Cocoa Pods Diseases Detection by MobileNet Confluence and Classification Algorithms

Diarra MAMADOU[1], Kacoutchy Jean AYIKPA[2], Abou Bakary BALLO[3], Brou Médard KOUASSI[4]
LaMI, Université Felix Houphouët-Boigny, Abidjan, CÔTE D'IVOIRE[1, 3, 4]
UREN, Université Virtuelle de Côte d'ivoire, Abidjan, CÔTE D'IVOIRE[2]
ImViA, Université de Bourgogne, Dijon, FRANCE[2]
LMI, Université Péléforo Gon Coulibaly, Korhogo, CÔTE D'IVOIRE[3]

***Abstract*—Cocoa cultivation is of immense importance to the people of Côte d'Ivoire. However, this culture is experiencing significant challenges due to diseases spread by various agents such as bacteria, viruses, and fungi, which cause considerable economic losses. Currently, the methods available to detect these cocoa diseases force farmers to seek the expertise of agronomists for visual inspections and diagnostics, a laborious and complex process. In the search for solutions, many studies have opted for using convolutional neural networks (CNNs) to identify diseases in cocoa pods. However, an essential advance is to develop hybrid approaches that combine the advantages of a CNN with sophisticated classification algorithms. This research stands out for its innovative contribution, combining MobileNetV2, a convolutional neural network architecture, with algorithms, such as Logistic Regression (LR), K Nearest Neighbors (KNN), Support Vector Machines (SVM), XGBoost, and Random Forest. The study was conducted in two distinct phases. First, each algorithm was evaluated individually, and then performance was measured when MobileNetV2 was merged with the algorithms mentioned. These hybrid approaches complement and amplify MobileNetV2's capabilities. To do so, they draw on MobileNetV2's inherent capabilities to extract key features and enhance information quality. By combining this expertise with the classification methods of these other models, hybrid approaches outperform individual techniques. Accuracy rates range from 72.4% to 86.04%.This performance amplitude underlines the effectiveness of the synergy between the extraction characteristics of MobileNetV2 and the classification skills of other algorithms.**

Keywords*—*Cocoa pods diseases; MobileNetV2; classification algorithms; machine learning; hybrid method

I. Introduction

Cocoa pod diseases are a significant problem for farmers and the cocoa industry [1]. These diseases can lead to yield loss, reduced cocoa bean quality, and higher production costs. Many methods are available to combat cocoa pod diseases, including pesticides, fungicides, and organic practices. However, these methods can be costly and difficult to apply.

Computer vision, generally based on machine learning and deep learning [2], is now being exploited for various agriculture, botany, and ecology tasks. These tasks include assessing the health of plants in our various crops. This new technology is proving necessary to improve crop yields in general. Researchers have already conducted numerous studies to identify disease risk factors, such as adverse weather conditions [3], insect pests, and weeds. By identifying these risk factors, farmers can take steps to reduce or eliminate them, thus helping to prevent disease.

Indeed, Machine Learning and Deep Learning algorithms have shown enormous potential in agriculture by helping to develop effective treatments for crop diseases. They provide an innovative and powerful approach to combating crop diseases. Their ability to process complex data, detect anomalies early, and customize treatments significantly benefits crop health, agricultural sustainability, and food security [4] [5].

Applying Machine Learning (ML) and Deep Learning (DL) algorithms for detecting cocoa pod diseases offers tangible benefits for improving cocoa quality, increasing selling prices, and improving farmers' incomes. This approach supports farmer profitability and contributes to the sustainability of the cocoa sector by enhancing the quality and reputation of the cocoa produced.

More specifically, this study contributes to the sustainability of the cocoa sector by improving the quality and reputation of the cocoa produced. Early detection of disease enables farmers to take corrective action, reducing production losses and improving cocoa quality

Our research focuses mainly on identifying and detecting diseases affecting cocoa pods. It has several important implications. Firstly, it shows that the hybrid approach is promising for cocoa pod disease detection. Secondly, it suggests that CNNs can be used to improve the performance of conventional machine learning algorithms. Finally, it paves the way for new applications of machine learning and deep learning in the cocoa sector. Our contributions will be broken down into the following aspects:

As a first step, we will extract images from each pod using data augmentation techniques. This step will be crucial to establish the dataset that will serve as the basis for our study.

We will combine classic Machine Learning algorithms and a convolutional neural network (CNN) such as MobileNet.

We will explore hybrid approaches aimed at merging the capabilities of the MobileNetV2 network with the set of classic Machine Learning algorithms already in use.

Finally, we will evaluate the performances of different methods by analyzing the impact of integrating MobileNet in the fusion of the approaches.

The organization of our study is outlined as follows. Following the introductory section outlining the issue tackled in this paper (Section I), we delve into existing research on identifying specific cocoa diseases (Section II). Subsequently, we present and expound upon the methodology in Section III. Elaborate findings are disclosed in Section IV.

A comprehensive discussion is laid out in Section V, and ultimately, Section VI addresses the research goals and derives conclusions from the conducted study.

II. Related Work

Cocoa farming is a vital sector for the Ivory Coast and an essential resource for our farmers. Unfortunately, this crop is sometimes threatened by diseases that cause enormous losses for our farmers and Côte d'Ivoire, the world's leading cocoa producer. To preserve the gains made and further improve the production of this crop, a great deal of work has been carried out.

Godmalin et al. [6] carried out a study based on a deep learning algorithm to tackle the automatic classification of the state of a cocoa pod. Their experimental research method relies on a convolutional neural network for training. The model can classify three states of a given cocoa pod image: healthy, attacked by black pod disease, and shot by a pest. The results of their experiment showed an accuracy of 94%. However, the study did not assess the impact of weather conditions on the algorithm's performance. The algorithm's performance may vary according to weather conditions. Please do not revise any of the current designations. Sandra Kumi et al. have proposed a method using machine learning techniques to detect and diagnose two significant diseases affecting cocoa production, namely Swollen Shoot and Black Pod [7]. A mobile application with integrated ML techniques is offered to cocoa farmers to take a photo of the pod and upload it for diagnosis, which takes place on a cloud backend service. Four CNN models were built and trained for automatic disease detection and diagnosis. The results of this study showed that the MobileNet V2 SSD gave the best accuracy rate, with a score of 80%. However, the study was conducted in a controlled environment, and the algorithm may need to be more accurate in a real-life setting, where conditions can vary.

Amoako et al. [8] studied a model based on VGG19 to classify cocoa diseases using images. Then, other pre-trained models, such as VGG16 and ResNet50, are compared. Their study is a step in the right direction toward developing a technology that could positively impact the cocoa industry. With further research, it is possible to improve the accuracy of machine learning models and make them more applicable to a wide range of environments. This could lead to a significant reduction in losses due to cocoa diseases and an improvement in the quality of the cocoa produced. Basri et al. [9] proposed a study comparing the results of four feature extraction models in the case of early recognition of disease attacks on cocoa fruits. The image extraction models used include:

Local binary pattern (LBP).

Gray level co-occurrence matrix (GLCM).

Hue saturation value (HSV).

Gray level co-occurrence histograms (GLCH).

In addition, the support vector machine (SVM) model was applied to the classification technique to measure the extraction results from the cocoa image dataset. SVM classification results revealed the best performance for HSV feature extraction for all types of SVM kernels applied (linear, RBF, and polynomial), with the highest accuracy being 80.95% for the RBF kernel.

Baba et al. [10] have studied a model based on image processing techniques for identifying early symptoms of pests and diseases in cocoa fruit from mobile applications. This research showed that the system's accuracy in recognizing cocoa fruit-shaped objects reached 100% in identifying cocoa fruit and 83% database for images of normal conditions at 83.75%, disease attack at 84.87%, and pest attack at 80.80%. Their study showed that image-processing techniques can accurately identify early symptoms of pests and diseases in cocoa fruit. However, the study also revealed areas for improvement in this approach, namely that many images can drag the model. In addition, the images need to be of high quality and well-lit. If this is not the case, the model may need to identify pests and diseases correctly. Gunawan et al. set up an expert system based on the Certainty Factor (CF) algorithm to identify the factor of cocoa pests and diseases that cannot be identified and prevented in advance [11]. Based on the results of the accuracy test, the proposed model can produce an accuracy rate of 86.67% in diagnosing pests and diseases on cocoa plants. However, the study used a simple algorithm, the certainty factor, which may not be able to detect all pests and diseases. Mohammad Yazdi et al. [12] have proposed a model for building a system to detect pest and disease types in cocoa pods. This study uses digital image processing techniques to extract color characteristics from digital images of cocoa pods. The method used to extract hue, saturation, and value (HSV) color characteristics and the classification algorithm used is K-Nearest Neighbor (KNN). One hundred fifty images were divided into 70% training data and 30% test data. Based on test results using k values of 5, 7, 11, and 13 in the restraint method, the best accuracy is 84.44% with a k = 5 value. However, the images used in the study are limited to a single type of cocoa tree, which could limit the generalizability of the results to other types of cocoa. It should be noted that the KNN algorithm is used to classify the images, but this simple algorithm may not detect the more complex pests and diseases. Godmalin et al. studied an experimental search method to train a convolutional neural network capable of classifying three levels of cocoa pod infection: low, moderate, and severe [13]. The results of this model achieved 91% accuracy in correctly classifying the condition of cocoa pods.

Overall, this study is a step in the right direction. Still, more work needs to be done to develop automatic detection methods for cocoa pod infection that are reliable and usable in real-life conditions.

III. Material and Method

A. Materials

The database used for our study is Cocoa Diseases [14], an online database launched in 2020 by the Autonomous

University of Bucaramanga, Colombia. It comprises 312 images with 1,591 tagged objects belonging to three classes: healthy, phytophthora, and monilia. The experiments used Python programming on a THINPAD laptop with an Intel(R) Core i7-10700 processor running at 2.90 GHz, 32 GB memory, and a 512 GB SSD hard disk.

B. Methods

The machine learning methods we used in our study are as follows.

1) SVM : A support vector machine (SVM) is a supervised machine learning algorithm [15] that can be used to solve classification and regression problems. SVMs are a generalization of linear classifiers. The principle behind SVMs is to find a hyperplane that optimally isolates data from two classes. The optimal hyperplane corresponds to the one that optimizes the margin between the data of the two classes [16]. The data closest to the hyperplane are called support vectors. SVMs are effective for solving classification and regression problems with non-linear data. They also adapt well to new data. Effective for solving classification and regression problems with non-linear data, it is an algorithm capable of generalizing well to new data. SVMs are used in various applications, such as image classification, facial recognition, object detection, text classification, and regression.

2) Random forest : The Random Forest algorithm is a supervised machine learning algorithm that uses a set of decision trees to make predictions [17]. It is known for its accuracy and robustness and is used in various fields, including classification, regression, and anomaly detection. The algorithm starts by creating a set of decision trees. The number of trees in the set is a parameter the user can adjust. Each decision tree is trained on a random subset of the training data. Data that is not used to train a decision tree is called test data. The algorithm then uses the test data to evaluate the accuracy of each decision tree. The algorithm then averages the predictions of all the decision trees to make a final prediction. The advantage of the Random Forest algorithm is that it can make more accurate predictions than the individual decision trees. The decision trees are uncorrelated, so they don't make the same errors. The Random Forest algorithm is also robust to noisy data. This is because the decision trees in the ensemble can adapt to variations in the data. The Random Forest algorithm is a powerful tool that can be used to solve various machine learning problems. It is renowned for its accuracy, robustness, and ease of use.

3) XGBoost: XGBoost is an improved model of the Gradient Boost algorithm. This machine-learning algorithm can solve common commercial problems using minimal resources [18]. Extreme Gradient is a method used to reduce the number of errors in predictive data analysis. XGBoost is an assembly of decision trees (weak learners) that predict residuals and correct mistakes in previous decision trees. The unique feature of this algorithm lies in the decision tree used. This recently introduced machine learning algorithm has proved very powerful for modeling complex processes in other research fields [19]. It is a robust algorithm that can solve various machine learning problems. It is known for its accuracy, speed, and flexibility.

4) KNN: The KNN (k-nearest neighbor) algorithm is a supervised learning algorithm used for classification and regression [5]. It works by calculating the distance between an unknown point and known points in the training dataset. The k points closest to the unknown point are then used to predict the class or value of the unknown point. The KNN algorithm is non-parametric, meaning that it makes no assumptions about the distribution of the data [2]. It is also a computationally inexpensive algorithm, making it suitable for large quantities of data. It works by calculating the distance between an unknown point and known points in the training data set. The k points closest to the unknown point are then used to predict the class or value of the unknown point. The result is a computationally inexpensive algorithm, making it suitable for large amounts of data.

5) Logistic regression is a supervised learning algorithm [20] used to predict the probability of an observation belonging to a particular class. It is often used for binary and multiclass classification [21]. It is a linear model, which means that the probability of class membership is modeled as a linear combination of the input variables. The probability function is a logistic function, which is a non-linear function that takes a value between 0 and 1 [22]. Logistic regression is a robust algorithm that can solve many problems. The algorithm begins by calculating the model weights. Weights are coefficients that measure the importance of each input variable. The weights are then used to calculate the probability of belonging to each class. The observation is classified in the category with the highest probability. The choice of weights is important for model performance. Weights can be optimized using a technique called gradient descent. Gradient descent is an optimization technique for finding weights that minimize model error. Logistic regression is a robust algorithm that can solve many problems. It is easy to implement and understand, making it popular with data scientists.

6) MobileNetV2: MobileNet V2 is a lightweight convolutional neural network model developed by Google AI [23]. It was first introduced in a research paper published in 2018. MobileNet V2 is based on the original MobileNet architecture but introduces several improvements, namely the use of a new convolution block called "bottleneck" and that of a learning technique called "transfer learning". MobileNet V2 has been designed for mobile devices such as smartphones and tablets. It can achieve performance comparable to larger models while being much lighter and more energy-efficient [24]. The bottleneck convolution block is a convolution block [25] that reduces the width and height of the input signal while retaining the depth. This reduces the number of model parameters while maintaining performance. MobileNet V2 is a powerful and versatile model that can be used for various tasks, such as image classification, object detection, and facial recognition. It is lightweight and energy-efficient, making it ideal for mobile devices.

The methodology adopted in this study is based on several well-defined phases, providing a solid and rigorous foundation for our research. Detection techniques are studied to enable the system to detect and identify cocoa pod diseases. A comprehensive outline of these stages is provided in Fig. 1:

Our methodological approach is based on a series of carefully developed steps:

- Database preparation: In this phase, each image was segmented to identify each pod in different images precisely. To improve the quality of our data, each series of images associated with a pod was subjected to a data augmentation step.
- Data Augmentation: Data augmentation is an essential element of our approach. In this step, random transformations were applied to images. Each image was subjected to random horizontal and vertical flipping, allowing left and right as well as top and bottom inversions. In addition, a random rotation was applied to each image, with a maximum rotation angle of 0.4 radians (approx. 23 degrees) clockwise.
- Data Division: For the training and evaluation of our models, we have divided our data into three sets: the training set, the validation set, and the test set. This division enables us to evaluate and validate the performance of our models.
- Using Various Algorithms: To tackle our pod disease detection task, we adopted a varied approach using traditional Machine Learning algorithms such as SVM, Random Forest, KNN, Logistic Regression and XGBoost, and deep methods based on the MobileNetV2 network.
- Hybrid Methods: In recognition of the complementary power of the approaches, we have also explored hybrid methods that merge the potential of the MobileNetV2 network with the set of Machine Learning algorithms mentioned above.
- Performance Evaluation: To assess the effectiveness of each model, we carefully evaluated their ability to detect pod diseases. This crucial phase enabled us to quantify and compare the performance of each approach.

Our methodology is built on a solid foundation, from careful data preparation to thorough model evaluation to detect pod diseases accurately.

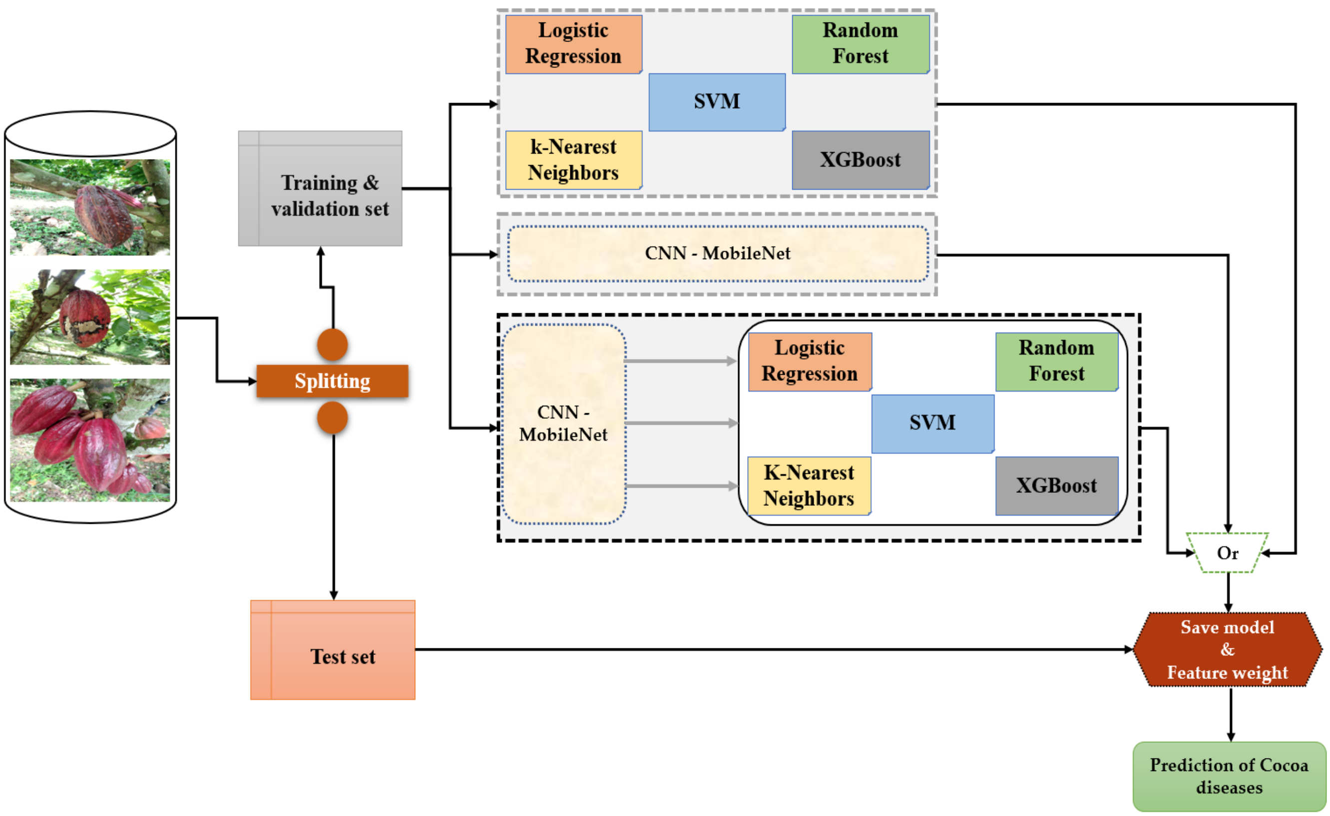

Fig. 1. Illustration of our methodology.

C. Evaluation Metrics

In evaluating the outcomes of our investigation, we employed various metrics. We included accuracy, precision, recall, F1 score, and Matthew's correlation coefficient (MCC) within this set of measures. The differential equations are outlined as follows:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

$$Precison = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1\ score = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (4)$$

$$MCC = \frac{TP*TN-FP*FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (5)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Yi - \hat{Y}i)^2 \quad (6)$$

The Receiver Operating Characteristic (ROC) curve is a graphical representation showcasing the performance of a binary classification model. It illustrates the relationship between the true positive rate (sensitivity) and the false positive rate across various classification thresholds. An ideal ROC curve aligns with the upper left corner of the graph, indicating high sensitivity and specificity.

The confusion matrix is a tabular summary that encapsulates the outcomes of a classification model's predictions. It assesses the model's predictions against the actual values in the dataset, categorizing them into four groups: true positives, true negatives, false positives, and false negatives. The confusion matrix compares the model's precision, recall, specificity, and accuracy.

IV. Results

Our findings will be presented in two distinct sections, each exploring a specific area in depth:

A. Individual Algorithm Performance

The first part of our results will be dedicated to evaluating and comparing the performance of the various Machine Learning algorithms we have used. These algorithms include SVM, Random Forest, KNN, Logistic Regression, XGBoost, and the MobileNetV2 Deep Learning algorithm. Each algorithm has been tested and analyzed, enabling us to determine its specific effectiveness in detecting pod diseases. We will evaluate each algorithm's key metrics, giving us a comprehensive view of its ability to meet this challenge. Table I shows the metric measurements for each model.

Fig. 2 shows the confusion matrices and shows the ROC curves for the three best accuracies.

TABLE I. Case of Individual Algorithm Metric Result

Models	**Accuracy (%)**	**Precision (%)**	**F1 score (%)**	**Recall (%)**	**MCC (%)**	**MSE**
SVM	67,09	66,78	66,78	67,09	48,49	0,81
Random Forest	59,53	63,28	63,28	64,17	43,54	0,91
KNN	46,96	39,54	39,54	46,96	18,07	1,47
Logistic Regression	54,65	53,71	53,71	54,65	29,06	1,08
XGBoost	68,61	67,79	67,79	68,61	50,95	0,77
MobileNetV2	**81,66**	**81,4**	**81,4**	**81,65**	**72,09**	**0,4**

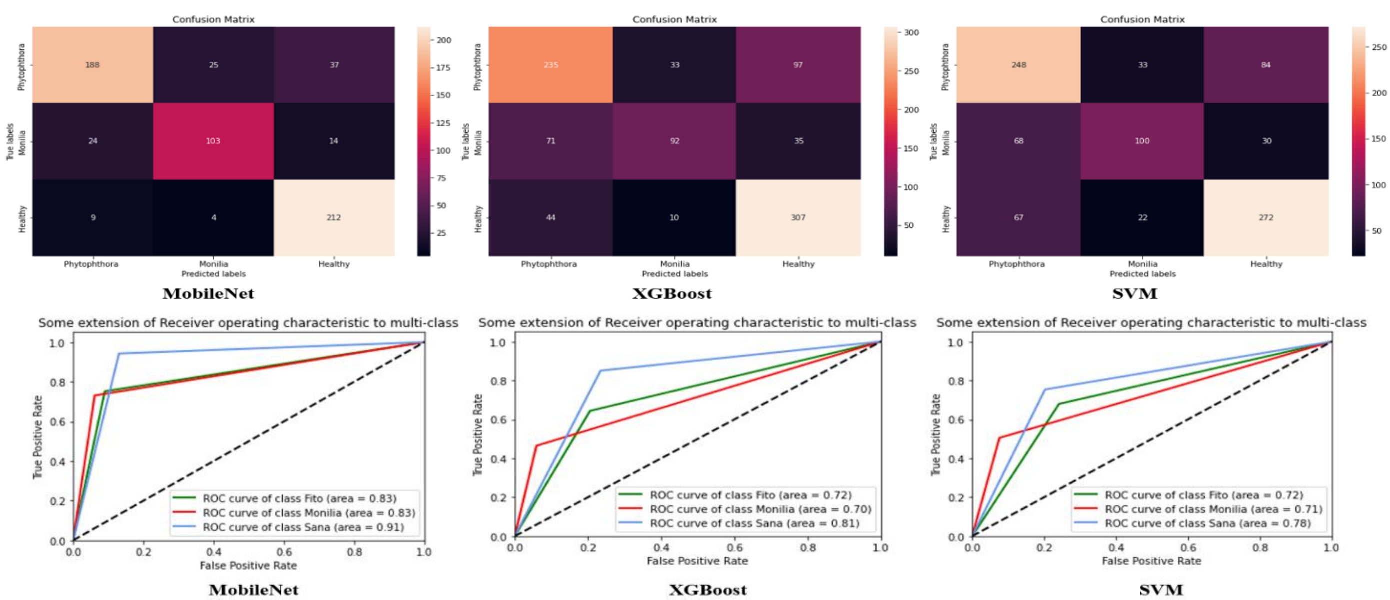

Fig. 2. Confusion matrix and ROC curve of individual algorithm.

The analysis reveals significant variations in the performance of each model. The MobileNetV2 model displays the highest accuracy at 81.66%, closely followed by logistic regression and SVM. However, when considering measures such as recall and F1 score, MobileNetV2 stands out, suggesting its ability to identify true positives well. Furthermore, the MCC, which assesses the overall prediction quality, puts MobileNetV2 in the lead with 72.09%.

On the other hand, KNN's performance could be better regarding precision, recall, and F1 score, suggesting difficulty discerning true positives. The lowest MCC also corroborates this among the models tested.

In sum, these results highlight the superiority of MobileNetV2 in terms of overall performance. Still, they also underline each model's specific strengths and weaknesses in the pod disease detection task.

B. Hybrid Method Performance

The other part of our results will focus on an in-depth examination of hybrid methods. This section will explore the synergies between the Deep Learning-based MobileNetV2 network and the previously mentioned machine learning algorithms. The performance of these hybrid methods will be rigorously analyzed, revealing whether their combination can achieve even higher detection levels. This evaluation will give us a significant perspective on the added value provided by the fusion of these approaches.

Table II shows the metric measurements for each model.

Fig. 3 shows the confusion matrices and shows the ROC curves for the three best accuracies

TABLE II. CASE OF HYBRID METHOD ALGORITHM METRIC RESULT

Models	**Accuracy (%)**	**Precision (%)**	**F1 score (%)**	**Recall (%)**	**MCC (%)**	**MSE**
MobileNetV2 - SVM	**86,04**	**85,88**	**85,88**	**86,03**	**78,53**	**0,33**
MobileNetV2 - RF	77,60	76,44	76,44	77,59	65,6	0,49
MobileNetV2 - KNN	72,4	71,59	71,59	72,4	61,03	0,76
MobileNetV2 - LR	85,88	85,71	85,71	85,87	78,29	0,31
MobileNetV2 - XGBoost	79,38	78,99	78,99	79,38	68,11	0,45
MobileNetV2	81,66	81,4	81,4	81,65	72,09	0,4

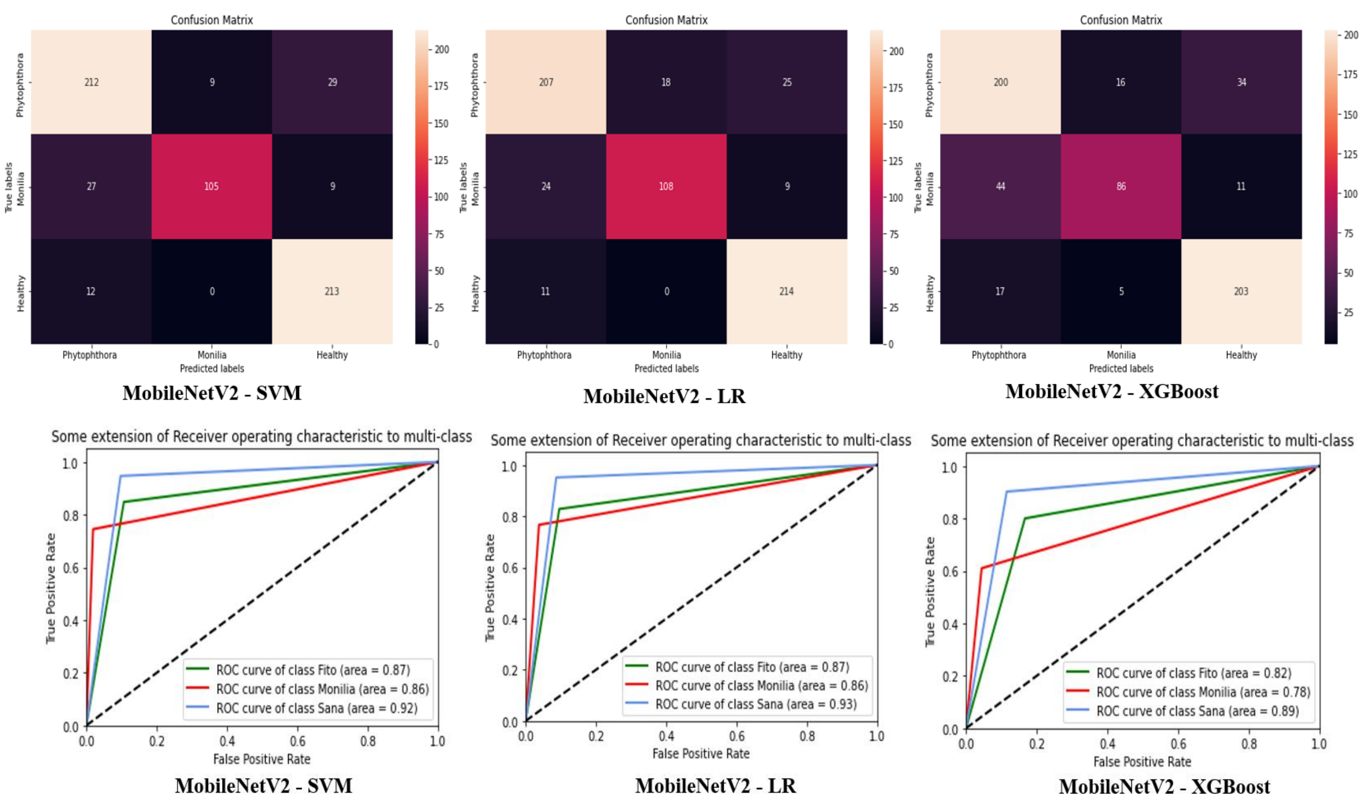

Fig. 3. Confusion matrix and roc curve of hybrid method algorithm.

Analysis of the table above highlights the performance and trends of the various models, including combinations of MobileNetV2 with multiple algorithms and the performance of MobileNetV2 as a stand-alone model. Each combination presents distinct characteristics regarding the accuracy, precision, F1 score, recall, Matthews Correlation Coefficient (MCC), and Mean Square Error (MSE).

The combination of MobileNetV2 with SVM is the most efficient, with an accuracy of 86.04%. This means that it can correctly classify 86.04% of images. Precision, F1 score, and recall are also high, indicating that the model can accurately identify true positives and minimize false positives and false negatives. The combination of MobileNetV2 with Random Forest is also effective, with an accuracy of 77.60%. However, the precision, F1 score, and recall measures are slightly lower compared with the MobileNetV2-SVM combination. This means the model is less accurate at identifying true positives but more likely to minimize false positives.

The combination of MobileNetV2 with KNN is less effective, with an accuracy of 72.4%. Precision, F1 score, and recall measures are all relatively similar, indicating that the model can detect true positives and negatives with comparable precision.

The combination of MobileNetV2 with LR is effective, with an accuracy of 85.88%. The high accuracy associated with this combination indicates an ability to avoid false positives. The high F1 and recall scores show that the model can effectively detect true positives. The combination of MobileNetV2 with XGBoost is comparable to other hybrids, with an accuracy of 79.38%. Precision, F1 score, and recall demonstrate a balanced ability to identify true positives and minimize false positives and negatives. As a stand-alone model, MobileNetV2 is effective, with an accuracy of 81.66%. Other measures, including precision, F1 score, recall, and MCC, also show balanced performance.

V. Discussion

The results reveal key aspects concerning the performance of the different models evaluated for pod disease detection. These results highlight distinct strengths and weaknesses of each model, providing important information to guide the optimal mode selection.

In terms of individual performance, MobileNetV2 had the best overall performance of all the models evaluated. Its high accuracy of 81.66% indicates its ability to deliver accurate predictions. In addition, its high F1 score and balanced recall (81.4% and 81.65%, respectively) indicate its ability to identify true positives while maintaining a balance with false negatives. The high MCC of 72.09% reflects an excellent correlation between predictions and actual observations. Although MobileNetV2 performs well, it may require higher computing resources due to its complex architecture.

SVM and Logistic Regression: These traditional Machine Learning models show stable and balanced performance, with precision, F1 score, and recall rates around 67%. They can be considered reliable options for the detection of pod diseases.

However, their inability to achieve the same levels of accuracy as MobileNetV2 may suggest limitations in their ability to capture subtle image features.

Random Forest and XGBoost: These ensemblistic models show comparable results, with F1 scores and balanced recalls. They show a certain robustness in detecting true positives.

However, their slightly lower accuracy could mean that they tend to identify some false positives.

Although KNN has the lowest accuracy among the models evaluated, its relatively high MCC reveals a specific relevance in detecting pod diseases.

However, its low recall and F1 score underline its difficulty correctly identifying all true positives.

Model performance varies according to method and architecture. While MobileNetV2 stands out for its high accuracy and correlation, traditional Machine Learning models offer a solid and stable option. By considering the strengths and weaknesses of each model, informed decisions can be made to maximize the quality of pod disease detection.

Fig. 4 presents the histogram of the metrics of classical algorithms.

Combining MobileNetV2 with models such as SVM, LR, or XGBoost improves overall performance compared with MobileNetV2 alone. This is due to how these models complement MobileNetV2's capabilities, leveraging the power of MobileNetV2's computer vision and the more traditional classification methods of these other models.

The combination with SVM seems particularly effective in terms of Accuracy and MCC, showing how the SVM approach can enhance MobileNetV2's classification capabilities. MobileNetV2 combined with KNN or Random Forest also shows improvements, but performance is still inferior to that obtained with SVM or XGBoost.

Fig. 5 presents the metric histogram of the hybrid methods.

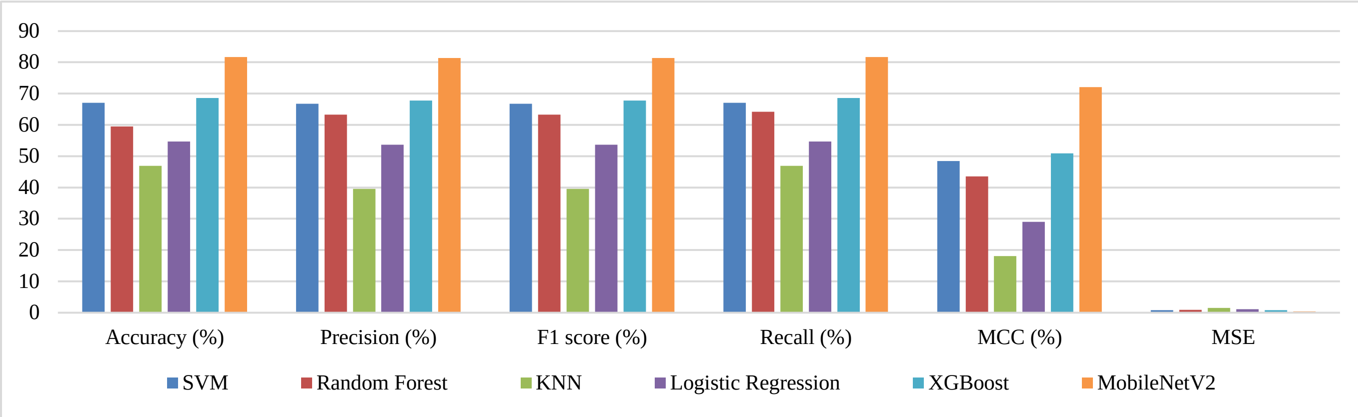

Fig. 4. Model performance histogram of the case of individual algorithm.

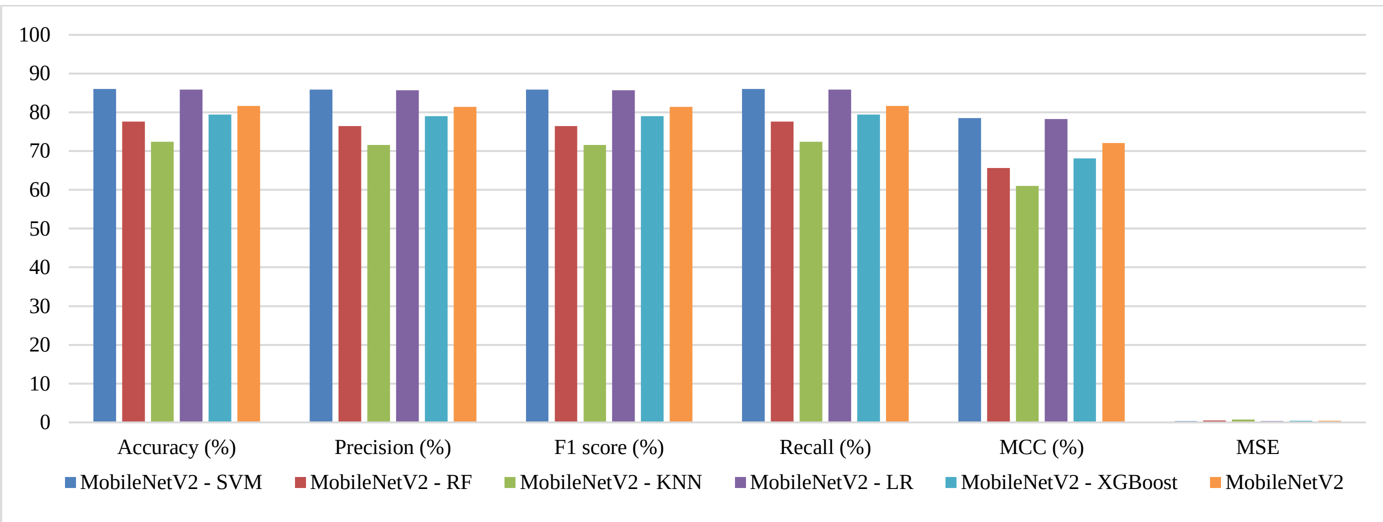

Fig. 5. Model performance histogram of the case of hybrid method algorithm

VI. Comparison with Existing Approaches

The results of our work exceeded those of several related studies, including previous research on the detection and identification of cocoa pests and diseases.

Table III shows the results. The results clearly show that our model outperforms that of the other two authors.

TABLE III. Comparison of Results with Previous Research

Method	Accuracy (%)
Basri B, et al [26]	82.50
RY Montesino et al[27]	83
Our method	86.04

VII. Conclusion

This study represents a significant step forward in the fight against cocoa diseases in Côte d'Ivoire. By combining the intelligence of MobileNetV2's convolutional neural networks with the advanced classification capabilities of algorithms such as Logistic Regression, K-nearest Neighbors, Support Vector Machines, XGBoost, and Random Forest, we have succeeded in significantly improving the accuracy of disease detection.

The results obtained are promising, highlighting the effectiveness of hybrid approaches in tackling complex agricultural problems. This innovative method offers an alternative to traditional methods of laborious visual inspection, enabling farmers to take more targeted measures to protect their crops.

However, it is important to note that further research is needed to refine these hybrid approaches and adapt them to the seasonal and environmental variations to which cocoa is exposed. In addition, additional validation on varied datasets and real field conditions would reinforce the conclusions' robustness. This study lays the foundations for a new era in cocoa disease monitoring and management, with potentially positive implications for farmers, local economies, and food security. As technology continues to evolve, this approach could serve as a model for solving similar problems in other areas of agriculture and beyond.

The prospects of this study could pave the way for the development of new applications of machine learning and deep learning in the cocoa sector. Indeed, the hybrid approach developed by the researchers proved effective in detecting disease in cocoa pods. This approach could be used to develop new applications, such as a mobile diagnostic tool that would enable farmers to detect cocoa pod diseases on the spot or to monitor the spread of cocoa diseases.

REFERENCES

[1] « Renforcer La Compétitivité de La Production de Cacao Et Augmenter Le Revenu Des Producteurs de Cacao en Afrique de L'ouest Et en Afrique Centrale | PDF | Solides de cacao | Afrique », Scribd. https://fr.scribd.com/document/505690969/2017-07-Renforcer-la-competitivite-de-la-production-de-cacao-et-augmenter-le-revenu-des-producteurs-de-cacao-en-Afrique-de-l-Ouest-et-en-Afrique-centr (consulté le 24 août 2023).

[2] « Boughaba_Boukhris.pdf ». Consulté le: 24 août 2023. [En ligne]. Disponible sur: https://dspace.univ-ouargla.dz/jspui/bitstream/123456789/17195/1/Boughaba_Boukhris.pdf

[3] « Réseaux Des Capteurs Sans Fil Intelligent Pour La Détection Des Mauvaises Herbes Dans Les Applications Agricoles ». https://theses-algerie.com/2729392144012869/memoire-de-master/universite-larbi-tebessi---tebessa/r%C3%A9seaux-des-capteurs-sans-fil-intelligent-pour-la-d%C3%A9tection-des-mauvaises-herbes-dans-les-applications-agricoles (consulté le 24 août 2023).

[4] T. Domingues, T. Brandão, et J. C. Ferreira, « Machine Learning for Detection and Prediction of Crop Diseases and Pests: A Comprehensive Survey », Agriculture, vol. 12, no 9, Art. no 9, sept. 2022, doi: 10.3390/agriculture12091350.

[5] K. J. Ayikpa, D. Mamadou, P. Gouton, et K. J. Adou, « Experimental Evaluation of Coffee Leaf Disease Classification and Recognition Based on Machine Learning and Deep Learning Algorithms », Journal of Computer Science, vol. 18, no 12, p. 1201-1212, déc. 2022, doi: 10.3844/jcssp.2022.1201.1212.

[6] R. A. Godmalin, C. J. Aliac, et L. Feliscuzo, « Classification of Cacao Pod if Healthy or Attack by Pest or Black Pod Disease Using Deep Learning Algorithm », in 2022 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET), Kota Kinabalu, Malaysia: IEEE, sept. 2022, p. 1-5. doi: 10.1109/IICAIET55139.2022.9936817.

[7] S. Kumi, D. Kelly, J. Woodstuff, R. K. Lomotey, R. Orji, et R. Deters, « Cocoa Companion: Deep Learning-Based Smartphone Application for Cocoa Disease Detection », Procedia Computer Science, vol. 203, p. 87-94, 2022, doi: 10.1016/j.procs.2022.07.013.

[8] P. Y. O. Amoako, G. Cao, et J. K. Arthur, « An Image-Based Cocoa Diseases Classification Based on an Improved Vgg19 Model », Springer Books, p. 711-722, 2023.

[9] Basri, Indrabayu, A. Achmad, et I. S. Areni, « Comparison of Image Extraction Model for Cocoa Disease Fruits Attack in Support Vector Machine Classification », in 2022 International Conference on Electrical and Information Technology (IEIT), Malang, Indonesia: IEEE, sept. 2022, p. 46-51. doi: 10.1109/IEIT56384.2022.9967910.

[10] B. Baba, R. Tamin, Indrabayu, I. Areni, et H. A. Karim, « MOBILE IMAGE PROCESSING APPLICATION FOR CACAO'S FRUITS PEST AND DISEASE ATTACK USING DEEP LEARNING ALGORITHM », ICIC Express Letters, vol. 14, p. 1025-1032, oct. 2020, doi: 10.24507/icicel.14.10.1025.

[11] R. D. Gunawan, I. Ahmad, Parjito, H. Anggono, E. H. Vernando, et F. Jaya, « Optimization of Certainty Factor Algorithm to Overcome Uncertainty in Expert System Identification of Pests and Diseases of Cocoa », in 2022 2nd International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS), Yogyakarta, Indonesia: IEEE, nov. 2022, p. 310-315. doi: 10.1109/ICE3IS56585.2022.10010123.

[12] Mohammad Yazdi Pusadan, Syahrullah, Merry, et Ahmad Imam Abdullah, « k-Nearest Neighbor and Feature Extraction on Detection of Pest and Diseases of Cocoa », J. RESTI (Rekayasa Sist. Teknol. Inf.), vol. 6, no 3, p. 471-480, juill. 2022, doi: 10.29207/resti.v6i3.4064.

[13] R. A. Godmalin, C. J. Aliac, et L. Feliscuzo, « Cacao Pod Infection Level Classification Using Transfer Learning », in 2023 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream), Vilnius, Lithuania: IEEE, avr. 2023, p. 1-6. doi: 10.1109/eStream59056.2023.10135062.

[14] « Cocoa Diseases (YOLOv4) ». https://www.kaggle.com/datasets/serranosebas/enfermedades-cacao-yolov4 (consulté le 24 août 2023).

[15] A. K. Jean, M. Diarra, B. A. Bakary, G. Pierre, et A. K. Jérôme, « Application based on Hybrid CNN-SVM and PCA-SVM Approaches for Classification of Cocoa Beans », IJACSA, vol. 13, no 9, 2022, doi: 10.14569/IJACSA.2022.0130927.

[16] C. Jabour, « Estimation de la résistance coronarienne par analyse du réseau vasculaire du fond de l'œil », phdthesis, Institut National des Sciences Appliquées de Lyon, 2023. Consulté le: 24 août 2023. [En ligne]. Disponible sur: https://hal.science/tel-04104565

[17] D. Jacob, R. Tievant, L. Cervoni, et M. Roudesli, « Prédiction des blessures au Foot 5 à l'aide d'une méthode de machine learning », Journal de Traumatologie du Sport, juin 2023, doi: 10.1016/j.jts.2023.06.001.

[18] B. Pan, « Application of XGBoost algorithm in hourly PM2.5 concentration prediction », vol. 113, p. 012127, févr. 2018, doi: 10.1088/1755-1315/113/1/012127.

[19] B. M. Kouassi, V. Monsan, A. B. Ballo, K. J. Ayikpa, D. Mamadou, et K. J. Adou, « Application of the Learning Set for the Detection of Jamming Attacks in 5G Mobile Networks », International Journal of Advanced Computer Science and Applications (IJACSA), vol. 14, no 6, Art. no 6, 30 2023, doi: 10.14569/IJACSA.2023.0140676.

[20] R. Benzitouni, L. Merabet, et S. Zertal, « La Proposition d'une Architecture basée Deep Learning pour la prédiction des maladies cardiaques dans un environnement IoT », 2022, Consulté le: 24 août 2023. [En ligne]. Disponible sur: http://localhost:8080/xmlui/handle/123456789/14486

[21] C. Moussaoui et A. (Encadreur) Mellit, « Conception et réalisation d'un système semi-automatique de diagnostic des générateurs photovoltaïques à l'aide de l'apprentissage profond et de l'imagerie thermique. », Thesis, Université de Jijel, 2022. Consulté le: 24 août 2023. [En ligne]. Disponible sur: http://dspace.univ-jijel.dz:8080/xmlui/handle/123456789/12453

[22] « (1) (PDF) Evaluation de l'impact de la territorialisation du Plan Maroc Vert sur le niveau de vie des phoeniciculteurs des zones oasiennes de la région Drâa- Tafilalet à travers une régression logistique binaire ». https://www.researchgate.net/publication/371812137_Evaluation_de_l'impact_de_la_territorialisation_du_Plan_Maroc_Vert_sur_le_niveau_de_vie_des_phoeniciculteurs_des_zones_oasiennes_de_la_region_Draa-_Tafilalet_a_travers_une_regression_logistique_binair (consulté le 24 août 2023).

[23] P. N. Srinivasu, J. G. Sivasai, M. F. Ijaz, A. K. Bhoi, W. Kim, et J. J. Kang, « Classification of skin disease using deep learning neural networks with mobilenet v2 and lstm », Sensors, vol. 21, no 8, avr. 2021, doi: 10.3390/s21082852.

[24] A. Ballo, M. Diarra, A. Kacoutchy Jean, K. Yao, A. Assi, et K. Fernand, « Automatic Identification of Ivorian Plants from Herbarium Specimens using Deep Learning », International Journal of Emerging Technology and Advanced Engineering, vol. 12, p. 56-66, mai 2022, doi: 10.46338/ijetae0522_07.

[25] « Sensors | Free Full-Text | Accelerating 3D Convolutional Neural Network with Channel Bottleneck Module for EEG-Based Emotion Recognition ». https://www.mdpi.com/1424-8220/22/18/6813 (consulté le 24 août 2023).

[26] Basri, B., Indrabayu, I., Achmad, A., & Areni, I. S. (2023). A Review of Image Processing Techniques for Pest and Disease Early Identification Systems on Modern Cocoa Plantation. International Journal of Computing and Digital Systems, 13(1), 1-1.

[27] RY Montesino, JA Rosales-Huamani et JL Castillo-Sequera, « Détection de phytophthora palmivora dans les fruits du cacao avec apprentissage profond », 2021 16e Conférence ibérique sur les systèmes et technologies d'information (CISTI) , Chaves, Portugal, 2021, pp. 1-4 , est ce que je : 10.23919/CISTI52073.2021.9476279.