Agric Res J **61** (4): 595-602, *August* 2024
DOI: 10.5958/2395-146X.2024.00073.5

CASSAVA LEAF DISEASES DETECTION AND CLASSIFICATION WITH ADVANCED MACHINE LEARNING TECHNIQUES

Vibha Aggarwal[1], Sandeep Gupta[2], Manjeet Singh Patterh[3] and Lovepreet Singh[1]

[1]*University College, Barnala-148 101, India*
[2]*College of Engineering and Management, Punjabi University Neighbourhood Campus, Rampura Phul-151 103, India*
[3]*Department of Electronics and Communication Engineering, Punjabi University, Patiala-147 001, India*

ABSTRACT

***Cassava* leaves offer therapeutic properties and is a great source of food and nourishment. For populations that depend on the crop, however, cassava leaf diseases like Cassava Blight Disease (CBD) and Cassava Mosaic Disease (CMD) can have significant economic and social repercussions. This research tries to identify these diseases in order to prevent losses accurately. The effectiveness of pre-processing techniques like backdrop removal is illustrated as the model is tested with and without background photos. The greatest accuracy for healthy and CMD leaves on the target picture stored in the database with backdrop is 97.32%, while without background, it rises to 99.11% for CMD leaves. For the target image outside the database with background, the maximum accuracy for the CBD leaf is 96.46% and increases to 99.11% for the CBD leaf. Overall, the development of this model could be very effective for the diagnosis and management of cassava diseases, which can have devastating effects on crop yields and food security in affected regions.**

Keywords: Cassava leaf diseases, Image processing, Machine learning, Cassava blight disease (CBD), Cassava mosaic disease (CMD)

The growing global population, coupled with both international and national food security policies, has compelled farmers and agricultural scientists to intensify crop production, often at the expense of environmental sustainability. Despite these efforts, numerous challenges, such as climate change, declining pollinator populations, and the emergence of plant diseases, continue to jeopardize food security. Plant diseases, in particular, can have a profound impact on both the yield and quality of crops, posing significant barriers to agricultural development. Over the years, various strategies have been introduced to mitigate crop losses caused by plant diseases. While pesticide use was once the dominant approach, integrated pest management (IPM) methods have gained prominence over the past few decades as a more sustainable alternative. Regardless of the approach chosen, early disease detection remains a critical component for effective disease management. Timely identification of plant diseases facilitates prompt intervention, reducing the risk of widespread crop loss and ensuring that resources are used efficiently. However, traditional methods of disease detection, which often rely on the expertise of farmers and agricultural technicians, can be challenging and prone to inaccuracies. In contrast, deep learning algorithms offer a promising solution by enabling rapid and accurate identification and classification of plant diseases through image analysis. These algorithms can process large volumes of data and detect subtle patterns in plant images, offering a more reliable and scalable method for early disease diagnosis.

Cassava is a widely cultivated crop and a staple food in regions such as Kerala and Andhra Pradesh in India. In Assam, it serves as a significant source of carbohydrates, particularly for communities in hilly areas. Among agricultural crops, cassava ranks third in terms of carbohydrate production per unit area, following sugarcane and sugar beet. As a perennial crop, cassava is well-suited to both irrigated and rain-fed conditions, and is primarily grown for its tubers. It is highly resilient to drought and can thrive in less fertile soils, making it an important crop for areas with challenging growing conditions. In addition to its role as a staple food, cassava is processed into a variety of industrial products, including starch, ethanol, glucose, and animal feed. Despite its economic and nutritional significance, cassava production in India is constrained by several diseases, including brown leaf spot, root rot, and Indian Cassava Mosaic Disease (CMD). Viral diseases such as CMD also affect cassava production in many countries worldwide, further threatening its potential as a reliable food and economic resource. Accurate identification of these diseases and informed management decisions are crucial for enhancing cassava yield. However, the challenge lies in the fact that many cassava diseases share similar symptoms and can co-occur, complicating diagnosis and treatment.

*Corresponding author: vibha_ec@pbi.ac.in
Date of receipt: 04.08.2023, Date of acceptance: 21.07.2024

Deep learning techniques offer a promising solution to this problem by providing precise and automated disease identification. By improving the accuracy and speed of disease detection, deep learning approaches can support better management practices, ultimately boosting cassava production and ensuring greater food security.

Cassava is highly susceptible to a range of diseases (Legg *et al.*, 2015), with the most significant being Cassava Bacterial Blight (CBB), Cassava Brown Streak Disease (CBSD), Cassava Mosaic Disease (CMD), and Cassava Green Mottle Disease (CGM). These diseases can result in substantial yield losses, posing a serious threat to smallholder farmers who depend on cassava as a primary crop. CBB, caused by the bacterium *Xanthomonas axonopodis* pv. *manihotis* (Xam), can severely affect the leaves, stems, and roots of the cassava plant, leading to significant damage. CBSD, caused by two members of the *Potyviridae* virus family, primarily affects cassava storage roots, resulting in substantial economic losses. CMD, caused by begomoviruses, leads to characteristic yellowing and curling of leaves, weakening the plant and reducing its productivity. Finally, CGM, caused by the Cassava Green Mottle Virus (CGMV), is known to cause considerable yield losses, particularly in regions where the disease is endemic. This study focuses primarily on the analysis of CBB, CMD, and healthy cassava leaves, aiming to enhance disease detection methods. In recent years, various studies have explored the use of machine learning algorithms to identify plant diseases, with a particular emphasis on leaf analysis. These machine learning techniques offer promising solutions for automating disease detection, enabling early diagnosis and facilitating more effective management strategies.

The following section reviews various studies that have applied machine learning approaches to detect and classify plant diseases, emphasizing their potential to transform disease management in cassava and other crops. Ai *et al.* (2020) developed a model to identify plant diseases using a large image dataset containing 47,363 images of 27 disease-related plant species and 10 crop varieties. The researchers trained the Inception-ResNet-V2 model on this dataset and evaluated its performance using metrics such as accuracy, precision, recall, and F1-score. The results showed high accuracy, demonstrating the model's effectiveness in recognizing plant diseases from images. In a similar line, Plested *et al.* (2021) advanced the use of Inception-based algorithms for fine-grained image classification tasks by integrating them with a hyperparameter optimization method. Leveraging transfer learning, their model excelled in classifying complex plant disease images, further highlighting the promise of deep learning for plant disease detection. Focusing on cassava specifically, Ramcharan *et al.* (2019) explored the potential of deep learning and mobile technologies for crop disease diagnosis. They introduced a lightweight Convolutional Neural Network (CNN) designed for efficient and portable disease detection, particularly in cassava crops. Their model demonstrated high accuracy, highlighting its potential for use not only in cassava but also in other crops and regions, offering a practical solution for widespread disease management. In another study, Oyewola *et al.* (2021) proposed the use of Deep Residual Convolutional Neural Networks (DRCNNs) with specialized computational blocks for detecting Cassava Mosaic Disease (CMD). Their model achieved high accuracy while maintaining computational efficiency, making it suitable for real-world applications in resource-limited environments where computational resources may be constrained. Emuoyibofarhe *et al.* (2019) demonstrated that machine learning models could effectively classify cassava diseases by distinguishing between healthy and diseased leaves. Their models showed remarkable accuracy in identifying specific disease types, offering valuable potential for early disease detection and timely interventions, which are critical for improving cassava yield and productivity. Sambasivam and Opiyo (2021) addressed the challenge of unbalanced datasets in cassava disease detection by applying the Synthetic Minority Over-sampling Technique (SMOTE). This approach helped balance the dataset, resulting in improved performance of the CNN model. The enhanced model demonstrated strong accuracy in detecting cassava diseases, a crucial factor for ensuring more effective disease control and minimizing crop losses, thereby boosting cassava productivity. Lastly, Ramcharan *et al.* (2017) developed a deep learning model utilizing transfer learning to classify three distinct types of cassava leaf diseases and insect damage. Trained on over 21,000 images, the model achieved an impressive 92.3% accuracy on a test dataset. Optimized for mobile devices, the model operates efficiently on low-end smartphones with an average processing time of just 2.16 seconds per image. This mobile optimization makes the model highly accessible, enabling real-time disease diagnosis and management in resource-limited settings, further contributing to the sustainable cultivation of cassava.

The review of literature discussed above indicates that many studies on cassava disease detection have relied on real-world photographs of cassava leaves, which often exhibit distortions, variations in texture, colour distributions, and uneven gradients, rather than high-quality, standardized images. These imperfections in the images can lead to misclassification of diseases, thus reducing the overall accuracy of the models. However, several studies suggest that this challenge can be mitigated by employing various methodologies such as data augmentation, transfer learning, and ensemble learning. For example, data augmentation techniques, such as flipping, rotating, and zooming, can effectively increase the size and variability of the training dataset, helping to prevent overfitting and improving the model's ability to generalize to new, unseen data. Transfer learning, which leverages pre-trained models on large, diverse datasets, can enhance the accuracy of cassava disease classification by enabling the model to transfer knowledge from broader image recognition tasks and adapt it to the specific characteristics of cassava diseases. Additionally, ensemble learning techniques, which combine the predictions of multiple models, can improve accuracy by reducing errors and providing a more robust decision-making process.

The dataset used in this study contains significant background noise, which can affect the accuracy of disease detection models. To evaluate the model's performance under different conditions, accuracy is assessed using a Support Vector Machine (SVM) both with and without background images of cassava leaves. For feature extraction, 13 distinct characteristics are derived from the cassava images. To select the most relevant features, the Whale Optimization Algorithm (WOA) is employed. WOA is particularly suitable for this task as it can be applied even when the detailed specifications of the problem are not fully known. Additionally, WOA strikes a balance between exploration and exploitation, making it effective for finding the optimal subset of features that contribute to improved model performance (Mirjalili and Lewis, 2016).

METHODOLOGY

This paper aims to develop a system capable of accurately predicting leaf diseases using image processing and machine learning techniques. The system will leverage advanced algorithms to detect and analyze disease patterns, while machine learning will be employed to train the system for precise disease classification. For this study, the dataset provided by Mendeley (Divyanth *et al.*, 2021) includes 228 images of cassava leaves: 91 images of healthy plants, 49 images of Cassava Bacterial Blight (CBB), and 88 images of Cassava Mosaic Disease (CMD). Although the dataset is relatively small, Support Vector Machine (SVM) is chosen for this task due to its effectiveness in handling smaller datasets, offering relatively low computational requirements while still delivering robust performance. SVM is particularly well-suited for classification problems with limited data and is capable of handling a wide range of specifications, making it an ideal choice for the current study.

The following processes are used by the system to detect diseases in cassava leaves:

The primary components of the system design are:

1) Image collection
2) Image Preprocessing
3) Feature extraction
4) Training
5) Classification using Multiclass Support Vector Machine
6) Feature Selection

Image collection

A collection of specimen images of cassava leaves is gathered for the purpose of categorizing cassava diseases. These datasets are crucial for improving the diversity and robustness of the disease classification system, as they typically include photographs taken under various lighting conditions, with different cameras, and across diverse environmental settings. The collection often features images of healthy cassava leaves as well as those exhibiting symptoms of various diseases, such as Cassava Bacterial Blight (CBB), Cassava Mosaic Disease (CMD), and other prevalent cassava diseases. To ensure acceptable accuracy and model robustness, the dataset is designed to be both balanced and diversified, with an adequate number of samples representing each disease class. This balance helps prevent bias in the classification process and enhances the model's ability to generalize across different disease types. Additionally, to maintain consistency and facilitate easier processing, the images are typically stored in standardized formats such as JPEG or PNG.

Image Pre-processing

Pre-processing techniques are essential for enhancing the quality and analytical value of the cassava leaf images in the dataset. Carefully selected and effectively implemented pre-processing steps improve the robustness and accuracy of the system for categorizing cassava diseases. Common pre-processing techniques include image normalization, colour correction, noise reduction, and image enhancement, all of which help in refining the input data for better analysis. One widely used pre-processing method is converting images to grayscale or black and white. This transformation emphasizes key features, such as borders and edges, which are critical for identifying disease markers in cassava leaves. Grayscale images also simplify the identification of statistical characteristics such as texture and contrast, which are important for distinguishing between healthy and diseased leaves. Alternatively, colour-based pre-processing techniques can be employed to highlight specific disease-related features, such as yellowing, necrosis, or other discolorations. Using formats like RGB (Red, Green, Blue) and other colour-based schemes allows the model to focus on colour variations that are indicative of specific disease symptoms. For example, subtle changes in leaf colour, such as yellowing associated with Cassava Mosaic Disease (CMD), can be more easily detected through these colour-focused methods.

Feature extraction

The process of extracting valuable information or characteristics from raw data, such as images, signals, or text, is known as feature extraction. In addition to the gray-level co-occurrence matrix (GLCM), other feature extraction methods include texture-based, shape-based, and colour-based features. Texture-based features can be extracted using techniques such as the gray-level co-occurrence matrix (GLCM), gray-level run-length matrix (GLRLM), and local binary patterns (LBP). These methods capture the surface properties of an image, such as roughness or smoothness. Shape-based features can be obtained using techniques like Fourier descriptors and moment invariants. These features describe the form and geometric properties of an object, such as its perimeter or compactness. Colour-based features can be extracted using techniques like colour histograms and colour moments, which represent the distribution of colours within an image. The 13 features considered in this work can be grouped as follows:

1. Statistical measures: Mean, Standard Deviation, Variance, Kurtosis, Skewness, RMS
2. Texture measures: Contrast, Correlation, Energy, Homogeneity, Smoothness, IDM
3. Information measures: Entropy

Statistical measures describe the distribution of pixel values within an image, providing insights into the overall intensity patterns. Texture measures, on the other hand, capture the spatial arrangement of pixel intensities and offer information about the image's surface characteristics, such as roughness or smoothness. Information measures assess the amount of information or randomness present in an image, reflecting its complexity and variability. By combining these different types of features, it is possible to extract meaningful and comprehensive information from images.

Training

To predict the leaf disease training process is as follows (Rajput *et al*., 2020):

1. Start with a set of images with known classes.
2. Extract features from each image and assign the labels.
3. Use a new image as input and extract its features.
4. Implement binary SVM to multi-class SVM process.
5. Train SVM with the chosen kernel function, and obtain the SVM structure along with support vector and bias values.
6. Determine the class of the input image.
7. Assign the label of the species to the next image based on the outcome, and add its feature set to the database.
8. Repeat steps 3-7 for all images in the database.
9. The testing procedure follows steps 3-6 of the training process, and the class of the input image is the predicted outcome.
10. To evaluate the accuracy of the SVM, a random set of input is selected from the database for training and testing purposes.

Classification

A binary classifier, such as Support Vector Machine (SVM), aims to identify the hyperplane that maximizes the margin between two classes of data. When the data is not linearly separable, SVM can employ various kernel functions, such as polynomial or radial basis function (RBF) kernels, to map the input data into a higher-dimensional space, where a hyperplane can effectively separate the classes. The data points closest to the optimal hyperplane, which maximizes the margin between the classes, are known as support vectors. SVM can be used for both binary and multiclass classification problems, with the final classification being determined by selecting the class with the highest output function.

The algorithm below is used to find diseases in cassava leaves:

Step 1: Obtain an original photo of the cassava leaf with or without noisy background.

Step 2: Process the image by resizing it, enhancing contrast, and converting it to the HSV colour space.

Step 3: Create a binary mask using colour thresholding to segment the leaf from the background.

Step 4: Crop the RGB image under the binary mask to isolate the leaf.

Step 5: Extract features from the cropped image, including Contrast, Correlation, Energy, Homogeneity, Mean, Standard Deviation, Entropy, RMS, Variance, Smoothness, Kurtosis, Skewness, and IDM.

Step 6: Classify the leaf image using SVM.

Step 7: Show the classified result, indicating whether a disease is detected in the cassava leaf or not.

Step 8: Check the accuracy of the classification results.

Step 9: Feature selection using WOA from extracted features.

Step 10: End the process.

The processes outlined above, which includes image pre-processing, feature extraction, and machine learning classification, can be utilized to identify diseases in cassava leaves.

RESULTS AND DISCUSSION

The original image data is pre-processed before training the model to eliminate any noise or irrelevant information that could impact the accuracy of the results. In this case, the proposed method is used to identify the type of disease affecting cassava leaves and involves two phases of model training. In the first phase, all 228 images in their original form, including background and noise, are used to train the model. This phase allows the model to learn to recognize the different types of cassava leaf diseases under various conditions, including varying levels of background noise. In the second phase, the background is removed from the 228 images, and the model is trained again. This phase helps the model focus more effectively on the key features of the cassava leaves that are directly related to the different types of diseases, eliminating any irrelevant information from the background.

The proposed method for cassava leaf disease recognition has demonstrated high levels of accuracy in both scenarios: when the target image is taken from outside the database and when it is from within the database.

The results presented in Tables 1-4 offer valuable insights into the accuracy of disease detection for cassava leaf diseases under different conditions. Table 1 indicates that when the target image is taken from outside the database and includes background, the model achieves an impressive accuracy of 96.46% for detecting Cassava Brown Disease (CBD), Cassava Mosaic Disease (CMD), and healthy leaves. This suggests that the model is capable of accurately classifying the type of disease even when it encounters new images containing background noise, highlighting its robustness and generalization capabilities. Table 2 reveals an even higher accuracy when the target image is taken from outside the database but without background. Specifically, the accuracy for CBD is

Table 1. Leaf recognition images from other than database images with background

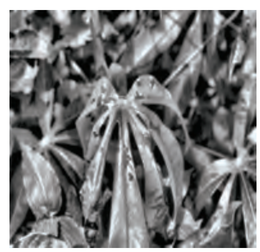
Sr. No.	Leaf Name	Input Leaf	Grayscale Image	Adaptive Histogram Image	Accuracy
1	Cassava Blight				96.46%
2	Cassava Healthy				96.46%
3	Cassava Mosaic				96.46%

Table 2. Leaf recognition images from other than database images without background

Sr. No.	Leaf	Input Leaf	Grayscale Image	Adaptive Histogram Image	Accuracy
1	Cassava Blight				99.11%
2	Cassava Healthy				98.23%
3	Cassava Mosaic				98.23%

Table 3. Leaf recognition images within the database images with background

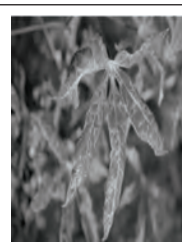
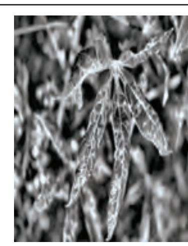
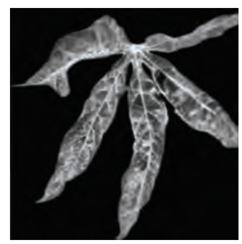
S. No.	Leaf	Input Leaf	Grayscale Image	Adaptive Histogram Image	Accuracy
1	Cassava Blight				95.53%
2	Cassava Healthy				97.32%
3	Cassava Mosaic				97.32%

Table 4. Leaf recognition images within the database images without background

S. No.	Leaf	Input Leaf	Grayscale Image	Adaptive Histogram Image	Accuracy
1	Cassava Blight				97.3214%
2	Cassava Healthy				98.2143%
3	Cassava Mosaic				99.1071%

99.11%, while for healthy and CMD leaves, it is 98.23%. These results suggest that removing the background enhances the model's accuracy, particularly for the detection of CBD, making it a valuable technique for improving disease classification performance in real-world applications where background noise can vary. Tables 3 and 4 further demonstrate the model's high accuracy when the target image is within the database, both with and without background. Table 3 shows an accuracy of 95.53% for CBD and 97.32% for both healthy and CMD leaves when background noise is present. In contrast, table 4 shows even better performance when the background is removed, with accuracies of 97.32%, 98.21%, and 99.11% for CBD, healthy, and CMD leaves, respectively. These results confirm that the model performs exceptionally well with images already present in the database, highlighting its ability to accurately classify known disease types under various conditions. Additionally, Step 9 of the cassava leaf disease detection algorithm incorporates the Whale Optimization Algorithm (WOA) for feature selection. This step uses a validation dataset consisting of 20% of the total data, with 10 whales and 100 iterations.

When the images contain background noise, the prominent features selected by the WOA were Contrast, Correlation, Homogeneity, Kurtosis, and Skewness. The accuracy of disease detection in this case was 93.33%. This suggests that these features are particularly important for accurately identifying the type of disease in cassava leaves, even when background noise is present. For images without background, the prominent features selected by the WOA were Correlation, Entropy, RMS, Kurtosis, and IDM (Inverse Difference Moment), and the accuracy of detection was 95.55%. This suggests that removing the background can improve the accuracy of disease detection, and that different features may become more relevant when the background is absent.

Interestingly, when considering both images with and without background, two common features, Correlation and Kurtosis, were found to be important for accurate detection. This suggests that these features may be particularly useful for identifying the type of disease in cassava leaves, regardless of the presence of background noise.

The results of the feature selection using the WOA suggest that different features may be important for accurate detection depending on the presence or absence of background noise. However, certain features, such as Correlation and Kurtosis, appear to be consistently important, regardless of the noise level.

CONCLUSION

In conclusion, this study aimed to develop a model capable of accurately detecting two major cassava leaf diseases, CBD and CMD. The model was tested on target images from both within and outside the database, with and without background noise. The results highlighted the significant impact of these parameters on accuracy. The highest accuracy for target images within the database was 97.32% for healthy and CMD leaves with background, while it reached 99.11% for CMD leaves without background. For images outside the database, the highest accuracy for CBD leaves with background was 96.46%, and it increased to 99.11% when the background was removed. The removal of background noise effectively improved the signal-to-noise ratio (SNR), allowing for clearer identification of the leaf object. Overall, the model demonstrated high accuracy in detecting cassava leaf diseases, particularly when background noise was absent. Although the dataset used in this study was limited, it can be expanded to improve the model's robustness. Future work could focus on enhancing the model's ability to detect additional leaf diseases and further improving its performance with background images, which would be beneficial for plant disease management and diagnosis.

Authors' contribution:

Conceptualization and design of the research work (VA, SG, MSP), Field and laboratory experimentation (VA, LS), Data collection, Data analysis, and Interpretation (VA, LS), Manuscript preparation (VA, SG, MSP).

Conflicts of interest: Authors declare that they have no conflicts of interest.

LITERATURE CITED

Ai Y, Sun C, Tie J and Cai X 2020. Research on recognition model of crop diseases and insect pests based on deep learning in harsh environments. *IEEE Access* **8**: 171686-93. https://dx.doi.org/10.1109/ACCESS.2020.3025325

Divyanth L G, Soni P and Machavaram R 2021. Cassava Leaf Disease Dataset, Mendeley Data, V1, https://dx.doi.org/10.17632/3832tx2cb2.1

Emuoyibofarhe O, Emuoyibofarhe J O, Adebayo S, Ayandiji A, Demeji O and James O 2019. Detection and classification of cassava diseases using machine learning. *Int J Comput Sci Software Engg* (*IJCSSE*). **8**(7): 166-76.

Legg J P, Kumar P L, Makeshkumar T, Tripathi L, Ferguson M, Kanju E, Ntawuruhunga P and Cuellar W 2015. Cassava virus diseases: Biology, epidemiology, and management. *Adv Virus Res* Vol. 91: 85-142. https://doi.org/10.1016/bs.aivir.2014.10.001

Mirjalili S and Lewis A 2016. The whale optimization algorithm. *Adv Eng Softw* **95**: 51-67. https://doi.org/10.1016/j.advengsoft.2016.01.008

Oyewola D O, Dada E G, Misra S and Damaševičius R 2021. Detecting cassava mosaic disease using a deep residual convolutional neural network with distinct block processing. *PeerJ Comput Sci* 7:e352. https://doi.org/10.7717/peerj-cs.352

Plested J, Shen X and Gedeon T 2021. Rethinking binary hyper parameters for deep transfer learning. **In**: Mantoro T, Lee M, Ayu M A, Wong K W and Hidayanto A N (eds.) Neural Information Processing, 28th International Conference, ICONIP 2021, Sanur, Bali, Indonesia, December 8-12, 2021, Proceedings, Part II 28, 2021, pp. 463-475, vol. 13109, Springer, Cham. https://doi.org/10.1007/978-3-030-92270-2_40

Rajput A S, Shukla S and Thakur S S 2020. Soybean leaf diseases detection and classification using recent image processing techniques. *Int J Students' Res Technol Manage* (*IJSTRTM*) **8**(3): 1-8. https://doi.org/10.18510/ijsrtm.2020.831

Ramcharan A, Baranowski K, McCloskey P, Ahmed B, Legg J and Hughes D P 2017. Deep learning for image-based cassava disease detection. *Front Plant Sci* **8**:1852. https://doi.org/10.3389/fpls.2017.01852

Ramcharan A, McCloskey P, Baranowski K, Mbilinyi N, Mrisho L, Ndalahwa M, Legg J and Hughes D P 2019. A mobile-based deep learning model for cassava disease diagnosis. *Front Plant Sci* **10:** 272. https://doi.org/10.3389/fpls.2019.00272

Sambasivam G A and Opiyo G D 2021. A predictive machine learning application in agriculture: Cassava disease detection and classification with imbalanced dataset using convolutional neural networks. *Egypt Inform J* **22**(1): 27-34. https://doi.org/10.1016/j.eij.2020.02.007