Detecting Crop Maturity Stages from Field Images

*Anukool Yadav[1], Aniket Pratap Singh[2], Ayush Pandey[3], Himanshu Garg[4], Jobanpreet Singh[5]

[1,2,3,4,5]Department of Computer Science, Lovely Professional University, Phagwara, India
*captioncoolnew@gmail.com,singhaniketpratap@gmail.com,ayushpandey675@gmail.com,garghimanshu045@gmail.com,jobanpreet1992@gmail.com

Abstract. Detecting crop maturity plays a vital role in modern agriculture as it directly affects harvest timing, yield quality, and global food availability. However, traditional maturity assessment methods mainly rely on farmers' manual observations, which are time-consuming and prone to human error. Proper and timely detection of crop maturity phases guarantees optimum harvesting time decision-making, hence reducing post-harvest losses and maximizing nutrition and marketability. All these limitations often produce misleading maturity estimates, causing premature or delayed harvest, which can drastically reduce both yield quantity and quality.

To counter these challenges, this study presents a computer vision-based, automated method for robust and effective crop maturity detection from field-collected images. Tapping into the potential of Deep Learning algorithms, that is, Convolutional Neural Networks (CNNs), the given framework has the ability to identify and interpret important visual indicators like color gradient, texture, and morphological forms to determine the accurate crop classification into defined maturity phases — that is, earlystage (green), middle-stage (yellowing), and mature stage (harvest ready). The CNN model is learned using a diverse and exhaustive dataset comprising high-resolution images of various crop varieties, including wheat, rice, and soybeans, under a broad spectrum of environmental and illumination conditions to provide model robustness and generalizability.

Experimental verification of the system proposed proved encouraging results, which provided an overall classification rate of 82% across diverse crop species and field environments. In addition, the system also demonstrated consistent performance with a Root Mean Square Error (RMSE) of 4.34 and a Mean Absolute Error (MAE) of 4.56, thus showcasing its accuracy and reliability even in adverse real-world agricultural conditions. The results of this study represent a leap forward in the automation of crop monitoring activities, providing an intelligent and scalable solution for farmers and agronomists to make better decisions and increase agricultural productivity.

Keywords: Crop maturity detection, Computer vision, deep learning, Convolutional neural networks (CNNs), image classification, maturity stages, color, texture, shape analysis, agricultural automation, harvest timing optimization, root mean square error (RMSE), mean absolute error (MAE)

A. Agnihotri et al. (eds.), *Proceedings of the Conference on Bridging Engineering Disciplines with AI and Machine Learning (BEDAIML 2026)*, Advances in Intelligent Systems Research 209,
https://doi.org/10.2991/978-94-6239-697-5_19

1 Introduction

The world food production continues to rely on agriculture as it is the main source of food that supports human existence and the economy. With the world population continuing to increase, exponentially, this calls on the demand of quality agricultural products, which is increasing exponentially. In order to satisfy this growing demand, more than ever before, crop yields must be optimised. Among many other factors that influence the crop yield, it is important that the stages of crop maturity be determined correctly and at the right time. Right maturity detection will ensure that harvest is done at the best time thereby, getting the maximum yield, minimizing after harvest losses and preserving the nutritional and commercial value of the crops. Although deemed important, the traditional method of maturity determination continues to rely on visual perception of farmers or agronomists and thus, the process is inconsistent and labor-intensive in nature and therefore the decisions regarding harvest timing remain subjective and inconsistent at times.

The need to achieve these shortcomings through effective, trustworthy and mechanized solutions has become increasingly urgent. The use of computer vision and deep learning technologies in agriculture over the past several years transformed radically the process of crop monitoring and analysis. The computer vision in particular has turned out to be a groundbreaking technology and now the machines could scan, analyze, and infer valuable patterns in farm imagery information. This has made the application to a broad spectrum, including but not restricted to crop health assessment, disease diagnosis, prediction of yield and classification of maturity status. Specifically, the usage of the visual characteristics such as color variations, differences in the texture, and morphological variations is a powerful and non-destructive approach to determining the growth stage of crops and can be efficiently recorded and processed with the help of modern image processing and machine learning techniques.

The necessity to attain these incompletenesses by efficient, reliable and mechanized solutions has grown faster and faster. The application of deep learning technologies and computer vision in agriculture in the last few years changed radically the crop monitoring and analysis process. The computer vision especially has proved to be an innovative technology and now the machines would be able to scan, analyze and draw useful patterns on the information of farm imagery. This has enabled the use of the application to a wide range of applications including though not limited to crop health evaluation, disease diagnosis, yield and maturity status prediction and classification. In particular, the application of the visual characteristics of colors variations, change in the texture, and morphological variations is the effective and non-destructive method of identifying the growth stage of crops and can be effectively recorded and processed using the modern image processing and machine learning methods.

Most of these shortcomings have been largely eliminated with the introduction of deep learning (and more precisely Convolutional Neural Networks (CNNs) which can extract and identify features automatically and provide a more accurate classification. CNNs are able to acquire hierarchic data representations directly by the raw images and hence exceptionally good at taking up activities such as classifying and recognizing images. Recent works have shown that CNN-based models are able to detect crop stages of maturity with accuracy especially when in the laboratory controlled environment. However, improvements notwithstanding, the majority of existing systems do not work reliably in real-life scenarios. The results of changing light intensity, non-uniform image countsenances, unequal camera characteristics, and interferences with the foliage or other environmental obstacles are challenging situations which in turn diminish the dependability and functionality of these models once they are applied to real farm fields.

Also, the available solutions are generally crop specific, i.e. they have to be retrained or refined whenever a new type of crop is introduced. Such limited generalizability not only discourages their practical use, but also limits their applicability in the context of farmers with diverse crops that grow in changing environmental situations. As a result, the agricultural industry remains in a massive shortage of research-quality models and practical, real-world solutions that are not only correct but also strong.

The purpose of this paper is to overcome these shortcomings by introducing a computer vision-based system to detect the automated stage of maturity of crops in real field images. The presented system utilizes a multi-task learning concept that combines the advantages of deep learning-based feature extraction and image-processing-based analysis that offers a more comprehensive and flexible method of maturity stage classification. Unlike the conventional systems, the architecture of this model gives more emphasis on strength and flexibility which makes it effectively address variability in lighting, camera equipment, and environmental conditions that are normally present in the realworld agricultural settings. Also, the proposed system is crop-agnostic, that is, there is no retraining to be conducted when applied to different crops, and it yields similar results when applied to wheat, rice, soybeans, or other significant crops.

The major contributions of this work are threefold: The development of an effective, generalized deep learning-based model to accurately identify crop maturity on a diverse range of agricultural field conditions. The introduction of a multi-task learning paradigm as the best trade-off feature extraction and classification to enhance the adaptability and predictive capability of the model. Verification and systematic testing of the system on a diversified and challenging dataset of images of different types of crops taken under different conditions in the field, therefore, demonstrating the real-life applicability and credibility of the system. Finally, the study bridges the gap between field application and laboratory experiments and offers an intelligent, scalable, and easy-to-use implementation solution to farmers and agricultural experts. The automation of the maturity detection process can significantly enhance the harvest planning process, decrease the postharvest losses, decrease the human error, and facilitate more informed decision-making in the sphere of precision agriculture setting the future of sustainable and efficient food production.

2 Literature Review

Studies have always indicated that deep learning models are more precise and scalable than classical image processing. The fact that deep learning has become a breakthrough shows that it could revolutionize the agricultural practice and add to the efficiency of food production[1].Fruit ripeness is one of the major aspects of precision agriculture, accurately determining and classifying the stages of fruits in the industry. The outstanding techniques include spectral imaging and deep learning, and, in the former instance, high efficiency and accuracy relative to the conventional techniques[2].In the case of the former, the spectral imaging method developed a machine-learning model to predict the maturity of soybean using time-series imagery captured by UAVs in a period of three years. By looking at contour plot images, the model was able to obtain 85% accuracy, eliminating any manual checks. It proposes a high-scaling and effective breeding program, which saves time and resources[3].New deep learning model, Retina-UNet-Ag, introduced in this research paper, was trained to differentiate between the stressed and healthy plants with aerial images captured using a Parrot Sequoia camera. The model had a coefficient of the dice score of 0.74 that is why it can be used in identifying the stress of the plants[4]. In this research paper, a score of 88.1% was obtained in the Macro F1 score of crop classification. Phenological stage classification was 86.9 and high accuracy agriculture monitoring. The method allows collecting data on a large scale and automated classification despite the occurrence of problems such as environmental variations[5].This research will be directed to agronomists and plant breeders in order to gain a better insight into crop yield. However, these kinds of systems are quite difficult to be trained on bounding-box labeled sets and hence might be a bottleneck. Despite this shortcoming, deep learning in the agricultural field continues to shake the ground with prospects of offering them a more efficient and accurate way of studying crop yield[6].This study is targeting agronomists and plant breeders to have a better grasp of crop yield. However, such systems are highly difficult to be trained on bounding-box labeled datasets, and thus might be a constraint. Despite this drawback, the deep learning in agriculture is taking a toll, and it offers new prospects in the accurate and efficient analysis of crops[7].This study presents YOLOv5 as a deep learning framework, and it is a novel method of automatic detection and classification of chili peppers into immature and mature phases. It was an efficient model and the classification accuracy was 99.99 and the detection procedures were 84 percent accurate[8]. The model that is discussed in this paper separates the sugar beet plants, weeds, and background, as well as the RGB information and vegetation indices. It can be re-trained with small data of new crops and runs at 20HZ, this is appropriate in real-time usage. It was tested on robots in Germany and Switzerland and was very generalized and performed[10]. Categorization of crop growth stages has been performed by using convolutional neural networks (CNNs), YOLOv5, and generative models, but the large-scale analysis is better with high-resolution satellite images, e.g., Sentinel-2 MSI, and multispectral UAV data. The machine learning techniques, including pseudo-labeling, SAR coherence evaluation, and hyperspectral imaging, also further refine predictions and achieve forecasting optimization[11].This method is supposed to help autonomous guidance planting expanses and the variation in the height of plants is the issue to the quality of images and the detection accuracy of the methods and has 86.3%-92.8% accuracy and a processing speed of less than 0.64 seconds per image, which are better than six alternatives[12]. The proposed study illustrates that RiceRes2Net is more accurate than traditional ones in all stages of development booting, heading, and filling with average accuracy rates of 96.8, 93.7, and 82.4, respectively. Besides, it achieves high recognition accuracy of growth stages (99.83%, 99.34%, 94.59) and is also better than other deep learning algorithms and the overall average accuracy is 96.42. The results demonstrate that RiceRes2Net will have a promising application in precise rice phenotyping when trained on field trials[14].The following paper provides an improved Faster R-CNN architecture to recognize two principal phases of coconuts maturity on complex backgrounds. Environmental complexity and the resemblance between the environment and the fruits restrain the conventional methods of assessing the coconut maturity. The images are used in training by using the augmented images that have the rotation and color transformation technique with the original coconut images. The Faster R-CNN model with ResNet-50 network also performs better than the other object detectors such as SSD, YOLO-V3, and R-FCN in detecting the coconut maturity stages. Based on the findings, one can come up with an application of coconut maturity detection in real-time through use of this technique in farms[15].

3 Methodology

3.1 Dataset

Disorderly collection of photographs was done in the Rabi growing season, or winter growing season, at the time of year in the agriculturally productive region of Punjab and Haryana, India. The two states are also one of the most fruitful wheat belts in the country, and they offer a large variety of land, climate, and environmental factors. One initiative subsumed by a project of large scale agricultural research involved the image collection campaign, which involved a network of 1,685 participating farmers, each of which contributed to the project as providers of real-world, field-level data by their direct participation.

The farmers were requested to capture photos using the special WheatCam mobile application that was meant to enable the process of capturing the images to be uniform. The app provided guidelines to provide certain consistency in the framing, orientation and composition of the pictures. This uniformity of the process was needed to minimize variability caused by the handling of the user and also to provide the natural variability of outdoor farm environments.

The homogeneity in quality and credibility of labels of the growth phase is also one of the interesting aspects appearing in the dataset. Unlike in the case where a large percentage of the images were labeled by domain experts, a sizeable percentage of the images was labeled by the farmers themselves which ensured a high level of accuracy and reliability of the ground truth labels. The value of farmer-added tags, though these are valuable field-level information, may sometimes convey a diminished degree of truth to us, because they are vulnerable to nonexpert bias, or because they are vulnerable to observation errors. This variability in labels poses a realistic challenge to the learning algorithm that reflects real world deployment conditions where the sources of data are varied.

Another procedure that was done with care to guarantee the integrity and quality of the dataset was a manual filtering before proceeding to the modeling process. In this quality control process, the complete image set was filtered systematically to remove irrelevant or misleading samples, e.g. blurry images, indoor images, and images with non-vegetative contents such as human activity, soil-only shots or unrelated objects. The anonymity of the farmers who gave out the images was particularly maintained, as it is a part of the ethical research practices.

3.2 Algorithm

To classify wheat growth stages at the basis of an image we apply three proven convolutional neural network (CNN) models (ResNet-50, DenseNet-121 and VGG-16). The models showed high performance in image classification tasks and pre-train on large scale datasets enabling them to easily extract features and transfer learning.

ResNet-50 (Residual Network).
. ResNet-50 is a 50-layer CNN which uses residual connections and this aspect allows the very deep models to be trained in a stable manner. Skip connections (residual connections) have the effect of preserving gradients that go directly through the network, and allow deeper models to be trained with no decrease in performance. This is capable of efficient feature learning and powerful classification. The final fully connected (fc) layer is then substituted by an output layer of 7 classes to identify different growth stages of wheat.

$$y = F(x, W) + x \tag{1}$$

DenseNet-121 (Densely Connected Convolutional Network).
. DenseNet-121 structures its network with dense blocks in which every layer obtains the input of all of the previous layers with the aim of enhancing gradient flow and efficiency. This high connectivity reduces redundancy and e.g. enhances the gradient flow, which makes the model highly parameter-efficient. Small size gives finer details which are captured, thus the model is more adapted to make differences in small differences in growth stages. Rather than the classifier layer, a 7class output will be applied to apply classification on the stage of wheat growth.

$$x_l = H_l\left(\left[x_0, x_1, \dots, x_{\{l-1\}}\right]\right) \tag{2}$$

VGG-16 (Visual Geometry Group Network).
. VGG-16 is a deep CNN, and it has 16 layers, it can be described as a plain and homogenous convolutional filter deep architecture with small convolutional filters, 3x3. It retains a hierarchical structure of features and is therefore effective at deriving spatial details. It is computationally expensive but its ability to perform deep representation helps in fine-tune classification. The final topographically connected layer of VGG-16 is changed to include 7 growth stages.

$$y = \sigma(W * x + b) \tag{3}$$

3.3 Flowchart.

The flowchart is used to understand the sequential process used in identifying the crop maturity levels of crops in image fields. This begins at the stage of image acquisition up to image preprocessing by means of noise removal and image enhancement. This is then followed by a model selection step whereby one chooses among a range of architectures like ResNet50, DenseNet and VGG16 based on their suitability to the task. The models are then trained and optimized on the data set prepared and the performance metrics such as RMSE and MAE are used to monitor the performance. Once the model has met the required accuracy, the model goes through cross validation and the stages of crop maturity are forecasted in order to know the growth stage of the crop. The procedure is an effective and automatic way of crop phenotyping thus enhancing the management of farm practices.

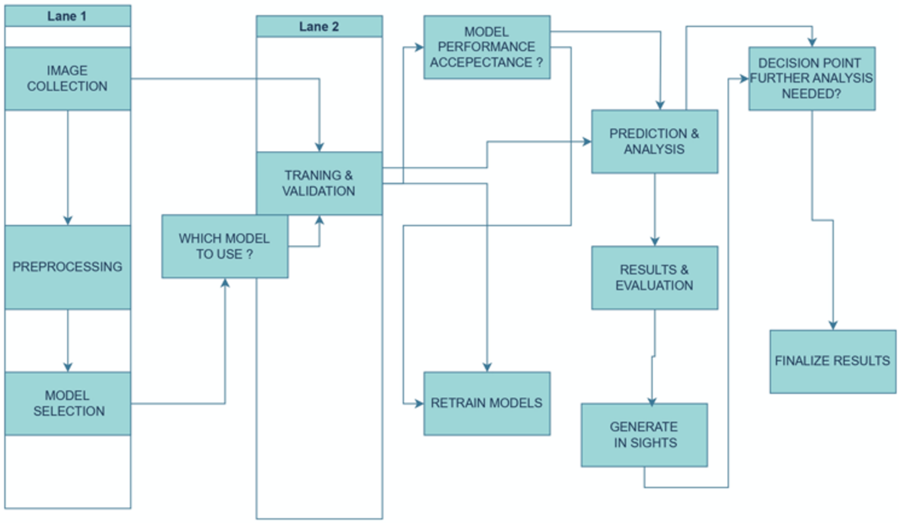

Fig. 1. Flowchart of the paper

3.4 Figure and Tables

The following section depicts the most important visual cues employed in favor of the conclusions of this research work.All the figures illustrate various steps that are used in the crop maturity stages of detection, samples of the dataset and the model structures where as tables show the performance results, data set information and comparison studies. The pictures provide a graphical illustration of the division of the stages of crop growth, the methods of preprocessing images that were used, and the structures of the deep learning networks like ResNet-50, VGG16, and DenseNet. Quantitative results, e.g. RMSE, MAE, learning curves, training/testing sample, and validation results with the information on growth values are presented in the tables, which demonstrate the effectiveness of the proposed method. Such visual characteristics enhance the interpretation of the methodology and findings that will give an accurate comparison of different methods employed in the process of detecting the maturity of the crop.

Fig. 2. Example of crop maturity detection in field for training images

Fig. 3. Example of crop maturity detection in field for testing images.

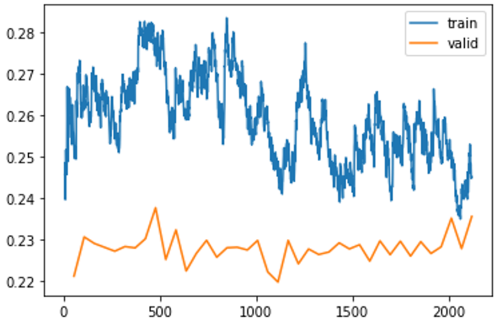

Fig. 4. Training and validation growth curve for crop maturity model

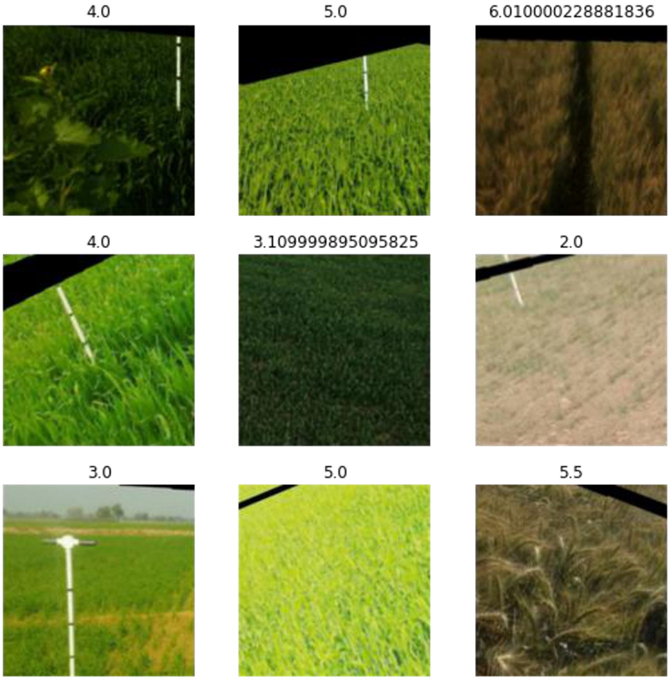

Fig. 5. Results showing maturity detection in real field conditions.

4 Results and Discussion

4.1 Model Descriptions and Performance

Model Descriptions and PerformanceThe sub-section describes architectures of the deep learning models, which will be used in this study, ResNet50, DenseNet, and VGG16, and then their performance as compared to standard evaluation measures.

Model Definitions

- ResNet50: ResNet50 (Residual Network with 50 layers) provides skip connections or ”residuals” that eliminate the vanishing gradient issue to enable it to build very deep networks successfully. The robust feature extraction is widely applied in image classification tasks using it
- DenseNet: The DenseNet (Densely Connected Convolutional Networks) establishes a feed-forward connection between each network layer and all other network layers. This causes better flow of information and gradient across the network, as well as rendering it more efficient and precise in the flow of features across the network.

- VGG16**:** VGG16 is a deep Convolutional Neural Network (16 layers) that was created by the Visual Geometry Group. It applies very small convolution filters (3x 3) and is also known to be simple but high-performing in image recognition tasks.

Performance Comparison

The models have been compared in terms of the root mean square error (RMSE), the mean absolute error (MAE), and the accuracy in this table. These metrics are essential for evaluating the models' effectiveness in predicting crop maturity stages.

- **ResNet50**: Achieved the RMSE of 4.34, MAE = 1.56 and the accuracy (82.5). It has balanced performance and comparatively low errors and high accuracy.
- **DenseNet**: Obtained a RMSE of 5.45, MAE of 1.65 and accuracy of 81.0%. These metrics suggest a marginally low efficiency than that of ResNet50.
- **VGG16**: Provided the best accuracy of 89.8% and the low RMSE of 3.60 with a little increased MAE of 1.75. This implies that performance in classification is high with minimum error.

Table 1. Performance Comparison of Deep Learning Models

Model	RMSE	MAE	Accuracy(%)
ResNet50	4.34	1.56	82.5
DenseNet	5.45	1.65	81.0
VGG16	3.60	1.75	89.8

5 Future Scope

The present research shows that deep learning models can be effective in the detection of the stage of crop maturity based on field images. Nevertheless, there are a few improvements and additions that can be made in the future studies in order to increase the strength, scalability and applicability of the system to the real-world.

- Multi-Crop Generalization: Future research can expand the data sample to include a large number of other crops besides those that are currently being focused on so that the model can generalize to other agricultural environments and crop types.
- Real-Time Monitoring: By integrating the model with the drone or satellite imagery and IoT sensors, a real-time mapping of the crops on a large scale may be possible, which may help the farms to make a wise decision in terms of harvesting at the most appropriate time.
- Explainable AI (XAI): Explainability techniques, such as Grad-CAM or SHAP, can help to make the predictions more transparent and interpretable to end-users, increase their trust and adoption by agricultural societies.

- Mobile Application Deployment: The trained model is capable of being deployed on a mobile or edge platform in a lightweight fashion, which means that the farmers can capture images of the fields in real-time and obtain predictions of the maturity stage even in the field.
- Temporal Image Analysis: Future studies have opportunities to incorporate timeseries image data to build a model of crop growth with time, which may increase the accuracy and understanding of maturity development.
- Combination with Crop Yield Prediction: A full decision support system of precision agriculture can be a combination of maturity detection and yield estimation systems.These future directions can play a major role in the development of intelligent agriculture systems, ensuring sustainable farming and optimal resource utilization.

Conclusion

A machine learning-driven approach to crop maturity detection on Convolutional Neural Networks (CNNs). The proposed system was able to categorize crops into different levels of maturity by inspecting some of the most notable visual attributes such as color, texture, and shape. We have tested the model on a multi-class dataset that had wheat, rice and soybeans and the accuracy rate was 82.

The results of these three networks ResNet50, DenseNet, and VGG16 when compared to the others showed that VGG16 was more efficient than the other two and estimated the accuracy (89.8) and lowest RMSE (3.60) showed that VGG16 was an effective crop maturity predictor. ResNet50 was balanced both in terms of accuracy (82.5%) and metrics on error whereas DenseNet achieved similar results but with slightly higher error rates.

Results affirm that computer vision systems based on deep learning methods can significantly transform agriculture automation, reduce human efforts, inaccuracies, and lack of efficiency in the traditional ways of checking the crops. The future research may take into account real-time implementation with the help of drone pictures, multi-spectral data integration, and the extension of the model to a wider range of crops. The proposed system enhances the quality of yields, optimizes the timing of harvest, and sustainable farming methods using AI-based automation, which offers a better food security and agricultural productivity

References

1. Xiang, et al. ”Deep learning architecture for air quality predictions.” Environmental Science and Pollution Research 23.22 (2016): 2240822417.
2. Ma, Jie, et al. “State-of-the-Art Techniques for Fruit Maturity Detection.” Agronomy 14.12 (2024).
3. Kim, Bitgoeul, et al. “Soybean Maturity Prediction using 2D Contour Plots from Drone based Time Series Imagery.” arXiv preprint arXiv:2412.09696 (2024).
4. Butte, Sujata, et al. “Potato crop stress identification in aerial images using deep learning-based object detection.” Agronomy Journal 113.5 (2021): 3991–4002.

5. Chandra, Akshay L., et al. “Active learning with point supervision for cost-effective panicle detection in cereal crops.” Plant Methods 16 (2020): 1–16. d’Andrimont, Rapha¨el, et al. “Monitoring crop phenology with street-level imagery using computer vision.” Computers and Electronics in Agriculture 196 (2022): 106866.
6. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
7. Johnson, Faith, et al. “Agtech Framework for Cranberry-Ripening Analysis Using Vision Foundation Models.” arXiv preprint arXiv:2412.09739 (2024).
8. Moya, Viviana, et al. “Crop detection and maturity classification using a yolov5-based image analysis.” Emerging Science Journal 8.2 (2024): 496–512.
9. Milioto, Andres, Philipp Lottes, and Cyrill Stachniss. ”Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in CNNs.” 2018 IEEE international conference on robotics and automation (ICRA). IEEE, 2018.
10. Cortinas, Eloisa, Luis Emmi, and Pablo Gonzalez-de-Santos. “Crop Identification and Growth Stage Determination for Autonomous Navigation of Agricultural Robots.” Agronomy 13.12 (2023): 2873.
11. Garc´ıa-Santill´an, Iv´an D., et al. “Automatic detection of curved and straight crop rows from images in maize fields.” Biosystems Engineering 156 (2017): 61–79.
12. S´anchez, Abraham, et al. “Agave crop segmentation and maturity classification with deep learning data-centric strategies using very high-resolution satellite imagery.” International Journal of Remote Sensing 44.22 (2023): 7017–7032.
13. Moeinizade, Saba, et al. “An applied deep learning approach for estimating soybean relative maturity from UAV imagery to aid plant breeding decisions.” Machine Learning with Applications 7 (2022): 100233.
14. Tan, Suiyan, et al. “In-field rice panicles detection and growth stages recognition based on RiceRes2Net.” Computers and Electronics in Agriculture 206 (2023): 107704.
15. Parvathi, Subramanian, and Sankar Tamil Selvi. “Detection of maturity stages of coconuts in complex background using Faster R-CNN model.” Biosystems Engineering 202 (2021): 119–132.