Deep learning for classification and severity estimation of coffee leaf biotic stress

José G.M. Esgario[a,b], Renato A. Krohling[a,b,c,*], José A. Ventura[d]

[a] Federal University of Espírito Santo, Brazil
[b] PPGI – Graduate Program in Computer Science, Brazil
[c] Production Engineering Department, Brazil
[d] Incaper, Rua Afonso Sarlo, 160, Bento Ferreira, 29052-010 Vitória, ES, Brazil

ABSTRACT

Biotic stress consists of damage to plants through other living organisms. The efficient control of biotic agents such as pests and pathogens (viruses, fungi, bacteria, etc.) is closely related to the concept of agricultural sustainability. Agricultural sustainability promotes the development of new technologies that allow the reduction of environmental impacts, greater accessibility to farmers and, consequently, increased productivity. The use of computer vision with deep learning methods allows the early and correct identification of the stress-causing agent. So, corrective measures can be applied as soon as possible to mitigate the problem. The objective of this work is to design an effective and practical system capable of identifying and estimating the stress severity caused by biotic agents on coffee leaves. The proposed approach consists of a multi-task system based on convolutional neural networks. In addition, we have explored the use of data augmentation techniques to make the system more robust and accurate. Computational experiments performed with the proposed system using the ResNet50 architecture obtained an accuracy of 95.24% for the biotic stress classification and 86.51% for severity estimation. Moreover, it was found that by classifying only the symptoms, the results were greater than 97%. The experimental results indicate that the proposed system might be a suitable tool to assist both experts and farmers in the identification and quantification of biotic stresses in coffee plantations.

1. Introduction

Among the sectors of the Brazilian economy, agriculture is one of the main pillars, contributing to employment generation, income and wealth for the country (Oliveira et al., 2014). According to International Coffee Organization (2019), Brazil is the world's largest producer of coffee and this is an important crop for the country. Biotic stress such as leaf miner, rust, brown leaf spot, and cercospora leaf spot, affects coffee plantations leading to the defoliation and reduction of photosynthesis, hence reducing the yield and quality of the final product (Ventura et al., 2017). According to the Incaper (2019), the state of Espírito Santo is the 2nd largest coffee producer in the country. Its expressive production of both arabica and conilon coffee varieties make it responsible for 22% from national production. Among the coffee producers, 73% are family-based.

For the efficient control of pests and diseases, it is important to know not only the causal agent but also the severity of the symptoms, since the diagnosis and quantification of plant stress are two equally important functions for phytopathology (Kranz, 1988). Plant disease severity measures the percentage of the plant tissue area that is symptomatic and is important to predict yield and recommend control treatments (Bock et al., 2010).

The task of identifying plant diseases is as important as it is challenging even for us humans with our remarkable ability to recognize and interpret patterns (Barbedo et al., 2016). The visual evaluation of symptoms in plant leaves is usually performed by specialists and has a high subjective character due to the similarity of the diseases and the diversity of variations in their characteristics that are influenced by external factors. Many farmers do not have access to specialists, which makes it harder to correctly identify plant diseases. As a consequence, the control performed tends to be inefficient and may lead to increased risk of crop yield losses.

Several efforts have been made using Artificial Intelligence to assist small farmers to correctly identify the diseases and pests that affect their production and the severity of the symptoms. Computer-aided diagnosis (CAD) systems allow any farmer with access to a smartphone to enjoy expert knowledge in a practical and low-cost way (Dehnen-Schmutz et al., 2016).

The area of Artificial Intelligence research has shown significant growth in the few years concerning the applications of Machine Learning, which led to the emergence of a new category of models called Deep Learning (LeCun et al., 2015). Deep Learning methods have shown state-of-the-art performance when compared to traditional Machine Learning approaches in different domains (Alom et al., 2019).

Among the methods of Deep Learning, Convolutional Neural Networks (CNN) have shown outstanding performance in image recognition tasks (Fuentes et al., 2017). CNN automatically learns the appropriate features from the training dataset while traditional approaches are based on handcrafted features, i.e., the features are calculated based on a priori knowledge of the problem. In addition, the segmentation step is intrinsic to the CNN convolutional filters, further simplifying its use.

According to Barbedo (2018b) CNNs are powerful tools that deal pretty well with the problem of plant diseases. However, there are still many challenges associated with the automatic diagnosis of plant diseases. In Barbedo (2016) the six main challenges encountered in this type of problem are addressed whereas these challenges were separated into extrinsic and intrinsic factors. Extrinsic factors are mainly related to the image acquisition in-field, i.e.: (1) Complex backgrounds, for example, leaves, soil, stones, etc. (2) Uncontrolled image capture conditions, e.g., variation of lighting, shadows, reflections, and blur. Intrinsic factors: (3) Symptoms boundary poorly defined presents a gradual fading between injury and healthy tissue. (4) Symptom variations for different stages of the disease such that the interaction between plant, disease, and environment leads to variations in their visual characteristics. (5) Multiple simultaneous disorders of biotic and abiotic origin may manifest on the same plant and even on the same leaf. (6) Similarities in the appearance of the symptoms caused by different types of stress.

To address the challenges pointed out by Barbedo (2016) new methods need to tackle issues like performance and reliability (Singh et al., 2018). Challenges (4) and (5) depend heavily on the construction of more representative and reliable datasets. According to Barbedo (2019), this is the main issue encountered by deep learning approaches in the automatic identification of plant diseases. The creation of these datasets is not a simple task, labeling the images is a costly process and often must be carried out by a specialist. Although there are initiatives that are using the concept of social networks to accelerate this process, this practice can lead to unreliable data (Barbedo, 2018a).

The largest public dataset currently was developed by Hughes and Salathe (2015) called PlantVillage and contains over 54 thousand images of leaves and 38 classes, among them, healthy and diseased leaves of different species. Several studies were carried out using the PlantVillage dataset (Mohanty et al., 2016; Too et al., 2019; Kaya et al., 2019), most of these studies presented results greater than 99% of accuracy. Another public dataset is the Digipathos, which includes many images of diseases affecting coffee and other crops (Barbedo et al., 2018). Also, it is common in the literature works that developed their own dataset applied to a specific type of crop such as Fuentes et al. (2017) that collected images of tomato leaves using conventional cameras. Johannes et al. (2017) developed a system capable of identifying diseases in photos of wheat leaves obtained by smartphones. Liu et al. (2017) created a dataset of apple leaves and proposed a new architecture based on AlexNet for the recognition of diseases and Ma et al. (2018) developed a dataset with images of common disease symptoms affecting cucumber leaves.

Although the problem of plant leaf disease has been addressed in several studies, few have focused on developing systems capable of estimating stress severity. Wang et al. (2017) proposed the use of Convolutional Networks to estimate the severity of plant diseases. The images of apple leaves of the dataset affected by black rot were labeled in four degrees of severity. The experimental results showed an accuracy greater than 90%.

Typically in Machine Learning, we focus on solving a single problem, such as diagnosing leaf diseases or estimating the stress severity. To do this, usually, a model or a set of models are trained to solve specific tasks. On the other hand, there is the concept of multi-Task Learning (MTL), which consists of training a system capable of solving multiple tasks using a shared architecture. MTL leads the model to prefer representations that serve for both tasks, in such a way that the initial layers of the network will learn joint generalized representations

allowing a better generalization ability of the model, preventing overfitting (Ruder, 2017).

Liang et al. (2019) proposed a multitasking CAD system able to diagnose diseases, recognizing the plant species and estimating the severity of diseases, called PD^2SE-Net. The PlantVillage dataset was used to perform the experiments. The estimation of stress severity consisted of classifying the leaves in one out of three classes: healthy, general and serious. The results presented an overall accuracy of 91% and 98% for disease severity estimation and plant disease classification, respectively.

Ghosal et al. (2018) developed a framework based on CNN able to identify and classify a set of 8 stresses (biotic and abiotic) affecting soybean leaves. In addition, they presented a mechanism for selecting and extracting the feature maps that best isolate visual symptoms in a similar way to that performed by humans and they had used such maps to quantify stress severity.

In a recent study, Barbedo (2019) explored the use of individual lesions instead of considering the entire leaf. The lesions were manually segmented, this allowed to significantly increase the dataset and to identify multiple and different lesions that affect the same leaf. However manual segmentation is a laborious process mainly in leaves with many small lesions.

Manso et al. (2019) developed a system capable of classifying the individual symptoms of coffee leaves whereas the segmentation of the symptoms was performed using a threshold-based method. The symptoms were classified into one out of two classes, leaf miner and rust, using handcrafted features. With the lesion segmentation results, it was possible to quantify the severity and to identify each symptom individually. Although the work presented good results, the segmentation with this type of approach does not take into account the positional relation between the pixels, being easily affected by factors such as illumination variation and specular reflection.

Few works in the literature have focused on quantifying the disease's severity such as Wang et al. (2017) and Liang et al. (2019). However, in these studies, subjective labels were attributed to the disease's severity without a concrete relationship with the severity percentages. In addition, most papers do not deal directly with individual symptoms, limiting them to detecting only one stress per leaf. In this work, we have significantly extended the work of Manso et al. (2019) by substantially increasing the dataset of coffee leaves and replacing conventional image processing methods with deep learning-based methods. The main contributions of this work are: 1) A larger and more representative image dataset of healthy and diseased coffee leaves has been built and will be made publicly available; 2) A multi-task system is proposed to classify and estimate the severity of biotic stress affecting coffee leaves; 3) Comparative studies of different deep learning architectures are performed.

The remainder of this paper is organized as follows. Section 2 is devoted to describing the details of the dataset and explains the proposed system architecture. Section 3 presents the experimental results and our findings. Session 4 discusses the important aspects of this paper. Section 5 draws some conclusions and suggests directions for future work.

2. Materials and methods

2.1. Image dataset

The dataset developed for this work contains images of arabica coffee leaves affected by the main biotic stresses that affect the coffee tree. The images were obtained using different smartphones (ASUS Zenfone 2, Xiaomi Redmi 5A, Xiaomi S2, Galaxy S8, and iPhone 6S). In addition, the leaves were collected at different times of the year in Marechal Floriano county, state of Espírito Santo, Brazil. The photos were taken from the abaxial (lower) side of the leaves under partially controlled conditions and placed on a white background. Besides that,

in-field images were not used due to occlusion of leaf parts and their inclination in relation to the camera, making the disease severity estimation infeasible in several cases. Also, the acquisition of the images was done without any criteria to make the dataset more heterogeneous.

A total of 1747 images of arabica coffee leaves were collected, including healthy leaves and diseased leaves, affected by one or more of the following biotic stresses: leaf miner, rust, brown leaf spot, and cercospora leaf spot. The process of biotic stresses recognition for dataset labeling was assisted by an expert and performed with the captured images. From the obtained photos were generated two datasets. A dataset with the original images of the entire leaves and a second one containing only symptoms images. Details of each dataset are described in the following.

Leaf dataset: It consists of the original images of the entire leaves labeled with the predominant biotic stress and its severity. A total of 372 images showed leaves with more than one kind of stress, among them, 62 leaves presented stresses with similar severity. This work is limited to recognizing a single stress per leaf in such a way that the classification of multiple stresses is beyond the scope of the current research. Therefore, these 62 images with similar severity were not used on this dataset, as it is difficult to visually distinguish which stress predominates. Stress severity was calculated using the symptom and leaf segmentation mask using automatic image processing methods presented in Manso et al. (2019). Segmentation results were visually validated for all images. The severity value for those that had poor segmentation was discarded and these images were analyzed separately by a specialist applying visual estimation methods. For certain severity ranges, labels were assigned as follows: healthy (<0.1%), very low (0.1% − 5%), low (5.1% − 10%), high (10.1% − 15%) and very high (>15%). Fig. 1 shows some examples of images contained in the dataset.

Symptom dataset: This dataset was created by cropping the isolated symptoms from the original images in a way that only single stress was present in each image. A total of 2147 symptom images were cropped. In addition to our images, 575 images made available by Barbedo (2019) were added to our dataset, accounting for 2722 symptom images. Fig. 2 presents some of the symptoms extracted from the original images.

Table 1 presents the number of images for each stress/leaf and severity/leaf pair of the above datasets.

2.2. Data augmentation

Complex models tend to suffer overfitting when trained with small datasets. To overcome this problem, data augmentation techniques are used to generate synthetic samples of the training data in order to increase the generalization ability of the model. In this work we used standard techniques in the literature (Inoue, 2018), we call it standard augmentation, and we explored the use of a new approach that aims to combine two randomly selected images generating a new one. Details of the used techniques are presented in the following.

- Standard augmentation: The use of standard data augmentation techniques are quite common in most works in the literature and present good performance gains. Although there are several ways to generate new images from the training samples, we focus only on the most common ones, being: horizontal and vertical mirroring, rotation and color variation (brightness, contrast, and saturation). Standard augmentation was applied to all experiments.
- Mixup: This approach was proposed by Zhang et al. (2017) consists of a simple linear combination of two images and their labels. Therefore, a new image is generated by $\{\tilde{x}, \tilde{y}\} = \lambda\{x_i, y_i\} + (1 − \lambda)\{x_j, y_j\}$ where $\lambda \in [0, 1]$ is a random value generated by the beta probability distribution, x represents the set of images and y the set of labels encoded in the one-hot format.

2.3. Deep learning architectures

Several CNN architectures are being proposed each year, each with its own particular characteristics. Despite the differences, they all share the same goal that is to increase accuracy and reduce the model complexity. Some architectures provide great performances on a wide range of applications. We selected some of the most common architectures used in the classification problem of plant diseases. Some characteristics of the used architectures are presented in Table 2.

The Leaf dataset has two tasks, biotic stress classification, and severity estimation. Since these tasks are closely related (problems of the same domain), we propose in this work the adaptation of CNN networks making them multi-task systems. Therefore, the architectures were modified by the addition of a new fully connected layer in parallel with the existing one. In this way, the classification blocks are individualized but the convolutional layers are shared, i.e., the model will learn joint features useful in classifying both problems. Fig. 3 presents the basic structure of systems with single-task and multi-task learning.

One can notice that multi-task learning reuses most of the network architecture. This simplification makes learning substantially faster since only a single model needs to be trained.

2.4. Experimental setup

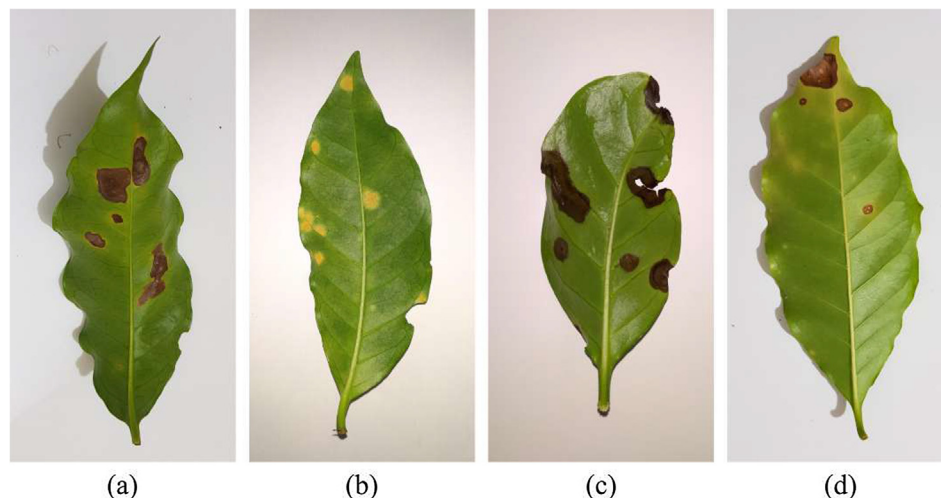To meet the input size requirements of CNN networks, images need

Fig. 1. Examples of coffee leaves affected by different biotic stresses: leaf miner (a), rust (b), brown leaft spot (c) and cercospora leaf spot (d).

Fig. 2. Examples of symptom images.

Table 1
Datasets details.

Biotic stress	Leaf dataset	Symptom dataset	Severity	Leaf dataset
Healthy	272	256	Healthy	272
Leaf miner	387	593	Very low	924
Rust	531	991	Low	332
Brown leaf spot	348	504	High	101
Cercospora leaf spot	147	378	Very high	56
Total	1685	2722	Total	1685

Table 2
Properties of the CNN architectures used in this work.

CNN architecture	Parameters (M)	Layers
AlexNet	61	8
GoogLeNet	6.9	22
VGG16	138	16
ResNet50	25	50
MobileNetV2	2.2	54

to be resized to $224 \times 224 \times 3$. Due to the size of the images, the network may find difficult to capture relevant characteristics in the dataset of leaves with very small symptoms. To alleviate this problem, a traditional thresholding method based on fixed threshold was used in the *S* channel of the *HSV* color space, in such a way that the leaf segmentation allows us to fit a bounding box for the removal of excesses that don't belong to the region of interest. Therefore, the images were cropped and resized allowing the leaf to cover much of the image. Fig. 4 shows the steps performed to crop and resize the leaves.

For the realization of all the experiments were used the proportions 70-15-15 for the training, validation and test datasets, respectively. Data augmentation techniques were applied online during training. Each new batch of images produces new images that are inputted into the CNN network and the network weights are adjusted until the network learns the most relevant discriminative features for a given problem. To make the training more efficient, transfer learning technique

was applied. The training was performed by adapting pre-trained networks with the ImageNet dataset. The models were trained end-to-end, without freezing the training of any layer. The hyper-parameters used during the models training are presented in Table 3.

During the training of the networks, the states (set of weights) in which the models presented the lowest loss value for the validation set were saved. The saved models were then evaluated with the test dataset and the results were computed in terms of Accuracy (acc), Precision (prc) and Recall (rec). All experiments were performed using PyTorch (Machine learning open source library) with an NVIDIA GeForce GTX 1060 and CUDA 10.0. The source code and the datasets are available at: https://github.com/esgario/lara2018.

3. Results

Experiments were conducted in order to evaluate how appropriate the models are for the problems of biotic stress classification and severity estimation. The results were divided by dataset (leaf and symptoms) and are presented in the following subsections.

3.1. Leaf dataset results

The obtained results for the Leaf dataset are presented in Table 4. The results were divided by task (biotic stress and severity), CNN architecture (AlexNet, GoogLeNet, VGG16, ResNet50 and MobileNetV2) and learning approach (single-task and multi-task). The best results are presented in bold.

The results show that ResNet50 architecture achieved an accuracy of 95.24% for biotic stress classification and 86.51% for severity estimation using multi-task learning. With the single-task learning, the obtained results for the same architecture were 95.63% and 84.13%. By comparing the individual results of each architecture, we notice that AlexNet, GoogLeNet, and ResNet50 have been benefited from multi-task learning. The addition of mixup in ResNet50 training presented similar results for the biotic stress classification problem, however, it worsened when applied to the severity estimation. The performance of the mixup is discussed in the following subsections. In addition to the
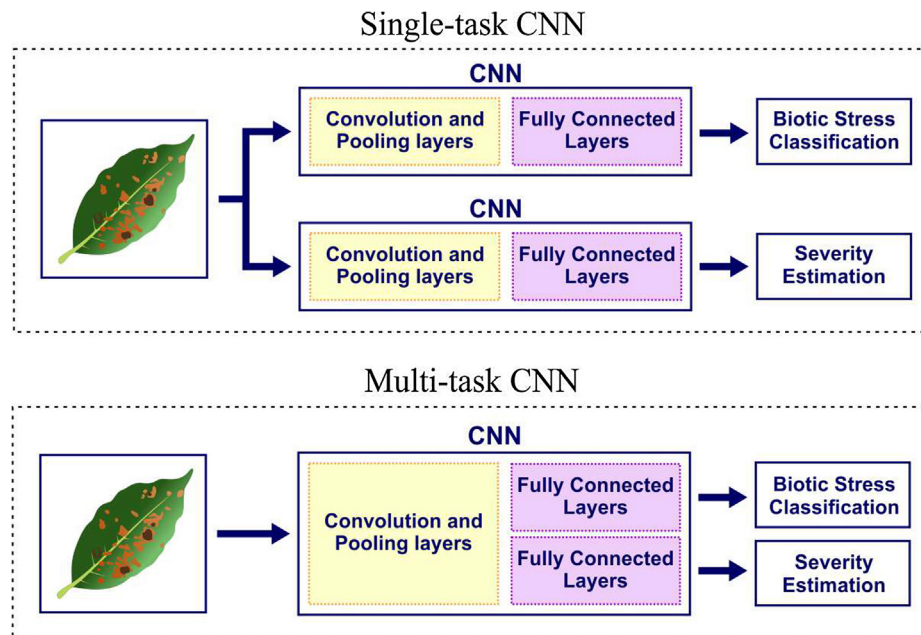
Single-task CNN

Multi-task CNN

Fig. 3. Framework used for single-task and multi-task learning.
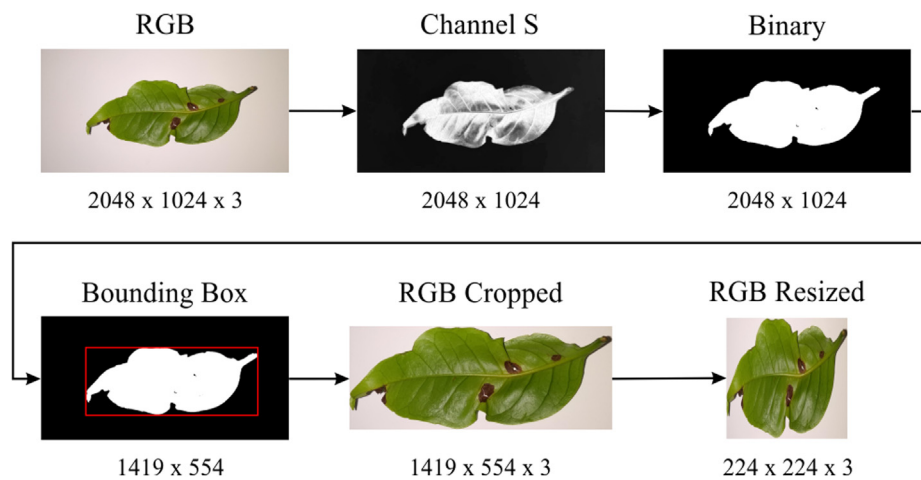
Fig. 4. Cropping and resizing leaf images.

Table 3
CNN training hyper-parameters.

Parameter	Value
Optimizer	Stochastic Gradient Descent
Loss function	Cross-Entropy
Epochs	100
Batch size	32
Learning rate[*]	0.01
Momentum	0.9
Weight decay	0.0005

[*] Decreases by a factor of 1/2 or 1/5 every 20 epochs, alternately.

performance results of the models, we present in Table 5 the average time in seconds for the training of each architecture.

Among the used architectures ResNet50 presented a tradeoff between computational cost, performance, and reliability superior to the competitors. Fig. 5 presents the confusion matrices associated with the prediction results obtained with ResNet50. The classification results of biotic stress were consistent for most of the stresses except cercospora leaf spot which presented a considerable amount of classification errors.

This result corroborates with the experiments carried out in Barbedo (2019) whose class with the largest number of samples misclassified was also the cercospora leaf spot. These misclassifications may be associated with similarity with other diseases and also with the dataset imbalance.

The severity estimation with the ResNet50 obtained an accuracy of 84.13%. This problem is clearly more challenging compared to biotic stress classification. The model presented no difficulty in separating healthy leaves from diseased, however, due to the imbalance of the data set, the classification of leaves with high and very high severity presented the highest misclassification rates. Although the model has made a considerable amount of mistakes, we notice by the confusion matrix that these errors are located close to the main diagonal, so most errors may be considered mild. Due to the ordinal nature of severity labels, estimating the stress severity as high rather than very high, for example, can be considered a minor error for this problem.

Another architecture that stood out in these experiments was MobileNetV2 with an accuracy of 94.84% for biotic stress and 84.52% for severity estimation. Although its results are lower than ResNet50, its computational and disk storage requirements to train and store a model of this architecture is substantially lower than other investigated

Table 4
Test results obtained with different architectures for the Leaf dataset.

Task	Architecture	Single-task			Multi-task		
		acc(%)	prec(%)	rec(%)	acc(%)	prec(%)	rec(%)
BioticStress	AlexNet	92.46	89.37	90.02	91.67	88.14	88.67
	GoogLeNet	91.67	88.33	89.38	94.05	92.56	90.87
	VGG16	95.47	**96.73**	91.94	94.44	94.62	89.55
	ResNet50	**95.63**	94.12	**92.7**	95.24	**95.29**	**91.14**
	ResNet50*	**95.63**	95.79	92.32	94.84	92.80	91.48
	MobileNetV2	94.84	93.24	91.79	94.44	92.77	90.84
Severity	AlexNet	84.13	74.23	72.86	**86.9**	80.88	77.57
	GoogLeNet	82.94	74.27	73.76	82.94	75.39	73.76
	VGG16	**86.51**	**82.49**	**80.89**	86.51	79.50	76.31
	ResNet50	84.13	81.66	78.9	86.51	**82.38**	**80.9**
	ResNet50*	78.57	69.33	67.42	80.95	78.17	68.97
	MobileNetV2	84.52	73.29	77.36	84.13	77.15	79.53

* Model trained with standard augmentation and mixup.

Table 5
Average training time for one epoch for the architectures investigated.

Architecture	Time(sec/epoch)
AlexNet	7.32
MobileNetV2	11.14
GoogLeNet	11.76
ResNet50	21.90
VGG16	36.48

architectures, due to the small number of trainable parameters.

Table 6
Test results obtained with different architectures for the Symptom dataset.

Architecture	Biotic stress		
	acc(%)	prec(%)	rec(%)
AlexNet	96.58	96.12	96.59
GoogLeNet	96.82	96.56	96.64
VGG16	95.60	95.02	95.31
ResNet50	**97.07**	**96.85**	96.69
ResNet50*	**97.07**	**96.85**	**96.99**
MobileNetV2	96.33	95.65	96.41

* Trained model with Standard aug. and mixup.

3.2. Symptom dataset results

The experiments performed with the Symptom dataset achieved 97.07% accuracy with the ResNet50 architecture, as presented in Table 6. As expected, the classification results only of the symptoms are superior as compared to the classification using images of the entire leaves. This occurs because the symptoms images have a well-delimited region of interest which leads to the elimination of spurious elements. So, the network can then focus on and visualize a greater level of symptoms details. The ResNet50 architecture once again provided the best results while VGG16 showed no significant improvement when located symptoms were considered.

The addition of mixup to ResNet50 training presented a small gain of recall compared to the model trained only with standard augmentations. Although performance can be considered the same, it was realized that the use of this technique reduces the risk of overfitting during the training stage, due to the great diversity of new images that are generated throughout the learning process. Fig. 6 presents the confusion matrices with the prediction results obtained by ResNet50 trained with and without mixup. The higher recall value reflects better balanced results between classes. Therefore, the mixup was able to deal better with imbalanced data.

3.3. Visualization

The t-Distributed Stochastic Neighbor Embedding (t-SNE)
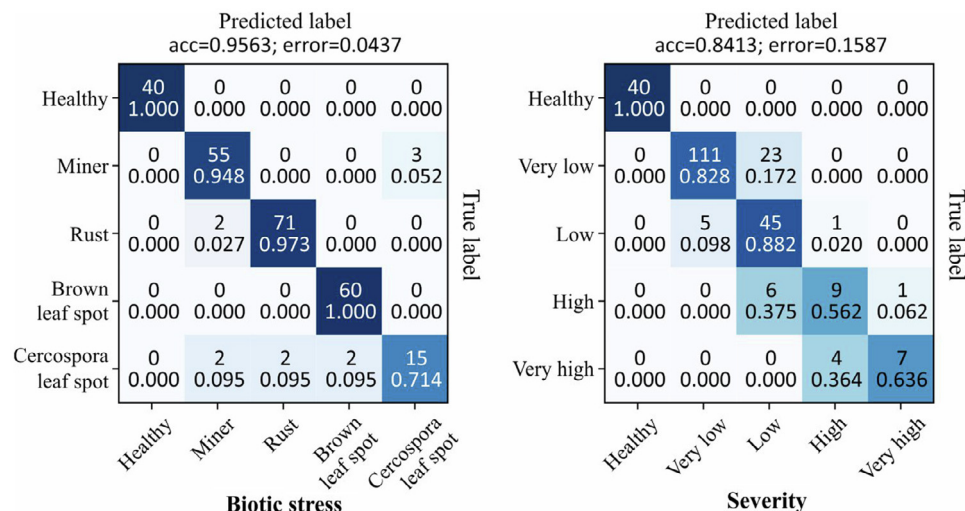
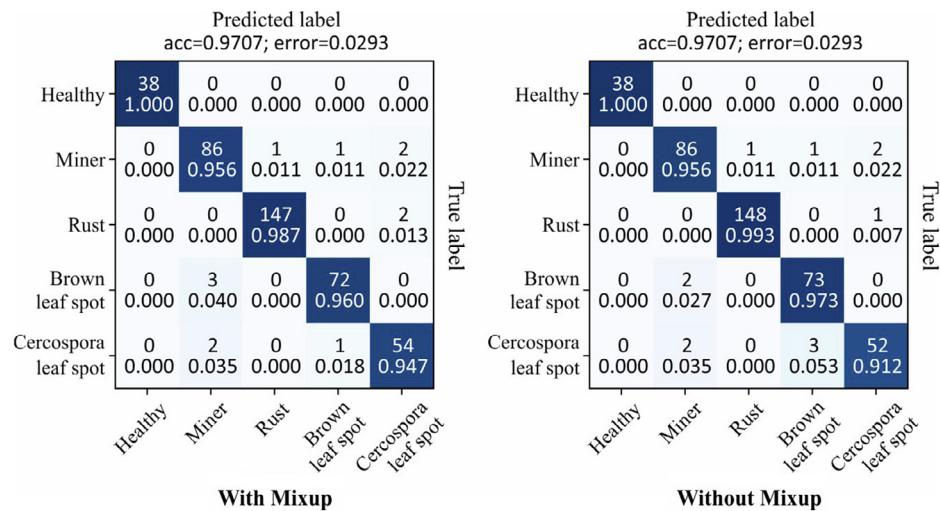**Fig. 5.** Confusion matrix for the ResNet50 prediction results using the Leaf Dataset.

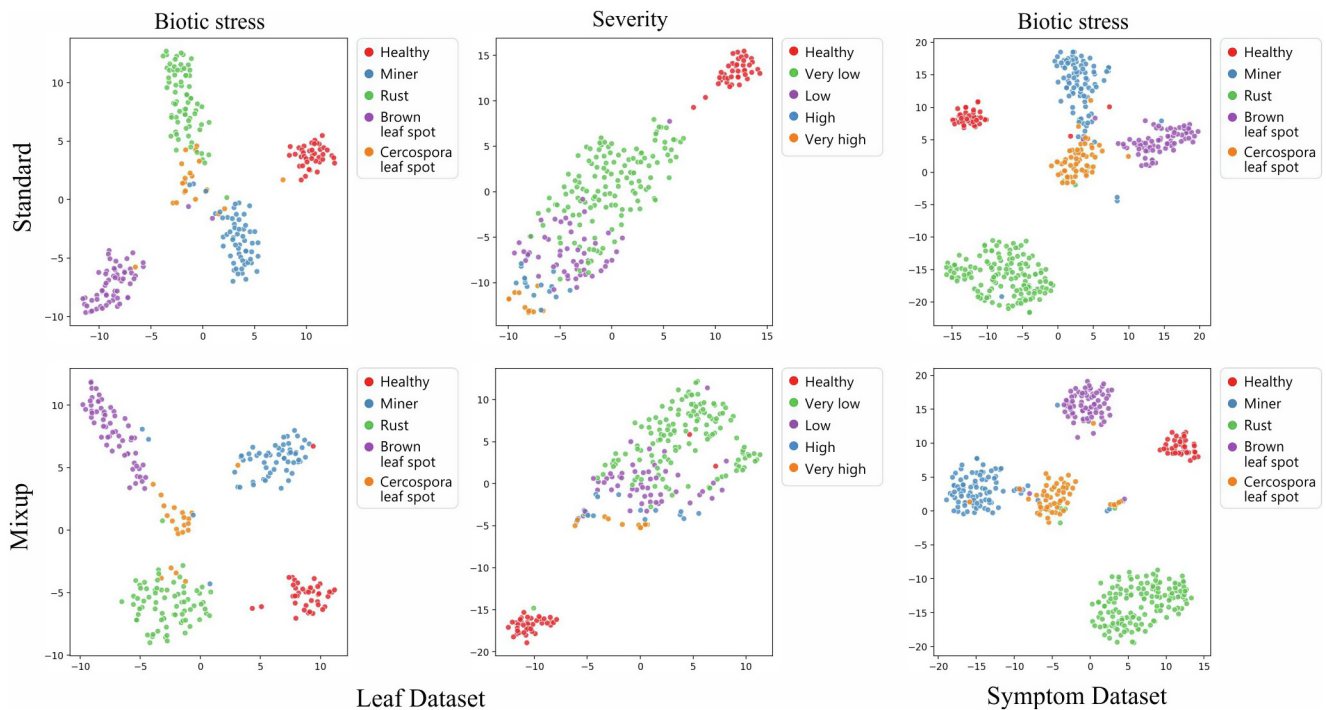**Fig. 6.** Confusion matrix for the ResNet50 prediction results using the Symptom Dataset.

Fig. 7. Visualization of the features spatial distribution extracted by ResNet50 from the test samples using t-SNE.

developed by Maaten and Hinton (2008) was used in the visualization of the high-dimensional features extracted by CNN as depicted in Fig. 7. In total, 2048 features were extracted from each test sample after the last ResNet50 convolutional layer. These features were reduced to two dimensions using t-SNE.

By visualizing the extracted features for the problem of biotic stress, we note that the cercospora leaf spot presented the highest overlap with other classes, showing the difficulty of the model to discriminate this class. Nevertheless, the model presented well-defined clusters of each class, showing that the model has a high generalization ability. Also, the severity estimation with standard augmentation showed a very linear relationship between the labels and their severity range.

The features learned with mixup presented a smaller within-class variance when compared to standard augmentation. Especially for the Symptom dataset where it is clear how the features learned by the mixup have a more spherical format. Although the mixup has presented

an appropriate behavior to the classification problem of biotic stress the technique worsened the classification results when applied to the severity estimation. The analysis of the results suggests that due to the ordinal nature of the severity labels, the combination of images with the mixup produced images with characteristics of median severity, which biased the network to learn more intermediate labels than extreme labels, causing a higher overlapping of the features.

4. Discussion

The experimental results obtained are not directly comparable with other methods in the literature. However, there is some consistency concerning the results obtained with those of similar works. For example, Liang et al. (2019) achieved an accuracy of 91% for disease severity estimation using three classes (healthy, general and serious) whereas we achieved 86.51% of accuracy when treating the problem

with five classes. Regarding the classification of biotic stress, the most similar work is that of Barbedo (2019), who conducted experiments to classify six biotic stresses that affect coffee trees. The experiments performed by Barbedo (2019) achieved an overall accuracy of 88.8% and in our experiments, we achieved an accuracy of 97.07%. Although the results were quite superior, it should be taken into account that in our study only four biotic stresses were used in the classification experiments.

An important aspect of this work that should be considered is that the images used were captured under controlled conditions, which could be considered a limitation for the practical application of the proposed system. In contrast, photos in controlled environments provide much more reliability than in-field photos, because fewer external disturbances influence the photo quality. An appropriate practice for using the proposed approach would be to collect random leaf samples throughout the plantation and then individually analyze each leaf with the proposed method. With the results in hand, it would be possible to infer with the use of statistical methods the general severity of each disease that affects that crop, facilitating the decision-making process in the control of such biotic stresses.

The proposed method was trained with leaves of arabica coffee trees, however, the similarity of symptoms in both robusta and arabica are quite similar, which leads us to believe that the method will have a very similar behavior if applied to leaves of robusta coffee trees. The main difference is in the frequency of disease occurrence in each coffee variety. As the proposed method does not consider this type of information when making inference, this would not affect the system classification results.

Although the construction of the dataset has incorporated the most common biotic stresses with the greatest economic impact on coffee crop, it is still limited to only a small set of disorders. As a consequence, the system will make mistakes when trying to classify photos of leaves affected by different stresses. In addition to this limitation, it has been noted that the emergence of multiple stresses on a single leaf can significantly alter the characteristics of symptoms, especially when they overlap. These symptom overlaps are quite rare to capture and would likely cause the system to make misclassifications. One way to alleviate this problem would be to train a similarity-based architecture that can rule out those symptoms that are not similar to any of the stresses in the dataset by classifying them into a new class, for example the *others* class.

5. Conclusion

This work presented different approaches using Deep Learning for the problems of biotic stress classification and severity estimation of the most important coffee diseases and pests through leaf images. For the accomplishment of the experiments a new dataset of coffee leaves images was developed. Different CNN architectures were used in experiments, whereas the trained network ResNet50 was the one that obtained the best results. It was verified that the multi-task learning can make the model more effective to solve tasks of the same domain in both performance and computational cost. One limitation of this work is related to the low representativity of the dataset that covers only the main biotic stresses that affect coffee trees. However, the results using Deep Learning techniques have proven consistent, increasing the dataset with new kind of stresses and a higher number of images might help alleviate this problem. Since an App has been developed in our lab, we are embedding this new technology to be used in practice. The first tests have been carried out with promising results and will be reported in near future.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.compag.2019.105162.

References

Alom, M.Z., Taha, T.M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M.S., Hasan, M., Van Essen, B.C., Awwal, A.A., Asari, V.K., 2019. A state-of-the-art survey on deep learning theory and architectures. Electronics 8, 292.

Barbedo, J.G., 2018a. Factors influencing the use of deep learning for plant disease recognition. Biosyst. Eng. 172, 84–91.

Barbedo, J.G.A., 2016. A review on the main challenges in automatic plant disease identification based on visible range images. Biosyst. Eng. 144, 52–60.

Barbedo, J.G.A., 2018b. Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. Comput. Electron. Agric. 153, 46–53.

Barbedo, J.G.A., 2019. Plant disease identification from individual lesions and spots using deep learning. Biosyst. Eng. 180, 96–107.

Barbedo, J.G.A., Koenigkan, L.V., Halfeld-Vieira, B.A., Costa, R.V., Nechet, K.L., Godoy, C.V., Junior, M.L., Patricio, F.R.A., Talamini, V., Chitarra, L.G., et al., 2018. Annotated plant pathology databases for image-based detection and recognition of diseases. IEEE Latin Am. Trans. 16, 1749–1757.

Barbedo, J.G.A., Koenigkan, L.V., Santos, T.T., 2016. Identifying multiple plant diseases using digital image processing. Biosyst. Eng. 147, 104–116.

Bock, C., Poole, G., Parker, P., Gottwald, T., 2010. Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging. Crit. Rev. Plant Sci. 29, 59–107.

Dehnen-Schmutz, K., Foster, G.L., Owen, L., Persello, S., 2016. Exploring the role of smartphone technology for citizen science in agriculture. Agronomy Sustainable Dev. 36, 25.

Fuentes, A., Yoon, S., Kim, S., Park, D., 2017. A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. Sensors 17, 2022.

Ghosal, S., Blystone, D., Singh, A.K., Ganapathysubramanian, B., Singh, A., Sarkar, S., 2018. An explainable deep machine vision framework for plant stress phenotyping. Proc. Nat. Acad. Sci. 115, 4613–4618.

Hughes, D.P., Salathe, M., 2015. An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv e-prints, arXiv:1511. 08060.

Incaper, 2019. Cafeicultura. URL https://incaper.es.gov.br/cafeicultura accessed on 10-29-2019.

Inoue, H., 2018. Data augmentation by pairing samples for images classification. arXiv e-prints, (p. arXiv:1801.02929).

International Coffee Organization, 2019. Ico annual review. URL http://www.ico.org/documents/cy2018-19/annual-review-2017-18-e.pdf accessed on 10-29-2019.

Johannes, A., Picon, A., Alvarez-Gila, A., Echazarra, J., Rodriguez-Vaamonde, S., Navajas, A.D., Ortiz-Barredo, A., 2017. Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case. Comput. Electron. Agric. 138, 200–209.

Kaya, A., Keceli, A.S., Catal, C., Yalic, H.Y., Temucin, H., Tekinerdogan, B., 2019. Analysis of transfer learning for deep neural network based plant classification models. Comput. Electron. Agric. 158, 20–29.

Kranz, J., 1988. Measuring plant disease. In: Experimental Techniques in Plant Disease Epidemiology. Springer, Berlin, Heidelberg, pp. 35–50.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.

Liang, Q., Xiang, S., Hu, Y., Coppola, G., Zhang, D., Sun, W., 2019. PD2SE-Net: Computer-assisted plant disease diagnosis and severity estimation network. Comput. Electron. Agric. 157, 518–529.

Liu, B., Zhang, Y., He, D., Li, Y., 2017. Identification of apple leaf diseases based on deep convolutional neural networks. Symmetry 10, 11.

Ma, J., Du, K., Zheng, F., Zhang, L., Gong, Z., Sun, Z., 2018. A recognition method for cucumber diseases using leaf symptom images based on deep convolutional neural network. Comput. Electron. Agric. 154, 18–24.

Maaten, Lv.d., Hinton, G., 2008. Visualizing data using t-SNE. J. Machine Learn. Res. 9, 2579–2605.

Manso, G.L., Knidel, H., Krohling, R., Ventura, J.A., 2019. A smartphone application to detection and classification of coffee leaf miner and coffee leaf rust. arXiv e-prints, arXiv:1904.00742.

Mohanty, S.P., Hughes, D.P., Salathé, M., 2016. Using deep learning for image-based plant disease detection. Front. Plant Sci. 7, 1419.

Oliveira, C., Auad, A., Mendes, S., Frizzas, M., 2014. Crop losses and the economic impact of insect pests on Brazilian agriculture. Crop Protection 56, 50–54.

Ruder, S., 2017. An overview of multi-task learning in deep neural networks. arXiv e-prints, arXiv:1706.05098.

Singh, A.K., Ganapathysubramanian, B., Sarkar, S., Singh, A., 2018. Deep learning for plant stress phenotyping: Trends and future perspectives. Trends Plant Sci. 23, 883–898.

Too, E.C., Yujian, L., Njuki, S., Yingchun, L., 2019. A comparative study of fine-tuning deep learning models for plant disease identification. Comput. Electron. Agric. 161, 272–279.

Ventura, J.A., Costa, H., Lima, I.d.M., 2017. Manejo de pragas do café conilon. In: Ferrão, R.G., da Fonseca, A.F.A., Ferrão, M.A.G., de Muner, L.H. (Eds.), Café Conilon, 2nd ed. Incaper, Vitória, ES, pp. 434–479.

Wang, G., Sun, Y., Wang, J., 2017. Automatic image-based plant disease severity estimation using deep learning. Comput. Intell. Neurosci. 2017.

Zhang, H., Cisse, M., Dauphin, Y.N., & Lopez-Paz, D. (2017). mixup: Beyond empirical risk minimization. arXiv e-prints, arXiv:1710.09412.