

Use of Images of Leaves and Fruits of Apple Trees for Automatic Identification of Symptoms of Diseases and Nutritional Disorders

Lucas Garcia Nachtigall, Center for Technological Advancement, Federal University of Pelotas, Pelotas, Brazil

Ricardo Matsumura Araujo, Center for Technological Advancement, Federal University of Pelotas, Pelotas, Brazil

Gilmar Ribeiro Nachtigall, Embrapa Grape & Wine, Vacaria, Brazil

ABSTRACT

Rapid diagnosis of symptoms caused by pest attack, diseases and nutritional or physiological disorders in apple orchards is essential to avoid greater losses. This paper aimed to evaluate the efficiency of Convolutional Neural Networks (CNN) to automatically detect and classify symptoms of diseases, nutritional deficiencies and damage caused by herbicides in apple trees from images of their leaves and fruits. A novel data set was developed containing labeled examples consisting of approximately 10,000 images of leaves and apple fruits divided into 12 classes, which were classified by algorithms of machine learning, with emphasis on models of deep learning. The results showed trained CNNs can overcome the performance of experts and other algorithms of machine learning in the classification of symptoms in apple trees from leaves images, with an accuracy of 97.3% and obtain 91.1% accuracy with fruit images. In this way, the use of Convolutional Neural Networks may enable the diagnosis of symptoms in apple trees in a fast, precise and usual way.

KEYWORDS

Apple, Apple Disorders, Artificial Intelligence, Automatic Disease Identification, Classifications, Convolutional Neural Networks, Disorders, Machine Learning

1. INTRODUCTION

Approximately 25% of apple production is lost by attack of pests, diseases and nutritional disorders of plants. A rapid and efficient diagnosis of these situations is essential to avoid greater losses. It is estimated that 80 to 90% of the damage caused by pests and diseases which attack the culture of the apple tree occurs in the leaves and fruits. Among these diseases, scab of apple tree and the spot of *Glomerella* are the most important ones (Valdebenito-Sanhueza et al., 2008). In the case of pests, where leaves and fruits serve as food source or hosts, the major issues are due to the attacks of the fruit fly, fruit moth and big caterpillars (Kovaleski, 2004). On the other hand, the disturbances caused by the excess or lack of nutrients are visible mainly in the leaves during the vegetative growth phase (Nachtigall et al., 2004).

A correct diagnosis is essential in order to define strategies of management and control, and consequently for the rational use of fertilizers and pesticides. One main obstacle towards a quick

DOI: 10.4018/IJMSTR.2017040101

diagnosis is the need for trained experts, making it costly to cover large areas in a timely manner. Moreover, experts often specialize in different issues, increasing the rate of misdiagnosis.

Some approaches exist to try and reduce the dependency on experts. A widely used one is a simple printed guide containing photos and explanations on how to diagnose a wide range of issues (Valdebenito-Sanhueza et al., 2008). Expert Systems, often built on top of Case-Based Reasoning algorithms, are applied to some cultures - e.g. vine (Fialho et al., 2012). These systems still require considerable amounts of training and are often not very accurate, mainly due to the typically very large number of questions required to be answered by the user and the sensitivity to wrong answers.

The concept of using machine learning to detect symptoms in plants has been shown to be a promising alternative in recent years, where several studies using different approaches have been carried out to identify or classify symptoms in cultivated plants. Rumpf et al. (2010) aimed to discriminate diseased from non-diseased sugar beet leaves, to differentiate between the three types of diseases and to identify diseases even before specific symptoms became visible. The authors used Support Vector Machines with a radial basis function as kernel to perform the identification and classification of symptoms of healthy or unhealthy leaves. As input they used nine spectral vegetation indexes, related to physiological parameters as features for an automatic classification, resulting in classification accuracies up to 97% on sugar beet leaves and diseased leaves, up to 86% classification accuracy between the three diseases symptoms and accuracy between 65% and 90% for pre-symptomatic detection of plant diseases.

Al-Hiary et al. (2011) proposed a methodology to automatically detect and classify plant leaf diseases from images. The process consists of six main phases: image acquisition, image pre-processing, image segmentation, feature extraction, statistical analysis and classification by a MLP. The authors used 32 samples for each of the six classes of leaves. Features were manually defined as 10 texture features extracted from the image.

A Perceptron Multilayer for classifying grape leaf diseases was used by Sannakki et al. (2013). Clustering was used to segment the image into groups, followed by lesion and manually defined feature extraction. They used a very small dataset (33 examples) and were able to achieve perfect accuracy in an also-small (4 examples) test set.

Revathi and Hemalatha (2014) focus on cotton leaf spot diseases. The authors used a dataset with 270 images divided into 6 disease classes. Features were manually defined, consisting of leaf edge, color and texture features. A Cross Information Gain Deep Forward Neural Network was used to perform the classification, resulting in an overall accuracy of 95%. Mohanty et al. (2016) proposed using GoogLeNet for the classification of leaf diseases in different cultures, using a dataset with 54306 images from different laboratories, divided in 38 classes, and a cluster computer to process the data, resulting in accuracy of 99.35%.

With the popularization of the techniques of artificial intelligence and machine learning for the classification of images, large banks of images were created, which were used to test the efficiency of the new techniques developed. Thus, several techniques were searched for a lower error rate in the classification of the images present in this image bank. One of the techniques that stood out for this classification task was the Convolutional Neural Network (CNN).

Krizhevsky; Sutskever; and Hinton (2012) conducted the training of a deep convolutional neural network to classify ImageNet and obtained the best result ever reported using the sub-sets presented in the ILSVRC-2010 and ILSVRC-2012 competitions. The architecture used had five convoluted layers and three completely connected layers, with a training time between five and six days. This architecture obtained an error rate in the top-1 and top-5 sets of 38.1% and 16.4%, respectively, using a set of five networks.

Based on the results obtained by Krizhevsky; Sutskever; And Hinton (2012), Simonyan and Zisserman (2014) made modifications to the proposed architecture using small convolutional filters (3x3) in order to perform better than those previously proposed. In addition, another focus was on the depth of the network, which could be increased by the fact that the filters were smaller in size. In

this architecture were present sixteen convoluted layers and three completely connected layers. The error rate in the top-1 and top-5 sets were 23.7% and 6.8%, respectively, using a set of two networks and multi-crop and dense evaluation techniques.

Moreover, CNNs have been shown to be good alternatives in the most different problems of computer vision, such as character recognition (Lecun et al., 1998), face recognition (Lawrence et al., 1997) and object categorization (Yu; Xu; and Gong, 2009).

On this paper, we extend the work developed in (Nachtigal et al., 2016), as we continue to analyze the possibility to automatically identify symptoms of important disorders present in apple orchards (*Malus domestica* Borkh) using only photographs of leaf and fruits, developing a new dataset of labeled images, containing five common disorders in leaves (deficiency of potassium and magnesium, apple scab, glomerella stains and damage caused by herbicide) and five important disorders in fruits (scab, alternaria rot, bull's eye rot, penicillium rot and calcium deficiency – bitter pit).

2. MATERIAL AND METHODS

The methodology consists of building a dataset containing labeled examples of ten types of issues commonly affecting apple orchards. This dataset was randomly partitioned into training, validation and test subsets. The training and validation subsets were used to train and optimize a Convolutional Neural Network and a Multilayer Perceptron (as a baseline) techniques. The test set was then used to assess the performance of the resulting classifiers and was also presented to experts for classification.

2.1. Dataset

The datasets were built by harvesting leaves and fruits from three species of apple trees (*Maxigala*, *Fuji Suprema* and *Pink Lady*) and photographing each leaf or fruit over a white background. Each sample was then subjected to laboratory tests to properly identify the underlying disorder which was then used to label the image. Harvesting occurred between January and April 2015 from orchards located in the southern part of Brazil, at Embrapa Uva e Vinho – Estação Experimental de Fruticultura de Clima Temperado.

Healthy leaves and five disorders were selected among those collected, as they are the most prevalent in the region. Selected symptoms represent two damages caused by nutritional imbalances (deficiency of potassium and magnesium), two diseases damage (apple scab - caused by the fungus *Venturia inaequalis* and Glomerella stains - caused by the fungus *Glomerella cingulata*) and damage caused by herbicide (glyphosate). Table 1 summarizes the leaf dataset. As for the fruits, five symptoms were also selected, based as they are the most prevalent in the region and the damage caused have a superior economy importance, along with healthy fruits. These symptoms are: one disorder caused by nutritional imbalances (calcium deficiency - bitter pit) and four disorders caused by diseases (scab, alternaria rot, penicillium rot and bull's eye rot). Table 1 summarizes the fruit dataset.

The identification of the disorders was conducted by a group of professional agronomist researchers specialized in these symptoms and with ample experience in plant nutrition and plant pathology. Also, to further support these professionals, published specialized books (Nachtigal et al., 2004; Valdebenito-Sanhueza et al., 2008) were used, providing technical information on the symptoms. Figures 1 and 2 exemplify the evaluated symptoms on leaves and fruits, and the Table 1 identify the number of samples collected for each symptom and healthy samples.

After the harvest and identification of fruit and leaves, a camera with the resolution of 12 MP was used to capture the images of each sample. A white background was used to photograph each leaf or fruit separately. An agronomist further analyzed the images, in order to verify if the symptoms classifications were possible, based on what was present on the images. A few images which presented defects or were outside the capture standards used were discarded.

In order to reduce errors and properly establish a ground-truth, three strategies were employed by the experts to properly diagnose each issue:

Figure 1. Leaves with symptoms of: (A) Potassium deficiency; (B) Magnesium deficiency; (C) Scab damage; (D) Glomerella damage; and (E) Damage caused by herbicide

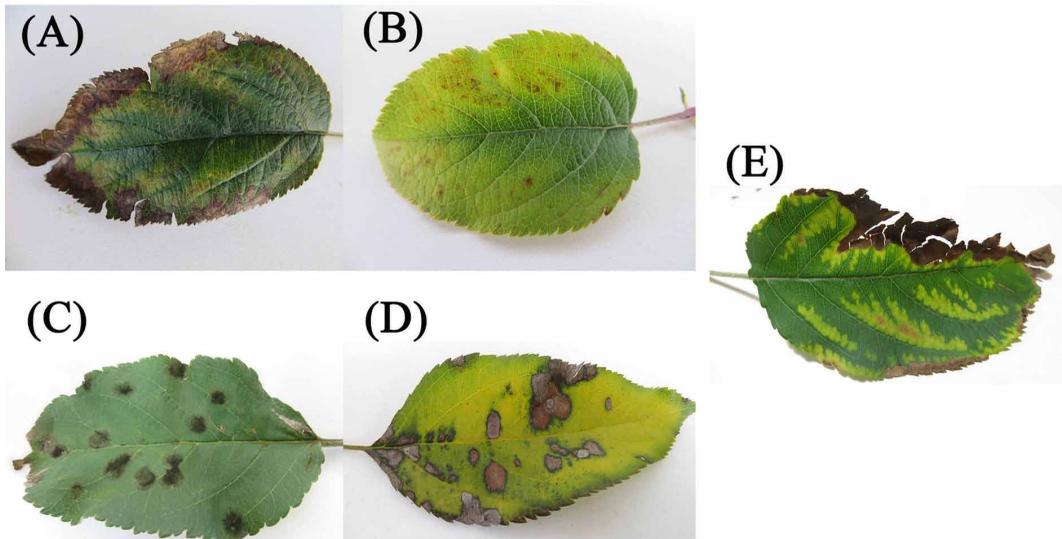
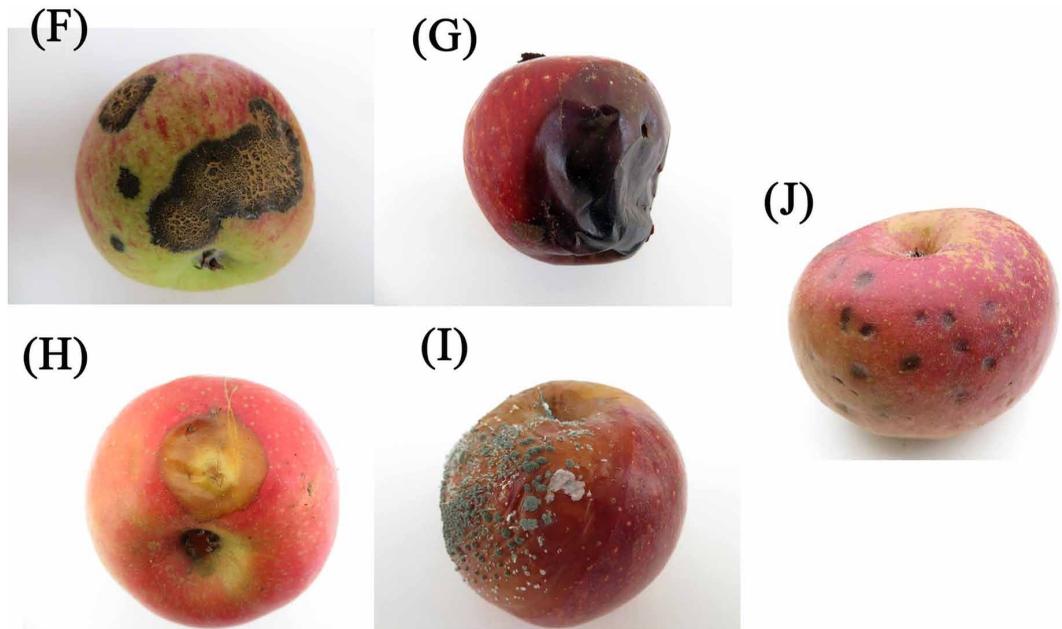


Figure 2. Fruits with the symptoms of: (F) Scab; (G) Alternaria rot; (H) Bull's eye rot; (I) Penicilium rot; and (J) Calcium deficiency - bitter pit



1. For symptoms caused by nutritional imbalances, samples of normal leaves and leaves with potassium and magnesium deficiency symptoms were selected, each consisting of 100 leaves. The samples then were dried at 60°C in a forced air circulation greenhouse until a constant weight and forwarded to the laboratory for chemical analyzes in order to quantify the total

Table 1. Number of leaves and fruits collected for each class

Issue	Number of Leaves Collected	Number of Fruits Collected
Potassium deficiency	341	-
Magnesium deficiency	355	-
Scab damage	391	-
Glomerella stain	558	-
Herbicide damage	325	-
Healthy leaves	569	-
Alternaria rot	-	490
Calcium deficiency	-	434
Penicillium rot	-	702
Bull's eye rot	-	486
Scab	-	931
Healthy fruits	-	1327

concentrations of nutrients (potassium and magnesium). The analysis results show that samples with symptoms effectively represent the deficiencies of potassium and magnesium. In the analytic results for samples of normal leaves and potassium deficiency ones, it was proved that the chosen leaves with visual symptoms of potassium deficiency (which originated the images for this symptom in the dataset) contained only 38% of the potassium concentration compared with ordinary and healthy leaves. The analytic results for samples of normal leaves and magnesium deficient leaves (at two levels of severity), verified that the sample leaves with visual symptoms of magnesium deficiency (also originating the images for this symptom in the dataset) contained 63% and 40% of the magnesium concentration in relation to normal leaves. As for the fruits, 100 samples of pulp and peels were collected, both normal fruits (without symptoms) and fruits with symptoms of calcium deficiency. The samples were then sent to the lab, for the chemical analysis, aiming at the quantification of total calcium concentrations. The analytic results of the samples represent effectively the calcium deficiency. These results of the pulp and peel samples, which possess calcium deficiency symptoms, demonstrated that these samples had only 78% and 65% of the calcium concentration found in healthy pulp and peel fruits, respectively;

- For symptoms caused by disease damage on leaves (apple tree scab - caused by the fungus *Venturia inaequalis* - and Glomerella's stains - caused by the fungus *Glomerella cingulata*), and on fruits (apple tree scab - caused by the fungus *Venturia inaequalis*, alternaria rot – caused by the fungus *Alternaria alternata*, bull's-eye rot induced by *Cryptosporiopsis perennans* and Penicilium rot caused by the fungus *Penicillium expansum*), samples were selected with symptoms previously identified for the each of the diseases. These samples were incubated for multiplication of the causative agent (fungus) and after was performed the isolation of fungi and their characterization and identification using a microscope, allowing the proof of causal agents and their damage on apple tree leaves and fruits. The obtained results proved that samples with diseases symptoms effectively represent the selected diseases;
- For symptoms caused by herbicide damage (glyphosate), it was decided to conduct chemical analysis in order to quantify the total concentrations of nutrients which could possibly cause confounding of symptoms in cases where the nutrients concentrations were below normal. This decision was due the fact that the analysis of the herbicide's active principle is difficult to characterize, once it is rapidly degraded on the plant after its absorption and origin of toxicity

symptoms. Then, samples with herbicide damage symptoms were put through the same protocols for nutritional analysis, which results showed these samples did not effectively present any nutritional disorders. Since the samples of leaves with symptoms caused by the herbicide glyphosate (in three levels of severity) were not different from the nutrient concentrations of normal leaves, all nutrients are within the range considered as standards for apple orchards (Nachtigall et al., 2004).

2.2. Pre-Processing, Training and Evaluation

All images were re-sized to a resolution of 256x256 pixels. In order to have a balanced dataset, we randomly selected 290 examples (labeled images) for each of the five classes containing symptoms. The resulting 1450 images were divided into three subsets. A test subset (hold-out set) was created by randomly choosing 15 images of each class. The remaining examples were further partitioned in a training set (192 examples, 70%) and validation set (83 examples, 30%). This division of subsets was due to the fact that when comparing machine learning techniques to experts, large test subsets would become exhausting for experts to classify, possibly increasing human error to the results.

As for the fruit dataset, 430 examples (labeled images) were randomly selected for each of the five classes containing symptoms and for the healthy fruits. The resulting 2580 images were divided into three subsets. A test subset (holdout set) was created by randomly choosing 43(10%) images of each class. The remaining examples were further partitioned in a training set (271 examples, 70%) and validation set (116 examples, 30%). Since there was no classification by experts on the fruit dataset, a larger hold-out test set was possible.

The tested learning algorithms were trained over the training set using different parameters and configurations and then applied to the validation set. The best performing (over the validation set) parameters of each algorithm were then trained over training and validation examples and applied to the test set. This procedure was adopted in order to avoid over fitting the results to the training or validation set. The test set was also used by the experts for them to provide their classifications, so that a direct comparison was possible. Hence, all results reported in this paper are for the test set.

Furthermore, in order to analyze the number of samples needed for a satisfactory classification, smaller training subsets were created, randomly selecting 5, 10, 20, 50, 100, 150, 200, and 250 samples from each class containing symptoms. These training subsets were then tested using the previous test subset, without any changes in the network configuration after the validation process was finished.

After evaluating the CNN approach by comparing its results with MLP networks and experts, 275 randomly selected healthy leaves were then added to the train subset, and 15 randomly selected healthy leaves were added to the test subset, in order to evaluate the network capacity to distinguish healthy leaves from the five symptoms chosen.

The evaluation of the results was done by calculating the overall accuracy of each classifier and analyzing the resulting confusion matrix, also the metrics recall, precision and kappa (Landis and Koch 1977) were used.

2.3. Convolutional Neural Network

For this paper, Caffe (Jia et al., 2014) and DIGITS (NVIDIA 2015) were the tools used to help in building, training and testing Convolutional Neural Networks. Multiple architectures were tested, ranging from shallow networks with 4 layers to deep networks using small (3x3) convolution filters as shown in (Simonyan and Zisserman 2014).

The best results over the validation set were obtained by the AlexNet architecture (Krizhevsky et al., 2012). This network consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers followed by soft-max and dropout regularization. The training parameters were as follows: batch size of 2; a step learning policy was added with a gamma of 0.2; training through a maximum of 300 epochs.

One of the issues faced, was the hardware limitations, where in large datasets it is common to use multiple high end GPU cards for faster and deeper learning (Mohanty et al., 2016), the hardware used for this research is the following: Geforce GTX 760 GPU, Intel(R) core(TM) i7 860 @ 2.80GHz CPU, 16 GB RAM.

2.4. Baseline Algorithms

In order to provide a baseline comparison to CNNs, was applied Multilayer Perceptrons to the dataset, using the same training methodology applied for CNNs. Multilayer Perceptron (MLP), was chosen for being closely related to CNNs. Shallow MLPs were able to achieve high accuracy on tasks such as digit recognition (Cho 1997), image classification (Hara et al., 1994) and feature extraction (Ruck et al., 1990).

DIGITS was used to build and test the MLPs. Different shallow architectures were tested. The best configuration consisted of two hidden layers with 200 and 500 neurons, respectively, with one unit for each class at the output layer and 65536 input units (256x256 pixels).

2.5. Classification by Experts

To collect the classification by experts, we asked 7 volunteer researchers specialized in apple trees to classify the images in the test set, including the agronomist who collected the leaves for the database, since he had a greater knowledge on the chosen disorders. These experts are further specialized in different fields of research, such as plant pathology or plant nutrition.

Each expert was given a form (Figure 3) to choose, for each image, one of the five classes. Healthy leaves were not shown to the experts, because in the field leaves that seem healthy are not chosen in order to classify a disorder. Each expert was shown the images in a succession over a random permutation. They were not given the correct answers and no time limit was enforced.

After storing the form answers, a voting system was developed in order to best evaluate the machine learning algorithms with experts. The motivation to create this voting system was because not all experts have knowledge of different research areas, and in real case scenarios more than one expert could be called to diagnose possible leaf disorders.

In order to create a final answer, the voting system analyzed which symptom had the most votes for each image in the form. Also, in case of a tie between symptoms, the votes of experts which had a higher overall accuracy were chosen.

Figure 3. Example of image presented in the form shown to experts



- A) Glomerella
- B) Herbicide damage
- C) Magnesium deficiency
- D) Potassium deficiency
- E) Scab damage
- F) No reply

3. RESULTS AND DISCUSSION

Figure 4 presents the overall CNN accuracy when different number of images per class are used for training and tested against the hold-out test set. The graph shows a logarithmic increase on accuracy when training set is increased. The curve largely levels off for samples larger than 200 images, evidencing that the number of samples collected was adequate.

Table 2 shows the final CNN confusion matrix when applied to the hold-out test set. The overall accuracy was of 97.3%, with only two incorrect classifications. The MLP, applied to the same test set, resulted in an accuracy of 77.3%.

Table 3 shows the individual accuracy obtained from the 7 consulted experts, also showing their specific field of research. We can observe that accuracy varies considerably across experts. The best result (93.3%) is worse than the result obtained by the CNN, but the average (71.9%) was much worse and below that obtained by the MLP. Table 4 presents the confusion matrix of the experts when aggregated by voting, where it can be seen that the overall accuracy improves significantly.

Table 5 summarizes the results. The best accuracy was obtained by the CNN, with a 97.3% accuracy, followed by the voting system with 96.0% accuracy, the best expert with 93.3% accuracy, and the MLP network with 77.3% accuracy. Figure 5 shows the Confidence interval ($IC1-\alpha(p)$) for the classifiers, using a 99% confidence level. It is possible to observe that all techniques are much better than random choice and that aggregated experts and the CNN have comparable performance and both are better than the MLP. A second experiment was conducted, introducing healthy leaves to the dataset. In this case, only the CNN was tested. The same distribution of 275 images for training and 15 images for testing was used.

No changes were made in the configuration of the CNN network in order to avoid an over fitting to the results. With healthy leaves in the training and test groups, the CNN was able to achieve accuracy of 96.67%, as shown in Table 6. It is possible to observe that the trained CNN is able to attain perfect accuracy when distinguishing between healthy and unhealthy leaves. This is expected,

Figure 4. Relation between the number of samples used for learning and the accuracy obtained in the symptoms classification. The solid line represents a logarithmic best-fit function over the data points (circles).

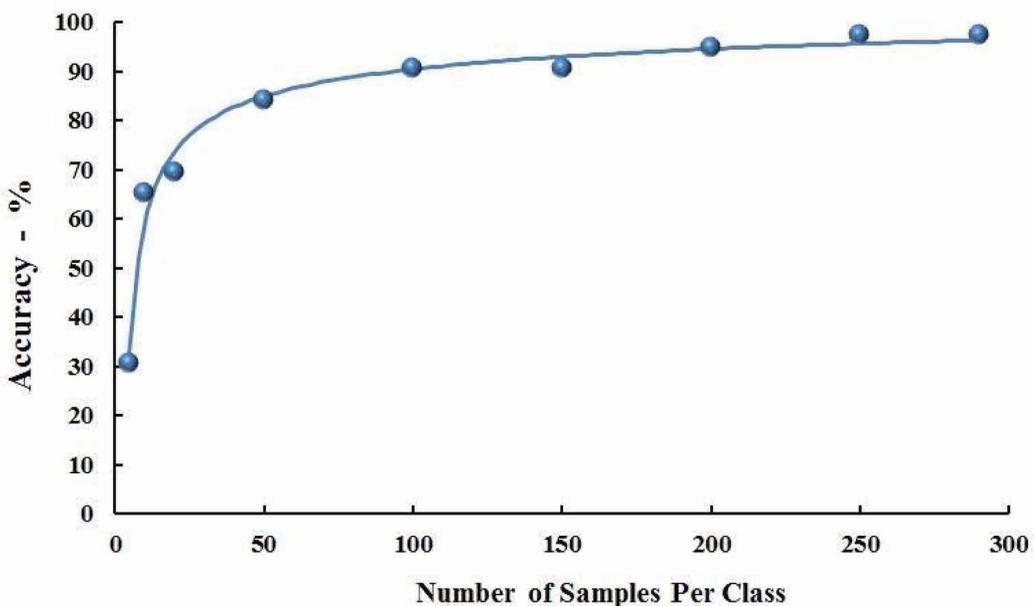


Table 2. Confusion Matrix resulting from CNN classification on the hold-out set

		CNN						
Label image	Symptom in leaves	Glomerella	Herbicide	Magnesium def.	Potassium def.	Scab	Recall	Precision
	Glomerella	15	0	0	0	0	100.0%	93.3%
	Herbicide	0	15	0	0	0	100.0%	100.0%
	Magnesium def.	0	0	15	0	0	100.0%	100.0%
	Potassium def.	0	0	0	14	1	93.3%	100.0%
	Scab	1	0	0	0	14	93.3%	93.3%
Accuracy								97.3%
Kappa								0.97

Table 3. Field of research and accuracy obtained by each expert when classifying images in the hold-out test set

Subject	Field of Research	Accuracy
1	Soil	93.3%
2	Plant pathology	92.0%
3	Post harvest	90.6%
4	Plant pathology	70.6%
5	Plant nutrition	60.0%
6	Crop science	60.0%
7	Environmental management	37.3%
Average	-	71.9%

Table 4. Confusion Matrix resulting from the aggregation of the classifications provided by human experts

		Voting System - Experts						
Label image	Symptom in leaves	Glomerella	Herbicide	Magnesium def.	Potassium def.	Scab	Recall	Precision
	Glomerella	15	0	0	0	0	100.0%	100.0%
	Herbicide	0	14	1	0	0	93.3%	100.0%
	Magnesium def.	0	0	14	0	1	93.3%	93.7%
	Potassium def.	0	0	0	14	1	93.3%	100.0%
	Scab	0	0	0	0	15	100.0%	88.2%
Accuracy								96.0%
Kappa								0.95

Table 5. Summary of the results of the leaf dataset, ordered by accuracy

Technique	Accuracy
CNN	97.3%
MLP	77.3%
Voting system	96.0%
Highest Accuracy Expert	93.3%

Figure 5. Confidence intervals with 99% confidence level, with normal approximation for each share of accuracy

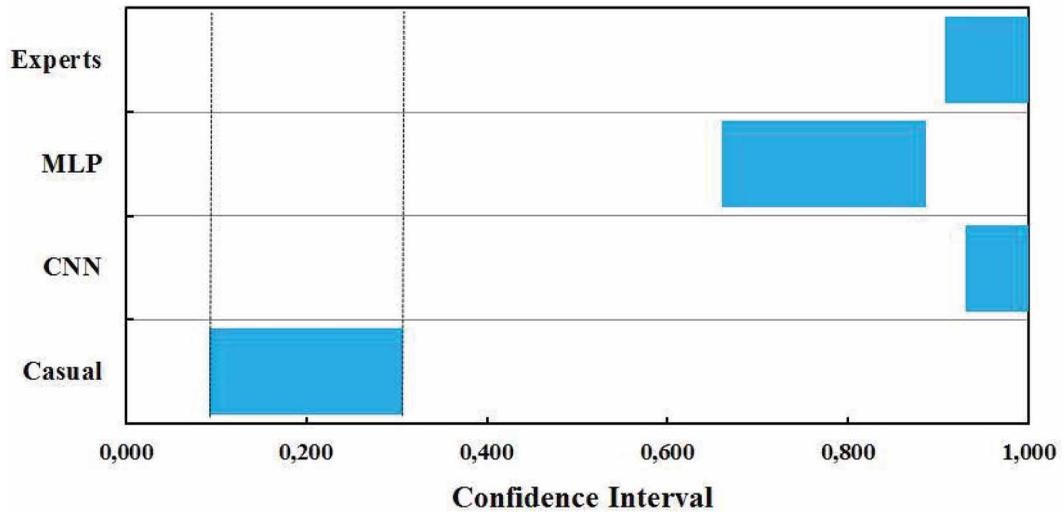


Table 6. Confusion Matrix resulting from CNN classifications on a hold-out test set when healthy leaves are added to the dataset

Label image	Symptom in leaves	CNN						Recall	Precision
		Glomerella	Herbicide	Magnesium def.	Potassium def.	Scab	Healthy leaves		
	Glomerella	15	0	0	0	0	0	100.0%	93.7%
	Herbicide	0	15	0	0	0	0	100.0%	100.0%
	Magnesium def.	0	0	15	0	0	0	100.0%	93.7%
	Potassium def.	0	0	0	14	1	0	93.3%	100.0%
	Scab	1	0	1	0	13	0	86.6%	92.8%
	Healthy leaves	0	0	0	0	0	15	100.0%	100.0%
Accuracy									96.6%
Kappa									0.96

as the disorders all display strong symptoms on the leaves, but an additional classification error is now made when distinguishing between disorders.

One possibility to improve classification is to train a CNN to first distinguish between healthy and unhealthy and then another to further distinguish between disorders. Indeed, a CNN trained on a binary healthy/unhealthy class is also able to attain 100% accuracy, hence allowing the use of the disorders-only CNN.

Table 7 shows the CNN confusion matrix when applied to the fruit train and hold-out test set. The overall accuracy was of 91.1%. Lower results on fruit classification were expected, as the fruit images were taken from different angles, allowing a better view of the symptom, but possibly making the learning task more challenging. As seen on Table 6, the applied CNN on the fruit hold-out test set were also able to achieve 100% accuracy when distinguishing healthy fruits.

4. CONCLUSION AND FUTURE WORK

In this paper we proposed the use Convolutional Neural Networks to assist on the task of identification and classification of apple tree disorders from leaf images. We used a novel dataset consisting of leaf images containing five known disorders, all confirmed by laboratory tests, and compared the results of a CNN to that of a MLP and experts.

Our results show that a CNN based on the AlexNet architecture is able to significantly outperform the baseline MLP, showing comparable performance to that of a group experts and outperforming any single expert. When applying the CNN to the fruit dataset, we were also able to achieve desirable results. Moreover, perfect accuracy was obtained when only distinguishing between healthy and unhealthy leaves or healthy and unhealthy fruits.

We conclude that CNNs compose a viable and useful option for this task, with more robust classifications than single human experts. In this sense, an automated system based on the trained model could contribute towards diagnosis reliability and cost reduction.

Compared to previous works, our approach does not require specialized equipment to capture the images or any sort of feature extraction or engineering. The CNN is able to learn relevant features

Table 7. Confusion Matrix resulting from CNN classifications on the fruit hold-out test set

		CNN - Fruits							
Label image	Symptom in fruits	Alternaria rot	Calcium def.	Penicillium rot	Bull's eye rot	Scab	Healthy fruits	Recall	Precision
	Alternaria rot	40	0	1	1	1	0	93.0%	87.0%
	Calcium def.	0	43	0	0	0	0	100.0%	89.6%
	Pinicillium rot	4	2	35	2	0	0	81.4%	87.5%
	Bull's eye rot	0	3	4	35	0	1	93.3%	100.0%
	Scab	2	0	0	1	39	1	90.7%	97.5%
	Healthy fruits	0	0	0	0	0	43	100.0%	95.5%
	Accuracy								
Kappa									0.89

from the data, to which we attribute the improved performance. This also allows for the general approach to be used in different disorders or even cultures with changes only to the dataset. This is important to allow for the automatic improvement of the model when more data is made available.

Although the SVM technique showed high accuracies in the related work, when applied the same methodology as the other techniques, using no pre-processing or feature extraction, the results did not achieve more than 60% accuracies, therefore they were not included in this article.

Several lines of future work are being planned. We are expanding the current dataset to make available more diverse examples. While we have shown that more examples obtained in the same way will only provide marginal improvements, the introduction of more diversity (e.g. different backgrounds and light conditions) could allow for better performance. We are also introducing additional disorders and cultures to test how the approach scales with these settings.

Different architectures are also being considered. We believe that a combination of more examples and improved architecture could lead to a system that can consistently outperform experts. Finally, we aim at integrating our methodology into working systems that can be used on the field, in less controlled conditions.

REFERENCES

- Al-Hiary, H. A., Ahmad, S. B., Reyalat, M., Braik, M., & Alrahamneh, Z. (2011). Fast and Accurate Detection and Classification of Plant Diseases. *International Journal of Computers and Applications*, 17(1), 31–38. doi:10.5120/2183-2754
- Cho, S. (1997). Neural-network classifiers for recognizing totally unconstrained handwritten numerals. *IEEE Transactions on Neural Networks*, 8(1), 43–53. doi:10.1109/72.554190 PMID:18255609
- Fialho, F., Garrido, L., Botton, M., de Melo, G., Fajardo, T., & Naves, R. (2012). Diagnostico de doenças e pragas na cultura da videira usando o sistema especialista uzum. Retrieved February 02, 2016, from <http://ainfo.cnptia.embrapa.br/digital/bitstream/item/73877/1/cot128.pdf>
- Hara, Y., Atkins, R., Yueh, S., Shin, R., & Kong, J. (1994). Application of neural networks to radar image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 32(1), 100–109. doi:10.1109/36.285193
- Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., & Darrell, T. (2014). Caffe. In *Proceedings of the ACM International Conference on Multimedia - MM '14*.
- Kovaleski, A. (2004). Pragas. In KOVALESKI, A. *Maçã: fitossanidade* (pp. 11–33). Bento Gonçalves: Embrapa Uva e Vinho.
- Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* (pp. 1097–1105).
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159–174. doi:10.2307/2529310 PMID:843571
- Lawrence, S., Giles, C. L., Tsoi, A. C., & Back, A. D. (1997). Face recognition: A convolutional neural-network approach. *IEEE Transactions on Neural Networks*, 1(8), 98–113.
- Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. In *Proceedings of the IEEE* (pp. 2278–2324).
- Mohanty, S. P., Hughes, D., & Salathe, M. (2016). Inference of plant diseases from leaf images through deep learning. arXiv:1604.03169
- Nachtigall, G. R., Basso, C., & Freire, C. J. S. (2004). Nutrição e adubação de pomares. In G.R. Nachtigall, (Ed.). *Maçã: produção*. Brasília, Embrapa Informação Tecnológica (pp. 63-77). Frutas do Brasil.
- Nachtigall, L. G., Araujo, R. M., & Nachtigall, G. R. (2016). Classification of Apple Tree Disorders Using Convolutional Neural Networks. In *Proceedings of the 2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI)*.
- NVIDIA. (2015). Nvidia digits – interactive deep learning GPU training system. Retrieved February 02, 2016, from <https://developer.nvidia.com/digits>
- Revathi, P., & Hemalatha, M. (2014). *Identification of cotton diseases based on cross information gain deep forward neural network classifier with PSO feature selection*. *International Journal of Engineering and Technology (IJET)*.
- Ruck, D., Rogers, S., & Kabrisky, M. (1990). Feature selection using a multilayer perceptron. *Journal of Neural Network Computing*, 2(2), 40–48.
- Rumpf, T., Mahlein, A., Steiner, U., Oerke, E., Dehne, H., & Plümer, L. (2010). Early detection and classification of plant diseases with Support Vector Machines based on hyperspectral reflectance. *Computers and Electronics in Agriculture*, 74(1), 91–99. doi:10.1016/j.compag.2010.06.009
- Sannakki, S. S., Rajpurohit, V. S., Nargund, V. B., & Kulkarni, P. (2013). Diagnosis and classification of grape leaf diseases using neural networks. In *Proceedings of the Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, 1-5.
- Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. Retrieved February 02, 2016, from <https://arxiv.org/abs/1409.1556>

Valdebenito-Sanhueza, R. M., Nachtigall, G. R., Kovaleski, A., dos Santos, R. S. S., & Spolti, P. (2008). *Manual de identificação e controle de doenças, pragas e desequilíbrio nutricional da macieira* (pp. 45–58). Bento Gonçalves: Embrapa Uva e Vinho.

Yu, K., Xu, W., & Gong, Y. (2009). Advances in Neural Information Processing Systems: Vol. 1889-1896. *Deep learning with kernel regularization for visual recognition*.

Ricardo Matsumura Araujo is a Computer Science Professor at the Center for Technological Advancement in the Federal University of Pelotas (UFPEl). He holds a PhD and Master Degrees in Computer Science from UFRGS and lectures at the Computer Science, Computer Engineering undergrad courses and the Graduation Program in Computer Science. His research interests include machine learning, computational social science and data science.

Gilmar Ribeiro Nachtigall holds the title of PhD in Agronomy. Gilmar currently works as a researcher in Embrapa Wine & Grapes.