

Article

Impact of Environmental Conditions on Renewable Energy Prediction: An Investigation Through Tree-Based Community Learning

Ferdî Doğan ^{1,*} , Saadin Oyucu ¹ , Derya Betul Unsal ² , Ahmet Aksöz ³  and Majid Vafaeipour ⁴ 

¹ Department of Computer Engineering, Adiyaman University, Adiyaman 02040, Türkiye; saadinoyucu@adiyaman.edu.tr

² Department of Electrical and Electronics Engineering, Cumhuriyet University, Sivas 58140, Türkiye; dbunsal@cumhuriyet.edu.tr

³ MOBILERS Research Team, Sivas Cumhuriyet University, Sivas 58580, Türkiye; aaksoz@cumhuriyet.edu.tr

⁴ Electric Vehicle and Energy Research Group (EVERGI), Mobility, Logistics and Automotive Technology Research Centre (MOBI), Department of Electrical Engineering and Energy Technology, Vrije Universiteit Brussel, 1050 Brussels, Belgium; majid.vafaeipour@vub.be

* Correspondence: fdogan@adiyaman.edu.tr

Abstract: The real-time prediction of energy production is essential for effective energy management and planning. Forecasts are essential in various areas, including the efficient utilization of energy resources, the provision of energy flexibility services, decision-making amidst uncertainty, the balancing of supply and demand, and the optimization of online energy systems. This study examines the use of tree-based ensemble learning models for renewable energy production prediction, focusing on environmental factors such as temperature, pressure, and humidity. The study's primary contribution lies in demonstrating the effectiveness of the bagged trees model in reducing overfitting and achieving higher accuracy compared to other models, while maintaining computational efficiency. The results indicate that less sophisticated models are inadequate for accurately representing complex datasets. The results evaluate the effectiveness of machine learning methods in delivering valuable insights for energy sectors managing environmental conditions and predicting renewable energy sources

Keywords: environmental factors; renewable energy sources; machine learning; EL; decision tree



Academic Editors: Roberto Romano, Alessandro Lampasi and Sabino Pipolo

Received: 4 November 2024

Revised: 6 December 2024

Accepted: 12 December 2024

Published: 1 January 2025

Citation: Doğan, F.; Oyucu, S.; Unsal, D.B.; Aksöz, A.; Vafaeipour, M. Impact of Environmental Conditions on Renewable Energy Prediction: An Investigation Through Tree-Based Community Learning. *Appl. Sci.* **2025**, *15*, 336. <https://doi.org/10.3390/app15010336>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, scientists have been working on the use and development of systems based on machine learning principles. Machine learning systems achieve almost 100% accuracy in predicting results based on input data with advanced algorithms. Renewable energy production prediction from alternative energy sources is becoming increasingly critical in revealing economic potential and promoting sustainable practices. Machine learning (ML) is a powerful tool that provides almost optimum precision in energy production prediction applications in renewable energy sources. Prediction has significant problems, including data quality fluctuations, environmental uncertainty, and complexities of energy systems. These problems require the investigation of sophisticated ML techniques designed specifically for the renewable energy sector [1]. Such problems have led to various algorithms and artificial intelligence models over time. New algorithms, new approaches emerge, and solutions are produced day by day. ML models use different methods for

classification, prediction, and detection. In these methods, certain paths are followed for training the data. Generally, ML techniques are divided into four categories: supervised, unsupervised, semi-supervised, and reinforcement learning [2–4].

Tree-based models play an important role in predictions of renewable energy resources. Tree-based models are important in improving the energy prediction accuracy in renewable energy resources [5]. Ensemble learning (EL) has shown significant effectiveness in improving the prediction stability using multiple models. EL approaches such as bagging, boosting, and stacking have increased the accuracy and robustness in various applications including renewable energy prediction [6]. Although ML approaches have shown versatility in various fields, their use in real-time renewable energy prediction has not yet been sufficiently explored. Precise prediction models are essential for improving energy production, storage, and integration in smart grids, especially for renewable resources.

This study aims to fill this gap by using tree-based ensemble learning methodologies to increase the accuracy and reliability of renewable energy forecasts, especially targeting photovoltaic production. The effect of the methodologies used in increasing the accuracy of energy production forecasts under different weather environmental conditions is highlighted. The effects of weather conditions are an important step in revealing the connections between the factors affecting energy production, and these connections are critical for making accurate forecasts. While traditional econometric methods, such as regression analysis, have been widely used in renewable energy prediction, they often struggle to capture the non-linear and dynamic interactions between environmental factors and energy production. This study highlights the advantages of machine learning approaches, particularly ensemble learning methods, in addressing these challenges through improved accuracy and robustness. In addition, the novelty of this study lies in its systematic evaluation of tree-based ensemble learning methods, such as bagged and boosted trees, for renewable energy prediction. Unlike prior works that focus on standalone models, this study integrates environmental factors and compares computational efficiency, thereby providing practical insights for optimizing energy management systems.

1.1. Literature Review

The literature indicates significant advancements in ML, especially in EL and tree-based algorithms. The methods, including random forests (RF), Gradient Boosting Machines (GBM), and their derivatives, have reliably attained elevated accuracy in predicting endeavors [7]. The study [8] demonstrated that RF algorithms have exhibited 99% accuracy on datasets such as MNIST, signifying their reliability in classification jobs.

In renewable energy, the precision of prediction is crucial; nonetheless, difficulties remain due to data variability and the unpredictable nature of energy sources. Current research has predominantly concentrated on theoretical progress rather than practical implementation [9–14]. Recent studies in renewable energy prediction highlight the necessity for tailored models to tackle domain-specific issues [15]. The proposed model in [16] employs sophisticated ML approaches, including genetically optimized Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Bidirectional LSTM (BiLSTM), to categorize strokes. Wildwood is a revolutionary RF methodology that offered efficient subtree aggregation for batch-supervised learning, presenting potential applications in renewable energy contexts. Likewise, sophisticated ML models such as genetically optimized CNN and LSTM networks have improved predicted accuracy and potentially fostered advancements in energy forecasting [17,18]. There are also several studies on enhancing security predictions with tree-based applications [19–24]. Table 1 provides detailed highlights of different tree-based approaches and features used in the literature toward identifying the research gaps. EL optimizes Mutual Information Criterion

(MIC) strategies in DL with consortia of classifiers trained from the same dataset to provide improved classification performance [25].

EL is a method that combines multiple models to enhance the overall generalization and prediction ability of a classification or prediction model [26]. Yang and Chen aimed to produce an often more accurate ensemble than any individual model. Their practical performance fueled extensive interest in the development of a variety of applications [24]. In essence, EL methods often yield higher predictive accuracy and generalize better on new datasets than any of the single models. There can be marginal effectiveness over single models but they require additional computational resources and algorithm sophistication [27,28]. There have been efforts using EL to provide more robust global support while handling local and regional details using model training based on regional features [29,30]. Many factors can impact EL's performance, including the choice of base learners, how they are combined, how to manage base learner learning processes, and how to refine the final model, among others [31]. Several theoretical and numerical deep forest studies have shown the existence of a difficult counterpart between the difficulties paramount for EL models' functioning and the characteristics of the training data [32]. Each of these techniques applies to any learning algorithm, making them suitable for classification as wrapper methods. In addition, the final prediction for regression, classification, or probability prediction tasks can be obtained using a weighted average, majority voting, or product of probability values [33].

Table 1. Highlights of the literature comparison.

Ref.	ML Explanation	EL Explanation	Real Time Data Usage	Proposed Algorithm	Tree Based Methods Comparison	Enhancing Results	Suggestions
[34]	✓	X	X	RF	X	X	X
[35]		✓	✓	SVM, MLPNN, DT, GBM, KNN	X	✓	X
[7]	X	X	✓ Victim Detection	Decision Tree	✓	✓	X
[36]	✓	✓	X	X	X	X	X
[15]	✓	X	✓ University of Wisconsin Hospitals	CNN, LSTM, biLSTM	X	✓	X
[16]		X	X	SVM With Kernel	X	✓	X
[18]	X	X	✓ Telecom Industry	Decision Tree	X	✓	X
[24]	X	X	X	Random Forest, SVM, Naive Bayes	X	✓	✓ Spam Detection
[19]	X	X	X	Random Forest, SVM, Neural Networks	X	✓	X
This Study	✓	✓	✓	✓	✓	✓	✓

Table 1 highlights the need for advanced ML and ensemble learning techniques to address the challenges of renewable energy forecasting. It connects the theoretical background to the practical methodologies used in the study, laying the foundation for the research. Table 1 is not just a summary of prior research but also a tool for critically analyzing the field of renewable energy forecasting. It provides a clear rationale for the methodologies adopted in the study and serves as a benchmark for evaluating the performance of tree-based and ensemble learning models in energy prediction tasks. Also, it demonstrates the evolution of renewable energy forecasting methods, transitioning from traditional econometric models to machine learning-based approaches. The results presented in this study align with the literature, confirming the superiority of ensemble learning methods over traditional techniques in terms of accuracy and robustness.

Upon examining its past, the EL traces back to the method of bagging proposed by Leo Breiman [37], which combines multiple decision tree models trained on different sub-

sets of the training data using different random instances of feature subsets. The method, also called Bagging Bootstrap Aggregating, was introduced in 1996 [38]. It is known for reducing the variance, overfitting, and noise in the dataset. In the beginning, several bootstrap samples for the training dataset are generated, and a separate model is trained on each of the datasets by its own training set. After the method of boosting by Robert Schapire in 1990 was proposed [39], EL entered its increasing period. In boosting, multiple weak models are combined to give a good model. It is a method that is used to reduce bias. The weak model in derivative boosting is the one that is better than random classification. The error in boosting is minimized by training another model with a dataset that has more instances whose prediction is not correct. During the prediction of an observation, the weighted majority vote of the modeling is an example of an EL method. The term GBM is described by Friedman with his regression technique [40]. Multiple additive regression trees (MART) is an advancement of gradient-boosted models as also developed by Friedman et al. [40]; with Breiman's report [41], Mason, Baxter et al., and Friedman referred to the generalized abstract class of algorithms on different applications [42–44]. Elith et al. generalize that method as “Boosted Regression Trees” (BRT) [45]. GBM is the name given to it by a well-known open-source R implementation [46], whereas packages that enhance this work use BRT [47]. Another term for it is TreeNet, named for an early commercial implementation made by Dan Steinberg of Salford Systems, one of the academics who first used tree-based approaches [48].

Nonetheless, practitioners have since derived various ensemble methods and steadily expanded their applicability to many areas in the context of different learning algorithms. For example, well-known ensemble methods like random forests, Adaboost, and Gradient Boosting Decision Tree (GBDT), and their variants have been extensively used and studied. Recent experimental results have shown that ensembles can provide significant gambling enhancement when used as classifiers [49]. The study shows that ensemble learners have been successfully used in a wide variety of applications, analyzing data from images, audio, video, and text databases [50]. The EL method derives from empirical research, spanning several application domains and analytics tasks, showing that in practice [51], a group of weak learners (not too weak, not too strong) can outperform a single model or a collection of stronger learners.

1.2. Impact of Environmental Conditions on Renewable Energy Forecasting in the Literature

Environmental conditions play a pivotal role in renewable energy production. Factors such as temperature, humidity, solar irradiance, and wind patterns introduce variability that complicate accurate forecasting. Studies emphasize the necessity of incorporating real-time environmental data into ML models to enhance prediction accuracy. For instance, the research demonstrates that integrating weather forecasts with ML algorithms significantly improves the reliability of solar energy predictions and highlights the effectiveness of these models in improving prediction accuracy [52].

Research on case studies on renewable energy predictions are worldwide:

Mexico by 2050: A study by the International Renewable Energy Agency (IRENA) outlines Mexico's potential to achieve a 100% renewable energy system by 2050. The research emphasizes the importance of integrating solar and wind energy, considering the country's diverse climatic conditions. Advanced forecasting models are recommended to manage variability and ensure grid stability [53].

Latin America and the Caribbean by 2050: IRENA's report on the region highlights the feasibility of a renewable energy transition, focusing on the integration of hydropower, solar, and wind resources. The study discusses the need for robust forecasting tools to address environmental challenges such as seasonal variability and extreme weather events [53].

Spain: Research published in the journal *Renewable Energy* examines Spain's path to a 100% renewable energy system. The study utilizes machine learning models to predict energy production from solar and wind sources, accounting for environmental factors like temperature and wind speed. The findings suggest that accurate forecasting is crucial for effective energy management and policy planning [54].

Japan: A study in the journal *Energy* explores Japan's renewable energy prospects, emphasizing the role of advanced forecasting models in managing the country's unique environmental conditions, including typhoons and seasonal changes. The research highlights the integration of machine learning techniques to improve the accuracy of energy production predictions [55].

Island Case Studies (Canary and Galapagos Islands): Research on the Canary Islands investigates the challenges of achieving 100% renewable energy in isolated systems. The study emphasizes the importance of accurate forecasting models to manage the variability of renewable sources and ensure energy security [56]. Similarly, a study on the Galapagos Islands discusses the integration of renewable energy sources and the necessity of reliable prediction models to address environmental challenges unique to island ecosystems [57].

Southeast Asia by 2050: An analysis by IRENA focuses on Southeast Asia's renewable energy outlook, highlighting the region's potential for solar and wind energy. The report underscores the need for advanced forecasting tools to manage environmental factors such as monsoons and tropical storms, which significantly impact energy production [52].

In conclusion, the literature research shows that achieving 100% renewable energy by 2050 in various regions necessitates the development and implementation of advanced forecasting models that account for environmental conditions. Machine learning techniques offer a viable solution to enhance prediction accuracy, thereby facilitating the integration of renewable energy sources into existing power systems. Also, renewable energy generation prediction presents numerous challenges such as the following:

Data Quality and Variability: Includes variable environmental conditions that impact data reliability.

Intricacy of Energy Systems: The interplay of various variables, including meteorological conditions and energy consumption, complicates modeling.

Integration with Smart Grids: Precise predictions are crucial for equilibrating energy supply and demand in real time.

Confronting these problems necessitates resilient ML models that can adjust to changing settings and yield interpretable outcomes. In addition, prediction photovoltaic (PV) energy has emerged as a vital element of renewable energy systems. ML approaches, such as support vector machines (SVM), artificial neural networks (ANN), and EL methods, have been extensively utilized. Research demonstrates that ensemble methods, including GBM and RF, surpass conventional statistical techniques in predicting PV production by effectively capturing non-linear correlations and minimizing errors.

The incorporation of real-time data, including solar irradiance and meteorological variables, improves the precision of photovoltaic forecasting models. Nonetheless, obstacles persist in expanding these models to accommodate greater datasets and integrating long-term forecasts. Subsequent studies ought to concentrate on enhancing machine learning techniques to mitigate these constraints and augment the incorporation of photovoltaic predictions into smart grid systems. This study confirms the supremacy of ML methods, particularly bagged trees, over traditional econometric approaches in renewable energy prediction. By effectively handling non-linear relationships and complex environmental interactions, ML models provide a significant advancement in accuracy and practicality for real-world applications

The reasons why DL models are superior to decision trees can be stated as follows: Since it can extract features from data, it learns more complex relationships and patterns. Model performance increases in large datasets. It has a higher performance than decision trees. This is because it contains more parameters. However, the low number of parameters in the dataset examples given in our study has led to the preference for decision trees instead of DL models, since they contain little information such as temperature, humidity, and pressure. This allows important information to be obtained in terms of explaining how the model makes decisions. It enables the interpretation of the given parameters and the emergence of relationships between them. In addition, having low computational costs allows it to produce the results more quickly. Another factor in preferring decision trees is that there is no need for data pre-processing.

This research combines ML techniques with specialized knowledge to improve the precision of renewable energy predictions. The study seeks to deliver practical insights for optimizing energy production and integration by utilizing advanced ML approaches such as EL models. The results are anticipated to enhance the wider implementation of ML in renewable energy, tackling both theoretical and practical deficiencies.

Notwithstanding the considerable progress in machine learning methodologies for renewable energy forecasting, there is a deficiency in studies that thoroughly assess tree-based ensemble algorithms in real-time contexts. Moreover, the majority of the research neglects the computational efficiency of these models, which is essential for practical use. Also, this study addresses this gap by performing a thorough investigation of tree-based algorithms, highlighting their accuracy, resilience, and computational viability for photovoltaic energy prediction across diverse environmental conditions.

2. Materials and Methods

Using decision trees to estimate energy production from solar panels provides significant advantages. Decision trees are easy to use and interpretable, do not require data preprocessing, and can model non-linear relationships. Non-linear energy production and the factors affecting it affect energy production in different environmental conditions. Decision trees can model this complex relationship flexibly and detail the effects of environmental factors. In addition, since these models do not have the risk of over-learning, they can demonstrate their generalization abilities. Providing flexibility, accuracy, interpretability, low cost, and high speed reveal the advantages of these models used for the purpose of the present study. In this section, the methodology used to examine the effects of environmental conditions on renewable energy estimates will be discussed in detail. The accurate prediction of renewable energy resources including solar is very important for providing energy management and sustainability insights. The present study provides detailed information on improving the performance of tree-based learning methods linked to the effects of environmental variables. Different tree structures such as fine tree, medium tree, coarse tree, bagged tree, and boosted tree are among the methods used make it possible to effectively model the complexity and variability in the dataset. In-depth analyses will be made on how each of these methods responds to the characteristics of the dataset and environmental conditions, and the reliability and validity of the results will be evaluated. In addition, information about the scope and content of the datasets used will be provided, and the training processes and performance criteria of each model will be detailed. Consequently, the role of environmental factors in renewable energy estimates will be realized while having the literature of the field contributed.

2.1. Data Description

The study used a data collection that includes temperature, pressure, humidity, and solar energy production. Table 2 includes the maximum, minimum, and median data of temperature, pressure, humidity, and energy amounts in the dataset used from 17,544 data points. The dataset includes real-time solar energy production amounts originating from the EPIAŞ platform, which is the authorized data supplier for Turkey's energy provider [58]. The dataset, taken every hour for 729 days for 24 months, was collected from a 1 MW solar power plant located in Sivas City Center for the period of January 2020–December 2021. Environmental data are accessed simultaneously from Meteonorm and Weatherspark databases through the same interface.

Table 2. Statistical information about the dataset.

	Max	Min	Median
Temperature (°C)	39.2	−22.9	10.5
Pressure (Psi)	888.2	854.6	872
Humidity (RH)	97	−22.9	62
Solor Production (MWe)	10.04	854.6	0

2.2. Variable Description

The variables used in this study are as follows:

Temperature (°C): Affects the efficiency of solar panels, with higher temperatures often leading to reduced energy conversion efficiency.

Pressure (Psi): Represents atmospheric conditions that influence solar irradiance and, consequently, energy generation.

Humidity (%): Determines atmospheric clarity, impacting the amount of solar radiation reaching the panels.

Solar Energy Production (MW): The target variable, representing the energy output of the solar plant.

The K-Fold cross-validation [59] method was used in the study. The data were made using 5-fold cross-validation and was set to 20% for 80% test for training. K-Fold cross-validations provide the opportunity to divide the dataset into k equal parts and use each part as a training-test set, respectively. Consequently, the assessment of the complete dataset is guaranteed. In other methodologies [60], the random selection of training and test sets may, in certain instances, result in the test set providing a misleading assessment of the model's overall performance. Although the Stratified K-Fold method [61] performs better in cases where the classes are imbalanced, the overall performance of K-Fold is quite valuable in terms of evaluating the overall performance. Furthermore, the K-Fold method reduces the risk of over-learning and allows performing the test on different data subsets. This increases the generalization ability of the model. Methods such as Leave-One-Out [62] can be effective for very small datasets; however, it is not favorable in terms of computational load when it comes to large datasets. In terms of computational efficiency, K-Fold cross-validation allows the training time of the model to be averaged for each fold. Other methods may cause the same model to be trained repeatedly in each evaluation, which leads to a loss of time. K-Fold cross-validation is chosen for its efficacy in dataset representation, model generalization capability, and computing economy.

2.3. Decision Trees

Decision trees are rule-based algorithms used to solve classification and regression problems. It is one of the most appropriate algorithms used in classification and prediction processes because of its ease of understanding and interpretation [63]. An example of the

decision trees consisting of roots, knots, and leaves is shown in Figure 1. The root nodes are at the top, the internal nodes are in the middle, and the leaf nodes are located at the bottom.

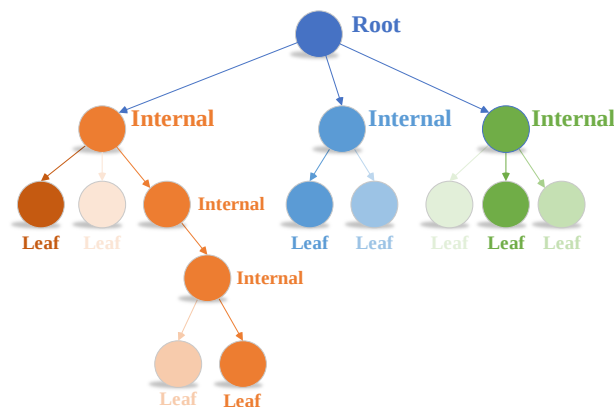


Figure 1. Decision tree diagram.

2.4. Fine Tree

A fine tree contains a low complexity of branches and partitions leading to simpler rules. It is especially used to test the complexity of decision trees and the generalization of the model. It is a decision tree used when datasets are low. When resources are limited, it is preferred for real-time applications. The fine tree has fewer fading points. Each branch contains a small number of properties and rules. Thanks to their simplicity, the training and testing process of fine trees is relatively fast with low overfitting probability [64].

2.5. Medium Tree

Medium tree is a decision tree of medium complexity. The complexity level is higher than that of the fine tree. It is usually used to optimize modeling. The complexity of the trees involves a structure where the fissures and divisions are moderate. Every decision point and rule has a certain depth. Each twist represents a certain level of complexity. Medium tree captures middle-level qualities in each data and is designed to work on both simple and complex data. Therefore, this method is suitable for data processing in different dimensions and usually has an average training and test duration even in cases of data change [65].

2.6. Coarse Tree

Coarse tree is a structure in which branches and divisions of trees have a higher degree of complexity with deeper fissures. The level of complexity is higher because of the high twisting and depth. It is more convenient to use it for the evaluation of larger dataset details. Decision points can make the details come to light better. However, there is a risk of over-adaptation in the dataset. This increases the likelihood of overfitting against noise and non-pattern data in the data. It is used both in high-resolution datasets and in complex structures leading to lengthy training and test periods [66]. The complexity of the fine trees, medium trees, and coarse tree trees, the number of leaves, the training and test time, and the overfitting comparison are shown in Figure 2.

The ensemble tree method seeks to enhance the effectiveness and predictability of a model by combining numerous decision trees. These strategies encompass two prominent types of models, boosted trees and bagged trees. These strategies are designed to merge trees in various ways to enhance their performance.



Figure 2. Decision trees features comparison.

2.7. Boosted Trees

The aim of this approach is to advance by methodically correcting errors in the tree. Sequencing is employed to train trees. Each tree corrects the shortcomings of its predecessor. This model is trained on trees and is primarily based on the fundamental model. The weight of each data point is calculated, and attempts are made to correct the mistaken prediction points. The recently established trees are trained using a dataset of this magnitude. The errors of the previous model have been rectified. Once the model training is over, the outcomes manifest through the aggregation of weights. Operations in this model are executed concurrently. Therefore, the operations are simultaneously trained. The bug extraction feature is a prominent aspect of this model. It possesses the ability to effectively extrapolate patterns and construct intricate models. Due to the potential for overfitting, it is essential to appropriately alter the settings. AdaBoost, Gradient Boosting, XGBoosting, LightGBM, and CatBoost are boosting algorithms that have been specifically designed and improved to enhance their performance and efficiency [67].

2.8. Bagged Trees

This model aims to decrease the variability of the tree. It extracts multiple sub-samples from the data collected and trains a decision tree on each. These operations are executed sequentially. A novel tree arises by amalgamating the outcomes of all the individual trees. This tree attempts to generate predictions as the definitive tree. This process occurs in three distinct steps. Sampling, modeling, and consolidation. This decision tree operates on diverse samples, hence diminishing the variance of the model.

Additionally, it serves as a highly efficient approach for mitigating overfitting in decision trees. Random forest is a type of bagging technique. In these three variants, the attributes used for splitting at each node are chosen randomly. Figure 3 illustrates the operational mechanism of bagging and boosting on trees. Parallel bagging executes each tree concurrently, whereas boosting is carried out sequentially [68].

The flow diagram illustrating the application of the models in the investigations is provided in Figure 4.

Data such as temperature, pressure, and humidity measurements obtained from solar panels are used as inputs in our predictive model to estimate PV energy production. These measured environmental parameters are important to estimate the potential production output through the applied methodology using machine learning algorithms.

Before training the models, several modifications must be implemented to ensure their usability. Table 3 includes details about these nuances. Additionally, it provides information on the duration of training and the specific sections that are stored in memory. The purpose of these decision trees is to enhance performance by augmenting correlation [63].

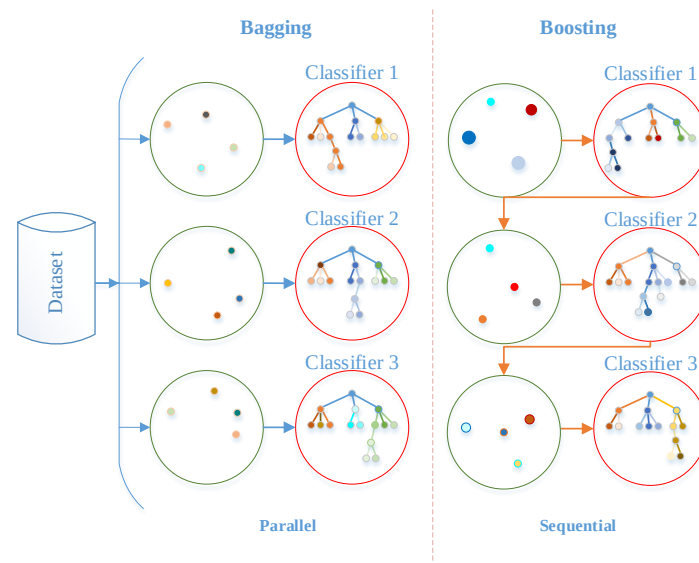


Figure 3. The concepts of parallel and sequential processing.

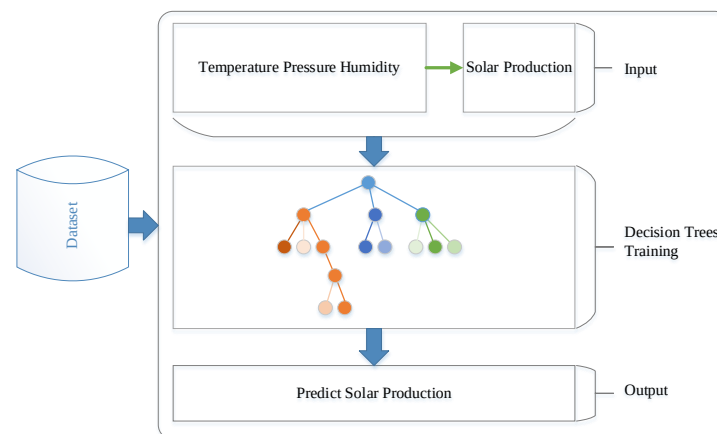


Figure 4. Graphical illustration of the predictive algorithm of decision trees.

Table 3. Hyperparameters, training time, and size used for the models.

Model	Training Time (Sec)	Prediction Speed (Obs/Sec)	Model Size (Bytes)	Hyperparameters
Custom Tree	9.207	72,806	337,475	Minimum leaf size: 7 Surrogate decision splits: Off.
Fine Tree	7.008	67,993	475,771	Minimum leaf size: 4 Surrogate decision splits: Off.
Medium Tree	6.509	86,341	21,2367	Minimum leaf size: 12 Surrogate decision splits: Off.
Coarse Tree	5.673	90,821	75,919	Minimum leaf size: 36 Surrogate decision splits: Off.
Boosted Trees	11.720	95,299	152,608	Minimum leaf size: 8 Number of learners: 30 Learning rate: 0.1 Number of predictors to sample: Select All
Bagged Trees	13.733	41,320	7,909,624	Minimum leaf size: 8 Number of learners: 30 Number of predictors to sample: Select All

2.9. Evaluation Measures

The outcomes achieved using the specified hyperparameters for decision trees are detailed in the results section, employing evaluation metrics.

The evaluation metrics used are *RMSE* (Root Mean Square Error), *MSE* (Mean Squared Error), and R^2 (R-squared) [69].

RMSE is a quantitative statistical metric that assesses the degree of deviation between estimated and actual values. This relationship is mathematically expressed in Equation (1).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (1)$$

Here, y_i is the actual value and $\hat{y}_i$ is the predicted value, and n represents the number of samples.

MSE is a statistical metric that quantifies the average squared difference between an estimated value and the true value as below. This calculation is formally represented in Equation (2).

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2)$$

In Equation (3), R^2 , often known as RSquared, is a statistical metric that yields a value ranging from 0 to 1. This statistic is crucial for both explaining and predicting data points.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

In Equation (4) *MAE* quantifies the absolute difference between the predicted value and the true value. It quantifies the magnitude of the medium, irrespective of the orientation of the errors.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4)$$

3. Experimental Results

The hyperparameters employed in the fine tree, medium tree, coarse tree, bagged tree, and boosted tree models were provided, along with corresponding parameters and details regarding training time and memory usage. Figure 5 displays the current prediction data and the corresponding outcomes for the verification estimate, based on the information provided.

The temperature, pressure, and humidity parameters within the dataset have an impact on the production quantity. Each parameter in this effect carries a distinct weight. These values have a direct impact on the level of production. Figure 6 illustrates the quantity of energy generated carrying the influence of these independent variables.

The findings are evaluated based on the information provided in each row of the training validation dataset. The study was performed in MATLAB Toolboxes were used. The evaluation metrics findings are provided in Table 4. Based on these findings, bagged tree shows superior performance compared to the other examined decision trees while fine tree ranked last.

The box plot obtained according to the data of error indices is shown in Figure 7. Using the data provided, we achieved results by independently using decision tree models in the study. The prediction of solar energy production is derived from factors that are reliant on temperature, humidity, and pressure. Decision trees are the most interpretable regression models that can be derived from this type of data. The decision trees, namely

the fine tree, medium tree, and coarse tree, exhibit similarities but also possess varying levels of complexity. The accuracy of the outcomes derived using these decision trees has been proven to be inferior to that of the ensemble decision tree.

Empirical evidence suggests that higher levels of complexity are positively correlated with improved outcomes. Studies employing econometric methods for energy prediction often report significant limitations, such as reduced accuracy when dealing with non-linear and high-dimensional datasets. For example, linear regression models typically achieve R^2 values between 0.3 and 0.5 in similar applications, whereas the bagged trees model in this study achieved an R^2 of 0.5650, demonstrating its superior performance. This comparison underscores the potential of ML approaches to address the inherent complexities of renewable energy forecasting.

Table 4. Results of statistical indices from decision trees.

Model Type	Preset	Status	RMSE (Validation)	MSE (Validation)	R^2 (Validation)	MAE (Validation)	MAE (Test)	MSE (Test)	RMSE (Test)	R^2 (Test)
Tree	Fine Tree	Tested	1.9696	3.8793	0.4264	1.1718	1.1978	4.0324	2.0081	0.4007
Tree	Medium Tree	Tested	1.8213	3.3172	0.5095	1.1198	1.1256	3.3111	1.8196	0.5079
Tree	Coarse Tree	Tested	1.7658	3.1180	0.5390	1.1104	1.1253	3.1815	1.7837	0.5272
Ensemble	Boosted Trees	Tested	1.7566	3.0857	0.5437	1.1462	1.1657	3.1479	1.7742	0.5322
Ensemble	Bagged Trees	Tested	1.7152	2.9420	0.5650	1.0666	1.0781	2.9928	1.7300	0.5552

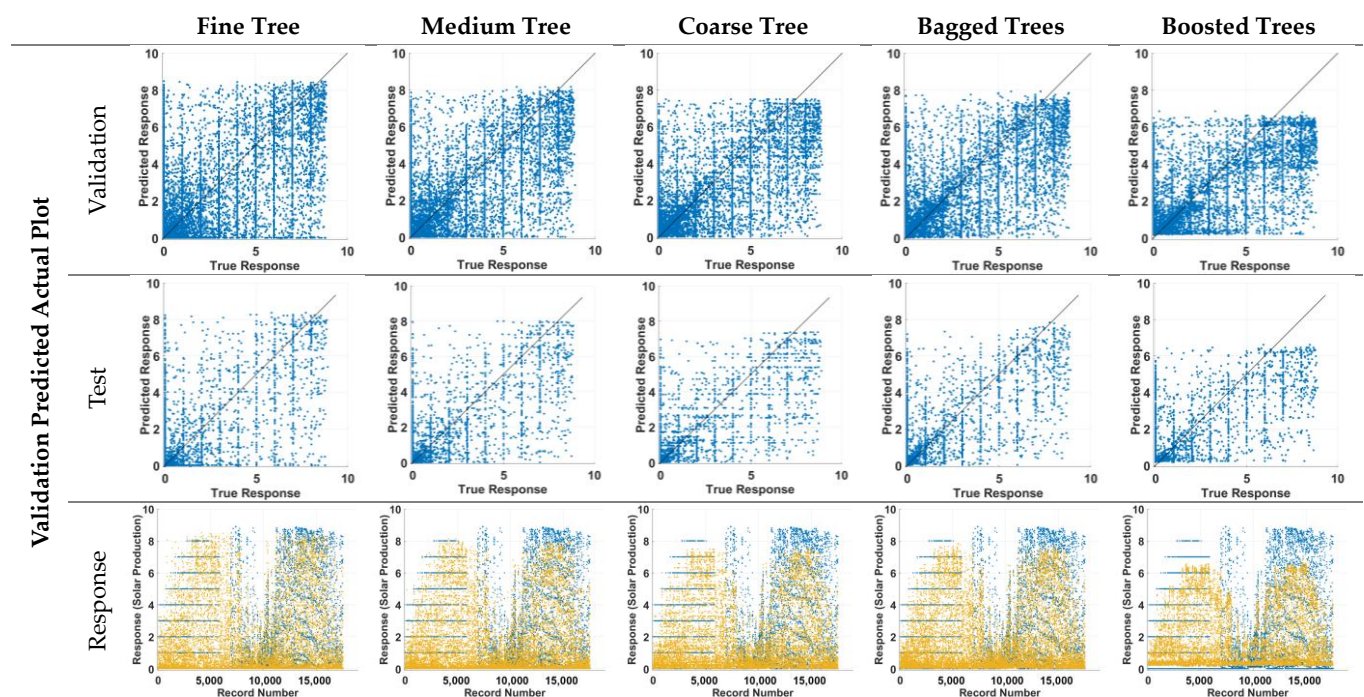


Figure 5. Tests, verification, and responses obtained after training.

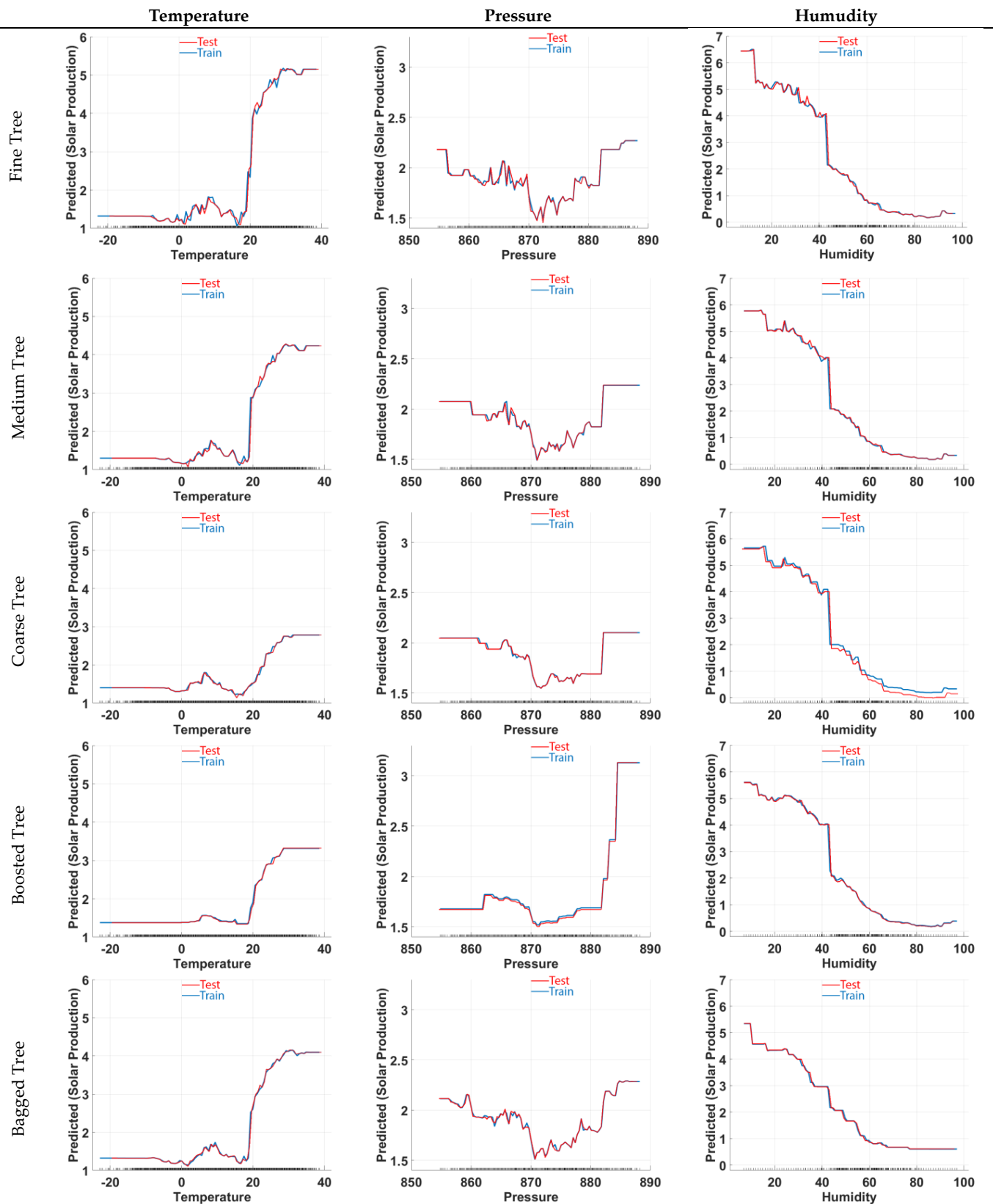


Figure 6. Temperature, pressure, and humidity effect graphs in energy production with datasets.

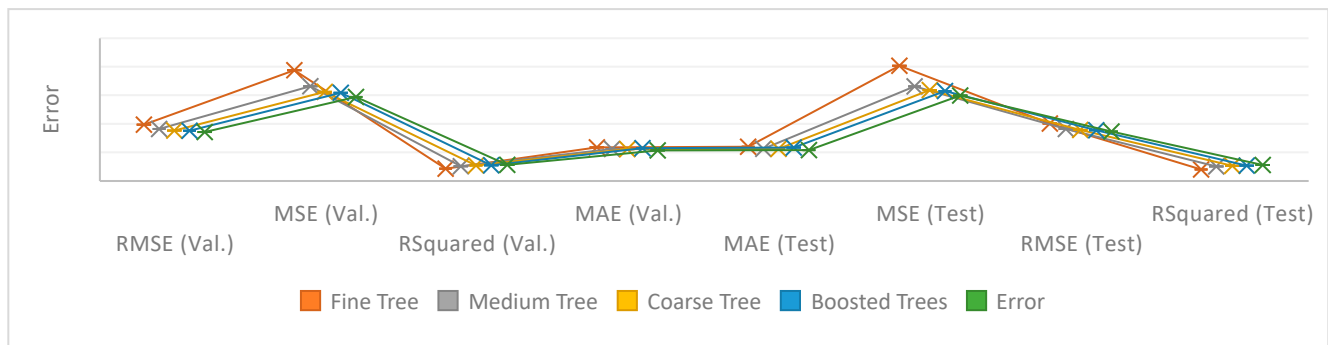


Figure 7. Box plot showing the distribution of errors between different models.

4. Discussion and Conclusions

This study aimed to evaluate the effectiveness of various tree-based EL models in predicting photovoltaic (PV) energy production by analyzing the effects of environmental factors such as temperature, pressure, and humidity. The results revealed that bagged trees and boosted trees models exhibit favorable prediction results on complex datasets, which highlight the advantages of using advanced tree-based algorithms for similar use cases. The bagged trees model was shown to be the most effective model, producing an *RMSE* of 1.7152 and an R^2 value of 0.5650. These metrics showed that the model effectively reduced dataset noise and minimized overfitting, thus increasing the reliability of the predictions. Regarding the learning results, bagged tree and boosted tree models achieved the best performances among tree-based models with a minimum of eight leaves. Bagged tree trained each decision tree on many different subsamples taken from the dataset and combined the results of all the trees to produce a new tree. This tree was tried to produce predictions as the final tree. At this point, it has been seen that the bagged tree produces successful results and exhibits superior performance than other tree structures. The working principle has shown that it produces better results than other tree structures. The fine tree model showed the weakest performance with an *RMSE* value of 1.9696 and an R^2 value of 0.4264. This highlights the limitations of simpler decision tree models when faced with complex datasets, suggesting that they may not be equipped to capture the complex relationships found in PV energy production. Our findings are consistent with previous literature suggesting that EL approaches such as bagged and boosted trees tend to outperform individual decision trees. Figure 5 visually presents the validation, testing, and response results of the fine tree, medium tree, coarse tree, bagged tree, and boosted tree. As shown in Table 4, extracted through a systematic review of the literature, the superiority of these models reinforces the idea that more sophisticated algorithms are required to accurately predict energy production under changing environmental conditions. However, it is important to acknowledge that training these EL models requires significant computational resources and time. When we look at the *MAE* validation and *MAE* test results, the fact that the difference between these two values is very small indicates that there is no overfitting and that the data and models work correctly. Researchers working on this topic should consider the size and complexity of the datasets when selecting appropriate models for real-world applications. Despite significant advances in tree-based, random tree, and random forest algorithms, the current body of research focuses primarily on survey studies rather than practical applications in the energy sector. In addition, although tree-based algorithms are widely used in healthcare, especially in disease identification through image processing, their application in processing real-time data for energy prediction remains limited. This study addresses this gap by providing a comparative analysis of different methods, thus providing valuable insights into the existing literature.

The findings show that solar energy forecasting models have the potential to significantly increase the accuracy of production estimates. The effects of temperature, pressure, and humidity on energy production are presented in Figure 6. It shows how the temperature, pressure, and humidity data produced results during the training and testing period. The figures showing a comparative result provide suggestive indicators. The models developed in our study could apply to effective management and strategic planning of renewable energy resources. Future research should investigate the validation of these models under various environmental conditions and geographical regions, thereby expanding their applicability and improving the prediction accuracy under various climatic conditions. The results reveal that temperature, pressure, and humidity have significant impacts on PV energy production, requiring models capable of capturing their non-linear interactions. Policies supporting localized weather monitoring and adaptive forecasting systems can enhance the accuracy of renewable energy management and planning.

Investigating the impact of these models on the management and strategic planning of renewable energy resources can further strengthen their practical benefits in the energy sector. Strengthening the algorithms and working with real-time data can lead to new working approaches that will improve energy efficiency and sustainability.

This study not only addressed the gaps identified through the literature but also highlighted the critical role of advanced tree-based EL models in the future of PV energy production forecasting. Proving that bagged trees offer the optimal balance between accuracy and computing efficiency. Establishing a comprehensive framework for the incorporation of environmental data into prediction models. Continuing research in this area merits further efforts through the evaluation of accuracy, computational cost, and applicability of such models in a holistic manner for developing robust renewable energy management and planning tools.

Author Contributions: Conceptualization, F.D. and S.O.; methodology, F.D.; software, F.D.; validation, A.A., S.O. and D.B.U.; formal analysis, S.O., F.D. and M.V.; investigation, S.O.; resources, D.B.U.; data curation, F.D.; writing—original draft preparation, F.D.; writing—review and editing, M.V.; visualization, F.D.; supervision, M.V.; project administration, A.A.; funding acquisition, A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the European Union’s Horizon Europe research and innovation program under grant number: 101172877.

Institutional Review Board Statement: This study does not require an ethics statement. It does not contain any information involving humans or animals.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset used in this study was obtained from EPIAŞ (Energy Exchange Istanbul). The data can be accessed through their official website: <https://www.epias.com.tr>, accessed on 6 December 2024. Please note that specific datasets may require proper authorization or permissions as per EPIAŞ’s data access policies.

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

References

1. Prasad, R.; Joseph, L.; Deo, R.C. Modeling and forecasting renewable energy resources for sustainable power generation: Basic concepts and predictive model results. In *Translating the Paris Agreement into Action in the Pacific*; Springer: Cham, Switzerland, 2020; pp. 59–79.
2. Vogt, J. Supervised & Unsupervised Transfer Learning. University of Basel. 2013. Available online: http://edoc.unibas.ch/diss/DissB_10347 (accessed on 6 December 2024).
3. Dai, K.; Yu, H.-Y.; Li, Q. A semisupervised feature selection with support vector machine. *J. Appl. Math.* **2013**, *2013*, 416320. [CrossRef]

4. Ernst, D.; Louette, A. Generative AI. *Bus. Inf. Syst. Eng.* **2024**, *66*, 111–126.
5. Lu, T.; Ai, Q.; Lee, W.-J.; Wang, Z.; He, H. An aggregated decision tree-based learner for renewable integration prediction. In Proceedings of the 2018 IEEE Industry Applications Society Annual Meeting (IAS2018), Portland, OR, USA, 23–27 September 2018.
6. Strobl, C.; Malley, J.; Tutz, G. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol. Methods* **2009**, *14*, 323. [[CrossRef](#)] [[PubMed](#)]
7. Seeja, G.; Doss, A.S.A.; Hency, V.B. A Novel Approach for Disaster Victim Detection Under Debris Environments Using Decision Tree Algorithms with Deep Learning Features. *IEEE Access* **2023**, *11*, 54760–54772. [[CrossRef](#)]
8. Devyatkin, D.A.; Grigoriev, O.G. Random kernel forests. *IEEE Access* **2022**, *10*, 77962–77979. [[CrossRef](#)]
9. Zounemat-Kermani, M.; Batelaan, O.; Fadaee, M.; Hinkelmann, R. Ensemble machine learning paradigms in hydrology: A review. *J. Hydrol.* **2021**, *598*, 126266. [[CrossRef](#)]
10. Mahadevkar, S.V.; Khemani, B.; Patil, S.; Kotecha, K.; Vora, D.R.; Abraham, A.; Gabralla, L.A. A review on machine learning styles in computer vision—Techniques and future directions. *IEEE Access* **2022**, *10*, 107293–107329. [[CrossRef](#)]
11. Halbouni, A.; Gunawan, T.S.; Habaebi, M.H.; Halbouni, M.; Kartiwi, M.; Ahmad, R. Machine learning and deep learning approaches for cybersecurity: A review. *IEEE Access* **2022**, *10*, 19572–19585. [[CrossRef](#)]
12. Duarte, V.; Zuniga-Jara, S.; Contreras, S. Machine learning and marketing: A systematic literature review. *IEEE Access* **2022**, *10*, 93273–93288. [[CrossRef](#)]
13. Tursunaliyeva, A.; Alexander, D.L.J.; Dunne, R.; Li, J.; Riera, L.; Zhao, Y. Making Sense of Machine Learning: A Review of Interpretation Techniques and Their Applications. *Appl. Sci.* **2024**, *14*, 496. [[CrossRef](#)]
14. Afifi, H.; Pochaba, S.; Boltres, A.; Laniewski, D.; Haberer, J.; Paeleke, L.; Poorzare, R.; Stolpmann, D.; Wehner, N.; Redder, A.; et al. Machine Learning with Computer Networks: Techniques, Datasets, and Models. *IEEE Access* **2024**, *12*, 54673–54720. [[CrossRef](#)]
15. Puspitasari, D.I.; Kholdani, A.F.R.; Dharmawati, A.; Rosadi, M.E.; Dhuhita, W.M.P. Stroke disease analysis and classification using decision tree and random forest methods. In Proceedings of the 2021 Sixth International Conference on Informatics and Computing (ICIC), Jakarta, Indonesia, 3–4 November 2021; IEEE: Piscataway, NJ, USA.
16. Saleem, M.A.; Javeed, A.; Akarathanawat, W.; Chutinet, A.; Suwanwela, N.C.; Asdornwised, W.; Chaitusaney, S.; Deelertpaiboon, S.; Srisiri, W.; Benjapolakul, W.; et al. Innovations in stroke identification: A machine learning-based diagnostic model using neuroimages. *IEEE Access* **2024**, *12*, 35754–35764. [[CrossRef](#)]
17. Gaïffas, S.; Merad, I.; Yu, Y. WildWood: A new random forest algorithm. *IEEE Trans. Inf. Theory* **2023**, *69*, 6586–6604. [[CrossRef](#)]
18. Amin, A.; Shah, B.; Khattak, A.M.; Moreira, F.J.L.; Ali, G.; Rocha, A.; Anwar, S. Cross-company customer churn prediction in telecommunication: A comparison of data transformation methods. *Int. J. Inf. Manag.* **2018**, *46*, 304–319. [[CrossRef](#)]
19. Phua, C.; Lee, V.; Smith, K.; Gayler, R. A comprehensive survey of data mining-based fraud detection research. *arXiv* **2010**, arXiv:1009.6119.
20. Mehdiyev, N.; Majlatow, M.; Fettke, P. Quantifying and explaining machine learning uncertainty in predictive process monitoring: An operations research perspective. *Ann. Oper. Res.* **2024**, 1–40. [[CrossRef](#)]
21. Sadeghi, K.; Ojha, D.; Kaur, P.; Mahto, R.V.; Dhir, A. Explainable artificial intelligence and agile decision-making in supply chain cyber resilience. *Decis. Support Syst.* **2024**, *180*, 114194. [[CrossRef](#)]
22. MahendraVardhan, A.; Sridhar, S. Determining False Positive Analysis of Software Vulnerabilities with Predefined Scan Rules using Random Forest Classifier and Decision Tree Technique. In Proceedings of the 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 16–17 December 2022.
23. Vaccari, I.; Carlevaro, A.; Narteni, S.; Cambiaso, E.; Mongelli, M. eXplainable and reliable against adversarial machine learning in data analytics. *IEEE Access* **2022**, *10*, 83949–83970. [[CrossRef](#)]
24. Arya, V.; Almomani, A.A.D.; Mishra, A.; Peraković, D.; Rafsanjani, M.K. Email spam detection using naive Bayes and random forest classifiers. In *International Conference on Cyber Security, Privacy and Networking*; Springer: Berlin/Heidelberg, Germany, 2021.
25. Yang, Y.; Lv, H.; Chen, N. A Survey on ensemble learning under the era of deep learning. *Artif. Intell. Rev.* **2022**, *56*, 5545–5589. [[CrossRef](#)]
26. Jia, J.; Liang, W.; Liang, Y. A review of hybrid and ensemble in deep learning for natural language processing. *arXiv* **2023**, arXiv:2312.05589.
27. Edeh, M.O.; Dalal, S.; Ben Dhaou, I.; Agubosim, C.C.; Umoke, C.C.; Richard-Nnabu, N.E.; Dahiya, N. Artificial intelligence-based ensemble learning model for prediction of hepatitis C disease. *Front. Public Health* **2022**, *10*, 892371. [[CrossRef](#)]
28. Huang, B.; Huang, H.; Zhang, S.; Zhang, D.; Shi, Q.; Liu, J.; Guo, J. Artificial intelligence in pancreatic cancer. *Theranostics* **2022**, *12*, 6931–6954. [[CrossRef](#)] [[PubMed](#)]
29. Saleh, H.; Mostafa, S.; Alharbi, A.; El-Sappagh, S.; Alkhalifah, T. Heterogeneous ensemble deep learning model for enhanced Arabic sentiment analysis. *Sensors* **2022**, *22*, 3707. [[CrossRef](#)] [[PubMed](#)]
30. Bleichrodt, A.; Bourouiba, L.; Chowell, G.; Lofgren, E.T.; Reed, J.M.; Ryan, S.J.; Fefferman, N.H. Assembling ensembling: An adventure in approaches across disciplines. *arXiv* **2024**, arXiv:2405.02599.

31. Ortega, L.A.; Cabañas, R.; Masegosa, A. Diversity and generalization in neural network ensembles. In Proceedings of the International Conference on Artificial Intelligence and Statistics, Virtual, 28–30 March 2022.
32. Zhou, Z.-H.; Feng, J. Deep forest. *Natl. Sci. Rev.* **2019**, *6*, 74–86. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Rahman, A.; Tasnim, S. Ensemble classifiers and their applications: A review. *Int. J. Comput. Trends Technol.* **2014**, *10*, 31–35. [\[CrossRef\]](#)
34. Zhang, J.M.; Harman, M.; Ma, L.; Liu, Y. Machine learning testing: Survey, landscapes and horizons. *IEEE Trans. Softw. Eng.* **2020**, *48*, 1–36. [\[CrossRef\]](#)
35. Althubiti, S.A.; Paul, S.; Mohanty, R.; Mohanty, S.N.; Alenezi, F.; Polat, K. Ensemble learning framework with GLCM texture extraction for early detection of lung cancer on CT images. *Comput. Math. Methods Med.* **2022**, *2022*, 2733965. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Chaudhary, A.; Singh, S.S. Lung cancer detection on CT images by using image processing. In Proceedings of the 2012 International Conference on Computing Sciences (ICCS), Phagwara, India, 14–15 September 2012.
37. Breiman, L. Heuristics of instability and stabilization in model selection. *Ann. Stat.* **1996**, *24*, 2350–2383. [\[CrossRef\]](#)
38. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [\[CrossRef\]](#)
39. Schapire, R.E. The strength of weak learnability. *Mach. Learn.* **1990**, *5*, 197–227. [\[CrossRef\]](#)
40. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [\[CrossRef\]](#)
41. Breiman, L. *Arcing the Edge Technical Report 486*; Statistics Department. University of California: Berkeley, CA, USA, 1997.
42. Mason, L.; Baxter, J.; Bartlett, P.; Frean, M. Boosting algorithms as gradient descent. *Adv. Neural Inf. Process. Syst.* **1999**, *12*, 512–518.
43. Friedman, J.H. Stochastic gradient boosting. *Comput. Stat. Data Anal.* **2002**, *38*, 367–378. [\[CrossRef\]](#)
44. Friedman, J.H.; Meulman, J.J. Multiple additive regression trees with application in epidemiology. *Stat. Med.* **2003**, *22*, 1365–1381. [\[CrossRef\]](#)
45. Elith, J.; Leathwick, J.R.; Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **2008**, *77*, 802–813. [\[CrossRef\]](#)
46. Ridgeway, G. Generalized Boosted Models: A guide to the gbm package. *Update* **2007**, *1*, 2007.
47. Elith, J.; Leathwick, J. Boosted Regression Trees for Ecological Modeling. R Documentation. 2017. Available online: <https://citeseeerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=3fa2b8826c881d732169995869b9d356c6996029> (accessed on 6 December 2024).
48. Steinberg, D. (n.d.). Overview of Data Mining and Its Applications [Lecture Notes]. Available online: <https://citeseeerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=3b3b3eeca6d56a0241efe98e13c36625de6f3f33> (accessed on 6 December 2024).
49. Hsu, K.-W. A theoretical analysis of why hybrid ensembles work. *Comput. Intell. Neurosci.* **2017**, *2017*, 1930702. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Ganaie, M.A.; Hu, M.; Malik, A.K.; Tanveer, M.; Suganthan, P.N. Ensemble deep learning: A review. *Eng. Appl. Artif. Intell.* **2022**, *115*, 105151. [\[CrossRef\]](#)
51. Yasmin, A.; Haider Butt, W.; Daud, A. Ensemble effort estimation with metaheuristic hyperparameters and weight optimization for achieving accuracy. *PLoS ONE* **2024**, *19*, e0300296. [\[CrossRef\]](#)
52. Mohabbati, M. A Survey on Machine Learning Applications in Renewable Energies Forecasting. In *Artificial Intelligence in the Operation and Control of Digitalized Power Systems*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 305–326.
53. International Renewable Energy Agency. Home. n.d. 2 December 2024. Available online: <https://www.irena.org> (accessed on 6 December 2024).
54. Renewable Energy World News. 2 December 2024. Available online: <https://www.renewableenergyworld.com/news/> (accessed on 6 December 2024).
55. Energy Rich Japan. Welcome. n.d. 2 December 2024. Available online: <https://www.energyrichjapan.info/en/welcome.html> (accessed on 6 December 2024).
56. Gils, H.C.; Simon, S. Carbon neutral archipelago—100% renewable energy supply for the Canary Islands. *Appl. Energy* **2017**, *188*, 342–355. [\[CrossRef\]](#)
57. Icaza-Alvarez, D.; Jurado, F.; Tostado-Véliz, M.; Arevalo, P. Decarbonization of the Galapagos Islands. Proposal to transform the energy system into 100% renewable by 2050. *Renew. Energy* **2022**, *189*, 199–220. [\[CrossRef\]](#)
58. Enerji Piyasaları Isletme, A.Ş. 2024. Available online: <https://www.epias.com.tr/> (accessed on 26 October 2024).
59. Anguita, D.; Ghelardoni, L.; Ghio, A.; Oneto, L.; Ridella, S. The ‘K’ in K-fold Cross Validation. In Proceedings of the ESANN, Bruges, Belgium, 25–27 April 2012; Volume 102, pp. 441–446.
60. Kim, J.-H. Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. *Comput. Stat. Data Anal.* **2009**, *53*, 3735–3745. [\[CrossRef\]](#)
61. Prusty, S.; Patnaik, S.; Dash, S.K. SKCV: Stratified K-fold cross-validation on ML classifiers for predicting cervical cancer. *Front. Nanotechnol.* **2022**, *4*, 972421. [\[CrossRef\]](#)
62. Kocaguneli, E.; Menzies, T. Software effort models should be assessed via leave-one-out validation. *J. Syst. Softw.* **2013**, *86*, 1879–1890. [\[CrossRef\]](#)

63. Xu, N.; Lovreglio, R.; Kuligowski, E.D.; Cova, T.J.; Nilsson, D.; Zhao, X. Predicting and assessing wildfire evacuation decision-making using machine learning: Findings from the 2019 kincade fire. *Fire Technol.* **2023**, *59*, 793–825. [[CrossRef](#)]
64. Ahakonye, L.A.C.; Nwakanma, C.I.; Lee, J.-M.; Kim, D.-S. Efficient classification of enciphered SCADA network traffic in smart factory using decision tree algorithm. *IEEE Access* **2021**, *9*, 154892–154901. [[CrossRef](#)]
65. Ren, G.; Wang, Y.; Ning, J.; Zhang, Z. Using near-infrared hyperspectral imaging with multiple decision tree methods to delineate black tea quality. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2020**, *237*, 118407. [[CrossRef](#)] [[PubMed](#)]
66. Charbuty, B.; Abdulazeez, A. Classification based on decision tree algorithm for machine learning. *J. Appl. Sci. Technol. Trends* **2021**, *2*, 20–28. [[CrossRef](#)]
67. Nagaraj, P.; Mansour, R.F.; Almazroa, A. Artificial flora algorithm-based feature selection with gradient boosted tree model for diabetes classification. *Diabetes, Metab. Syndr. Obes. Targets Ther.* **2021**, *14*, 2789–2806.
68. James, G.; Witten, D.; Hastie, T.; Tibshirani, R.; Taylor, J. Tree-based methods. In *An Introduction to Statistical Learning: With Applications in Python*; Springer: Berlin/Heidelberg, Germany, 2023; pp. 331–366.
69. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* **2014**, *7*, 1247–1250. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.