



WEB APPLICATION BASED ON CNN FOR CLASSIFICATION OF BIOTIC AND ABIOTIC STRESSES IN COFFEE LEAVES

Daniel Henrique Leite^{1*}, Domingos Sárvio Magalhães Valente¹, Gabriel Dumbá Monteiro de Castro¹, Pedro Maya Ferreira Arruda², Fábio Daniel Tancredi³ and Daniel Marçal de Queiroz¹

¹ Department of Agricultural Engineering, Federal University of Viçosa, Viçosa 36571-900, MG, Brazil; valente@ufv.br; gabriel.dumba@ufv.br; queiroz@ufv.br

² Department of Agricultural Engineering and Environment, Federal Fluminense University, Niterói 24210-240, RJ, Brazil; pedromaya@id.uff.br

³ Minas Gerais Agricultural Research Agency (EPAMIG-Sudeste), Viçosa 36570-000, MG, Brazil; fabio.tancredi@epamig.br

* Correspondence: daniel.h.leite@ufv.br

**A paper from the Proceedings of the
17th International Conference on Precision Agriculture and 11th
Brazilian Congress on Precision and Digital Agriculture
13-16 July 2026
Porto Alegre, Brazil**

Abstract.

The use of digital systems can assist coffee growers and professionals in diagnosing stresses that affect coffee plantations, ensuring that crop management is carried out correctly and efficiently. Therefore, the aim of this study was to develop a web application based on a pre-trained Convolutional Neural Network to classify coffee leaf images exhibiting symptoms of biotic and abiotic stresses. Initially, a dataset consisting of coffee leaf images affected by biotic and abiotic stresses was constructed. Image acquisition was performed at the field level and through a review of publicly available datasets to build the dataset. The dataset comprises 15 classes: Boron Deficiency, Calcium Deficiency, Cercospora, Healthy, Iron Deficiency, Magnesium Deficiency, Manganese Deficiency, Miner, More Deficiencies, Nitrogen Deficiency, Phoma, Phosphorus Deficiency, Potassium Deficiency, Rust, and Sunburn. The total number of images was 2,539, and the dataset was split into training, validation, and test sets at 80%, 10%, and 10%, respectively, yielding 2,047 for training, 246 for validation, and 246 for testing. To increase the diversity and size of the training set, Data Augmentation was applied using a Keras sequential pipeline that included random horizontal and vertical flipping, rotation by 0.3, translation by 0.3 in both height and width, and Gaussian noise with a standard deviation of 0.3. These transformations were applied exclusively to the training set with an augmentation factor of 20% per class, expanding it from 2,047 to 2,453 images. The model was compiled using the AdamW optimizer with a learning rate of 0.0001 and a weight decay of 0.0001, employing the Sparse

The authors are solely responsible for the content of this paper, which is not a refereed publication. Citation of this work should state that it is from the Proceedings of the 17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture. EXAMPLE: Last Name, A. B. & Coauthor, C. D. (2026). Title of paper. In Proceedings of the 17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture (unpaginated, online). Monticello, IL: International Society of Precision Agriculture and Brazilian Congress on Precision and Digital Agriculture.

Categorical Crossentropy loss function, and trained for up to 50 epochs with a batch size of 4. The following callbacks were employed: ModelCheckpoint to save the best weights based on validation loss, EarlyStopping, and ReduceLROnPlateau with a reduction factor of 0.1 and a patience of 6 epochs. For training, the DenseNet201 Convolutional Neural Network architecture was selected and implemented using the Transfer Learning technique. To integrate the trained model into a web application, a Flask-based web application was developed, using HTML and CSS for the front-end and the trained model for the back-end. After training, the best model was saved using the ModelCheckpoint callback at the point of lowest validation loss, and subsequently loaded to evaluate the test set. The model achieved an accuracy of 0.9268, precision of 0.9110, recall of 0.8825, and F1-score of 0.8920. Once the model was trained and evaluated, it was deployed in the web application for coffee leaf stress classification. Therefore, it can be concluded that the web application developed in this study is useful and effective for coffee growers and professionals, potentially improving leaf diagnosis and enhancing crop performance.

Keywords: Deep Learning, Transfer Learning, DenseNet201, Plant Disease Detection, Nutritional Deficiency Diagnosis, Digital Agriculture, Flask, Data Augmentation.

1 INTRODUCTION

Coffee is one of Brazil's main crops, and the country ranks as the world's largest producer and exporter of the commodity. This activity plays a major role in the national economy, contributing to job creation, sector income, and the country's Gross Domestic Product (Ferreira 2023). Maintaining coffee yield and quality depends largely on correctly identifying the factors that affect plant health throughout the production cycle. For this reason, diagnosing leaf stress is an important step for crop management and loss reduction.

Coffee leaves are subject to stresses of different natures, generally grouped into biotic stresses, caused by living organisms, and abiotic stresses, resulting from unfavorable environmental conditions (Andrade et al. 2020). In addition to these factors, there are nutritional disorders associated with a shortage of nutrients required by the plant. Together, these stresses impair plant functions and can reduce yield and product quality, as well as raise production costs (Cerdeira et al. 2017; Esgario et al. 2020). Since most of these symptoms manifest in the leaf, the leaf is the main organ observed during diagnosis.

Among the biotic stresses are fungal diseases and insect damage. Rust (*Hemileia vastatrix*) is among the diseases with the greatest impact on production. At the same time, Cercospora (*Cercospora coffeicola*) and phoma leaf spot (*Phoma spp.*) also affect leaves and fruits, with lesions of variable appearance (Andrade et al. 2016; Catarino et al. 2016). As for pests, the coffee leaf miner (*Leucoptera coffeella*) reduces photosynthetic area by forming mines on leaves and can cause premature leaf drop, resulting in losses that can reach half of the production (Dantas et al. 2021).

Among abiotic stresses, sunscald, caused by sun exposure, leads to chlorosis and necrosis in the most exposed regions of the leaf under conditions of high temperature and low water availability (Borgo et al. 2024). Nutritional disorders, in turn, result from the deficiency of macronutrients and micronutrients, such as nitrogen, phosphorus, potassium, calcium, magnesium, boron, iron, and manganese, and can occur in isolation or simultaneously in the same leaf (Tuesta-Monteza et al. 2023). These deficiencies alter the color, shape, and integrity of the leaf tissue, and their symptoms often resemble one another and also those of some diseases.

The diversity of causes and the similarity of symptoms make diagnosis difficult, since stresses of different origins can produce similar visual responses in the leaf. Assessment is generally carried out visually by trained professionals, but it is subjective and prone to error when symptoms are confused with one another (Esgario et al. 2020). Since experts are not always available in the field, digital tools to support diagnosis can broaden access to this type of analysis and contribute

to better-informed management decisions.

In this context, Artificial Intelligence techniques, especially Deep Learning methods, have been applied to pattern recognition in plant images. Hassan et al. (2021) identified 38 disease classes in 14 species, and Yang et al. (2019) diagnosed cold damage in maize using Convolutional Neural Networks. In coffee, Sorte et al. (2018) applied Deep Learning to disease detection, and Leite (2018) addressed mineral deficiencies and pests using computer vision. Most of these studies, however, focus on a small number of classes or on a single type of stress, leaving room for approaches that integrate diseases, pests, abiotic stress, and nutritional deficiencies into a single tool.

In light of this, the objective of this study was to develop a web application based on a pre-trained Convolutional Neural Network to classify coffee leaf images showing symptoms of biotic and abiotic stresses. Specifically, the aim was to develop and evaluate pre-trained Convolutional Neural Network models for the classification of 15 classes of leaf stresses and, based on the best-performing model, to implement an application built on the Flask microframework, capable of returning the most likely class and generating a PDF report to support diagnosis, intended for coffee growers, researchers, and other professionals in the coffee production chain.

2 MATERIALS AND METHODS

2.1 Dataset

The dataset used in this study consists of coffee leaf images affected by biotic and abiotic stresses, compiled from two sources. The first corresponds to field acquisition, conducted in localities in the state of Minas Gerais, Brazil, at different times of year, owing to the occurrence of symptoms throughout the crop cycle. The second corresponds to the publicly available CoLeaf dataset (Tuesta-Monteza et al. 2023) which was used to increase the number of images and the number of classes related to nutritional deficiencies.

The full dataset comprises 15 classes. Nine of them come from CoLeaf: eight correspond to nutritional deficiencies associated, respectively, with the nutrients Boron, Calcium, Iron, Magnesium, Manganese, Nitrogen, Phosphorus, and Potassium, and the ninth, named More Deficiencies, gathers the occurrence of more than one deficiency in the same leaf. The remaining six classes were acquired in the field and correspond to Cercospora, Phoma, and Rust (diseases); Miner (pest); Sunburn (sun-induced burn); and Healthy (healthy leaves). The dataset comprises 2,539 images in total.

The dataset was partitioned into training, validation, and test sets, in proportions of 80%, 10%, and 10%, respectively. This split resulted in 2,047 images for training, 246 for validation, and 246 for testing. The distribution of images by class and by set is presented in Table 1.

Table 1. Distribution of images by class and by dataset

Class	Training	Validation	Test	Total
Boron	81	10	10	101
Calcium	134	16	16	166
Cercospora	200	25	25	250
Healthy	206	25	25	256
Iron	56	6	6	68
Magnesium	71	7	7	85
Manganese	70	8	8	86
Miner	200	25	25	250
More Deficiencies	89	10	10	109
Nitrogen	55	6	6	67

Phoma	200	25	25	250
Phosphorus	198	24	24	246
Potassium	87	9	9	105
Rust	200	25	25	250
Sunburn	200	25	25	250
Total	2,047	246	246	2,539

2.2 Data Augmentation and Preprocessing

Owing to the small number of images in the dataset, especially in some classes, the Data Augmentation technique was adopted to increase the diversity of the training set and reduce the risk of overfitting. The transformations were applied only to the training set using a Keras sequential pipeline that included random horizontal and vertical flipping, rotation by 0.3, translation by 0.3 in both height and width, and the addition of Gaussian noise with a standard deviation of 0.3. The augmentation was applied per class with a 20% factor, expanding the training set from 2,047 to 2,453 images. The validation and test sets did not undergo Data Augmentation. Finally, all images were resized to 224×224 pixels and normalized to the range $[0, 1]$.

2.3 Models Based on Convolutional Neural Networks

Owing to the small number of images, even after applying Data Augmentation, the Transfer Learning technique was adopted, leveraging models pre-trained on the ImageNet database. Two architectures were selected, DenseNet201 (Huang et al. 2017) and VGG16 (Simonyan and Zisserman 2014), both implemented with the Keras library. In each architecture, the pre-trained weights were frozen and, on top of them, a classification structure was added, composed of a Flatten layer, a Dense layer with 256 neurons and ReLU activation, a Dropout layer with a rate of 0.2, and an output Dense layer with 15 neurons, corresponding to the number of classes.

The models were compiled with the AdamW optimizer, a learning rate of 0.0001, and a weight decay of 0.0001, using the Sparse Categorical Crossentropy loss function. Training was carried out for up to 50 epochs, with a batch size of 4. The following callbacks were used: ModelCheckpoint to save the best weights based on validation loss; EarlyStopping to halt training in the absence of improvement; and ReduceLROnPlateau with a reduction factor of 0.1 and a patience of 6 epochs. The two models were compared to identify the best-performing one. The configuration and parameter count for each model are presented in Table 2.

Table 2. Configuration and number of parameters of the trained models

Model	Input dimension	Total parameters	Trainable parameters
DenseNet201	$224 \times 224 \times 3$	42,410,582	24,088,591
VGG16	$224 \times 224 \times 3$	21,141,334	6,426,639

2.4 Evaluation Metrics

To evaluate and compare model performance, the accuracy, precision, recall, and F1-score metrics were used. These metrics were calculated using the confusion matrix, which compares the actual and predicted classes and enables detailed performance analysis. In the equations, TP corresponds to true positives, TN to true negatives, FP to false positives, and FN to false negatives.

Accuracy measures the proportion of correct predictions relative to the total number of predictions, as in Equation (1). Precision is the fraction of correct positive predictions, as in Equation (2). Recall is the fraction of positive instances correctly identified, as in Equation (3). The F1-score is the harmonic mean of precision and recall and combines these two metrics into a single value, as in Equation (4).

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

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

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

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

2.5 Web Application Development

The purpose of the web application is to integrate the trained model into the classification of coffee leaf images, so that coffee growers, researchers, and other professionals in the field can identify leaf symptoms from a computer or smartphone with internet access.

The application was developed with the Flask microframework, written in Python, and intended for web application development. The front-end was built with HTML and CSS, and the back-end incorporates the trained model. In operation, the user uploads an image, which is resized to the model's input dimension. Upon triggering the classification, the model is run and returns the probability that the image belongs to each class; the application highlights the class with the highest probability and lists the six most likely classes. In addition, the application allows generating a PDF report that combines the analyzed image, the identified class, the associated probability, and a descriptive text about the identified class.

3 RESULTS AND DISCUSSION

3.1 Model Training and Validation

The two models were trained for up to 50 epochs and evaluated on the validation set. During training, the best weights of each model were saved based on validation loss. Both models reached a training accuracy above 99%. However, since the goal of a model is to generalize correctly to new data, performance on the validation set is a more appropriate indicator. At this stage, both models achieved an accuracy close to 84%, indicating similar performance. The difference between the training and validation accuracies suggests a tendency toward overfitting, which Data Augmentation and the Dropout layer mitigate. The small number of images in some classes may have contributed to classification errors, suggesting the need to increase the quantity and diversity of images.

3.2 Model Performance on the Test Set

After training, the best weights for each model were selected for evaluation on the test set, which comprised 10% of the total images and was not used in the training or validation stages. A confusion matrix was generated for each model, presented in Figures 1 and 2. In these matrices, each row represents the actual class and each column the predicted class; the concentration of the highest values along the main diagonal indicates the tendency toward correct classification.

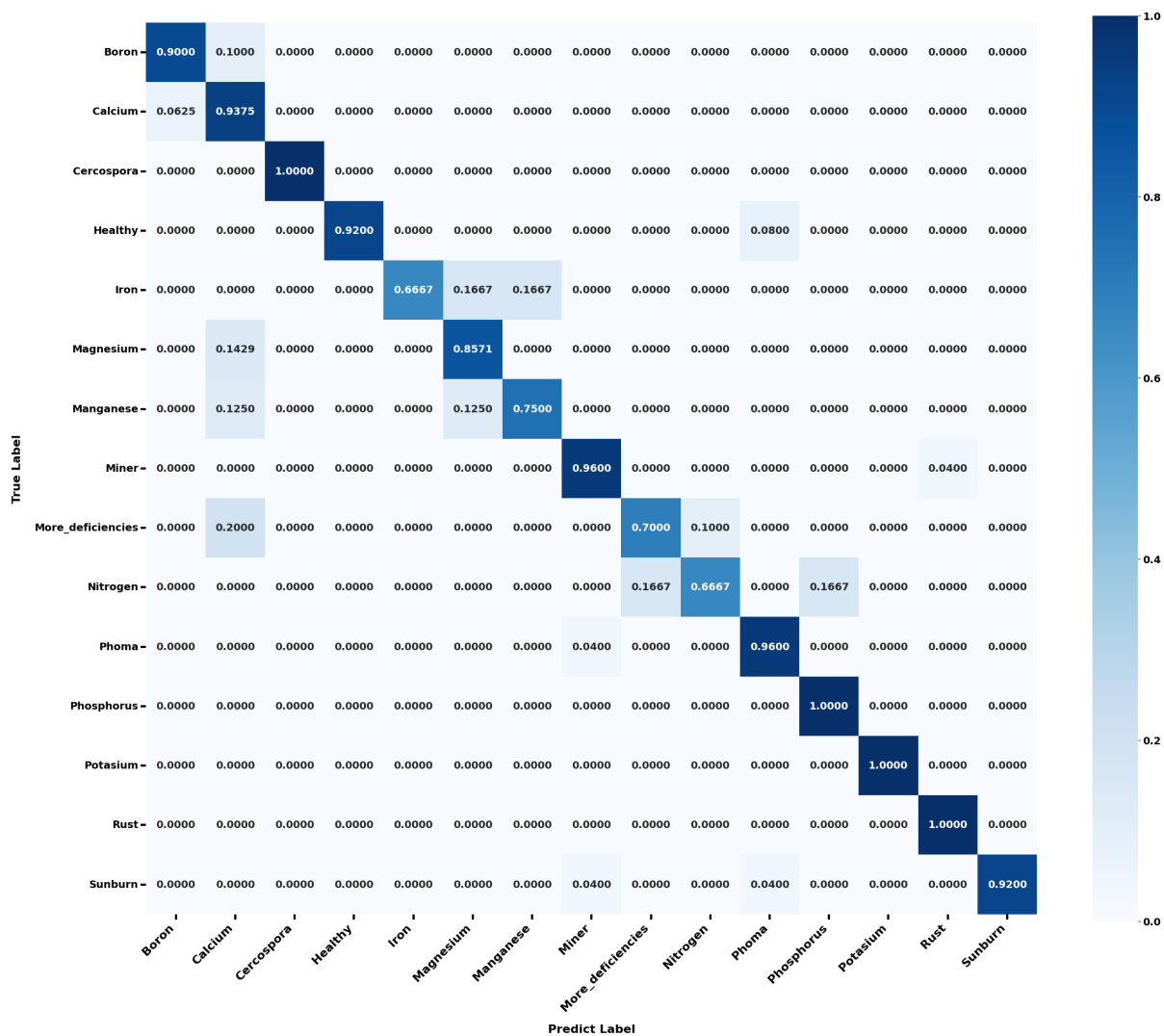


Figure 1. Confusion matrix with the prediction results of the DenseNet201 model.

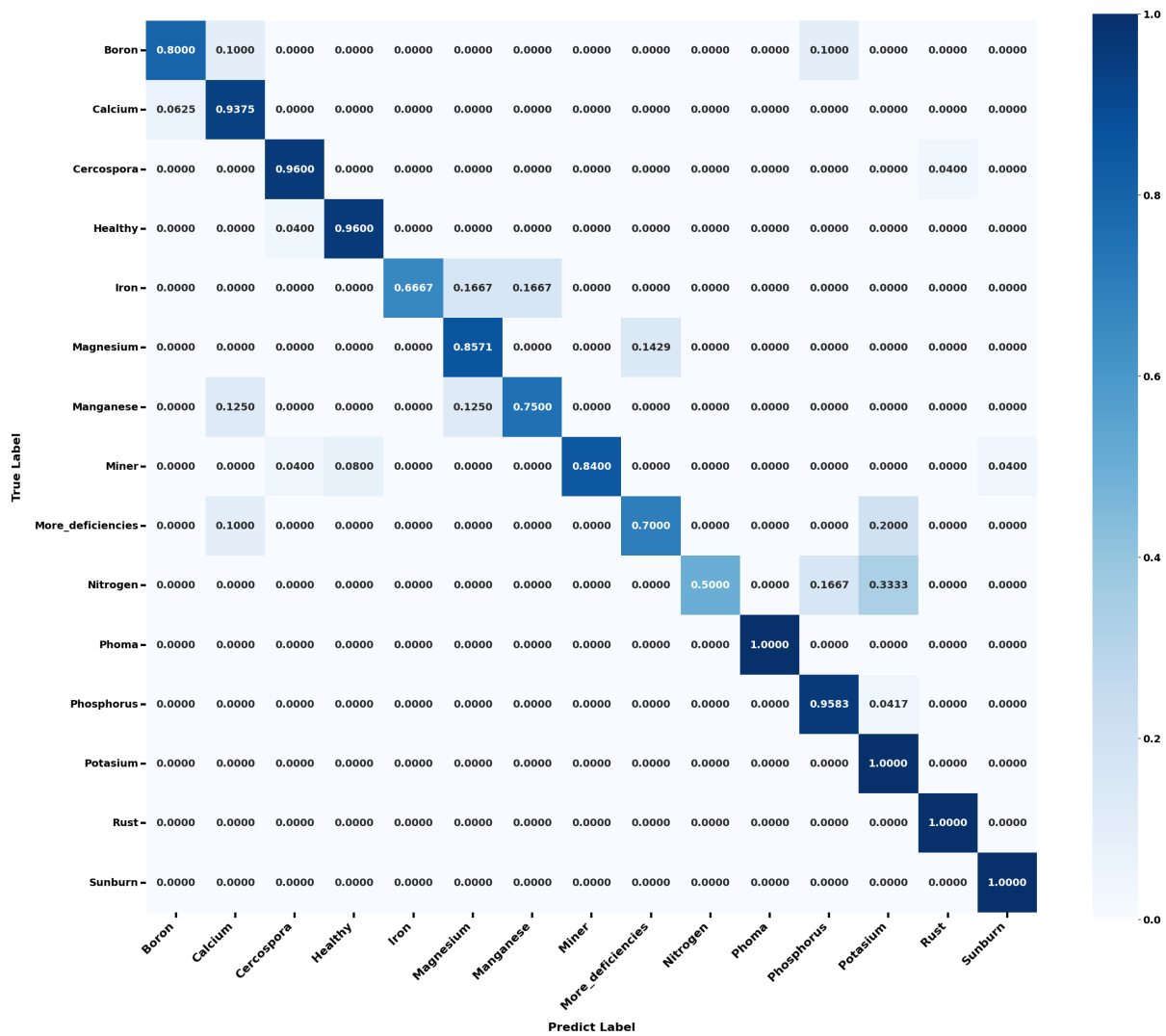


Figure 2. Confusion matrix with the prediction results of the VGG16 model.

The metric values obtained on the test set for each model are presented in Table 3. The DenseNet201 model showed the best performance, with an accuracy of 0.9268, precision of 0.9110, recall of 0.8825, and F1-score of 0.8920. The VGG16 model showed an accuracy of 0.9146, precision of 0.9024, recall of 0.8620, and F1-score of 0.8698. Thus, the DenseNet201 model was superior to VGG16 across all adopted metrics.

Model	Accuracy	Precision	Recall	F1-score
DenseNet201	0.9268	0.9110	0.8825	0.8920
VGG16	0.9146	0.9024	0.8620	0.8698

The per-class metrics of the DenseNet201 model are presented in Table 4. The model correctly classified all images of the Cercospora and Potassium classes, with an F1-score of 1.0000. The Phosphorus, Rust, Healthy, and Sunburn classes also showed high F1-Score, all above 0.95. On the other hand, the Nitrogen and More Deficiencies classes showed the lowest F1-score values, of 0.7273 and 0.7778, respectively. These two classes are among the fewest in the training set, with 55 and 89 images, which may explain their lower performance. A similar result was observed for the Iron class, which showed a recall of 0.6667 and is also among the classes with the fewest images.

Class	Precision	Recall	F1-score
Boron	0.9000	0.9000	0.9000
Calcium	0.7500	0.9375	0.8333
Cercospora	1.0000	1.0000	1.0000
Healthy	1.0000	0.9200	0.9583
Iron	1.0000	0.6667	0.8000
Magnesium	0.7500	0.8571	0.8000
Manganese	0.8571	0.7500	0.8000
Miner	0.9231	0.9600	0.9412
More Deficiencies	0.8750	0.7000	0.7778
Nitrogen	0.8000	0.6667	0.7273
Phoma	0.8889	0.9600	0.9231
Phosphorus	0.9600	1.0000	0.9796
Potassium	1.0000	1.0000	1.0000
Rust	0.9615	1.0000	0.9804
Sunburn	1.0000	0.9200	0.9583

Overall, the DenseNet201 model outperformed VGG16 in classifying biotic and abiotic stresses in coffee on the test set. For this reason, the DenseNet201 model was selected for the web application.

3.3 Web Application Implementation

The web application uses the DenseNet201 model and runs in a browser with internet access. After capturing an image of a crop leaf, the user uploads it to the website and triggers the classification, as shown in Figure 3.

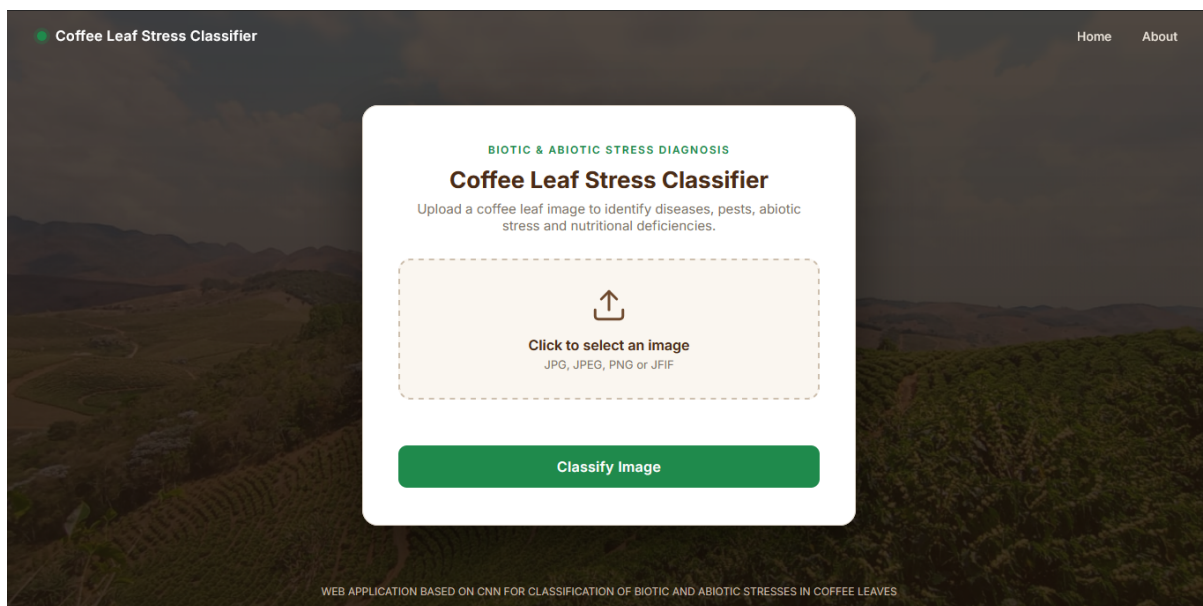


Figure 3. View of the home page of the web application.

To verify its operation, an image obtained from an internet search was used, belonging to the Miner class and not part of the training dataset. After the upload and classification trigger, the application returned the probabilities per class and displayed the results as shown in Figure 4. For the tested image, the application correctly classified the leaf as belonging to the Miner class, with a probability of 96.75%.

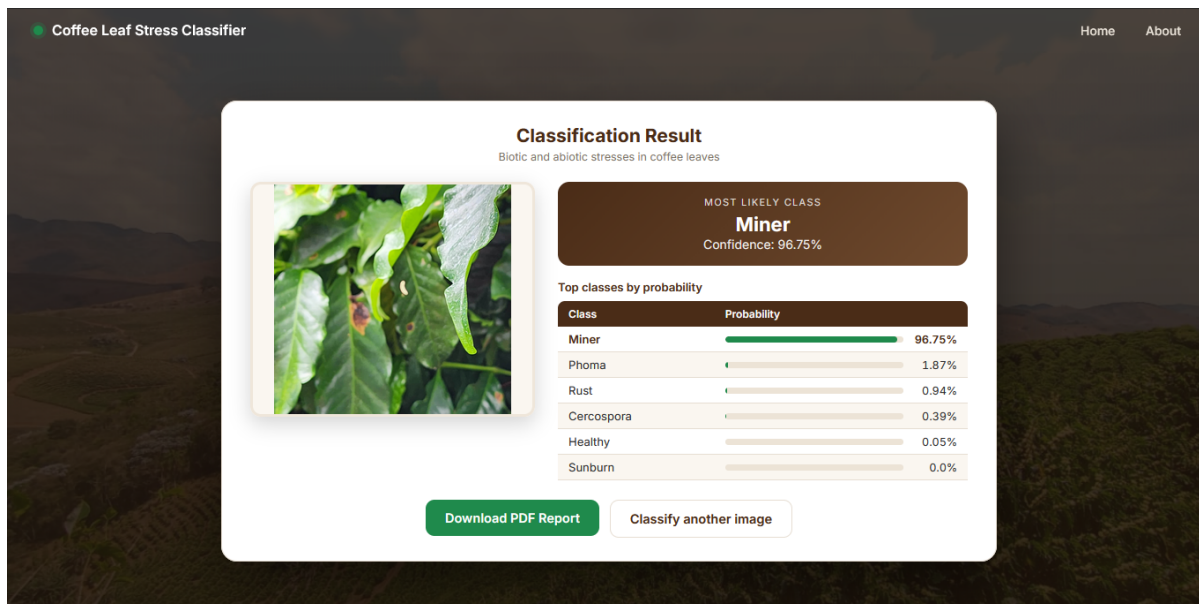


Figure 4. Prediction result for an image of the Miner class by the web application.

Finally, the application generates a PDF report that combines the analyzed image, the identified class, the associated probability, and a descriptive text corresponding to the class, as shown in Figure 5. In this way, the application delivers to the user, in addition to the classification, supporting information about the stress identified on the leaf.

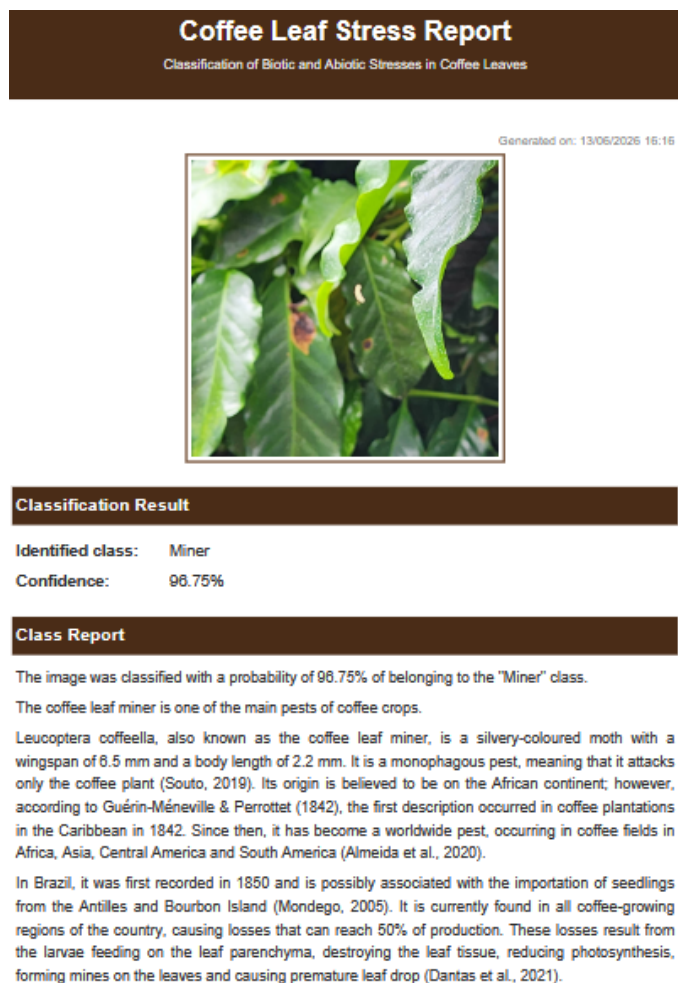


Figure 5. PDF report generated by the web application for the classified image.

4 CONCLUSION

This study evaluated two convolutional neural network architectures, DenseNet201 and VGG16, with Transfer Learning, for the classification of 15 classes of biotic and abiotic stresses in coffee leaves. On the test set, the DenseNet201 model achieved the best performance, with an accuracy of 0.9268 and an F1-score of 0.8920, outperforming VGG16 across all metrics. Based on this result, DenseNet201 was integrated into a web application developed in Flask that allows coffee growers, researchers, and other professionals to classify leaf symptoms and generate a PDF report to support decision-making. The main limitations relate to the small number of images in some classes, such as Nitrogen, More Deficiencies, and Iron, which performed worst. As future work, it is suggested to increase the quantity and diversity of the images, especially in the minority classes, and to evaluate other architectures and class-balancing strategies.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-Brasil (CAPES)-Finance Code 001, FAPEMIG (Research Support Foundation of Minas Gerais State, Funding Codes PPE-00047-21 and APQ-00750-23), and the National Council for Scientific and Technological Development (CNPq). We sincerely thank these institutions for their financial support to conduct this study.

REFERENCES

- Andrade, T., Silva, R. A., Matos, C. de S. M. de, Volpato, M. M. L., Pereira, A. B., & Ferreira, D. D. (2020). Predição de pragas e doenças no cafeeiro utilizando Redes Neurais Artificiais. *Congresso Brasileiro de Automática - CBA*, 2(1). <https://doi.org/10.48011/ASBA.V2I1.1286>
- Andrade, Vicentin, R. P., Costa, J. R., Perina, F. J., Resende, M. L. V. De, & Alves, E. (2016). Alterations in antioxidant metabolism in coffee leaves infected by *Cercospora coffeicola*. *Ciência Rural*, 46(10), 1764–1770. <https://doi.org/10.1590/0103-8478CR20150938>
- Borgo, L., Rabêlo, F. H. S., Marchiori, P. E. R., Guilherme, L. R. G., Guerra-Guimarães, L., & Resende, M. L. V. de. (2024). Impact of Drought, Heat, Excess Light, and Salinity on Coffee Production: Strategies for Mitigating Stress Through Plant Breeding and Nutrition. *Agriculture* 2025, Vol. 15, Page 9, 15(1), 9. <https://doi.org/10.3390/AGRICULTURE15010009>
- Catarino, A. de M., Pozza, E. A., Pozza, A. A. A., Santos, L. S. dias, Vasco, G. B., & Souza, P. E. de. (2016). Calcium and potassium contents in nutrient solution on Phoma leaf spot intensity in coffee seedlings. *Revista Ceres*, 63(4), 486–491. <https://doi.org/10.1590/0034-737X201663040008>
- Cerda, R., Avelino, J., Gary, C., Tixier, P., Lechevallier, E., & Allinne, C. (2017). Primary and Secondary Yield Losses Caused by Pests and Diseases: Assessment and Modeling in Coffee. *PLOS ONE*, 12(1), e0169133. <https://doi.org/10.1371/JOURNAL.PONE.0169133>
- Dantas, J., Motta, I. O., Vidal, L. A., Nascimento, E. F. M. B., Bilio, J., Pupe, J. M., et al. (2021). A comprehensive review of the coffee leaf miner leucoptera coffeella (Lepidoptera: Lyonetiidae)—a major pest for the coffee crop in brazil and others neotropical countries. *Insects*, 12(12), 1130. <https://doi.org/10.3390/INSECTS12121130/S1>
- Esgario, J. G. M., Krohling, R. A., & Ventura, J. A. (2020). Deep learning for classification and severity estimation of coffee leaf biotic stress. *Computers and Electronics in Agriculture*, 169, 105162. <https://doi.org/10.1016/J.COMPAG.2019.105162>
- Ferreira, T. L. (2023). Cafés do Brasil tem valor bruto da produção total estimado em R\$ 48,8 bilhões no ano-cafeeiro 2023. *A Embrapa - Portal Embrapa*.
- Hassan, S. M., Maji, A. K., Jasiński, M., Leonowicz, Z., & Jasińska, E. (2021). Identification of [Proceedings of the 17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture: 13-16 July, 2026, Porto Alegre, Brazil](#)

- Plant-Leaf Diseases Using CNN and Transfer-Learning Approach. *Electronics* 2021, Vol. 10, Page 1388, 10(12), 1388. <https://doi.org/10.3390/ELECTRONICS10121388>
- Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. <https://github.com/liuzhuang13/DenseNet>. Accessed 14 June 2026
- Leite, M. da S. (2018). IDENTIFICAÇÃO DE SINTOMAS DE DEFICIÊNCIAS MINERAIS E PRAGAS NA CULTURA CAFEEIRA UTILIZANDO VISÃO COMPUTACIONAL. <https://www.eng-automacao.araxa.cefetmg.br/wp-content/uploads/sites/152/2019/02/TCC-MARLY-LEITE.pdf>. Accessed 13 June 2026
- Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. <https://arxiv.org/pdf/1409.1556>. Accessed 14 June 2026
- Sorte, L. X. B., Saito, J. H., & Goulart, R. dos R. (2018). Detecção de Doenças do Cafeeiro com Deep Learning e Caracterização de Textura GLCM. *Anais do WCF*, 5. Accessed 13 June 2026
- Tuesta-Monteza, V. A., Mejia-Cabrera, H. I., & Arcila-Diaz, J. (2023). CoLeaf-DB: Peruvian coffee leaf images dataset for coffee leaf nutritional deficiencies detection and classification. *Data in Brief*, 48, 109226. <https://doi.org/10.1016/J.DIB.2023.109226>
- Yang, W., Yang, C., Hao, Z., Xie, C., & Li, M. (2019). Diagnosis of Plant Cold Damage Based on Hyperspectral Imaging and Convolutional Neural Network. *IEEE Access*, 7, 118239–118248. <https://doi.org/10.1109/ACCESS.2019.2936892>