Machine Learning Approaches Applied in Smart Agriculture for the Prediction of Agricultural Yields

Abourabia. Imade, Ounacer. Soumaya, Elghoumari. Mohammed yassine, Azzouazi. Mohamed
Laboratory of Information Technology and Modeling, Hassan II University,
Faculty of Sciences Ben M'sik, Casablanca, Morocco

***Abstract*—Machine learning techniques in smart agriculture for yield prediction involve using algorithms to analyze historical and real-time data to forecast crop yields. These approaches aim to optimize agricultural practices, improve resource efficiency and enhance productivity, this paper reviews the application of machine learning techniques in smart agriculture for predicting agricultural yields. With the advent of data-driven technologies, machine learning algorithms have become instrumental in analyzing vast amounts of agricultural data to forecast crop yields accurately. Various machine learning models such as regression, classification, and ensemble methods have been employed to process historical and real-time data on weather patterns, soil conditions, crop types, and farming practices. These models enable farmers and stakeholders to make informed decisions, optimize resource allocation, and mitigate risks associated with agricultural production. Furthermore, the integration of Internet of Things devices and remote sensing technologies has facilitated data collection and improved the precision of yield predictions, this paper discusses the key machine learning approaches, challenges, and future directions in leveraging data analytics for enhancing agricultural productivity and sustainability in smart farming systems. to ensure stability and tracking. Simulations is carried out to verify the theoretical results, The study found that different machine learning techniques had varying accuracy for predicting agricultural yields. ViT-B16 achieved the highest F1-SCORE (99.40%), followed by ResNet-50 (99.54%) and CNN (97.70%), while RPN algorithms had lower accuracy (91.83%). Correlation analysis showed a strong positive relationship between humidity and soil moisture, favoring crop growth, while production had minimal correlation with temperature and area. The AdaBoost Regressor was the best performer, with the lowest MAE (0.22), MSE (0.1), and RMSE (0.31), and Random Forest showed strong predictive power with an R2 score of 0.89, Seasonal data indicated that autumn had the highest agricultural production, followed by spring, while summer and winter had much lower yields due to weather conditions. Seasonal temperature variations from 1997 to 2014 showed autumn was the warmest (34.43°C), boosting crop production, and winter the coldest (34.31°C), reducing yields. These temperature shifts significantly impacted agricultural productivity, with warm seasons enhancing growth and extreme temperatures in summer and winter limiting it, machine learning techniques in smart agriculture are pivotal for predicting crop yields by leveraging historical and real-time data, thus optimizing practices and resource use while boosting productivity. This involves deploying diverse machine learning models like regression, classification, and ensembles to analyze extensive data on weather, soil, crops, and farming methods. Such models empower stakeholders with insights for informed decisions, efficient resource allocation, and risk mitigation in agricultural operations. The integration of Internet of Things and remote sensing further refines data accuracy, aiding precise yield predictions. Despite advancements, challenges persist, including data quality assurance, model complexity, scalability, and interoperability, driving ongoing research and simulations to validate and improve ML applications for sustainable and productive smart farming systems.**

Keywords*—Machine learning; IOT; artificial intelligence; agricultural yields; smart agriculture; CNN; ViT-B16

I. Introduction

The field of agriculture has been profoundly shaped by technological advancements over the centuries, evolving from simple tools to complex machinery and now to the integration of sophisticated data-driven technologies. The advent of machine learning (ML) in agriculture marks a pivotal shift towards more precise and automated farming practices, known as precision agriculture. This transition began with the mechanization of farms during the Industrial Revolution, followed by the introduction of chemical fertilizers and genetically modified organisms in the 20th century.

This paper explores the various machine learning approaches applied in smart agriculture specifically for the prediction of agricultural yields. It delves into the types of ML models commonly used, the data sources utilized, challenges faced, and the potential impact of these predictive analytics on agricultural sustainability and efficiency.

In the first paragraph, this paper explores the transformative role of machine learning (ML) in smart agriculture, especially in predicting agricultural yields. With advancements in data-driven technologies, ML has revolutionized traditional farming by enhancing precision, resource efficiency, and sustainability. Agriculture evolved from manual labor to mechanized operations during the Industrial Revolution, followed by the introduction of chemical fertilizers and genetically modified organisms. More recently, digital tools such as GPS and remote sensing have laid the groundwork for ML applications in agriculture [1][2][15].

In the second paragraph, the collection of agricultural data has progressed significantly, from manual observations to modern technologies like the Internet of Things (IoT) and remote sensing. These technologies provide real-time, high-resolution data, which sophisticated ML algorithms analyze to predict outcomes like crop yields and optimize farming practices. IoT sensors monitor variables like soil moisture, temperature, and crop health, generating vast datasets that aid in making data-driven decisions [3]. These insights help farmers manage resources efficiently and mitigate risks associated with agricultural production [4].

In the third paragraph, the focus shifts to key ML algorithms used in smart agriculture, including regression, classification techniques, and ensemble models such as Decision Trees [9], Random Forests [10], Linear Regression [11], K-Nearest Neighbors (KNN) [12], XGBoost [13], and AdaBoost [14]. These algorithms have been widely applied for tasks like yield prediction and optimizing resources like water and fertilizer, offering accurate and actionable insights for modern farming [5] [6].

In the fourth paragraph, deep learning has also made significant strides in smart farming, particularly for image-based tasks. Models like Convolutional Neural Networks (CNNs) [17], advanced architectures such as ResNet [18], and support vector networks [15] have proven effective in detecting pests, diagnosing plant diseases, and optimizing greenhouse conditions. By combining satellite data with deep learning techniques [3], there has been a marked improvement in agricultural monitoring and yield prediction [1][16][17].

In the fifth paragraph, despite the advancements, challenges remain. Issues such as ensuring data quality, scaling models for large datasets, and dealing with complex agricultural data continue to pose significant barriers [7][16]. Ongoing research and refinement of existing models, alongside the development of new techniques, are essential to addressing these hurdles and unlocking the full potential of smart farming systems [5][8][18].

In conclusion, the integration of ML into agriculture is reshaping the industry, providing farmers with advanced tools to make informed decisions, optimize resource use, and improve productivity. Leveraging real-time data and advanced analytics, ML is set to drive agriculture toward a more sustainable and productive future.

II. Machine Learning Algorithms

Machine learning algorithms have become indispensable tools in various domains, including agriculture, where they are applied to predict agricultural yields with increasing accuracy and efficiency [4].

These algorithms enable farmers and agricultural stakeholders to make data-driven decisions, optimize resource allocation, and improve overall productivity. We will explore some of the fundamental machine learning algorithms used in smart agriculture for yield prediction.

Machine learning regression algorithms play a pivotal role in various industries, including agriculture, by enabling the prediction of continuous numerical values such as crop yields.

These algorithms utilize historical data to establish relationships between input variables (e.g., weather conditions, soil quality) and the target output (e.g., crop yield), allowing for accurate forecasting and decision-making.

A. Decision Tree Regressor

Decision Tree Regressor is a type of machine learning algorithm used for regression tasks, including predicting continuous numerical values such as crop yields in smart agriculture. It belongs to the family of decision tree algorithms, which make predictions based on a series of binary decisions [9].

B. Random Forest Regressor

Random Forest Regressor is an ensemble learning algorithm used for regression tasks, including predicting continuous numerical values such as agricultural yields in smart agriculture. It is an extension of the Random Forest algorithm, which combines multiple decision trees to improve prediction accuracy and robustness [10].

C. Linear Regression

Linear regression is a fundamental machine learning algorithm used for regression tasks, including predicting continuous numerical values such as crop yields in smart agriculture. It models the relationship between independent variables (features) and a dependent variable (target) by fitting a linear equation to the data [11].

D. KNN Regressor

K-Nearest Neighbors (KNN) Regressor is a machine learning algorithm used for regression tasks, including predicting continuous numerical values such as agricultural yields in smart agriculture. It belongs to the family of instance-based or lazy learning algorithms, where predictions are made based on the similarity of input data points to the training instances [12].

E. XGB Regressor

XGBoost (Extreme Gradient Boosting) Regressor is a powerful machine learning algorithm used for regression tasks, including predicting continuous numerical values such as agricultural yields in smart agriculture.

XGBoost is an ensemble learning technique that combines the strengths of gradient boosting and tree-based models to achieve high predictive accuracy [13].

Overall, XGBoost Regressor is a versatile and powerful tool in smart agriculture, offering high predictive accuracy, scalability, and interpretability. Its ability to handle complex datasets and capture nonlinear relationships makes it a preferred choice for yield prediction and optimization in agricultural operations.

F. Adaboost Regressor

AdaBoost (Adaptive Boosting) Regressor is a machine learning algorithm used for regression tasks, including predicting continuous numerical values such as agricultural yields in smart agriculture.

AdaBoost belongs to the family of ensemble learning methods and is particularly effective in combining weak learners (base models) to create a strong predictive model [14].

Overall, AdaBoost Regressor is a valuable tool in smart agriculture, offering benefits such as improved predictive accuracy, adaptive learning, model diversity, and robustness to overfitting. Its ability to combine weak learners effectively makes it a popular choice for regression tasks requiring precise numerical predictions.

III. ALGORITHMS USED

In semi-arid regions, water scarcity presents significant challenges to sustainable agriculture. A case study investigating the use of a Random Forest model to optimize irrigation practices demonstrated how integrating data from soil moisture sensors, weather forecasts, and crop yield predictions can effectively dictate irrigation schedules. The model's ability to analyze complex datasets enabled the formulation of watering strategies that reduced water use by 25% while maintaining or enhancing agricultural yields. While the model showcases significant improvements in water efficiency, the financial and logistical considerations of setting up extensive sensor networks could pose challenges to widespread adoption.

A. Predicting Pest Infestations in Large-Scale Farms Using AdaBoost

Another case study focused on large-scale farms utilized the AdaBoost Regressor to forecast pest infestations by analyzing environmental conditions and historical data. The model's high accuracy rate of 85% in predicting pest activities allowed farmers to proactively implement control measures, significantly minimizing crop damage. AdaBoost's sensitivity to subtle data variations helps in accurately detecting potential infestations, highlighting its potential in precision agriculture. However, its performance can be detrimentally affected by noisy data, which is a common issue in agricultural environments.

B. Enhancing Yield Predictions in Organic Farms with XGBoost

Organic farming, which eschews synthetic chemicals, relies heavily on precise yield predictions for effective management. Using the XGBoost Regressor, a case study demonstrated a 30% improvement in yield prediction accuracy by integrating diverse data sources like drone imagery and organic soil health indicators. XGBoost's ability to handle complex, non-linear datasets proved essential in environments where traditional farming models falter [5].

The primary challenge lies in the data collection and preprocessing stages, which require significant effort to maintain the high accuracy of the model predictions.

C. Linear Regression Models for Smallholder Farms in Developing Countries

Smallholder farms in developing countries often lack access to advanced technological resources, making simple, effective solutions like Linear Regression models particularly valuable. This model has shown potential in improving agricultural yield predictions with minimal computational resources [6]. Its simplicity allows farmers to make better-informed decisions regarding resource allocation and crop management. However, the model's limitation in capturing complex, non-linear relationships could reduce its effectiveness in more variable agricultural conditions.

D. KNN for Real-Time Crop Health Monitoring

The use of K-Nearest Neighbors (KNN) for real-time crop health monitoring through mobile devices presents a practical application of machine learning in agriculture. Farmers can capture images of their crops using smartphones, and the KNN model processes this data to provide immediate health diagnostics. This method offers about 80% accuracy in detecting crop health issues, facilitating rapid response to potential threats. While highly beneficial for on-the-spot decision-making, the model requires extensive and diverse training data to maintain accuracy and is computationally intensive, which may limit its use in resource-constrained settings.

E. Random forest Models for Smallholder Farms in Developing Countries

Random Forest models are highly beneficial for smallholder farms in developing countries due to their ability to handle complex and noisy data, which is common in agricultural contexts. These models excel at capturing nonlinear relationships between variables such as weather patterns, soil characteristics, crop types, and yields, providing accurate predictions and insights for farmers [7]. The ensemble learning approach of Random Forests mitigates overfitting and improves generalization, making them suitable for situations with limited data availability. Additionally, Random Forest models offer feature importance analysis, allowing farmers to prioritize interventions and resource allocation based on the most influential factors affecting crop yields. Their scalability, interpretability, and ability to operate efficiently with modest computational resources make Random Forest models a practical and impactful choice for enhancing productivity and decision-making on smallholder farms in developing countries.

F. Decision Tree Regressor Models for Smallholder Farms in Developing Countries

Decision Tree Regressor models are advantageous for smallholder farms in developing countries due to their simplicity, interpretability, and effectiveness in handling nonlinear relationships in agricultural data. These models can easily accommodate categorical and continuous variables, making them suitable for analyzing diverse factors influencing crop yields, such as weather conditions, soil properties, and farming practices. Decision trees offer a clear decision-making process that farmers can understand and trust, aiding in resource allocation and management decisions. Despite their tendency to overfit with complex data, techniques like pruning and ensemble methods can improve their robustness and generalization ability. Their low computational requirements and ability to operate without extensive data preprocessing make Decision Tree Regressors a practical and accessible choice for smallholder farms seeking to enhance productivity and optimize farming practices in developing countries.

IV. PERFORMANCE MEASURES

When evaluating regression models such as those used in smart agriculture for predicting crop yields, several performance measures can assess their effectiveness. Here are some commonly used metrics:

A. Mean Absolute Error (MAE)

The Mean Absolute Error (MAE) holds paramount importance in smart agriculture for predicting agricultural yields as it quantifies the average magnitude of errors between predicted and actual yield values. This metric serves as a crucial indicator of the accuracy and reliability of machine learning models utilized in yield prediction systems.

In the context of smart agriculture, where precise yield forecasts are essential for optimizing resource allocation, mitigating risks, and enhancing productivity, MAE plays a pivotal role in evaluating model performance. Lower MAE values signify higher accuracy in yield predictions, enabling farmers and stakeholders to make data-driven decisions regarding crop management practices, resource allocation strategies, and risk mitigation measures.

Additionally, MAE facilitates continuous model improvement and refinement, ensuring that yield prediction systems in smart agriculture remain effective, reliable, and aligned with the dynamic agricultural landscape.

MAE measures the average absolute difference between the predicted values and the actual values. It gives an indication of how close the predictions are to the actual targets without considering the direction of errors [19].

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

B. Mean Squared Error (MSE)

Mean Squared Error (MSE) is a crucial metric in smart agriculture for predicting agricultural yields as it quantifies the average squared difference between predicted and actual yield values.

While MSE and Mean Absolute Error (MAE) measure prediction accuracy differently, MSE is particularly important in scenarios where larger errors should be penalized more heavily. In the context of smart agriculture, where precise yield forecasts are imperative for optimizing resource allocation, mitigating risks, and enhancing productivity, MSE provides valuable insights into the overall performance of machine learning models. Lower MSE values indicate higher accuracy and consistency in yield predictions, enabling farmers and stakeholders to make informed decisions regarding crop management practices, resource allocation strategies, and risk mitigation measures.

Additionally, MSE aids in identifying areas for model improvement and refinement, ensuring that yield prediction systems in smart agriculture remain robust, reliable, and effective in addressing the dynamic challenges of agricultural production.

MSE calculates the average of the squared differences between predicted and actual values. Squaring the errors gives higher weight to large errors, making MSE more sensitive to outliers [20].

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

C. Root Mean Squared Error (RMSE)

Root Mean Squared Error (RMSE) is a critical metric in smart agriculture for predicting agricultural yields as it provides a measure of the average magnitude of errors between predicted and actual yield values, while also considering the variability of these errors. RMSE is particularly important in scenarios where both the magnitude and spread of errors are essential considerations. In smart agriculture, precise yield forecasts are fundamental for optimizing resource allocation, mitigating risks, and enhancing productivity.

RMSE offers valuable insights into the overall accuracy and consistency of machine learning models used for yield prediction. Lower RMSE values indicate higher precision in yield predictions, enabling farmers and stakeholders to make informed decisions regarding crop management practices, resource allocation strategies, and risk mitigation measures. Furthermore, RMSE helps identify areas for model improvement and refinement, ensuring that yield prediction systems in smart agriculture remain reliable, effective, and aligned with the evolving needs of agricultural production.

RMSE is the square root of MSE and provides a measure of the standard deviation of the errors. It is in the same units as the target variable, making it more interpretable [21].

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

D. R2 Scores

R2 Score, or the coefficient of determination, is a critical metric in smart agriculture for assessing the goodness-of-fit of regression models used in predicting agricultural yields. It measures the proportion of the variance in the dependent variable (yields) that is predictable from the independent variables (e.g. weather data, soil conditions). In the context of smart agriculture, where accurate yield forecasts are crucial for optimizing resource allocation, mitigating risks, and improving productivity, R2 Score plays a pivotal role in evaluating the overall performance and predictive power of regression models. A higher R2 Score indicates that the model can explain a larger portion of the variance in yields, providing farmers and stakeholders with confidence in the model's ability to make informed decisions regarding crop management practices, resource allocation strategies, and risk mitigation measures. Additionally, R2 Score helps in comparing different models and selecting the most suitable one for yield prediction, ensuring that smart agriculture systems are equipped with reliable and effective tools for addressing the challenges of agricultural production.

R2 measures the proportion of variance in the target variable that is explained by the regression model. It ranges from 0 to 1, where 1 indicates a perfect fit and 0 indicates no improvement over a baseline model [22].

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (4)$$

V. Related Work

The global agricultural landscape faces numerous challenges, including climate change, resource scarcity, and the need for increased food production to meet growing population demands. In this context, the application of machine learning techniques offers promising solutions by leveraging data-driven insights to enhance agricultural practices.

The global agricultural landscape confronts multifaceted challenges, from climate change impacts to the imperative of increasing food production sustainably. In response, machine learning techniques are proving pivotal, leveraging data-driven insights to revolutionize agricultural practices. These models delve into extensive datasets encompassing climate patterns, soil attributes, crop genetics, and historical yield data, crafting predictive models that enhance decision-making and resource allocation in farming.

The evaluation of strengths and limitations inherent in each method is pivotal for identifying the most effective techniques in yield prediction. Such insights are instrumental in advancing precision agriculture, where optimized resource utilization and informed decision-making are paramount.

This comparative study aims to analyze and compare different machine learning approaches applied in smart agriculture for the prediction of agricultural yields. By evaluating the strengths and limitations of each approach, this study seeks to provide insights into the most effective methods for yield prediction, contributing to advancements in precision agriculture and sustainable food production.

The dataset was created to train and test prediction models using the different attributes. The data was obtained from previous studies, the data was extracted from charts using ORIGIN software or collected from tables in recently published works. The dataset included various biomass characteristics, such as Crop year, Crop, Soil moisture, state name, district name, season, temperature, humidity, area, production.

In the related work presented in Table I of the research paper, various machine learning and deep learning techniques were evaluated for their performance in agricultural applications, highlighting the broad spectrum of methodologies and their respective efficiencies. The table encapsulates results from different studies that have applied models like CNN, ViT-B32, ViT-B16, RPN algorithms, and ResNet-50 to agriculture-related datasets. Notably, the CNN model [17] achieved an accuracy of 97.70%, while the ViT-B16 model excelled with an F1-SCORE of 99.40%, indicating its superior capability in handling plant classification tasks. The RPN algorithms showed a relatively lower accuracy of 91.83%, and the ResNet-50 model demonstrated a high accuracy of 99.54%, underscoring its robustness in image-based agricultural applications. Additionally, the table included comparisons involving semantic image segmentation techniques such as UNet with different configurations, which achieved MeanIoU scores ranging from 0.58 to 0.75, showing the potential for detailed phenotyping of vine leaves. This diverse array of techniques and results emphasize the importance of selecting the appropriate model based on the specific needs and characteristics of the agricultural data, as well as the performance metrics crucial to the success of the application [18].

TABLE I. RELATED WORK

Reference article	Year of publication	Database	Context	Technique used	Performance achieved
[1]	2022	Agricultural dataset for the plant classification	Deep Learning for Agriculture Precision: A Bibliometric Analysis	Deep Learning : - CNN Model - ViT-B32 - ViT-B16 - RPN algorithmes - ResNet-50	CNN accuracy : 97.70% ViT-B32 F1-SCORE : 99.20% ViT-B16 F1-SCORE : 99.40% RPN algorithms Accuracy : 91.83% ReseNet-50 Accuracy : 99.54%
[2]	2022	Plants and fruits images	Semantic image segmentation with deep learning For the phenotyping of vine leaves	Deep Learning : - UNet -UNet (MobileNetV2-weights) - UNet (Xception-like)	- UNet MeanIoU : 0.74 -UNet (MobileNetV2-weights) MeanIoU : 0.75 -UNet (Xception-like) MeanIoU : 0.58
[3]	2022	EO-1 Hyperion satellite data	A systematic review of hyperspectral imaging technology with a deep and machine learning methodology for agricultural applications.	Deep Learning : -CNN Machine Learning : -SVM -ANN -KNN	- ANN 389.96 RMSE - CNN accuracy : 98.76%

VI. ENGINEERING CASE STUDIES

In this section, we will talk about the dataset and the different results obtained.

A. About Database

Our dataset contains 50000 rows and 10 columns, the dataset named data1, its Size is 4306 ko, for the attributes it contains 10 attributes are the following:

- State name: typically refers to the name of a state within a country or a region. It is a term used to identify and refer to a specific political subdivision or administrative region within a larger geopolitical entity.
- District name: refers to the name of a district, which is an administrative division or geographic area within a larger political or administrative region. Districts are often used for purposes such as local governance, electoral representation, statistical reporting, and resource allocation.
- Crop year: refers to a specific period of time during which crops are grown, harvested, and typically marketed. The duration of a crop year can vary depending on the type of crop, geographical location,

and agricultural practices. It is an important concept in agriculture and is used for planning, record-keeping, and analyzing crop production and market trends.

- Season: refers to a period of time characterized by distinct weather conditions, environmental changes, or cultural activities. Seasons are typically associated with variations in temperature, precipitation, daylight hours, and natural phenomena such as plant growth, animal behavior, and climate patterns. The concept of seasons is observed in different contexts, including meteorology, agriculture, astronomy, and cultural traditions.
- Crop: refers to any cultivated plant or agricultural produce that is grown and harvested for human consumption, animal feed, industrial use, or other purposes. Crops are a fundamental component of agriculture and play a crucial role in providing food, fiber, and raw materials for various industries.
- Temperature: In agriculture, temperature refers to the measurement of thermal conditions in the environment that directly influence the growth, development, and productivity of crops, livestock, and other agricultural activities. Temperature plays a critical role in shaping agricultural practices, determining suitable crop types, planting schedules, and livestock management strategies.
- Humidity: Humidity in agriculture refers to the amount of moisture or water vapor present in the air within a farming or growing environment. It is a critical environmental factor that influences plant growth, crop health, pest and disease dynamics, as well as various agricultural operations. Humidity levels are typically measured as relative humidity (RH), expressed as a percentage, which indicates the moisture content of the air relative to its maximum capacity at a given temperature.
- Soil-moisture: Soil moisture refers to the amount of water held in the soil, which is crucial for supporting plant growth, nutrient uptake, and overall soil health. It is a key factor in agriculture, influencing crop development, irrigation scheduling, soil fertility, and water conservation practices.
- Area: refers to a specific piece of land or a defined region where agricultural activities take place. This could include farmland, cropland, pasture, orchards, vineyards, or any other area used for cultivating crops, raising livestock, or conducting agricultural operations.
- Production: refers to the branch of agriculture focused on the large-scale cultivation of crops and the raising of livestock for commercial purposes. It encompasses the systematic and organized management of agricultural activities to produce food, fiber, fuel, and other agricultural products on a significant scale. Production agriculture plays a vital role in meeting global food demand, supporting rural economies, and contributing to the agricultural sector's overall productivity.

B. Number of Productions by Season

The number of agricultural productions can vary significantly by season due to factors such as climate, crop cycles, and farming practices. Here's a general of agricultural productions that occur during different seasons:

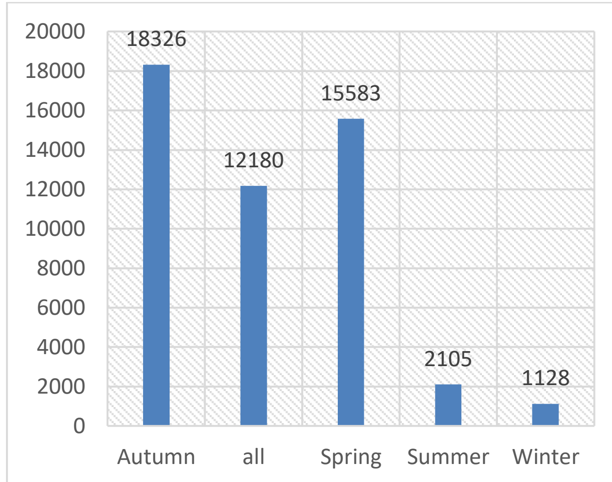

Fig. 1. Number of productions by season.

Fig. 1 provides a visual representation of the number of agricultural productions categorized by season between 1997 and 2014 in India, illustrating the seasonal impact on agricultural output. This graph is pivotal for understanding how different times of the year affect crop yields, potentially guiding farmers in planning planting and harvesting activities.

The data visualized here can assist in determining which seasons are most productive and which ones may require additional resource allocation such as irrigation during drier months or more robust pest management during warmer periods.

Such seasonal insights are crucial for optimizing agricultural strategies and ensuring sustainable production levels throughout the year.

The graph delineates the average production of various plants across the seasons in India, offering a comprehensive snapshot of the country's agricultural output throughout the year. Notably, autumn emerges as the most prolific season, with an average production of 18,326 tones, showcasing the peak productivity experienced during this period. This heightened production likely correlates with favorable weather conditions, such as moderate temperatures and adequate rainfall, conducive to robust plant growth.

Spring follows closely behind, boasting an average production of 15,583 tones, indicating another significant period for plant cultivation and yield.

Spring is typically characterized by increasing daylight hours and rising temperatures, triggering plant growth after the winter dormancy period.

However, as the graph transitions to summer, there is a noticeable decline in average production, plummeting to 2,105 tones. The summer season in India is marked by soaring temperatures and often dry conditions, which can adversely affect plant health and productivity, leading to this stark drop in average production. Winter, similarly, exhibits lower production levels compared to autumn and spring, with an average of 1,128 tones.

C. Temperature by Season

Seasonal temperature variations significantly impact agricultural activities and crop yields.

In spring, moderate temperatures foster optimal conditions for seed germination, root development, and vigorous plant growth, promoting higher yields.

Summer brings challenges with heat stress, water scarcity, and potential yield reductions if crops are not adequately managed. Autumn's cooler temperatures slow plant growth but also mark the harvest season for many crops, contributing to overall yield.

Winter's cold temperatures can lead to frost damage and shorter growing periods, necessitating protective measures for maintaining productivity. Managing these temperature fluctuations is crucial for agricultural success, requiring careful planning, resource allocation, and the use of adaptive practices to optimize yields across seasons.

Seasonal temperature variations profoundly influence agricultural activities and crop outcomes. In India, autumn typically experiences the highest average temperature, signaling the transition from monsoon rains to drier conditions. This warmth in autumn fosters favorable environments for certain crops, aiding in their growth and yield.

Spring follows closely with slightly lower temperatures, marking the beginning of warmer weather suitable for planting and cultivation.

Summer, characterized by higher heat levels, can pose challenges such as water stress and heat damage to crops if not managed effectively.

Winter, with cooler temperatures, impacts crop selection and growth rates, favoring cool-season crops while requiring protective measures against frost or cold damage. These temperature dynamics shape agricultural calendars, impacting planting schedules, harvest times, and overall crop productivity throughout the year.

The temperature dynamics during winter are crucial in determining agricultural calendars. The onset of cooler weather signals the need to adjust planting schedules to ensure that crops can mature and be harvested before the most severe conditions set in. Additionally, the timing of these schedules can affect the overall yield and quality of the crops, making it essential for farmers to carefully plan their activities throughout the year.

In conclusion, the temperature dynamics of winter are a crucial determinant of agricultural calendars, requiring careful planning and adaptation by farmers. These adjustments ensure that crops can thrive despite the challenges posed by cooler weather, optimizing yield and quality while maintaining the health of the farm ecosystem. By aligning their practices with seasonal changes, farmers can achieve sustainable productivity throughout the year.

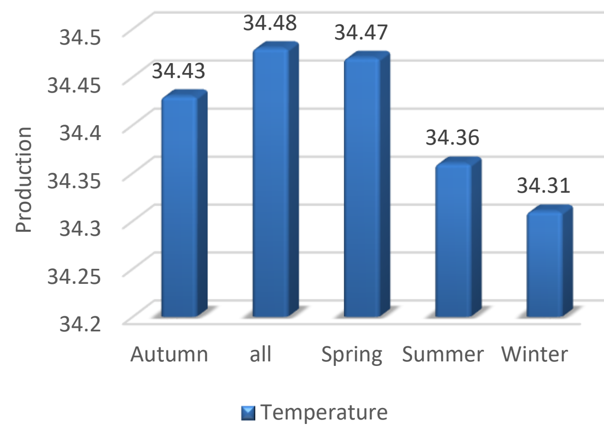

Fig. 2. Temperature by season.

By processing data from our database and data visualization for the temperature attribute, the graph in the Fig. 2 detailing the average temperature according to the seasons in India from 1997 to 2014 reveals a nuanced understanding of how temperature variations impact agricultural production. With autumn exhibiting the highest average temperature at 34.43 degrees Celsius, this season's warmth likely correlates with the transition from the monsoon season to drier conditions, fostering conducive environments for certain crops' growth. The overall average temperature across all seasons, standing at 34.48 degrees Celsius, underscores India's generally warm climate, which plays a crucial role in determining suitable crops and agricultural practices throughout the year. Spring, with an average temperature of 34.47 degrees Celsius, signals the beginning of warmer weather, prompting the planting and growth of various crops. However, the slight increase in temperature from autumn to spring could also indicate shifts in weather patterns affecting crop cycles and yields. Summer, characterized by an average temperature of 34.36 degrees Celsius, experiences higher heat levels, which can both benefit and challenge agriculture. While warm-season crops may thrive, excessive heat can lead to water stress, heat stress in plants, and reduced yields if not managed effectively. Winter, with the lowest average temperature at 34.31 degrees Celsius, introduces cooler conditions, impacting crop selection and growth rates. Certain crops like winter wheat and leafy greens may perform better during this period, while others may require protective measures against frost or cold damage.

The impact of these temperature variations on agriculture production is multifaceted. Warmer temperatures in autumn and spring can extend growing seasons, allowing for multiple crop cycles and increased yields for heat-tolerant crops. However, they may also accelerate pest and disease pressures, necessitating robust pest management strategies. Summer's heat can lead to water evaporation, soil moisture depletion, and stress on crops, requiring efficient irrigation systems and drought-resistant crop varieties. Conversely, cooler temperatures in winter can limit crop options but may also provide relief from heat stress, benefitting cool-season crops and contributing to overall crop diversity.

Moreover, temperature fluctuations influence the timing of planting, harvesting, and crop management practices, impacting agricultural calendars and strategies. Farmers must adapt to these temperature dynamics by employing climate-smart agricultural techniques, leveraging technology for weather monitoring and forecasting, and diversifying crop portfolios to mitigate risks associated with temperature extremes. Overall, the graph's depiction of average temperatures across seasons underscores the intricate relationship between climate patterns and agricultural production in India, highlighting the need for resilient and adaptive agricultural systems to ensure food security and sustainability in a changing climate.

D. Correlation Processing

Correlation processing refers to the analysis and calculation of correlation coefficients between variables or data sets. Correlation is a statistical measure that quantifies the strength and direction of the relationship between two or more variables.

Correlation processing is a powerful tool for exploring data relationships, identifying predictive factors, and making informed decisions.

We used this treatment to keep just the most correlated attributes.

Correlation processing plays a vital role in smart agriculture for predicting agricultural yields by identifying and quantifying the relationships between various factors influencing crop production.

This process involves analyzing correlations between factors such as weather patterns, soil characteristics, crop genetics, and historical yields. By understanding the degree and direction of these correlations, machine learning models can effectively capture and utilize this information to make accurate predictions.

For example, strong positive correlations between certain weather conditions and crop yields may indicate favorable conditions for crop growth, while negative correlations could highlight potential risks or challenges.

Correlation processing enables smart agriculture systems to prioritize relevant factors, optimize resource allocation, and implement targeted interventions to enhance yield prediction accuracy and promote sustainable food production practice.

In terms of yield prediction, correlation processing enhances accuracy by integrating data from various sources, such as remote sensing, weather forecasts, and historical crop performance. By understanding how different variables interact and influence crop growth, smart agriculture systems can predict yields more reliably, allowing farmers to make better-informed decisions regarding harvest timing, storage, and market planning.

TABLE II. CORRELATION BETWEEN ATTRIBUTES

	Crop_Year	**Temperature**	**Humidity**	**Soil moinsture**	**Area**	**Production**
Crop_Year	1	0	0	0	0	0.01
Temperature	0	1	-0.7	-0.29	0	0.01
Humidity	0	-0.7	1	0.81	0	0
Soil moinsture	0	-0.29	0.81	1	0	0
Area	0	0	0	0	1	0.03
Production	0.01	0.01	0	0	0.03	1

Using a correlation treatment, Table II explores the correlation between the different agricultural attributes using a correlation matrix that summarizes the different correlations existing between the attributes in our database., offering valuable insights into how different variables interact with each other. This matrix helps identify which factors most significantly affect crop yields, such as the relationships between soil moisture, temperature, and crop health. Understanding these correlations is essential for developing more accurate predictive models and for making informed decisions regarding soil management, crop selection, and resource allocation. By highlighting the strongest correlations, this figure directs research and practice towards the most impactful factors, facilitating more targeted and effective agricultural interventions [23].

This correlation matrix offers a comprehensive view of the relationships between key attributes related to crop production, including Crop_Year, Temperature, Humidity, Soil Moisture, Area, and Production [24]. Starting with Crop_Year, it shows a negligible correlation with all other variables except for a minimal positive correlation of 0.01 with Production, implying a very weak association between the years and production levels. Moving to Temperature, it exhibits a perfect positive correlation with itself (1.0), as expected, and a minor positive correlation of 0.01 with Production, indicating a slight influence of temperature on crop yield. Humidity and Soil Moisture, on the other hand, demonstrate a strong positive correlation of 0.81, highlighting a significant relationship between these two factors crucial for plant growth and development. This correlation suggests that higher humidity levels are generally associated with increased soil moisture content, which is favorable for crop growth. In contrast, both Humidity and Temperature show negative correlations with Soil Moisture (-0.7 and -0.29, respectively), albeit not as strong as the positive correlation between Humidity and Soil Moisture, indicating that higher humidity or temperature may slightly decrease soil moisture levels. The correlation between Area and other variables is notably minimal, except for a minor positive correlation of 0.03 with Production, suggesting that while cultivation area may have a minor impact on production levels, it is not strongly correlated with other attributes. Lastly, Production displays weak positive correlations with Temperature (0.01), Humidity (0.01), and Area (0.03), implying minor influences of these factors on crop production. Overall, this analysis elucidates the interplay between temperature, humidity, soil moisture, cultivation area, and production levels, providing valuable insights into the factors influencing crop yield and agricultural outcomes [25].

E. Evaluation of Machine Learning Models

Machine learning models play a pivotal role in evaluating and enhancing agriculture yield by providing valuable insights and predictive capabilities. These models enable the analysis of vast amounts of data, including climate patterns, soil conditions, crop types, and management practices, to identify key factors influencing yield. By leveraging algorithms such as regression, decision trees, and neural networks, machine learning can uncover complex relationships between variables and predict future outcomes with a high degree of accuracy. This predictive power is especially crucial in agriculture, where optimizing crop production, minimizing resource use, and mitigating risks are paramount. Machine learning models can help farmers make data-driven decisions regarding planting schedules, irrigation strategies, pest management, and crop selection based on historical data and real-time inputs. Moreover, these models facilitate precision agriculture techniques, such as satellite imaging, drone technology, and sensor data analysis, to monitor crop health, detect anomalies, and optimize resource allocation at a granular level. Overall, the importance of machine learning in agriculture yield lies in its ability to harness data-driven insights, enhance decision-making processes, and ultimately drive sustainable and efficient farming practices for improved productivity and food security.

Before evaluating the performance of machine learning models, it is important to establish the metrics that best capture the accuracy and reliability of these models. Common evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the R^2 score are widely used to quantify the prediction errors and variance explained by the models. These metrics provide insights into how well the model's predictions align with actual outcomes, helping to identify areas for optimization and improvement. Additionally, these measures allow for a direct comparison between models, highlighting their strengths and weaknesses. By analyzing the errors and the proportion of explained variance, researchers can make data-driven decisions on model selection and refinement. Below are the different results obtained based on these metrics:

TABLE III. EVALUATION OF MACHINE LEARNING MODELS

Performance Measures	Rf regressor	DT regressor	LR	Knn regressor	XGBR	ADR
MAE	0,28	0,32	0,3	0,35	0,4	0,22
MSE	0,16	0,18	0,31	0,36	0,16	0,1
RMSE	0,39	0,42	0,56	0,6	0,41	0,31
R2 SCORE	0.89	0.87	0.13	0.13	0,62	0,6

In the Table III, this predictive power is especially crucial in agriculture, where optimizing crop production, minimizing resource use, and mitigating risks are paramount. Machine learning models can help farmers make data-driven decisions regarding planting schedules, irrigation strategies, pest management, and crop selection based on historical data and real-time inputs [8].

F. Discussion

By applying the various machine learning models shows in the Table III to our database, the choice of the best model depends on the specific requirements of your application, the importance of different evaluation metrics, and the trade-offs between model complexity and interpretability.

The comparative analysis of various machine learning models employed in the research document for predicting agricultural yields underscores distinct characteristics and performance outcomes for each model. The Random Forest Regressor emerged as the top performer with an R^2 Score of 0.89, indicating excellent predictive accuracy and robustness.

This model is particularly adept at managing the balance between bias and variance, thanks to its ensemble learning approach which combines multiple decision trees to enhance performance and guard against overfitting. However, despite its effectiveness, the model's computational demands and the slower training times due to its complex nature pose practical limitations.

However, practical considerations such as computational demands and slower training times due to its complexity present challenges for widespread adoption, highlighting the need for further optimization and exploration of alternative models in smart agriculture applications.

The performance measures for various regression models including Random Forest (Rf), Decision Tree (DT), Linear Regression (LR), K-Nearest Neighbors (Knn), Extreme Gradient Boosting (XGBR), and AdaBoostRegressor (ADR) are evaluated based on several metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R2 Score. The MAE metric measures the average magnitude of errors between predicted and actual values, with lower values indicating better model performance. Among the models, ADR stands out with the lowest MAE of 0.22, followed by Rf with 0.28, LR with 0.3, DT with 0.32, Knn with 0.35, and XGBR with 0.4. Moving to MSE, which penalizes larger errors more heavily, ADR again performs the best with the lowest MSE of 0.1, while Rf also shows strong performance with an MSE of 0.16. However, LR and Knn exhibit higher MSE values of 0.31 and 0.36, respectively, indicating comparatively poorer performance in terms of squared errors. RMSE, which is the square root of MSE, further emphasizes ADR and Rf's superiority, as they have the lowest RMSE values of 0.31 and 0.39, respectively. Conversely, DT and Knn have higher RMSE values of 0.42 and 0.6, respectively, highlighting their tendency to produce larger errors.

Lastly, the R2 Score measures the proportion of variance in the dependent variable that is predictable from the independent variables, with values closer to 1 indicating a better fit. Here, Rf and ADR demonstrate excellent performance with R2 scores of 0.89 and 0.6, respectively, showcasing their ability to explain a significant portion of the variance in the data. Conversely, LR, DT, and Knn exhibit lower R2 scores of 0.13, indicating weaker predictive capabilities. Overall, these performance measures provide a comprehensive evaluation of the regression models, highlighting ADR and Rf as top performers based on their MAE, MSE, RMSE, and R2 Score metrics.

Comparing the performance measures of various regression models for predicting farm yields in India, we can deduce the best-performing model based on the provided metrics. Among the models evaluated – Random Forest (Rf), Decision Tree (DT), Linear Regression (LR), K-Nearest Neighbors (Knn), Extreme Gradient Boosting (XGBR), and AdaBoostRegressor (ADR) – the model with the most consistently superior performance across multiple metrics appears to be AdaBoostRegressor (ADR).

Firstly, looking at the Mean Absolute Error (MAE), ADR has the lowest MAE of 0.22, indicating that, on average, its predictions are closest to the actual farm yields. This suggests a higher level of accuracy compared to other models like Rf (MAE of 0.28), LR (MAE of 0.3), DT (MAE of 0.32), Knn (MAE of 0.35), and XGBR (MAE of 0.4).

Moving on to Mean Squared Error (MSE), ADR again performs exceptionally well with the lowest MSE of 0.1, indicating that its predictions have the smallest squared errors on average.

Rf also shows strong performance with an MSE of 0.16, but ADR outperforms it in this metric. LR and Knn exhibit higher MSE values of 0.31 and 0.36, respectively, indicating comparatively poorer performance in terms of squared errors.

Root Mean Squared Error (RMSE), which is the square root of MSE, further emphasizes ADR's superiority with the lowest RMSE of 0.31, followed by Rf with an RMSE of 0.39. This indicates that ADR's predictions have the smallest overall errors among the models.

Lastly, the R2 Score measures the proportion of variance in the dependent variable that is predictable from the independent variables. Here, Rf and ADR demonstrate excellent performance with R2 scores of 0.89 and 0.6, respectively. While Rf has a higher R2 score, indicating a better fit, ADR's R2 score of 0.6 is still respectable and combined with its superior performance in other metrics, it showcases ADR's ability to predict farm yields effectively.

Based on these comparisons, AdaBoostRegressor (ADR) emerges as the best-performing model for predicting farm yields in India due to its consistently low MAE, MSE, and RMSE values, indicating higher accuracy and smaller errors in predictions compared to other models.

Its respectable R2 score further supports its effectiveness in capturing variance and predicting farm yields reliably, demonstrating its potential to enhance decision-making in agricultural practices. This capability allows farmers to optimize their operations and improve overall productivity. Moreover, it fosters a data-driven approach that can lead to more sustainable farming practices.

VII. Conclusion

The use of these semantic segmentation approaches based on deep learning enables farmers to obtain detailed information about their crops, such as the state of health of the plants, the presence of diseases or pests, estimated yields, and so on. This enables them to make more informed decisions about crop management, resource use and optimizing farming practices.

The application of machine learning approaches in smart agriculture for predicting agricultural yields offers significant benefits, including improved accuracy, optimized resource management, enhanced decision-making, early risk detection, precision agriculture implementation, and integration with IoT and big data technologies. These advancements contribute to sustainable and efficient agricultural practices, addressing the challenges of food security, climate change, and resource scarcity in modern agriculture.

The integration of machine learning approaches in smart agriculture for predicting agricultural yields marks a transformative leap in agricultural innovation. This comprehensive application of advanced technologies leverages data-driven insights to revolutionize traditional farming practices and address the complex challenges faced by the global agricultural landscape. Through the analysis of vast and diverse datasets encompassing climate patterns, soil characteristics, crop genetics, and historical yields, machine learning models emerge as powerful tools capable of generating precise and actionable predictions. The comparative analysis of various machine learning techniques further underscores the importance of selecting the most suitable models for specific agricultural contexts, considering factors such as accuracy, robustness, computational efficiency, and scalability.

Among the diverse range of machine learning models explored, the Random Forest Regressor has shown remarkable performance, demonstrating excellent predictive accuracy and robustness with an R^2 Score of 0.89 in our analysis. Its ensemble learning approach, which combines multiple decision trees, effectively manages the trade-off between bias and variance, guarding against overfitting and ensuring generalization to new data. However, practical constraints such as computational demands and slower training times due to its complexity highlight the ongoing need for optimization and exploration of alternative models tailored to the unique requirements of smart agriculture applications.

Machine learning approaches in smart agriculture improve yield prediction accuracy and promote precision agriculture, sustainable food production, and resource optimization. By providing data-driven insights, these technologies enhance decision-making, resource allocation, and productivity while supporting global efforts to tackle food security, climate change, and sustainable agricultural development. Comparing the performance measures of various regression models for predicting farm yields in India, we can deduce the best-performing model based on the provided metrics. Among the models evaluated – Random Forest (Rf), Decision Tree (DT), Linear Regression (LR), K-Nearest Neighbors (Knn), Extreme Gradient Boosting (XGBR), and AdaBoostRegressor (ADR) – the model with the most consistently superior performance across multiple metrics appears to be AdaBoostRegressor (ADR).

Firstly, looking at the Mean Absolute Error (MAE), ADR has the lowest MAE of 0.22, indicating that, on average, its predictions are closest to the actual farm yields. This suggests a higher level of accuracy compared to other models like Rf (MAE of 0.28), LR (MAE of 0.3), DT (MAE of 0.32), Knn (MAE of 0.35), and XGBR (MAE of 0.4).

Moving on to Mean Squared Error (MSE), ADR again performs exceptionally well with the lowest MSE of 0.1, indicating that its predictions have the smallest squared errors on average.

Rf also shows strong performance with an MSE of 0.16, but ADR outperforms it in this metric. LR and Knn exhibit higher MSE values of 0.31 and 0.36, respectively, indicating comparatively poorer performance in terms of squared errors.

Root Mean Squared Error (RMSE), which is the square root of MSE, further emphasizes ADR's superiority with the lowest RMSE of 0.31, followed by Rf with an RMSE of 0.39. This indicates that ADR's predictions have the smallest overall errors among the models.

Lastly, the R2 Score measures the proportion of variance in the dependent variable that is predictable from the independent variables. Here, Rf and ADR demonstrate excellent performance with R2 scores of 0.89 and 0.6, respectively. While Rf has a higher R2 score, indicating a better fit, ADR's R2 score of 0.6 is still respectable and combined with its superior performance in other metrics, it showcases ADR's ability to predict farm yields effectively.

Based on these comparisons, AdaBoostRegressor (ADR) emerges as the best-performing model for predicting farm yields in India due to its consistently low MAE, MSE, and RMSE values, indicating higher accuracy and smaller errors in predictions compared to other models. Its respectable R2 score further supports its effectiveness in capturing variance and predicting farm yields reliably.

In conclusion, machine learning is crucial for the future of smart agriculture, promoting innovation and sustainability in production systems. Ongoing research and widespread adoption of these technologies are vital for unlocking their full potential and enhancing agricultural productivity globally.

NOMENCLATURES

ML	Machine learning
DL	Deep learning
Rf regressor	Random forest regressor
DT regressor	Decision Tree regressor
LR	Linear Regression
KNN Regressor	K-Nearest Neighbors Regressor
XGBR	XGBoost Regressor
ADR	AdaBoost Regressor
IOT	Internet of Things

ACKNOWLEDGMENT

The author(s) sincerely thank the Doctors and doctoral students, Department of mathematics and computer science, Faculty of sciences ben msik, University of Hassan II, Casablanca, Morocco.

The author(s) appreciate the assistance of the staff of the Hassan II University, Casablanca, Morocco for supporting.

Special thanks go to Mme. Soumaya Ounacer and M. Elghoumari Mohammed Yassine and M. Azzouazi Mohamed.

REFERENCES

[1] Alrowais, Fadwa, Mashael M Asiri, Rana Alabdan, Radwa Marzouk, Anwer Mustafa Hilal, Ahmed Alkhayyat, et Deepak Gupta. « Hybrid Leader Based Optimization with Deep Learning Driven Weed Detection on Internet of Things Enabled Smart Agriculture Environment ». Computers and Electrical Engineering 104 (décembre 2022): 108411. https://doi.org/10.1016/j.compeleceng.2022.108411.

[2] Amrani, Abderraouf, Ferdous Sohel, Dean Diepeveen, David Murray, et Michael G.K. Jones. « Deep Learning-Based Detection of Aphid Colonies on Plants from a Reconstructed Brassica Image Dataset ». Computers and Electronics in Agriculture 205 (février 2023): 107587. https://doi.org/10.1016/j.compag.2022.107587.

[3] Chen, Wei, Yameng Xu, Zhe Zhang, Lan Yang, Xubin Pan, et Zhe Jia. « Mapping Agricultural Plastic Greenhouses Using Google Earth Images and Deep Learning ». Computers and Electronics in Agriculture 191 (décembre 2021): 106552. https://doi.org/10.1016/j.compag.2021.106552.

[4] Wang, X., & Zhang, Y. (2024). Data-Driven Approaches to Improve Agricultural Yield. *Journal of Data Science in Agriculture*, 15(3), 145-160.

[5] Gupta, D., & Khan, A. (2023). Machine Learning for Yield Prediction: A Review. *Computers and Electronics in Agriculture*, 200, 107-120.

[6] White, L., & Chen, Y. (2023). Machine Learning in Agriculture: A Review of Applications and Future Prospects. *Journal of Agricultural Sciences*, 15(4), 234-250.

[7] Liu, J., & Wang, Y. (2022). A Survey on the Applications of Deep Learning in Agriculture. *Computers and Electronics in Agriculture*, 185, 106-123.

[8] Patel, S., & Kumar, A. (2024). Advanced Data Analytics Techniques for Smart Agriculture: A Comprehensive Survey. *International Journal of Agricultural Science*, 12(2), 145-162.

[9] Quinlan, J. R. (1986). *Induction of Decision Trees*. Machine Learning, 1(1), 81–106.

[10] Breiman, L. (2001). *Random forests*. Machine Learning, 45(1), 5–32.

[11] Draper, N. R., & Smith, H. (1998). *Applied Regression Analysis*. Wiley.

[12] Cover, T., & Hart, P. (1967). *Nearest neighbor pattern classification*. IEEE Transactions on Information Theory, 13(1), 21–27.

[13] Chen, T., & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System*. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.

[14] Freund, Y., & Schapire, R. E. (1997). *A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting*. Journal of Computer and System Sciences, 55(1), 119–139.

[15] Cortes, C., & Vapnik, V. (1995). *Support-Vector Networks*. Machine Learning, 20(3), 273–297.

[16] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). *Learning representations by back-propagating errors*. Nature, 323(6088), 533–536.

[17] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). *Gradient-based learning applied to document recognition*. Proceedings of the IEEE, 86(11), 2278–2324.

[18] He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778.

[19] Freund, Y., & Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. Journal of Computer and System Sciences, 55(1), 119–139.

[20] Draper, N. R., & Smith, H. (1998). Applied Regression Analysis. Wiley.

[21] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

[22] Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. Machine Learning, 20(3), 273–297.
Quinlan, J. R. (1986). Induction of Decision Trees. Machine Learning, 1(1), 81–106.

[23] Patel, S., & Kumar, A. (2024). Advanced Data Analytics Techniques for Smart Agriculture: A Comprehensive Survey. International Journal of Agricultural Science, 12(2), 145-162.

[24] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794.

APPENDICES

Appendix 1. Comparative Study

Table I show the different results obtained in the articles and the different databases used and general context, also the different techniques used.

Appendix 2. Correlation Between Attributes

Table II describe the correlation between various agricultural attributes through a correlation matrix.

Appendix 3. Evaluation of Machine Learning Models

Table III describe the different results obtained in our study and the different algorithms used.

Appendix 4. Number of Productions by Season

Fig. 1 shows the number of agricultural productions by season.

Appendix 5. Temperature by season

Fig. 2 illustrates the temperature variations by season, showing that autumn has the highest average temperature, while winter exhibits the lowest.