR data provide and what regional applications require. Statistical downscaling methods have frequently been used in several domains to obtain FR parameters since they require less computation and running time, and are more accurate than other downscaling methods such as dynamic downscaling [104].

The statistical downscaling procedure for retrieving FR data from CR data is as follows: (1) develop models relating CR parameters and predictor variables or ancillary variables at a coarse resolution; (2) apply the CR models to the FR data, assuming that the relationships between target parameters and predictor variables remain unchanged at different spatial scales; and (3) obtain the target FR parameters from the models and FR predictor variables at a fine resolution. A variety of machine learning algorithms, especially the increasingly used ensemble learning algorithms, have been used as the relationships between target parameters and predictor variables are often nonlinear and complex.

RF has been widely used to upscale large-scale precipitation, land surface temperature (LST), and soil moisture remote sensing data. For example, Shi et al. [91] established nonparametric relationships between precipitation and six indicators (EVI, altitude, slope, aspect, latitude, and longitude) using RF and spatially downscaled annual precipitation for 2001–2012 from 0.25° pixels to 1 km pixels over the Tibetan Plateau. Zhao et al. [97] used RF to downscale the Tropical Rainfall Measuring Mission (TRMM) monthly precipitation product from 25 km resolution to produce monthly precipitation data for China at 1 km resolution. Hutengs and Vohland [80] developed a model relating LST to digital elevation data, land cover type, and surface reflectance in the red and near-infrared bands using RF; they downscaled the spatial resolution of LST from 1 km to 250 m, with the RMSEs from 1.41 to 1.92 K. When compared to the widely-adopted TsHARP sharpening method, downscaling accuracy using RF improved up to 19%. Karami et al. [83] created an RF-based regression tree to downscale the daily SMAP soil moisture product at 9 km resolution and created a 1 km soil moisture product using Sentinel-1 data, MODIS NDVI, land cover, and auxiliary topography and soil properties. Xu et al. [94] combined wavelet analysis with machine learning algorithms to create a wavelet support vector machine (WSVM) and a wavelet random forest (WRF) algorithm to downscale North American multimodel ensemble (NMME) precipitation forecasts. Their results showed improvement over quantile mapping, with an average decrease in RMSE of 18–40 mm (21–33%).

Several ensemble boosting algorithms have also been used for spatial downscaling. Wei et al. [92] created high resolution soil moisture maps covering the entire Tibetan Plateau using GBRT with SMAP soil moisture data and related variables derived from MODIS and DEM. Using GBRT, Asadollah et al. [64] yielded a significant improvement in downscaling global climate model predictions, compared to SVR that was previously found as the most suitable for downscaling climate in Iran. Xu et al. [95] developed a multifactor geographically weighted machine learning algorithm using Sentinel-2A data that combined the results from three base learners (XGBoost, MARS, and Bayesian ridge regression) and downscaled LST at 30 m resolution to 10 m resolution.

*4.5. Other Applications*

Other applications have used ensemble learning algorithms. In this section, we briefly mention some studies that recognized the importance of stacking.

Wu et al. [93] developed a two-layer stacking and blending ensemble method to predict daily reference evapotranspiration in which level 0 models included RF, SVR, multilayer perceptron neural network (MLP), and kNN. Both stacking and blending were significantly more accurate than the base models, and this approach is thus highly recommended for predicting reference evapotranspiration. Cho et al. [46] used a stacking model to predict daily maximum air temperature that consisted of multiple linear regression (MLR) and support vector regression (SVR), and RF was optimized by SVR. The stacking ensemble method produced more accurate predictions than cokriging, three distinct data-driven methods, and a simple average ensemble model. Divina et al. [71] showed that stacking was a suitable approach for short term electricity consumption forecasting. Healey et al. [79] showed that stacking increased the accuracy of detection of forest change. They investigated a stacking model using both parametric and RF-based image fusion rules as the meta learner to combine several forest disturbance detection algorithms and found that stacking using an RF model to build the meta learner reduced the rates of errors of omission and errors of commission by 50% in some instances when compared to individual change detection methods and by 25% when compared with stacking using a logistic regression model as the meta learner.

## 5. Discussion and Future Directions

*5.1. Combining Feature Selection with Ensemble Learning*

Many studies have shown that using selected important features or variables instead of using all extracted variables can result in more accurate and robust predictions [105–107].

Ensemble algorithms are essentially black box models that carry the risk of overfitting, and the underlying physical mechanisms can be obscure [108]. It is therefore critical to identify important variables by selecting features before training a model; filter, wrapper, and embedded algorithms are just a few of the feature selection techniques that have been proposed [106,109].

Some studies have highlighted the importance of combining feature selection with ensemble learning algorithms in practical applications. For example, Luo et al. [43] found that recursive feature elimination (RFE) for feature selection in combination with a Cat-Boost model produced the most accurate prediction of forest biomass for all forests in Jilin province, China with RMSE = 25.77 t/ha. There have also been studies that used a variance inflation factor (VIF) to quantify multicollinearity between independent variables, and only variables with VIF values < 10 were finally included for modeling [62,110]. Of the three types of ensemble learning methods, tree-based bagging and boosting algorithms have identified important variables mainly by permutation importance [111,112]. Some other indicators, such as mean decrease in accuracy and mean decrease in impurity, have also been used in association with tree-based algorithms (e.g., RF) [113]. In contrast, stacking ensemble algorithms appear to have difficulty in selecting important variables due to working with a set of models rather than an individual model, and it can be difficult to interpret the ensemble results [88,114]. Feature selection should thus be implemented with care when stacking, but this aspect of the technique has been little explored in published studies.

### 5.2. Other Ensemble Learning Algorithms

In this study, we primarily reviewed bagging, boosting and stacking ensemble algorithms. Other ensemble learning algorithms have been developed in addition to these well-known algorithms, such as dynamic ensemble learning [115] and Bayesian additive regression trees [116]. The dynamic ensemble method, unlike static ensemble algorithms which combine fixed base learners, selects the single best learner or combines a subset of learners from the pool using a just-in-time condition that depends on the particular input pattern from which a prediction is to be made when making a prediction [117–119].

Blending is another ensemble technique that is derived from stacking. Blending differs from stacking in that it does not use k-fold cross-validation to generate training data for the meta learner but instead uses a one-holdout set. This technique results in only a small portion of the training dataset being used to generate predictions to be used as inputs to a meta model [93,120].

### 5.3. Deep Learning Algorithms

Deep learning algorithms are used in many fields, including agriculture and remote sensing. The base learners in current ensemble models are mostly statistical and conventional ML methods, and the possibility of combining deep learning models in several ways is worthy of investigation. Deep learning algorithms have been used as base learners in some studies with the intention of increasing prediction accuracy. For example, de Oliveira e Lucas et al. [70] used three CNNs to predict reference evapotranspiration time series and developed ensemble models consisting of the three CNNs. The CNN ensembles produced predictions with high accuracy and low variance. Lv et al. [47] developed a heterogeneous ensemble learning approach that combined three deep learning models (a deep belief network (DBN), a CNN and a deep residual network (ResNet)) to map landslide susceptibility in the Three Gorges reservoir area in China.

Deep learning algorithms well capture nonlinear relationships between target and sensor signals, so an ensemble of various deep learning algorithms should produce more accurate predictions than a single algorithm. The three principal ensemble methods we described in this study, particularly stacking, provide a framework for leveraging different algorithms. Future studies will investigate optimal combinations of deep learning algorithms in various applications.

## 6. Conclusions

In this paper, we reviewed bagging, boosting, and stacking ensemble learning algorithms and their typical applications in the use of remote sensing data. RF was the most often adopted algorithm in several fields that used remote sensing data. In contrast, the other ensemble algorithms were often not considered for specific applications. Despite recent progress in increasing the prediction accuracy of ensemble algorithms, there are still some gaps in our knowledge, such as how to effectively combine feature selection with ensemble algorithms and how to incorporate deep learning algorithms in an ensemble to increase prediction accuracy. The understanding of ensemble algorithms deserves to be the main focus of future study and will enable us to incorporate more advanced and diverse algorithms in practical applications.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Nomenclature

| | |
|---|---|
| AdaBoost | adaptive boosting |
| ANN | artificial neural networks |
| ARMA | autoregressive–moving-average |
| ARIMA | autoregressive integrated moving average |
| BPNN | back propagation neural network |
| BRT | boosted regression tree |
| CART | classification and regression trees |
| CatBoost | categorical boosting |
| CNN | convolutional neural networks |
| DBN | deep belief network |
| DL | deep learning |
| DT | decision tree |
| EBF | evidence belief function |
| ELM | extreme learning machine |
| ENR | elastic net regression |
| ERT | extremely randomized trees |
| FDA | flexible discriminant analysis |
| GBDT | gradient boosting decision tree |
| GBRT | gradient boosting regression tree |
| GLM | generalized logistic models |
| GP | Gaussian process |
| kNN | k-nearest neighbor |
| LMT | logistic model tree |
| LR | linear regression |
| LSSVM | least square support vector machine |
| LWLR | locally weighted linear regression |
| MADT | multiclass alternating decision trees |
| MARS | multivariate adaptive regression splines |
| MLP | multilayer perceptron |
| MLR | multiple linear regression |
| PRF | parallel random forest |
| REPT | reduced error pruning tree |
| ResNet | residual neural network |
| RF | random forest |
| RR | ridge regression |
| RRF | regularized random forest |
| RT | regression tree |

| SGB | stochastic gradient boosting |
| SVM | support vector machine |
| SVR | support vector regression |
| VFI | voting feature interval |
| XGBoost | extreme gradient boosting |

## References

1. Navalgund, R.R.; Jayaraman, V.; Roy, P.S. Remote sensing applications: An overview. *Curr. Sci.* **2007**, *93*, 1747–1766.
2. Weiss, M.; Jacob, F.; Duveiller, G. Remote sensing for agricultural applications: A meta-review. *Remote Sens. Environ.* **2020**, *236*, 111402. [CrossRef]
3. Roy, P.S.; Behera, M.D.; Srivastav, S.K. Satellite Remote Sensing: Sensors, Applications and Techniques. *Proc. Natl. Acad. Sci. India Sect. A Phys. Sci.* **2017**, *87*, 465–472. [CrossRef]
4. Myneni, R.B.; Maggion, S.; Iaquinta, J.; Privette, J.L.; Gobron, N.; Pinty, B.; Kimes, D.S.; Verstraete, M.M.; Williams, D.L. Optical remote sensing of vegetation: Modeling, caveats, and algorithms. *Remote Sens. Environ.* **1995**, *51*, 169–188. [CrossRef]
5. Bonan, G.B.; Levis, S.; Sitch, S.; Vertenstein, M.; Oleson, K.W. A dynamic global vegetation model for use with climate models: Concepts and description of simulated vegetation dynamics. *Glob. Change Biol.* **2003**, *9*, 1543–1566. [CrossRef]
6. Jacquemoud, S.; Verhoef, W.; Baret, F.; Bacour, C.; Zarco-Tejada, P.J.; Asner, G.P.; François, C.; Ustin, S.L. PROSPECT+SAIL models: A review of use for vegetation characterization. *Remote Sens. Environ.* **2009**, *113*, S56–S66. [CrossRef]
7. Pan, M.; Sahoo, A.K.; Wood, E.F. Improving soil moisture retrievals from a physically-based radiative transfer model. *Remote Sens. Environ.* **2014**, *140*, 130–140. [CrossRef]
8. Schulze, E.-D.; Beck, E.; Buchmann, N.; Clemens, S.; Müller-Hohenstein, K.; Scherer-Lorenzen, M. Dynamic Global Vegetation Models. In *Plant Ecology*; Springer: Berlin, Heidelberg, Germany, 2019; pp. 843–863.
9. García-Haro, F.J.; Gilabert, M.A.; Meliá, J. A radiosity model for heterogeneous canopies in remote sensing. *J. Geophys. Res. Atmos.* **1999**, *104*, 12159–12175. [CrossRef]
10. Lary, D.; Remer, L.; MacNeill, D.; Roscoe, B.; Paradise, S. Machine learning and bias correction of MODIS aerosol optical depth. *IEEE Geosci. Remote Sens. Lett.* **2009**, *6*, 694–698. [CrossRef]
11. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [CrossRef]
12. Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259. [CrossRef]
13. Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
14. Sheykhmousa, M.; Mahdianpari, M.; Ghanbari, H.; Mohammadimanesh, F.; Ghamisi, P.; Homayouni, S. Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 6308–6325. [CrossRef]
15. Wolpert, D.H.; Macready, W.G. No free lunch theorems for optimization. *IEEE Trans. Evol. Comput.* **1997**, *1*, 67–82. [CrossRef]
16. Lasisi, A.; Attoh-Okine, N. Machine Learning Ensembles and Rail Defects Prediction: Multilayer Stacking Methodology. *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part A Civ. Eng.* **2019**, *5*, 04019016. [CrossRef]
17. Galar, M.; Fernandez, A.; Barrenechea, E.; Bustince, H.; Herrera, F. A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches. *IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.)* **2012**, *42*, 463–484. [CrossRef]
18. Zhou, Z.-H. Ensemble Learning. In *Encyclopedia of Biometrics*; Li, S.Z., Jain, A., Eds.; Springer: Boston, MA, USA, 2009; pp. 270–273.
19. Mendes-Moreira, J.; Soares, C.; Jorge, A.M.; Sousa, J.F.D. Ensemble approaches for regression: A survey. *ACM Comput. Surv.* **2012**, *45*, 10. [CrossRef]
20. Ting, K.M.; Witten, I.h. Issues in stacked generalization. *J. Artif. Intell. Res.* **1999**, *10*, 271–289. [CrossRef]
21. Sagi, O.; Rokach, L. Ensemble learning: A survey. *WIREs Data Min. Knowl. Discov.* **2018**, *8*, e1249. [CrossRef]
22. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [CrossRef]
23. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
24. Kuter, S. Completing the machine learning saga in fractional snow cover estimation from MODIS Terra reflectance data: Random forests versus support vector regression. *Remote Sens. Environ.* **2021**, *255*, 112294. [CrossRef]
25. Chen, L.; Wang, Y.; Ren, C.; Zhang, B.; Wang, Z. Assessment of multi-wavelength SAR and multispectral instrument data for forest aboveground biomass mapping using random forest kriging. *For. Ecol. Manag.* **2019**, *447*, 12–25. [CrossRef]
26. Liu, Y.; Cao, G.; Zhao, N.; Mulligan, K.; Ye, X. Improve ground-level PM2.5 concentration mapping using a random forests-based geostatistical approach. *Environ. Pollut.* **2018**, *235*, 272–282. [CrossRef] [PubMed]
27. Geurts, P.; Louppe, G. Learning to rank with extremely randomized trees. In Proceedings of the Learning to Rank Challenge, PMLR; 2011; Volume 14, pp. 49–61.
28. Zhang, Y.; Ma, J.; Liang, S.; Li, X.; Li, M. An Evaluation of Eight Machine Learning Regression Algorithms for Forest Aboveground Biomass Estimation from Multiple Satellite Data Products. *Remote Sens.* **2020**, *12*, 4015. [CrossRef]

29. Wei, J.; Li, Z.; Cribb, M.; Huang, W.; Xue, W.; Sun, L.; Guo, J.; Peng, Y.; Li, J.; Lyapustin, A. Improved 1 km resolution PM 2.5 estimates across China using enhanced space–time extremely randomized trees. *Atmos. Chem. Phys.* **2020**, *20*, 3273–3289. [CrossRef]

30. Azpiroz, I.; Oses, N.; Quartulli, M.; Olaizola, I.G.; Guidotti, D.; Marchi, S. Comparison of climate reanalysis and remote-sensing data for predicting olive phenology through machine-learning methods. *Remote Sens.* **2021**, *13*, 1224. [CrossRef]

31. Cao, Y.; Li, M.; Zhang, Y. Estimating the Clear-Sky Longwave Downward Radiation in the Arctic from FengYun-3D MERSI-2 Data. *Remote Sens.* **2022**, *14*, 606. [CrossRef]

32. Galelli, S.; Castelletti, A. Assessing the predictive capability of randomized tree-based ensembles in streamflow modelling. *Hydrol. Earth Syst. Sci.* **2013**, *17*, 2669–2684. [CrossRef]

33. Shang, K.; Yao, Y.; Li, Y.; Yang, J.; Jia, K.; Zhang, X.; Chen, X.; Bei, X.; Guo, X. Fusion of Five Satellite-Derived Products Using Extremely Randomized Trees to Estimate Terrestrial Latent Heat Flux over Europe. *Remote Sens.* **2020**, *12*, 687. [CrossRef]

34. Elith, J.; Leathwick, J.R.; Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **2008**, *77*, 802–813. [CrossRef]

35. Chen, T.Q.; Guestrin, C.; Assoc Comp, M. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.

36. Zhao, Y.; Yin, X.; Fu, Y.; Yue, T. A comparative mapping of plant species diversity using ensemble learning algorithms combined with high accuracy surface modeling. *Environ. Sci. Pollut. Res.* **2022**, *29*, 17878–17891. [CrossRef]

37. Joharestani, M.Z.; Cao, C.X.; Ni, X.L.; Bashir, B.; Talebiesfandarani, S. PM2.5 Prediction Based on Random Forest, XGBoost, and Deep Learning Using Multisource Remote Sensing Data. *Atmosphere* **2019**, *10*, 373. [CrossRef]

38. Li, Y.; Li, C.; Li, M.; Liu, Z. Influence of Variable Selection and Forest Type on Forest Aboveground Biomass Estimation Using Machine Learning Algorithms. *Forests* **2019**, *10*, 1073. [CrossRef]

39. Ma, M.; Zhao, G.; He, B.; Li, Q.; Dong, H.; Wang, S.; Wang, Z. XGBoost-based method for flash flood risk assessment. *J. Hydrol.* **2021**, *598*, 126382. [CrossRef]

40. Ke, G.L.; Meng, Q.; Finley, T.; Wang, T.F.; Chen, W.; Ma, W.D.; Ye, Q.W.; Liu, T.Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In Proceedings of the 31st Annual Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 4–9 December 2017.

41. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased boosting with categorical features. In Proceedings of the 32nd International Conference on Neural Information Processing System, Montréal, Canada, 3–8 December 2018.

42. Hancock, J.T.; Khoshgoftaar, T.M. CatBoost for big data: An interdisciplinary review. *J. Big Data* **2020**, *7*, 94. [CrossRef]

43. Luo, M.; Wang, Y.; Xie, Y.; Zhou, L.; Qiao, J.; Qiu, S.; Sun, Y. Combination of Feature Selection and CatBoost for Prediction: The First Application to the Estimation of Aboveground Biomass. *Forests* **2021**, *12*, 216. [CrossRef]

44. Huang, G.; Wu, L.; Ma, X.; Zhang, W.; Fan, J.; Yu, X.; Zeng, W.; Zhou, H. Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. *J. Hydrol.* **2019**, *574*, 1029–1041. [CrossRef]

45. Wolpert, D.H. Stacked generalization. *Neural Netw.* **1992**, *5*, 241–259. [CrossRef]

46. Cho, D.; Yoo, C.; Im, J.; Lee, Y.; Lee, J. Improvement of spatial interpolation accuracy of daily maximum air temperature in urban areas using a stacking ensemble technique. *GISci. Remote Sens.* **2020**, *57*, 633–649. [CrossRef]

47. Lv, L.; Chen, T.; Dou, J.; Plaza, A. A hybrid ensemble-based deep-learning framework for landslide susceptibility mapping. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *108*, 102713. [CrossRef]

48. Naimi, A.I.; Balzer, L.B. Stacked generalization: An introduction to super learning. *Eur. J. Epidemiol.* **2018**, *33*, 459–464. [CrossRef] [PubMed]

49. Nath, A.; Sahu, G.K. Exploiting ensemble learning to improve prediction of phospholipidosis inducing potential. *J. Theor. Biol.* **2019**, *479*, 37–47. [CrossRef] [PubMed]

50. Dai, Q.; Ye, R.; Liu, Z. Considering diversity and accuracy simultaneously for ensemble pruning. *Appl. Soft Comput.* **2017**, *58*, 75–91. [CrossRef]

51. Kuncheva, L.I.; Whitaker, C.J. Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy. *Mach. Learn.* **2003**, *51*, 181–207. [CrossRef]

52. Rooney, N. A weighted combiner of stacking based methods. *Int. J. Artif. Intell. Tools* **2012**, *21*, 1250040. [CrossRef]

53. Zhang, H.; Cao, L. A spectral clustering based ensemble pruning approach. *Neurocomputing* **2014**, *139*, 289–297. [CrossRef]

54. Tang, E.K.; Suganthan, P.N.; Yao, X. An analysis of diversity measures. *Mach. Learn.* **2006**, *65*, 247–271. [CrossRef]

55. Ma, Z.; Dai, Q. Selected an Stacking ELMs for Time Series Prediction. *Neural Process. Lett.* **2016**, *44*, 831–856. [CrossRef]

56. Breiman, L. Stacked Regressions. *Mach. Learn.* **1996**, *24*, 49–64. [CrossRef]

57. Fei, X.; Zhang, Q.; Ling, Q. Vehicle Exhaust Concentration Estimation Based on an Improved Stacking Model. *IEEE Access* **2019**, *7*, 179454–179463. [CrossRef]

58. Zhang, Y.; Ma, J.; Liang, S.; Li, X.; Liu, J. A stacking ensemble algorithm for improving the biases of forest aboveground biomass estimations from multiple remotely sensed datasets. *GISci. Remote Sens.* **2022**, *59*, 234–249. [CrossRef]

59. Zheng, H.; Cheng, Y.; Li, H. Investigation of model ensemble for fine-grained air quality prediction. *China Commun.* **2020**, *17*, 207–223. [CrossRef]

60. Tyralis, H.; Papacharalampous, G.; Burnetas, A.; Langousis, A. Hydrological post-processing using stacked generalization of quantile regression algorithms: Large-scale application over CONUS. *J. Hydrol.* **2019**, *577*, 123957. [CrossRef]

61. Ang, Y.; Shafri, H.Z.M.; Lee, Y.P.; Abidin, H.; Bakar, S.A.; Hashim, S.J.; Che'Ya, N.N.; Hassan, M.R.; Lim, H.S.; Abdullah, R. A novel ensemble machine learning and time series approach for oil palm yield prediction using Landsat time series imagery based on NDVI. *Geocarto Int.* **2022**, 1–32. [CrossRef]

62. Arabameri, A.; Saha, S.; Mukherjee, K.; Blaschke, T.; Chen, W.; Ngo, P.T.T.; Band, S.S. Modeling Spatial Flood using Novel Ensemble Artificial Intelligence Approaches in Northern Iran. *Remote Sens.* **2020**, *12*, 3423. [CrossRef]

63. Arabameri, A.; Chandra Pal, S.; Santosh, M.; Chakrabortty, R.; Roy, P.; Moayedi, H. Drought risk assessment: Integrating meteorological, hydrological, agricultural and socio-economic factors using ensemble models and geospatial techniques. *Geocarto Int.* **2021**, 1–29. [CrossRef]

64. Asadollah, S.B.H.S.; Sharafati, A.; Shahid, S. Application of ensemble machine learning model in downscaling and projecting climate variables over different climate regions in Iran. *Environ. Sci. Pollut. Res.* **2022**, *29*, 17260–17279. [CrossRef]

65. Band, S.S.; Janizadeh, S.; Chandra Pal, S.; Saha, A.; Chakrabortty, R.; Melesse, A.M.; Mosavi, A. Flash Flood Susceptibility Modeling Using New Approaches of Hybrid and Ensemble Tree-Based Machine Learning Algorithms. *Remote Sens.* **2020**, *12*, 3568. [CrossRef]

66. Cao, J.; Wang, H.; Li, J.; Tian, Q.; Niyogi, D. Improving the Forecasting of Winter Wheat Yields in Northern China with Machine Learning–Dynamical Hybrid Subseasonal-to-Seasonal Ensemble Prediction. *Remote Sens.* **2022**, *14*, 1707. [CrossRef]

67. Cartus, O.; Kellndorfer, J.; Rombach, M.; Walker, W. Mapping Canopy Height and Growing Stock Volume Using Airborne Lidar, ALOS PALSAR and Landsat ETM+. *Remote Sens.* **2012**, *4*, 3320–3345. [CrossRef]

68. Chapi, K.; Singh, V.P.; Shirzadi, A.; Shahabi, H.; Bui, D.T.; Pham, B.T.; Khosravi, K. A novel hybrid artificial intelligence approach for flood susceptibility assessment. *Environ. Model. Softw.* **2017**, *95*, 229–245. [CrossRef]

69. Corte, A.P.D.; Souza, D.V.; Rex, F.E.; Sanquetta, C.R.; Mohan, M.; Silva, C.A.; Zambrano, A.M.A.; Prata, G.; Alves de Almeida, D.R.; Trautenmüller, J.W.; et al. Forest inventory with high-density UAV-Lidar: Machine learning approaches for predicting individual tree attributes. *Comput. Electron. Agric.* **2020**, *179*, 105815. [CrossRef]

70. de Oliveira e Lucas, P.; Alves, M.A.; de Lima e Silva, P.C.; Guimarães, F.G. Reference evapotranspiration time series forecasting with ensemble of convolutional neural networks. *Comput. Electron. Agric.* **2020**, *177*, 105700. [CrossRef]

71. Divina, F.; Gilson, A.; Goméz-Vela, F.; García Torres, M.; Torres, J.F. Stacking Ensemble Learning for Short-Term Electricity Consumption Forecasting. *Energies* **2018**, *11*, 949. [CrossRef]

72. Du, C.; Fan, W.; Ma, Y.; Jin, H.-I.; Zhen, Z. The Effect of Synergistic Approaches of Features and Ensemble Learning Algorithms on Aboveground Biomass Estimation of Natural Secondary Forests Based on ALS and Landsat 8. *Sensors* **2021**, *21*, 5974. [CrossRef] [PubMed]

73. Dube, T.; Mutanga, O.; Abdel-Rahman, E.M.; Ismail, R.; Slotow, R. Predicting Eucalyptus spp. stand volume in Zululand, South Africa: An analysis using a stochastic gradient boosting regression ensemble with multi-source data sets. *Int. J. Remote Sens.* **2015**, *36*, 3751–3772. [CrossRef]

74. Feng, L.; Zhang, Z.; Ma, Y.; Du, Q.; Williams, P.; Drewry, J.; Luck, B. Alfalfa Yield Prediction Using UAV-Based Hyperspectral Imagery and Ensemble Learning. *Remote Sens.* **2020**, *12*, 2028. [CrossRef]

75. Fei, S.; Hassan, M.A.; He, Z.; Chen, Z.; Shu, M.; Wang, J.; Li, C.; Xiao, Y. Assessment of Ensemble Learning to Predict Wheat Grain Yield Based on UAV-Multispectral Reflectance. *Remote Sens.* **2021**, *13*, 2338. [CrossRef]

76. García-Gutiérrez, J.; Martínez-Álvarez, F.; Troncoso, A.; Riquelme, J.C. A comparison of machine learning regression techniques for LiDAR-derived estimation of forest variables. *Neurocomputing* **2015**, *167*, 24–31. [CrossRef]

77. Ghosh, S.M.; Behera, M.D.; Jagadish, B.; Das, A.K.; Mishra, D.R. A novel approach for estimation of aboveground biomass of a carbon-rich mangrove site in India. *J. Environ. Manag.* **2021**, *292*, 112816. [CrossRef] [PubMed]

78. Hakim, W.L.; Achmad, A.R.; Lee, C.-W. Land Subsidence Susceptibility Mapping in Jakarta Using Functional and Meta-Ensemble Machine Learning Algorithm Based on Time-Series InSAR Data. *Remote Sens.* **2020**, *12*, 3627. [CrossRef]

79. Healey, S.P.; Cohen, W.B.; Yang, Z.; Kenneth Brewer, C.; Brooks, E.B.; Gorelick, N.; Hernandez, A.J.; Huang, C.; Joseph Hughes, M.; Kennedy, R.E.; et al. Mapping forest change using stacked generalization: An ensemble approach. *Remote Sens. Environ.* **2018**, *204*, 717–728. [CrossRef]

80. Hutengs, C.; Vohland, M. Downscaling land surface temperatures at regional scales with random forest regression. *Remote Sens. Environ.* **2016**, *178*, 127–141. [CrossRef]

81. Kalantar, B.; Ueda, N.; Saeidi, V.; Ahmadi, K.; Halin, A.A.; Shabani, F. Landslide Susceptibility Mapping: Machine and Ensemble Learning Based on Remote Sensing Big Data. *Remote Sens.* **2020**, *12*, 1737. [CrossRef]

82. Kamir, E.; Waldner, F.; Hochman, Z. Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods. *ISPRS J. Photogramm. Remote Sens.* **2020**, *160*, 124–135. [CrossRef]

83. Karami, A.; Moradi, H.R.; Mousivand, A.; van Dijk, A.I.J.M.; Renzullo, L. Using ensemble learning to take advantage of high-resolution radar backscatter in conjunction with surface features to disaggregate SMAP soil moisture product. *Int. J. Remote Sens.* **2022**, *43*, 894–914. [CrossRef]

84. Li, Z.; Chen, Z.; Cheng, Q.; Duan, F.; Sui, R.; Huang, X.; Xu, H. UAV-Based Hyperspectral and Ensemble Machine Learning for Predicting Yield in Winter Wheat. *Agronomy* **2022**, *12*, 202. [CrossRef]

85. Pham, B.T.; Tien Bui, D.; Dholakia, M.; Prakash, I.; Pham, H.V. A comparative study of least square support vector machines and multiclass alternating decision trees for spatial prediction of rainfall-induced landslides in a tropical cyclones area. *Geotech. Geol. Eng.* **2016**, *34*, 1807–1824. [CrossRef]

86. Rahman, M.; Chen, N.; Elbeltagi, A.; Islam, M.M.; Alam, M.; Pourghasemi, H.R.; Tao, W.; Zhang, J.; Shufeng, T.; Faiz, H.; et al. Application of stacking hybrid machine learning algorithms in delineating multi-type flooding in Bangladesh. *J. Environ. Manag.* **2021**, *295*, 113086. [CrossRef]

87. Rahman, M.; Ningsheng, C.; Mahmud, G.I.; Islam, M.M.; Pourghasemi, H.R.; Ahmad, H.; Habumugisha, J.M.; Washakh, R.M.A.; Alam, M.; Liu, E.; et al. Flooding and its relationship with land cover change, population growth, and road density. *Geosci. Front.* **2021**, *12*, 101224. [CrossRef]

88. Ribeiro, M.H.D.M.; dos Santos Coelho, L. Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. *Appl. Soft Comput.* **2020**, *86*, 105837. [CrossRef]

89. Ruan, G.; Li, X.; Yuan, F.; Cammarano, D.; Ata-Ui-Karim, S.T.; Liu, X.; Tian, Y.; Zhu, Y.; Cao, W.; Cao, Q. Improving wheat yield prediction integrating proximal sensing and weather data with machine learning. *Comput. Electron. Agric.* **2022**, *195*, 106852. [CrossRef]

90. Sachdeva, S.; Bhatia, T.; Verma, A.K. A novel voting ensemble model for spatial prediction of landslides using GIS. *Int. J. Remote Sens.* **2020**, *41*, 929–952. [CrossRef]

91. Shi, Y.; Song, L. Spatial Downscaling of Monthly TRMM Precipitation Based on EVI and Other Geospatial Variables Over the Tibetan Plateau From 2001 to 2012. *Mt. Res. Dev.* **2015**, *35*, 180–194. [CrossRef]

92. Wei, Z.; Meng, Y.; Zhang, W.; Peng, J.; Meng, L. Downscaling SMAP soil moisture estimation with gradient boosting decision tree regression over the Tibetan Plateau. *Remote Sens. Environ.* **2019**, *225*, 30–44. [CrossRef]

93. Wu, T.; Zhang, W.; Jiao, X.; Guo, W.; Alhaj Hamoud, Y. Evaluation of stacking and blending ensemble learning methods for estimating daily reference evapotranspiration. *Comput. Electron. Agric.* **2021**, *184*, 106039. [CrossRef]

94. Xu, L.; Chen, N.; Zhang, X.; Chen, Z.; Hu, C.; Wang, C. Improving the North American multi-model ensemble (NMME) precipitation forecasts at local areas using wavelet and machine learning. *Clim. Dyn.* **2019**, *53*, 601–615. [CrossRef]

95. Xu, S.; Zhao, Q.; Yin, K.; He, G.; Zhang, Z.; Wang, G.; Wen, M.; Zhang, N. Spatial Downscaling of Land Surface Temperature Based on a Multi-Factor Geographically Weighted Machine Learning Model. *Remote Sens.* **2021**, *13*, 1186. [CrossRef]

96. Xu, X.; Lin, H.; Liu, Z.; Ye, Z.; Li, X.; Long, J. A Combined Strategy of Improved Variable Selection and Ensemble Algorithm to Map the Growing Stem Volume of Planted Coniferous Forest. *Remote Sens.* **2021**, *13*, 4631. [CrossRef]

97. Zhao, X.; Jing, W.; Zhang, P. Mapping Fine Spatial Resolution Precipitation from TRMM Precipitation Datasets Using an Ensemble Learning Method and MODIS Optical Products in China. *Sustainability* **2017**, *9*, 1912. [CrossRef]

98. Elavarasan, D.; Vincent, D.R.; Sharma, V.; Zomaya, A.Y.; Srinivasan, K. Forecasting yield by integrating agrarian factors and machine learning models: A survey. *Comput. Electron. Agric.* **2018**, *155*, 257–282. [CrossRef]

99. van Klompenburg, T.; Kassahun, A.; Catal, C. Crop yield prediction using machine learning: A systematic literature review. *Comput. Electron. Agric.* **2020**, *177*, 105709. [CrossRef]

100. Houghton, R.A.; Hall, F.; Goetz, S.J. Importance of biomass in the global carbon cycle. *J. Geophys. Res. Biogeosci.* **2009**, *114*, G00E03. [CrossRef]

101. Fontaine, M.M.; Steinemann, A.C. Assessing Vulnerability to Natural Hazards: Impact-Based Method and Application to Drought in Washington State. *Nat. Hazards Rev.* **2009**, *10*, 11–18. [CrossRef]

102. Arabameri, A.; Yamani, M.; Pradhan, B.; Melesse, A.; Shirani, K.; Bui, D.T. Novel ensembles of COPRAS multi-criteria decision-making with logistic regression, boosted regression tree, and random forest for spatial prediction of gully erosion susceptibility. *Sci. Total Environ.* **2019**, *688*, 903–916. [CrossRef]

103. Chowdhuri, I.; Pal, S.C.; Arabameri, A.; Saha, A.; Chakrabortty, R.; Blaschke, T.; Pradhan, B.; Band, S. Implementation of artificial intelligence based ensemble models for gully erosion susceptibility assessment. *Remote Sens.* **2020**, *12*, 3620. [CrossRef]

104. Wilby, R.L.; Wigley, T.M.L. Downscaling general circulation model output: A review of methods and limitations. *Prog. Phys. Geogr. Earth Environ.* **1997**, *21*, 530–548. [CrossRef]

105. Bolón-Canedo, V.; Sánchez-Maroño, N.; Alonso-Betanzos, A.; Benítez, J.M.; Herrera, F. A review of microarray datasets and applied feature selection methods. *Inf. Sci.* **2014**, *282*, 111–135. [CrossRef]

106. Cai, J.; Luo, J.; Wang, S.; Yang, S. Feature selection in machine learning: A new perspective. *Neurocomputing* **2018**, *300*, 70–79. [CrossRef]

107. Guyon, I.; Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182.

108. Li, C.; Li, Y.; Li, M. Improving Forest Aboveground Biomass (AGB) Estimation by Incorporating Crown Density and Using Landsat 8 OLI Images of a Subtropical Forest in Western Hunan in Central China. *Forests* **2019**, *10*, 104. [CrossRef]

109. Chandrashekar, G.; Sahin, F. A survey on feature selection methods. *Comput. Electr. Eng.* **2014**, *40*, 16–28. [CrossRef]

110. Kavzoglu, T.; Sahin, E.K.; Colkesen, I. Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. *Landslides* **2014**, *11*, 425–439. [CrossRef]

111. Auret, L.; Aldrich, C. Empirical comparison of tree ensemble variable importance measures. *Chemom. Intell. Lab. Syst.* **2011**, *105*, 157–170. [CrossRef]

112. Altmann, A.; Toloşi, L.; Sander, O.; Lengauer, T. Permutation importance: A corrected feature importance measure. *Bioinformatics* **2010**, *26*, 1340–1347. [CrossRef]

113. Gregorutti, B.; Michel, B.; Saint-Pierre, P. Correlation and variable importance in random forests. *Stat. Comput.* **2017**, *27*, 659–678. [CrossRef]

114. Qureshi, A.S.; Khan, A.; Zameer, A.; Usman, A. Wind power prediction using deep neural network based meta regression and transfer learning. *Appl. Soft Comput.* **2017**, *58*, 742–755. [CrossRef]

115. Alam, K.M.R.; Siddique, N.; Adeli, H. A dynamic ensemble learning algorithm for neural networks. *Neural Comput. Appl.* **2020**, *32*, 8675–8690. [CrossRef]

116. Chipman, H.A.; George, E.I.; McCulloch, R.E. BART: Bayesian additive regression trees. *Ann. Appl. Stat.* **2010**, *4*, 266–298. [CrossRef]

117. Conroy, B.; Eshelman, L.; Potes, C.; Xu-Wilson, M. A dynamic ensemble approach to robust classification in the presence of missing data. *Mach. Learn.* **2016**, *102*, 443–463. [CrossRef]

118. Rooney, N.; Patterson, D. A weighted combination of stacking and dynamic integration. *Pattern Recognit.* **2007**, *40*, 1385–1388. [CrossRef]

119. Ko, A.H.R.; Sabourin, R.; Britto, J.A.S. From dynamic classifier selection to dynamic ensemble selection. *Pattern Recognit.* **2008**, *41*, 1718–1731. [CrossRef]

120. Soman, G.; Vivek, M.V.; Judy, M.V.; Papageorgiou, E.; Gerogiannis, V.C. Precision-Based Weighted Blending Distributed Ensemble Model for Emotion Classification. *Algorithms* **2022**, *15*, 55. [CrossRef]