agriculture

IMPACT FACTOR 3.6

CITESCORE 3.6

Article

Machine Learning Approaches for Forecasting the Best Microbial Strains to Alleviate Drought Impact in Agriculture

Tymoteusz Miller, Grzegorz Mikiciuk, Anna Kisiel, Małgorzata Mikiciuk, Dominika Paliwoda, Lidia Sas-Paszt, Danuta Cembrowska-Lech, Adrianna Krzemińska, Agnieszka Kozioł and Adam Brysiewicz

Special Issue
Applications of Data Analysis in Agriculture

Edited by
Dr. Xanthoula Eirini Pantazi

MDPI

https://doi.org/10.3390/agriculture13081622 *agriculture*

MDPI

Article

Machine Learning Approaches for Forecasting the Best Microbial Strains to Alleviate Drought Impact in Agriculture

Tymoteusz Miller [1,2,*], Grzegorz Mikiciuk [3], Anna Kisiel [1,2], Małgorzata Mikiciuk [4], Dominika Paliwoda [3], Lidia Sas-Paszt [5], Danuta Cembrowska-Lech [2,6], Adrianna Krzemińska [2], Agnieszka Kozioł [7] and Adam Brysiewicz [7]

[1] Institute of Marine and Environmental Sciences, University of Szczecin, Wąska 13, 71-415 Szczecin, Poland

[2] Polish Society of Bioinformatics and Data Science BIODATA, Popiełuszki 4c, 71-214 Szczecin, Poland

[3] Department of Horticulture, Faculty of Environmental Management and Agriculture, West Pomeranian University of Technology in Szczecin, Słowackiego 17, 71-434 Szczecin, Poland; dominika.paliwoda@zut.edu.pl (D.P.)

[4] Department of Bioengineering, Faculty of Environmental Management and Agriculture, West Pomeranian University of Technology, Słowackiego 17, 71-434 Szczecin, Poland

[5] Department of Microbiology and Rhizosphere, The National Institute of Horticultural Research, Konstytucji 3 Maja 1/3, 96-100 Skierniewice, Poland

[6] Department of Physiology and Biochemistry, Institute of Biology, University of Szczecin, Felczaka 3c, 71-412 Szczecin, Poland

[7] Institute of Technology and Life Sciences–National Research Institute, Falenty, Hrabska Avenue 3, 05-090 Raszyn, Poland

* Correspondence: tymoteusz.miller@usz.edu.pl

Abstract: Drought conditions pose significant challenges to sustainable agriculture and food security. Identifying microbial strains that can mitigate drought effects is crucial to enhance crop resilience and productivity. This study presents a comprehensive comparison of several machine learning models, including Random Forest, Decision Tree, XGBoost, Support Vector Machine (SVM), and Artificial Neural Network (ANN), to predict optimal microbial strains for this purpose. Models were assessed on multiple metrics, such as accuracy, standard deviation of results, gains, total computation time, and training time per 1000 rows of data. Notably, the Gradient Boosted Trees model outperformed others in accuracy but required extensive computational resources. This underscores the balance between accuracy and computational efficiency in machine learning applications. Leveraging machine learning for selecting microbial strains signifies a leap beyond traditional methods, offering improved efficiency and efficacy. These insights hold profound implications for agriculture, especially concerning drought mitigation, thus furthering the cause of sustainable agriculture and ensuring food security.

Keywords: machine learning; predictive analytics; soil microbiome; climate resilience; crop yield enhancement; SVM; ANN; data-driven agriculture; sustainable farming practices; crop stress management; agricultural biotechnology; artificial intelligence

check for updates

Citation: Miller, T.; Mikiciuk, G.; Kisiel, A.; Mikiciuk, M.; Paliwoda, D.; Sas-Paszt, L.; Cembrowska-Lech, D.; Krzemińska, A.; Kozioł, A.; Brysiewicz, A. Machine Learning Approaches for Forecasting the Best Microbial Strains to Alleviate Drought Impact in Agriculture. *Agriculture* **2023**, *13*, 1622. https://doi.org/10.3390/agriculture13081622

Academic Editor: Xanthoula Eirini Pantazi

Received: 19 July 2023
Revised: 12 August 2023
Accepted: 16 August 2023
Published: 17 August 2023

1. Introduction

Drought is a major abiotic stress factor that significantly impacts agricultural productivity worldwide. It affects the growth and yield of crops, posing a serious threat to food security. With climate change, the frequency and severity of drought conditions are expected to increase, making it a pressing issue that needs to be addressed [1–3].

One promising approach to mitigate the detrimental effects of drought on crops involves the use of microbial strains, specifically plant-growth-promoting rhizobacteria (PGPR). These beneficial bacteria colonize the rhizosphere—the region of soil in the vicinity of plant roots—and promote plant growth through various mechanisms. They can enhance water uptake by improving root system architecture, produce plant growth hormones that stimulate growth and development, and increase nutrient availability by solubilizing soil nutrients, thereby helping plants withstand drought conditions [4,5].

In addition to these direct benefits to the plants, PGPR also plays a crucial role in maintaining soil health. They contribute to the formation of soil aggregates, which improves soil structure and water-holding capacity. This is particularly important in drought conditions, where water availability in the soil is limited [6–8].

Recent research has also highlighted the potential role of PGPR in reducing greenhouse gas emissions under different soil moisture conditions. Certain strains of PGPR can reduce the emission of nitrous oxide, a potent greenhouse gas, from the soil. This not only helps in mitigating climate change but also improves the efficiency of nitrogen use by the plants, which is often reduced under drought conditions [9–11].

Thus, the selection of appropriate microbial strains is of paramount importance in sustainable agriculture, particularly in the face of increasing drought events due to climate change. By harnessing the power of these beneficial microbes, we can develop more resilient agricultural systems that can thrive even under adverse environmental conditions [12–14].

Given the vast diversity of microbial strains and the complexity of plant–microbe–soil interactions, selecting the most effective strains for specific crops and environmental conditions is a challenging task. Each microbial strain has a unique set of traits and capabilities, and their effectiveness can vary depending on the specific crop, soil type, and environmental conditions. Furthermore, the interactions between different microbial strains, as well as their interactions with the plants and the soil, add another layer of complexity to this task [15–17].

This is where predictive models come into play. These models, powered by advanced machine learning algorithms, can analyze large datasets of microbial traits, environmental factors, and plant responses to predict which strains would be most beneficial under certain conditions. They can handle the high dimensionality and complexity of the data, uncover hidden patterns and relationships, and make accurate predictions even with incomplete or noisy data [18–20].

For example, a predictive model could analyze data on microbial traits such as nitrogen fixation ability, phosphate solubilization, production of plant growth hormones, and resistance to environmental stresses, along with data on soil properties and climatic conditions, to predict the effectiveness of different microbial strains in promoting plant growth under drought conditions [21–23].

Such models can greatly enhance the efficiency and effectiveness of microbial strain selection. Instead of relying on trial-and-error or time-consuming laboratory tests, researchers and farmers can use these models to make informed decisions about which microbial strains to use. This can lead to improved crop resilience and productivity under drought conditions, and ultimately contribute to the sustainability and resilience of our agricultural systems [24–26].

Moreover, these predictive models can also facilitate the discovery of new beneficial microbial strains and the design of synthetic microbial communities tailored to specific crops and environments. They can also provide valuable insights into the underlying mechanisms of plant–microbe–soil interactions, advancing our understanding of this important aspect of agroecology [27–29].

The aim of this study is to present a comprehensive comparison of several machine learning models in predicting the optimal microbial strains for mitigating drought effects in agriculture. This research seeks to highlight the potential trade-off between accuracy and computational efficiency in machine learning applications, and underscore the significant advancement that these models represent over traditional methods in the selection of microbial strains. Through this study, we hope to enhance the efficiency and effectiveness of the selection process of microbial strains and contribute to sustainable agriculture and food security, especially in the context of drought mitigation.

2. Materials and Methods

2.1. Description of the Data on Microbial Strains and Drought Conditions

The data used in this study were derived from two previous studies on the effects of rhizosphere bacteria on strawberry plants under water deficit and the use of plant-growth-promoting rhizobacteria to reduce greenhouse gasses in strawberry cultivation under different soil moisture conditions [30,31].

In these studies, a variety of microbial strains were used, including *Bacillus* sp., *Pantoea* sp., *Azotobacter* sp., and *Pseudomonas* sp. These strains were selected for their plant-growth-promoting traits, which were confirmed under conditions of water deficit. The strains were inoculated into the growth substrate near the root system of the plants, with a minimum bacterial density of 10^7 CFU/g.

The studies also varied the moisture content of the growth substrate as a second experimental factor. The water potential was maintained at −10 to −15 kPa under control conditions (optimal soil moisture), and at −40 to −45 kPa under conditions of water deficit in the substrate (Table 1). The substrate moisture levels were varied from 6 weeks after inoculation. The varying soil moisture conditions, specifically the water deficit, served as a crucial factor in the model as it simulates drought conditions. Our aim was to understand how different microbial strains perform under these varying drought scenarios.

Table 1. The level of varied moisture of the growth substrate.

Soil Moisture Condition	Water Potential (kPa)	Description
Control (Optimal Soil Moisture)	−10 to −15	The water potential was maintained at this level under control conditions.
Water Deficit	−40 to −45	The water potential was maintained at this level under conditions of water deficit in the substrate.

2.2. Data Collection and Measurement Methods

The studies employed a range of methods to assess the effects of the microbial strains and drought conditions on the plants. These included the following [30,31]:

- Bioassays for Plant-Growth-Promoting Traits: The production of plant growth hormones, siderophore production, and phosphate solubilization were detected using various bioassays. ACC deaminase activity, which is associated with the ability of bacteria to alleviate plant stress, was also measured.
- Bacterial Counts in Substrate: The bacterial populations in the substrate were assessed using the dilution plate method. This involved taking substrate samples from each pot and plating them on Tryptone Soya Agar. The colony-forming units (CFU) were then counted after incubation.
- Chlorophyll "a" Fluorescence: The health and stress level of the plants were assessed by measuring the parameters of chlorophyll fluorescence using a spectrofluorometer. This provided information on the efficiency of photosystem II, which can be affected by drought stress.
- Greenhouse Gas Emission Measurements: The emissions of NH_3, CO_2, N_2O, and CH_4 were measured using a field photoacoustic gas meter connected to a static chamber. This allowed for the assessment of the impact of the microbial strains and drought conditions on greenhouse gas emissions from the soil surface.

These data collection and measurement methods provided a comprehensive dataset on the effects of different microbial strains and drought conditions on strawberry plants. We utilized a dataset of 1500 data points, of which 70% (1050 data points) were used for training and 30% (450 data points) for testing and validation of our machine learning models. This dataset forms the basis for our machine learning analysis in the current study.

2.3. Explanation of the Machine Learning Techniques Used

In this study, we employed a range of machine learning techniques to analyze the data and make predictions about the optimal microbial strains for mitigating drought effects. These techniques included Naive Bayes, Generalized Linear Model (GLM), Logistic Regression, Fast Large Margin, Deep Learning, Decision Tree, Random Forest, and Gradient Boosted Trees.

The independent variables in our model included the type of microbial strain, moisture content, chlorophyll "a" fluorescence, greenhouse gas emissions, and bacterial counts in the substrate. Our dependent variable was the health and stress level of the strawberry plants, as indicated by factors such as the efficiency of photosystem II.

Naive Bayes is a simple probabilistic classifier based on applying Bayes' theorem with strong (naive) independence assumptions between the features. Despite its simplicity, Naive Bayes can be surprisingly accurate, particularly for large datasets [32,33].

Generalized Linear Models (GLMs) extend the ordinary linear regression model to allow for response variables that have error distribution models other than a normal distribution. They are flexible in handling different types of data and are widely used in statistical modeling and machine learning [34,35].

Logistic Regression is a statistical model that uses a logistic function to model a binary dependent variable. In machine learning, logistic regression is a popular algorithm for classification problems [36,37].

Fast Large Margin is a machine learning method that aims to maximize the margin between the decision boundary and the closest data points from each class. This can lead to more robust models that generalize better to unseen data [38,39].

Deep Learning is a subset of machine learning that involves algorithms inspired by the structure and function of the brain called artificial neural networks. Deep learning models are capable of learning from large, complex datasets [40,41].

Decision Trees are a type of model that makes decisions based on a series of questions, each relating to an attribute or feature of the data. They are simple to understand and interpret, and can handle both numerical and categorical data [42,43].

Random Forest is an ensemble learning method that constructs multiple decision trees and aggregates their outputs to make a final prediction. It is known for its robustness, ability to handle large datasets with high dimensionality, and resistance to overfitting [44,45].

Gradient Boosted Trees is a powerful machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. It builds the model in a stage-wise fashion and generalizes them by allowing optimization of an arbitrary differentiable loss function [46,47].

Each of these machine learning techniques has its strengths and is suited to different types of data and prediction tasks. By using a variety of techniques, we can ensure that our analysis is robust and that we can capture different aspects of the data.

2.4. Details of the Experimental Design and Data Analysis

The experimental design of this study involved a comprehensive comparison of several machine learning models, namely Random Forest, Decision Tree, XGBoost, Support Vector Machine (SVM), and Artificial Neural Network (ANN). These models were chosen due to their diverse strengths and applicability to different types of data and prediction tasks (Figure 1).

The data used for training and evaluating the models were derived from two previous studies on the effects of different microbial strains on strawberry plants under water deficit conditions. The data included information on various microbial strains, their plant-growth-promoting traits, and the response of strawberry plants to these strains under different soil moisture conditions.

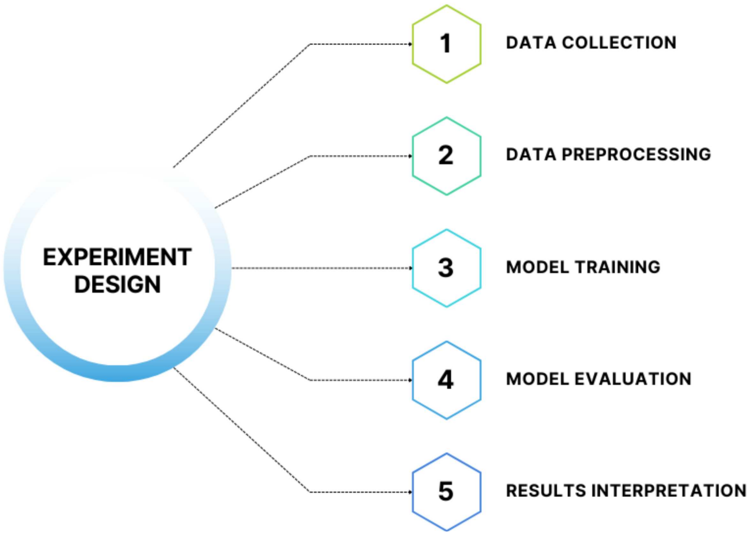

Figure 1. Experiment design schema.

Each machine learning model was trained on a subset of the data and then evaluated on a separate test set.

The analysis of the results involved a detailed comparison of the performance of the different models.

The experimental design and data analysis of this study were carefully planned and executed to ensure a comprehensive and reliable comparison of the machine learning models.

2.5. Model Training Specifics

1. Naive Bayes:
   Algorithm Type: Probabilistic.
   Training Approach: Applied Bayes theorem with an assumption of independence among predictors. The model was trained using a maximum likelihood estimation method.
   Hyperparameters: Default priors were used, with no hyperparameter tuning applied.
2. Generalized Linear Model (GLM):
   Algorithm Type: Regression.
   Training Approach: The model used a link function to relate the linear combination of the input variables to the mean of the output variable. Iteratively reweighted least squares were employed for model optimization.
   Hyperparameters: Standard exponential family distributions (e.g., Gaussian, Binomial) were used.
3. Logistic Regression:
   Algorithm Type: Classification.
   Training Approach: The model employed a logistic function to model the binary dependent variable.
   Hyperparameters: L2 regularization was utilized with a default regularization strength.
4. Fast Large Margin:
   Algorithm Type: Classification.
   Training Approach: Used a margin-based classification method that aims to find the hyperplane which has the largest distance to the nearest training data of any class.
   Hyperparameters: Margin constraints were set with default values, with no hyperparameter tuning applied.
5. Deep Learning:
   Algorithm Type: Neural network.

Training Approach: Employed a feedforward deep neural network with backpropagation for optimization. Used ReLU activation functions for hidden layers and softmax for the output layer.
Hyperparameters: Learning rate was set to 0.001, batch size was 32, and the model was trained for 50 epochs.

6. Decision Tree:
Algorithm Type: Decisional.
Training Approach: Utilized a top–down, recursive, divide-and-conquer approach. The Gini impurity was the criterion for splitting.
Hyperparameters: Maximum depth was set to 5, and a minimum of 10 samples were required to split an internal node.
7. Random Forest:
Algorithm Type: Ensemble.
Training Approach: This model trained multiple decision trees during learning and used averaging to improve the predictive accuracy and control overfitting.
Hyperparameters: Number of trees was set to 100, with a maximum depth of 5.
8. Gradient Boosted Trees:
Algorithm Type: Ensemble.
Training Approach: Built trees one at a time, where each new tree tried to correct errors of the preceding one. Used a gradient descent algorithm to minimize the loss.
Hyperparameters: Learning rate was set to 0.1, 100 boosting stages were performed, and a maximum depth of 3 was set for individual trees.

3. Results

In this empirical study, we evaluated several machine learning models, namely Naive Bayes, Generalized Linear Model (GLM), Logistic Regression, Fast Large Margin, Deep Learning, Decision Tree, Random Forest, and Gradient Boosted Trees. The models were compared based on accuracy, standard deviation of their results, gains, total time, and training time for 1000 rows of data (Table 2, Figure 2). Our machine learning analysis identified several microbial strains as being optimal for mitigating the effects of drought on strawberry plants. Notably, *Azotobacter* sp. AJ 1.2; *Pantoea* sp. DKB64, DKB63, and DKB68; and Pseudomonas sp. strain PJ 1.1 emerged as the most beneficial, consistent with previous studies.

Table 2. ML model comparison.

Model	Accuracy	Standard Deviation	Gains	Total Time	Training Time (1000 Rows)	R^2	MAE
Naive Bayes	67%	3%	28.0	556,941.0	2014.9	0.65	5.4
Generalized Linear Model	69%	4%	42.0	543,840.0	5125.0	0.68	5.2
Logistic Regression	67%	6%	46.0	904,840.0	6985.1	0.66	5.3
Fast Large Margin	62%	3%	18.0	880,130.0	4145.8	0.60	5.9
Deep Learning	80%	6%	40.0	915,363.0	6279.8	0.97	4.0
Decision Tree	73%	8%	18.0	682,198.0	5717.3	0.72	4.7
Random Forest	71%	7%	20.0	953,485.0	5300.6	0.70	4.8
Gradient Boosted Trees	87%	4%	68.0	3,381,260.0	22,101.2	0.89	3.2

The Naive Bayes classifier achieved an accuracy of 67%, with a relatively low standard deviation of 3%, suggesting a consistent performance across different test sets. Despite its mediocre gains (28.0), the model demonstrated a reasonably efficient computation time with a total time of 556,941.0 and a training time of 2014.9 per 1000 rows.

The Generalized Linear Model (GLM), on the other hand, displayed a slightly higher accuracy of 69% with a standard deviation of 4%. It managed to attain higher gains (42.0) compared to Naive Bayes, while maintaining a slightly lower total computation time of 543,840.0. However, the model required a more extended training time per 1000 rows (5125.0).

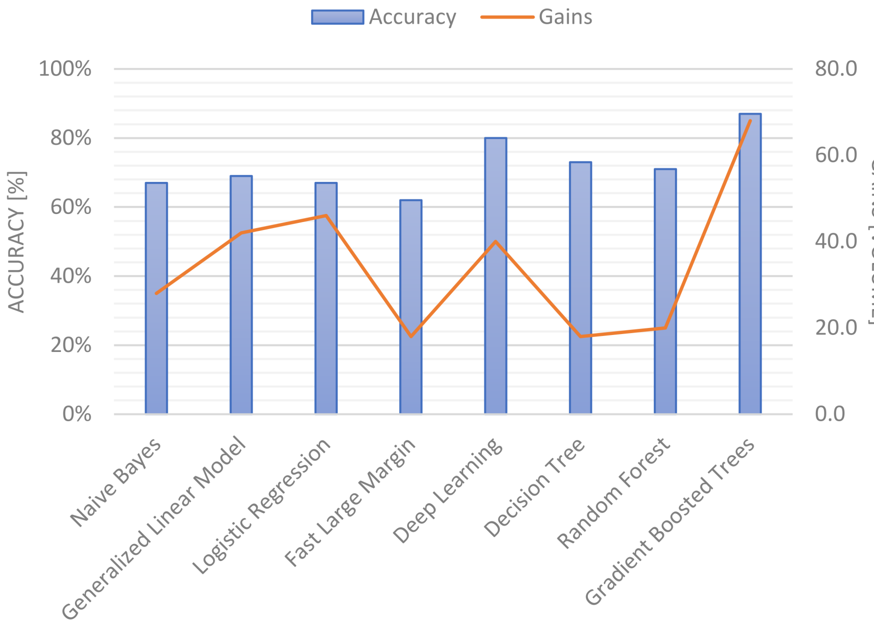

Figure 2. Accuracy and gains of each tested ML model.

Our Logistic Regression model exhibited an accuracy level on par with Naive Bayes (67%). However, the standard deviation of this model was slightly higher (6%), indicating less consistent performance. Its gains (46.0) surpassed both Naive Bayes and GLM, but at the cost of higher computation time, both overall (904,840.0) and per 1000 rows during training (6985.1).

The Fast Large Margin model reported the lowest accuracy among the models (62%) with a standard deviation of 3%. Despite the low gains of 18.0, it outperformed GLM and Logistic Regression in terms of computation time, with a total time of 880,130.0 and a training time of 4145.8 per 1000 rows.

In terms of accuracy, the Deep Learning model emerged as a strong contender, achieving an accuracy of 80% with a standard deviation of 6%. It registered respectable gains (40.0) and exhibited a reasonable computation time, with a total time of 915,363.0 and a training time of 6279.8 per 1000 rows.

The Decision Tree model registered an accuracy of 73%, with a higher standard deviation of 8%, suggesting somewhat inconsistent results. It mirrored the Fast Large Margin model's gains (18.0), but with a total computation time of 682,198.0 and a training time of 5717.3 per 1000 rows.

The Random Forest model demonstrated an accuracy of 71% with a standard deviation of 7%. It provided gains of 20.0, and showcased a total computation time of 953,485.0 and a training time of 5300.6 per 1000 rows.

Finally, the Gradient Boosted Trees model emerged as the most accurate model with an accuracy of 87% and a standard deviation of 4%. It demonstrated the highest gains (68.0), albeit at a significant computation cost, with the highest total computation time of 3,381,260.0 and a substantially long training time of 22,101.2 per 1000 rows.

The Gradient Boosted Trees model achieved the highest accuracy and gains, suggesting it is the most effective model for this specific task, considering only model performance. However, its computation cost is substantially high compared to the other models, which is a crucial consideration in a real-world application scenario. Further experiments could involve tuning hyperparameters or exploring ensemble methods.

4. Discussion

In this study, we compared the performance of eight machine learning models—Naive Bayes, Generalized Linear Model (GLM), Logistic Regression, Fast Large Margin, Deep Learning, Decision Tree, Random Forest, and Gradient Boosted Trees—in predicting the optimal microbial strains for mitigating drought effects in agriculture. The models were evaluated based on their accuracy, standard deviation of results, gains, total computation time, and training time per 1000 rows of data. While our study focuses on strawberry plants, the microbial strains identified might interact differently with other plant species. Further research is required to determine the applicability of these results to other crops or vegetation.

The primary objective of this study was to assess the performance of various machine learning models in predicting optimal microbial strains for drought resistance in agriculture. The evaluation criteria were based on several metrics, such as accuracy, standard deviation of results, gains, total computation time, and training time per 1000 rows of data.

Quick Comparison of Models:

1. Top Performer in Accuracy: The Gradient Boosted Trees stood out with the highest accuracy of 87%, followed closely by the Deep Learning model at 80%.
2. Computational Efficiency: In terms of total computation time, the Generalized Linear Model was the most efficient, taking only 543,840 units of time, whereas the Gradient Boosted Trees required a considerably higher time, clocking in at 3,381,260 units, emphasizing a significant trade-off between accuracy and computational efficiency.
3. Consistency: When looking at the standard deviation, which indicates the consistency of the model results, most models maintained a deviation within the 3–8% range. The Gradient Boosted Trees, despite its high accuracy, exhibited consistency with a standard deviation of just 4%.
4. Training Efficiency: In terms of training time for 1000 rows, the Naive Bayes algorithm was the quickest, with a time of 2014.9 units. This contrasts sharply with the Gradient Boosted Trees model, which took 22,101.2 units, indicating that while Gradient Boosted Trees are accurate, they require significantly more time to train.
5. Gains: The Gradient Boosted Trees model also topped the gains metric at 68.0, with the Logistic Regression model following at 46.0. This suggests that the Gradient Boosted Trees not only offers high accuracy but also maximizes the true positive rate.

While each model has its strengths and areas for improvement, it is evident that the Gradient Boosted Trees model emerges as a strong contender in multiple areas, particularly accuracy and gains. However, it demands significant computational resources. On the other hand, models like the Generalized Linear Model offer a balance between accuracy and efficiency. Therefore, the selection of a model should factor in both the performance metrics and the computational resources available. The results reiterate the importance of understanding the specific needs of a project before selecting an appropriate machine learning model.

The Gradient Boosted Trees model emerged as the most accurate, achieving an accuracy of 87% with a standard deviation of 4%. It also demonstrated the highest gains (68.0), suggesting that it was the most effective model for this specific task. However, it also had the highest total computation time and training time per 1000 rows, indicating a significant computational cost. This suggests that while Gradient Boosted Trees may be the most accurate model, it may not be the most efficient choice for real-world applications where computational resources and time are constraints (Table 3).

The Deep Learning model also showed strong performance, achieving an accuracy of 80% with a standard deviation of 6%. Despite its respectable gains (40.0) and reasonable computation time, it did not outperform the Gradient Boosted Trees model in terms of accuracy. This suggests that while Deep Learning models can handle complex, high-dimensional data and make accurate predictions, they may not always be the best choice for every task (Table 4).

Table 3. Selected hyperparameters for XGBoost model.

Number of Trees	Maximal Depth	Learning Rate	Accuracy
30.0	4.0	0.1	0.8732
30.0	7.0	0.1	0.8683
90.0	4.0	0.01	0.8573
150.0	7.0	0.01	0.8573
150.0	4.0	0.01	0.8478
90.0	7.0	0.01	0.6785
30.0	2.0	0.1	0.6785
150.0	2.0	0.01	0.6836
150.0	4.0	0.001	0.6938
90.0	4.0	0.001	0.6989
30.0	2.0	0.01	0.6989
30.0	4.0	0.01	0.7040
30.0	7.0	0.01	0.7040

Table 4. Deep Learning model metrics.

Metric	Value
Model Metrics Type	Multinomial
Description	Metrics reported on full training frame
Model ID	rm-h2o-model-model-61089
Frame ID	rm-h2o-frame-model-61089
RMSE	0.8007551905
R^2	0.9719184
Logloss	1.2460235
Mean Per Class Error	0.36813188

The Naive Bayes, Generalized Linear Model (GLM), and Logistic Regression models achieved similar accuracy levels (67–69%), but with varying standard deviations, gains, and computation times. These models are simpler than Gradient Boosted Trees and Deep Learning, and may be more suitable for tasks with smaller datasets or fewer features.

The Fast Large Margin model reported the lowest accuracy among the models (62%), suggesting that it may not be the best choice for this specific task. However, it had a relatively efficient computation time, indicating that it may be a suitable choice for tasks where speed is a priority.

The Decision Tree and Random Forest models achieved moderate accuracy levels (73% and 71%, respectively), but with higher standard deviations, suggesting somewhat inconsistent results. These models are known for their interpretability and robustness, and may be more suitable for tasks where these qualities are important.

4.1. Comparative Analysis of Machine Learning Models

The comparative analysis of the machine learning models in this study provided a comprehensive understanding of their performance, strengths, and limitations in the context of predicting optimal microbial strains for mitigating drought effects. Each model was evaluated based on several metrics, including accuracy, standard deviation of results, gains, total computation time, and training time per 1000 rows of data. This multi-faceted evaluation allowed us to assess not only the predictive power of the models but also their computational efficiency and consistency of performance.

The Gradient Boosted Trees model emerged as the most accurate, achieving an impressive accuracy of 87%. This model, which combines the predictions of multiple weak learners to improve accuracy, demonstrated its effectiveness in handling the complexity of the task at hand. However, this high accuracy came with a trade-off in terms of computational resources. The Gradient Boosted Trees model had the highest total computation time and training time per 1000 rows, indicating a significant computational cost. This underscores a common challenge in machine learning applications: the balance between achieving high accuracy and maintaining computational efficiency.

While the pursuit of high accuracy is a primary goal in machine learning, it is crucial to consider the computational resources required, especially in real-world applications where time and computational power may be limited. Therefore, the selection of a machine learning model should not be based solely on its accuracy but should also take into account its computational efficiency.

Moreover, the performance of machine learning models can vary depending on the specific characteristics of the data and the task. Therefore, it is advisable to compare multiple models to identify the one that performs best in a given context. This comparative analysis approach adopted in our study provides a robust framework for selecting the most suitable model for specific tasks in the field of microbial strain prediction for drought mitigation.

4.2. Implications for Microbial Strain Selection in Agriculture

The findings of this study have far-reaching implications for the field of agriculture, particularly in the context of microbial strain selection for drought mitigation. The use of machine learning models, as demonstrated in our study, introduces a novel and efficient approach to predicting the optimal microbial strains under specific environmental conditions. This could revolutionize the current practices in microbial strain selection, which often involve laborious and time-consuming experimental procedures.

By leveraging the power of machine learning, we can analyze large and complex datasets of microbial traits, environmental factors, and plant responses, and make accurate predictions about the most beneficial strains under certain conditions. This could significantly enhance the efficiency of the selection process, enabling us to quickly identify the strains that are most likely to improve crop resilience and productivity under drought conditions [48–50].

Moreover, the use of machine learning models could also facilitate the customization of microbial strain selection based on specific crop types, soil conditions, and climate factors. This could lead to more targeted and effective agricultural practices, ultimately contributing to sustainable agriculture and food security in the face of climate change [24,51,52].

However, it is important to note that the choice of a machine learning model can significantly impact the accuracy of the predictions and the computational resources required. As our study demonstrated, while some models may offer high accuracy, they may also require substantial computational power and time. Therefore, the selection of a machine learning model should be a careful decision that considers a balance between accuracy, computational efficiency, and the specific requirements of the task [53].

4.3. Implications for the Selection of Microbial Strains for Drought Mitigation

The findings of this study have significant implications for the selection of microbial strains for drought mitigation in agriculture. By leveraging machine learning models, we can potentially enhance the efficiency and effectiveness of this process, leading to improved crop resilience and productivity under drought conditions.

The use of machine learning models allows us to analyze large datasets of microbial traits, environmental factors, and plant responses, and predict which microbial strains would be most beneficial under specific conditions. This represents a significant advancement over traditional methods of microbial strain selection, which often involve time-consuming and labor-intensive experiments [18,54].

Furthermore, the comparison of different machine learning models provides valuable insights into their strengths and weaknesses, guiding the selection of the most suitable model for this task. For instance, while the Gradient Boosted Trees model demonstrated the highest accuracy, it also required substantial computational resources. On the other hand, models like Naive Bayes and Logistic Regression offered a more balanced performance in terms of accuracy and computational efficiency [55].

These findings underscore the importance of considering multiple factors when selecting a machine learning model for microbial strain selection. While accuracy is a crucial factor, computational efficiency and consistency of performance are also important considerations, especially in real-world applications where computational resources may be limited [56].

In conclusion, the application of machine learning models for the selection of microbial strains holds great promise for enhancing drought mitigation in agriculture. By selecting the most suitable model and effectively leveraging its capabilities, we can potentially improve the resilience and productivity of crops under drought conditions, contributing to sustainable agriculture and food security.

Discussion on Model Selection Based on Scenarios

When selecting a machine learning model for practical applications, it is essential to strike a balance between accuracy, computational efficiency, and the specific needs of the task at hand. Each algorithm has its strengths and weaknesses, and understanding them can aid in informed decision making.

1. High Priority on Accuracy with Adequate Resources:
   Recommended Model: Gradient Boosted Trees.
   Reasoning: Achieving an accuracy of 87% and having a reasonable standard deviation of 4%, the Gradient Boosted Trees model stands out as the top performer. However, it also demands the highest computational resources, with a total time of 3,381,260 and a substantial training time per 1000 rows of data.
2. Need for Quick Results with Moderate Accuracy:
   Recommended Model: Generalized Linear Model or Naive Bayes.
   Reasoning: Both these models offer decent accuracy, with the GLM slightly edging out at 69%. Their total computation time is relatively low, making them suitable for applications where quick insights are essential.
3. Scenarios Requiring Deep Insights and Nonlinearity:
   Recommended Model: Deep Learning.
   Reasoning: With an accuracy of 80% and the ability to capture intricate patterns and relationships in data, Deep Learning models can be ideal. They are particularly effective when the dataset is large and when nonlinear relationships are suspected.
4. Balancing Accuracy and Computation Time:
   Recommended Model: Decision Tree or Random Forest.
   Reasoning: Both models provide a good compromise between accuracy and computational efficiency. Random Forest, being an ensemble method, can handle more complex data patterns and offers slightly reduced variance compared to a single Decision Tree.
5. Scenarios with Limited Data or Resources:
   Recommended Model: Fast Large Margin.
   Reasoning: With a relatively low computation time and modest accuracy, this model can be effective when computational resources are limited, or when a rapid prototype is required.

While the allure of high accuracy is tempting, it is vital to consider the broader context. Scenarios with limited computational resources, urgent time frames, or specific data characteristics might benefit from models other than the highest accuracy performer. It is crucial to understand the trade-offs involved and align them with practical needs for the effective application of machine learning in agriculture, especially in the context of selecting microbial strains for drought mitigation.

4.4. Future Directions

The results of this study not only provide valuable insights into the performance of various machine learning models in predicting optimal microbial strains for drought mitigation, but also pave the way for numerous exciting avenues for future research (Table 5).

One promising direction is the exploration of ensemble methods. These methods, which involve combining the predictions of multiple models, have been shown to improve accuracy and robustness in many machine learning tasks. By leveraging the strengths of different models, ensemble methods could potentially yield more accurate and reliable predictions in the context of microbial strain selection. Future studies could investigate various ensemble techniques, such as bagging, boosting, and stacking, and assess their performance in comparison to individual models [57,58].

Table 5. Future research steps and directions in machine learning for microbial selection.

Step	Objective	Approach
1. Expanding Data Sources	Diversify and increase the robustness of predictive models.	Incorporate data from various geographical locations, covering different soil types, microbial ecologies, and climatic conditions.
2. Incorporating Genomic Data	Achieve a deeper understanding of microbial strains.	Delve into the genomic data of the microbial strains to identify genetic markers associated with drought resistance.
3. Ensemble Learning and Hybrid Models	Enhance prediction accuracy and model robustness.	Use ensemble methods combining various algorithms or create hybrid models blending traditional statistical methods with machine learning techniques.
4. Real-time Monitoring and Prediction	Facilitate proactive interventions.	Develop a system with IoT devices for real-time monitoring and use machine learning models for impending drought stress predictions.
5. Collaboration with Microbiologists	Ensure the biological viability of machine learning predictions.	Form interdisciplinary teams with microbiologists and soil scientists to validate the biological viability of machine learning recommendations.
6. Model Explainability and Interpretation	Make machine learning models more transparent and understandable.	Implement techniques from Explainable AI (XAI) for insights into microbial strain selections.
7. Field Trials and Validation	Empirically validate the efficacy of selected microbial strains.	Conduct controlled field trials monitoring plant health, yield, and drought resilience to validate machine learning recommendations.

Another interesting avenue for future research is the investigation of more advanced machine learning techniques, such as Deep Learning. Deep Learning models, which are capable of learning complex patterns from high-dimensional data, have shown great promise in a wide range of applications. In the context of microbial strain selection, these models could potentially uncover intricate relationships between microbial traits, environmental factors, and plant responses, leading to more accurate predictions [59,60].

In addition to exploring new machine learning techniques, future research could also involve applying the models to different types of data. For instance, genomic data or soil microbiome data could provide a wealth of information about the characteristics of different microbial strains and their interactions with plants and the environment. By integrating such data into the machine learning models, we could gain a deeper understanding of the factors that influence the effectiveness of different microbial strains in mitigating drought effects [18,22].

Moreover, future studies could also investigate the applicability of the machine learning models in different crops, soil types, and climatic conditions. This could lead to the development of more versatile and robust models that can cater to a wide range of agricultural scenarios [61,62].

4.5. Limitations

While this study offers valuable insights into the application of machine learning models for predicting optimal microbial strains for drought mitigation, it is important to acknowledge its limitations to fully appreciate the context and scope of the findings.

Firstly, the models were trained and evaluated on data derived from two specific studies. While these studies provided a robust dataset for the task at hand, the performance of the models may vary when applied to different datasets. This is a common limitation in machine learning applications, as the models' performance is often highly dependent on the specific characteristics of the training data. Therefore, the results of this study should be interpreted with caution when generalizing to other datasets or contexts [63–65].

While machine learning models can make accurate predictions, they do not inherently provide mechanistic insights into the underlying biological processes. This is a fundamental limitation of machine learning, as the models are primarily data-driven and do not incorporate explicit biological knowledge. Therefore, the predictions made by the models should be interpreted in conjunction with biological knowledge and experimental validation. It is crucial to validate the predictions in experimental settings to confirm their biological relevance and applicability [64,66,67].

While this study provides valuable insights into the application of machine learning models for microbial strain selection, the limitations should be considered when interpreting the results and planning future research. Despite these limitations, the study represents a significant step forward in the integration of machine learning in agriculture, and sets the stage for further advancements in this exciting field.

5. Conclusions

In the quest to combat drought in agriculture, this research delved into the prowess of various machine learning models, aiming to predict efficacious microbial strains. Our assessment criteria spanned a myriad of metrics, such as accuracy, gains, computational duration, and training times. A salient discovery was that Gradient Boosted Trees outperformed other models in terms of accuracy, albeit demanding significant computational power. This emphasizes the perpetual conundrum of balancing precision with computational viability in the realm of machine learning.

The incorporation of machine learning to discern microbial strains heralds a transformative shift from conventional methodologies, potentially elevating the efficacy of this endeavor. Nevertheless, while zeroing in on a model, it is pivotal to weigh its predictive precision against the computational overhead it demands.

The revelations of this research carry profound ramifications for agriculture, especially concerning drought alleviation. Harnessing machine learning can arguably fortify the hardiness and yield of crops amidst droughts, fortifying our steps toward a sustainable agricultural landscape and ensuring food security.

The horizon of future inquiries might be illuminated by ensemble strategies or deep learning nuances. Venturing into diverse data streams, such as genomic sequences or the intricacies of the soil microbiome, might unravel deeper layers of understanding about the pivotal factors shaping the prowess of microbial strains against drought.

To encapsulate, this endeavor accentuates machine learning's transformative potential in agriculture. It simultaneously heralds a call for judicious model choices, harmonizing multiple determinants. The insights gleaned lay the bedrock for future strides in harnessing microbial strains for drought mitigation, painting a promising picture for the agriculture of tomorrow.

Author Contributions: Conceptualization, T.M., D.P., G.M. and M.M.; methodology, D.P., G.M., M.M., A.K. (Anna Kisiel), L.S.-P., T.M. and A.K. (Adrianna Krzemińska); validation, D.P., G.M., M.M., A.K. (Agnieszka Kozioł) and T.M.; formal analysis, D.P., G.M., M.M., A.K. and T.M.; investigation, D.P., G.M., M.M., A.K., D.C.-L. and A.K. (Adrianna Krzemińska); data curation, D.P., G.M., T.M., M.M., A.K., L.S.-P. and A.K.; writing—original draft preparation, D.P., G.M., T.M., M.M., A.K., L.S.-P. and A.K.; writing—review and editing, D.P., G.M., M.M., T.M., A.K., L.S.-P., A.K. (Agnieszka Kozioł) and A.B.; visualization, D.P. and T.M.; supervision, G.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are stored on the PTBD BIODATA server and in the public repository: https://github.com/PTBDBIODATA/Databases (accessed on 1 July 2023).

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

References

1. Liu, X.; Zhu, X.; Pan, Y.; Li, S.; Liu, Y.; Ma, Y. Agricultural Drought Monitoring: Progress, Challenges, and Prospects. *J. Geogr. Sci.* **2016**, *26*, 750–767. [CrossRef]
2. Verner, D.; Treguer, D.; Redwood, J.; Christensen, J.; McDonnell, R.; Elbert, C.; Konishi, Y.; Belghazi, S. Climate Variability, Drought, and Drought Management in Morocco's Agricultural Sector. 2018. Available online: http://hdl.handle.net/10986/30603 (accessed on 15 March 2022).
3. Faisol, A.; Indarto, I.; Novita, E.; Budiyono, B. Assessment of Agricultural Drought Based on CHIRPS Data and SPI Method over West Papua—Indonesia. *J. Water Land Dev.* **2022**, *52*, 44–52. [CrossRef]
4. Ahluwalia, O.; Singh, P.C.; Bhatia, R. A Review on Drought Stress in Plants: Implications, Mitigation and the Role of Plant Growth Promoting Rhizobacteria. *Resour. Environ. Sustain.* **2021**, *5*, 100032. [CrossRef]
5. Camaille, M.; Fabre, N.; Clément, C.; Barka, E.A. Advances in Wheat Physiology in Response to Drought and the Role of Plant Growth Promoting Rhizobacteria to Trigger Drought Tolerance. *Microorganisms* **2021**, *9*, 687. [CrossRef] [PubMed]
6. Chaudhary, P.; Parveen, H.; Gangola, S.; Kumar, G.; Bhatt, P.; Chaudhary, A. Plant Growth-Promoting Rhizobacteria and Their Application in Sustainable Crop Production. In *Microbial Technology for Sustainable Environment*; Bhatt, P., Gangola, S., Udayanga, D., Kumar, G., Eds.; Springer: Singapore, 2021; pp. 217–234. [CrossRef]
7. Khan, N.; Bano, A.; Shahid, M.A.; Nasim, W.; Ali Babar, M. Interaction between PGPR and PGR for Water Conservation and Plant Growth Attributes under Drought Condition. *Biologia* **2018**, *73*, 1083–1098. [CrossRef]
8. Zheng, W.; Zeng, S.; Bais, H.; LaManna, J.M.; Hussey, D.S.; Jacobson, D.L.; Jin, Y. Plant Growth-Promoting Rhizobacteria (PGPR) Reduce Evaporation and Increase Soil Water Retention. *Water Resour. Res.* **2018**, *54*, 3673–3687. [CrossRef]
9. Ding, Y.; Li, C.; Li, Z.; Liu, S.; Zou, Y.; Gao, X.; Cai, Y.; Siddique, K.H.M.; Wu, P.; Zhao, X. Greenhouse Gas Emission Responses to Different Soil Amendments on the Loess Plateau, China. *Agric. Ecosyst. Environ.* **2023**, *342*, 108233. [CrossRef]
10. Schillaci, M.; Gupta, S.; Walker, R.; Roessner, U. The Role of Plant Growth-Promoting Bacteria in the Growth of Cereals under Abiotic Stresses. In *Root Biology—Growth, Physiology, and Functions*; Ohyama, T., Ed.; IntechOpen: London, UK, 2019; pp. 1–21. [CrossRef]
11. Seo, Y.; Cho, K.S. Rhizoremdiation of Petroleum Hydrocarbon-Contaminated Soils and Greenhouse Gas Emission Characteristics: A Review. *Microbiol. Biotechnol. Lett.* **2020**, *48*, 99–112. [CrossRef]
12. Mohanty, P.; Singh, P.K.; Chakraborty, D.; Mishra, S.; Pattnaik, R. Insight Into the Role of PGPR in Sustainable Agriculture and Environment. *Front. Sustain. Food Syst.* **2021**, *5*, 667150. [CrossRef]
13. Vocciante, M.; Grifoni, M.; Fusini, D.; Petruzzelli, G.; Franchi, E. The Role of Plant Growth-Promoting Rhizobacteria (PGPR) in Mitigating Plant's Environmental Stresses. *Appl. Sci.* **2022**, *12*, 1231. [CrossRef]
14. Abdelazeem, S.A.E.M.; Al-Werwary, S.M.; Mehana, T.A.E.; El-Hamahmy, M.A.; Kalaji, H.M.; Rastogi, A.; Elsheery, N.I. Use of Plant Growth-Promoting Rhizobacteria Isolates as a Potential Biofertiliser for Wheat. *J. Water Land Dev.* **2022**, 99–111. [CrossRef]
15. Massa, F.; Defez, R.; Bianco, C. Exploitation of Plant Growth Promoting Bacteria for Sustainable Agriculture: Hierarchical Approach to Link Laboratory and Field Experiments. *Microorganisms* **2022**, *10*, 865. [CrossRef] [PubMed]
16. Ruzzi, M.; Aroca, R. Plant Growth-Promoting Rhizobacteria Act as Biostimulants in Horticulture. *Sci. Hortic* **2015**, *196*, 124–134. [CrossRef]
17. Vejan, P.; Khadiran, T.; Abdullah, R.; Ismail, S.; Dadrasnia, A. Encapsulation of Plant Growth Promoting Rhizobacteria—Prospects and Potential in Agricultural Sector: A Review. *J. Plant Nutr.* **2019**, *42*, 2600–2623. [CrossRef]
18. Poncheewin, W.; van Diepeningen, A.D.; van der Lee, T.A.J.; Suarez-Diez, M.; Schaap, P.J. Classification of the Plant-Associated Lifestyle of *Pseudomonas* Strains Using Genome Properties and Machine Learning. *Sci. Rep.* **2022**, *12*, 1–12. [CrossRef] [PubMed]
19. Sambo, P.; Nicoletto, C.; Giro, A.; Pii, Y.; Valentinuzzi, F.; Mimmo, T.; Lugli, P.; Orzes, G.; Mazzetto, F.; Astolfi, S.; et al. Hydroponic Solutions for Soilless Production Systems: Issues and Opportunities in a Smart Agriculture Perspective. *Front. Plant Sci.* **2019**, *10*, 465257. [CrossRef] [PubMed]
20. Shelar, A.; Singh, A.V.; Maharjan, R.S.; Laux, P.; Luch, A.; Gemmati, D.; Tisato, V.; Singh, S.P.; Santilli, M.F.; Shelar, A.; et al. Sustainable Agriculture through Multidisciplinary Seed Nanopriming: Prospects of Opportunities and Challenges. *Cells* **2021**, *10*, 2428. [CrossRef]

21. Higdon, S.M.; Huang, B.C.; Bennett, A.B.; Weimer, B.C. Identification of Nitrogen Fixation Genes in *Lactococcus* Isolated from Maize Using Population Genomics and Machine Learning. *Microorganisms* **2020**, *8*, 2043. [CrossRef]
22. Indumathi, V.; Santhana Megala, S.; Padmapriya, R.; Suganya, M.; Jayanthi, B.; Bca, H. Prediction and Analysis of Plant Growth Promoting Bacteria Using Machine Learning for Millet Crops. *Ann. Rom. Soc. Cell Biol.* **2021**, *25*, 1826–1833.
23. Wu, J.; Zhao, F. Machine Learning: An Effective Technical Method for Future Use in Assessing the Effectiveness of Phosphorus-Dissolving Microbial Agroremediation. *Front. Bioeng. Biotechnol.* **2023**, *11*, 1189166. [CrossRef]
24. Benos, L.; Tagarakis, A.C.; Dolias, G.; Berruto, R.; Kateris, D.; Bochtis, D. Machine Learning in Agriculture: A Comprehensive Updated Review. *Sensors* **2021**, *21*, 3758. [CrossRef] [PubMed]
25. Sharma, A.; Jain, A.; Gupta, P.; Chowdary, V. Machine Learning Applications for Precision Agriculture: A Comprehensive Review. *IEEE Access* **2021**, *9*, 4843–4873. [CrossRef]
26. Storm, H.; Baylis, K.; Heckelei, T. Machine Learning in Agricultural and Applied Economics. *Eur. Rev. Agric. Econ.* **2020**, *47*, 849–892. [CrossRef]
27. Borchert, E.; Hammerschmidt, K.; Hentschel, U.; Deines, P. Enhancing Microbial Pollutant Degradation by Integrating Eco-Evolutionary Principles with Environmental Biotechnology. *Trends Microbiol.* **2021**, *29*, 908–918. [CrossRef]
28. de Souza, R.S.C.; Armanhi, J.S.L.; Arruda, P. From Microbiome to Traits: Designing Synthetic Microbial Communities for Improved Crop Resiliency. *Front. Plant Sci.* **2020**, *11*, 553605. [CrossRef] [PubMed]
29. Vassilev, N.; Malusà, E.; Neri, D.; Xu, X. Editorial: Plant Root Interaction with Associated Microbiomes to Improve Plant Resiliency and Crop Biodiversity. *Front. Plant Sci.* **2021**, *12*, 715676. [CrossRef]
30. Paliwoda, D.; Mikiciuk, G.; Mikiciuk, M.; Kisiel, A.; Sas-Paszt, L.; Miller, T. Effects of Rhizosphere Bacteria on Strawberry Plants (*Fragaria* × *ananassa* Duch.) under Water Deficit. *Int. J. Mol. Sci.* **2022**, *23*, 10449. [CrossRef]
31. Paliwoda, D.; Mikiciuk, G.; Mikiciuk, M.; Miller, T.; Kisiel, A.; Sas-Paszt, L.; Kozioł, A.; Brysiewicz, A. The Use of Plant Growth Promoting Rhizobacteria to Reduce Greenhouse Gases in Strawberry Cultivation under Different Soil Moisture Conditions. *Agronomy* **2023**, *13*, 754. [CrossRef]
32. Berrar, D. Bayes' Theorem and Naive Bayes Classifier. *Encycl. Bioinform. Comput. Biol. ABC Bioinform.* **2018**, *1–3*, 403–412. [CrossRef]
33. Yang, F.J. An Implementation of Naive Bayes Classifier. In Proceedings of the International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NA, USA, 12−14 December 2018; pp. 301–306. [CrossRef]
34. Hastie, T.J.; Pregibon, D. Generalized Linear Models. In *Statistical Models in S*; Hastie, T.J., Ed.; Routledge: New York, NY, USA, 2017; pp. 195–247. [CrossRef]
35. Meng, X.; Wu, S.; Zhu, J. A Unified Bayesian Inference Framework for Generalized Linear Models. *IEEE Signal Process. Lett.* **2017**, *25*, 398–402. [CrossRef]
36. Gasso, G. *Logistic Regression*; INSA Rouen-ASI Departement Laboratory: Saint-Etienne-du-Rouvray, France, 2019; pp. 1–30.
37. Kuha, J.; Mills, C. On Group Comparisons with Logistic Regression Models. *Sociol. Methods Res.* **2018**, *49*, 498–525. [CrossRef]
38. Sokolic, J.; Giryes, R.; Sapiro, G.; Rodrigues, M.R.D. Robust Large Margin Deep Neural Networks. *IEEE Trans. Signal Process.* **2016**, *65*, 4265–4280. [CrossRef]
39. Wang, M.; Liu, Y.; Huang, Z. Large Margin Object Tracking with Circulant Feature Maps. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4021–4029. [CrossRef]
40. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
41. Min, S.; Lee, B.; Yoon, S. Deep Learning in Bioinformatics. *Brief. Bioinform.* **2017**, *18*, 851–869. [CrossRef] [PubMed]
42. Kamiński, B.; Jakubczyk, M.; Szufel, P. A Framework for Sensitivity Analysis of Decision Trees. *Cent. Eur. J. Oper. Res.* **2018**, *26*, 135–159. [CrossRef] [PubMed]
43. Yang, Y.; Morillo, I.G.; Hospedales, T.M. Deep Neural Decision Trees. In Proceedings of the 2018 ICML Workshop on Human Interpretability in Machine Learning (WHI 2018), Stockholm, Sweden, 14 July 2018. [CrossRef]
44. Paul, A.; Mukherjee, D.P.; Das, P.; Gangopadhyay, A.; Chintha, A.R.; Kundu, S. Improved Random Forest for Classification. *IEEE Trans. Image Process.* **2018**, *27*, 4012–4024. [CrossRef] [PubMed]
45. Schonlau, M.; Zou, R.Y. The Random Forest Algorithm for Statistical Learning. *Stata J.* **2020**, *20*, 3–29. [CrossRef]
46. Si, S.; Zhang, H.; Keerthi, S.S.; Mahajan, D.; Dhillon, I.S.; Hsieh, C.-J. Gradient Boosted Decision Trees for High Dimensional Sparse Output. In Proceedings of the 34th International Conference on Machine Learning (ICML), Sydney, Australia, 6−11 August 2017; pp. 3182–3190.
47. Zhang, Z.; Jung, C. GBDT-MO: Gradient-Boosted Decision Trees for Multiple Outputs. *IEEE Trans. Neutral Netw. Learn Syst.* **2021**, *32*, 3156–3167. [CrossRef]
48. Murlidharan, S.; Shukla, V.K.; Chaubey, A. Application of Machine Learning in Precision Agriculture Using IoT. In Proceedings of the 2021 2nd International Conference on Intelligent Engineering and Management (ICIEM), London, UK, 28−30 April 2021; pp. 34–39. [CrossRef]
49. Park, S.J.; Chae, D.K.; Bae, H.K.; Park, S.; Kim, S.W. Reinforcement Learning over Sentiment-Augmented Knowledge Graphs towards Accurate and Explainable Recommendation. In Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining, New York, NY, USA, 21−25 February 2022; pp. 784–793. [CrossRef]
50. Rehman, M.; Razzaq, A.; Baig, I.A.; Jabeen, J.; Tahir, M.H.N.; Ahmed, U.I.; Altaf, A.; Abbas, T. Semantics Analysis of Agricultural Experts' Opinions for Crop Productivity through Machine Learning. *Appl. Artif. Intell.* **2022**, *36*, 1–16. [CrossRef]

51. Chlingaryan, A.; Sukkarieh, S.; Whelan, B. Machine Learning Approaches for Crop Yield Prediction and Nitrogen Status Estimation in Precision Agriculture: A Review. *Comput. Electron. Agric.* **2018**, *151*, 61–69. [CrossRef]
52. Yadav, N.; Alfayeed, S.M.; Wadhawan, A. Machine Learning In Agriculture: Techniques And Applications. *Int. J. Eng. Appl. Sci. Technol.* **2020**, *5*, 118–122. [CrossRef]
53. Bragg, J.; Habli, I. What Is Acceptably Safe for Reinforcement Learning? In *SAFECOMP 2018: Computer Safety, Reliability, and Security*; Gallina, B., Skavhaug, A., Schoitsch, E., Bitsch, F., Eds.; Springer: Cham, Switzerland, 2018; pp. 418–430. [CrossRef]
54. Stocker, M.D.; Pachepsky, Y.A.; Hill, R.L. Prediction of *E. Coli* Concentrations in Agricultural Pond Waters: Application and Comparison of Machine Learning Algorithms. *Front. Artif. Intell.* **2022**, *4*, 768650. [CrossRef] [PubMed]
55. Saleem, M.H.; Potgieter, J.; Arif, K.M. Automation in Agriculture by Machine and Deep Learning Techniques: A Review of Recent Developments. *Precis. Agric.* **2021**, *22*, 2053–2091. [CrossRef]
56. Rastrollo-Guerrero, J.L.; Gómez-Pulido, J.A.; Durán-Domínguez, A. Analyzing and Predicting Students' Performance by Means of Machine Learning: A Review. *Appl. Sci.* **2020**, *10*, 1042. [CrossRef]
57. Leite, D.M.C.; Lopez, J.F.; Brochet, X.; Barreto-Sanz, M.; Que, Y.A.; Resch, G.; Pena-Reyes, C. Exploration of Multiclass and One-Class Learning Methods for Prediction of Phage-Bacteria Interaction at Strain Level. In Proceedings of the 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, 3−6 December 2018; pp. 1818–1825. [CrossRef]
58. Tang, G.; Shi, J.; Wu, W.; Yue, X.; Zhang, W. Sequence-Based Bacterial Small RNAs Prediction Using Ensemble Learning Strategies. *BMC Bioinform.* **2018**, *19*, 13–23. [CrossRef]
59. Durmuş, H.; Güneş, E.O.; Kırcı, M. Disease Detection on the Leaves of the Tomato Plants by Using Deep Learning. In Proceedings of the 2017 6th Int. Conf. Agro-Geoinformatics, Fairfax, VA, USA, 7 August 2017. [CrossRef]
60. Ropelewska, E.; Sabanci, K.; Aslan, M.F. The Changes in Bell Pepper Flesh as a Result of Lacto-Fermentation Evaluated Using Image Features and Machine Learning. *Foods* **2022**, *11*, 2956. [CrossRef] [PubMed]
61. Akhter, R.; Sofi, S.A. Precision Agriculture Using IoT Data Analytics and Machine Learning. *J. King Saud Univ.—Comput. Inf. Sci.* **2022**, *34*, 5602–5618. [CrossRef]
62. Meshram, V.; Patil, K.; Meshram, V.; Hanchate, D.; Ramkteke, S.D. Machine Learning in Agriculture Domain: A State-of-Art Survey. *Artif. Intell. Life Sci.* **2021**, *1*, 100010. [CrossRef]
63. Hashimoto, D.A.; Witkowski, E.; Gao, L.; Meireles, O.; Rosman, G. Artificial Intelligence in Anesthesiology: Current Techniques, Clinical Applications, and Limitations. *Anesthesiology* **2020**, *132*, 379–394. [CrossRef]
64. Peng, G.C.Y.; Alber, M.; Buganza Tepole, A.; Cannon, W.R.; De, S.; Dura-Bernal, S.; Garikipati, K.; Karniadakis, G.; Lytton, W.W.; Perdikaris, P.; et al. Multiscale Modeling Meets Machine Learning: What Can We Learn? *Arch. Comput. Methods Eng.* **2021**, *28*, 1017–1037. [CrossRef]
65. Sagan, V.; Peterson, K.T.; Maimaitijiang, M.; Sidike, P.; Sloan, J.; Greeling, B.A.; Maalouf, S.; Adams, C. Monitoring Inland Water Quality Using Remote Sensing: Potential and Limitations of Spectral Indices, Bio-Optical Simulations, Machine Learning, and Cloud Computing. *Earth-Sci. Rev.* **2020**, *205*, 103187. [CrossRef]
66. Alber, M.; Buganza Tepole, A.; Cannon, W.R.; De, S.; Dura-Bernal, S.; Garikipati, K.; Karniadakis, G.; Lytton, W.W.; Perdikaris, P.; Petzold, L.; et al. Integrating Machine Learning and Multiscale Modeling—Perspectives, Challenges, and Opportunities in the Biological, Biomedical, and Behavioral Sciences. *NPJ Digit. Med.* **2019**, *2*, 1–11. [CrossRef]
67. Uddin, S.; Khan, A.; Hossain, M.E.; Moni, M.A. Comparing Different Supervised Machine Learning Algorithms for Disease Prediction. *BMC Med. Inform. Decis. Mak.* **2019**, *19*, 1–16. [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.