Australian Journal of Crop Science
AJCS
AJCS 15(08):1177-1185 (2021) ISSN:1835-2707
doi: 10.21475/ajcs.21.15.08.p3256

Soft computational techniques to identify cotton leaf damage

Rafael Faria Caldeira*[1,2], Wesley Esdras Santiago[2], Bárbara Teruel[1]

[1]**Faculty of Agricultural Engineering of the University of Campinas– FEAGRI/UNICAMP**
[2]**Institute of Agricultural Sciences of the Federal University of the Jequitinhonha and Mucuri Valleys– ICA/UFVJM**

***Corresponding author: rafael.faria@ufvjm.edu.br**

Abstract

The principal objective of agriculture is the production of a high yield of healthy crops. This yield may be improved by the automatic detection of diseases and the consequent reduction in the use of pesticides. A digital processing system for images was thus developed and used to identify lesions on the leaves of cotton plants. A collection of 60,659 images of sub-metric resolution showing samples of soil and both healthy and damaged leaves was obtained and processed with an algorithm for the extraction of texture from 102x102-pixel samples. Then they analyzed with a neuro-fuzzy classifier trained to discriminate the three types of regions (soil, healthy leaf, and lesioned leaf). The algorithm developed was able to recognize the three classes. It generated a great amount of information on recognition of background which was more consistent than leaf damage areas. Therefore, it surpassed the performance of areas of healthy leaves. A similar trend was found for sensitivity. The overall accuracy of the system was 71.2%, suggesting that the unbalanced data of the different classes had skewed the results of the algorithm, as the number of false positives for the less well represented classes was greater. The analysis of unbalance (F-Score) showed that, independent of the volume of data, the attributes of texture utilized yielded better results for the images containing areas of damage in relation to overall accuracy. Therefore, given the challenges involved in the automatic identification of lesions in agricultural crops, such as variations in illumination, color, and texture, as well as obstruction, overlapping, and complexity of the region of which the image was taken, the behavior of the model was deemed satisfactory. Given the hybrid nature of the model, it should contribute to the state of the art in the use of intelligent systems in agriculture. This algorithm is available at https://github.com/rafaeufg/Cotton-diseases

Keywords: Precision agriculture, Artificial neural networks, Diffuse inference system, Machine learning, Digital processing of images.
Abbreviations: NFA_Adaptive neuro-fuzzy; SBC1_Class I sub-images; SBC2_Class 2 sub-images; SBC3_Class 3 sub-images.

Introduction

The interaction between plant genotype, pathogens, and the environment makes the pathogenesis of a disease a quite complex process. Moreover, it can reduce the yield, as well as the quality of the crop produced. For crops such as soybeans, losses of US$100.00 per hectare have been estimated, at times make the production unfeasible (Allen et al., 2017; Bowen et al., 2018). However, understanding the reactions of crops to biotic stress (symptomatology) can make it possible for the rural producer to diagnose diseases early on, minimizing the effects of their occurrence (Wahabzada et al., 2015).
The conventional method for the diagnosis of agricultural pests and diseases requires the analysis of lesions by a trained technician. This puts the farmer in a difficult situation, because such an analysis can only be executed after the development of the symptoms. Moreover, it depends on the experience of the technician. In scenarios involving extensive cultivation, there is also a demand for speed in the diagnosis, recommendations for treatment, and administration, since the windows for intervention are ever more limited (Santos, 2015; Bourguet and Guillemaud, 2016).
For some time, alternative technological strategies designed to help the producer make decisions in the management of pesticides for crops (Peshin et al., 2009; Veisi, 2012; Pelzer et al., 2012). Information about plant physiology and health can be inferred from the electromagnetic spectrum reflected from the crop (Camargo e Smith, 2009; Mutka and Bart, 2015). The visible region of the spectrum reveals the content of pigment, while the near infrared reflects the structural characteristics of the plants and the short infrared waves reflect mainly chemical components and the water content of the leaves (Patil and Kumar, 2011; Barbedo, 2016).
Pathogenesis causes alterations in the biophysical and biochemical properties of plants, and consequently their color (Barbedo, 2013) and texture (Pujari et al., 2015). The digital processing of images can be used to implement systems capable of identifying such alterations and provide a trustworthy diagnosis of lesions (Camargo and Smith, 2009; Zhang et al., 2017).
A system for the digital processing of images depends basically on three factors: the quality, the type and resolution of the image, the type of descriptor used to define characteristics, and the specific technique of machine learning selected (Gonzalez e Woods, 2018; Pujari et al., 2015). The most common studied descriptors have utilized information about texture (Kaddar, 2017, Montoya-Zegarra et al., 2008), color (Garcia-Lamont et al. 2018, Petrushan et al., 2013,) and shape (Wang et al., 2017, Costa et al., 2011). Two basic techniques of machine learning are available:

supervised and unsupervised methods (Olaode et al., 2014, Saxena, 2017, Ma et al., 2017).

Pujari et al. (2015) required machine learning to utilize techniques of digital processing of images which are independent of the specific crop (fruit, vegetables and cereals), since the lesions arising from the attack of agricultural pests and diseases create a characteristic spectral signature. This can be identified using adequate descriptors. A similar observation was reported by Revathi and Hemalatha (2014), considering the attributes of color, texture and shape for identification of leaf diseases from their spectral signature.

The complexity of the data for agricultural application and the large number of analyses have led to the use of the techniques of machine learning on machines with support of vectors and application of artificial neural networks (Van de Vijver et al., 2020, Bakhshipour and Jafari, 2018). These techniques are robust and versatile and have led to promising results, such as those seen in Niell et al., (2018) and Kumar et al., (2017), who achieved success rates of 100% and 96%, respectively.

Despite the fact that digital image processing in agriculture has considerably advanced (Patrício and Rieder, 2018; Tian et al., 2020), Iqbal et al. (2018) have suggested that we are only at the beginning phases. The development of tools with a better performance in terms of success and fewer computational demands is important so that the challenges of complex scenarios and real time applications can be met.

Although the pests and disease (lesions) symptoms can be detected by the analysis of images, the successful application of this technology in precision agriculture is highly dependent to the spatial resolution and the quality of the images. Lesions caused by fungi, viruses and bacteria can be smaller than 1 cm in diameter and be confused with nutritional deficiencies of the crop. Therefore, the objective of this study was to develop a methodology based on the analysis of images to identify sub-metric lesions on cotton leaves. The implemented algorithm segmented the images and extracted the attributes of texture. A hybrid classifier to detect such lesions on cotton leaves was also developed and evaluated.

Results and discussion

The performance of the classifiers based on fuzzy logic and artificial neural networks for the recognition of objects and the classification of images has improved considerably in the past few years (Altaher, 2017; Belaout et al., 2018; Chlingaryan et al., 2018; Rangarajan et al., 2018). Although unusual, the application of a hybrid classifier for the recognition of lesions on leaves via images is seen as an improvement in the way of dealing with the problem, since most conventional classifiers are binary and allow little or no flexibility (Mohanty et al., 2016).

Global analysis

The performance of the algorithm proposed in this paper was evaluated based on the calculations of confusion matrix (a table showing the quantitative relationship between the images belonging to the class under study and those predicted by the algorithm). The results of the predictions of the algorithm as to the presence of lesions on cotton leaves and their derivations are presented in Tables 1 and 2.

The use of the proposed algorithm is feasible, since it presents an overall accuracy of 71.1% for the identification of lesions on cotton leaves (Chemura et al., 2017). However, this performance is far from the ideal. Mohanty et al. (2016) achieved 99.35% accuracy in the detection of leaf diseases, while Brahimi et al. (2017) achieved 99.18% for the identification of tomato diseases. Zhang et al. (2018) found an overall accuracy above 90% in the identification of damage to soybeans. All of these values are substantially greater than those found here. Despite the fact that the hybrid classifier adopted here is able to make predictions in complex scenarios, inference is still necessary. According to Behmann et al. (2015), the success of classification is directly related to the descriptor of the characteristics utilized. On the other hand, when dealing with the identification of symptoms of stress and disease in plants, both Masood and Khan (2016) and Singh et al., (2016) suggested that the selection of a method for machine learning is a factor for the success of classification.

Although the optimization of mathematical computational models for identification of agricultural pests and diseases provides a glimpse of the overall accuracy, relatively high rates of success do not necessarily mean that the performance of that classifier is good, since situations of over-adjustment of models and/or low generalizability are quite common (Davis and Goadrich, 2006; Zhang et al., 2018; Vasicek, 2019; Yeom et al., 2019).

Krawczyk, (2016) points out that in the use of machine learning, disproportional distribution of the classes in the data set can mask the results of the classifier, causing commission errors (when images attributed to a class belong to a different class), i.e., false positives are common. In Table 2, it can be seen that the false positives (commission errors) substantially affect all of the classes, although the background class has the smallest proportion of false positives. This indicates that the unbalance of the classes minimized the errors of the algorithm for the class of background, i.e., the algorithm is biased for this class, since it has more than 50% of the total set of data (Figure 1).

For certain applications, unbalanced data is to be expected, due to the high cost of obtaining data, low availability of data, or even problems related to labeling in the manual classification of data due to noise. In these cases, the adoption of strategies for dealing with the unbalanced data set it is crucial. The more traditional methods used in the literature include increasing the data artificially by creating new data (Mikolajczyk and Grochowski, 2018; Huang et al., 2019; Cha et al., 2020), resampling the data in the training set (Koziarski and Wozniak, 2017; Sun et al., 2018; Nguyen et al., 2019), and regularization of the classifier (Yuan et al., 2018; Vasicek, 2019). However, in recent years, various proposals have been made for using deep neural networks. This has improved the performance of classification systems (Acharya et al., 2017; Yu et al., 2017; Yuan et al., 2018).

Analysis by class

In this study, we attempted to go beyond an overall analysis of the results to include a study of the behavior of the classifier for each class, since this can identify limitations and the contribution of the individual classes, as well as ignoring the unbalance in data (Saito e Rehmsmeier, 2017; Chelli and Boileau, 2020). In the literature of machine learning, the analysis of classifier behavior includes tests of sensitivity (recall), specificity, accuracy, and precision, as derived from the confusion matrix (Goutte and Gaussier, 2005; Nazarenko et al., 2016). The relevant equations are presented in Table 3.

Table 1. Confusion matrix based on the algorithm for NFA classification.

		Predicted Class		
		Background	Leaf with lesion	Healthy leaf
True class	Background	12903	1525	1237
	Leaf with lesion	1278	3086	390
	Healthy leaf	1782	794	1267

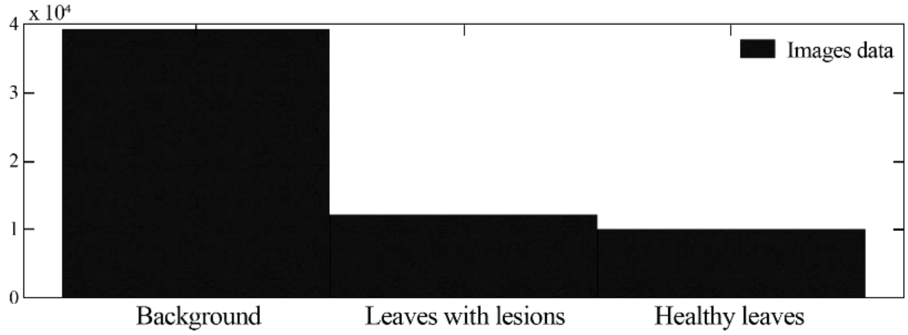

Figure 1. Histogram of the manual labeling of the classes in the dataset. This figure shows an imbalance of data between classes, with the fund class contributing more than 60% of the data, while healthy sheet contributes just over 15% of the data.

Table 2. Breakdown of confusion matrix

Variable	Background	Leaf with Lesion	Healthy leaf
TP	12903	3086	1267
FP	3060	2319	1627
FN	2762	1668	2576
TN	4393	14170	15989
Overall accuracy = 71,1%			

In the table, the following variables are included: true positive (TP), referring to an image correctly predicted to be in the target class; true negative (TN) for images which are correctly predicted as not belonging to the target class; False positive (FP) for images incorrectlypredicted as belonging to the target class, and False Negative (FN) for images belonging to the target class which were predicted to be in another class.

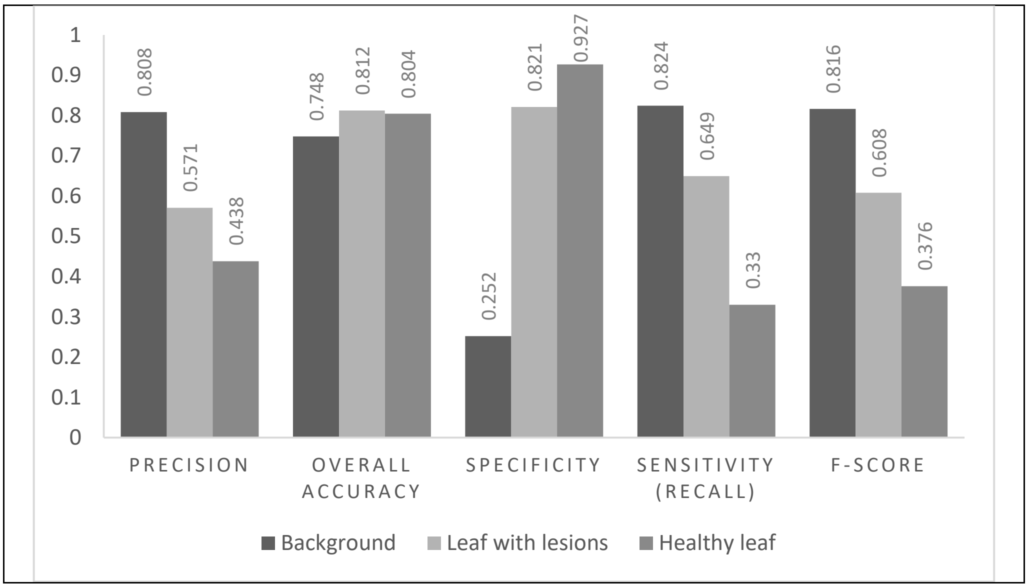

Figure 2. Comparison of performance metrics obtained with the hybrid NFA classifier.

Table 3. Performance measures derived from the confusion matrix.

Performance metric	Equation
Sensitivity (Recall)	TP/(TP+FN)
Specificity	TN/(TN+FP)
Overall accuracy	(TP+TN)/(TP+FP+TN+FN)
Precision	TP/(TP+FP)
F-Score	(2*Precision*Recall)/(Precision+Recall)

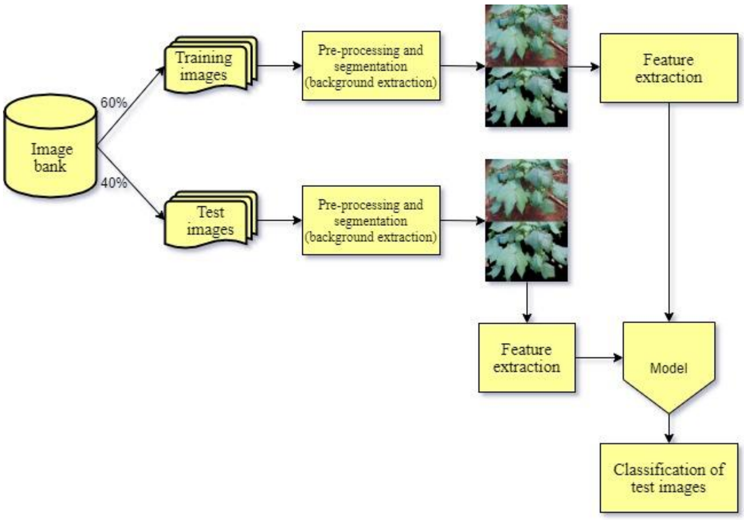

Figure 3. Fluxogram of system of recognition of lesions on cotton leaves.

Table 4. Proposed algorithm

```
//Input: digital image of leaf
//Saída: Diagnosis of the presence of the plant in the image and leaf status (healthy or injured)

//Algorithm for detection of leasions in cotton leaves

Function – diagnosisofplant(){
  //Step 1 – loading of image
  image <- read archive (image)

  //Step 2 – Preparation of image – PDI
  image processed <-  Process digital image (image)

  // Extraction of characteristics – PDI
 vector of characteristic <- extract data (processed image)

  //Step 3 - Analysis of image
  answer[3] <- analyze image (vector of characteristics)

  //Step 4 - Show answers
  print " Confirmation of presence of plant in image":*, answer[1]   //Positive or Negative
  print " status of plant leaf":*, answer[2]                    //Lesioned or healthy
  print " Probability of correct answer (precision)":*, answer[3]   //Precision of analysis
}
Function process_digital image (image)  {
  //Segmentacion of images
  image <- subdivision(image,[102 102])
  //Dessaturation
  image <- convert_RGB_to_gray(imagem)
  // Return  Image processed
  return image_processed
}

Function extract_data(){
  // Descriptors of chareacteristics
  descriptor[N] <- STATXTURE, SIFT, SURF, HOG, PHOW, HAAR,...
  //Extract Characteristics
  for descriptive index from 0 to N {
    vector_characteristic <- vector_characteristic + apply_descriptor[processed image, descriptor[descriptive index]]
  }
  return vector_characteristic
}

  Function analyze_image(vector_characteristic){
    //Load trained data bank - Dataset
    data_base <-read_archive (set_trained_data)
    //recognize Image
    answer[3] <- classify(vector_characteristic, set_trained_data)
    //Return array with answers
    return answer
  }
}
```

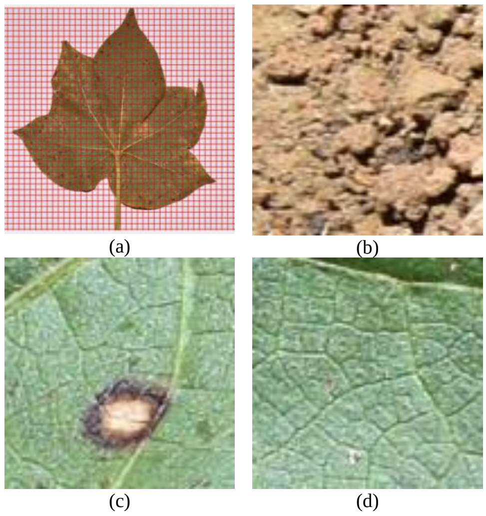

Figure 4. Pre-processing of images and manual identification of patterns: (a) grid for the formation of sub-images with a resolution of 102 x 102 pixels; (b) sub-image of *background*, (c) sub-image of a lesion, and (d) sub-image of healthy area of leaf.

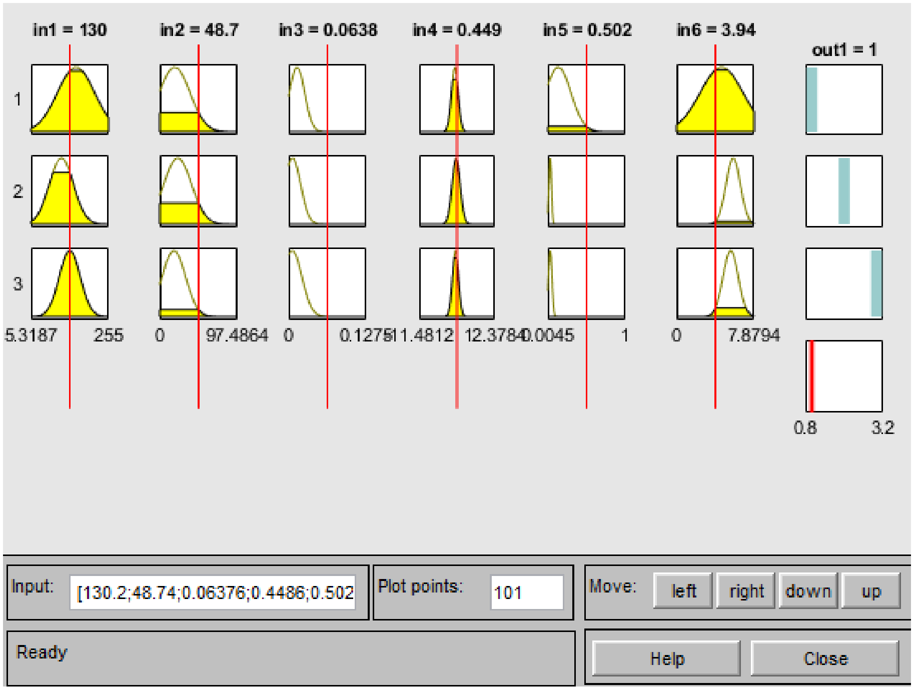

Figure 5. Pertinent functions resulting from the supervised learning of the NFA classifier. The adopted adaptive neuro-fuzzy classifier uses language barriers capable of changing the primary meaning of membership functions to a secondary meaning.

The performance metrics used in the analysis of the behavior of machine learning are defined as follows. Precision is the ratio of the images correctly classified as belonging to a class in relation to all the images predicted to be of this class. Accuracy is the proportion of images identified correctly as to class. Sensitivity or recall measures the fraction of the actual number of images in a class correctly identified as being of that class (a value which makes it possible to identify the class for which the method is least sensitive). Specificity measures the ratio of the images correctly identified as not belonging to a class in relation to those actually not in that class. The F-Score makes an equilibrated analysis of the system possible, whether or not the data are balanced, since it involves the relation between the actual members of the class as identified (precision) and those identified as being of that class in relation to actual number existing (recall). Figure 2 compares the classes in relation to the values obtained for sensitivity, specificity, precision and accuracy, as well as F-Score, using the hybrid adaptive neuro-fuzzy classifier (NFA) algorithm for the recognition of lesions on cotton leaves.

As shown by the confusion matrix (Tables 1 and 2) and reinforced by the grouping of the data (Figure 3), it seems that the proposed system is most accurate for identification of background images (80.8%), i.e., the predictions for this class present low dispersion. However, for the predictions indicating whether or not the leaves are healthy, the results for precision and the F score are less than the 71.1% obtained for overall accuracy (Table 2).

The equilibrated analysis of data shows that images with regions showing background information are correctly identified 25% more of the time than those with information about lesions and 54% more than those of healthy leaves. This compromises the use of the system for the identification of lesions in images of cotton leaves in complex scenarios.

It is the overall accuracy which measures how promising a system is for predicting the identification of classes in the images, since in addition to true positives it also considers true negatives. This value suggests that the system proposed is indeed promising for the identification of lesions on leaves, as well as healthy regions and background information, with rates of 81.2%, 80.4%, and 74.8%, respectively.

Materials and Methods

Based on the theory of the processing of images presented in Gonzalez and Woods (2018), the proposed system is divided into the stages of image acquisition, pre-processing, attribute extraction, training, and classification.

Image acquisition

The images of cotton leaves, both healthy and non-healthy (those with lesions) were captured using a digital Nikon camera (Model D5500) with an automatic setting. Images in the JPEG format, RGB color space, and a resolution of 4000 x 6000 pixels were selected and stored in a digital data bank for processing and attribute extraction.

The system presented here considers statistical attributes of texture to describe the classes of leaves (healthy or with lesions). A hybrid classifier obtained by the integration of fuzzy logic with artificial neural networks was implemented (Cetişli, 2010, Azar et al., 2015). Table 4 describes the algorithm proposed, followed by the fluxogram of the classification process (Figure 3).

The implementation of image processing and machine learning algorithms was performed using the Toolbox of Image Processing and Computer Vision and Toolbox of Machine Learning of the Matlab R2018a (Mathworks) software, installed on an HP Z800 computer with two Intel® Xeon® X5650 Cache processors. 12M, 2.66GHz, 6.40 GT/s Intel® QPI, 128Gb Ram, Nvidea Quadro FX 3800 graphics based on Windows 10 operating system.

Pre-processing

The pre-processing of the system proposed here consists of two steps. In the first, an attempt is made to identify lesions on cotton leaves, whatever the size in pixels. For this, each of the images stored in the data bank was divided into sub-images of 102 x 102 pixels.

In the second step, the sub-images were grouped into three classes: Class I sub-images (SBC1), Class 2 sub-images (SBC2) and Class 3 sub-images (SBC3). SBC1 corresponding to the information about soil and straw and any other background elements other than leaves; SBC2 corresponded to the area of leaves with lesions caused by agricultural pests and disease, and SBC3 corresponded to areas of healthy leaves (Figure 4).

Attribute extraction

Given the characteristics of the lesions and the promising results found in various studies of image processing (Hlaing e Zaw, 2017; Patil and Kumar, 2017; Pires et al., 2016), the extractors utilized here characterized the texture based on statistical measurement. The simplest aspect (primitive) of a digital image in grayscale space is a pixel, with the concentration of gray varying as a function of the depth of color of the image (Gonzalez et al., 2003).

The distribution of shades of gray of the pixels, or a histogram of these shades, defines the texture of the image. In this paper the statistical averages of the average texture of the shade of grey, the standard deviation, correlation, third moment, uniformity, and entropy are considered.

The average (m) refers to the average intensity of each region, despite the texture as such in not described.

$$m = \sum_{i=0}^{L-1} z_i \, p(z_i) \qquad \text{Equation 1}$$

Where Z_i is a random variable indicating light intensity; $p(z_i)$ is the histogram of the level of intensity of the region, and L is the possible number of levels or shades of grey for each pixel.

The standard deviation (σ) corresponds to information about the average texture of contrast, i.e., the average of the variation in the shade of grey of the image, obtained from the square root of the "second moment" (μ_2) or variance.

$$\sigma = \sqrt{\mu_2(z)} = \sqrt{\sigma^2} \qquad \text{Equation 2}$$

The attribute R measures the relative softness of the shades of grey in a region. R is 0 for a region with a constant intensity and 1 for those with great variation in grayscale. For practical purposes, the variance used in the measurement is normalized between 0 and 1.

$$R = 1 - \frac{1}{1+\sigma^2} \qquad \text{Equation 3}$$

The "third moment" (μ_3) measures the asymmetry of the histogram. This attribute is 0 for symmetrical histograms, positive for those which are symmetrical, but dislocated to the right (above average) and negative for those dislocated to the left. The resultant values are normalized between 0 and 1.

$$\mu_3 = \sum_{i=0}^{L-1} (z_i - m)^3 \, p(z_i) \qquad \text{Equation 4}$$

The measure of uniformity is at its maximum when all the shades of grey are the same (maximum uniformity) and decreases as variations in shades of gray are introduced.

$$U = \sum_{i=0}^{L-1} p^2(z_i) \qquad \text{Equation 5}$$

Entropy furnishes dispersion, i.e., randomness in the level of the grey in the image.

$$e = -\sum_{i=0}^{L-1} p(z_i) log_2 \, p(z_i) \qquad \text{Equation 6}$$

Classification

The process of classification begins by the random division of the images into two sub-sets, one for training (60% of the images) and the other for testing (40% of the images).

In this paper, the adaptive neuro-fuzzy classifier proposed by Cetişli (2010) was used. The innovation of this kind of classifier results from its hybrid nature, since it combines the flexibility of fuzzy logic with the speed and adaptability of artificial neural networks (Ghosh et al., 2009; Pradhan, 2013; Khoshnevisan et al., 2014).

The adaptive Neuro-fuzzy classifier adopted here utilizes linguistic barriers capable of reducing the principal significance of pertinent functions (Figure 5).

To improve the meaning of the fuzzy rules and the accuracy of the classifier, a layer defining linguistic barriers was added to the neural network proposed. The linguistic barriers were trained with other parameters of the network using a conjugated gradient training algorithm, with the values of the linguistic barriers tuned (syntonized) to those of the fuzzy sets, thus making the sets more flexible and improving the rate of distinction from overlying classes (Cetişli, 2010).

The evaluation of the technical applicability of the proposed algorithm for the identification of lesions on cotton leaves was made using an analysis of the confusion matrix and other metrics derived from it, such as overall accuracy, precision, sensitivity, specificity, and F-Score.

Conclusion

A system using an adaptive neuro-fuzzy classifier for the processing of digital images based on data of texture was developed (https://github.com/rafaeufg/Cotton-diseases). A hit rate of 71.1% of correct identifications was obtained, a performance considered satisfactory for use as a tool to assist making decisions for integrated management of a cotton crop in presence of agricultural pests and disease. The system proved to be sensitive in relation to the identification of background information. It was least sensitive for healthy areas of the leaves. The performance for images containing areas with lesions was better in terms of overall accuracy.

Acknowledgements

The authors would like to thank the Brazilian National Council for Scientific and Technological Development (CNPq) for the grant (process 140719/2017-2) awarded to the research worker R.F.C. during the execution of the research.

References

Acharya UR, Oh SL, Hagiwara Y, Tan JH, Adam M, Gertych A, Tan RS (2017) A deep convolutional neural network model to classify heartbeats. Comput Biol Med. 89, 389–396.

Allen T, Bradley C, Sisson AJ, Byamukama E, Chilvers M, Coker CM, Collins AA, Damicone JP, Dorrance A, Dufault N, Esker P, Faske TR, Giesler L, Grybauskas AP, Hershman D, Hollier C, Isakeit T, Jardine D, Kelly H, Wrather J (2017) Soybean yield loss estimates due to diseases in the United States and Ontario, Canada, from 2010 to 2014. Plant Health Prog. 18, 19–27.

Altaher A (2017) An improved Android malware detection scheme based on an evolving hybrid neuro-fuzzy classifier (EHNFC) and permission-based features. Neural Comput Appl. 28, 4147–4157.

Azar AT, Hassanien AE (2015) Dimensionality reduction of medical big data using neural-fuzzy classifier. Soft Comput. 19, 1115–1127.

Bakhshipour A, Jafari A (2018) Evaluation of support vector machine and artificial neural networks in weed detection using shape features. Comput Electron Agric. 145, 153–160.

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

Behmann J, Mahlein A-K, Rumpf T, Römer C, Plümer L (2015) A review of advanced machine learning methods for the detection of biotic stress in precision crop protection. Precis Agric. 16, 239–260.

Belaout A, Krim F, Mellit A, Talbi B, Arabi A (2018) Multiclass adaptive neuro-fuzzy classifier and feature selection techniques for photovoltaic array fault detection and classification. Renew. Energ. 127, 548–558.

Bowen KL, Hagan AK, Pegues M, Jones J, Miller HB (2018) Epidemics and Yield Losses due to Corynespora cassiicola on Cotton. Plant Dis. 102, 2494–2499.

Brahimi M, Boukhalfa K, Moussaoui A (2017) Deep Learning for Tomato Diseases: Classification and Symptoms Visualization. Appl. Artif Intell. 31, 299–315.

Camargo A, Smith JS (2009) Image pattern classification for the identification of disease causing agents in plants. Comput Electron Agr. 66, 121–125.

Cetisli B (2010) Development of an adaptive neuro-fuzzy classifier using linguistic hedges: Part 1. Expert Syst Appl. 37, 6093–6101.

Cha KH, Petrick N, Pezeshk A, Graff CG, Sharma D, Badal A, Sahiner B (2020) Evaluation of data augmentation via synthetic images for improved breast mass detection on mammograms using deep learning. J Med Imaging (Bellingham) 7, 12703.

Chelli M, Boileau P (2020) Let's "recall" sensitivity and specificity: Letter to the Editor regarding Gowd et al: "Construct validation of machine learning in the prediction of short-term postoperative complications following total shoulder arthroplasty. J Shoulder Elb Surg. 29, 98–99.

Chemura A, Mutanga O, Dube T (2017) Separability of coffee leaf rust infection levels with machine learning methods at Sentinel-2 MSI spectral resolutions. Precis Agric. 18, 859–881.

Chlingaryan A, Sukkarieh S, Whelan B (2018) Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. Comput Electron Agr 151, 61–69.

Costa C, Antonucci F, Pallottino F, Aguzzi J, Sun DW, Menesatti P (2011) Shape Analysis of Agricultural Products: A Review of Recent Research Advances and Potential Application to Computer Vision. Food Bioprocess Tech. 4, 673–692.

Davis J, Goadrich M (2006) The Relationship between Precision-Recall and ROC Curves. In Proc. 23rd Int. Conf. Mach. Learn., New York, NY, USA. ICML '06. 233–240.

Garcia-Lamont F, Cervantes J, López A, Rodriguez L (2018) Segmentation of images by color features: A survey. Neurocomputing 292, 1–27.

Ghosh A, Uma Shankar B, Meher SK (2009) A novel approach to neuro-fuzzy classification. Neural Netw. 22, 100–109.

Gonzalez RC, Woods RE, Eddins SL (2003). *Digital Image Processing Using MATLAB*. New Jersey, Prentice Hall.

Gonzalez RC, Woods RE, (2018). *Digital Image Processing*. New York City:Person, 4 rd.

Goutte C., Gaussier E. (2005) A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. In: Losada D.E., Fernández-Luna J.M. (eds) Advances in Information Retrieval. ECIR 2005. Lecture Notes in Computer Science, vol 3408. Springer, Berlin, Heidelberg.

Guillemaud T, Bourguet Denis (2016) The hidden and external costs of pesticide use. Sustainable Agriculture Reviews 19: 35-120. Sustain Agric Res. 19, 35–120.

Hlaing CS, Zaw SMM (2017) Plant diseases recognition for smart farming using model-based statistical features. 2017 IEEE 6th Global Conference on Consumer Electronics, GCCE 2017, 1–4.

Huang J, Zeng M, Li W, Meng X (2019) Application of Data Augmentation and Migration Learning in Identification of Diseases and Pests in Tea Trees Written for presentation at the 2019 ASABE Annual International Meeting Sponsored by ASABE. 2019 ASABE-AIM American Society of Agricultural and Biological Engineers Annual Inrenational Meeting.

Iqbal Z, Khan MA, Sharif M, Shah JH, ur Rehman MH, Javed K (2018) An automated detection and classification of citrus plant diseases using image processing techniques: A review. Comput Electron Agr. 153, 12–32.

Kaddar B, Fizazi H, Boudraa AO (2018) Texture features based on an efficient local binary pattern descriptor. Comput Electr Eng. 70, 496–508.

Khoshnevisan B, Rafiee S, Omid M, Mousazadeh H (2014) Prediction of potato yield based on energy inputs using multi-layer adaptive neuro-fuzzy inference system. Measurement 47, 521–530.

Koziarski M, Wozniak M (2017) CCR: A combined cleaning and resampling algorithm for imbalanced data classification. Int J Ap Mat Com-Pol. 27, 727–736.

Krawczyk B (2016) Learning from imbalanced data: open challenges and future directions. Prog. Artif. Intell. 5, 221–232.

Kumar S, Mishra S, Khanna P, Pragya (2017) Precision Sugarcane Monitoring Using SVM Classifier. Procedia Comput. Sci. 122, 881–887.

Ma L, Li M, Ma X, Cheng L, Du P, Liu Y (2017) A review of supervised object-based land-cover image classification. ISPRS J Photogramm, 130, 277–293.

Masood R, A. Khan S, N.A. Khan M (2016) Plants Disease Segmentation using Image Processing. Int. J. Mod. Educ. Comput. Sci. 8, 24–32.

Mikołajczyk A, Grochowski M (2018) Data augmentation for improving deep learning in image classification problem. In: 2018 International Interdisciplinary PhD Workshop (IIPhDW), p. 117-122.

Mohanty SP, Hughes DP, Salathé M (2016) Using deep learning for image-based plant disease detection. Front Plant Sci. 7, 1–10.

Montoya-Zegarra JA, Papa JP, Leite NJ, Da Silva Torres R, Falc̃o AX (2008) Learning how to extract rotation-invariant and scale-invariant features from texture images. Eurasip J Adv Sig Pr. 691924.

Mutka AM, Bart RS (2015) Image-based phenotyping of plant disease symptoms. Front Plant Sci. 5, 1–8.

Nazarenko D V., Kharyuk P V., Oseledets I V., Rodin IA, Shpigun OA (2016) Machine learning for LC-MS medicinal plants identification. Chemometr Intell Lab. 156, 174–180.

Nguyen S, Niu G, Quinn J, Olinsky A, Ormsbee J, Smith R, Bishop J (2019) Detecting Non-injured Passengers and Drivers in Car Accidents: A New Under-resampling Method for Imbalanced Classification. In: Advances in Business and Management Forecasting pp. 93–105.

Niell S, Jesús F, Díaz R, Mendoza Y, Notte G, Santos E, Gérez N, Cesio V, Cancela H, Heinzen H (2018) Beehives biomonitor pesticides in agroecosystems: Simple chemical and biological indicators evaluation using Support Vector Machines (SVM). Ecol Indic. 91, 149–154.

Olaode A, Naghdy G, Todd C (2014) Unsupervised Classification of Images: A Review. Int. J. Image Processing 8, 2014–2325.

Patil JK, Kumar R (2017) Analysis of content based image retrieval for plant leaf diseases using color, shape and texture features. Eng Agric Environ. Food 10, 69–78.

Patrício DI, Rieder R (2018) Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. Comput Electron Agr. 153, 69–81.

Pelzer E, Fortino G, Bockstaller C, Angevin F, Lamine C, Moonen C, Vasileiadis V, Guérin D, Guichard L, Reau R, Messéan A (2012) Assessing innovative cropping systems with DEXiPM, a qualitative multi-criteria assessment tool derived from DEXi. Ecol Indic 18, 171–182.

Peshin R, Dhawan AK (2009). Integrated pest management. Vol.1. 10.1007/978-1-4020-8992-3_1.

Petrushan M, Vermenko Y, Shaposhnikov D, Anishenko S (2013) Comparative analysis of color- and grayscale-based feature descriptions for image recognition. Pattern Recognit. Image Anal. 23, 415–418.

Pires RDL, Gonçalves DN, Oruê JPM, Kanashiro WES, Rodrigues JF, Machado BB, Gonçalves WN (2016) Local descriptors for soybean disease recognition. Comput Electron Agr. 125, 48–55.

Pradhan B (2013) A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. COMPUT GEOSCI 51, 350–365

Pujari JD, Yakkundimath R, Byadgi AS (2015) Image Processing Based Detection of Fungal Diseases in Plants. Procedia Comput. Sci. 46, 1802–1808.

Rangarajan AK, Purushothaman R, Ramesh A (2018) Tomato crop disease classification using pre-trained deep learning algorithm. Procedia Comput. Sci. 133, 1040–1047.

Revathi P, Hemalatha M (2014) Identification of Cotton Diseases Based on Cross Information Gain _ Deep Forward Neural Network Classifier with PSO Feature Selection. 5, 4637–4642.

Saito T, Rehmsmeier M (2017) Precrec: Fast and accurate precision-recall and ROC curve calculations in R. Bioinformatics 33, 145–147.

Santos, WJ. (2015) Manejo das pragas do algodão, com destaque para o cerrado brasileiro. In: FREIRE, E. C. (Ed.). Algodão no cerrado do Brasil. 3. ed. Brasília, DF: Associação Brasileira dos Produtores de Algodão. p. 267-364.

Saxena A, Prasad M, Gupta A, Bharill N, Patel OP, Tiwari A, Er MJ, Ding W, Lin CT (2017) A review of clustering techniques and developments. Neurocomputing 267, 664–681.

Singh A, Ganapathysu B, Singh AK, Sarkar S (2016) Machine Learning for High-Throughput Stress Phenotyping in Plants. Trends Plant Sci. 21, 110–124..

Sun L, Song J, Hua C, Shen C, Song M (2018) Value-Aware Resampling and Loss for Imbalanced Classification. In: CSAE '18: Proceedings of the 2nd International Conference on Computer Science and Application Engineering, 1-6.

Tian H, Wang T, Liu Y, Qiao X, Li Y (2020) Computer vision technology in agricultural automation —A review. Inf Process Agric. 7, 1–19.

Van de Vijver R, Mertens K, Heungens K, Somers B, Nuyttens D, Borra-Serrano I, Lootens P, Roldán-Ruiz I, Vangeyte J, Saeys W (2020) In-field detection of Alternaria solani in potato crops using hyperspectral imaging. Comput Electron Agr. 168, 105106.

Vasicek D (2019) Artificial intelligence and machine learning: Practical aspects of overfitting and regularization. Inf Serv Use, 39, 281–289.

Veisi H (2012) Exploring the determinants of adoption behaviour of clean technologies in agriculture: A case of integrated pest management. Asian Journal of Technology Innovation - Asian J Technol Innov. 20, 67–82.

Wahabzada M, Mahlein A-K, Bauckhage C, Steiner U, Oerke E-C, Kersting K (2015) Metro Maps of Plant DiseaseDynamics—Automated Mining of Differences Using Hyperspectral Images. PLoS ONE 10, e0116902.

Wang Z, Li H, Zhu Y, Xu T (2017) Review of Plant Identification Based on Image Processing. Arch Comput Method E. 24, 637–654.

Yeom S, Giacomelli I, Menaged A, Fredrikson M, Jha S (2019) Overfitting, robustness, and malicious algorithms: A study of potential causes of privacy risk in machine learning. J. Comput. Secur. 28, 35–70.

Yu X, Wu X, Luo C, Ren P (2017) Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework. Gisci Remote Sens. 54, 1–18.

Yuan X, Xie L, Abouelenien M (2018) A regularized ensemble framework of deep learning for cancer detection from multi-class, imbalanced training data. Pattern Recognit. 77, 160–172.

Zhang C, Vinyals O, Munos R, Bengio S (2018) A Study on Overfitting in Deep Reinforcement Learning. 1–25. arxiv

Zhang S, Wu X, You Z, Zhang L (2017) Leaf image based cucumber disease recognition using sparse representation classification. Comput Electron Agr. 134, 135–141.