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

Sarath Mukundan Adavakkat
School of Computing
Dublin City University
Dublin, Ireland
sarath.mukundanadavakkat2@mail.dcu.ie

Suhara Fathima Shahul Hameed
School of Computing
Dublin City University
Dublin, Ireland
suhara.shahulhameed2@mail.dcu.ie

Claudia Mazo
School of Computing
Dublin City University
Dublin, Ireland
ORCID: 0000-0003-1703-8964
claudia.mazo@dcu.ie

***Abstract*—Coffee plants, which belong to the coffee genus, play a crucial role globally, not only as the source of a popular beverage but also due to their cultural and economic significance. They serve as a vital agricultural commodity, sustaining millions of farmers, especially in underdeveloped nations, and contributing significantly to various countries' economies, playing a pivotal role in global trade. These plants require vital elements such as nitrogen, phosphorus, potassium, magnesium, iron, and zinc for proper growth and bean formation. 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 — MobileNetV2, InceptionV3, VGG19, EfficientNetV2, and ResNet50 — and data augmentation strategies. The study performs five experiments in the CoLeaf dataset V2.0: (E1) without any data augmentation, the unbalanced, and raw dataset; (E2) the unequal distribution of classes was preserved although data augmentation methods including rotation, shearing, and rescaling were applied; (E3) to create a balanced dataset we used oversampling; (E4) undersampling was used to decrease the number of samples in the majority classes to balance the dataset; (E5) we used data augmentation with oversampling to establish a balanced dataset and increase its diversity. Out of all experiments on all models, ResNet50 on E5 was the high-performing model with F1-scores of 0.96 (B), 0.93 (Ca), 0.98 (Fe), and 0.98 (K), which demonstrate its general efficiency and dependability in dividing the four nutritional classes. The results show that the ResNet50 model performs much better when data augmentation and oversampling are used.**

***Keywords*— Coffee Plants, Nutritional Deficiencies, CoLeaf Dataset, Data Augmentation, Transfer Learning, ResNet50**

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 daily 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 the 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 [6] for the classification of nutritional deficiencies [7]–[9]. 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, and Potassium 29. Later in 2023, the images in the dataset were updated to a 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]. 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 dataset V1.0 to V2.0 indicates an effort to enhance the dataset's size and diversity, addressing potential limitations in earlier versions. 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 using Otsu algorithm, median filter, and image size reduction. 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 classify: 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, for selected 269 images for four nutrition, Boron -69, Iron -70, Potassium -56, and Calcium -74, in total were proposed. Initially, image pre-processing using Otsu's thresholding method for 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%, 65.05%, and 59.11% with 1-NN, Naïve Bayes, and neural network classifiers, respectively. For the GLCM descriptor, an overall accuracy of 46.84%, 46.09%, and 49.81% was achieved by the 1-NN classifier, Naïve Bayes classifier, and neural network-based classifier, respectively. Naïve Bayes classifier with BSM descriptor achieves the best results for all nutritional deficiencies. 1-NN classifier with GLCM descriptor performs best for Boron, Iron, and Potassium deficiencies. For Calcium deficiencies, the 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. However, the overall accuracy remained relatively low. In [9] a deep CNN (ResNet50) to automatically detect nutritional deficiencies considering the shape and colour of coffee leaves was used. This was performed on the newer version of V2.0 of the CoLeaf dataset. The dataset was divided into an 80% training set and a 20% test set. One study's shortcoming is that detailed performance metrics for particular classes are not provided; instead, the study gives just an overall accuracy of 87.75% on the test set.

On the other hand, there are few works as [5], [10] using the Brazilian Arabica Coffee Leaf (BRACOL) [11] dataset which is publicly available and consists 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. In [10] 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. 70% of the images are used for training the model, 15% 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 dataset [11]. Data augmentation techniques, such as rotating and rescaling the images were applied. The proposed model achieves a testing accuracy of 98.51% and a loss of 0.2482 after being trained for 96 epochs. However, disease detection models can be misled by focusing only 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] a deep learning model for identifying and categorizing coffee leaf disease was developed. Data collection involved gathering 1,120 images from Wolaita Sodo agricultural research and Kaggle datasets [13], categorised into Healthy 46, 224; Rust 49, 236; Wilt 72, 206; Brown eye spot 22, 265 for 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, and 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 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 CNN.

Additionally, there are numerous related studies exist, encompassing various datasets, diverse plant species, and employing different strategies within the realm of machine learning [4], [14]–[16]. In [4], a transfer learning strategy for detecting numerous diseases in tomato plants was proposed. This proposal uses a dataset with ten classes — nine for tomato diseases and one for healthy plants — and 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. VGG16, InceptionV3, and ResNet50 models were used on a dataset of 15 distinct varieties of plant leaves. All three models exhibited testing accuracy 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 in terms of performance. In [16] a deep CNN strategy for plant disease classification and detection was proposed using 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.

III. Method

Figure 1 represents the graphical abstract of the proposed work. It illustrates a step-by-step process for implementing transfer learning techniques to classify coffee leaf nutrient deficiencies. 1) Training Dataset comprising four classes of images, namely Boron (B), Potassium (K), Calcium (Ca), and Iron (Fe). 2) data augmentation which uses methods like 2.1) rotation, 2.2) scaling, and 2.3) shearing to add diversity to the collection. 3) Transfer Learning model for effective training to classify nutritional deficits in coffee plants. Using well-known deep CNNs, we have used transfer learning to identify nutritional deficiencies. 4) The trained model is accurate and resilient. 5) Testing Dataset, which is independent from the training dataset. 6) Classification, identifying leaves lacking in four nutrients 6.1) boron, 6.2) potassium, 6.3) calcium, and 6.4) iron. This methodology improves the model's ability to recognize and categorize nutritional deficits in coffee leaves, guaranteeing accurate outcomes.

A. Dataset

The publicly available CoLeaf dataset V2.0 [6] which includes 1006 images of the leaves of coffee plants that have been grouped according to different nutritional deficits, including 10 classes: Healthy leaf -6, Nitrogen (N) -64, Phosphorus (P) -246, Potassium (K) -96, Magnesium (Mg) -79, Boron (B) -101, Manganese (Mn) -83, Calcium (Ca) -162, Iron (Fe) -65 and more than one deficiency -104 was used. The images in the dataset were of size 4000×3000

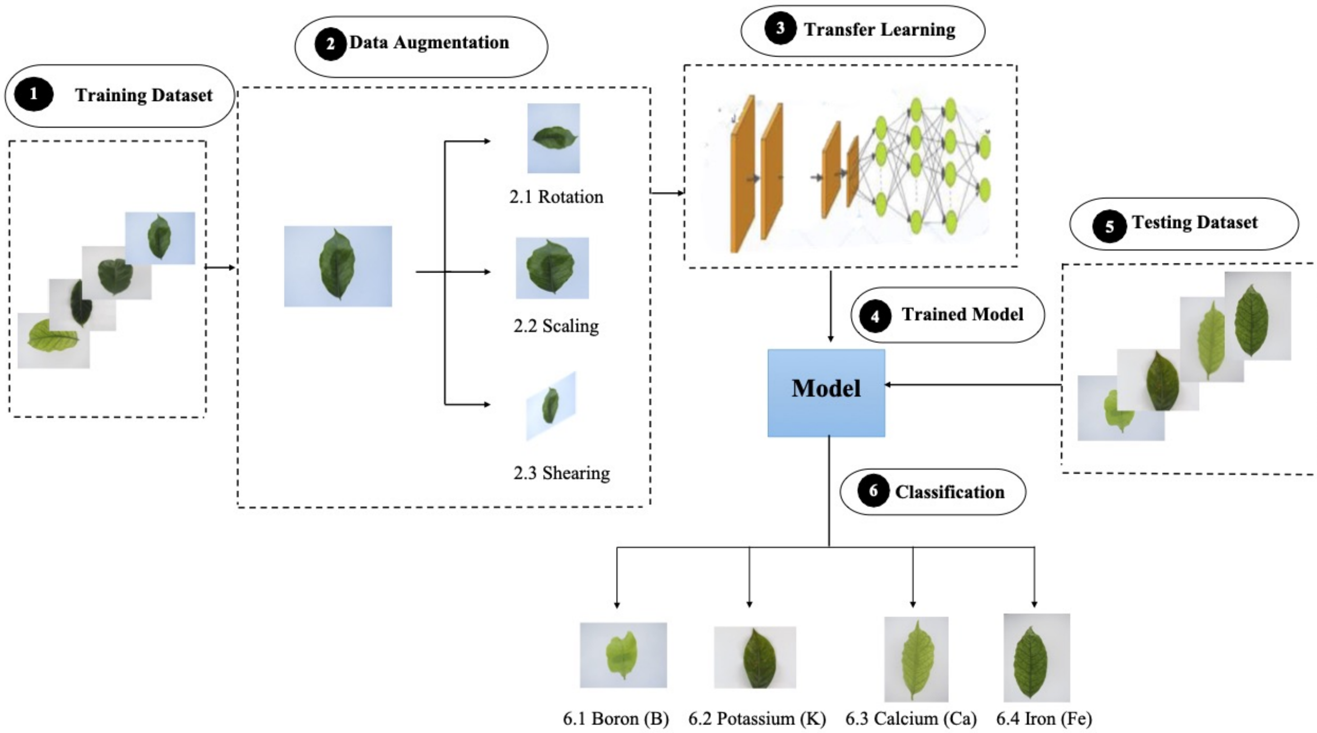

Fig. 1. Graphical Abstract of Proposed Work: (1) training dataset with leaf images. (2) data augmentation. (2.1),(2.2),(2.3), data augmentation strategies. (3) transfer learning models. (4) trained model. (5) testing dataset with test images of leaves. (6) classification of nutrients. (6.1), (6.2),(6.3), and (6.4) classified nutrients

pixels which is stored in JPEG format. In our research, four nutritional deficiencies (see Table I) will be classified: Boron (B), Calcium (Ca), Iron (Fe), and Potassium (K), in a 70:30 ratio for training and testing, respectively.

TABLE I
NUTRITIONAL DEFICIENCIES AND SYMPTOMS

Nutrition	Symptoms
Potassium (K)	Older leaves are impacted first; chlorosis along the leaf margins and tips, leading to necrosis
Boron (B)	Stunted development, interveinal chlorosis, necrosis of the leaf tips, and thick, brittle, malformed young leaves that exhibit the ”brittle leaf” sign
Calcium (Ca)	Young leaves and growing tips are mainly impacted; new leaves are twisted, hooked, or have unusual forms. Leaf tips may show necrosis
Iron (Fe)	In young leaves, interveinal chlorosis causes the veins to stay green while turning the surrounding tissue yellow. Severe cases can result in total leaf yellowing and decreased growth

B. Data Augmentation

Data Augmentation is performed on the dataset to increase the number of images. Producing more augmented images allows models to use additional training data, particularly for uncommon conditions where source data variances are absent or dataset size is small.

A range of methods to improve the dataset's diversity and resilience is used. One such method is rotation [5], in which rotations are applied to the leaf images to replicate orientations frequently found in real-world situations. This ensures that our model is effective from various viewpoints and makes it invariant to changes in leaf orientation.

Additionally, scaling [5] is performed, which entails resizing the leaf images to various sizes. This enables us to replicate differences in leaf size, which can improve the model's ability to generalize to different-sized leaves. Furthermore, shearing is carried out, which is a method that moves the pixels along one axis while maintaining the other axis stationary, thereby distorting the leaf images. Shearing can improve the model's capacity to identify and categorize leaves under various deformations by introducing geometric distortions akin to those brought on by elements like wind or physical damage to leaves.

The goal of using these data augmentation approaches is to add a variety of variations to the dataset so that our deep learning model may acquire strong features and perform better in identifying and categorizing nutritional deficits in coffee plant leaves.

In the first experiment (E1), the dataset used is the original one without any data augmentation applied. This means that the dataset remains in its raw form as initially collected. Consequently, the class distribution in this dataset is imbalanced. The second experiment (E2)

involves using the original dataset with data augmentation techniques applied. Data augmentation techniques like rotation, shearing, and rescaling were applied, so all the images were resized into 224×224 pixels. Despite the augmentation, the class distribution remains imbalanced. In the third experiment (E3), the original dataset is balanced through the process of oversampling. Oversampling involves duplicating samples from the minority class to ensure that all classes have a similar number of samples. Balancing the dataset in this way helps mitigate the bias toward the majority class, promoting fairer training across all classes. The fourth experiment (E4) also aims to balance the dataset, but it uses undersampling. Undersampling reduces the number of samples in the majority class to match the minority class, resulting in a balanced dataset. In the fifth experiment (E5), oversampling with data augmentation is applied to achieve a balanced dataset to ensure that each class has an equal number of samples, leading to a balanced class distribution.
Table II presents the five different experiments on the same dataset. Additionally, Table III shows information about the number of images used in all the experiments.

TABLE II
COMPARISON OF DATASETS BASED ON BALANCE/IMBALANCE CLASSES AND DATA AUGMENTATION

Experiment	**Balanced/Imbalanced**	**Data Augmentation**
E1: Original Dataset	Imbalanced	No
E2: Original Dataset with Data Augmentation	Imbalanced	Yes
E3: Oversampling Original Dataset	Balanced	No
E4: Undersampling Original Dataset	Balanced	No
E5: Oversampling Data Augmented Dataset	Balanced	Yes

TABLE III
NUMBER OF IMAGES IN THE FIVE EXPERIMENTS

Experiment	**Total Images**	**K**	**B**	**Ca**	**Fe**
E1	424	96	101	162	65
E2	1275	288	306	486	195
E3	648	162	162	162	162
E4	260	65	65	65	65
E5	1944	486	486	486	486

C. Transfer Learning

In contrast to classical machine learning, which applies its learning process to tasks without accounting for prior experience, Transfer Learning considers prior tasks to comprehend new tasks. Using transfer learning techniques, it provides an invaluable tool for scholars and professionals working on systems for the detection and categorization of these shortcomings of classical machine learning [5].

We test several transfer learning models employing MobileNetV2 [17], InceptionV3 [18], VGG19 [19], EfficientNetV2 [20], and ResNet50 [21] architectures with the CoLeaf dataset in our study on identifying and categorizing nutritional deficiencies. The transfer learning models were trained on the ImageNet dataset (over 14 million high-resolution images that are sorted into thousands of item types and are ranged according to the WordNet hierarchy). The convolutional layers remain unchanged during the training phase to prevent any modifications. Meanwhile, the fully connected layers are adjusted to fit the new problem by changing the output vector length to six. Following these adjustments, the training phase is carried out. Further specifics of the implementations can be found in Table IV.

TABLE IV
DETAILS OF CNN ARCHITECTURES USED IN OUR DOMAIN.

Item	**MobileNetV2**	**InceptionV3**	**VGG19**	**EfficientNetV2**	**ResNet50**
General					
Parameters	32.4M	23.3M	20.4M	7.9M	26.7M
Channels	3	3	3	3	3
Input size	224×224	299×299	224×224	224×224	224×224
Number of layers					
Convolutional	53	94	16	55	53
Max pooling	1	4	5	1	1
Fully connected	1	1	3	2	1
Presence of modules					
Batch normalization	yes	yes	no	yes	yes
Residual connections	no	no	no	no	yes

For image classification, MobileNetV2, InceptionV3, VGG19, EfficientNetV2, and ResNet50 are frequently used due to their ability to balance accuracy, efficiency, and [4], [5]. Pre-trained weights, great precision, and adaptability are all features of these models that make them appropriate for a range of applications.

IV. EXPERIMENTS AND RESULTS

This section outlines the experiments that underpin the proposed approach. We present the outcomes in three steps: (i) Comparison of Experiments, (ii) Comparison of Transfer Learning Models on E5, and (iii) Comparison with Pre-existing Models.

Accuracy, F1-score, Precision, and Recall are the four metrics obtained from the confusion matrix that was used to evaluate this work's performance in classifying B, Ca, Fe, and K. These measures are expressed mathematically as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{3}$$

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{4}$$

where TP is True Positive, FP is False positive, TN is True Negative, and FN is False Negative.

A. Comparison of Experiments

Table V presents an overview of the accuracy and Figure 2 presents the F1-score of five distinct machine learning models (MobileNetV2, InceptionV3, VGG19, EfficientNetV2, and Resnet50) with the five different experiments.

TABLE V
OVERALL ACCURACY OF FIVE MODELS ACROSS EXPERIMENTS

Model	**E1**	**E2**	**E3**	**E4**	**E5**
MobileNetV2	0.862	0.882	0.862	0.888	**0.959**
InceptionV3	0.815	0.772	0.832	0.762	**0.916**
VGG19	0.692	0.780	0.796	0.688	**0.842**
EfficientNetV2	0.869	0.885	0.872	0.862	**0.933**
Resnet50	0.869	0.869	0.872	0.812	**0.961**

The best-performing experiment is E5, exhibited by the overall accuracy in Table Vand F1-score from Figure 2 for the five models

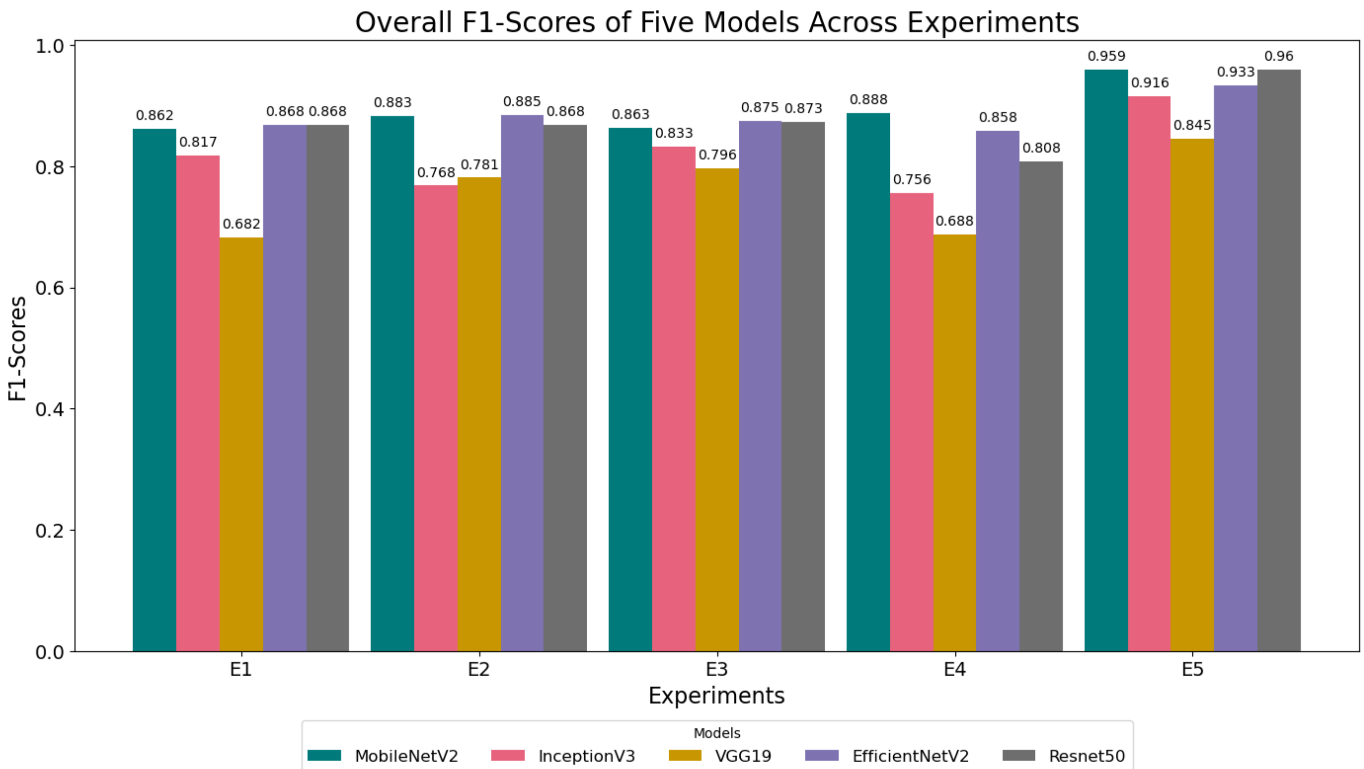

Fig. 2. Overall F1-Score on Experiments

across experiments. An overall accuracy of 0.959, 0.916, 0.842, 0.933, 0.961 and overall F1-score of 0.959, 0.916, 0.845, 0.933, 0.960 across MobileNetV2, InceptionV3, VGG19, EfficientNetV2, and ResNet50, respectively, consistently showing the greatest scores in this experiment. The utilization of data augmentation and oversampling approaches in E5 is accountable for its remarkable performance, which probably improved the models' capacity to generalize and function effectively on a variety of data sets.

In contrast, E4 is the experiment that performs low due to a lesser number of images as it considers undersampling the original dataset. In E4, the results are significantly lower, especially for InceptionV3 and VGG19 models with accuracy scores of 0.762 and 0.688, respectively. The F1-scores also illustrate this trend with values of 0.756 and 0.688 respectively, with E4 exhibiting the least effectiveness out of all the trials.

B. Comparison of Transfer Learning Models on E5

This section outlines the comparison of transfer learning models implemented on E5. We select the most efficient transfer learning model comparing MobileNetV2, VGG19, InceptionV3, EfficientNetV2, and ResNet50 architectures to classify Boron, Calcium, Iron, and Potassium in coffee leaves nutrients in terms of precision, recall, and F1-score measurements.

Table VI and VII illustrate the precision and recall achieved by the five models on E5. Figure 3 represents F1-score of five transfer learning models on E5.

From comparing the transfer learning models on E5, ResNet50 was the best model in terms of F1-scores of four class. Eventhough, they aren't best values from precision and recall out of five models, it performs consistently and follows top models with a little difference. EfficientNetV2 achieves the maximum precision of 0.98 and maintains a recall of 0.88 for B, while MobileNetV2 and ResNet50, slightly outperform EfficientNetV2 in recall at 0.95. When it comes to Ca, EfficientNetV2 performs better in recall 0.94 but MobileNetV2 leads in precision 0.98. VGG19 has the highest precision of 0.98 for Fe, however, EfficientNetV2 exhibits a recall of 1.00. EfficientNetV2 scores the greatest in precision at 0.99 for K, whereas Resnet50 and MobileNetV2 match this performance in recall at 0.99. ResNet50 follows the top models by a small margin but consistently exhibits great recall and precision across four nutrient classes. In terms of precision, ResNet50 comes in extremely close to the best models: for B, K it lags EfficientNetV2 by just 0.02;for Ca, it lags MobileNetV2 by 0.04 and for Fe, it lags VGG19 by 0.01.Likewise, in terms of recall, ResNet50, along with MobileNetV2 is highest for B and K with a value of 0.95 and 0.99 respectively. For Ca and Fe, ResNet50

TABLE VI
PRECISION OF FIVE MODELS ACROSS NUTRIENTS

Class	MobileNetV2	InceptionV3	VGG19	EfficientNetV2	Resnet50
B	0.96	0.96	0.82	**0.98**	0.96
Ca	**0.98**	0.91	0.71	0.91	0.94
Fe	0.95	0.90	**0.98**	0.87	0.97
K	0.95	0.89	0.94	**0.99**	0.97

TABLE VII
RECALL OF FIVE MODELS ACROSS NUTRIENTS

Class	MobileNetV2	InceptionV3	VGG19	EfficientNetV2	Resnet50
B	**0.95**	0.88	0.87	0.88	**0.95**
Ca	0.91	0.86	0.92	**0.94**	0.92
Fe	0.99	0.97	0.77	**1.00**	0.98
K	**0.99**	0.95	0.82	0.91	**0.99**

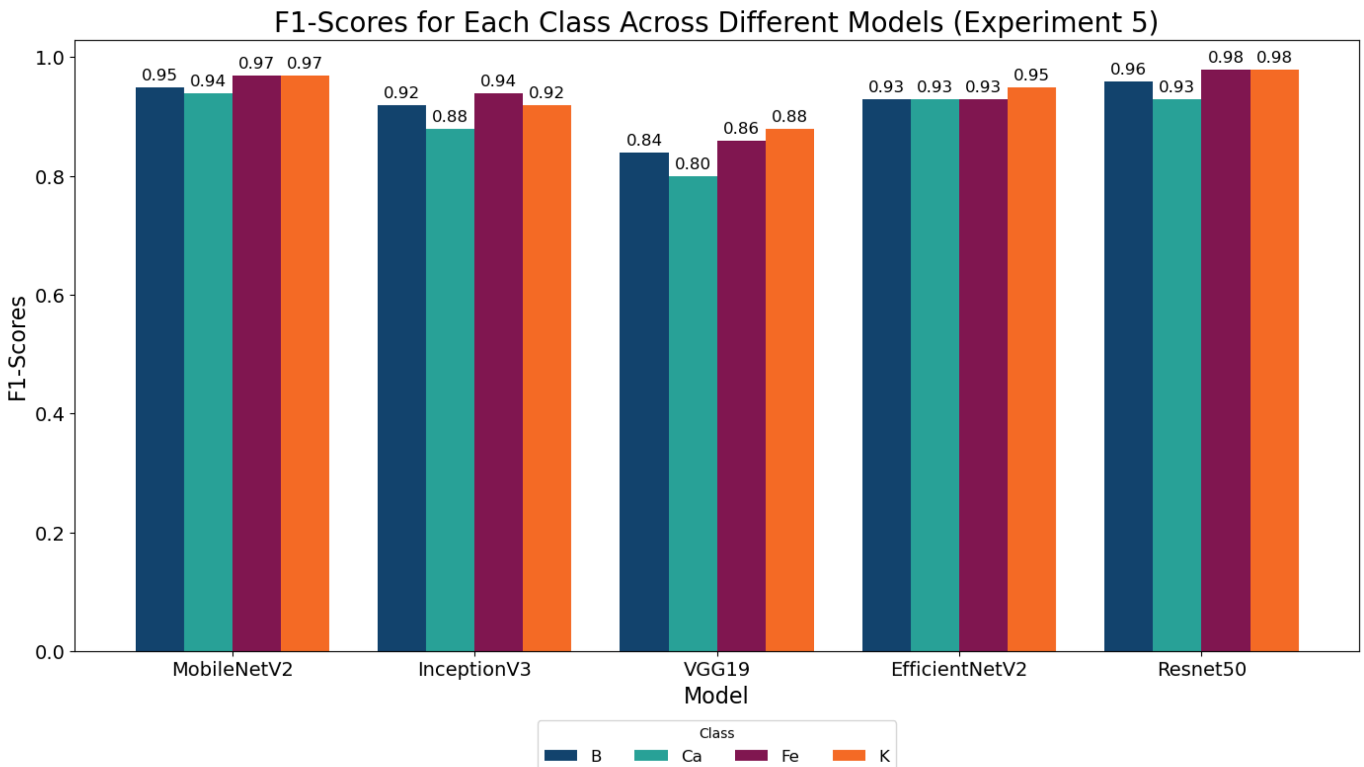

Fig. 3. F1-Score for five models using E5

lags EfficientNetV2 by 0.02. ResNet50 achieved precision scores of 0.96, 0.94, 0.97, and 0.97 for B, Ca, Fe, and K, and recall scores of 0.95, 0.92, 0.98, and 0.99 for the same. According to Figure 3, top results were obtained with F1-scores of 0.96, 0.93, 0.98, and 0.98 for B, Ca, Fe, and K, respectively, suggesting its consistency in F1-scores in ResNet50. Fe and K, have the highest F1-score for Resnet50 with 0.98 and Ca is the lowest performing class for Resnet50 with an F1-score of 0.93, despite the relatively minor difference in performance between classes. Conversely, VGG19 is shown to be the least effective model, with lower precision and recall values: it has lower F1-scores of 0.84, 0.80, 0.86, and 0.88 for B, Ca, Fe, and K, respectively; Precision scores of 0.82, 0.71, 0.98, and 0.94; Recall scores are 0.87, 0.92, 0.77, and 0.82.

C. Comparison with Pre-existing Models

In [7], the CoLeaf dataset V1.0 with eight classes Boron (43), Magnesium (20), Nitrogen (39), Manganese (23), Calcium (42), Phosphorus (43), Iron (26), and Potassium (19) were used. They used Random Forest, a decision tree-based machine learning technique, to classify nutritional deficiencies in coffee plants. For local feature extraction, the SIFT algorithm is utilized and defines global features based on shape and colour. An experimental evaluation with 335 images of coffee leaves showed that global features performed better than local features, achieving an accuracy of 67.5%.

In [8] an automatic detection of nutritional deficiencies in coffee leaves using the CoLeaf dataset V1.0, which comprises 269 photos categorized into 4 categories — Boron (69), Calcium (74), Iron (70), and Potassium (54)— was proposed. The Blurred Shape Model and Gray-Level Co-occurrence Matrix are the main shape and texture descriptors, while classifiers like KNN, Naïve Bayes, and Neural Networks are used for detection. The best results for identifying the four nutritional deficits are obtained using the Naïve Bayes classifier with BSM descriptor. In our proposal, we use the same four nutrients depicted in 424 photos from the CoLeaf dataset V2.0 — boron (101), calcium (162), iron (65), and potassium (96). We used transfer learning techniques using architectures such as MobileNetV2, InceptionV3, VGG19, EfficientNetV2, and Resnet50 combined with data augmentation. Table VIII depicts the comparison between our model and the proposed method in [8].

TABLE VIII
PRECISION AND RECALL COMPARISON BETWEEN NAÏVE BAYES AND RESNET50

Class	Naïve Bayes		ResNet50	
	Precision	Recall	Precision	Recall
B	0.67	0.69	**0.96**	**0.95**
Ca	0.64	0.56	**0.94**	**0.92**
Fe	0.68	0.70	**0.97**	**0.98**
K	0.59	0.64	**0.97**	**0.99**
Overall	0.64	0.65	**0.96**	**0.96**

In terms of precision, recall, from Table VIII and F1-score from Figure 4, the performance comparison between Naïve Bayes and our proposed ResNet50 model shows that ResNet50 performs better than Naïve Bayes in every class and measurement. Using ResNet50, the highest precision scores for B, Ca, Fe, and K were 0.96, 0.94, 0.97, and 0.97, respectively. For these classes, Naïve Bayes scored lower on precision 0.67,0.64,0.68, and 0.59, for B, Ca, Fe, and K, respectively. ResNet50 performed better in recall than Naïve Bayes, achieving the highest recall scores of 0.95, 0.92, 0.98, and 0.99 for B, Ca, Fe, and K, respectively. Naïve Bayes yielded recall scores of 0.69, 0.56, 0.70, and 0.64 for B, Ca, Fe, and K, respectively. With an overall

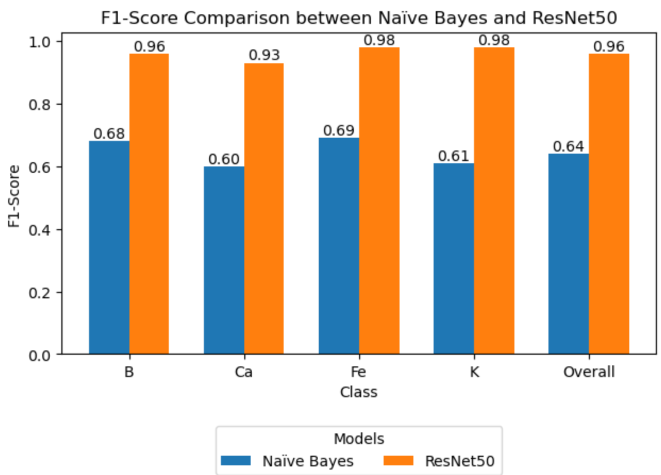

Fig. 4. F1-Score Comparison between Naive Bayes and ResNet50

precision and recall of 0.96 for both of them, ResNet50 achieved the highest values. ResNet50 performs better than Naïve Bayes in the four classes, according to the F1-score. In comparison to Naïve Bayes, ResNet50 produces notable improvements of 26% to 38% in precision and recall rates for four classes. This demonstrates that ResNet50 outperforms Naïve Bayes in identifying nutritional deficits in coffee plants. With a value of 0.96, our model ResNet50 obtained the best overall F1-score. They used CoLeaf Dataset V1.0 with less number of images compared to CoLeaf Dataset V2.0 with more images which we used.

Later in [9] the CoLeaf dataset V2.0 with 10 classes- Healthy leaf (6), Nitrogen (64), Phosphorus (246), Potassium (96), Magnesium (79), Boron (101), Manganese (83), Calcium (162), Iron (65), More than one deficiency (104) ResNet50 was used to classify all coffee nutritional deficiencies in the Coleaf-DB dataset with an overall accuracy of 87.75%. In our research, the best model ResNet50 on E5 achieved an overall accuracy of 96%.

A drawback of [7], [9] is that only an overall accuracy is provided and does not provide a per-class result, which restricts our ability to understand how well the model performs in various classes.

V. Conclusions

In this paper, we applied transfer learning models and data augmentation approaches for the classification of nutritional deficiencies in B, Ca, Fe, and K in coffee plant leaves. The CoLeaf dataset, which is publicly available and contains images of coffee leaves with a range of nutritional deficits, was used in the study to train and verify models for classifying four deficiencies (B, Ca, Fe, and K). By applying data augmentation techniques, such as rotation, scaling, and shearing, model performance was improved and the dataset's diversity was significantly expanded.

Firstly, out of five different experiments conducted, the E5 showed the highest accuracy and F1-scores across all models ranging from 0.86 to 0.96. Secondly, ResNet50 was the best-performing model across MobileNetV2, InceptionV3, VGG19, and EfficientNetV2 with an overall F1-score of 0.96. The results highlight how crucial balanced datasets and data augmentation are to improving deep learning models' capacity for generalization. In addition to providing a reliable method for identifying nutritional deficits B, Ca, Fe, and K in coffee plants, this research offers insightful information about the application of transfer learning and data augmentation for coffee plants. This information may help with crop management and yield optimization.

Three studies have classified nutritional deficits in coffee plants using the CoLeaf dataset. In [7], they used Random Forest with global features based on shape and color and local features derived using the SIFT technique. With 335 images from eight classes from CoLeaf Dataset V1.0, they achieved an accuracy of 67.5%, with global features outperforming local features. In [8], 269 images taken from CoLeaf V1.0 were classified into four classes -B, Ca, Fe and K. The best results were obtained by the Naïve Bayes classifier using the Blurred Shape Model (BSM) descriptors with an accuracy of 65.05%. In [9], ResNet50 obtained an accuracy of 87.75% using CoLeaf V2.0 with 10 classes. They incorporated all classes in their research. In studies by [7] and [9], however they did not provide per-class performance information and instead concentrated on total accuracy 67.5% and 87.75%, respectively. We achieved 96% overall accuracy with our ResNet50 model on E5, our study produced a more thorough performance analysis and specific per-class information.

In the future, we will extend this proposal by working on the following three research lines. Firstly, by adding new classes to the classification process from the original dataset, we want to expand the CoLeaf dataset's classification by incorporating all eight nutritional deficiency classifications as we have only considered four classes in our study. Secondly, increasing the dataset using strategies like data augmentation techniques or synthetic image generation using GANs. Thirdly, ensemble approaches that integrate various transfer learning models to make use of each one's unique advantages.

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] 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

[9] 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

[10] 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.

[11] 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

[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.

[17] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," 2019. [Online]. Available: https://arxiv.org/abs/1801.04381

[18] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," 2015. [Online]. Available: https://arxiv.org/abs/1512.00567

[19] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015. [Online]. Available: https://arxiv.org/abs/1409.1556

[20] M. Tan and Q. V. Le, "Efficientnetv2: Smaller models and faster training," 2021. [Online]. Available: https://arxiv.org/abs/2104.00298

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015. [Online]. Available: https://arxiv.org/abs/1512.03385