Automatic Classification of Nutritional Deficiencies in Coffee Leaves Images Using Transfer learning

Sarath Mukundan Adavakkat
MSc in Computing (with Major Options)
Dublin City University
Dublin, Ireland
sarath.mukundanadavakkat2@mail.dcu.ie

Suhara Fathima Shahul Hameed
MSc in Computing (with Major Options)
Dublin City University
Dublin, Ireland
suhara.shahulhameed2@mail.dcu.ie

***Abstract*—Coffee plants, which belong to the Coffee genus, require vital elements such as nitrogen, phosphorus, potassium, magnesium, iron, and zinc for proper growth and bean formation. Maintaining a balanced soil nutrient profile, as well as paying attention to pH and organic matter, is essential for growing healthy coffee plants and producing tasty beans. If left untreated, nutrient deficiencies in crops can result in reduced yields, poor quality, and financial losses for farmers. The capacity to discover these problems quickly and reliably using automated picture analysis represents a proactive approach to agricultural management. The focus of this study is on the automatic detection of nutritional deficiencies such as potassium, boron, calcium, and iron in coffee plants using transfer learning techniques and data augmentation strategies.**

I. Introduction

Peruvian coffee has become well-known for its high-quality, specialty coffee. Peru is now the seventh largest producer of coffee beans in the world [1]. Coffee production in Peru significantly strengthens rural employment, providing vital opportunities for farmers and agricultural workers. Its robust export volume enhances the country's trade balance, while the cultivation of coffee acts as a crucial driver in elevating rural incomes [2]. Several important nutrients are required for the healthy growth of coffee plants, and deficits in these minerals can manifest apparent symptoms. These deficiencies manifest in various forms, including leaf discolourations, stunted growth, limited fruiting, and twisted leaves, resulting from insufficient quantities of critical nutrients such as nitrogen, phosphorus, potassium, magnesium, iron, calcium, sulphur, zinc, and boron [3]. Coffee crop productivity can deteriorate as a result of delayed disease identification in crops. Detecting nutritional deficits in coffee plants requires paying close attention to leaf symptoms such as yellowing, shape deformation, discoloration, as well as abnormal growth patterns which is quite a challenging and laborious task for farmers on a daily basis as its time consuming process to analyse and take corrective actions is required. Early identification gives farmers the foresight that they need to make informed decisions, boosting effective resource management and increasing overall output in coffee cultivation.

The potency of Deep Learning has been harnessed in the past for identifying diseases in various plants, including coffee leaves. Given the distinct visual symptoms exhibited by coffee leaf diseases, images of these conditions can be used as input for training deep learning models [4]. Convolutional Neural Networks (CNN) have been employed as an effective approach in the analysis of coffee plants, demonstrating remarkable precision and efficiency in disease classification [5]. Our research focuses on the automatic detection of nutritional deficiencies such as potassium, boron, calcium, and iron in coffee plants using transfer learning techniques and data augmentation strategies on CoLeaf dataset [6].

II. Literature Review

There are few studies in the literature on the image processing of coffee leaves using the publicly available CoLeaf-DB dataset for classification of nutritional deficiencies. Initially, the first version (V1.0) CoLeaf-DB dataset consisted of 335 images with eight nutritional deficiencies classes [7]. The number of images per class was Boron -53, Nitrogen -49, Calcium -52, Iron -36, Magnesium -30, Manganese -33, Phosphorus -53, Potassium -29. Later in 2023, the images in dataset was updated to new version (V2.0) to 1006 images, consisting of Healthy leaf -6, Nitrogen -64, Phosphorus -246, Potassium -96, Magnesium -79, Boron -101, Manganese -83, Calcium -162, Iron -65, More than one deficiency -104 [6].

In [7] a machine learning technique based on decision trees and Random Forest was used to classify automatically nutritional deficiencies in coffee plants, using local and global features extracted from the leaves to build the classification model. The process started with leaf pre-processing: Otsu algorithm for segmentation, median filter for noise removal, and image size reduction for focused leaf analysis. Local features were extracted using SIFT and Bag-of-Features, including a colour descriptor. Global features encompassed basic measurements and shape features for nutritional deficiency identification. Four models were built with Random Forest: 1) Using local features; 2) Using global features; 3) Combining local and global features; and 4) Combining SIFT local features and global features. The CoLeaf dataset V1.0 with 255 images for training and 80 images (10 images per class) for testing was used. The proposed model correctly classified: Boron-10, Nitrogen-8, Calcium-9, Magnesium-3, Manganese-4, Iron-5, Phosphorus-10 and Potassium-5. Out of the four models, global features have better performance than local features with an overall accuracy of 67.5% in classifying nutritional deficiencies in coffee leaves. Furthermore, the dataset helped identify areas for improvement, like enhancing vein representation and refining necrosis detection on the leaf tip, to boost the proposed methodology's accuracy and robustness. In [8] digital image processing techniques and supervised classification algorithms on the CoLeaf dataset V1.0, 269 images for four nutrition, Boron -69, Iron -70, Potassium -56, Calcium -74, was proposed. Initially, image pre-processing using the Otsu's thresholding method for digital image segmentation. Feature extraction is then performed using the Blurred Shape Model(BSM) and the Gray Level Co-occurrence Matrix(GLCM). The extracted features were used to train supervised classifiers such as K-nearest neighbors, Naïve Bayes, and a Neural Network. BSM descriptor obtained an overall accuracy of 46.09% (124 samples correctly classified), 65.05%(175 samples correctly classified), and 59.11%(classifying correctly 159 samples) with 1-NN, Naïve Bayes, and neural network classifiers, respectively. For GLCM descriptor, an overall accuracy of 46.84% (126 samples correctly classified), 46.09%(124 samples correctly classified), and 49.81% (correctly classifying 134 samples) was achieved by 1-NN classifier, Naïve Bayes classifier and neural network-based classifier, respectively. No clear superiority between BSM and GLCM descriptors in direct performance comparison. Naïve Bayes classifier with BSM descriptor achieves best results for all nutritional deficiencies. 1-NN classifier with GLCM descriptor performs best for Boron, Iron, and Potassium deficiencies. For Calcium deficiencies, BSM descriptor outperforms in 1-NN classifier. Neural network-based classifier with BSM descriptor excels in Boron, Potassium, and Calcium deficiencies. GLCM descriptor performs better in detecting Iron deficiencies for the neural network-based classifier.Overall, Naïve Bayes with BSM descriptor outperforms in accuracy for all deficiencies in supervised classification. GLCM descriptor does not show a clear superiority among classifiers, with the best results being shared across the three classifiers for each nutritional deficiency. Despite showcasing the potential of the proposed method and digital image processing. However, the overall accuracy remained relatively low. In [9] a deep CNN (ResNet50) to automatically detect nutritional deficiencies considering the shape and color of coffee leaves was used. This was performed on the newer version V2.0 of CoLeaf dataset. The dataset was divided into a training set with 800 images and a test set with 200 images, i.e. from each group of nutritional deficiencies 80% of the images have been taken for training and 20% for evaluating the model. The model achieved an accuracy of 87.75% on the test set.

We need to highlight that the dataset size in V1.0 is relatively small, which might limit the model's ability to generalize well to diverse cases. The transition from CoLeaf-DB V1.0 to V2.0 indicates an effort to enhance the dataset's size and diversity, addressing potential limitations in earlier versions.

On the other hand, the Brazilian Arabica Coffee Leaf (BRACOL) dataset which is publicly available and consisting of 1747 images divided into two parts - a leaf dataset consisting of original images of the leaves and another symptom dataset containing cropped images of symptomatic parts of leaves. The number of images are: Healthy Leaf- 284, Leaf Miner-400, Leaf Rust – 543, Phoma- 361, Cercospora Spots- 159 in Leaf Dataset; Healthy Leaf- 251, Leaf Miner- 541, Leaf Rust – 612, Phoma- 433, Cercospora Spots- 310 in Symptom dataset [10].

In [11] a CNN model for the detection and classification of coffee leaf diseases from healthy and unhealthy coffee plants was proposed. Using transfer learning techniques, InceptionV3, data augmentation, and Mini-Batch Gradient Descent(MBGD) as an optimizer it was executed. Data augmentation techniques are applied to enhance the generalization power of the model, creating a varied and enlarged dataset. The 70% of images are used for training the model, 15% each for validation, and 15% for testing. The proposed CNN model has achieved an overall testing accuracy of 97.61% with a loss value of 0.35. This high level of accuracy is a significant contribution to the field, as it enables the accurate and timely detection of diseases in coffee plants. In [5] a CNN model with transfer learning, MobileNetV2, is used to timely detect leaf diseases in coffee plants on the same BRACOL dataset. Data augmentation techniques, such as rotating and rescaling the images, are applied to create a new set of unobserved data, thereby enlarging the dataset. The proposed model achieves a testing accuracy of 98.51% and a loss of 0.2482 after being trained for 96 epochs. The model outperforms previously existing methods with higher accuracy, demonstrating its effectiveness in detecting and classifying coffee leaf diseases. Disease detection models can be misled by focusing just on accuracy and loss because of issues like misinterpreted results and class imbalance. Contextual considerations and other metrics are necessary for a more complete assessment of the model's performance.

In [12], developed a deep learning model for identifying and categorizing Coffee leaf disease, another approach in a different dataset of coffee plants. Data collection involved gathering 1,120 images from Wolaita Sodo agricultural research and Kaggle datasets [13], categorized into healthy images and those with brown spots, wilt, and rust. Data Source – Online – 189, Onsite – 931. Healthy – 46, 224; Rust – 49, 236; Wilt – 72, 206; Brown eye spot – 22, 265 online and onsite respectively. Data augmentation techniques like rotation and flipping were used to address data imbalance and overfitting. After data augmentation, a total of 3,360 images consisting of Healthy-810, Rust-855, Wilt-834, Brown eye spot-861.Two approaches were compared: training from scratch and transfer learning. Results showed a 97.92% test accuracy and 5.25% test loss for training from scratch using CNN, Test accuracies were 99.89% for Resnet50 and 98.96% for Mobilenet, demonstrating promising results for Coffee leaf disease detection and classification using convolutional neural networks.

Additionally, there are numerous related studies exist, encompassing various datasets, diverse plant species, and employing different strategies within the realm of machine learning. In [4], a transfer learning strategy for detecting numerous diseases in tomato plants was shown to be quite successful. Using a dataset with ten classes—nine for tomato diseases and one for healthy plants—the study used pre-trained models including ResNet50, VGG16, and InceptionV3.InceptionV3 demonstrated the highest accuracy, reaching 92%, outperforming ResNet50 and VGG16. In [14], the Plant Village dataset obtained from Kaggle was used to use a transfer learning technique for plant disease prediction. Models like as VGG16, InceptionV3, and ResNet50 were used on a dataset of fifteen distinct varieties of plant leaves. All three models exhibited testing accuracies above 90%, with ResNet50 demonstrating particularly strong performance in detecting tomato and pepper bell disease. In [15] several pre-existing architecture models, such as NASNetMobile, MobileNetV2, Inception-Resnet, VGG16, and VGG19, were used for plant disease identification on the "New Plant Diseases Dataset" as a comparative research. Among the models, VGG19 had the highest accuracy, indicating that it is a useful tool for classifying plant diseases. ResNet performed better than the ResNet152 V2 model when the study evaluated its performance with that model. In [16], a deep CNN strategy for plant disease classification and detection was used, utilizing a pooled dataset of 54,000 photos from the EdenLibrary and PlantVillage datasets. Transfer learning was used to improve three cutting-edge image classifiers (ResNet34, DenseNet121, and AlexNet) that had been pre-trained on ImageNet. The study demonstrated the value of combining datasets, with DenseNet121 proving to be the most accurate model for classifying plant diseases.

REFERENCES

[1] BrandonBrandon, "Top 10 largest producer of coffee in the world 2024," Jan 2024. [Online]. Available: https://worldunfolds.com/largest-producer-of-coffee-in-the-world/

[2] ElevaFinca, "Peruvian coffee: Economic importance, sustainable practices, and market demand," Oct 2023. [Online]. Available: https://www.elevafinca.com/post/peruvian-coffee-economic-importance-sustainable-practices-and-market-demand

[3] Feb 2020. [Online]. Available: https://vikaspedia.in/agriculture/crop-production/integrated-pest-managment/ipm-for-commercial-crops/ipm-strategies-for-coffee/coffee-nutritional-deficiencies-disorders

[4] H. Singh, U. Tewari, and S. Ushasukhanya, "Tomato crop disease classification using convolution neural network and transfer learning," in *2023 International Conference on Networking and Communications (ICNWC)*, 2023, pp. 1–6.

[5] L. Dutta and A. K. Rana, "Disease detection using transfer learning in coffee plants," in *2021 2nd Global Conference for Advancement in Technology (GCAT)*, 2021, pp. 1–4.

[6] V. A. Tuesta-Monteza, "Coleaf dataset," Mar 2023. [Online]. Available: https://data.mendeley.com/datasets/brfgw46wzb/1

[7] D. Monsalve, M. Trujillo, and D. Chaves, "Automatic classification of nutritional deficiencies in coffee plants," in *6th Latin-American Conference on Networked and Electronic Media (LACNEM 2015)*, 2015, pp. 1–6.

[8] V. A. Tuesta-Monteza, H. I. Mejia-Cabrera, and J. Arcila-Diaz, "Coleaf-db: Peruvian coffee leaf images dataset for coffee leaf nutritional deficiencies detection and classification," *Data in Brief*, vol. 48, p. 109226, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352340923003451

[9] M. Vassallo-Barco, L. Vives-Garnique, V. A. Tuesta-Monteza, H. I. Mejía-Cabrera, and S. de Sipán, "Automatic detection of nutritional deficiencies in coffee tree leaves through shape and texture descriptors," 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:202707906

[10] R. A. Krohling, "Bracol - a brazilian arabica coffee leaf images dataset to identification and quantification of coffee diseases and pests," Nov 2019. [Online]. Available: https://data.mendeley.com/datasets/yy2k5y8mxg/1

[11] M. Kumar, P. Gupta, P. Madhav, and Sachin, "Disease detection in coffee plants using convolutional neural network," in *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, 2020, pp. 755–760.

[12] E. B. Paulos and M. M. Woldeyohannis, "Detection and classification of coffee leaf disease using deep learning," in *2022 International Conference on Information and Communication Technology for Development for Africa (ICT4DA)*, 2022, pp. 1–6.

[13] Jtaglione, "Coffee leaf diseases," Jan 2022. [Online]. Available: https://www.kaggle.com/code/jtaglione/coffee-leaf-diseases/data

[14] A. Lakshmanarao, N. Supriya, and A. Arulmurugan, "Plant disease prediction using transfer learning techniques," in *2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*, 2022, pp. 1–5.

[15] H. Pendhari, R. Virkar, B. Edakkalathur, and A. Jadhav, "A comparative study on algorithms for plant disease detection using transfer learning," in *2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA)*, 2023, pp. 528–533.

[16] A. Aarizou and M. Merah, "Transfer learning for plant disease detection on complex images," in *2022 7th International Conference on Image and Signal Processing and their Applications (ISPA)*, 2022, pp. 1–6.