# *Proceedings of* **DAAfrica'2024 -** Data Science for Agriculture in Africa

***Workshop affiliated with CARI'2024***

***November 23, 2024, Bejaia, Algeria (hybrid)***

**Editors:**

- **Paulin Melatagia Yonta,** *University of Yaoundé, Cameroun / IRD, UMMISCO, F-9314 Bondy, France*
- **Mathieu Roche,** *CIRAD, UMR TETIS, F-34398 Montpellier, France*

*Version 1 – January 2025*

# DAAfrica'2024: Data Science for Agriculture in Africa *

**Workshop affiliated with CARI'2024,**

**November 23, 2024, Bejaia, Algeria (hybrid)**

Paulin Melatagia Yonta[1,2][0000−0003−3479−2627] and Mathieu Roche[3,4][0000−0003−3272−8568]

[1] University of Yaoundé I, Cameroun
[2] IRD, UMMISCO, F-93143, Bondy, France
[3] CIRAD, UMR TETIS, F-34398 Montpellier, France
[4] TETIS, Université de Montpellier, AgroParisTech, CIRAD, INRAE, Montpellier, France

## Scope

Data science in agriculture has evolved with the increasing accessibility of data to farmers, which allows them to analyze and make decisions. Today, new technologies such as the Internet of Things (IoT) enable the collection and storage of farm and environmental data (e.g., soil data and water data) in dedicated databases and/or data warehouses. These agricultural data can be combined with other data sources (e.g., remote sensing, weather stations, and social media), highlighting the need to address new challenges, such as the use of heterogeneous data.

Data science in agriculture aims to explore and mine agricultural data via different techniques, such as machine learning, deep learning, computer vision, text mining, and large language models (LLMs). For example, data science can be used to predict crop yields and plant and animal diseases with different variables, including rainfall, temperature fluctuations, and soil conditions, by using a variety of data sources (e.g., sensor data, texts, satellite images, and plant images).

Therefore, agriculture professionals and decision-makers can use data science to obtain information and knowledge on which they can base decisions about agricultural activities in Africa.

During the DAAfrica'2024 workshop that brought around 50 attendees, nine abstracts were submitted and seven short papers are finally published in this proceedings volume.

* Supported by ASDS (https://asds.africa/), #DigitAg (https://www.hdigitag.fr/) and the MOOD project (https://mood-h2020.eu/)

## Topics of interest

The topics of the workshop encompass all aspects concerning the intersection of data science and agriculture in Africa with different applications:

- Smart Farming
- Yield and production
- Plant species identification
- Land cover monitoring
- Crop recommendation
- Crop monitoring and forecasting
- Monitoring of animal and plant health
- Agroecology and water management
- Food safety and security
- etc.

## Papers in these proceedings



## Accepted extended abstracts

- *Kinyua Gikunda and Nicolas Jouandeau*, Optimized Lightweight Deep Learning Techniques for Precise Weather Prediction and Forecasting in Africa.
- *Marouane Kihal, Lamia Hamza and Mohammed Charif Kihal*, Advanced Binary Classification for Disease Detection in Trees Using a novel Machine-Deep Learning method.

- *Djènaton Kpadonou Ambroise Houedjissin, Sèmèvo Arnaud R. M. Ahouandjinou, Manhougbé Probus A. F. Kiki, François-Xavier Ametepe and Marc Kokou Assogba*, Lightweight segmentation of UAV images for early detection of maize leaf diseases.
- *Agouanet Franklin Platini and Tewa Jeanjules*, Effectiveness of Image Processing Techniques in the Cultural Control of Fungal Plant Leaf Diseases.
- *Belayneh Dejene, Gizachew Setegn and Selamawit Belay*, Explainable and Interpretable Dry Beans Classification using Soft Voting Classifier.
- *Akouyo Yvette Gbedevi, Kossi Atchonouglo, Marie-Ange Manier and Sid Lamrous*, Crop Rotation Problem, The significance of the planning horizon duration and Fertilization with plants.
- *Andé Mbairanodji, Founadoudou, Daquin Cédric Awouafack and Garba Bouba Boiga*, Towards the digitalization of Cameroonian agriculture: current situation, challenges and prospects.
- *DVinablo Kodjo Doinique Dagbelou, Souleymane Bah, A. Souleymane A. Adekambi and Jacob A. Yabi*, Bibliometric and economic analysis in precision agriculture.
- *Moses Muga and Ruth Njoroge*, Seed Pelleting Technology on Spider Plant Seed.

## Scientific Committee

- Paulin Melatagia Yonta, University of Yaounde I, Cameroun – *co-chair*
- Mathieu Roche, CIRAD, TETIS, Montpellier, France – *co-chair*
- Marcellin Atemkeng, Rhodes University, South Africa
- Sandrine Auzoux, CIRAD, AIDA, Saint-Denis, La Réunion, France
- Roberto Interdonato, CIRAD, TETIS, Montpellier, France
- Julien Sarron, CIRAD, HortSys, Madagascar
- Mehtab Alam Syed, CIRAD, TETIS, Montpellier, France
- Maguelonne Teisseire, INRAE, TETIS, Montpellier, France

## Acknowledgements

This workshop was supported by The **African Society in Digital Science (ASDS)** - https://asds.africa/.

This workshop was partially funded by the French National Research Agency under the Investments for the Future Programme #DigitAg, referred to as ANR-16-CONV-0004 and the EU grant 874850 MOOD.

# Lightweight Deep Learning for Weather Prediction and Forecasting in Africa

Kinyua Gikunda[1][0000−0001−7962−2168] and Nicolas Jouandeau[2][0000−0001−6902−4324]

[1] Dedan Kimathi University of Technology, Nyeri, Kenya
patrick.gikunda@dkut.ac.ke
[2] Université Paris 8, Vincennes - Saint-Denis
[3] n@up8.edu

**Abstract.** Weather forecasting in Africa is hampered by sparse meteorological data and limited computational resources. This paper addresses these challenges by proposing lightweight deep learning (DL) for weather prediction and forecasting. We integrate active learning and transfer learning methods to enhance model training efficiency and accuracy. By focusing on the informativeness and representativeness of training samples, our approach significantly reduces the need for extensive and costly labeling. After training on a source dataset, model skills are transferred to target datasets, allowing for effective weather variable predictions with minimal data. Extensive experiments on three weather datasets demonstrate that our hybrid Transfer Active Learning method achieves similar classification accuracy compared to existing methods, using only 20% of the training samples. This study highlights the potential of advanced DL techniques to improve weather forecasting in Africa, despite the constraints of data scarcity and limited computational infrastructure.

**Keywords:** Weather Forecasting · Deep Learning · Transfer Learning · Active Learning

## 1 Introduction

Weather forecasting is a critical component in managing and adapting to environmental changes, particularly in Africa [1]. The continent faces unique challenges due to its vast geographical diversity and limited availability of meteorological data. Many regions in Africa have sparse weather station networks, resulting in uneven and incomplete datasets [2]. Additionally, the computational resources required for advanced weather prediction models are often scarce, further complicating accurate forecasting efforts. These challenges necessitate innovative approaches that can leverage available data and computational resources efficiently. Deep learning (DL) combined with strategies like active learning and transfer learning offers promising solutions to enhance weather prediction and forecasting accuracy in Africa. By utilizing lightweight DL models, it is possible to achieve weather forecasts even in data-scarce and resource-constrained environments, ultimately aiding in better decision-making and resource management across the continent.

## 2 Deep Learning for Weather Prediction

The non-linear behavior of meteorological data poses significant challenges for weather prediction, even with state-of-the-art numerical models [3]. This complexity has led researchers to explore emerging Artificial Intelligence (AI) approaches, which have demonstrated impressive performance in various fields [4]. Traditional parametric models, such as linear models, struggle with meteorological data due to their limited expressive power and inability to stack linear operations for more abstract representations [5]. Non-parametric learners like Gaussian kernels offer flexibility but are hindered by their reliance on local generalization and the exponential growth of input dimensionality.

Deep Learning (DL) methods address these challenges by stacking multiple feature learning layers to form deep representations, enhancing both computational and statistical efficiency. Recent advancements have improved the representation of inputs with fewer parameters, allowing for effective feature learning using both labeled and unlabeled data. Transfer Learning (TL), a process within DL, leverages learned features to apply knowledge from one domain to another related domain, improving learning efficiency and effectiveness. This makes DL particularly suitable for complex and dynamic fields like weather prediction.

Deep learning methods, especially convolutional neural network (CNN)-based time series classifiers, have proven highly effective for extracting temporal and spatial features from spatio-temporal weather data [7]. These methods offer faster and more accurate predictions and can handle large, complex datasets from weather satellites and IoT devices [8]. Unlike traditional models, DL do not require extensive feature engineering, making them more adaptable and practical for weather forecasting applications.

The flexibility and robustness of DL approaches make them well-suited for the complexities of weather data, which often exhibit non-linear and chaotic behavior. DL models, leveraging distributed and sparse representations, can capture intricate data structures that traditional parametric and non-parametric models struggle to represent effectively. This capability is crucial for processing high-dimensional meteorological datasets, where capturing subtle patterns and correlations can significantly enhance prediction accuracy.

DL's superior feature learning capabilities allow for better representation and understanding of weather patterns, leading to improved prediction accuracy and reliability [9]. These techniques reduce the need for manual data preprocessing and feature extraction, streamlining the forecasting process. Moreover, DL methods excel at learning from vast amounts of data, continually improving predictive performance as more data becomes available. Their scalability ensures that forecasting systems remain efficient and effective even as data volumes grow, making DL particularly beneficial for weather forecasting.

## 3 Transfer Learning and Active Learning

To address the challenge of sparse training data in time series datasets, the proposed model incorporates two primary DL techniques: Transfer Learning and Active Learning.

TL allows the model to leverage pre-existing knowledge from a related source task and apply it to the target task. This technique enhances the model's ability to generalize and perform well even with limited data by re-using model skills. AL dynamically queries and selects the most informative samples to add to the training set. It uses labeled data to provide critical information about class labels or boundaries, while unlabeled data helps in understanding the base data distribution. This iterative process improves the efficiency of the learning process by focusing on the most useful data points.

Before delving into the specifics of these techniques, it is essential to define the Time Series Classification (TSC) problem.

**Definition 1.** *An univariate time series* $U_t = [x_1, x_2, ..., x_T]$ *is an ordered set of real values. The length of* $U_t$ *is equal to the number of observable time-points* $T$*.*

**Definition 2.** *A multivariate time series* $M_t = U_t^1, U_t^2, ...., U_t^n$ *consist of* $n$ *observations per time-point with* $U_t^i \in R^T$

**Definition 3.** *A dataset* $D = (X_1, Y_1), (X_2, Y_2), ..., (X_N, Y_N)$ *is a collection of pairs* $(X_i, Y_i)$ *where* $X_i$ *could either be Ut or Mt with* $Y_i$ *as its corresponding label. For a dataset containing* $K$ *classes, the label vector* $Y_i$ *is a vector of length* $K$ *where each element* $j \in [1, K]$ *is equal to 1 if the class of* $X_i$ *is* $j$ *and 0 otherwise.*

We can define Time Series Classification (TSC) as the task of mapping time-based inputs to a probability distribution over a set of labels. This can be mathematically represented by the following equation:

$$C_t = f(w * U_{t-l/2:t+l/2} + b) | \forall t \in 1, T \tag{1}$$

$C$ denotes the convolution result on a univariate time series $U_t$ of length $T$ with a filter $w$ of length $l$, a bias parameter $b$ and a non-linear function $f$. Applying several filters on a time series will result in a multivariate time series whose dimensions are equal to the number of filters used. Using the same filter values $w$ and $b$ in ConvNets its possible to find the results for all time stamps $t \in [1, T]$. This is possible by using weight sharing that enables the model to learn feature detectors that are invariant across the time array

## 4 Deep Transfer Active Learning

During target training, the model's parameters are initialized using weights from a previous task, represented as $\Theta \leftarrow \vartheta_\theta$. After initializing the weights, a forward pass through the model is performed using the function $f(\theta, x_i)$, which computes the output for an input $x_i$. The output is a vector of estimated probabilities for $x_i$ belonging to each class. The prediction loss is then computed using a cost function, such as the negative log likelihood. Using gradient descent, the weights are updated in a backward pass to propagate the error. This iterative process of forward pass followed by backpropagation updates the model's parameters to minimize the loss on the training data. During testing, the model is evaluated on unseen data. A forward pass is performed on the new input, followed by class prediction. The predicted class corresponds to the one with the highest probability. For this, categorical cross-entropy is applied as the loss function, denoted as:

$$L(y, \hat{y}) = -\sum_{i=1}^{N} y_i \log(\hat{y}_i) \tag{2}$$

where $y_i$ is the true label and $\hat{y}_i$ is the predicted probability for class $i$. This loss function helps to measure the performance of the classification model by comparing the predicted probabilities with the actual labels.

AL is used to select instances a model is most uncertain about to improve learning efficiency. In uncertainty sampling, the model aims to identify and learn from the most informative data points. Three primary metrics used to define uncertainty are least confidence, sample margin, and entropy. To take consideration of the entire output distribution, entropy is used as a metric which is defined as:

$$f_u(x) = \arg\max_i -\sum_i P(y_i|x_i) \log P(y_i|x_i) \tag{3}$$

Here, $P(y_i|x_i)$ is the posterior probability of instance $x_i$ belonging to class $i$. For binary classification, the most uncertain instances are those with nearly equal probabilities for both classes.

Besides uncertainty, considering the distribution of instances can enhance AL performance. Instance diversity helps in selecting the most representative samples, thus improving query performance and avoiding outliers.

The correlation measure assesses the pairwise similarities of instances. The informativeness of an instance is determined by its average similarity to its neighbors. For two instances $x_i$ and $x_j$, the correlation measure $f_c$ is defined as:

$$f_c(x) = \frac{1}{DU} \sum_{x_j \in DU/x_i} f_c(x_i, x_j) \tag{4}$$

The value of $f_c(x_i)$ represents the density of $x_i$ in the unlabeled set. Higher values indicate that an instance is closely related to others, while lower values suggest outliers, which should be avoided for labeling.

To select the most informative and representative samples, a heuristic combination of correlation and uncertainty measures is employed. The most effective instance to label can be expressed as:

$$\hat{x} = \arg\max_i (f_u(x) \cdot f_c(x)) \tag{5}$$

This approach ensures that the selected samples are both uncertain and representative, enhancing the learning process.

## 5 Results

Three datasets were used in the experiments namely: a) RAUS[4] dataset contains daily weather observations from various Australian weather stations for a period of 10 years, b) KenCentralMet (Kenya Meteorological Department[5] privately acquired daily weather observations covering Central Kenya for a period of 3 years from 2012-2014 ) and c) MeteoNet[6] a meteorological dataset developed and made available by the French national meteorological service. For each of the dataset, less than 20% of the labeled samples was used as the initial training set. We present comparison of the proposed DTAL method, as detailed in the previous section, against: i) Random selection of data samples to query, iii) QUIRE method inspired by the margin based active learning from the minimax viewpoint with emphasize on selecting unlabeled instances that are both informative and representative [10], iv) DFAL method that selects unlabeled samples with the smallest perturbation. The distance between a sample and its smallest adversarial example better approximates the original distance to the decision boundary [11], v) Core-Set non-uncertainty based AL method [12].

| | **RAUS** | | | **KenCentralMet** | | | **MeteoNet** | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\mathbb{P}$ | $\mathbb{R}$ | $\mathbb{A}$ | $\mathbb{P}$ | $\mathbb{R}$ | $\mathbb{A}$ | $\mathbb{P}$ | $\mathbb{R}$ | $\mathbb{A}$ |
| **Random** | 81 | 80 | 79 | 64 | 67 | 62 | 89 | 85 | 91 |
| **DTAL** | 80 | **85** | **85** | **68** | 64 | 67 | **91** | 90 | **93** |
| **QUIRE** | **89** | 84 | 81 | 67 | **68** | 67 | 87 | 88 | 86 |
| **DFAL** | 83 | 82 | 80 | 60 | 62 | 64 | **91** | 88 | **93** |
| **Core-Set** | 79 | 83 | 84 | 65 | 65 | **68** | 90 | **91** | 91 |

**Table 1.** Experimental results with Precision $\mathbb{P}$, Recall $\mathbb{R}$ and Accuracy $\mathbb{A}$.

Table 1 shows that DTAL generally outperforms a bit other methods (except QUIRE that is better with RAUS), especially in terms of precision and recall, demonstrating the effectiveness of the proposed hybrid strategy in selecting the most valuable training samples from the distribution. However, performance varies depending on the dataset, highlighting the importance of dataset characteristics in the efficacy of active learning methods and demonstrates that results can be equivalent even with less samples.

## 6 Conclusion

This paper demonstrates the efficacy of lightweight deep learning, integrating active and transfer learning, for weather prediction in Africa. Our hybrid Transfer

[4] https://www.kaggle.com/datasets/jsphyg/weather-dataset-rattle-package
[5] https://meteo.go.ke/
[6] https://www.kaggle.com/datasets/katerpillar/meteonet

Active Learning method significantly enhances forecasting accuracy with minimal data, using only small portion of the training samples compared to existing methods. Despite challenges of data scarcity and limited computational resources, our approach shows promise in providing good weather forecasts essential for effective decision-making and resource management in Africa. Future work will focus on refining these techniques and validating their practical benefits in real-world applications.

## References

1. Cooper, P. J., Dimes, J., Rao, K. P. C., Shapiro, B., Shiferaw, B., & Twomlow, S. (2008). Coping better with current climatic variability in the rain-fed farming systems of sub-Saharan Africa: an essential first step in adapting to future climate change?. Agriculture, ecosystems & environment, 126(1-2), 24-35.
2. Radeny, M., Desalegn, A., Mubiru, D., Kyazze, F., Mahoo, H., Recha, J., ... & Solomon, D. (2019). Indigenous knowledge for seasonal weather and climate forecasting across East Africa. Climatic Change, 156, 509-526.
3. Benavides Cesar, L., Amaro e Silva, R., Manso Callejo, M. Á., & Cira, C. I. (2022). Review on spatio-temporal solar forecasting methods driven by in situ measurements or their combination with satellite and numerical weather prediction (NWP) estimates. Energies, 15(12), 4341.
4. Das, M., & Ghosh, S. K. (2018). Data-driven approaches for meteorological time series prediction: a comparative study of the state-of-the-art computational intelligence techniques. Pattern Recognition Letters, 105, 155-164.
5. Cohen, N., Sharir, O., & Shashua, A. (2016, June). On the expressive power of deep learning: A tensor analysis. In Conference on learning theory (pp. 698-728). PMLR.
6. Langer, M., He, Z., Rahayu, W., & Xue, Y. (2020). Distributed training of deep learning models: A taxonomic perspective. IEEE Transactions on Parallel and Distributed Systems, 31(12), 2802-2818.
7. Torres, J. F., Hadjout, D., Sebaa, A., Martínez-Álvarez, F., & Troncoso, A. (2021). Deep learning for time series forecasting: a survey. Big Data, 9(1), 3-21.
8. Chen, L., Han, B., Wang, X., Zhao, J., Yang, W., & Yang, Z. (2023). Machine learning methods in weather and climate applications: A survey. Applied Sciences, 13(21), 12019.
9. Huang, G., Wang, Y., Ham, Y. G., Mu, B., Tao, W., & Xie, C. (2024). Toward a learnable climate model in the artificial intelligence era. Advances in Atmospheric Sciences, 1-8.
10. Huang, S. J., Jin, R., & Zhou, Z. H. (2010). Active learning by querying informative and representative examples. Advances in neural information processing systems.
11. Ducoffe, M., & Precioso, F. (2018). Adversarial active learning for deep networks: a margin based approach. arXiv preprint arXiv:1802.09841.
12. Sener, O., & Savarese, S. (2017). Active learning for convolutional neural networks: A core-set approach. arXiv preprint arXiv:1708.00489.

# Advanced Binary Classification for Disease Detection in Trees Using a novel Machine-Deep Learning method

KIHAL Marouane[1][0000−0002−6675−7087], HAMZA Lamia[1][0000−0002−5436−3099], and KIHAL Mohammed charif[2][0000−0002−2075−2491]

[1] Laboratory of Medical Informatics (LIMED), Faculty of Exact Sciences, University of Bejaia, 06000 Bejaia, Algeria
marouane.kihal@univ-bejaia.dz
lamia.hamza@univ-bejaia.dz
[2] Laboratory of electrical engineering and industrial electronics (L2EI), Faculty of Science and Technology, Department of Electrical Engineering, University of Jijel, 18000 Jijel, Algeria
mc.kihal@univ-jijel.dz

**Abstract.** Detecting plant health is crucial to prevent losses in the productivity and quality of agricultural products. This study focuses on identifying plant diseases through the visual examination of leaf patterns. Specifically, we aim to efficiently determine the health status (diseased or healthy) of lemon trees by analyzing the condition of their leaves using nine different machine learning algorithms optimized with a deep learning approach. Our experimental results demonstrate that this method achieves a high accuracy rate of 93%, surpassing other machine learning techniques. The integration of multiple machine learning algorithms followed by deep learning proves to be a promising solution for effective detection of tree diseases.

**Keywords:** Trees diseases detection · Binary detection · Machine learning · Deep Learning · Citrus fruit

## 1 Introduction

Before the advent of AI-based methods, early detection of disease in trees was often hampered by rudimentary, empirical methods. Observers often had to rely on visible signs such as changes in leaf color or obvious external symptoms, limiting the ability to identify diseases at an early stage of development. In addition, the diversity of diseases and the variability of forest environments made it difficult to implement uniform and reliable detection protocols. These challenges highlighted the urgent need for innovative solutions to improve the efficiency and accuracy of tree disease monitoring. Machine Learning (ML) in the field of tree disease detection involves using algorithms to analyze data such as images or sensory data to identify characteristic signs of disease. This approach enables computer systems to learn from data without being explicitly programmed, thus improving the accuracy and efficiency of diagnosis. On the other hand, Deep Learning (DL), an advanced branch of machine learning, uses artificial neural networks to perform complex recognition and classification tasks. In tree disease detection, deep learning enables in-depth analysis of high-resolution images, facilitating early detection of infection or structural damage thanks to its ability to extract significant features and patterns from large quantities of data. Thus, detecting plant health with ML and DL algorithms is crucial to prevent losses in yield and quality of agricultural products by examining visually observable patterns on plants, such as leaves, stems, and fruits.

In this paper, we aim at efficiently binary detection of the health status of lemon trees (diseased or healthy) from the state of the leaves using an approach based on deep learning optimization of nine machine learning algorithms. The main contribution of this article is to design the following:

- Binary detection of the health status of trees from the state of the leaves using an approach based on deep learning optimization of nine machine learning algorithms.
- Comparison of our approach with nine Machine Learning algorithms.
- Application of four different evaluation metrics to compare results.

The rest of this paper is organized as follows: Section 2 reviews related works. Section 3 introduces our proposed approach for detecting diseased trees. In Section 4, we evaluate our results. Finally, Section 5 concludes the paper and suggests potential directions for future research.

## 2 Related works

Numerous studies have focused on detecting diseases from the leaves of various plants. For tomato, Prajwala et al.[1] proposed a variation of the convolutional neural network model, LeNet, to detect and identify diseases in tomato leaves. For rice, Kawcher et al.[2] introduced a rice leaf disease detection system utilizing machine learning techniques. Additionally, a study on potatoes [3] suggests a model that employs pre-trained models for fine-tuning to extract relevant features from the dataset, followed by a logistic regression classifier. For lemon tree, Banni and Sksvmacet [4] proposed a model that utilises GLCM (Grey Level Co-Occurrence Matrix) algorithms for the detection of citrus leaf disease. However, this study was unable to obtain appropriate outcomes in order to classify the image data. This study yielded an accuracy rate of approximately 85.71%. Recently, more work has been based on machine learning and deep learning algorithms. Pramanik et al. [5] used Transfer Learning-based Deep Learning models, specifically DenseNet-201, ResNet-50, ResNet-152V2, and Xception, to classify lemon leaf diseases. Xception outperformed all other models in terms of accuracy, with 94.34%. Khattak et al. [6] suggested the use of a CNN model to distinguish between healthy fruits and leaves and those that have prevalent citrus diseases, including black spot, canker, scab, greening, and Melanose. This CNN model has a test accuracy of 94.55%. Hassam et al. [7] proposed a single-stream convolutional neural network architecture to identify illnesses in citrus fruits. The expanded citrus dataset (Citrus Fruits, Leaves, and Hybrid Datasets) were employed in the experiment, and the accuracy was 99.4%, 99.5%, and 99.7% respectively. However, the study reveals little redundant information in the collected deep features. Yuan [8] evaluated and compared two deep learning models, DenseNet and MobileNet, for the case study of lemon leaf image classification. This study indicated that MobileNet is more promising in practice. Islam et al. [9] used InceptionV3 and VGG16 deep learning models to classify diseases in citrus leaves, including melanoses, canker, scab, and black spot. InceptionV3 outperforms VGG16 in terms of accuracy. Despite numerous studies in the field of agronomy, no previous work has employed a comprehensive set of ML algorithms, including DL, to optimize the detection of tree diseases.

## 3 Our approach

In this Section, we propose the use of nine ML algorithms and DL techniques to enhance the detection of tree diseases from leaf images. Various machine learning techniques that we employed for this task, including:

1. **Ada Boost :** A technique of grouping a number of individual weak classifiers all together in a single powerful classifier.
2. **Logistic Regression :** A model that maximises the probability for a binary dependent variable.
3. **Decision Trees :** A technique that divides the data into subsets according to specific values of the input dimensions, it can reveal the patterns correlated with plant diseases.
4. **Random Forests :** A learning algorithm that builds many decision trees in the training process.
5. **Support Vector Machines :** A method aims to select the hyperplane that provides the maximum distance between classes of healthy and diseased leaves in the space of features.
6. **k-Nearest Neighbors :** Categorizes a leaf based on the results of a majority vote on the k nearest neighbors using distance.
7. **Naive Bayes :** Utilize the Bayesian model with strong (naive) hypothesis of feature's independence.
8. **Linear Discriminant Analysis :** A method aims to determine the best way of dividing the different classes.
9. **Extreme Gradient Boosting :** An optimized gradient boosting.

As mentioned the result of the nine ML algorithms will be passed by a deep learning model to make and optimize the finale decision as shown in Fig. 1.

## 4 Experimentation and results

In this Section, we will present the details and results of the experiments conducted on images of lemon trees. The images used for this study were obtained from the *Collection of Different Category of Leaf Images*[10]

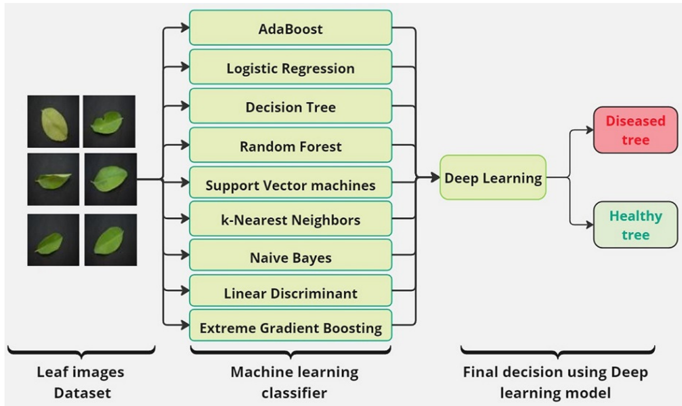



**Fig. 1.** General architecture of our proposed method

### 4.1 Methodology

We employed the nine machine learning algorithms discussed in the previous section, followed by implementing a deep learning model. Specifically, our DL model is a sequential model developed using Keras[11] consisting of five dense layers. The first dense layer has 64 units with a ReLU activation function, taking input data of dimension. The following layers have 128, 256, and 512 units, each also using the ReLU activation. The last dense layer has a single unit with a sigmoid activation function, suitable for binary classification tasks. The model is compiled with the 'adam' optimizer, 'binary crossentropy' loss function. Finally, the model is trained on data for 50 epochs with a batch size of 32.

### 4.2 Results

The result of experimentation are shown in Table 1. This outcomes demonstrate a clear comparison of various algorithms in terms of Accuracy, Precision, Recall, and F1-Score. Adaboost and SVM both achieve an accuracy of 85%, with Adaboost showing a high recall of 99% and an F1-Score of 90%, indicating strong performance in identifying true positives but a slightly lower precision of 82%. Logistic Regression and XGBoost both achieve an accuracy of 87.5%, with high precision 87% and recall 96%, resulting in an F1-Score of 92%, highlighting their balanced performance. Decision Trees and Random Forests show lower accuracy at 82.5% and 80% respectively, k-NN achieves also 82.5% of accuracy. Linear Discriminant Analysis achieves 85% accuracy, while Naive Bayes has the lowest accuracy at 75%, reflecting its limited effectiveness in this context. Notably, our proposed ML-DL approach outperforms all other algorithms with an accuracy of 93%, precision of 90%, recall of 99%, and an F1-Score of 95%, indicating superior overall performance in terms of both identifying true positives and minimizing false positives. Moreover, we trained the VGG16 model proposed by Islam et al. [9] on the same datasete used to train our model, the VGG16 model obtained about 85% in all evaluation metrics. Furthermore, the loss and accuracy curves show that the loss steadily decreases while the accuracy consistently increases until reaching 93%, indicating good model learning, as shown in Fig. 2.

| Algorithm | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| Adaptative boosting | 85% | 82% | 99% | 90% |
| Logistic Regression | 87,5% | 87% | 96% | 92% |
| Decision Trees | 82,5% | 86% | 89% | 88% |
| Random Forests | 80% | 78% | 99% | 88% |
| Support Vector Machines | 85% | 82% | 99% | 90% |
| k-Nearest Neighbor | 82,5% | 80% | 99% | 89% |
| Linear Discriminant Analysis | 85% | 87% | 93% | 90% |
| Naive Bayes | 75% | 76% | 93% | 84% |
| eXtreme Gradient Boosting | 87,5% | 87% | 96% | 92% |
| Islam et al. [9] | 85% | 85% | 85% | 85% |
| **Our approach** | **93%** | **90%** | **99%** | **95%** |

**Table 1.** Performance of Different Algorithms

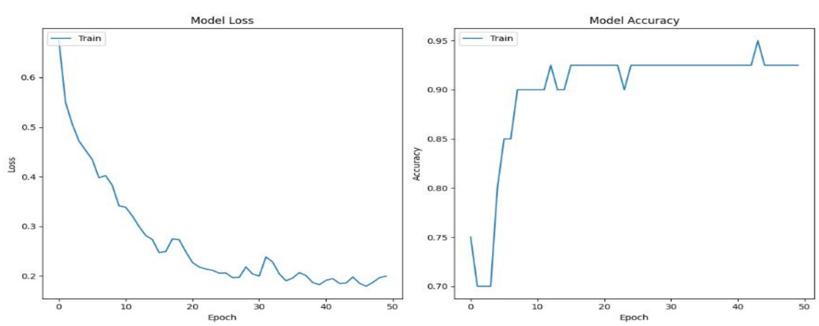



**Fig. 2.** Loss and accuracy model curves

## 5 Conclusion

In this paper, we have proposed a new approach based on machine learning followed by deep learning to efficiently detect the health status of tree leaves, using nine powerful machine learning algorithms, namely adaboost, logistic regression, decision tree, random forest, support vector machines, k-nearest neighbors, naive bayes, linear discriminant analysis, and extreme gradient boosting. The results presented in the experiment demonstrate that the proposed model outperformed the individual machine learning algorithms on four evaluation measures, achieving accuracy of 93%, precision of 90%, recall of 99%, and an F1 measure of 95%. These results indicate the effectiveness and robustness of the proposed approach, which can be used as an effective solution for tree disease control. Our future research will explore the generalization of this approach to other domains and datasets.

## 6 Aknowledgment

This work has been sponsored by the General Directorate for Scientific Research and Technological Development, Ministry of Higher Education and Scientific Research (DGRSDT), Algeria.

## References

1. P. Tm, A. Pranathi, K. SaiAshritha, N. B. Chittaragi and S. G. Koolagudi: Tomato Leaf Disease Detection Using Convolutional Neural Networks, In: Eleventh International Conference on Contemporary Computing (IC3), Noida, India, pp. 1-5 (2018) doi: 10.1109/IC3.2018.8530532.
2. K. Ahmed, T. R. Shahidi, S. M. Irfanul Alam and S. Momen: Rice Leaf Disease Detection Using Machine Learning Techniques. In: International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh, pp. 1-5, (2019) doi: 10.1109/STI47673.2019.9068096
3. D. Tiwari, M. Ashish, N. Gangwar, A. Sharma, S. Patel and S. Bhardwaj: Potato Leaf Diseases Detection Using Deep Learning. In: 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2020, pp. 461-466, (2020) doi: 10.1109/ICICCS48265.2020.9121067.
4. R. Banni and L. Sksvmacet: Citrus leaf disease detection using image processing approaches. International Journal of Pure and Applied Mathematics, 2018, vol. 120, no 6, pp. 727-735.
5. A. Pramanik, A. Z. Khan, A. A. Biswas and M. Rahman: Lemon leaf disease classification using CNN-based architectures with transfer learning. In : International Conference on Computing Communication and Networking Technologies (ICCCNT), Khargpur, India, pp. 1-6, (2021) doi: 10.1109/ICCCNT51525.2021.9579586
6. A. Khattak, M. U. Asghar, U. Batool, M. Z. Asghar, H. Ullah, M. Al-Rakhami and A. Gumaei: Automatic detection of citrus fruit and leaves diseases using deep neural network model, IEEE access, 2021, vol. 9, pp. 112942-112954.
7. M. Hassam, M. A. Khan, A. Armghan, S. A. Althubiti, M. Alhaisoni, et al: A single stream modified mobilenet v2 and whale controlled entropy based optimization framework for citrus fruit diseases recognition. IEEE access, 2022, vol. 10, pp. 91828-91839.
8. Y. Yuan, Computer Vision and Deep Learning for Precise Agriculture: A Case Study of Lemon Leaf Image Classification. Journal of Physics: Conference Series, 2023, vol. 2547, No 1, pp. 012024.
9. S. F. Islam, N. Chakrabarty, and M. S. Uddin: Citrus leaf disease detection through deep learning approach. Bulletin of Electrical Engineering and Informatics, 2024, vol.13, no 4, pp. 2643-2649.

10. M. Nagadia: Collection of Different Category of Leaf Images. Available: https://www.kaggle.com/datasets/meetnagadia/collection-of-different-category-of-leaf-images, last accessed 2024/07/25.
11. F. Chollet and others: Keras. GitHub repository, 2015. Available: https://github.com/fchollet/keras,last accessed 2024/07/25.

# Lightweight segmentation of UAV images for early detection of maize leaf diseases.

*Ambroise D. K. Houedjissin [1], Arnaud Ahouandjinou[2], Manhougbé Probus A. F. KIKI [1], François Xavier Amétépé [2], and Kokou M. Assogba[1]*

[1] Ecole Doctorale des Sciences de l'Ingénieur
University of Abomey-Calavi, Abomey-Calavi, Benin

[2] Institut de Formation et de Recherche en Informatique,
University of Abomey-Calavi, Abomey-Calavi, Benin

abhouedjissin@gmail.com

**Abstract.** Unmanned Aerial Vehicles (UAVs) equiped with RGB cameras have emerged as effective tools for monitoring agricultural crops. However, motion blur in UAV images can affect the accuracy of subsequent image analysis tasks, such as disease detection in plant leaves. This study proposes a real-time image segmentation approach for analyzing UAV-captured maize leaf images. The algorithm evaluates image blur using the Laplacian variance, applies an adaptive Wiener filter for deblurring, segments maize leaves from the background using color transformations, and identifies diseased regions through Canny edge and contour detection. Experimental results demonstrate the lightweight and effectiveness of proposed approach with less than 1s runtime, improving image quality and allowing accurate disease detection of maize crops for real-time purpose.

**Keywords:** UAV, RGB images, motion blur, maize diseases detection, segmentation.

## 1 Introduction

Plant disease detection is a key application of UAVs and has been extensively researched [1]. One of the advantages of using UAVs is its ability to detect diseases early and prevent their spread, thereby reducing crop losses [2]. Decision-support systems that incorporate UAVs can lead to better decision-making, increased production, improved product quality, and labor savings [3]. UAVs are utilized across various crop types and for detecting multiple diseases. Some diseases present visible symptoms, while others require temperature measurements for detection [4]. Early detection of pests and crop diseases provides farmers and other stakeholders with enough time to prevent potential epidemics and minimize yield losses. However, motion blur in UAV images generally caused by the camera movement during image capture, the combined effects of atmospheric turbulence, the shaking of the UAV platforms, high altitude or operation errors can affect the accuracy of subsequent image analysis tasks, such as disease detection in crop plant leaves [5]. This represents a common issue in UAV imagery and various methods have been proposed to address motion blur.

On the other hand, recent advancements in deep learning (DL) have produced various methods for detecting and classifying plant diseases using images of infected plants [6]. However, they require huge datasets for advanced approaches such as CNN to produce good results and large image datasets result in increased accuracy rates [7].

This study presents a real-time algorithm for the segmentation of UAV images, specifically targeting the detection of maize plant leaf diseases. The proposed method leverages motion blur detection, adaptive Wiener filtering, color conversion combined with morphological operations, Canny edge detection, Otsu color thresholding and contour area detection so that to isolate infected regions. The results demonstrate the algorithm's efficacy in identifying unhealthy plant areas in less than 1 second runtime, thereby providing a robust tool for precision agriculture in real-time. The rest of the paper is organized as follows : in section 2, we present the related works, section 3 outlines the proposed approach, the experimentation is introduced in section 4 and results and discussion are presented in section 5. Finally, section 6 provides a conclusion.

## 2 Related works

Successful disease estimation have been demonstrated in many UAV-based imagery applications such as [8], [9], [10]. These studies often used either the mean value of the vegetation index or the count of pixels below a certain threshold within a plot to estimate the disease score.

Table 1. gives a synthetic comparative analysis of classical image classification techniques in plant leaves healthy and unhealthy area detection.

**Table 1.** Comparison of classical image classification techniques.

| Classification Techniques | Description | Advantages | Limits | References |
|---|---|---|---|---|
| Color thresholding | Compares the distribution of colors in an image using a threshold. | - Simple to implement<br>- Fast computation | - Limited discriminative power<br>- Sensitive to changes in lighting conditions | [11] |
| Texture Analysis | Analyzing textural patterns in an image to characterize healthy and unhealthy areas. | - Captures subtle differences in texture<br>- Robust against changes in lighting and color | - Parameter tuning required<br>- May be computationally intensive | [12] |
| Machine Learning Models | Utilizing machine learning algorithms (e.g., SVM, Random Forest, CNN) to learn features and classify healthy and unhealthy areas. | - High accuracy and robustness<br>- Can automatically learn complex patterns | - Requires large amounts of labeled data<br>- Training and inference can be computationally costly | [13] |

According to Table 1. the choice of classification algorithm depends on various factors, including the desired level of accuracy, computational resources, and the availability of labeled data. Color thresholding is simple and efficient but may lack the discriminative power of more complex methods like texture analysis and machine learning models. Texture analysis can capture subtle differences in texture but may require more computational resources. Machine learning models offer high accuracy but come with higher complexity and resource requirements, especially during training.

## 3 Proposed Approach

Figure 1 describes the proposed approach.

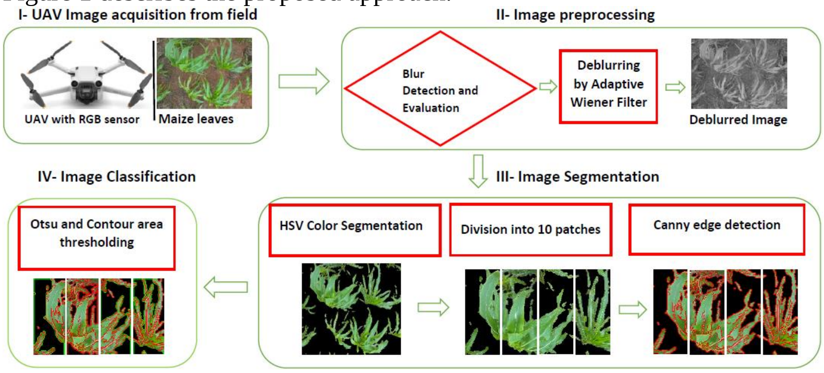



**Fig1.** Scheme of the proposed method.

The proposed algorithm involves several key steps and each step is based on specific mathematical operations and image processing techniques to ensure accurate segmentation and detection.

### 3.1 UAV Image Acquisition from field

First, UAVs equipped with high-resolution RGB cameras capture images of maize fields. The UAV images were collected between 4:45 p.m. and 6:00 p.m., on July 14, 2024, whereas it was sunny and windless. The DJI Mini 3 Pro, a quadcopter UAV system, was used to collect aerial images of maize leaves from a field. This system carried an automated RGB sensor (Quad Bayer CMOS camera) which was developed for agricultural applications.

**Image Preprocessing.** The second step in the algorithm involves preprocessing the UAV-captured images to enhance their quality and reduce noise. This is crucial for improving the accuracy of subsequent image analysis steps.

*Motion Blur Detection and Assessing.* The Laplacian variance [14] is used to detect motion blur. When detected, an adaptive Wiener filter is applied to deblur the images. To assess the degree of motion blur in the UAV images, we calculate the Laplacian variance of the grayscale image :

$$Laplacian(I) = \sum_{i,j}\left(\frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}\right) \quad (1)$$

$$Variance = \frac{1}{N}\sum_{i,j}(Laplacian(i,j) - \mu)^2 \quad (2)$$

Where $I$ is the grayscale image, $\mu$ is the mean of the Laplacian image, and $N$ is the total number of pixels.

The variance provides a measure of image sharpness, with lower values indicating higher blur levels.

*Image Deblurring by Adaptive Wiener Filter.* If the image is found to be blurred, an adaptive Wiener filter is applied to reduce noise and enhance details. The Wiener filter operates as follows:

*Normalize grayscale image:*

$$I_{norm} = \frac{I_{gray}}{255.0} \tag{3}$$

*Create averaging kernel:*

$$K = \frac{1}{k^2} 1_{k\times k} \tag{4}$$

*Compute local mean and variance:*

$$\mu_{local} = convolve2d(I_{norm}, K, 'same') \tag{5}$$

*Compute overall variance:*

$$\sigma^2_{local} = convolve2d(I^2_{norm}, K, 'same') - \mu^2_{local} \tag{6}$$

*Apply Wiener filter*

$$I_{wiener} = (I_{norm} - \mu_{local}) \cdot \left(\frac{\sigma^2_{overall}}{\sigma^2_{local}+\sigma^2_{overall}}\right) + \mu_{local} \tag{7}$$

*Denormalize:*

$$I_{deblurred} = I_{wiener} \cdot 255 \tag{8}$$

**Image Segmentation.** The third step in the algorithm relates to segmentation which plays a crucial role in identifying regions of interest within UAV images, such as healthy and infected areas of maize plant leaves, for further analysis.

*HSV Color Segmentation.* The deblurred image is then converted to the HSV (Hue, Saturation, Value) color space to facilitate segmentation based on color characteristics [15]. The green color range corresponding to healthy maize leaves is defined in the HSV space, and a binary mask is created to isolate these regions:

$$Mask = \begin{cases} 1 & if\ range1 \leq HSV(x,y) \leq range2 \\ 0 & otherwise \end{cases} \tag{9}$$

where $range1$ and $range2$ define the HSV range for green color.

Morphological operations, including opening and closing, are applied to the binary mask to remove noise and smooth the segmented regions.

$$Morphology(I) = (I \odot K) \oplus K \tag{10}$$

where $\odot$ denotes morphological opening and $\oplus$ denotes closing, and $K$ is the structuring element.

*Division into Patch.* The preprocessed image is divided into a specified number of equal-sized patches for localized analysis of disease symptoms. The number of patches here is 10.

Calculate patch dimensions :

$$h_p = \frac{H}{2} \quad (11)\ ; \qquad w_p = \frac{W}{5} \quad (12)$$

where $H$ is the height and $W$ is the width of the patch

*Canny Edge Detection.* Each patch undergoes further analysis to detect and classify objects of interest. The edges of potential diseased regions are detected using the Canny edge detection algorithm [16], which identifies boundaries based on gradients in the image :

$$Edges(x,y) = \begin{cases} 1 & if\ G(x,y) > threshold \\ 0 & otherwise \end{cases} \tag{13}$$

where G(x,y) is the gradient magnitude at pixel (x,y).

**Image Classification**. The detected edges are used to identify contours, which are then classified as healthy or unhealthy based on their area. A threshold is applied to differentiate between large healthy regions and smaller unhealthy spots. Contours are then extracted and classified based on their area, with larger areas typically indicating healthy regions and smaller areas potentially indicating diseased regions.

Following equations explain the classification process :

$$Area = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} I_{ij} \tag{14}$$

where $I_{ij}$ is the pixel value within a contour.

$$Classification = \begin{cases} Healthy & if\ Area > A_{threshold} \\ Unhealthy & otherwise \end{cases} \tag{15}$$

where $A_{threshold}$ is fixed here to 500.

Below is the detailed algorithm :

***Algorithm 1***: Lightweight Segmentation for maize leave disease detection

**Input**: x,y, $I_{(x, y)}$, ctg(), clv(), avf(), rhsv(), sgr(), mo(), dpp(), ced(), fdc() --image of leaves
Begin
Initialize Image Processing
**Output**: Result -- Segmented and classified plant regions
$I_{(x, y)} \leftarrow$ ctg($I_{(x, y)}$)) -- *Convert the UAV image to a grayscale image*
$\sigma^2 \leftarrow$ clv($I_{(x, y)}$) -- *Calculate the Laplacian variance to assess image sharpness*
**if** $\sigma^2$ < Threshold **then**
    $I_{(x, y)} \leftarrow$ avf($I_{(x, y)}$) -- *Apply the adaptive Wiener filter if Variance < threshold*
**end if**
$I_{HSV} \leftarrow$ rhsv($I_{(x, y)}$) --*Resize the image and convert it to the HSV color space*
$I_{Leaves} \leftarrow$ sgr($I_{HSV}$) -- *Segment green plant regions using predefined HSV ranges*
$I_{mo} \leftarrow$mo($I_{Morph}$--*Apply morphological operations to enhance segmentation*
Patch [ ] $\leftarrow$ dpp($I_{mo}$) -- *Divide the processed image into smaller patches*
**for each** patch **in** Patch **do**
    Edge [ ] $\leftarrow$ ced(patch) -- *Canny edge detection of each patch*
    Cntrs [ ] $\leftarrow$ fdc(Edge [ ]) -- *Classify detected edges based on health status*
**end For**

# 4 Experimentation

The characteristics of the UAV employed for the flight mission and the computer used for testing resulting aerial images are shown in Table 2 below.

## 4.1 Characteristics of UAV used for data acquisition

The mini-sized, mega-capable DJI Mini 3 Pro is just as powerful as it is portable. Weighing less than 249 g and with upgraded safety features, it is not only regulation-friendly but also the safest in its series [17]. With a 1/1.3-inch sensor and top-tier features, it redefines what it means to fly Mini.

Table 1. shows the specifications of the small UAV used for acquiring the images of the maize plants leaves used for constructing the dataset [18].

**Table 2.** Characteristics of the DJI Mini 3 Pro used for experimentation.

| Characteristics | Specifications |
|---|---|
| Model | DJI Mini 3 Pro |
| Weight | Under 250g |
| Camera | 1/1.3" (0.77") 48MP f1.7 Quad Bayer CMOS Sensor |
| Frame per second (fps) | 4K 60 fps with HDR; 1080p 120fps |
| Camera Orientation | Horizontal and Fully Vertical camera orientations |
| Sensors | Obstacle Avoidance Sensors |
| Flight Mode | Intelligent |
| Flight times | Up to 47 minutes (with the optional larger batter) or 25-30 minutes with the standard supplied batteries. |
| Controller | New Smart with 1080p 30fps |
| Maximum flight speed | 16 m/s |
| Maximum flight height | 4000m |
| Maximum horizontal range | 8KM |

## 4.2 Experimentation site localisation

The images were accessed on 14 July 2024 and acquired on board the UAV, in the village of Dodji-Sehe inside Sekou in the town of Allada, Benin Republic. Following Fig. 2 and Fig. 3 give aditionnal details on the site of study.

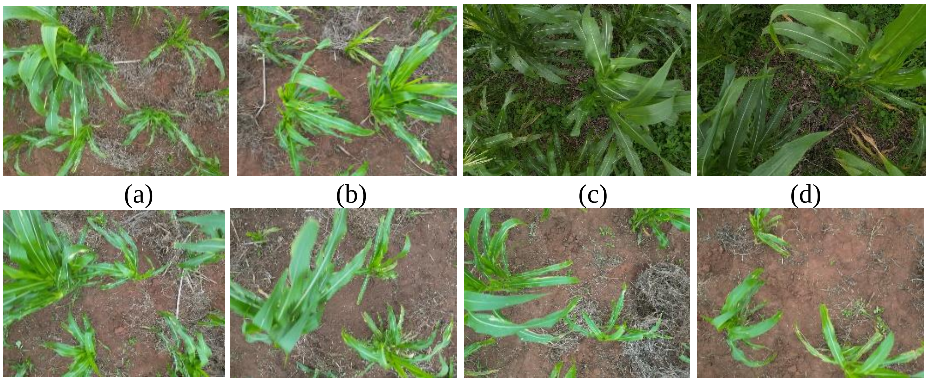
**Fig. 2.** Acquisition site localization, Benin Republic Map

**Fig. 3.** Dodji-Sèhè village, Sékou District

### 4.3 Dataset

The dataset consists of 266 images. When the images are captured in the state of stabilization of the device they are usually clear. On the other hand, the images captured during flight time are subject to motion blur.

The selected images were in the JPG file format and 4032 x 3024 pixels (see Figure 2).



(a) (b) (c) (d)

(e) (f) (g) (h)

**Fig. 2.** Example of UAV images of maize leaves from the dataset. Image from **(a)** to **(d)** appear to be sharp whereas those from **(e)** to **(h)** are motion-blurred.

## 5 Results and discussion

### 5.1 Preprocessing steps

Below we present a sample of image from the dataset with preprocessing steps as follows:


(a) (b) (c)

**Fig. 5.** (a) Original image, (b) Adaptive Wiener deblurred image of (a), (c) histogram of b

Table 2 indicates the performances characteristics of the deblurred image.

**Table 2.** Performances of the Adaptive Wiener Filter image deblurring step.

| Image Quality Index | Value |
|---|---|
| Image Entropy | 6.508706806116296 |
| MSE | 704.7281638886699 |
| PSNR | 19.65058732797596 |
| SSIM | 0.853058838351285 |

This deblurred image is characterized by an Entropy of 6,50 ; a Minimun Square Error of 704,72 ; a Peak Signal to Noise Ratio of 19,65 and a Structural Similarity Index Measure of 0,85, which indicates generation of an image of better quality.

### 5.2 HSV color segmentation

We perform here plant leaves segmentation using color transformations based on HSV color space combined with morphological operations. Figure 3.d) shows the result of extracted maize leaves from background.

**Fig. 6.** Original image of maize leaves

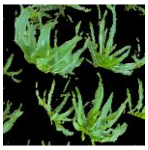


**Fig. 7.** Maize Leaves Extracted from background

We perform here plant leaves segmentation using color transformations based on HSV color space combined with morphological operations.

Color transformations provides a reliable means to segment maize leaves from the background, a critical step for accurate disease detection.

### 5.3 Division into patches and Canny Edge Detection

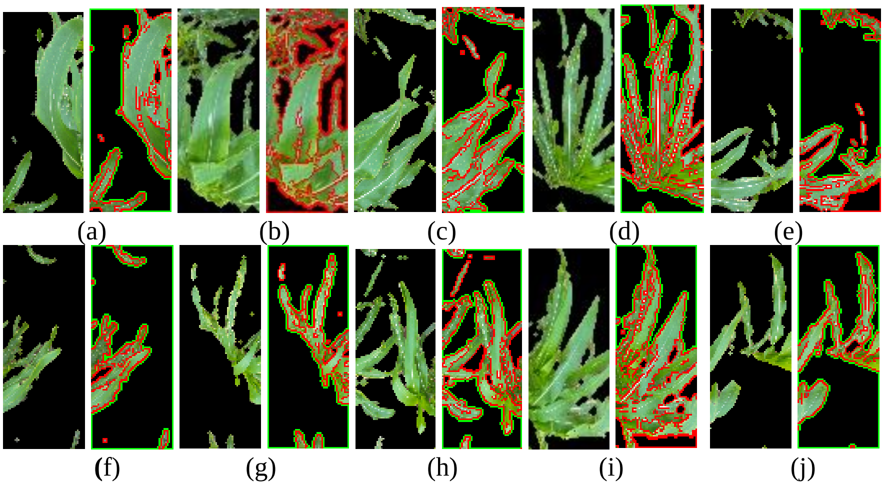



**Fig. 8.** Division into 10 patches and Canny Edge Detection

The preprocessed image is divided into a specified number of equal-sized patches for localized analysis of disease symptoms. The number of patches here is 10. Each patch undergoes Canny edge algorithm to detect objects of interest.

### 5.4 Health classification

Figure 5. below shows the health classification results generated for the 10 patches of the sample image used above.

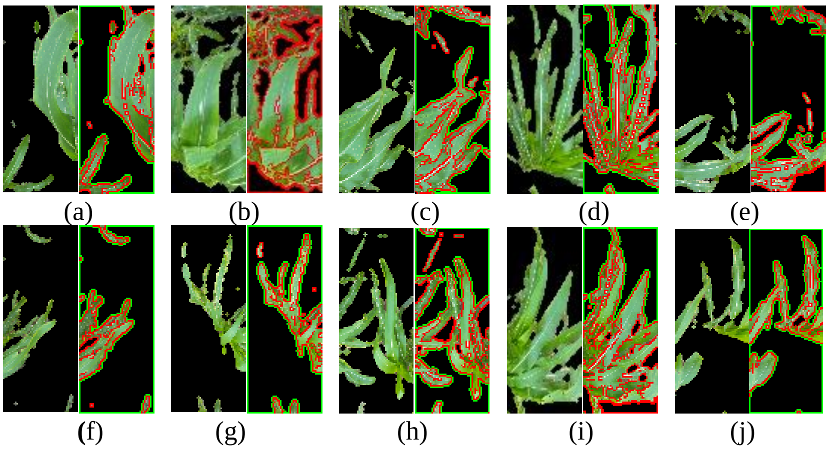



**Fig. 5.** Health classification results. For each pacth generated from a to j, note a = (Patch1and Res(Patch1)), j = (Patch10, Res(Patch10)) where Res(Patch1) is the result after Canny edge detetion and classification as healthy (green pixels) and unhealthy (red spots (non-green diseased areas) and double red contours for leaves damaged on edges.

Table 3. below presents the runtime of the program for a previously clear image.

**Table 3.** Performance Analysis of classification results of tested UAV maize leaves images.

| **Performance criteria** | **Value** |
|---|---|
| Runtime of the program | 0.863659143447876 seconds |

According to Table 3, the classification results are otained in 0,86 second less than 1 second runtime, which ensures low computation performance of the proposed approach, crucial condition for real-time application and decision-making.

The proposed approach addresses several key challenges in UAV-based crop monitoring. By evaluating and correcting image blur, we ensure that the subsequent segmentation and analysis steps are based on high-quality data. The use of adaptive Wiener filtering is particularly beneficial for real-time applications due to its efficiency and effectiveness in varying noise conditions. Color transformations provide a reliable means to segment maize leaves from the background, a critical step for accurate disease detection. The division of the image into patches allows for detailed localized analysis, making it possible to detect early signs of disease that might be missed in a full-image analysis. Canny edge detection and color thresholding leverages both structural and color information, enhancing speed and robustness of disease detection in 0,86 second.

## 6 Conclusion

This study presents a comprehensive real-time image processing algorithm for analyzing UAV-captured images of maize leaves. First, we leverage adaptive Wiener filtering to address image motion blur, then use color space transformation to segment maize leaves and finally patch-based analysis to detect diseased regions through a combination of edge detection and color analysis. The proposed method offers a promising solution for automated crop health monitoring, enabling timely interventions and improving agricultural productivity, forming a fast and effective tool for precision agriculture with less than 1 second. Future work will focus on optimizing the algorithm for different crop types and integrating it into a real-time UAV-based monitoring system.

**Acknowledgments.** This work was not supported by any funding. The first author would like to acknowledge the Benin Ministry of Higher Education and Scientific Research.

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## References

[1] F. Veroustraete, "The rise of the drones in agriculture," *EC agriculture*, vol. 2, no. 2, pp. 325–327, 2015.

[2] V. B. C. Calou, A. dos Santos Teixeira, L. C. J. Moreira, C. S. Lima, J. B. de Oliveira, and M. R. R. de Oliveira, "The use of UAVs in monitoring yellow sigatoka in banana," *biosystems engineering*, vol. 193, pp. 115–125, 2020.

[3] A. del-Campo-Sanchez *et al.*, "Quantifying the effect of Jacobiasca lybica pest on vineyards with UAVs by combining geometric and computer vision techniques," *PLoS One*, vol. 14, no. 4, p. e0215521, 2019.

[4] C. H. W. de Souza, R. A. C. Lamparelli, J. V. Rocha, and P. S. G. Magalhães, "Mapping skips in sugarcane fields using object-based analysis of unmanned aerial vehicle (UAV) images," *Computers and Electronics in Agriculture*, vol. 143, pp. 49–56, 2017.

[5] Z. Zhan, X. Yang, Y. Li, and C. Pang, "Video deblurring via motion compensation and adaptive information fusion," *Neurocomputing*, vol. 341, pp. 88–98, 2019.

[6] S. V. Militante, B. D. Gerardo, and N. V. Dionisio, "Plant leaf detection and disease recognition using deep learning," in *2019 IEEE Eurasia conference on IOT, communication and engineering (ECICE)*, IEEE, 2019, pp. 579–582. Accessed: Aug. 16, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8942686/

[7] S. Uchida, S. Ide, B. K. Iwana, and A. Zhu, "A further step to perfect accuracy by training CNN with larger data," in *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, IEEE, 2016, pp. 405–410. Accessed: Aug. 16, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7814098/

[8] S. Sarkar, A. F. Ramsey, A.-B. Cazenave, and M. Balota, "Peanut leaf wilting estimation from RGB color indices and logistic models," *Frontiers in plant science*, vol. 12, p. 658621, 2021.

[9] A. Guo *et al.*, "Wheat yellow rust detection using UAV-based hyperspectral technology," *Remote Sensing*, vol. 13, no. 1, p. 123, 2021.

[10] A. Patrick, S. Pelham, A. Culbreath, C. C. Holbrook, I. J. De Godoy, and C. Li, "High throughput phenotyping of tomato spot wilt disease in peanuts using unmanned aerial systems and multispectral imaging," *IEEE Instrumentation & Measurement Magazine*, vol. 20, no. 3, pp. 4–12, 2017.

[11] M. J. Swain and D. H. Ballard, "Color indexing," *Int J Comput Vision*, vol. 7, no. 1, pp. 11–32, Nov. 1991, doi: 10.1007/BF00130487.

[12] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Transactions on systems, man, and cybernetics*, no. 6, pp. 610–621, 1973.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[14] R. Bansal, G. Raj, and T. Choudhury, "Blur image detection using Laplacian operator and Open-CV," in *2016 International Conference System Modeling & Advancement in Research Trends (SMART)*, IEEE, 2016, pp. 63–67. Accessed: Aug. 17, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7894491/

[15] I. Patel, S. Patel, and A. Patel, "Analysis of various image preprocessing techniques for denoising of flower images," *International Journal of Computer Sciences and Engineering*, vol. 6, no. 5, pp. 1111–1117, 2018.

[16] J. Canny, "A computational approach to edge detection," *IEEE Transactions on pattern analysis and machine intelligence*, no. 6, pp. 679–698, 1986.

[17] "https://www.dji.com/global/mini-3-pro - Recherche Google." Accessed: Aug. 16, 2024. [Online]. Available: https://www.google.fr/search?q=https%3A%2F%2Fwww.dji.com%2Fglobal%2Fmini-3-pro&sca_esv=6306a503b65c9e12&sca_upv=1&ei=uxq_ZsOcI_HXhbIPnveN-yAM&ved=0ahUKEwiD5qKtmPmHAxXxa0EAHZ57AzkQ4dUDCA8&uact=5&oq=https%3A%2F%2Fwww.dji.com%2Fglobal%2Fmini-3-pro&gs_lp=Egxnd3Mtd2l6LXNlcnAiJWh0dHBzOi8vd3d3LmR-qaS5jb20vZ2xvYmFsL21pbmktMy1wcm9I_yNQAF-gAcAB4AJABAJgB4QGgAeEBqgEDMi0xuAEDyAEA-AEC-AEBmAIAoAIAmAMAkgcAoAdZ&sclient=gws-wiz-serp

[18] "https://dronegear.co.za/products/dji-mavic-mini-3-pro-with-smart-controller - Recherche Google." Accessed: Aug. 16, 2024. [Online].

# Explainable and Interpretable Dry Beans Classification using Soft Voting Classifier

Belayneh Dejene[1], Gizachew Setegn[2], Selamawit Belay [3]

[13]University of Gondar, Gondar, 196, Ethiopia

[2]Debark University, Debark, 90. Ethiopia

belzman2011@gmail.com

**Abstract.** Dry beans, integral to the Fabaceae family, boast global significance with their diverse genetic heritage tracing back to their dissemination from America centuries ago. This study endeavours to develop an explainable dry bean classification model using a soft voting classifier, juxtaposing its performance against classic and ensemble machine learning algorithms. Data preprocessing ensured suitability for classification algorithms, with feature selection employing information gain and variance inflation factors. The class imbalance was addressed via SMOTE + Tomek methods. Evaluation metrics encompassed accuracy, precision, recall, and F1-score. XGBoost led with 92.5065% accuracy, while soft voting classifiers (LGBM, XGB, CatBoost, RF, and DT) closely followed at 92.691%. The soft voting classifier proved optimal for dry bean classification, aiding in model interpretation and decision-making processes.

**Keywords:** Classification, Dry bean, Explainable, Machine learning, voting classifier.

## 1. Introduction

Dry beans belong to the diverse Fabaceae family, sometimes referred to as Leguminosae, and they are the most important and the most produced pulse in the world [1]. It originated in America, while there is a wide genetic diversity in the world since, in the 15th and 16th centuries, they were transported to Europe and Africa and quickly spread to the rest of the globe[1]. The selection of dry beans plays an important role in the economy of agriculture-based countries like Bangladesh, India, Pakistan, etc. throughout the winter season. Currently, Dry bean is a staple food for many regions of the world and processing enables the consumption and incorporation of this nutrient-dense food in daily diets. Dry beans are the most known source of protein. In addition, they are low in fat and a rich source of fiber and other important nutrients [2][3]. Dry beans are important for environmental and human health benefits, such as improved soil fertility, reduced risk of chronic disease, and improved or promoted glycemic control [1]. There are several genetic diversities of dry beans, and it is the most produced one among the edible legume crops in the world. According to the Turkish Standards Institution, dry beans are classified as Barbunya, Battal, Bombay, Calı, Dermason, Horoz, Tombul, Selanik, and Seker” based on their botanical characteristics [4][5][6]. Plants are sensitive to the effect of climatic changes and they have a variety of resistance. Finding high-quality seed is the primary challenge facing dry bean producers and distributors or marketers. Using a lower quality seed in production will induce to lower quantity even if all the cultivation conditions are provided. A wide range of computational tools are available to regulate food and agricultural product quality. But most of them are done with the use of conventional techniques of the professionals. For example, different seed categorization is conducted based on human understanding, and determining the type of dry beans requires a skillful person to take a huge time manually, and passes a challenging process [6]. In particular, the color of various dry bean species varies, and geometrical data does not reveal this color variation. Due to this reason, it is vital in economically technical aspects to build an automated technique to detect as well as categorize seed features rapidly and repeatedly. Even, it is difficult for a human operator to understand or handle the seeds except for specific tools or automatic software procedures. The main problem dry bean producers and marketers face is in ascertaining good seed quality. Lower quality of seeds leads to lower quality of produce. Seed quality is the key to bean cultivation in terms of yield and disease. In today's world, the inspection of the quality of seeds, fruits, and vegetables along with the examination and categorization of seeds and grains have been performed worldwide to meet these demands with the help of machine learning and computer vision [4]. This is why we try to use a soft voting classifier and compare it with individual algorithms to classify dry beans. In recent years, machine learning algorithms have been used in the inspection, classification, prediction, and segmentation of food product quality. Classification techniques are becoming more popular in the fields of medicine, biostatistics, bioinformatics, agriculture, business, etc. as machine learning applications [7]. Machine learning is a subfield of artificial intelligence that enables computers to understand existing data and estimate the existence of unidentified targets. Seed quality is influential in crop production. Seed classification is important for both producers and marketers to provide the values of sustainable agricultural systems. By applying predictive analysis to agricultural data, significant decisions can be taken and classifications can be made.

Besides the classification model conducting explainability and interpretability of the classification model provide the professionals with insights into how the classifications are made, fostering trust in the model's decisions [8]. The explainable machine learning model impacts professionals more likely to trust and adopt understand and interpret the reasoning behind the model's recommendations by solving the black box nature of the algorithms. To handle this problem, several studies have been conducted to detect the quality of dry beans using various machine-learning techniques. For example, [4][5][6][7] conducted on dry bean classifications. The previous research on dry bean classification has largely neglected the crucial aspect of explainability and interpretability in their models. Instead, researchers predominantly focused on employing various algorithms without addressing the black box nature inherent in these methods. Classic machine learning approaches were commonly utilized, often with default parameter settings, despite evidence suggesting that optimizing these parameters could enhance classification performance [9]. Additionally, while some studies attempted to tackle class imbalance issues, they typically employed simplistic oversampling methods, which could lead to the generation of redundant data. Advanced techniques for addressing class imbalance were rarely explored. Furthermore, previous research overlooked feature selection methods, which could potentially improve model efficiency and interpretability. The absence of studies utilizing explainable techniques to handle black-box models, as well as the scarcity of research employing soft voting classifiers and tuned parameters, underscored the need for this study. Motivated by these gaps, this study endeavors to develop an explainable and interpretable classification model for dry beans. It seeks to utilize soft voting classifiers, a technique not extensively explored in previous research, and compare its performance with individual machine learning algorithms. By incorporating explainable and interpretable methods, this study aims to classify dry beans accurately while providing insights into the decision-making process, thus facilitating evidence-based policies and interventions in the selection of appropriate dry bean classes.

## 2. Related works

Several studies such as [4] [5][6] and [7], investigated the dry bean classifications using machine learning algorithms. However, most of the previous researchers didn't consider the explainability and the interpretability of the dry beans' classification model, most of these previous studies developed a classification model by handling the class imbalance problem on the whole data and developing the classification model without tuning relevant parameters. These studies did not conduct any feature selection methods, they developed the classification model by using all the features in the dataset. M. Koklu and I. A. Ozkan [4] develop multi-class dry bean classifiers using MLP, SVM, kNN, and DT, classification models. The overall correct classification rates have been determined as 91.73%, 93.13%, 87.92%, and 92.52% for MLP, SVM, kNN, and DT, respectively. The SVM classification model has the highest performance with the accuracy of the Barbunya, Bombay, Cali, Dermason, Horoz, Seker, and Sira bean varieties 92.36%, 100.00%, 95.03%, 94.36%, 94.92%, 94.67%, and 86.84%, respectively. However, this researcher didn't consider the explainability and the interpretability of the dry beans' classification model. G. Słowiński [5] tried to classify dry beans using machine learning techniques: Multinomial Bayes, Support Vector Machines, Decision Trees, Random Forests, and Voting Classifier. The overall accuracies obtained were in the range: of 88.35 - 93.61%. However, this researcher didn't consider the explainability and the interpretability of the dry beans' classification model. M. Salauddin Khan *et al* [7] aimed to construct a multiclass dry bean classification model using the eight most popular classifiers and compare their performances. The algorithms they used, were LR, NB, KNN, DT, RF, XGB, SVM, and MLP with balanced and imbalanced classes. The XGB classifier performed better than other classifiers with the balanced and imbalanced dataset of dry beans within each class. It performed an accuracy of 93.0% and 95.4% in imbalanced and balanced classes respectively. The overall performance is better than the previous studies, however, the researchers didn't consider the explainability and the interpretability of the dry beans' classification model. The researcher develops the model without tuning the parameters and developing the model without those parameters faces overfitting. Not only this but also, the researcher handles the class imbalance problems on the whole dataset before splitting it, and evaluating the model using those fabricated datasets.

## 3. Materials and Methods

### 3.1. Data collection methods

To conduct this study, we have used the publicly available dataset in the Kaggle repository. The extracted datasets consist of a total of 13,611 grains of 7 different registered dry beans with a total of 17 features including the class level (see table 1 here below for the dataset descriptions)

**Table 1.** Dataset descriptions

| No | Feature | Type | Description |
|---|---|---|---|
| 1 | Area | Integer | The area of a bean zone and the number of pixels within its boundaries |
| 2 | Perimeter | float | Bean circumference is defined as the length of its border |
| 3 | Major axis length | float | The distance between the ends of a dry bean can be drawn from a bean the longest line that |

| | | | |
|---|---|---|---|
| 4 | Minor axis length | float | The longest line that can be drawn from the bean while standing perpendicular to the main axis |
| 5 | Aspect ratio | float | Defines the relationship between L and l |
| 6 | Eccentricity | Real | The eccentricity of the ellipse having the same moments as the region |
| 7 | Convex area | Integer | Number of pixels in the smallest convex polygon that can contain the area of a bean seed |
| 8 | Equivalent diameter | float | The seed area diameter of a circle is the same area as a bean |
| 9 | Extent | float | The ratio of the pixels in the bounding box to the bean area |
| 10 | Solidity | float | The ratio of the pixels in the convex shell to those found in beans |
| 11 | Roundness | float | Calculated with the following formula |
| 12 | Compactness | float | Measures the roundness of an object |
| 13 | ShapeFactor1 | float | Shape factor 1 |
| 14 | ShapeFactor2 | float | Shape factor 2 |
| 15 | ShapeFactor3 | float | Shape factor 3 |
| 16 | ShapeFactor4 | float | Shape factor 4 |
| 17 | Class | Nominal | Target class of the dry bean |

### Data preprocessing methods

Data preparation involves data selection, data cleaning, data integration, feature selection, handling imbalances, and data transformation to make it available to extract value from those data [10][11]. In this subsection, we have detected the missing values, removed redundancies, detected outliers, and handled class imbalance problems from the dataset using statistical methods

### Data cleaning

This is a way of removing noise, inconsistencies, redundancy, and missing values to carefully develop the model. Without cleaning the collected data, we can’t get an accurate result [12][13]. In the dataset, there are no missing values, though we have not applied any methods to handle the missing values. From the data, we have removed 68 redundant records using drop redundant methods. Most of the variables have a higher proportion of outliers including Area, Perimeter, Minor Axis Length, Eccentricity, Convex Area, EquivDiameter, and ShapeFactor4. To handle this outlier, we have used interquartile range and boxplot methods.

### Data transformation

Where data are transformed and consolidated into forms appropriate for extracting by performing summary or aggregation operations. The data are transformed into forms appropriate for mining [14][15]. In these datasets, only the class level needs to be transformed for mining purposes, but all the remaining features don’t need to transform and we have used it as it is. To transform the class level, we have used the level encoding methods and encoded them into numeric values. We have encoded as ‘DERMASON’ = 0, ’SIRA’ = 1, ’SEKER’ = 2, ’HOROZ’ = 3, ’CALI’ = 4, ’BARBUNYA’ = 5, and ’BOMBAY’ = 6.

### Feature selection

In this method, we have checked the importance of all the features by using information gain, (see Fig 1 here below), from the 16 features the last three, features (ShapeFactor4, Solidity, and Extent) were the least important, but it is not mean that they are not valuable for the model. We have checked the multicollinearity of the feature using the variance inflation factor, and the variance inflation factor shows that all of the features were significant to the model. Due to this, we have not dropped them for their usefulness and we used all of the 16 features for developing the classification model.

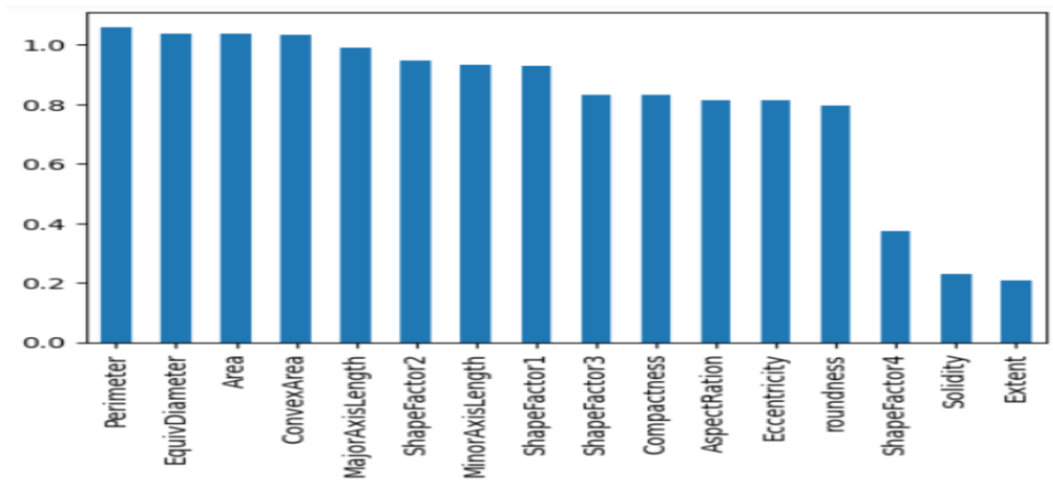



**Fig. 1.** Feature importance

### Handling class imbalance

By nature, the class level of the collected data is imbalanced see Figure 2 here below. To overcome the imbalanced class distributions problem, we can add samples to or remove samples from the data set [16]. Sampling can be achieved in two ways, Undersampling, randomly removing the majority class, oversampling the minority class, or by combining over and under-sampling techniques [16][17]. The extracted dataset class level has 7 values, from these values, some of them have the least values see Figure 2 here below. In the class distribution, the “BOMBAY” class has the least value when we compare it with other classes. To conduct this research, we used the synthetic minority over-sampling technique (SMOTE) + Tomek methods to handle the class imbalance of the class levels of the dataset. The main reason that we use SMOTE + Tomek is, it avoids the loss of valuable information [16][17]. In SMOTE + Tomek, the SMOTE combines the SMOTE ability to generate synthetic data for the minority class and the Tomek ability to remove the data that are identified as Tomek links from the majority class [18][19].

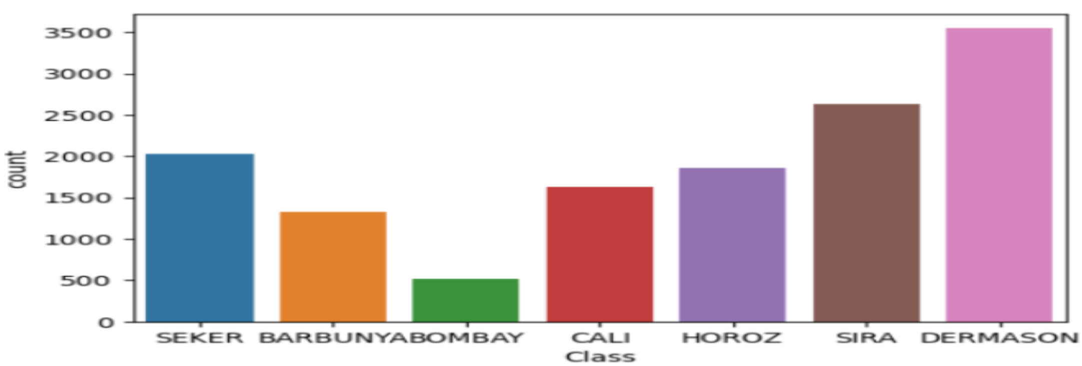



**Fig. 2.** Class imbalance

### Train test split

In model building, the researcher needs to develop datasets for training and testing to learn and evaluate the machine appropriately [20][21]. To conduct this study, we used the stratified splitting technique to split the whole dataset to train and test data and split the dataset into 80:20 train test ratios.

### Parameter tuning

In the process of machine learning and deep learning algorithms, the performance of the algorithm highly depends on the selection of hyperparameters, which has always been a crucial step in the process of machine learning [22][23][24]. To improve the performance rate for each algorithm a collection of hyperparameters has been tuned using grid search methods. Gird search is commonly used as an approach to hyper-parameter tuning that will methodologically build and evaluate a model for each combination of algorithm parameters specified in a grid [24]. Here, we used the grid-search with GridSearchCV for selecting tuning parameters for a homogeneous ensemble machine learning algorithm.

**Table 2.** Tuned parameters

| No | Algorithms | Parameters |
|---|---|---|
| 1 | Soft voting classifier | Default parameters |
| 2 | LGBM classifier | Default parameters |
| 3 | Random Forest | criterion='entropy',max_features='sqrt',min_samples_split=3,n_estimators=500,random_state=0,max_depth=20, max_leaf_nodes=400, n_jobs=-1 |
| 4 | Cat boost | random_state=42, learning_rate=0.1, l2_leaf_reg=4, iterations=600, depth= 6 |
| 5 | xgboost | random_state=42, verbosity=0, min_child_weight=2, max_depth=4, learning_rate=0.15, gamma=0.22, colsample_bytree=0.5 |
| 6 | Decision tree | max_depth=20, criterion='gini',max_features='sqrt',splitter='best', max_leaf_nodes=100 ,min_samples_split=3 |

### Classification model

In this study, to construct a dry bean classification model we have used a soft voting classifier in both the balanced and the unbalanced dataset. To compare that the soft voting classifier can perform better than other machine learning algorithms, another model was developed using decision tree algorithms and other ensemble learning classifiers namely random forest, catboost XGBoost, and LGBM classifiers. To improve each algorithm's performance rate, a collection of hyperparameters has been tuned using grid search methods. The performance of each classification model was evaluated using accuracy, precision, recall, and F1-score.

### Model explainability

To enhance the explainability of the classification model, we have employed various feature relevance explanation techniques like Local Interpretable Model-agnostic Explanations (LIME) and Shapley Additive Explanation (SHAP) to highlight the most influential features and regions in the input data, and to explain the quality of the inner functioning of deep learning models and decisions by calculating the influence of each input variable and producing relevant scores. Global interpretability techniques, such as feature importance analysis or rule extraction, are employed to reveal the underlying patterns and decision rules learned by the model [8].

## 4. Result and discussion

Experiments have been carried out to develop a dry bean classification model by using a soft voting classifier and comparing it with other classic and ensemble machine learning algorithms. To construct a classification model for dry beans, we conducted two experiments on the imbalance data and the balanced data using a soft voting classifier, RF, cat boost, XGB, LGBM, and DT. Each experiment was conducted using 16 features and by using all the tuned parameters using grid search (see Table 2). This experiment is multiclass classification because the dataset by nature has seven class levels. In these experiments, we evaluated all the classification models using accuracy, precision, recall, and f1_score evaluation metrics. Finally, we have explained the model using LIME and SHAP feature relevancy explanation techniques.

**Experiment# 1: Imbalanced dataset**

This experiment was conducted by using the imbalanced dataset or without applying any data imbalance handling methods. We have developed the model by using DT, RF, Catboost, XGB, LGBM, and a soft voting classifier. We have also evaluated those models' using accuracy, precision, recall, and f1_score (see Table 3 here below)

**Table 3.** Model performance using the imbalanced dataset

| Algorithms | Metrics | | | |
|---|---|---|---|---|
| | Accuracy | Precision | Recall | F1_score |
| Decision tree | 0.910668 | 0.920002 | 0.920301 | 0.920069 |
| Random forest | 0.923588 | 0.936018 | 0.933484 | 0.934621 |
| Cat boost | 0.927649 | 0.939536 | 0.939268 | 0.938353 |
| XGBoost | 0.928756 | 0.940888 | 0.938008 | 0.939343 |
| LGBM classifier | 0.92285 | 0.936619 | 0.935014 | 0.935765 |
| Soft voting classier (LGBM, cat boost, XGB, RF, DT) | 0.92691 | 0.940701 | 0.93913 | 0.939856 |
| Soft voting classier (cat boost, XGB) | 0.927649 | 0.939835 | 0.93874 | 0.93922 |
| Soft voting classier (LGBM, cat boost) | 0.925065 | 0.93902 | 0.936754 | 0.937815 |
| Soft voting classier (RF, DT) | 0.90993 | 0.920326 | 0.919321 | 0.919589 |

As we see from Table 3 above, the XGBoost algorithm outperforms the best result with accuracy precision, and f1_score of 0.928756%, 0.940888%, and 0.938008% respectively. But in the case of recall cat boost algorithm performs the best with 0.939268%. When we see the soft voting classifiers, the soft voting of the algorithms LGBM, cat boost, XGB, RF, and DT performs better than the soft voting with other algorithms. In the soft voting algorithms, the voting that contains cat boost and XGBoost algorithm performs a better result.

**Experiment# 2: balanced dataset**

This experiment is conducted by balancing the dataset using SOMTE + Tomek methods on the training set only and developing the model using DT, RF, Catboost, XGB, LGBM, and soft voting classifiers. We have also evaluated those models' using accuracy, precision, recall, and f1_score (see Table 4 below).

**Table 4.** Model performance using the balanced dataset

| Algorithms | Metrics |
|---|---|

| | Accuracy | Precision | Recall | F1_score |
|---|---|---|---|---|
| Decision tree | 0.921004 | 0.932629 | 0.933324 | 0.932835 |
| Random forest | 0.921004 | 0.932629 | 0.933324 | 0.932835 |
| Cat boost | 0.925434 | 0.9367 | 0.938996 | 0.937794 |
| XGBoost | 0.925065 | 0.936134 | 0.93809 | 0.937017 |
| LGBM classifier | 0.924695 | 0.937783 | 0.937861 | 0.937776 |
| Soft voting classier (LGBM, cat boost, XGB, RF, DT) | 0.926541 | 0.936982 | 0.938738 | 0.937805 |
| Soft voting classier (cat boost, XGB) | 0.925065 | 0.936067 | 0. 93803 | 0.93695 |
| Soft voting classier (LGBM, cat boost) | 0.926541 | 0.938991 | 0.940395 | 0.939642 |
| Soft voting classier (RF, DT) | 0.906238 | 0.916778 | 0.919568 | 0.917991 |

Finally, in this experiment developing the model by handling the imbalance problem is not always a good solution to get a better performance.

### Model comparison

As a result, the researcher compared the performance of algorithms to classify the dry bean using a soft voting classifier and other classic and ensemble machine learning algorithms using both imbalanced and balanced datasets. The dataset has seven classes. Then, the researcher used overall accuracy, precision, recall, and f1_score as an evaluation for classification model comparison. According to the overall performance, the classification algorithm that registered the highest performance is selected as the best algorithm for the classification model for the dry bean. As indicated in Table 3 and Table 4 above, the experiments are conducted on classification algorithms for classifying the dry bean. The XGB algorithms registered the highest accuracy of 92.8756% in the imbalanced dataset and the soft voting classifiers of the algorithm LGBM, cat boost, XGB, RF, and DT performed an accuracy of 92.6541% using the imbalanced datasets. The soft voting classifiers of LGBM, XGB, Cat boost, RF, and DT perform the best result next to XGBoost algorithms with overall accuracy, precision, recall, and f1_score of 92.691%, 94.0701%, 93.913%, and 93.986% respectively. The decision tree algorithm is registered with the lowest performance in both the imbalanced and the balanced datasets, see Table 3 and Table 4. Therefore, the XGBoost algorithm is selected as the best classifier as compared to other classic and ensemble machine learning algorithms, and the soft voting classifiers of LGBM, XGB, Cat boost, RF, and DT are selected as the best classifier where we compared with other voting classifiers.

### Model explainability

To enhance the explainability of the classification model, we have employed various techniques. We have explained and interpreted the classification model developed with each algorithm to make the trust of how it achieves the result. The explainable AI approach with LIME and SHAP frameworks is implemented to understand how the model predicts the final results. To explain the model, we have randomly selected the rows 100, 150, 200, 250, and 300 in the dataset. But this row was selected randomly and we can select any other rows in the dataset.

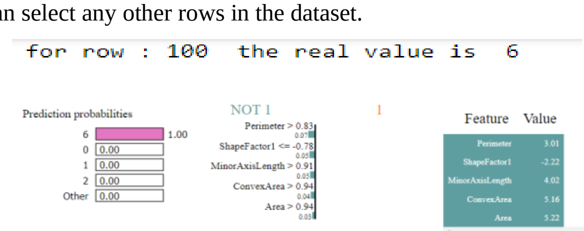



**Fig. 3.** Model Explanation with LIME for row 100

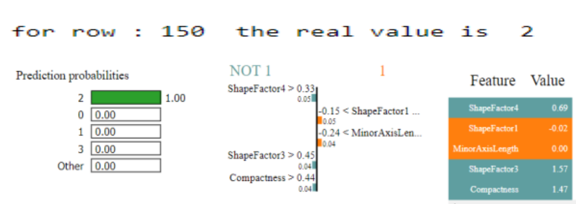
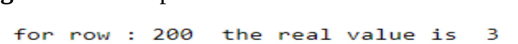



**Fig.5**. Model Explanation with LIME for row 150



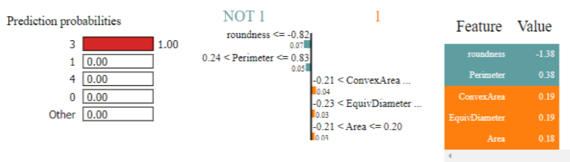



**Fig. 4.** Model Explanation with LIME for row 200

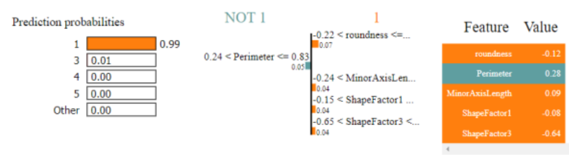



**Fig.6.** Model Explanation with LIME for row 250

```
for row : 300  the real value is  5
```

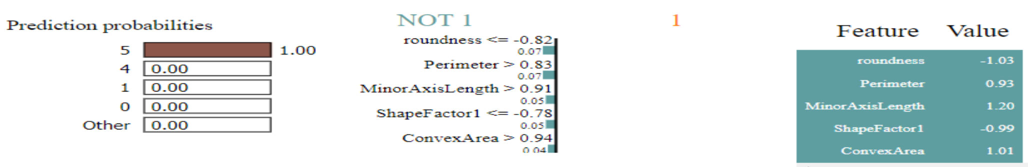



**Fig. 7.** Model Explanation with LIME for row 300

The figures 3, 4, 5, 6, and 7 above depict interpretations of an XGBoost model using the LIME explainable AI method for classifying specific types of dry beans. In each case, the model achieves 100% accuracy in classifying the beans into their respective classes. Here are the key findings from each interpretation:

**Class 'BOMBAY' (Figure 3):** The model identifies dry beans as 'BOMBAY' based on specific features such as perimeter, shape factors, minor axis length, convex area, and area. For instance, the beans are classified as 'BOMBAY' when perimeter > 0.83, ShapeFactor1 <= 0.78, MinorAxisLength > 0.91, Convex Area > 0.94, and Area > 0.94.

**Class 'SEKER' (Figure 4):** The model correctly classifies dry beans as 'SEKER' by considering features like shape factors, minor axis length, and compactness. For instance, beans are categorized as 'SEKER' when ShapeFactor4 > 0.33, ShapeFactor1 < -0.15, MinorAxisLength < -0.24, ShapeFactor3 > 0.45, and Compactness > 0.44.

**Class 'HOROZ' (Figure 5):** Dry beans are accurately classified as 'HOROZ' based on features such as roundness, perimeter, convex area, equivalent diameter, and area. For example, beans are classified as 'HOROZ' when roundness <= -0.82, Perimeter > 0.24 & <= 0.83, ConvexArea > -0.21 & <= 0.19, EquivDiameter > -0.23 & <= 0.19, and Area > -0.21 & <= 0.20.

**Class 'SIRA' (Figure 6):** The model identifies dry beans as 'SIRA' considering attributes like perimeter, roundness, minor axis length, shape factors, and shape factor 3. For instance, beans are classified as 'SIRA' when Perimeter > 0.24 & <= 0.83, roundness > -0.22 & <= -0.12, MinorAxisLength > -0.24 & <= 0.09, ShapeFactor1 > -0.15 & <= -0.08, and ShapeFactor3 > -0.65 & <= -0.64.

**Class 'BARBUNYA' (Figure 7):** Dry beans are correctly classified as 'BARBUNYA' based on features like roundness, perimeter, minor axis length, shape factor 1, and convex area. For example, beans are categorized as 'BARBUNYA' when roundness <= -0.82, Perimeter > 0.83, MinorAxisLength > 0.91, ShapeFactor1 <= -0.78, and ConvexArea > 0.94.

These interpretations provide insights into how the model makes its predictions, highlighting the specific features that are influential in classifying different types of dry beans.

Figures 8, 9, 10, 11, and 12 below show the decisions generated by the XGBoost model for the randomly selected rows of 100, 150, 200, 250, and 300 respectively. Based on the decisions generated by the XGBoost model, the class value for rows 100, 150, 200, 250, and 300 is 6, 2, 3, 1, and 5 respectively. to check the name of the class, see section 3.2.2.

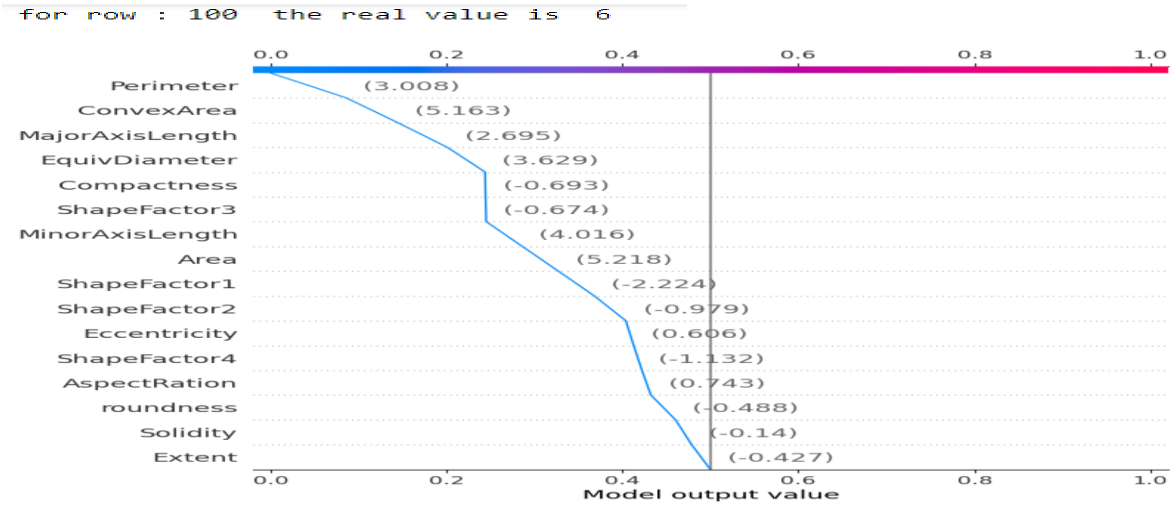



**Fig.8.** Decisions for row 100

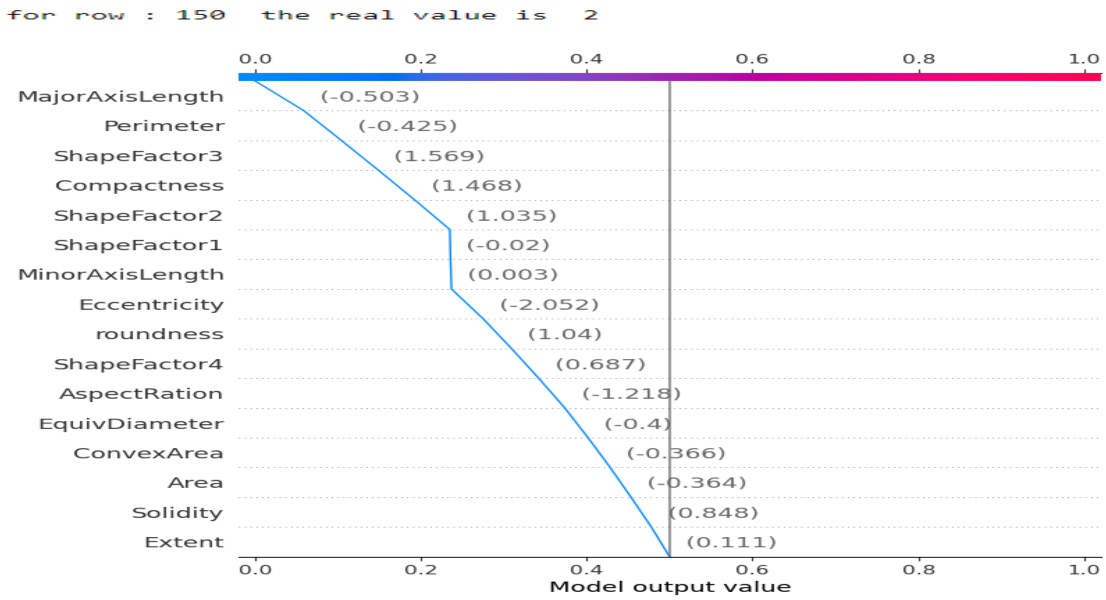



**Fig. 9.** Decisions for row 150

for row : 200 the real value is 3

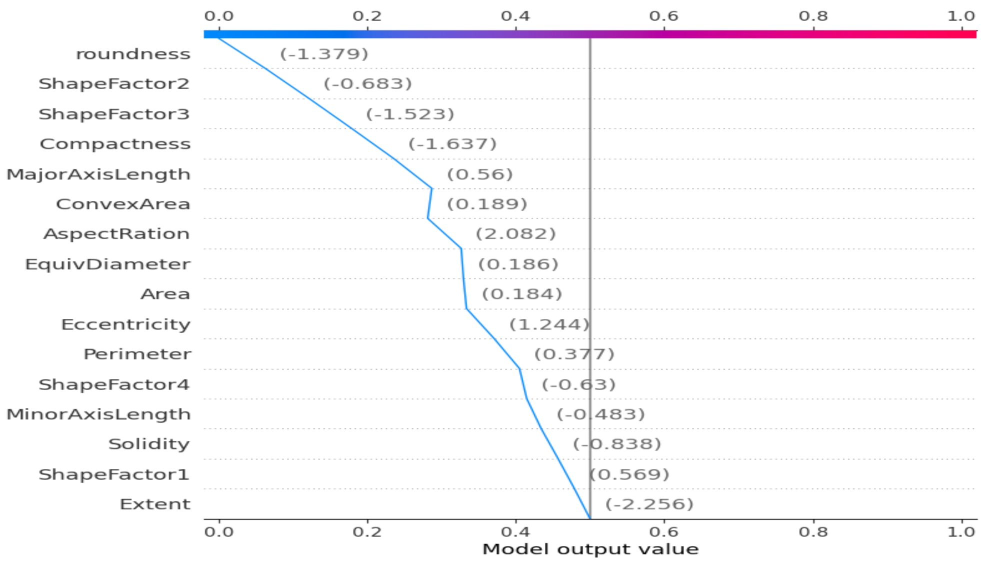



**Fig. 10.** Decisions for row 200

for row : 250 the real value is 1

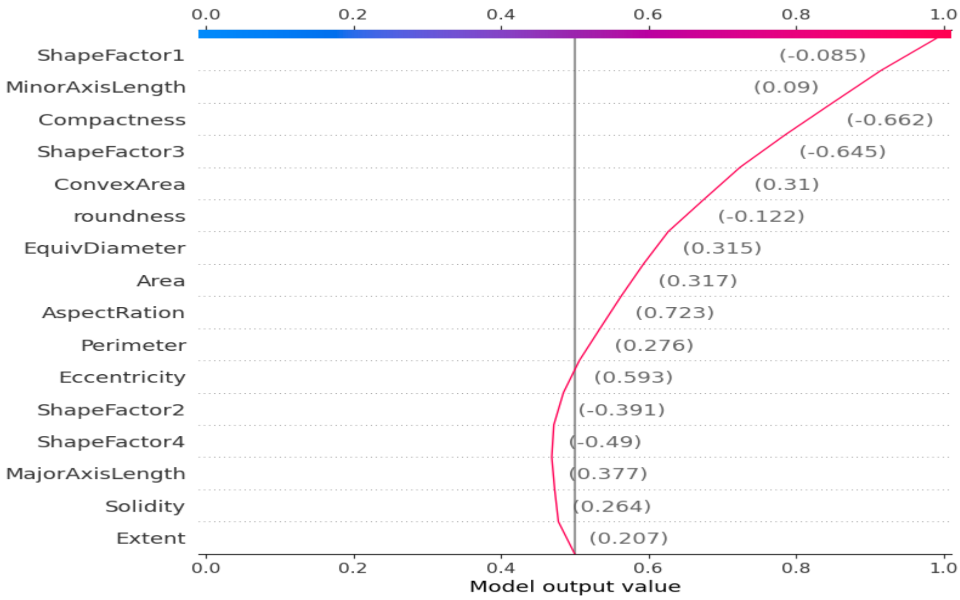



**Fig. 11.** Decisions for row 250

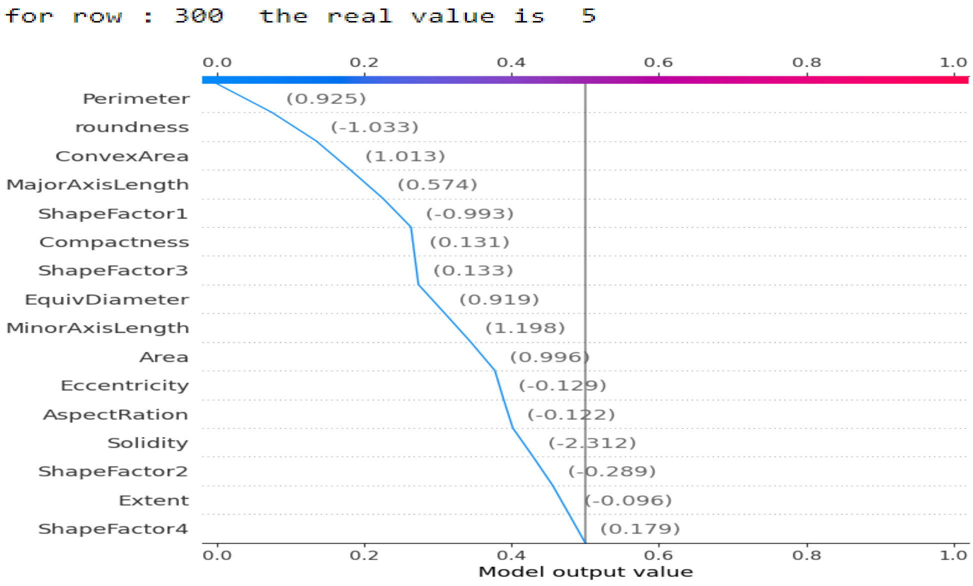



**Fig. 12.** Decisions for row 300

The figure 13 below shows the importance of each feature for each class in constructing the classification model. Based on the result above we have decided that XGBoost is the best classification model to classify the dry beans. So, we have explained the XGBoost model using SHAP explainable AI methods that explain the model using the feature relevancy in the model. As we see here below the figure shows the feature importance of each feature for each class.

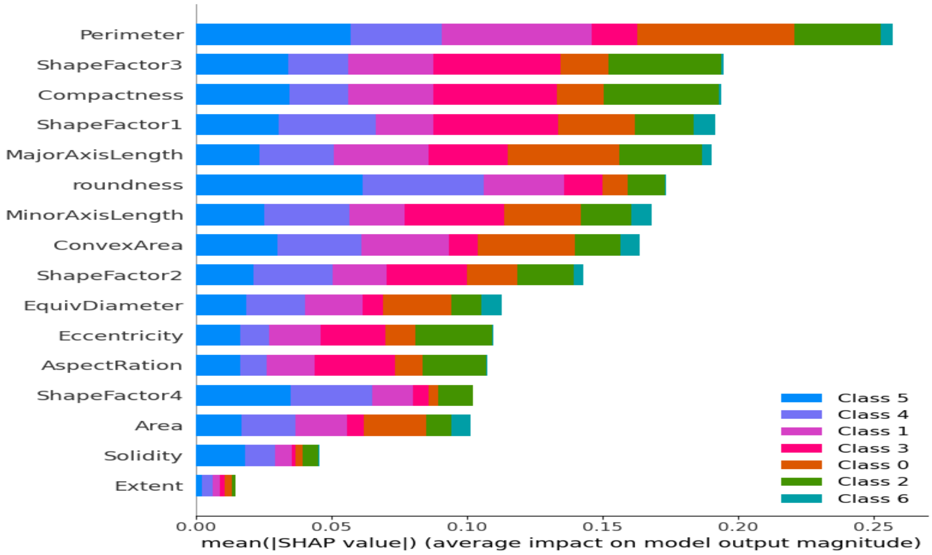


**Fig. 13.** Explainable AI with SHAP

## 5. Conclusion and Recommendation

Dry beans belong to the diverse Fabaceae family, sometimes referred to as Leguminosae, and they are the most important and the most produced pulse in the world. It is originally from America, while there is a wide genetic diversity in the world since, in the 15th and 16th centuries, they were transported to Europe and Africa and quickly spread to the rest of the globe. There are numerous genetic diversities of dry beans, and it is the most produced one among the edible legume crops in the world. According to the Turkish Standards Institution, dry beans are classified as Barbunya, Battal, Bombay, Calı, Dermason, Horoz, Tombul, Selanik, and Seker" based on their botanical characteristics. This study aimed to develop an explainable and interpretable classification model for dry beans using a soft voting classifier and compare the performance with other classic and ensemble machine learning algorithms. The data source for this research is publicly available datasets on Kaggle. After applying the data preprocessing task, out of 13611 instances with 16 features and one class level, 13543 instances with 16 features were used for developing the classification model, and after handling class imbalance using SMOTE + Tomek, 7655 instances were used for the model. We checked the multicollinearity of each feature using variance inflation factors to check the significance of each feature, and we concluded that all the features were significant. The proposed model was constructed using soft voting classifiers, decision trees, random forests, extreme gradient boosting, cat boost, and LGBM algorithms using the balanced and unbalanced dataset. To conduct this study, we have done a total of twelve experiments. The performances of the models are evaluated using accuracy, precision, recall, and f1_score evaluation metrics. We have also explained the classification model using LIME and SHAP feature relevancy explanation techniques, to enhance the explainability and interpretability of the classification model by solving the black-box nature of the algorithms. In this study, the best classification model is identified using the accuracy of the developed classification model. Then, XGBoost is selected as the best algorithm that classifies the dry bean using the balanced dataset with *92.5065%* accuracy. At the end of this conclusion, the researcher recommended that other researchers do: A dry bean classification model by including additional features of the dry bean like 3D features or the suture axis of the bean. The future researcher can also conduct a dry bean classification model using any other advanced algorithms to improve the performances and develop a mobile application.

**Abbreviations.HTML:** Hypertext Markup Language; SMOTE: Synthetic Minority Over-sampling Technique.

**Acknowledgment.** We would like to acknowledge the University of Gondar research and community service president's office for supporting us with funds to conduct this study.

**Author contributions.** Belayneh conceived and designed the study, participated in data analysis, wrote the report, finished the model refinements, carried out a deep analysis of the experiment results, drafted and revised the initial manuscript, and revised the manuscript; Gizachew designed the study, managed the quality and progress of the whole study, and revised the manuscript; Selamawit designs the study and revised the manuscript; all authors read and approved the final manuscript.

**Availability of data and materials.** The datasets analyzed during the current study are available at https://www.kaggle.com/datasets/muratkokludataset/dry-bean-dataset.

**Consent for publication.** Not applicable.

**Competing interests.** The authors report that they have no conflicts.

## References

[1] A. Desole, "Dry Bean Dataset Analysis," Math. Mach. Learn., 2022.

[2] S. K. Sathe, "Dry bean protein functionality," Crit. Rev. Biotechnol., vol. 22, no. 2, pp. 175–223, 2002, doi: 10.1080/07388550290789487.

[3] Y. Long, A. Bassett, K. Cichy, A. Thompson, and D. Morris, "Bean split ratio for dry bean canning quality and variety analysis," IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work., vol. 2019-June, pp. 2665–2668, 2019, doi: 10.1109/CVPRW.2019.00323.

[4] M. Koklu and I. A. Ozkan, "Multiclass classification of dry beans using computer vision and machine learning techniques," Comput. Electron. Agric., vol. 174, no. May, p. 105507, 2020, doi: 10.1016/j.compag.2020.105507.

[5] G. Słowiński, "Dry beans classification using machine learning," CEUR Workshop Proc., vol. 2951, pp. 166–173, 2021.

[6] M. Moshinsky, “Dry Bean Classification,” Nucl. Phys., vol. 13, no. 1, pp. 104–116, 1959.

[7] M. Salauddin Khan et al., “Comparison of multiclass classification techniques using dry bean dataset,” Int. J. Cogn. Comput. Eng., vol. 4, no. March 2022, pp. 6–20, 2023, doi: 10.1016/j.ijcce.2023.01.002.

[8] P. E. D. Love, W. Fang, J. Matthews, S. Porter, H. Luo, and L. Ding, “Explainable Artificial Intelligence ( XAI ): Precepts , Methods , and Opportunities for Research in Construction Explainable Artificial Intelligence ( XAI ): Precepts , Methods , and Opportunities for Research in Construction,” pp. 1–58, 2022.

[9] B. E. Dejene, T. M. Abuhay, and D. S. Bogale, “Predicting the level of anemia among Ethiopian pregnant women using homogeneous ensemble machine learning algorithm,” BMC Med. Inform. Decis. Mak., vol. 22, no. 1, pp. 1–11, 2022, doi: 10.1186/s12911-022-01992-6.

[10] Anynomous, “Data Preprocessing Techniques for Data Mining,” Science (80-. )., p. 6, 2011.

[11] A. M. Dymond, R. W. Coger, and E. A. Serafetinides, “Data preprocessing applied to human average visual evoked potential P100-N140 amplitude, latency, and slope,” Psychiatry Res., vol. 3, no. 3, pp. 315–322, 1980, doi: 10.1016/0165-1781(80)90061-X.

[12] N. H. Son, “Data cleaning and Data preprocessing,” 2011, [Online]. Available: http://www.mimuw.edu.pl/~son/datamining/DM/4-preprocess.pdf

[13] S. B. Kotsiantis and D. Kanellopoulos, “Data preprocessing for supervised leaning,” Int. J. …, vol. 1, no. 2, pp. 1–7, 2006, doi: 10.1080/02331931003692557.

[14] S. Manikandan, “Data transformation,” J. Pharmacol. Pharmacother., vol. 1, no. 2, p. 126, 2010, doi: 10.4103/0976-500x.72373.

[15] J. W. Osborne, “Notes on the use of data transformations,” Pract. Assessment, Res. Eval., vol. 8, no. 6, 2003.

[16] I. Journal and C. Science, “Class Imbalance Problem in Data Mining : Review,” vol. 2, no. 1, 2013.

[17] R. P. Ribeiro, “SMOTE for Regression,” no. October 2015, 2013, doi: 10.1007/978-3-642-40669-0.

[18] E. F. Swana, W. Doorsamy, and P. Bokoro, “Tomek Link and SMOTE Approaches for Machine Fault Classification with an Imbalanced Dataset,” Sensors, vol. 22, no. 9, 2022, doi: 10.3390/s22093246.

[19] “Imbalanced Classification in Python: SMOTE-Tomek Links Method | by Raden Aurelius Andhika Viadinugroho | Towards Data Science.” Accessed: Mar. 30, 2023.

[Online]. Available: https://towardsdatascience.com/imbalanced-classification-in-python-smote-tomek-links-method-6e48dfe69bbc

[20] "Training, Validation and Testing Data Explained | Applause." Accessed: Aug. 16, 2021. [Online]. Available: https://www.applause.com/blog/training-data-validation-data-vs-test-data

[21] M. K. Uçar, M. Nour, H. Sindi, and K. Polat, "The Effect of Training and Testing Process on Machine Learning in Biomedical Datasets," Math. Probl. Eng., vol. 2020, 2020, doi: 10.1155/2020/2836236.

[22] M. J. Healy, "Statistics from the inside. 15. Multiple regression (1).," Arch. Dis. Child., vol. 73, no. 2, pp. 177–181, 1995, doi: 10.1136/adc.73.2.177.

[23] R. G. Mantovani, A. L. D. Rossi, E. Alcobaça, J. C. Gertrudes, S. B. Junior, and A. C. P. de L. F. de Carvalho, "Rethinking Defaults Values: a Low Cost and Efficient Strategy to Define Hyperparameters," 2020, [Online]. Available: http://arxiv.org/abs/2008.00025

[24] B. H. Shekar and G. Dagnew, "Grid search-based hyperparameter tuning and classification of microarray cancer data," 2019 2nd Int. Conf. Adv. Comput. Commun. Paradig. ICACCP 2019, no. February, pp. 1–8, 2019, doi: 10.1109/ICACCP.2019.8882943.

# The effect of planning horizon length and green manure on net income in Crop Rotation Problem

Akouyo Yvette Gbedevi[1], Kossi Atchonouglo[1], Marie-Ange Manier[2], and Sid Ahmed Lamrous[2]

[1] Université de Lomé, Laboratoire d'Analyse et de Modélisation Mathématiques , Lomé, Togo
[2] Université de Technologie Belfort-Montbéliard , FEMTO-ST , Belfort, France

**Abstract.** The world population is increasing rapidly, and recent awareness of the limits of natural resources and the pollution of soil, air and water, is pushing towards a new form of agriculture, sustainable agriculture. Sustainable agriculture can be defined as an agriculture that combines environmental, social and economic objectives for the well-being of farmers and the soil. Crop rotation took hold during this decade by emphasizing the management of soil resources, while improving economic, environmental and social factors. The goal of the crop rotation planning problem discussed in this paper is to maximize the total net return, with a particular emphasis on incorporating a plant that contributes green manure to the soil alongside nutrient amendments by choosing the best horizon for plannification. The rotations generated are of fixed duration for all the plots and the objective is to maximize the income of the farmers. The results showed that the determined algorithm was feasible.

**Keywords:** Sustainable Agriculture · Crop Rotation Problem · Mixed-Integer Linear Programming · Optimization.

## 1 Introduction

Agriculture is the main occupation in many countries around the world and with a growing population, which the UN projects will increase from 7.5 billion to 9.7 billion in 2050 [21]. This means that farmers will have to do more with less. According to the same survey, food production will have to increase by 60% to feed two billion more people. However, traditional methods are not enough to handle this huge demand. This drives farmers and agricultural businesses to find new ways to increase production while conserving soil resources. The challenge is to increase global food production by 50% by 2050 [1] to feed two billion more people by practicing sustainable agriculture.

Crop rotation is a fundamental feature of all organic farming systems. Crop rotation means changing the type of crop grown in a particular piece of land from year to year. There are cyclical rotations, which repeat the same sequence indefinitely, and non-cyclical rotations which allow for changes in crop sequence, adapting to management decisions and evolving as market opportunities arise [20]. Greater soil fertility, fewer pests and crop diseases, and higher yields are just a few of the positive outcomes of rotation. When compared to continuous monoculture practices, according to [26] , rotation increased crop yields by 20 % on average and that the benefits are highly context-dependent.

The agrarian sector's financial stability could be improved by including a variety of cropping strategies. The profits of the farm would not be based solely on a single primary cash crop; rather, they would be based on a diverse collection of commercial crops that were evenly distributed over a number of periods. This could eventually lead to an improvement in cash flow by introducing regular incomes into the agribusiness [23].

In order to assist the farmer in choosing the appropriate crops for the rotation cycle, we proposed a Mixed-Integer Linear Programming (MILP) model to solve the crop rotation problem by exploring the benefit of including green manure. We studied the importance of including fertilizer plants in crop rotation schedule by using exact methods such as linear programming to determine the best solution of the proposed agricultural model and to test it in a real context. The experiment was conducted for a real planting area of average size with 9 plots, considering 8 crops from different botanical families and a two-year, three year and four year planting rotation.

## 2 Literature Review

Maximizing the overall net return is the aim of the crop rotation planning problem that they address in [11]. The branch and bound approach is used to optimize the crop rotation plans in an integer linear programming model.

In addition to the agronomic, water supply, and seasonal demand constraints, the suggested model includes a new temporal preference constraint.

In order to optimize the efficiency of organic farming in the Philippines' second and third largest agricultural land areas, the authors suggested using a Mixed Integer Programming model as a decision-making tool. This tool considers a number of organic farming-related factors. With the matching profit organic farms may make from planting them, the best-selling and most popular produce in the area is used [22].

Using a mixed integer linear programming model, their work [10] examines the crop rotation problem with water supply/demand and net return uncertainties that fluctuate within the permitted rotation cycle. It renders the developed model numerically feasible, mainly when applied to intricate agricultural issues. Determining the best cropping plans, providing a fair income for the farmer, and strategically accounting for water uncertainties are the primary goals of this endeavor.

In order to optimize crop rotation in conventional organic farms with plot adjacency constraints and nutrient amendments, this work [9] makes use of real-world data and the CBC solver. The goal of the created rotations is to maximize farmers' income, and they have set durations for each plot. The suggested agricultural model's solution is found using a linear programming technique.

In this study [18], the authors examine the Crop Rotation Problem and its applicability to the combination of farm management and Precision Agriculture. They increased the problem's appeal for sustainability by presenting a novel mathematical method for the CRP based on crop requirements and nutrient balance. To optimize the CRP, a real-encoded genetic algorithm was created. The findings show that mid- and long-term crop scheduling performed well.

A binary nonlinear bi-objective optimization model is presented in [4] to address the issue of agricultural cultivation planning that is sustainable by using a meta-heuristic technique based on a genetic algorithm and constructive heuristics. A planting schedule for crops to be grown in designated plots to limit the likelihood of pests proliferating and increase the process's profitability.

An integrated strategic-tactical planning model for the supply chain issue involving sugar beets is presented by the authors this study. In order to minimize the overall operational costs, including transportation and inventory of processed and unprocessed beets, a binary integer programming model is developed . To enable crop rotation planning across many cropping seasons, a special temporal dimension was introduced to the planning horizon [12].

## 3 Methodology

### 3.1 Study Area

According to FAO's annual report in 2006, nearly 69% of the 56,600 km2 of the Togolese territory is agricultural land, and 38% of this land is exploited. Rain-fed agriculture is practiced in Togo, and nearly 75% of the country's working population works as small farmers using traditional farming techniques. The two principal crops are maize in the south and focus of the nation and sorghum/millet in the north. 70% of the country's population, or 5.75 million people in 2010, is supported by agriculture, which also makes up nearly 40% of the country's GDP. From the south to the north, Togo is divided into five regions: the Maritime, Plateaux, Central, Kara, and Savanes. From one region to the next, pedoclimatic conditions and the availability of agricultural production factors, particularly land, vary significantly [16]. Togo's climate is tropical, and it varies a lot from south to north. The climate is sub-Sahelian (hot and dry) in the north, sub-Sahelian (rainy) in the center, and Guinea-Sudanian (hot and humid) in the south. The regions' intra-annual rainfall distribution is also unique. There are two rainy seasons in the Maritime and Plateaux regions each year *March/April to July and September to mid-November* and two dry seasons *August and mid-November to March/April* with 900 and 1500 mm of precipitation per year. There is a dry season from *November to April* in the Central, Kara, and Savanes regions, with annual rainfall ranging from 1200 to 1500 mm. As a result, there are two agricultural seasons in the southern parts of Togo, while there is only one in the northern parts. Despite the significance of agriculture to the Togolese economy, it has remained traditional and lacked access to organic and mineral fertilizers. Agricultural production includes:

- **Cereals**: Maize, Millet, and Sorghum
- **Tubers**: Yams and Cassava
- **Legumes**: Beans, Groundnuts, and Soya

Our research focuses primarily on the subtropical climate of the region, taking into account the potential crops.

### 3.2 Problem Description

Several factors, including market demands, soil characteristics, and crop nutrient requirements from a climatic perspective, are taken into consideration when choosing a cropping sequence. Plots are used to divide an agricultural area. Different kinds of crops can be grown on each plot [8]. Planning crops on agricultural land while taking into account the primary factors that affect crop yields (economic, environmental, and ecological) presents a challenge when developing a crop rotation system [9]. Utilizing the principles of crop rotation to make farmers as much money as possible while taking into consideration restrictions based on demand, crop characteristics, production times, and plot conditions is the primary objective of this work. The proposed model takes into account the following constraints:

1. **Sowing Period :** It is essential to observe each crop's sowing and production times.
2. **Cultural continuity among families:** Different botanical families contain the cultures. It is not recommended to grow cultures from the same botanical family in succession on the same plot [23]. This issue is primarily brought about by the fact that crops in the same family have similar deficiencies (the risk of acquiring the same diseases or weeds) and nutrient requirements. The cropping system's ability to last is jeopardized as a result of this.
3. **Crops belonging to the same family's neighbors:** On two adjacent plots, two crops belonging to the same botanical family cannot be sown simultaneously [23]. Characteristics are shared by cultures of the same family. Therefore, staking them simultaneously on two plots that are adjacent to one another is the same as staking the same crop on these two plots. This makes more resources available to pests, which in turn increases their population and the damage they can cause.
4. **Needs for nutrients:** The quantity of soil nitrogen, phosphorus, and potassium required to start a crop at any given time is referred to as its nutrient requirements. We define the nutrient of a crop as the quantity of nitrogen, phosphorus, and potassium that this crop requires, and the nutrient of a plot as the amount of nitrogen, phosphorus, and potassium applied to this plot over a specific time period.
5. **Fertilization with plants:** In the same rotation cycle, combining legumes with other crops has a positive effect on the soil and, as a result, increases yield.
6. **Fallow Period:** To allow the soil to regain its moisture and fertility, each cycle ought to include one or more periods of fallowness.
7. **Market demand:** The distribution of cash crops is significantly restricted as a result of this. Each culture has a preexisting market; the demand needs to be met.

The crop rotation problem is a complicated combinatorial optimization problem that changes depending on the model's scope, considering the aforementioned constraints. Along the rotation cycle, it is desirable to determine the best crop combinations to plant in each plot at each time. The problem is solved using the constructed mathematical model, and the following decisions are made:

- Determine the crop-specific area needed to satisfy demand.
- During the rotation cycle, determine the crop sequence in each plot.
- Obtain the various agricultural operations' calendar.

The crop rotation problem is a complicated combinatorial optimization problem that changes depending on the model's scope, considering the aforementioned constraints. Along the rotation cycle, it is desirable to determine the best crop combinations to plant in each plot at each time. The problem is solved using the constructed mathematical model, and the following decisions are made:

- Determine the crop-specific area needed to satisfy demand.
- During the rotation cycle, determine the crop sequence in each plot.
- Obtain the various agricultural operations' calendar.

## 4 Mathematical Modeling of the problem

The planting area is thought to be divided into plots by us. If two parcels share a boundary that is not reduced to a discrete set of points, they are neighbors. We will consider the opposite plots to be adjacent, for instance, if the planting area is divided into four plots.

A crop rotation schedule with a clearly defined planning horizon (T) is established by a producer with the intention of maximizing profits following an agricultural season. It has a number of crops (N) that are part of various botanical families (p), where F(p) is the number of crops in the family (p) and surfaces that are divided into plots (K) and naturally contain certain amounts of nutrients (Nitrogen, Phosphorus, and Potassium). Profitability (l), production cost (CP), production time (Z), average production (Q) per hectare, market demand (D), and nutrient requirements are all associated with each crop. Additionally, the farmer is restricted by the following restrictions: sowing time, continuity and neighbor for crops belonging to the same family, green fertilization, and fallow time.

The tables define all indices 1, parameters/data, and variables 2 for the proposed model.

Table 1: List of Index.

| Index | Description |
|---|---|
| $i$ | relatifs aux cultures |
| $j$ | relatifs aux parcelles |
| $t$ | période de l'horizon de planification |
| $p$ | lié à la famille botanique des plantes |
| $\alpha$ | lié à l'intervalle de fertilisation, $\alpha \in \Omega, \Omega = \{\alpha \in N^* \mid \alpha \cdot \theta \leq T, \theta \in N^*\}$ |

The index of fertilization ($\alpha$) is determined by the model parameters fertilization interval ($\theta$) and planning horizon (T). For instance, if the planning horizon is 24 periods and the fertilization interval is 12 periods ($\theta = 12$), then the set $\Omega$ is 1, 2 because $\alpha = 1$ and $\alpha = 2$ satisfy the definition of $\alpha \cdot \theta \leq T$.

Table 2: List of Parameters.

| Parameters | Description |
|---|---|
| N | number of crops to plant(N>=2) |
| K | number of plots (lot) available |
| T | the horizon (Duration) of planification |
| V | Set of crops for green fertilization |
| n =N+1 | represents a fictitious crop imposing a fallow |
| $F_P$ | Set of family cultures $p$, $p$= 1, …, $N_f$ |
| $N_f$ | Number of crop families |
| $Surf_j$ | Area of the plot $j$ in $ha$ |
| $l_i$ | Crop profitability $i$ per $ha$ (FCFA XOF) |
| $z_i$ | Crop production cycle $i$ including the sowing period and the harvest period |
| $Q_i$ | Crop Production Average $i$ per ha |
| $I_i$ | Crop sowing interval including earliest and latest period $i$ $\{I_i1, \ldots, I_in$ |
| $D_i$ | Crop demand $i$ (unit/period) |
| $S_j$ | Adjacent plots to the plot $j$ |
| $F_N\alpha ij$ , $F_P\alpha ij$ , $F_K\alpha ij$ | Dose of Nitrogen, Phosphorus and Potassium to bring to the plot $j$ on interval $\alpha$ according to crop $i$ . |
| $B_Nj$ , $B_Pj$ , $B_Kj$ | Initial composition of the soil in Nitrogen, Phosphorus and Potassium of the plot $j$ per ha. |
| $R_Ni$, $R_Pi$, $R_Ki$ | Crop requirements $i$ in Nitrogen,Phosphorus and Potassium |
| $C_N$, $C_P$, $C_K$ | Cost of fertilisation in (FCFA par ha) |
| $F_{mini}$ , $F_{maxi}$ | Limit of fertilization for crop $i$ |
| $\theta$ | Fertilization equilibrium interval (Compensation) |
| $OC_i$ | Other production costs incurred on plot $j$ at period $t$ due to crop $i$. (Preparation of the land,purchase of seeds for sowing, transport, labor) |

The entire model is displayed below. The benefits of crop planning (calendar), the costs of fertilization, and other production costs (land preparation, transportation, labor) are evaluated by the objective function in Equation (1). Fertilization costs are also included in, with the goal of reducing reliance on external chemical fertilizers.

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{K} Surf_j * l_i * x_{itj} - \sum_{\alpha\in\Omega}\sum_{j=1}^{K}(F_{N\alpha ij} * C_N + F_{P\alpha ij} * C_P + F_{K\alpha ij} * C_K) - \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{K} OC_i * x_{itj}. \quad (1)$$

Subject to (1)-(15).

$$\sum_{i=1}^{N}\sum_{r=0}^{zi-1} x_{i(t-r)j} \leq 1. \quad (2)$$

*With* $t$=1,...,T; $j$ = 1,...,K.

It is not allowed to plant two crops in the same plot [24, 13, 18]. Indeed, a crop occupies the entire plot throughout its growth and harvest. The constraint defined in equation (2) forbids scheduling more than one culture during the same period. There is a spatial constraint.

$$\sum_{i\in F_{(p)}}\sum_{r=0}^{z_i-1}\sum_{v\in S_j} x_{i(t-r)v} \leq K\left(1 - \sum_{i\in F_{(p)}}\sum_{r=0}^{z_i-1} x_{i(t-r)j}\right). \quad (3)$$

*With* $p$=1,...,$N_f$; $t$=1,...,T; $j$ = 1,...,K.

It is not recommended to plant two crops from the same botanical family on adjacent plots at the same time [24, 9, 13, 18]. The fact that the cultures of the same family share characteristics is the source of this issue. Therefore, planting them simultaneously on two plots that are adjacent to one another amounts to planting the same crop on both plots, which does not maximize crop distribution on the various plots. If a crop $i$ of the same botanical family $p$ has already been sown on plot $k$, constraint (3) states that the number of crops on all plots $S_j$ adjacent to plot $j$ during their production periods $Z_i$ must be equal to zero; otherwise, the number of crops is at most equal to the number of independent plots.

$$\sum_{i\in F_{(p)}}\sum_{r=0}^{z_i} x_{i(t-r)j} \leq 1. \quad (4)$$

With $p$=1,...,$N_f$, $t$=1,...,T, $j$=1,...,K.

On the same plot, cultures belonging to the same botanical family cannot be grown immediately [24, 13]. This issue is primarily brought about by the fact that crops in the same family have similar deficiencies (the risk of acquiring the same diseases or weeds) and nutrient requirements. The cropping system's agronomic viability is jeopardized as a result of this. Constraint in equation (4) sets a maximum of one crop as the sum of all crops $i$ in the botanical family $p$ over their production period $Z_i$.

$$\sum_{i\in V}\sum_{t=1}^{T} x_{itj} \geq 1 \quad (5)$$

With $j$=1,...,K.

Additionally, green manure improves the structure and fertility of the soil while also enhancing its organic matter enrichment [13]. Green manure adds nitrogen to the rotation by planting legumes [7]. Constraint (5) guarantees that every plot gets at least one implementation of green manure (legumes).

$$\sum_{t=1}^{T} x_{ntj} \geq 1. \quad (6)$$

With $n$=N+1, $j$=1,. . . ,K .

The period of frost or fallow enables the plot to restock its production capacity, water reserves, and other resources. as well as to restrict excessive agricultural production. Constraint defined in equation (6) ensures that each plot has at least one freezing period. During the frost period, there are no restrictions on neighborhood and consecutive planting.

$$\sum_{j=1}^{K}\sum_{t\notin I_i} x_{itj} = 0. \tag{7}$$

With $i$=1,. . . ,N.

Following the recommended planting date is critical for allowing the crop to express its yield potential and lowering crop protection costs. By preventing allocation outside of this window, the constraint outlined in equation (7) ensures that crop scheduling takes place only during the appropriate sowing period.

When nutrients are present in sufficient quantities and in mineral forms that can be absorbed by plants, a soil is more fertile [9]. They deplete the soil of the essential nutrients they require as they grow.

$$F_{N\alpha j} - \sum_{i=0}^{N}\sum_{t=1+(\alpha-1)*\Theta}^{\alpha*\Theta} x_{itj} * Surf_j * (R_{Ni} - B_{Nj}) \geq 0. \tag{8}$$

with $t$=1,. . . ,T ; $\alpha \in \Omega$.

$$F_{P\alpha j} - \sum_{i=0}^{N}\sum_{t=1+(\alpha-1)*\Theta}^{\alpha*\Theta} x_{itj} * Surf_j * (R_{Pi} - B_{Pj}) \geq 0. \tag{9}$$

with $t$=1,. . . ,T ; $\alpha \in \Omega$.

$$F_{K\alpha ij} - \sum_{i=0}^{N}\sum_{t=1+(\alpha-1)*\Theta}^{\alpha*\Theta} x_{itj} * Surf_j * (R_{Ki} - B_{Kj}) \geq 0. \tag{10}$$

with $t$=1,. . . ,T ; $\alpha \in \Omega$.

The quantity of nitrogen, phosphorus, and potassium that the soil requires to initiate a crop at any given time is referred to as the minimum nutrient requirement. A plot's nutrient amendment is defined here as the amount of nitrogen, phosphorus, and potassium applied to it over a specific time period. Let $\alpha$ be the size of the

interval that is appropriate for sowing crop $i$, the interval that is used to apply the amendments $F_{N\alpha j}$ ,$F_{P\alpha j}$ and $F_{K\alpha j}$, $B_{Nj}$,$B_{Pj}$ and $B_{Kj}$, which represent the initial composition of the Nitrogen, Phosphorus, and Potassium soil of plot $j$ per ha [18, 9] , $R_{Ni}$ ,$R_{Pi}$ and $R_{Ki}$, which represent the minimum amount of nutrients that crop $i$ requires.

Equations (8), (9),(10) are used to calculate fertilization balances based on the surface nutrient budget.

$$\sum_{j=1}^{K}\sum_{t=1}^{T} Surf_j * Q_i * x_{itj} \geq D_i. \tag{11}$$

$i$ = 1,. . . ,N.

Demand from the market is another significant constraint. The farm's crop yields are constrained by this constraint to meet the anticipated demand for the crop $i$. To avoid issues like prolonged product storage or conservation, which can result in additional costs and increase the risk of income variability, we believe that the farmer should not exceed the estimated demand. The constraint outlined in the equation (11) is used to evaluate the crop's production requirements.

Each of the Boolean decision variables $x_{itj}$ represents the schedule (planning) of crop $i$ during period $j$ on plot $t$. When and where crops are planted are tracked by them. The variables of fertilization $F_{N\alpha j}$ ,$F_{P\alpha j}$ and $F_{K\alpha j}$ are actual variables.

$$x_{itj} \in \{0, 1\}\,. \tag{12}$$

With $i$=1,...,N ; $t$=1,...,T ; $j$=1,...,K .

$$F_{N\alpha j} = \left\{F_{N\alpha j} \in R^{+} \mid F_{Nmax} \geq F_{N\alpha j} \geq F_{Nmin}\right\}. \tag{13}$$

With $j$=1,...,K; $\alpha \in \Omega$ .

$$F_{P\alpha j} = \left\{F_{P\alpha j} \in R^{+} \mid F_{Nmax} \geq F_{P\alpha j} \geq F_{Nmin}\right\}. \tag{14}$$

With $j$=1,...,K; $\alpha \in \Omega$ .

$$F_{K\alpha j} = \left\{F_{K\alpha j} \in R^{+} \mid F_{Nmax} \geq F_{K\alpha j} \geq F_{Nmin}\right\}. \tag{15}$$

With $j$=1,...,K; $\alpha \in \Omega$ .

## 5 Model discussion and analysis

### 5.1 Tools

Python-MIP, a collection of Python tools for modeling and solving Mixed-Integer Linear Programs (MIPs), was used for the model's implementation and evaluation and CBC Solver as Solver since it does not require a license unlike Gurobi. The solver uses the branch and bound algorithm to find the optimal solution for the crop rotation problem. The study will focus on eight crops from five botanical families. Crop data, as well as production parameters like planting, harvesting dates and nutrient requirements, are listed in Table 4. The planning horizon is divided into month. At each rotation schedule, each area adopted at least one fallow period and one green manure crop. The proposed model can be solved in less than a minute and includes 1998 variables and 23049 linear constraints.

### 5.2 Study cases

The proposed model will be evaluated in two scenarios. Firstly, we present the impact of including nutrient amendment in the crop rotation schedule on the net profit (Instance 1). We evaluate the income and crop rotation schedule in this experiment, taking into consideration whether or not soil amendments are present. Besides, we present the alteration of the design of the plots on the establishing region which restricts the crop allocation. Then, we evaluate our model using a variety of two, three, and four years planning horizons (Instance 2).

**Instance 1** Table 4 presents details about the crop, planting and harvest times and Nitrogen, Phosphorus and Potassium requirement for each crop . In this study, there are a total of eight crops (N = 8) and five families of crops (Nf = 5). Our planning horizon is two years, divided into 24 periods (T = 24). Table 5 describes the adjacency among plots and the cultivable area of each plot. We do not take into consideration the main diagonal, but the other positions in the table filled in Green represent the plots that are adjacent to each other. Details such as number of variables and linear constraint about the two models are displayed in Table 3.

Table 3: Variables and Linear Constraints of each model

| Models | Variables | Linear Constraints |
|---|---|---|
| Without Nutrient Amendment | 1944 | 22995 |
| With Nutrient Amendment | 1998 | 23049 |

In the first instance of our problem, the solver reached the value of the objective function in **9,66** seconds for the model without nutrient amendments which is **30,639,998 XOF** and in **61,21** seconds for the model with nutrient amendments which is **22,087,660 XOF**. Planning is presented differently in each model. The representation of the CBC solver-derived solutions for the two models is depicted in Figures 1 and 2. Figures 1 and 2 both demonstrate compliance with the adjacency constraint.

Table 4: Crop's attributes: seeding, harvesting and Nutrient Demand.

| Index | Crops | Botanical Family | Sowing | Harvest | Cycle(Month) | N(kg/ha) | P(kg/ha) | K(kg/ha) |
|---|---|---|---|---|---|---|---|---|
| 1 | Maize | Poaceae | April-May-September | (July-August)-December | 4 | 27-34 | 10-12 | 26-37 |
| 2 | Groundnuts | Fabaceae | March | June | 4 | | | |
| 3 | Beans | Fabaceae | July-August | October-November | 4 | 69 | 19 | 51 |
| 4 | Sorghum | Poaceae | Mid-June-Mid-July | September-October | 4 | 33 | 10 | 34 |
| 5 | Soya | Fabaceae | Mid-June-Mid-July | November-December | 6 | 67-71 | 16-26 | 18-53 |
| 6 | Cassava | Euphorbiaceae | June-July | June-July | 12 | 2-6 | 1-2 | 1-9 |
| 7 | Cotton | Malvaceae | July-August | November-December | 5 | 120-214 | 43-86 | 87-223 |
| 8 | Yam | Dioscoreaceae | November-March-April-May | July-November-December-January | 9 | | | |

Table 5: Plot's Adjacency matrix corresponding to the given farm.

| **Plots** | **Plot 1** | **Plot 2** | **Plot 3** | **Plot 4** | **Plot 5** | **Plot 6** | **Plot 7** | **Plot 8** | **Plot 9** | **Area in ha** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Plot 1** | | | | | | | | | | 2 |
| **Plot 2** | | | | | | | | | | 3 |
| **Plot 3** | | | | | | | | | | 1 |
| **Plot 4** | | | | | | | | | | 1 |
| **Plot 5** | | | | | | | | | | 2 |
| **Plot 6** | | | | | | | | | | 3 |
| **Plot 7** | | | | | | | | | | 2 |
| **Plot 8** | | | | | | | | | | 1 |
| **Plot 9** | | | | | | | | | | 3 |

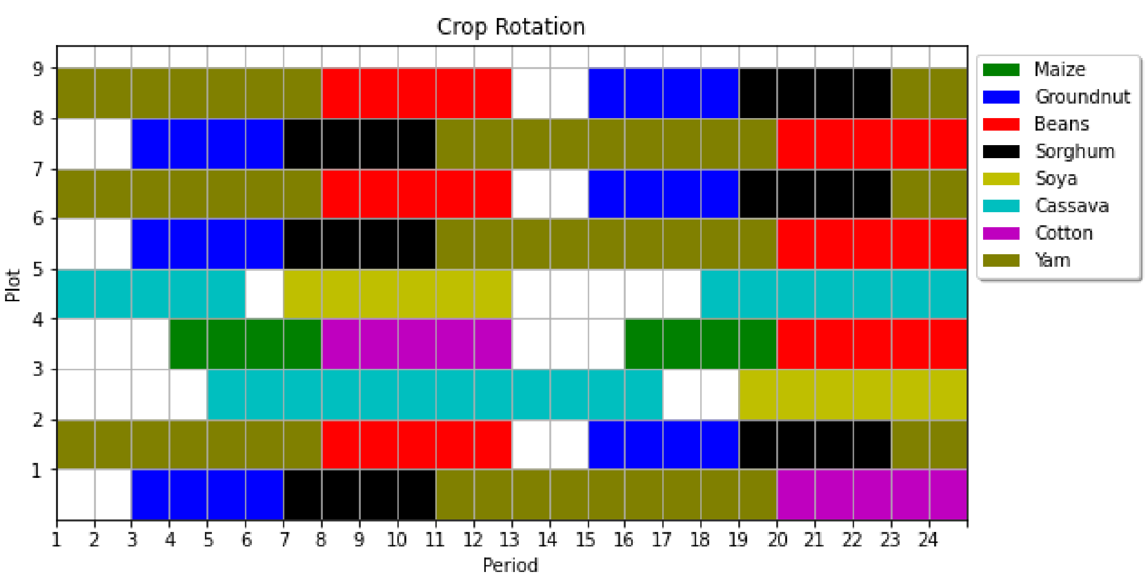



Fig. 1: Illustration of crop rotation schedule proposed from the model without the addition of nutrients. The white cells represents fallow period. Yam is cultivated in $T = 23$ and ended-up in $T = 7$ of the following year on plot $j = 2$ for a total duration of 9 months $Z_{yam} = 9$. As Groundnut and Beans are from the same botanical family, there is a fallow period of 2 to avoid an immediate succession.

**Instance 2** Instance 2, we want to compare the net returns from the crop rotation planning so we consider for experiment two-year, three-year and four-year period for the horizon of planning with the same crop data as

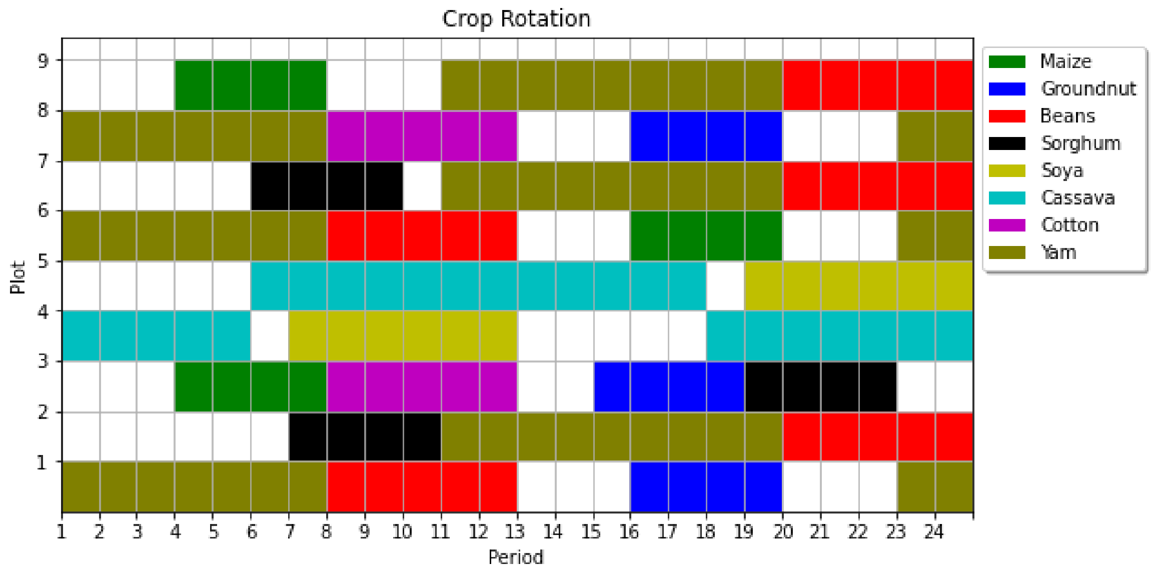
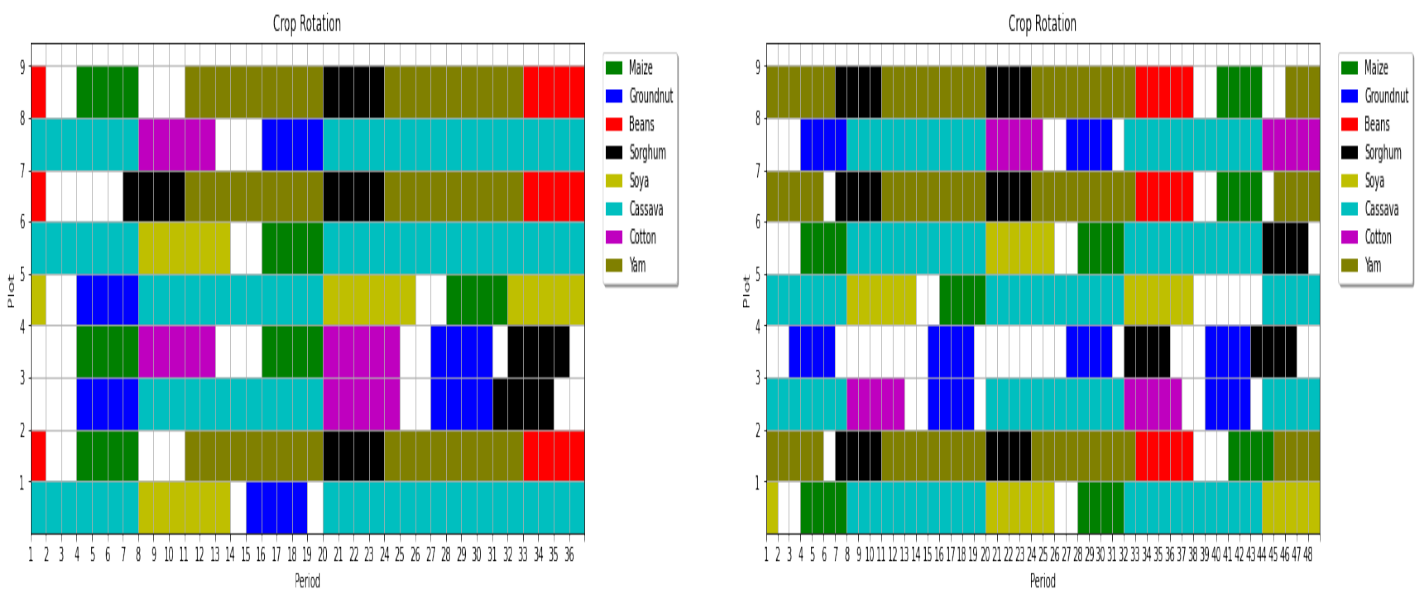



Fig. 2: Illustration of crop rotation schedule proposed from the model including soil nutrients amendments. The white cells signify fallow periods. Yam is cultivated in $T = 23$ and ended-up in $T = 7$ of the following year on plot $j = 1$ for a total duration of 9 months $Z_{yam} = 9$. As Groundnut and Beans are from the same botanical family, there is a fallow period of 2 to avoid an immediate succession.




(a) Illustration of crop rotation schedule proposed from the model with a planning horizon of three years.

(b) Illustration of crop rotation schedule proposed from the model with a planning horizon of four years

Fig. 3: Illustration of the crop schedule according to the planning horizon of three and four years respectively 36 and 48 periods

Table 4 . The Cbc solver respectively a value of **31.366.080 XOF** and **88.327.840 XOF** after **30 seconds** and **28 seconds** of running time, and the associated rotation for each plot is given in Figure 3.

As can be seen in Figure 4, the net profit increases with the planning horizon. This necessitates consideration of the horizon selection and crop selection for the rotation. But planning crops ahead of time over several years can be risky because there are many things that can affect market demand, like the weather, and market prices can change a lot from season to season.

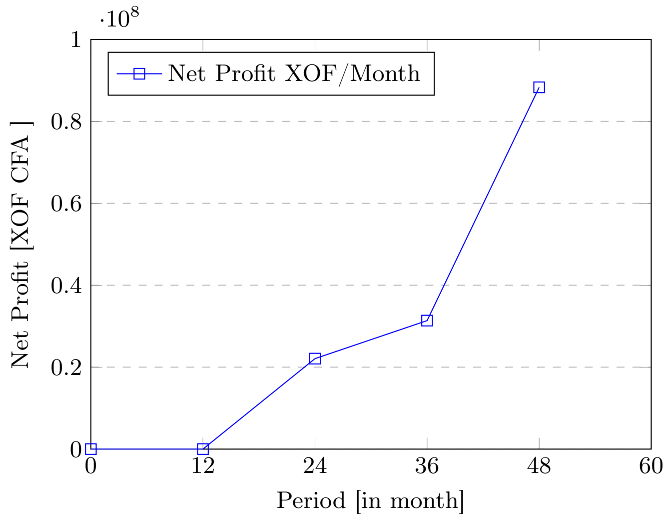



Fig. 4: Variation of net profit based on planning Horizon

## 6 Conclusions

This article highlights the significant role of legumes in modeling a crop rotation system, considering constraints such as contiguity and the use of mineral fertilizers for soil amendment. To address the problem of increasing the net income of farmers in a crop rotation system, We used a mixed integer linear programming model (MILP). Our model proposes the best possible solution by maximizing the objective, in this case farmers' income, while respecting a set of constraints, such as crop rotation rules. The best solution of the proposed model was tested in an experiment conducted for a medium-sized real plantation area with nine plots, considering eight crops from five botanical families and a two-year plantation rotation. The results of the study indicate that farmer's incomes improve when a longer planning period is considered.

However, the solutions proposed by MILP may lack the flexibility to adapt to rapid changes in the field, such as a sudden drought or market price fluctuations. This aspect will be taken into consideration in our future work by introducing a model based on stochastic linear programming model. Additionally, integrating a dynamic subdivision could further boost farmers' incomes.

## References

1. Proceedings of the expert meeting on how to feed the world in 2050: 24-26 june 2009, fao headquarters, rome. FAO, Rome (2009), oCLC: 458810459
2. Agriculture & Land-Based Training Association (ALBA): Crop planning and production (May 2012)
3. Alarcos Llorach, E.: Gramática de la Lengua Española. Editorial Espasa Calpe, Madrid (1999)
4. Aliano Filho, A., de Oliveira Florentino, H., Pato, M.V., Poltroniere, S.C., da Silva Costa, J.F.: Exact and heuristic methods to solve a bi-objective problem of sustainable cultivation. Annals of Operations Research **314**(2), 347–376 (Jul 2022). https://doi.org/10.1007/s10479-019-03468-9, https://link.springer.com/10.1007/s10479-019-03468-9
5. Arslan, L.M., Hansen, J.H.L.: Language accent classification in American English. Speech Communication **18**(1), 353–367 (1996)
6. Bartkova, K., Jouvet, D.: Selective prosodic post-processing for improving recognition of French telephone numbers. In: Smith, J. (ed.) Proc. of the 7th European Conference on Speech Communication and Technology. pp. 267–270. Taylor and Francis (1999)
7. Céspedes-León, C., Espinoza, S., Maass, V.: Nitrogen transfer from legume green manure in a crop rotation to an onion crop using 15 n natural abundance technique. Chilean Journal of Agricultural Research **82**, 44 – 51 (01 2022). https://doi.org/10.4067/S0718-58392022000100044

8. El Adoly, A.A., Gheith, M., Nashat Fors, M.: A new formulation and solution for the nurse scheduling problem: A case study in Egypt. Alexandria Engineering Journal **57**(4), 2289–2298 (Dec 2018). https://doi.org/10.1016/j.aej.2017.09.007, https://linkinghub.elsevier.com/retrieve/pii/S111001681730282X
9. Fendji, J.L.E.K., Kenmogne, C.T., Fotsa-Mbogne, D.J., Förster, A.: Improving farmers' revenue in crop rotation systems with plot adjacency constraints in organic farms with nutrient amendments. Applied Sciences **11**(15), 6775 (Jul 2021). https://doi.org/10.3390/app11156775, https://www.mdpi.com/2076-3417/11/15/6775
10. Fikry, I., Eltawil, A., Gheith, M.: A robust crop rotation optimization model with water scarcity and net return uncertainty considerations. IEEE Access **9**, 128938–128950 (2021). https://doi.org/10.1109/ACCESS.2021.3113125, https://ieeexplore.ieee.org/document/9539176/
11. Fikry, I., Gheith, M., Eltawil, A.: A new mathematical model for the deterministic crop rotation planning problem. In: 2019 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM). pp. 314–318. IEEE, Macao, Macao (Dec 2019). https://doi.org/10.1109/IEEM44572.2019.8978672, https://ieeexplore.ieee.org/document/8978672/
12. Fikry, I., Gheith, M., Eltawil, A.: An integrated production-logistics-crop rotation planning model for sugar beet supply chains. Computers & Industrial Engineering **157**, 107300 (Jul 2021). https://doi.org/10.1016/j.cie.2021.107300, https://linkinghub.elsevier.com/retrieve/pii/S0360835221002047
13. Filho, A., Florentino, H., Pato, M.: A genetic algorithm for crop rotation. ICORES 2012 - Proceedings of the 1st International Conference on Operations Research and Enterprise Systems pp. 454–457 (01 2012)
14. Ibarra Echeverría, M., Núñez Solís, E., Huerta Ibáñez, J.: Manual de Aceros inoxidables. INDURA S.A., Cerritos, Chile (2010)
15. Jacobs, E., Ward, K., Pinkerton, R.: The characteristics of 78 related airfoils sections from tests in the variable-density wind tunnel. Tech. Rep. 460, National Advisory Commitee for Aeronautics (NACA), Washington D.C. (1935)
16. Kokou, K.: ANALYSE ET MODELISATION DE L'EVOLUTION DES INDICATEURS DE LA FERTILITE DES SOLS CULTIVES EN ZONE COTONNIERE DU TOGO. Ph.D. thesis, UNIVERSITE DE BOURGOGNE - FRANCE UNIVERSITE DE LOME- TOGO (Nov 2011)
17. Krause, G.: Análisis teórico y numérico de atractores e intermitencia en la ecuación DNLS. Ph.D. thesis, Facultad de Ciencias Exactas, Físicas y Naturales; Universidad Nacional de Córdoba, Córdoba, Argentina (2014), tesis doctoral
18. Miranda, B., Yamakami, A., Berbert, P.: A new approach for crop rotation problem in farming 4.0. pp. 99–111 (04 2019). https://doi.org/10.1007/978-3-030-17771-3_9
19. Mitchell, J.K.: Geotechnical and Geoenvironmental Engineering Handbook, chap. Physicochemistry of Soils for Geoenvironmental Engineering, pp. 691–710. Kluger Academic Publishers, Boston (2001)
20. Mohler, C.L., Johnson, S.E., Natural Resource, Agriculture, and Engineering Service (eds.): Crop rotation on organic farms: a planning manual. No. 177 in NRAES, Natural Resource, Agriculture, and Engineering Service (NRAES) Cooperative Extension, Ithaca, NY (2009), oCLC: ocn320131151
21. Nations, U.: Population, https://www.un.org/en/global-issues/population
22. Ong, N.R., Castro, M.A.R., Osorio, P.S., Tycangco, J.G.D., Umali, A.M.F.: Application of mixed integer linear programming approach on crop rotation practices in organic farms in central luzon and calabarzon, philippines. Journal of Physics: Conference Series **1529**(3), 032050 (Apr 2020). https://doi.org/10.1088/1742-6596/1529/3/032050, https://iopscience.iop.org/article/10.1088/1742-6596/1529/3/032050
23. Santos, L., Michelon, P., Arenales, M., Santos, R.: Crop rotation scheduling with adjacency constraints. Annals OR **190**, 165–180 (10 2011). https://doi.org/10.1007/s10479-008-0478-z
24. dos Santos, L.M.R., Michelon, P., Arenales, M.N., Santos, R.H.S.: Crop rotation scheduling with adjacency constraints. Annals of Operations Research **190**(1), 165–180 (Oct 2011). https://doi.org/10.1007/s10479-008-0478-z, http://link.springer.com/10.1007/s10479-008-0478-z
25. Wang, H., Liu, S., Gong, J., Wu, N., Lin, J.: Contribution of velocity vortices and fast shock reflection and refraction to the formation of EUV waves in solar eruptions. The Astrophysical Journal pp. 114–129 (2015)
26. Zhao, J., Yang, Y., Zhang, K., Jeong, J., Zeng, Z., Zang, H.: Does crop rotation yield more in china? a meta-analysis. Field Crops Research **245**, 107659 (2020). https://doi.org/https://doi.org/10.1016/j.fcr.2019.107659, https://www.sciencedirect.com/science/article/pii/S0378429019307464

# Towards the digitalization of Cameroonian agriculture: current situation, challenges and prospects

André MBAIRANODJI[1], FOUNADOUDOU[1], Daquin Cédric AWOUAFACK[*1] and Garba BOUBA BOÏGA[1]

[1] Department of Agricultural Surveys and Statistics, Ministry of Agriculture and Rural Development, Yaoundé-Cameroon
*Email: dcawouafack@yahoo.com

**Abstract.** The aim of this review article is to highlight the digital solutions available to Cameroonian agriculture. It also looks at the strengths, weaknesses, opportunities and threats of these solutions. In view of the constraints weighing on this sector, notably climate change, prospects are outlined.

**Keywords:** Digitalization, farmers, new-generation agriculture, sustainability.

## 1 Introduction

In a world where digital transformation is redefining the contours of every industry, agriculture is at the heart of a silent but essential revolution. Although digitalization offers promising prospects for growth and sustainability on a global scale, it remains largely under-exploited, particularly in Africa, where structural challenges abound. While the FAO (Food and Agriculture Organization of the United Nations) estimates that global food production must increase by 60% by 2050 to meet the needs of a growing population [1, 2], African agriculture remains well behind in the adoption of digital technologies, with only 10% of farms using digital technologies to manage their activities [3]. This digital divide threatens food security and economic development in developing countries, including Cameroon, where agriculture accounts for over 20% of GDP and employs nearly 60% of the working population [4]. According to the United Nations' Sustainable Development Report [5], the integration of digital technologies in agriculture can help achieve the Sustainable Development Goals (SDGs), notably SDG 2 (Zero Hunger) and SDG 8 (Decent Work and Economic Growth). Faced with these challenges, the digitization of Cameroon's agriculture appears to be an imperative necessity to increase productivity, improve efficiency and strengthen the resilience of agricultural systems to climate change and economic fluctuations. In this context, strategic documents such as the National Development Plan 2020-2030 [6] and the National Strategy for the Development of Agriculture [7], which aim for a profound transformation of the sector, emphasize the importance of increased digitalization. This vision of a modernized agricultural sector is based on optimizing agricultural value chains, digitizing agricultural services and promoting precision farming. However, the challenges that lie ahead are many and complex: geographical disparities in access to technology, the need to train players and insufficient funding to support farmers in this digital revolution [8]. This article explores the current state of digitalization in Cameroonian agriculture, analyzing the forces at work, the challenges to be overcome and the strategies underway to remove the obstacles that are holding back this digital transition. Through this investigation, we will attempt to identify the prospects for a more productive, more resilient and more competitive agriculture, where digital technologies, far from being a mere trend, become the driving force behind a profound and sustainable transformation. For beyond the challenges, the stakes are high: we need to give Cameroon's land the opportunity to bear fruit, not just for one country, but for an entire region in search of food security and development.

## 2 Overview of the current situation

Digitizing Cameroon's agricultural sector is now a strategic imperative for modernizing the sector, boosting competitiveness and guaranteeing sustainable growth. Several innovative initiatives have been launched to support this digital transition, backed by national and international projects. For example, the Agricultural Production Support Program in Cameroon in short PARPAC [9], has deployed a platform for registering agricultural producers, facilitating access to agricultural and financial services and reinforcing the traceability of agricultural activities [1]. This project is part of a digital inclusion logic, in line with the objectives of the Digital Transformation of Cameroon Project in short PATNUC [10], which aims to promote access to agricultural technologies for producers through accessible and adapted digital solutions [3]. At the same time, the SIMC Project [11] has set up a call center enabling farmers to receive real-time information on weather conditions and forecasts, a crucial tool in the face of climatic hazards. This initiative enables producers to better anticipate the risks associated with climate change and adopt more resilient management strategies [5]. Similarly, the Cameroon Agropastoral Portal, in short CamAgro, facilitates networking between producers, suppliers and consumers [12], optimizing agricultural value chains and strengthening cooperation [13]. The Observatory of agricultural inputs, meanwhile, aims to provide a platform dedicated to real-time dissemination of prices and availability of agricultural inputs, a valuable tool for combating price fluctuations and guaranteeing a stable supply [14]. However, despite these advances, digitalization still remains timid in sub-Saharan Africa, where only 10% of farms benefit from digital solutions [3]. Challenges persist: digital infrastructure, access to financing and training for agricultural players remain major obstacles to be overcome. Nevertheless, with the rise of these digital solutions and PATNUC's support for digital inclusion, Cameroon's agriculture is entering a promising new era, offering prospects for increased productivity, resilience in the face of crises and a boost to the country's economic growth. The road ahead is still strewn with pitfalls, but the integration of digital technologies now seems unavoidable if we are to meet the challenges of the future and guarantee the country's food security and sustainable development.

# 3 Available solutions and outlook

## 3.1 On agricultural inputs

The quality and quantity of agricultural production depend on inputs, in particular agricultural inputs. In this respect, digitization focuses on the following elements:

**Production and marketing of seeds and seedlings.** It's well known that when seeds and seedlings are available, of good quality, accessible to growers and well used by them, agricultural production is boosted. Digitization appears to be a catalyst for the development of sustainable agriculture. In fact, the use of drones and sensors offers numerous advantages in terms of monitoring plots, calculating dosages when spraying, and anticipating invasions of bio-aggressors and plant diseases [15]. The possibility of planting cover crops while protecting crops and soil, facilitating access to difficult plots and exploring vast areas. Once the seeds have been produced, the existence of a virtual market via platforms is an asset for interconnecting seed suppliers and demanders, reducing information asymmetries and seed losses.

**Monitoring phytosanitary operations.** For a more resilient agriculture, it is imperative to monitor phytosanitary interventions and therefore elements that require information systems on warning, monitoring, pest inventories, exploitation and dissemination of phytosanitary information. For new-generation agriculture, digitalization is still called upon at this level. For example, the implementation of an electronic pest monitoring and alert system enables control of one aspect of phytosanitary intervention. Similarly, the digitized collection and electronic dissemination of phytosanitary information are major assets in the fight against pests and the anticipation of national plagues.

**Monitoring of agricultural input markets (fertilizers and crop protection products).** When it comes to monitoring the markets for agricultural inputs, in this case fertilizers and crop protection products, digital transformation seems unavoidable given the evolution of information and communication technologies. In this respect, the existence of a digital platform that provides key information on market monographs as well as information on prices and commercial availability of the inputs sold. With this in mind, Cameroon's Ministry of Agriculture and Rural Development has launched the National Observatory for Agricultural Inputs. The aim of the observatory is to produce, analyze and disseminate information on the legal aspects, availability, accessibility, use and access to aid for agricultural inputs in Cameroon. The aim of the observatory is to provide a platform for disseminating information on the price and availability of agricultural inputs, an essential means of improving the economic and geographical accessibility of agricultural inputs.

**Soil management and mapping (land vocation).** In a context where resources are limited and the environment is under heavy pressure, the use of cartography in agricultural production is more than necessary [15]. In the literature, it has been clearly demonstrated that the use of cartography in agricultural production is open to optimal management of resources and limiting waste. For example, Geographic Information System (GIS) maps can be used to analyze topography and soils, and process meteorological data in a more concrete way to decide which crops are best suited to a given plot of land. The existence of a platform that identifies the potential of agricultural land, maps its vocation and highlights the norms for its use is a real opportunity for modern agriculture.

### 3.2 On professional agricultural organizations

With the rise in power of Information and Communication Technologies (ICT) vitalized by Artificial Intelligence, farmer support, agricultural extension and advisory services have entered a new dynamic [16]. This dynamic, characterized by the availability of information and easy access to agricultural data, makes professional agricultural organizations a key link in the digitization of agriculture. Indeed, in developing countries, it is precisely through professional agricultural organizations that farmers, mostly smallholders, update their know-how and share their experiences. Setting up virtual digital spaces for learning and agricultural extension for the benefit of farmers facilitates the promotion of good agricultural practices, thereby increasing their income. In addition, such spaces provide an awareness-raising base to help farmers limit the induction of disasters and climate change and/or protect themselves in return.

**Agricultural extension.** With digitization, the methods used to disseminate extension themes have been revamped, and the management of the research-extension relationship, designed to test research results and then add value to them, has been enhanced. For example, in conducting participatory diagnoses or school fields, digitization makes it possible to achieve objectives over a wider geographical area, at lower cost and in a shorter time. It also makes it easier to mobilize the results of agricultural research. It is also natural to think that the digitization of a personalized agricultural calendar (by agro-ecological zone and by production basin) and made dynamic by regular updating, in view of climatic changes, is an undeniable added value of digital technology.

**Farm management.** With the advent of new technologies, farm management - whether large, medium or small - is taking on a whole new dimension. Thanks to the Internet of Things, geographic information systems and artificial intelligence, farmers can monitor weather conditions, soil moisture levels and crop yields in real time. Integrated management platforms enable crop rotation planning. In addition, advisory support for farms and the strengthening of their managerial capacity is made easier with digital technologies such as e-learning, which offers a panorama of training courses enriched with tutorials and modern communication tools. In Cameroon, there is an observatory of professional organizations which sets up and regularly updates databases on professional organizations. The observatory, housed at the Ministry of Agriculture, uses this data to monitor and analyze the evolution of professional organizations for dissemination and decision-making purposes. As data storage and exploitation are the first heirs of the digital age, this observatory is now being given a new lease of life and is regaining its full consistency.

**Agricultural mechanization.** In most developing countries, the rate of agricultural mechanization is still low, and not very responsive to digital technology. Yet digital technology offers an unprecedented opportunity for a new generation of agriculture, where technology and nature come together to feed a growing world [6, 17]. Today's agriculture in Cameroon is characterized by the coexistence of peasant, artisanal and conventional mechanical systems, with the use of traditional equipment such as animal traction. It is gradually migrating towards modern mechanical systems, but digital inking is residual or even non-existent.

### 3.3 Regulation and quality control of agricultural inputs and products

When it comes to the regulation and quality control of agricultural inputs and products, digital technology is once again called upon for efficiency and effectiveness in the process of certification, homologation and attestation of quality. To ensure that the outputs of Cameroon's agriculture meet conventional regulatory and quality standards, seeds and seedlings used as inputs must be certified. Pesticides for agricultural use must be registered, as must the varieties and species of seeds and seedlings. Phytosanitary products must also be controlled at Cameroon's borders.

To make the most of digitalization, it would not be too much to ask for platforms dedicated to regulation and quality control. A perfect illustration of this is the creation of a digital database accessible online, presenting standards and regulations relating to fertilizers and phytosanitary products, and enabling crop declaration and monitoring of the seed and plant certification process. In addition, Cameroon's agricultural services have a national laboratory for the diagnostic analysis of agricultural products and inputs, which tests, among other things, the attributes of seeds and seedlings that are not visible to the naked eye. These attributes generally relate to germination capacity, specific purity, sanitary condition, moisture content, weight, viability and so on. It also helps strengthen the technical capabilities of inspectors, analysts, controllers and seed laboratory technicians. Such a laboratory will be even more useful if it is equipped with a digital platform that facilitates dissemination of the results obtained, and offers e-learning services for capacity-building of resource persons.

### 3.4 On the living environment of farmers

Whether it's a question of technical and topographical studies, agricultural hydraulics or improving the living environment of farmers, we can't help but say that digitization is an opportunity offered to these aspects. In fact, Cameroon has an observatory for

improving the living environment in rural areas. This observatory is in a position to benefit from the opportunities offered by digitization in the same way as the previously mentioned observatories. As a digitized observatory, it collects, processes, analyzes and disseminates data relating to the improvement of farmers' living environment. As such, it offers a double advantage: on the one hand, farmers use this platform to acquire the skills they need for an ideal living environment, and on the other, public authorities rely on its indicators to implement policies that generate well-being.
Digitization also makes it easier to maintain and update hydraulic databases, enabling more efficient management of hydro-agricultural projects. Water quality standards for agricultural use are now better disseminated, and in general, technology transfer in irrigation and drainage is better assured, as digitization is a portal that is wide open to learning. As far as technical and topographical studies are concerned, Geographic Information Systems have revolutionized the agricultural sector, facilitating topographical surveys, plot measurements and, in general, the planning of agricultural areas.
Under similar conditions and with the same level of investment, production targets are better met with digitization than in a non-digitized context. Time, yield and quality are all improved.

### 3.5 On agricultural surveys and statistics

Digitization, the driving force behind data science, has propelled agricultural statistics and surveys into a new era characterized by unprecedented precision, speed and efficiency. Data collection is digitized, data processing methods are innovative and adaptable to massive data.
The digitalization of Cameroonian agriculture, within the framework of agricultural surveys and statistics, marks a major step towards more efficient management and better planning of the sector. Indeed, the increased use of digital technologies in the collection, analysis and dissemination of agricultural data makes it possible to respond to the structural challenges facing Cameroonian agriculture, in particular the weakness of traditional statistical systems. Through the Department of Agricultural Surveys and Statistics of the Ministry of Agriculture and Rural Development (MINADER), digital platforms are being deployed to facilitate the collection of data on agricultural production, climatic conditions and farm profitability. These tools allow, for example, real-time data entry via mobile applications by field agents, who transmit the information directly to the databases. This represents a real opportunity to quantitatively and qualitatively improve the supply of agricultural data to decision-makers and researchers.

## 4 Key challenge

Developing countries are, at the foot of a mountain, ready to undertake the ascent towards the digital agricultural revolution. But the challenges are colossal [8]. In general, limited budgets, failing energy and telecommunications infrastructure and isolated rural areas are all ills that hinder the march towards the digitalization of their agriculture.

Though agricultural digitization in Cameroon presents undeniable opportunities for modernizing the sector and improving resource management, it faces several major challenges. Firstly, limited access to the Internet and adequate technological infrastructure in rural areas remains a considerable obstacle. Secondly, producers' digital illiteracy is a further obstacle. The majority of farmers, especially those on small farms, lack adequate training in the use of new technologies, which limits the effectiveness of the digital solutions on offer. Finally, there are growing concerns about data security, the confidentiality of agricultural information and the management of the large quantities of data collected. These challenges, though numerous, are not insurmountable; they require a strategic vision to truly make digitalization a lever for sustainable and inclusive development.

## 5 Conclusion

The results of our investigation show that numerous digitization initiatives have been undertaken for Cameroon's agricultural sector, and their implementation initiated by the government and its partners. However, progress in terms of transferring digital technology to farmers, professional agricultural organizations, agro-industries and other stakeholders remains embryonic. Limited access to energy and telecoms infrastructures in some parts of the country has so far hampered the transmission and appropriation of digital know-how. Looking at the organizational chart of the Ministry of Agriculture and Rural Development [18], as well as the activities implemented in its projects and programs, we can see that Cameroon is planning to set up a number of innovative platforms with a digital outlook. These platforms, housed in observatories, laboratories and information systems, are designed to manage every link in the agricultural chain, from seeds to the marketing of agricultural products. In this way, they connect the players involved in each link of the agricultural sector. By implementing these platforms, Cameroon will be taking a giant step towards digitizing its agriculture**.**

## 6 References

[1] Foods and Agriculture Organizations of the United Nations, “The future of food and agriculture - drivers and triggers for transformation.” 2017. [Online]. Available: https://www.fao.org/agrifood-economics/publications/detail/en/c/1622632/

[2] Foods and Agriculture Organizations of the United Nations, “The future of food and agriculture - trends and challenges.” 2022. Accessed: Sep. 12, 2024. [Online]. Available: https://www.fao.org/global-perspectives-studies/resources/detail/en/c/458158/

[3] The World Bank, “Digital agriculture: profil du secteur.” 2022.

[4] ONU Commerce et Développement (CNUCED), “Rapport sur le Développement Economique en Afrique.” 2022. [Online]. Available: https://unctad.org/fr/publication/rapport-2022-sur-le-developpement-economique-en-afrique

[5] United Nations Development Programme, “UNDP Annual Report 2022.” 2022. [Online]. Available: https://www.undp.org/publications/undp-annual-report-2022

[6] C.-E. NEPAD, “Accélération de la transformation numérique inclusive des chaînes de valeur agricoles,” Nov. 2020.
[7] Ministry of Economy, Planning and Regional Development (MINEPAT), “NATIONAL DEVELOPMENT STRATEGY 2020-2030 - Cameroon.” 2020. [Online]. Available: https://minepat.gov.cm/fr/ova_doc/strategies-sectorielles-de-developpement/
[8] D. Mhlanga and E. Ndhlovu, “Digital Technology Adoption in the Agriculture Sector: Challenges and Complexities in Africa,” *Hum. Behav. Emerg. Technol.*, vol. 2023, pp. 1–10, Sep. 2023, doi: 10.1155/2023/6951879.
[9] Ministry of Agriculture and Rural Development, “Agricultural Production Support Program in Cameroon (PARPAC).” 2022.
[10] Ministry of Posts and Telecommunications of Cameroon, “Digital Transformation of Cameroon Project.” 2021. [Online]. Available: https://documents1.worldbank.org/curated/en/099725001202242928/pdf/PPA0P1732400SE00Cameroon02021011017.pdf
[11] Ministry of Agriculture and Rural Development, “Central Africa BACKBONE (CAB) Project.” 2020. [Online]. Available: https://www.afdb.org/fileadmin/uploads/afdb/Documents/Project-and-Operations/Cameroon-_Central_Africa_Backbone__CAB__Project_Cameroon_Component.pdf
[12] The Ministry of Trade - Cameroon, “Cameroon Agropastoral Portal CamAgro.” 2015.
[13] African Development Bank Group, “Agriculture and agribusiness.” 2022. [Online]. Available: https://www.afdb.org/en/agribusiness
[14] Foods and Agriculture Organizations of the United Nations, *The state of food security and nutrition in the world 2022*. FAO; IFAD; WHO; WFP; UNICEF;, 2022. doi: 10.4060/cc0639fr.
[15] A.-R. Abdulai, J. Pulido-Castanon, E. R. Duncan, S.-L. Ruder, K. B. K. C., and E. Fraser, “Will agricultural digitalization deliver relative advantages in quality of work, productivity, profitability, return on investments, and reliability? Perceptions of Canadian producers,” *Cogent Food Agric.*, vol. 10, no. 1, p. 2422529, Dec. 2024, doi: 10.1080/23311932.2024.2422529.
[16] M. Carolan, “Digital agriculture killjoy: Happy objects and cruel quests for the good life,” *Sociol. Rural.*, vol. 63, no. S1, pp. 37–56, Feb. 2023, doi: 10.1111/soru.12398.
[17] P. K. Dutta and S. Mitra, “Application of Agricultural Drones and IoT to Understand Food Supply Chain During Post COVID-19,” in *Agricultural Informatics*, 1st ed., A. Choudhury, A. Biswas, M. Prateek, and A. Chakrabarti, Eds., Wiley, 2021, pp. 67–87. doi: 10.1002/9781119769231.ch4.
[18] République du Cameroun, “Décret n° 2005/118 du 15 avril 2005 portant organisation du Ministère de l’Agriculture et du Développement Rural.” Apr. 2005.

# Bibliometric and economic analysis in precision agriculture

Vinablo K. Dominique DAGBELOU[1]; Souleymane BAH[2]; Souleymane A. ADEKAMBI[1] and Jacob A. YABI[3]

[1] Institut Universitaire de Technologie, Université de Parakou, Parakou, Benin
[2] Institut de Mathématiques et de Sciences Physiques, Université d'Abomey-Calavi, Dangbo, Benin
[3] Faculté d'Agronomie, Université de Parakou, Parakou, Benin
dagbeloud@gmail.com

**Abstract.**

This article aims to better understand as well as the evolution of research status through the available literature on economic analysis in Precision Agriculture (PA). PA is used to improve agriculture processes. Economic analysis of Precision Agriculture scientific papers' are reported and discussed. The Scopus and Dimensions data based were used to obtain the research records under study. Indicators of scientific productivity; collaboration between counties and research impact were evaluate through economic analysis.

The keywords included in the publications and subjects' areas under which the research was published were also evaluated through PA economic analysis. A total of 112 articles were analyzed from 1995 to 2024. The most productive journal were Precision agriculture (13); Computers and electronics agriculture (4); Fields crops research (4); and Agricultural Water research (3). The most keywords were precision agriculture (61); economic analysis (59), Agriculture (23); Irrigation (20); and crop yield (19). Citation countries were classified with United States at first place; and Australia; Malaysia; Spain and India were arrived in second to fifth place respectively. There is no collaborative study between countries.

**Keywords:** Precision agriculture; economic analysis; cost analysis.

## 1 Introduction

Human subsistence increase pressure for food security and sustainability as well as a need to halt environmental degradation has focus attention on the efficient use of farm resources (Tey and Brindal; 2012). Smart Agriculture, known as Smart Farm Technology for sustainable farmland, is the farm management approach where decision making relies on information-based knowledge. It describes an advanced type of farming technology utilizing robot technology and Information and Communication Technology (ICT) to promote labor saving, precision and high-quality production Ochiai (2023); Iba and Lilavanichakul (2023). For, Ochiai (2023); smart agriculture, is expected to reduce the labor load and working time for farm production, improve farm profit through expanding farm size and enable sustainable agriculture farmland Avolio et al. (2014); Lieder and Schroter-Schlaack (2021). For example; in Japan; Smart Agriculture improves farm productivity by reducing labor costs and working time Ochiai (2023); Iba and Lilavanichakul (2023). GPS machines such as tractors or plantors, remote controlled weeding machines, drones for spraying chemicals and equipment can make more efficient for farmers Ochiai (2023).

## 2 Materials and methods

The bibliometric database was compiled by searching the Scopus abstract and citation databases for key words and variants on 17 October 2023, using the following query string (TITLE-ABS-KEY("economic analysis") OR TITLE-ABS-KEY("economic study") OR TITLE-ABS-KEY("economic evaluation") OR TITLE-ABS-KEY("financial analysis") OR TITLE-ABS-KEY("costs and benefits evaluation") OR TITLE-ABS-KEY("commercial analysis")) AND (TITLE-ABS-KEY("precision agriculture") OR TITLE-ABS-KEY("site-specific crop management") OR TITLE-ABS-KEY("site specific crop management") OR TITLE-ABS-KEY("precision crop management") OR TITLE-ABS-KEY("site-specific agriculture") OR TITLE-ABS-KEY("site specific agriculture") OR TITLE-ABS-KEY("site-specific farming") OR TITLE-ABS-KEY("site specific farming") OR TITLE-ABS-KEY("as-needed farming") OR TITLE-ABS-KEY("prescription farming") OR TITLE-ABS-KEY("smart farming")).

The selection and structure of keyword used during the search was an iterative process guided by the authors' experiences in this particular research focus area and previous literature identified through preliminary searches in google scholar. The search was conducted without applying any constraints on the timespan; however, articles that were not published in accredited peer-reviewed journals and not written in English were exclude. The search in Scopus and Dimension results returned 112 articles.

We replaced 15th International Congress on Agricultural Mechanization and Energy in Agriculture, ANKAgEng 2023 by two articles such as 1- Analysis of Factors Affecting Farmers' Intention to Use Autonomous Ground Vehicles Johnny Waked, Gabriele Sara, Giuseppe Todde, Daniele Pinna, Georges Hassoun, Maria Caria; 2- Economic Analysis of Subsurface Drainage Systems in North Central Iowa Kapil Arora, Kelvin Leibold.

We removed some articles which are not relied on the area:

Antecedents of smart farming adoption to mitigate the digital divide – extended innovation diffusion model

2- Comparison of uniform and variable rate nitrogen and phosphorus fertilizer application for grain sorghum

3-Developing and testing an algorithm for site-specific N fertilization of winter oilseed rape

4- Development of an automated slope measurement and mapping system

5- Fossil energy usage for the production of baby leaves

6- Multidisciplinary studies on sustainable nitrogen fertilisation considering the potential of satellite-based precision agriculture; [Multidisziplinäre Untersuchungen zur nachhaltigen Stickstoffdüngung unter Berücksichtigung der Möglichkeiten der satellitengestützten Präzisionslandwirtschaft]

7- Smart green house for controlling & monitoring temperature, soil & humidity using IoT.

At the end of the process of screening 112 articles are retained for this bibliometrics analysis.

For data analysis, we use R bibliometrix package.

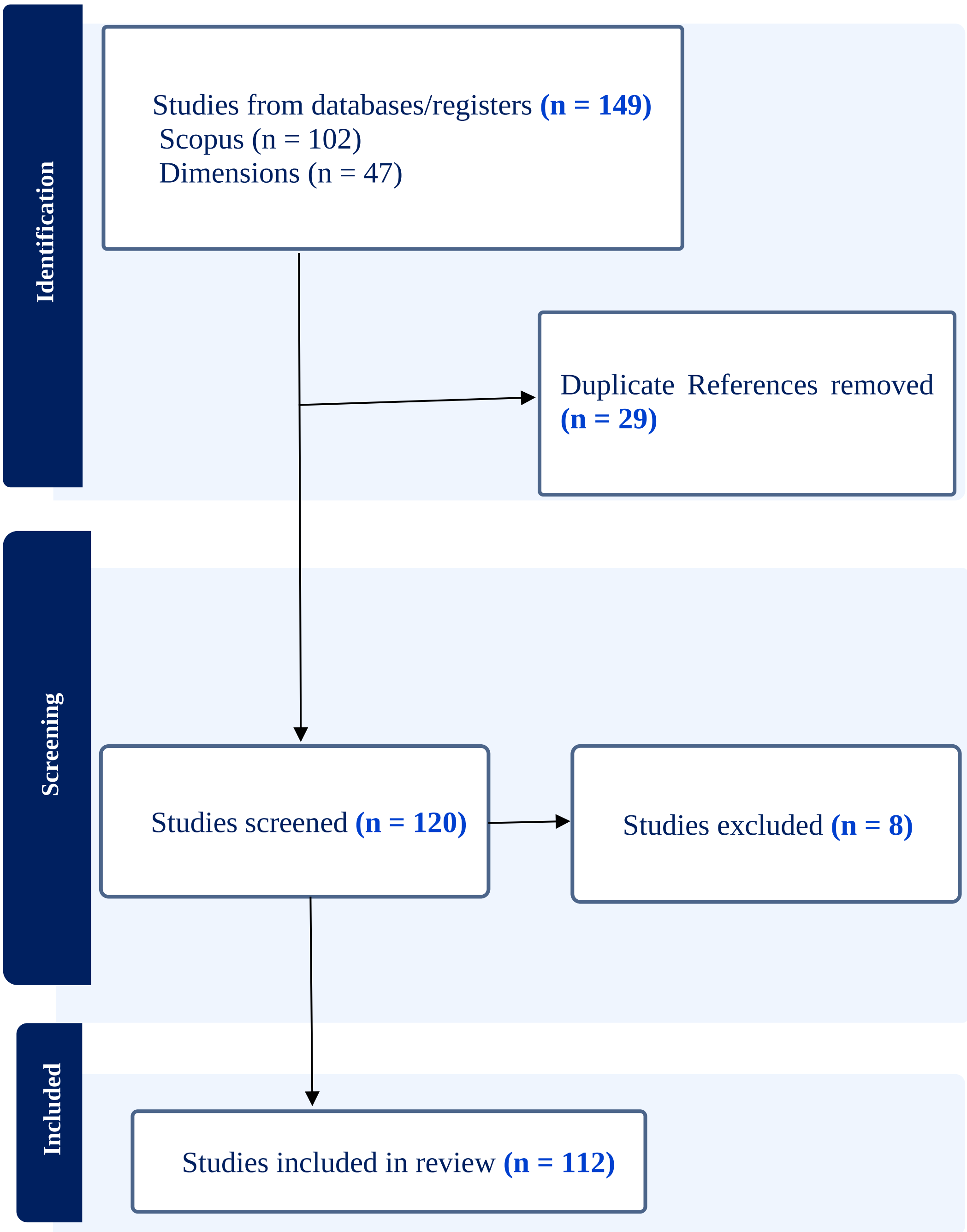



**Fig. 1.** Screening process

# 3 Results

## 3.1 Historical evolution

A summary of the key statistics regarding the final literature dataset is provided in table 1. Research on smart agriculture economic analysis first engaged in 1995 and has been gradually increased with a compound annual growth rate of 2,42%. The highest production was achieved in 2012, representing 46% (For confirmation) of the total publication Figure 2. The period of 2005; 2009 and 2012 had the highest average total at 123 citations per publication, peaking at 138, 33 - 80,33 - 62, 67. In comparison, the highest average TCs per year occurred in 2012 at 138,33.

## 3.2 Most influential journals

The final literature database consisted of 64 journals with 93 publications on economic analysis of precision agriculture. The journal Precision agriculture, Computers and electronics and Field crop research have the highest number of articles accounting for 65% of the total publications. Precision agriculture also retain their position at the top of the ranking for TCs with 138. Therefore, this is the dominant journal in this particular domain research focus area.

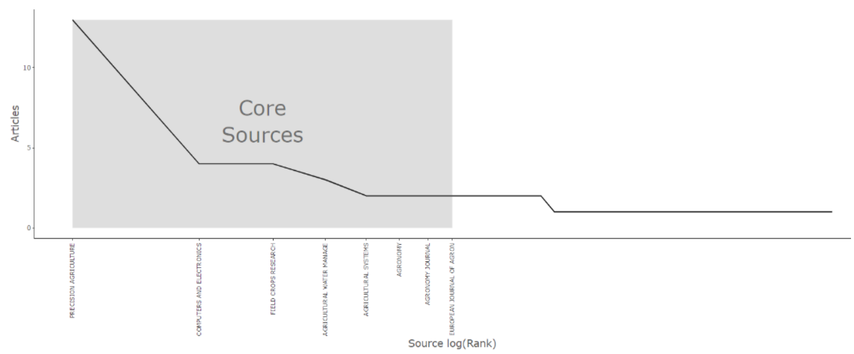



**Fig. 2.**

According to Bradford's law, the are broadly three zones that can categorize the frequency of citations emanating from journals for particular research focus area. Zone 1 represents the most influential journals as they are cited most frequently in that subject area and likely attract the greatest interest from researchers. Zone 2 and zone 3 represent the journals with the average and least citations, respectively (Abafe, Bahta and Jordan 2022).

### 3.3 Analysis of publications by country

Regarding the geographic distribution of published research on the use of economic analysis of precision agriculture; 30 countries have been involved in precision agriculture economic analysis. USA (26), Australia (9), Brazil (8), Italia (7), Spain (4) and Danmark (3) are the only countries to have produced more than 2 publications on the economic analysis of precision agriculture and account for 63% of the total number of the publications. Only USA features consistently in the tops countries. Collaborations between authors have been mostly restricted to the countries in which they reside, however there are no international collaboration between authors while economic analysis is very important to be analyzed in smart agriculture domain.

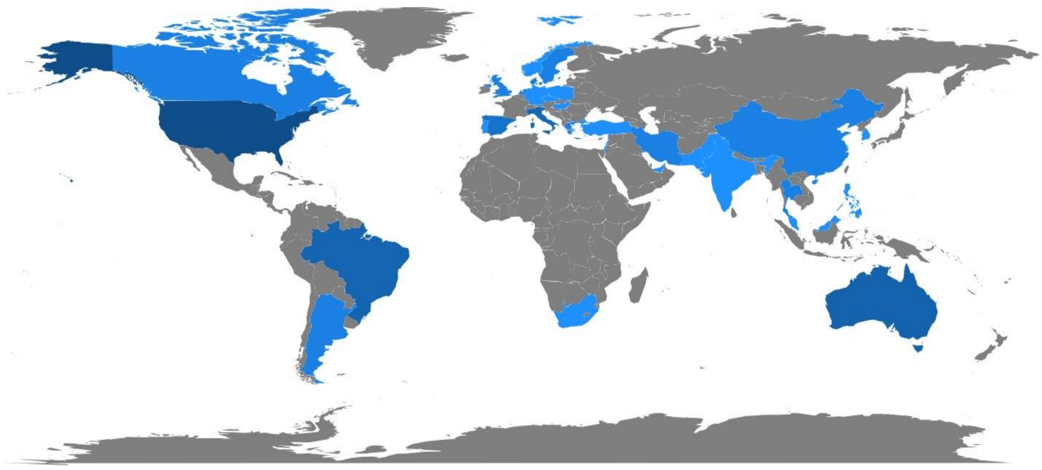


**Fig. 3.** Global publications and collaborations, where by the darker the shade of blue, the larger the number of publications.

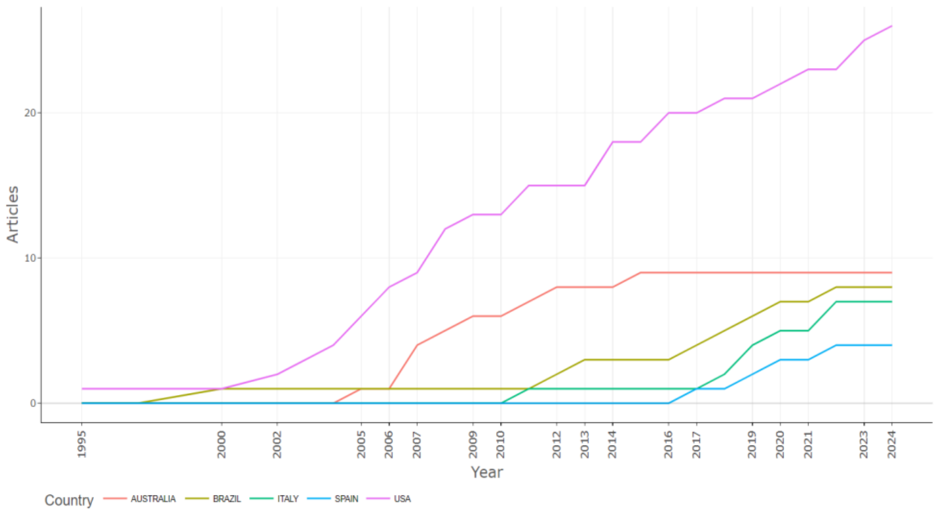



**Fig. 4.** Countries' productions over times

### 3.4 Most influential Authors and citation analysis

Some authors have been known by the citation index. The first author, Robertson has written more than four articles, second by Gandorfer, Lite and Sadler known with three articles. For two articles, some authors are known. We have Abuzar; Best; Bullock; Buschermohle; Camp and Chandra.

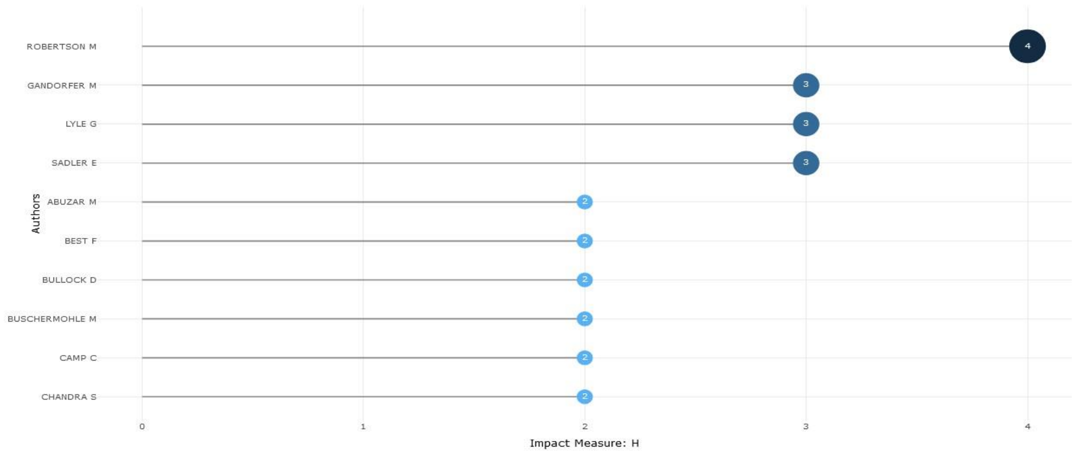



**Fig. 5.**

### 3.5 Analysis of keywords frequency, growth and co-occurrence

Perceptual cart based on two axes such as relevance degree and development degree show the following results. The first category is relative to the most important research themes which are sustainable agriculture and development; crop yields; precision agriculture; economic analysis; soils; decision support systems; land use and human procedures. The second category concerns cost benefit analysis; earnings drainage; agricultural technology; technology adoption; investment; ground vehicles and pesticides. Through these results, we are in force to thank that agricultural and development sustainability are important research theme in the literature in one hand Avolio et al. (2014); Lieder and Schroter-Schlaack (2021) and the agricultural technology' cost benefit analysis in second hand Ochiai (2023); [1].

.

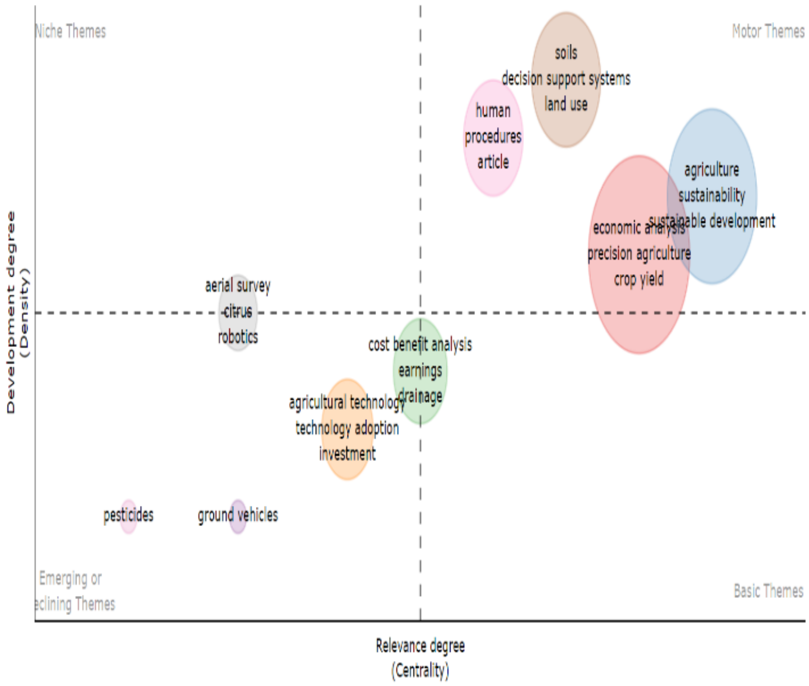



**Fig. 6.** Perceptual cart

Thematic map show important themes which take authors attention in this field.


**Fig. 7.** Thematic map

### 3.6 Precision agriculture economic analysis

The economic analysis of precision agriculture, supported by bibliometric research, highlights the significant potential of these technologies to increase productivity, reduce costs and efficiency gains, and improve environmental sustainability.

One of the most compelling economic reasons for adopting precision agriculture is the reduction in input costs. Precision Agriculture technologies enable more efficient use of resources such as water, fertilizers, pesticides, and labor, leading to cost savings. Precision agriculture can lead to increased crop yields by enhancing resource management and improving the ability to monitor crop health efficiency [2-4]; [8]. This is particularly important as the global demand for food is rising due to population growth. Another important economic benefit is the positive impact of precision agriculture on environmental sustainability, which can have indirect long-term financial benefits for farmers [5-7].

However, the adoption of PA depends on various factors, including farm size, access to capital, the region's technological infrastructure, and the ability to integrate new systems with existing practices [13].

A thorough economic analysis, based on both cost-benefit studies and ROI models, is essential for farmers to make informed decisions about investing in precision agriculture. Additionally, policy makers play a crucial role in supporting the adoption of these technologies through incentives, subsidies, and training programs.

In the coming years, the economic landscape of precision agriculture will likely continue to evolve, with more focus on affordability, scalability, and the broader economic and environmental impact of these technologies. [7-9]. As bibliometric analysis reveals, the growing body of research will further clarify how precision agriculture can be economically integrated into diverse farming systems around the world, ultimately making agriculture more efficient, sustainable, and profitable [5-7].

## 4 Discussions

Economic analysis of smart agriculture remained capital for sustainable agriculture [2]. Bibliometric review show that sustainability of agriculture and sustainable development are important and need smart technologies tools for agricultural efficiency [2-4] ; [5-7]. Many scientists are very preoccupied of agricultural effect on environment degradation; climate change; degradation of soil cover; destruction of natural ecosystem etc. [7-9]; [5]. For [10] increasing production of crops, livestock and aquatic products for food security must be the objective of smart agriculture which haven't a negative impact on environment [10-12]. For productivity and yields improvement; some research focused on smart farming technology adoption [13]. Economic reasons are main reason of smart agriculture adoption [13]. For them; adoption of smart agriculture in two regions (North and south) in Germany provides economic gains for both technical (sensor' based technology and mapping based technology) for farmers; however they don't find the neighboring farmers adopted the technologies [13]; [14].

## 5 Conclusion

In conclusion; this bibliometric analysis illuminates the dynamic evolution of smart agriculture economics analysis. Since smart agriculture is known as important for farmland efficiency; it was become an obligation to examine his economic contribution Ochiai (2023) [1]. This bibliometric analysis sheet the light on what is known and what is to be known for scientific aspect. The economic contribution of smart agriculture is very important because smallholder struggle for their subsistence, they have little resources to face Smart Farming Technologies charges such as drone acquisition as soon as its' utility is recognized by researchers and agrobusiness actors. As we can imagine smart agriculture need collaboration of scientists, agrobusiness men and government to overcome smart agriculture' obstacles Ochiai [1].

## References

1. Iba, H., Lilavanichakul, A.: Farm Business Model on Smart Farming Technology for Sustainable Farmland in Hilly and Mountainous Areas of Japan, https://www.scopus.com/inward/record.uri?eid=2-s2.0-85151441164&doi=10.3390%2fland12030592&partnerID=40&md5=93b20094cace78ad0808a4ff0c402180, (2023). https://doi.org/10.3390/land12030592.
2. Lampridi, M.G., Kateris, D., Vasileiadis, G., Marinoudi, V., Pearson, S., Sørensen, C.G., Balafoutis, A., Bochtis, D.: A case-based economic assessment of robotics employment in precision arable farming, https://www.scopus.com/inward/record.uri?eid=2-s2.0-

85064178566&doi=10.3390%2fagronomy9040175&partnerID=40&md5=aa614a27b7c0af5fe8914bcd31bfa4e5, (2019). https://doi.org/10.3390/agronomy9040175.
3. Morales-Rodríguez, P.A., Cano, E.C., Villena, J., López-Perales, J.A.: A Comparison between Conventional Sprayers and New UAV Sprayers: A Study Case of Vineyards and Olives in Extremadura (Spain), https://www.scopus.com/inward/record.uri?eid=2-s2.0-85131565555&doi=10.3390%2fagronomy12061307&partnerID=40&md5=527074aa29ab5b9bdc970e111440e5f4, (2022). https://doi.org/10.3390/agronomy12061307.
4. Paraforos, D.S., Vassiliadis, V., Kortenbruck, D., Stamkopoulos, K., Ziogas, V., Sapounas, A.A., Griepentrog, H.W.: A Farm Management Information System Using Future Internet Technologies, https://www.scopus.com/inward/record.uri?eid=2-s2.0-84994716608&doi=10.1016%2fj.ifacol.2016.10.060&partnerID=40&md5=432d8e82d3a79f21d3c01a7817c1b01e, (2016). https://doi.org/10.1016/j.ifacol.2016.10.060.
5. Casson, A., Ortuani, B., Giovenzana, V., Brancadoro, L., Corsi, S., Gharsallah, O., Guidetti, R., Facchi, A.: A multidisciplinary approach to assess environmental and economic impact of conventional and innovative vineyards management systems in Northern Italy, https://www.scopus.com/inward/record.uri?eid=2-s2.0-85131430131&doi=10.1016%2fj.scitotenv.2022.156181&partnerID=40&md5=ec30ba159382b5593652d85912cbe77d, (2022). https://doi.org/10.1016/j.scitotenv.2022.156181.
6. Pedersen, M.F., Gyldengren, J.G., Pedersen, S.M., Diamantopoulos, E., Gislum, R., Styczen, M.E.: A simulation of variable rate nitrogen application in winter wheat with soil and sensor information - An economic feasibility study, https://www.scopus.com/inward/record.uri?eid=2-s2.0-85107600026&doi=10.1016%2fj.agsy.2021.103147&partnerID=40&md5=fca9de3378b10225b4d812aff3f9e398, (2021). https://doi.org/10.1016/j.agsy.2021.103147.
7. Rulev, A., Rulev, G.: Ecological and Economic Aspects of Land Desertification. Vestnik Volgogradskogo gosudarstvennogo universiteta. Ekonomika. 158–169 (2019). https://doi.org/10.15688/ek.jvolsu.2019.3.14.
8. Robertson, M.J., Llewellyn, R.S., Mandel, R., Lawes, R., Bramley, R.G.V., Swift, L., Metz, N., O'Callaghan, C.: Adoption of variable rate fertiliser application in the Australian grains industry: Status, issues and prospects, (2012). https://doi.org/10.1007/s11119-011-9236-3.
9. Robertson, M.J., Lyle, G., Bowden, J.W.: Within-field variability of wheat yield and economic implications for spatially variable nutrient management. Field Crops Res. 105, 211–220 (2008). https://doi.org/10.1016/j.fcr.2007.10.005.
10. Mowla, K.E.A.E., Aziz, H.H.A.E.: Economic analysis of climate-smart Agriculture in Egypt. Egyptian Journal of Agricultural Research. 98, 52–63 (2020). https://doi.org/10.21608/ejar.2020.101423.
11. Maine, N., Nell, W.T., Lowenberg-DeBoer, J., Alemu, Z.G.: Economic analysis of phosphorus applications under variable and single-rate applications in the bothaville district, (2007). https://doi.org/10.1080/03031853.2007.9523785.
12. Gonçalves, J.R.M.R., Ferraz, G.A.S., Reynaldo, É.F., Marin, D.B., Ferraz, P.F.P.: Comparative economic analysis of soil sampling methods used in precision agriculture, (2020). https://doi.org/10.1590/0001-3765202020190277.
13. Meyer-Aurich, A., Gandorfer, M., Heißenhuber, A.: Economic analysis of precision farming technologies at the farm level: Two german case studies, https://www.scopus.com/inward/record.uri?eid=2-s2.0-84892138568&partnerID=40&md5=1695204c609c22410d56524f43fa508e, (2008).

14. de Amorim, F.R., Patino, M.T.O., de Abreu, P.H.C., Santos, D.F.L.: Economic and risk evaluation of fertilizer application systems in the sugar cane crop: Fixed rate by average and variable rate technology, (2019).