Available online at http://www.journalijdr.com

IJDR

ISSN: 2230-9926

International Journal of Development Research
Vol. 08, Issue, 10, pp. 23547-23552, October, 2018

ORIGINAL RESEARCH ARTICLE

OPEN ACCESS

ALGORITHM TO DETECT NITROGEN FOLIAR DEFICIENCY IN BEAN CROPS APPLYING DIGITAL IMAGE PROCESSING AND DATA MINING

***Mollo Neto, Mario, Rezende Filho, João Paulo and Putti, Fernando Ferrari**

São Paulo State University (UNESP), School of Sciences and Engineering, Tupã Rua Domingos da Costa Lopes, 780 - Jd. Itaipu - Tupã/SP, Brasil - CEP 17602-496

ARTICLE INFO

Article History:
Received 14th July, 2018
Received in revised form
21st August, 2018
Accepted 09th September, 2018
Published online 30th October, 2018

Key Words:

Data Mining, Image Processing,
Precision Farming,
Vigna unguiculata (L.) Walp.

ABSTRACT

The determination of the Nitrogen correct application in bean crops, aiming to increase productivity and reduce costs with fertilizers, is a great challenge, which has a solution based on the use of precision farming tools. This work proposes the use of digital image processing techniques coupled with a data mining process aiming at the construction of a decision tree based algorithm for the definition of dose for the maximum production Nitrogen dosage to be applied by means of cover fertilization at focal points of the beans cultures. From the digital images collection of the bean plants leaves, these are processed obtaining the color channels RGB. From these images, the histograms are extracted, to be compared with the reference histograms already obtained from other witnesses' sheets with various degrees of known nitrogen deficiency from a set of previously prepared and studied portions. The database generated through the comparison between these histograms was processed in a data mining tool using the classification algorithm J48 (C4.5) and, the process allowed the construction of the classification tree. In the obtained results it was possible to classify the Nitrogen dosage to be applied to the focal points of the cultures under analysis, with 76.33% accuracy.

Citation: Mollo Neto, Mario, Rezende Filho, João Paulo and Putti, Fernando Ferrari. 2018. "Algorithm to detect nitrogen foliar deficiency in bean crops applying digital image processing and data mining", *International Journal of Development Research*, 8, (10), 23547-23552.

INTRODUCTION

In the work of MANCUSO *et al.* (2016), it is noteworthy that the *Vigna unguiculata (L.) Walp* beans are of great socioeconomic importance for the north and northeast regions of Brazil, essentially due to its adaptation to edaphoclimatic conditions and present a relatively short cycle, between 60 and 80 days, and develop well under conditions of low water availability and low soil fertility, its cultivation has expanded in a large scale in the center-west region. According to BERTOLDO *et al.* (2015), the soils of tropical regions are generally deficient in nitrogen (N), due to the low organic matter content, being considered one of the limiting factors for agricultural production. For this reason, the complementation is necessary by means of cover fertilization. The growing concern with the more efficient management of agribusiness inputs and outputs has led to the development of a new concept of agriculture, called Precision Agriculture (PA) (PINHO *et al.,* 2015). From this context of PA, the use of optical sensors, as a technological tool, is one of the alternatives to the traditional N recommendation as indicated in the research of SAMBORSKI *et al.*(2016), where it is emphasized that nitrogen management strategies developed based on small-scale plot research are not always meaningful for large-scale farm conditions using sensor-based systems incorporating active optical sensor (AOS), a GPS receiver, data loggers and appropriate software, allow for very intensive data acquisition. Also, the image processing is another technology that advances in stride in this field of research (PA), as can be seen in the work of GABRIEL *et al.* (2017) where the authors say that some rapid and non-destructive ways to obtain multiple measurements are optical readings that can provide indicators of crop nutritional status, and the greenness of plants is strongly related to leaf chlorophyll content and to N status, so it has been used as an indicator of N availability. SIMÕES *et al.* (2015), studies on the parameterization of morphophysiological attributes of plant canopies, made with methods based on the optical properties

***Corresponding author:* Mollo Neto, Mario,**
Orcid: 0000-0002-8341-4190 São Paulo State University (UNESP), School of Sciences and Engineering, Tupã Rua Domingos da Costa Lopes, 780 - Jd. Itaipu - Tupã/SP, Brasil - CEP 17602-496

of plants, have gained prominence. The authors point out that the analytical processing of the images captured with the cameras allows the construction of several indexes, related to the vegetation cover, the accumulation of biomass, and the nitrogen contents in several cultivated species. They also say that access to this information has traditionally been done by cutting and drying the biomass, followed by laboratory analysis, which complements that such procedures, although precise, have limitations on costs and execution time, which highlights the importance of indirect methodologies such as image processingevaluations in real time, such as the use of information obtained from digital cameras. Another important contribution is the application of the techniques of Machine-learning in the live stock production, techniques among which stands out the one known as KDD (Knowledge Discovery in Databases).The KDD process, as statedby PÖLÖNEN *et al.* (2013) is the joining of several consecutive and interconnected processes, which enables the researcher to obtain a knowledge containing a database with numerous data contained in it, guaranteeing a high precision in its response. Within this process is contained another technique called "Data Mining". The use of Data Mining technique can improve the findings of knowledge as stated by MOI *et al.* (2014). LIMA *et al.* (2017), describe that among the data mining techniques, decision tree induction is a method very simple and efficient because it allows the classification of “datasets” consisted of numerical and categorical variables. In addition, the knowledge found in a decision tree is represented by rules and the algorithm achieves satisfactory results when compared with other more sophisticated approaches available in the literature. The authors also say that a decision tree is a flowchart-like structure that shows the various outcomes from a series of decisions. It is formed of nodes connected by branches and leaves and, the main advantages of using a decision tree include the support in a decision-making process, which considers the most relevant attributes, and the facility of interpretation and understanding of the results because the classification is obtained explicitly, simplifying its interpretation and allowing users to know which attributes influence the variables. They also state that in addition, the results are usually supplied quickly due to the computational efficiency as explored in the research of DAI *et al.* (2016). The aim of this research was to organize the data collected by cameras installed in the crop to enable the construction of a decision tree-based algorithm for the definition of a correct Nitrogen dosage to be applied by means of cover fertilization at focal points of the beans *Vigna unguiculata (L.) Walp* cultures.

MATERIAL AND METHODS

Sampling: The experiment for collecting data on N levels was carried out from February to April 2017, in the experimental greenhouse of the Science and Engineering Faculty of UNESP Campus of Tupã / SP, located at latitude 21o 55 '39' 'S, 50o 29 '30' 'W, altitude 495 m, with humid temperate climate with dry winter and hot summer. The culture was sown in *the Vigna unguiculata (L.) Walp*, in which the experiment was conducted in a randomized block, with the treatments being T0 = 0 kg of $N.ha^{-1}$, T1 = 25 kg of $N.ha^{-1}$, T2 = 50 kg of $N.ha^{-1}$, T3 = 75 kg of $N.ha^{-1}$, T4 = 100 kg of $N.ha^{-1}$, T5 = 125 kg of $N.ha^{-1}$, with 4 replicates. The fertilization was carried out along the cycle with 4 applications according to the recommendation of the agronomic institute of the Secretariat of Agriculture and Supply of the State of São Paulo. The culture planting *Vigna unguiculata (L.) Walp* in an area covered with the straw of peanut bark to avoid the growth of weeds, sown at 0.50 m spacing, and approximately 12 to 18 viable seeds per linear meter of the furrow. The analysis of soil samples from the portions, carried out in the laboratory for agronomic analysis of the Shunji Nishimura Foundation of Technology in the city of Pompéia/SP, and the methodology for extraction of P, K, Ca, Mg - Resin was used; for S-SO4 - Calcium Phosphate; Fe, Mn, Zn and Cu-DTPA-TEA and for B, hot water.

The initial characteristics of the soil where the bean seeds were cultivated (T0), obtained from the analysis of soil, were as follows:

pH $CaCl_2$ = 4,60; pH H_2O= 5,40; MO = 7,00 $g.dm^{-3}$; P = 6,00 $mg.dm^{-3}$; K = 0,90 $mmolc.dm^{-3}$; Ca = 5,00 $mmolc.dm^{3-3}$; Mg = 3,00 $mmolc.dm^{-3}$; Al = 4,00 $mmolc.dm^{-3}$; H+Al = 16,00 $mmolc.dm^{-3}$; SB = 9,00 $mmolc.dm^{-3}$; T = 25,00 $mmolc.dm^{-3}$; V% = 36,00; S = 3,00 $mg.dm^{-3}$; Fe = 11,00 $mg.dm^{-3}$; Mn = 11,60 $mg.dm^{-3}$; Zn = 0,20 $mg.dm^{-3}$; Cu = 0,50 $mg.dm^{-3}$; B = 0,07 $mg.dm^{-3}$.The relations were: Ca/Mg = 1,7; Ca/K= 5,6 and Mg/K = 3,3.

The portions had the dimensions of 4 m x 1 m. Before the sowing, a composite soil sample was taken for analysis. The factorial design composed of 9 treatment (3x3) separated by 1-meter intervals was prepared. For each portion, different rates of N were tested during fertilization.

Data Collection

The images of the bean leaves of the portions treated with different N rates were captured through a camera with SONY CCD (Charge Coupled Device) 1/3 "with a resolution of 1000 lines of TV (Horizontal) with Gamma = 0.45 and SNR greater than 48dB, generating the digitization of the frames corresponding to the leaves of each portion. The images of each bean treatment were obtained using different doses of N, varying between 0 and 125 kg $N.ha^{-1}$. The amount of image collected was 15 frames of images for each of the treatments and replication. From the data collection, a file was generated containing all the collected images of the bean, being a total of 360 images. These scanned frames were processed in the application software developed in the MATLAB programming environment, with the development of algorithms that allowed, through the coloring of the leaves and the separation of their formation channels (RGB). To make the identification of colors and to construct their respective histograms that in the future will be compared to the histograms of the samples of the parcels to determine the classification of the candidate images as to their N content in the leaves for the proposals of complementation of the fertilization.

In the sequence, the developed MATLAB algorithm generated the desired histograms to the comparison,following as recommendations of the toolbox.The image processing mechanism was subdivided into four stages: acquisition of image frames, preprocessing, color separation of digital images, and construction of the histograms for the comparisons. After constructing the histograms, the difference between the histograms is found, considering as analysis pairs two by two, always subtracting the histogram that received the Nitrogen reinforcement treatment from the histogram control, which did not receive extra nitrogen. These difference values corresponding to the number of pixels present in the captured image of the nitrogen addition treatment sheet for each of the 256 tones of the study color channel subtracted from the values of the amounts of pixels present in the captured image of the untreated sheet. The resulting pixel count was used as the predictive attributes for classification in the data mining process. The original data set generated to the data mining process was composed by 4 attributes (3 predictive attributes and one target-attribute or response variable), which were added to the set of data, totaling 300 instances for each attribute. To summarise, the description ofthe studied predictive attributes and target attribute are: The first predictive attribute is called "R", the number of pixels of the treated Red channel subtracted from the number of pixels of the Red channel without treatment;

The second predictive attribute is called "G", the number of pixels of the treated Green channel subtracted from the number of pixels of the Green channel without treatment;

The third and the last elected predictive attribute is called "B", The number of pixels of the treated Blue channel subtracted from the number of pixels of the Blue channel without treatment and;

Finally, the target attribute is called "Classification" (%). The target attribute refers to the Nitrogen dosage and is the target of the classification. For identifying correct Nitrogen dosage data was processed using the software Weka® 3.5 Freeware, from The University of Waikato, New Zealand using the classification induction algorithm C4.5, known as J48 to generate the decision tree. The methodology adopted for the mining process for classification was the same one for Induction and validation adopted by LIMA *et al.* (2017).

Measurement

The induced models with a variation in the number of instances (or observations), per leaf, were assessed using the evaluate on training data test mode. The selection of the best model was made based on the measures: accuracy; the number of leaves (rules) and the *Kappa* coefficient. As a result of the induction of the decision tree model, a data analyst obtains the confusion matrix of 2x2 size, which according to HAN *et al.* (2011) is widely used in statistical analysis of agreement. In the confusion matrix, the observed class A leads to a predicted class A = TP, class B = FN with the total = P; the observed class B leads to a predicted class A = FP, class B = TN with the total = N; the observed Total leads to a predicted class A = P', class B = N' with the total = P+N; From the confusion matrix, according to LIMA *et al.* (2017), it is possible to get the measures of performance evaluation. As stated by LIMA *et al.* (2017), accuracy is the percentage of examples that were correctly classified by the classifier and can be expressed as in Equation 1.

$$Acuracy = (TP + TN).(P + N)^{-1} \quad \text{..............Eq. 1}$$

The *Kappa* coefficient is used to describe the measure of agreement between the predicted and observed classes. Such a coefficient ranges from 0 to 1, representing poor and excellent ranking results, respectively. It can be defined by Equation 2 (LIMA *et al.,* 2017) as follows (WITTEN *et al.,* 2011), where Pr(a) is a relative agreement observed for a given class in the confusion matrix; Pr(e) is the probability of the expected agreement in this same class.

$$K = \Pr(a) - \Pr(e).(1 - \Pr(e))^{-1} \quad \text{..............Eq. 2}$$

The *Kappa* coefficient is calculated considering all the classes available in a "dataset" (LIMA *et al.,* 2017). A possible interpretation of model's performance from the *Kappa* statistic method was introduced by Landis and Koch (1977) and LIMA *et al.* (2017), follow the classes: *Kappa* statistics< 0.00 Quality Very bad; *Kappa* statistics 0.00 – 0.20 Quality bad; *Kappa* statistics 0.21 – 0.40 Quality Average; *Kappa* statistics0.41 – 0.60 Quality Good; *Kappa* statistics 0.61 – 0.80 Quality Very good; *Kappa* statistics0.81 – 1.00QualityExcellent.

For data classification, as in the research of LIMA *et al.* (2017), are used the method of binary decision tree available in the Weka software 3.6 (WITTEN *et al.,* 2011). The induction algorithm used was the J48, widely known as C4.5 and developed by Quinlan (1993). Tree pruning techniques were also used to reduce the number of internal nodes, generating smaller and less complex trees and, therefore, easier to be understood as indicated in the LIMA *et al.* (2017) article. The comparison of the results obtained for the *Kappa* indicators allowed to evaluate the performance of the algorithm generated from the decision tree obtained.

RESULTS

For the study parcels, complementary N doses were added in proportions of T0 = 0 kg of N.ha^{-1}, T1 = 25 kg of N.ha^{-1}, T2 = 50 kg of N.ha^{-1}, T3 = 75 kg of N.ha^{-1}, T4 = 100 kg of N.ha^{-1}, T5 = 125 kg of N.ha^{-1}. Soon after the procedures of image collection according to MURAKAMI *et al.* (2005) in the instructional guide for leaf color analysis, the images are processed in the application software developed in the MATLAB programming environment, in a personal computer. The separation of color channels are reached using a 3 dimensions array, of size [x,y,3] were x and y are the dimensions of the collected image and 3 is for the R, G and B components. The set of base instructions used for coding in the MATLAB environment was:

R = reshape(I(:,:,1),[],1);G = reshape(I(:,:,2),[],1);B = reshape(I(:,:,3),[],1);

To obtain the construction of histograms with the intensity image I and the values for each one of the 256 bins, are followed the recommendations of the image processing toolbox. In the MATLAB environment, the set of base instruction used was: imhist(I). The histograms, for each of the color channels, were then subtracted by applying the following procedure recommended in the image processing toolbox. The MATLAB set of base instructions used were:

a=imread('image1'); b=imread('image2'); c=imhist(a); d=imhist(b); z=c-d;

After this, the text file for the mining and classification of data was constructed. With this data from the subtractions applied to the histograms, the text file was constructed as a ".arff file", (Attribute-Relation File Format) an ASCII text file that describes a list of instances sharing a set with the declaration of the selected predictive attributes and the target attribute and part of the collected data. Below the statements have listed the data for the 300 instances collected using the procedure described by Waikato University. The ".arff" data file was processed using the software Weka® 3.5 using the classification induction algorithm C4.5, known as J48 to generate the decision tree. The decision tree (J48 pruned) obtained from the processing of this ".arff file" in the Weka software with the use of a training set, has a final size of 131 branches and 66 Leaves.

The pruned tree generated in the software Weka®3.5 was:

```
B <= 0.827
|   R <= 0.7899
|   |   G <= 0.6894
|   |   |   G <= 0.1425
|   |   |   |   R <= 0.1473
|   |   |   |   |   R <= 0.1353: 100 (9.0/4.0)
|   |   |   |   |   R > 0.1353: 125 (2.0)
|   |   |   |   R > 0.1473: 100 (4.0)
|   |   |   G > 0.1425
|   |   |   |   R <= 0.1811
|   |   |   |   |   R <= 0.1508: 25 (3.0)
|   |   |   |   |   R > 0.1508: 125 (3.0/1.0)
|   |   |   |   R > 0.1811
|   |   |   |   |   B <= 0.3729
|   |   |   |   |   |   G <= 0.3578
|   |   |   |   |   |   |   G <= 0.3083
|   |   |   |   |   |   |   |   B <= 0.0791: 25 (2.0/1.0)
|   |   |   |   |   |   |   |   B > 0.0791: 75 (11.0/4.0)
|   |   |   |   |   |   |   G > 0.3083: 125 (7.0/2.0)
|   |   |   |   |   |   G > 0.3578
|   |   |   |   |   |   |   R <= 0.5832
|   |   |   |   |   |   |   |   B <= 0.3147: 25 (3.0/1.0)
|   |   |   |   |   |   |   |   B > 0.3147: 75 (4.0/1.0)
|   |   |   |   |   |   |   R > 0.5832: 75 (5.0)
|   |   |   |   |   B > 0.3729
|   |   |   |   |   |   G <= 0.436
|   |   |   |   |   |   |   R <= 0.4483: 125 (7.0/4.0)
|   |   |   |   |   |   |   R > 0.4483: 50 (3.0)
|   |   |   |   |   |   G > 0.436
|   |   |   |   |   |   |   B <= 0.445: 25 (3.0/1.0)
|   |   |   |   |   |   |   B > 0.445
|   |   |   |   |   |   |   |   G <= 0.6312: 100 (11.0/5.0)
|   |   |   |   |   |   |   |   G > 0.6312: 25 (2.0/1.0)
|   |   G > 0.6894: 25 (7.0/1.0)
|   R > 0.7899: 75 (18.0/5.0)
B > 0.827
|   G <= 0.7951
|   |   R <= 0.6374
|   |   |   R <= 0.4861: 50 (2.0)
|   |   |   R > 0.4861: 25 (3.0)
|   |   R > 0.6374: 50 (9.0/1.0)
|   G > 0.7951
|   |   B <= 1.0468
|   |   |   R <= 0.8166: 25 (4.0/1.0)
|   |   |   R > 0.8166
|   |   |   |   B <= 0.9383
|   |   |   |   |   R <= 0.9796: 75 (4.0)
|   |   |   |   |   R > 0.9796
|   |   |   |   |   |   R <= 1.069: 100 (2.0)
|   |   |   |   |   |   R > 1.069: 75 (2.0/1.0)
|   |   |   |   B > 0.9383
|   |   |   |   |   R <= 1.1328: 25 (5.0/1.0)
|   |   |   |   |   R > 1.1328: 100 (4.0/1.0)
|   |   B > 1.0468
|   |   |   G <= 3.327
|   |   |   |   R <= 1.7811
|   |   |   |   |   B <= 1.748
|   |   |   |   |   |   G <= 1.6043
|   |   |   |   |   |   |   G <= 1.1845: 125 (11.0/6.0)
|   |   |   |   |   |   |   G > 1.1845
|   |   |   |   |   |   |   |   B <= 1.3484
|   |   |   |   |   |   |   |   |   R <= 1.3191: 25 (4.0)
|   |   |   |   |   |   |   |   |   R > 1.3191: 75 (3.0/1.0)
|   |   |   |   |   |   |   |   B > 1.3484
|   |   |   |   |   |   |   |   |   R <= 1.5037
|   |   |   |   |   |   |   |   |   |   G <= 1.3587: 50 (4.0/1.0)
|   |   |   |   |   |   |   |   |   |   G > 1.3587
|   |   |   |   |   |   |   |   |   |   |   B <= 1.5028: 25 (2.0/1.0)
|   |   |   |   |   |   |   |   |   |   |   B > 1.5028: 75 (2.0)
|   |   |   |   |   |   |   |   |   R > 1.5037
|   |   |   |   |   |   |   |   |   |   G <= 1.5374: 125 (3.0)
|   |   |   |   |   |   |   |   |   |   G > 1.5374: 50 (3.0)
|   |   |   |   |   |   G > 1.6043: 100 (5.0/1.0)
|   |   |   |   |   B > 1.748
|   |   |   |   |   |   G <= 1.5491: 50 (7.0)
|   |   |   |   |   |   G > 1.5491
|   |   |   |   |   |   |   R <= 1.6667: 25 (5.0/1.0)
|   |   |   |   |   |   |   R > 1.6667: 50 (4.0)
|   |   |   |   R > 1.7811
|   |   |   |   |   B <= 3.0858
|   |   |   |   |   |   G <= 2.5704
|   |   |   |   |   |   |   B <= 2.37
|   |   |   |   |   |   |   |   G <= 1.9491
|   |   |   |   |   |   |   |   |   R <= 1.9524
|   |   |   |   |   |   |   |   |   |   B <= 1.6397: 50 (2.0/1.0)
|   |   |   |   |   |   |   |   |   |   B > 1.6397: 125 (5.0/1.0)
|   |   |   |   |   |   |   |   |   R > 1.9524
|   |   |   |   |   |   |   |   |   |   B <= 1.9547
|   |   |   |   |   |   |   |   |   |   |   G <= 1.6251: 75 (2.0)
|   |   |   |   |   |   |   |   |   |   |   G > 1.6251: 100 (3.0/1.0)
|   |   |   |   |   |   |   |   |   |   B > 1.9547: 50 (2.0)
|   |   |   |   |   |   |   |   G > 1.9491
|   |   |   |   |   |   |   |   |   G <= 2.0554: 125 (4.0/1.0)
|   |   |   |   |   |   |   |   |   G > 2.0554
|   |   |   |   |   |   |   |   |   |   R <= 2.235: 100 (5.0/1.0)
|   |   |   |   |   |   |   |   |   |   R > 2.235
|   |   |   |   |   |   |   |   |   |   |   R <= 2.3869: 125 (2.0)
|   |   |   |   |   |   |   |   |   |   |   R > 2.3869: 100 (2.0)
|   |   |   |   |   |   |   B > 2.37
|   |   |   |   |   |   |   |   G <= 2.3788: 50 (7.0/1.0)
|   |   |   |   |   |   |   |   G > 2.3788
|   |   |   |   |   |   |   |   |   G <= 2.4222: 125 (2.0)
|   |   |   |   |   |   |   |   |   G > 2.4222: 75 (3.0/1.0)
|   |   |   |   |   |   G > 2.5704
|   |   |   |   |   |   |   R <= 2.8492
|   |   |   |   |   |   |   |   G <= 2.9385
|   |   |   |   |   |   |   |   |   G <= 2.7505
|   |   |   |   |   |   |   |   |   |   G <= 2.6089: 25 (3.0/1.0)
|   |   |   |   |   |   |   |   |   |   G > 2.6089: 100 (4.0/1.0)
|   |   |   |   |   |   |   |   |   G > 2.7505: 25 (3.0)
|   |   |   |   |   |   |   |   G > 2.9385: 125 (2.0)
|   |   |   |   |   |   |   R > 2.8492
|   |   |   |   |   |   |   |   B <= 2.8144: 75 (2.0)
|   |   |   |   |   |   |   |   B > 2.8144: 100 (6.0)
|   |   |   |   |   B > 3.0858
|   |   |   |   |   |   R <= 3.3706
|   |   |   |   |   |   |   R <= 2.834: 50 (2.0)
|   |   |   |   |   |   |   R > 2.834
|   |   |   |   |   |   |   |   B <= 3.6134: 125 (5.0)
|   |   |   |   |   |   |   |   B > 3.6134: 50 (2.0)
|   |   |   |   |   |   R > 3.3706: 75 (3.0)
|   |   |   G > 3.327
```

```
|   |   |   |   G <= 3.7037: 25 (7.0/1.0)
|   |   |   |   G > 3.7037
|   |   |   |   |   R <= 8.3917
|   |   |   |   |   |   B <= 6.3825
|   |   |   |   |   |   |   G <= 3.8249: 50 (2.0)
|   |   |   |   |   |   |   G > 3.8249: 125 (19.0/11.0)
|   |   |   |   |   |   B > 6.3825: 50 (5.0/1.0)
|   |   |   |   |   R > 8.3917: 100 (4.0/2.0)
```

The summary of the results obtained from the processing of the ".arff" file in the Weka software of the generated tree: Correctly Classified Instances: 229 (76.3333 %); Incorrectly Classified Instances: 71 (23.6667 %); *Kappa* statistic: 0.7042; Mean absolute error: 0.1275; Root mean squared error: 0.2525; Relative absolute error: 39.8432 %; Root relative squared error: 63.1215 %.

DISCUSSION

As can be seen in the results, the value of the *Kappa* statistic obtained was equal to 0.7042, and this value allows us, in comparison with the classification performance evaluation model presented by Landis and Koch (1977), to infer that the constructed classifier is " very good ", since it is in the *Kappa* range 0.61 - 0.80.

For each of the classes, the accuracy was detailedas follows:
Class 25: TP Rate = 0,750; FP Rate= 0,046; Precision = 0,804; Recall = 0,750; F-Measure = 0,776; MCC = 0,723; ROC Area = 0,961 and PRC Area = 0,845.

Class 50: TP Rate = 0,817; FP Rate= 0,021; Precision = 0,907; Recall = 0,817; F-Measure = 0,860; MCC = 0,829; ROC Area = 0,981 and PRC Area = 0,928.

Class 75: TP Rate = 0,767; FP Rate= 0,054; Precision = 0,780; Recall = 0,767; F-Measure = 0,773; MCC = 0,717; ROC Area = 0,960 and PRC Area = 0,841.

Class 100: TP Rate = 0,717; FP Rate= 0,067; Precision = 0,729; Recall = 0,717; F-Measure = 0,723; MCC = 0,654; ROC Area = 0,940 and PRC Area = 0,793.

Class 125: TP Rate = 0,767; FP Rate= 0,108; Precision = 0,639; Recall = 0,767; F-Measure = 0,697; MCC = 0,617; ROC Area = 0,939 and PRC Area = 0,792.

The accuracy of class weighted average Is:

TP Rate = 0,763; FP Rate= 0,059; Precision = 0,772; Recall = 0,763; F-Measure = 0,766; MCC = 0,708; ROC Area = 0,956 and PRC Area = 0,840.

It was possible to classify stressful situations evaluated with 76.3333% accuracy using only the tree predictive attributes and the intensity of the Nitrogen dosage. As we can see in the results, SAMBORSKI *et al.* (2016), the nitrogen management strategies developed based on small-scale plot research are meaningful using sensor-based systems incorporating active optical sensor (AOS). The image processing technology applied in this work, as can be seen in the work of GABRIEL *et al.* (2017) results as a rapid and non-destructive way to obtain optical readings the Nstatus, so it can be used as an indicator of N availability and reinforces SIMÕES *et al.* (2015) studies on the parameterization of morphophysiological attributes of plant canopies, made with methods based on the optical properties of plants, showing that is not necessary access to information by cutting and drying the biomass, followed by laboratory analysis, which complements that leads to more costs and execution time. This highlights the importance of indirect methodologies such as image processing evaluations in real time, such as the use of information obtained from digital cameras. The use of Data Mining technique improve the findings of knowledge as stated by MOI *et al.* (2014) and the KDD process, as stated by PÖLÖNEN *et al.* (2013) enables the researcher to obtain a knowledge containing a database, guaranteeing a high precision in theobtained responses.the data mining techniques, described by LIMA *et al.* (2017), leads to an efficient decision tree induction,In addition, the knowledge found in the decision tree generate rules and the algorithm achieves satisfactory results when compared with other more sophisticated approaches available in the literature. And finally, we can also state that in addition, the results are usually supplied quickly due to the computational efficiency as cited by DAI *et al.* (2016).

Conclusion

The achieved results show that the data collected by cameras installed in the crop have enabled the construction of a decision tree-based algorithm for the definition of a correct Nitrogen dosage to be applied by means of cover fertilization at focal points of the beans *Vigna unguiculata (L.) Walp* cultures. The developed technological tool was presented as a good alternative to the management strategies to determine the correct application of Nitrogen in bean crops. The images captured with the cameras allows the construction of indexes, related to the vegetation cover and was possible access information traditionally been done by cutting and drying the biomass, followed by laboratory analysis. This reveals the importance of indirect methodologies such as image processing evaluations in real time, obtained from digital cameras with the use of Data Mining technique to improve the findings increasing the knowledge to support a better decision-making process. Applying this method was possible to classify the Nitrogen dosage to be applied to the focal points of the cultures, with 76.33% accuracy, using only the tree predictive attributes and the intensity of the Nitrogen dosage.

Acknowledgments

The authors wish to thank CNPq for granted scholarship.

REFERENCES

Bertoldo, J. G.; Pelisser, A.; Silva, R. P.; Favreto, R.; Oliveira, L. A. D. 2015. Alternativas na fertilização de feijão visando a reduzir a aplicação de N-ureia1. Pesquisa AgropecuáriaTropical, Goiânia, v. 45, n. 3, p. 348-355.

Dai, Q. Y.; Zhang, C. P.; Wu, H. 2016. Research of Decision Tree Classification Algorithm in Data Mining. *International Journal of Database Theory and Application,* v. 9, n. 5, p. 1-8.

Gabriel, J. L.; Zarco-Tejada, P. J.; López-Herrera, P. J.; Pérez-Martín, E.; Alonso-Ayuso, M.; Quemada, M. 2017. Airborne and ground level sensors for monitoring nitrogen status in a maize crop. *Biosystems Engineering,* v.160, p.124-133.

Han, J.; Kamber, M.; Pei, J. 2011. Data mining: concepts and techniques. 3rd edition. San Francisco: Morgan Kaufmann Publishers.

Landis, J.R.; Koch, G.G. 1997. The measurement of observer agreement for categorical data. *Biometrics,* v. 33, n.1, p. 159-174.

Lima, E. S., Souza, Z.M., Montanari, R., Oliveira, S.R. M., Lovera, L. H., & Farhate, C. V. V. (2017). Classification of the initial development of eucalyptus using data mining techniques. Cerne, Lavras, v. 23, n. 2, p. 201-208.

Mancuso, M. A. C.; Aires, B. C.; Negrisoli, E.; Corrêa, M. R.; Soratto, R. P. 2016. Seletividade e eficiência de herbicidas no controle de plantas daninhas na cultura do feijão-caupi. Revista Ceres, Viçosa, v. 63, n. 1, p. 25-32.

Murakami, P. F.; Turner, M. R.; Van Den Berg, A. K.; Schaberg, P. G 2005. An Instructional Guide for Leaf Color Analysis using Digital Imaging Software. USDA Forest Service, *Northeastern Research Station*, General Technical Report NE-327.

Pinho, T.; Boaventura-Cunha, J.; Morais, R 2015. Tecnologias da eletronicae da computação na recolha e integração de dados em agricultura de precisão. Revista de Ciências Agrárias**,**Lisboa, v. 38, n. 3, p. 291-304.

Pölönen, H. Saari, J. Kaivosoja, E. Honkavaara, L. Pesonen 2013. "Hyperspectral imaging based biomass and nitrogen content estimations from light-weight UAV", Proc. SPIE 8887, Remote Sensing for Agriculture, Ecosystems, and Hydrology XV, 88870J. 10.1117/12.2028624.

Samborski, S. M.; Gozdowski, D.; Stępień, M.; Walsh, O. S.; Leszczyńska, E. 2016. On-farm evaluation of an active optical sensor performance for variable nitrogen application in winter wheat. *European Journal of Agronomy,* V.74, p.56-67.

Simões, C. R.; Rossiello, R. O. P.; Graciosa, M. G.; Machado, M. L.; Silva, C. F. 2015. Imagens multiespectrais para avaliação de índice de área foliar e massa seca do capim 'Tifton 85', sob adubação nitrogenada. CiênciaRural, SantaMaria, v. 45, n. 4, p. 697-703.

Witten, I.H.; Frank, E.; Hall, M.A 2011. Data mining: practical machine learning tools and techniques. São Francisco, CA: The Morgan Kaufmann series in data management systems, 665p.

****