

ANALYSIS OF VHR IMAGE CLASSIFICATION BY SINGLE AND ENSEMBLE OF CLASSIFIERS

M. G. Lacerda^{1,2,*}, E. H. Shiguemori^{1,3}, A. J. Damião^{1,2}, C. S. Anjos^{4,**}, M. Habermann¹

¹ Institute for Advanced Studies - IEAv, São José dos Campos-SP, Brazil - (marielcio,elcio,damiao, habermann)@ieav.cta.br

² Aeronautics Institute of Technology - ITA, São José dos Campos-SP, Brazil

³ National Institute for Space Research - INPE, São José dos Campos-SP, Brazil

⁴ Federal Institute of Education, Science and Technology of South of Minas Gerais - IFSULDEMINAS, Inconfidentes-MG, Brazil
- (camila.lacerda@ifsuldeminas.edu.br)

* IEEE Student Member ** IEEE Member

ICWG

KEY WORDS: RPAS, Very High Resolution Images, Classifiers, Majority Voting, Computational Time

ABSTRACT:

Given the wide variety of image classifiers available nowadays, some questions remain about the accuracy and processing time of Very High Resolution (VHR) images. Another question concerns the use of a Single or Ensemble Classifiers. Of course, the main factor to consider is the quality of the classified image, but computational cost is also important, especially in applications that require real-time processing. Given this scenario, this paper aims to relate the accuracy of seven single classifiers and the ensemble of the same classifiers with the processing time. In this paper the ensemble of classifiers had the best results in terms of accuracy, however, it comes to processing time, the decision tree had the best performance.

1. INTRODUCTION

When it comes to image classification and image interpretation, this classification must depict reality as good as possible (Foody, 2002).

For high spatial resolution images, normally the classification process of a given input data begins with the segmentation, whose parameters must be carefully chosen. The segmentation process merges the pixels that share similar characteristics, this way the groups end up well distributed in the feature space (Geotecnologias, 2010). All the pixels in a group bear the same label, so there are fewer groups than pixels in an image, consequently, the segmentation-based classification has less work to do when compared to a pixel-wise approach (Gonzalez, Woods, 2010).

Once the segmentation parameters are defined and the image is segmented, the next step is the choice of either supervised or unsupervised classification. In this work, we opt for supervised classification. Thus, it is necessary to have training data to appropriately represent all the classes of the classification problem (Lacerda et al., 2019).

In the sequel, it is necessary to proceed to the land-cover classification. Some classification algorithms are largely used and well accepted by the scientific community, however, new approaches are popping up (Chaves, 2012). Those new methods have a large range of accuracy and processing time to generate the classified image.

In this context, this work aims at comparing eight classifiers in terms of accuracy and processing time. The input is an orthophoto mosaic with very high spatial resolution, set up from images acquired by a small-sized drone. The orthophoto mosaic embraces rural and urban zones, with different kinds of land-cover.

2. IMAGE ACQUISITION AND PROCESSING

2.1 Study area

For this work, we chose the municipality of Inconfidentes, located to the south of the state of Minas Gerais. Figure 1 presents the study area.

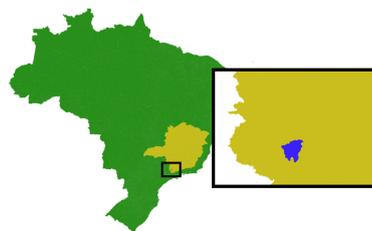


Figure 1. Study area - Inconfidentes/MG.

2.2 RPA Used for Survey and Flight Planning

To acquire the images used in this paper, we proceeded to flights using a small RPA, namely, DJI Phantom 4. This drone features a 12 MPixel visible-range (RGB) camera, coupled with a gimbal, which ensures you get your images at the desired angle and with minimal vibration. The camera also allows shooting at up to 4K resolution ((Lacerda et al., 2019) and (DJI, 2016).

For flight planning, the Pix4DCapture (Pix4D, 2019) mobile app was used. The experiment was performed on November 11, 2018, with the conditions:

- Flight Height: 120m (from takeoff point);
- Imaged Area: 1,100 x 1,600 m;
- Front Cover: 80% and Side Cover: 60%;

This contribution has been peer-reviewed.

<https://doi.org/10.5194/isprs-archives-XLII-3-W12-2020-37-2020> | © Authors 2020. CC BY 4.0 License.
Primary publication at IEEE Xplore: <https://doi.org/10.1109/LAGIRS48042.2020.9165621>

- Flight mode: 2D - One-way flight;
- Aircraft speed: 15 m/s;
- Camera Direction: 90° (NADIR); and

During the flight, 721 images were grabbed, and they were processed in the appropriate software. (Pix4D, 2019). We used 8 control points and 3 check-points, which yielded an orthophoto mosaic with about 6 centimeter GSD, as shown in Figure 2.

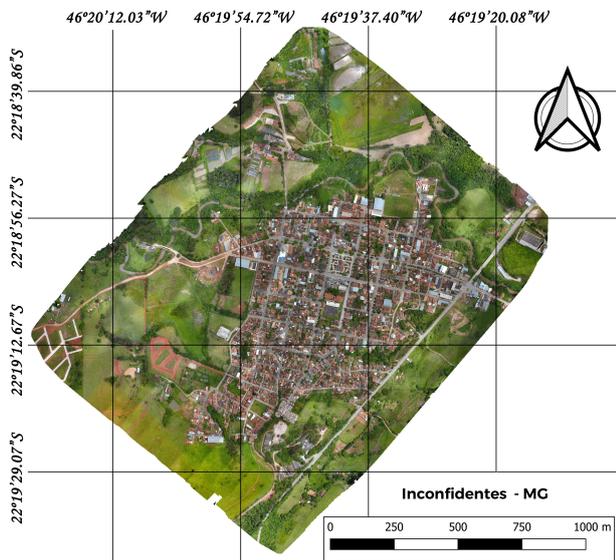


Figure 2. Orthophoto mosaic

2.3 Methodology

In Figure 3, a summary of the methodology that has been used in this work is presented.

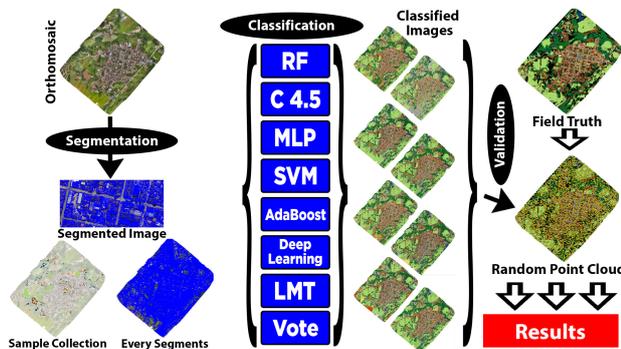


Figure 3. Methodology Summary

2.3.1 Image Segmentation. Segmentation seeks to subdivide the image into smaller regions (or objects). Thus, the stopping criterion of the algorithm depends on how good the objects are distinguished according to the necessity of the problem (Gonzalez, Woods, 2010).

So, segmentation plays a very important role in the realm of digital image processing, for the refinement degree of the segmentation enables a proper characterization of the objects in the image under analysis (Gonzalez, Woods, 2010).

Taking into consideration the results from other scientific papers (Anjos et al., 2017b), we proceed to the segmentation of the orthophoto mosaic (item 2.2) in two steps. The first step

is the Multi-resolution Segmentation (MRS), whose parameters are: **Scale** = 50; **Shape** = 0.1; and **Compactness** = 0.5.

The second step is Spectral Difference Segmentation (SDS), whose parameter Maximum Spectral Difference = 10. The output of these two steps is an image with 333,596 segments.

Figure 4 shows the resulting segmented image.

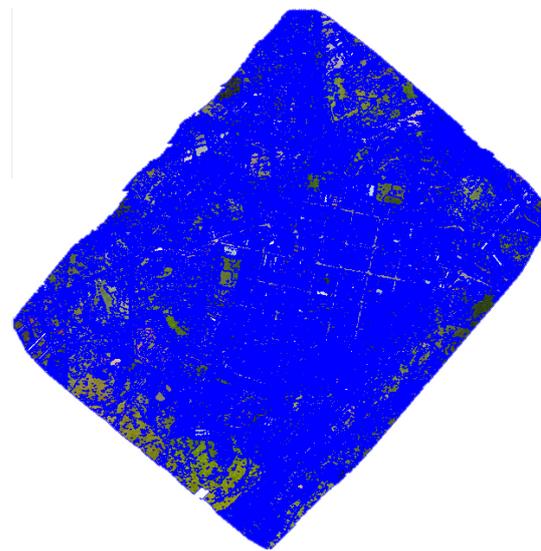


Figure 4. Segmented Image

A. Multi-Resolution Segmentation (MRS). The multiresolution segmentation technique analyzes the image objects in different resolution levels. It considers that smaller objects may be adequately analyzed in finer resolutions, whereas bigger elements may be handled in coarser resolution. The segmentation may be accomplished in multiple resolutions, by varying the segmentation algorithm parameters, which change the processing resolution (Anjos et al., 2015) and (Leonardi, 2010).

Basically, the parameters that have influence on MRS are (Definiens, 2007):

- **Scale parameter** determines the maximum heterogeneity for the objects in the resulting image. For heterogeneous data, the resulting objects according to a given scale parameter are smaller than the objects from homogeneous data. When we alter the scale parameter value, the size of the objects in the resulting image also changes.
- **Homogeneity Composition Criterion** is used as a synonym of minimized heterogeneity. In the software ((eCognition, 2019), two parameters define this value: **shape** e **compactness**.

B. Spectral Difference Segmentation (SDS). SDS refines pre-existing segmentation, and it is not used to generate new segmentation in the pixel level (Definiens, 2007). SDS analyzes objects and neighboring segments of the first segmentation and merges them according to the absolute difference among their mean grey level. This difference is the maximum value or a threshold defined by the user according to the characteristics and features to be distinguished in the image (Bartani, Rossetti, 2014). This parameter is called **Maximum Spectral Difference (MSD)**.

2.3.2 Classes In this work, we consider 21 classes: **1.** Water - Swimming Pool **2.** Water - River and Lake **3.** Water Box **4.** Ceramic Coverage **5.** Concrete Coverage **6.** Fiber Cement Coverage **7.** Metallic Coverage **8.** Rusty Metallic Coverage **9.** Plastic and Canvas **10.** Asphalt Paving **11.** Painted Asphalt Paving **12.** Block Paving **13.** Painted Block Paving **14.** Ceramic Paving **15.** Concrete Paving **16.** Mining Stone **17.** Solar Power Plates **18.** Exposed Ground **19.** Grass **20.** Arboreal Vegetation **21.** Shadow

All of those classes are present in the image to be analyzed. However, we emphasize that this image has only RGB information, which makes the distinction of some classes with similar spectral signature more complicated, like Block Paving and Concrete Paving and Concrete Coverage; or Mining Stone and Ceramic Coverage.

2.3.3 Sample Collection The classifiers used in this work perform supervised classification. Thus, it is necessary to have training data. In total, 951 samples were collected, spread all over the image and embracing all classes. The samples are shown in Figure 5.



Figure 5. Sample Collection

2.3.4 Attribute Selection After segmentation, the following attributes were chosen to be exported for use in the image classification process: Mean; Standard Deviation; Brightness; Maximum Difference; Mode; Ratio; Minimum Pixel Value; Maximum Pixel Value; Contrast to Neighbor Pixels; HSI; Area, Border Length, Length, Length/Width; Width; Asymmetry; Border Index; Compactness; Main Direction; Rectangular Fit; Shape Index; Texture after Haralick (GLCM Homogeneity, GLCM Dissimilarity, GLCM Entropy, GLCM Ang. 2nd Moment and GLCM Correlation) and Class Name. As most of those features are related to the channels of the image, and there are three channels and DSM, therefore there are 68 attributes.

Those features were selected by the fact of being considered adequate according to their characteristics (Definiens, 2007).

2.3.5 Data set From the segmented image (Item 2.3.1) and taking into account the features (Item 2.3.4), we exported the data from both sets used during the process of image classification: one set containing the attributes for all segments,

but without information about class membership; and a training set, containing the sample collection segments with attributes and label information.

Both sets were used during the image classification by the classifiers presented in item 2.3.6.

2.3.6 Image Classification. The classifiers used are thus:

- Random Forest (RF);
- Decision Tree (C 4.5);
- Multilayer Perceptron (MLP);
- Support Vector Machine (SVM);
- AdaBoost;
- Deep Learning;
- Logistic Model Tree (LMT); and
- Majority Voting.

Those classifiers were chosen because they have been used in much recent scientific paper concerning image classification. All of them are implemented in the software we use, namely, (WEKA, 2019).

A. Random Forest (RF) . RF is a classification algorithm devised by (Breiman, 2001). It uses lots of unrelated decision trees. From the whole training data, different data sets are randomly created, one set for each tree (Anjos et al., 2017a). According to the literature, the RF accuracy outperforms other classifiers. It also has good performance in practical problems (Biau et al., 2008).

B. Decision Tree (C 4.5). Decision Tree assigns a class to an input sample by using one, or more, decision function in a successive fashion. This strategy may be illustrated by a tree-like diagram (Swain, Hauska, 1977). In general, a decision tree is composed of an initial node, internal nodes, and terminal nodes. Its more important characteristic is the capacity of decomposing a complex decision-making process in a bunch of simpler decisions, which turns the solution easier to interpret (Safavi, Landgrebe, 1990).

C. Multilayer Perceptron (MLP). Multilayer Perceptron may be used for data classification. In a network, each neuron performs a simple function. The input is a feature vector to be classified, and the output is the class label. The weights that connect the neurons are adjusted during the training phase of the algorithm (Ruck et al., 1989).

D. Support Vector Machine (SVM). SVM is based on the idea of structural risk minimization. Firstly, SVM maps the input samples in the feature space and finds a separating hyperplane in order to maximize the margin between classes. The solution of the best hyperplane may be written as a combination of some input samples, called support vectors (Lin, Wang, 2002).

E. AdaBoost. AdaBoost is acronym for *Adaptive Boosting*. The algorithm stands out due to its simplicity, flexibility, and potential (Chaves, 2012). AdaBoost is one of the most popular algorithms to construct a strong classifier with a linear combination of member classifiers. The member classifiers are selected to minimize the errors in each iteration step during the training process. AdaBoost provides a very simple and useful method to generate ensemble classifiers. The performance of the ensemble depends on the diversity among the member classifiers as well as the performance of each

member classifiers. However, the existing AdaBoost algorithms are focused on error minimization problems (An, Kim, 2010).

F. Deep Learning. Deep Learning has been extensively used in the area of pattern recognition. It consists of a neural network with more than 3 hidden layers. The weights are used to initialize a fine-tuning stage for the training of the deep network. (Revathi, Kumar, 2016). The Weka Deep Learning package used provides Graphics Processing Unit (GPU) support and enables Graphical User Interface (GUI) based training of deep neural networks such as convolutional and recurrent neural networks (Lang et al., 2019).

G. Logistic Model Tree (LMT). LMT combines logistic regression and decision tree learning. The logistic regression may be seen as a statistical technique that aims at creating a value prediction model. From a sample set, the algorithm uses a categorical variable, usually binary, to represent continuous or binary variables. Logistic regression models permit the selection of relevant features naturally. When compared to known regression techniques, especially the linear one, logistic regression has the advantage of having a categorical output (da Silva Ruiz et al., 2018).

H. Ensemble of Classifiers. Ensemble methods are learning algorithms that construct a set of classifiers and then classify new data points by taking a vote of their predictions. Ensembles can often perform better than any single classifier (Dietterich, 2000). In this paper, the class label is defined by the majority voting of seven classifiers. That is, one must define which classifiers participate in voting. Each classifier defines the class of the input sample. Majority voting has demonstrated great ability in producing very precise classification rules (James, 1998).

2.3.7 Ground Truth To validate the results, the ground-truth data was generated by local inspection and visual interpretation by a specialist team. Extra flights were necessary to get further information.

With the ground-truth data, the data were analyzed and we defined 21 classes.

Figure 6 presents the ground truth and the 21 classes.

2.3.8 Random Points To validate the classification results, we used the software QGIS 2.18 (QGIS, 2019) to spread 4,000 points over the ground-truth image. 2,500 of them lie over the image, whereas the others are concentrated on the region without ground-truth.

By intersecting the points of the ground-truth image with the points of the classified image, it is possible to proceed to the validation. After that, we calculate the confusion matrix, from which the Overall Accuracy and Kappa Index are evaluated.

2.3.9 Evaluation of the Classification When we work with automatic image classification it is necessary to guarantee the reliability of the output. One way to do so is by using the confusion matrix, from which it is possible to calculate various metrics to assess the classification (Bernardes, 2006) and (Figueiredo, Vieira, 2007).

By analyzing the confusion matrix (Figueiredo, Vieira, 2007) and (Foody, 2002), it is possible to calculate the Overall Accuracy, Kappa Index, Producer Accuracy, User Accuracy,



Figure 6. Ground Truth and classes

Inclusion Error, Omission Error, Kappa Index for class, etc. In this work, we use the Overall Accuracy and Kappa Index.

Overall Accuracy is calculated by the quotient between the sum of the main diagonal elements and the total number of data samples (Figueiredo, Vieira, 2007). That is, for the calculation of Overall Accuracy we take into consideration only the number of right answers (main diagonal) and the total number of samples.

Kappa Index (K) is obtained by considering all the entries of the confusion matrix. Kappa is considered adequate for the verification of classification precision because it considers the whole confusion matrix (including the off-diagonal elements). Therefore, it is different from the overall accuracy (Congalton, 1991).

3. RESULTS AND DISCUSSION

As previously stated, eight classifications were performed. Figure 7 shows all the classified images from each process.

For each classification, there is a confusion matrix, from which the Overall Accuracy and Kappa Index are calculated.

Table 1 presents the results: Overall Accuracy, Kappa Index and the processing time, which is composed of two parts: time to generate the model and time for classification.

Figure 8 shows a plot that relates to accuracy in each classification with the processing time of each algorithm.

4. CONCLUSIONS

We can analyze the data from Table 1 under three different perspectives:

- According to accuracy (Overall Accuracy and Kappa Index);

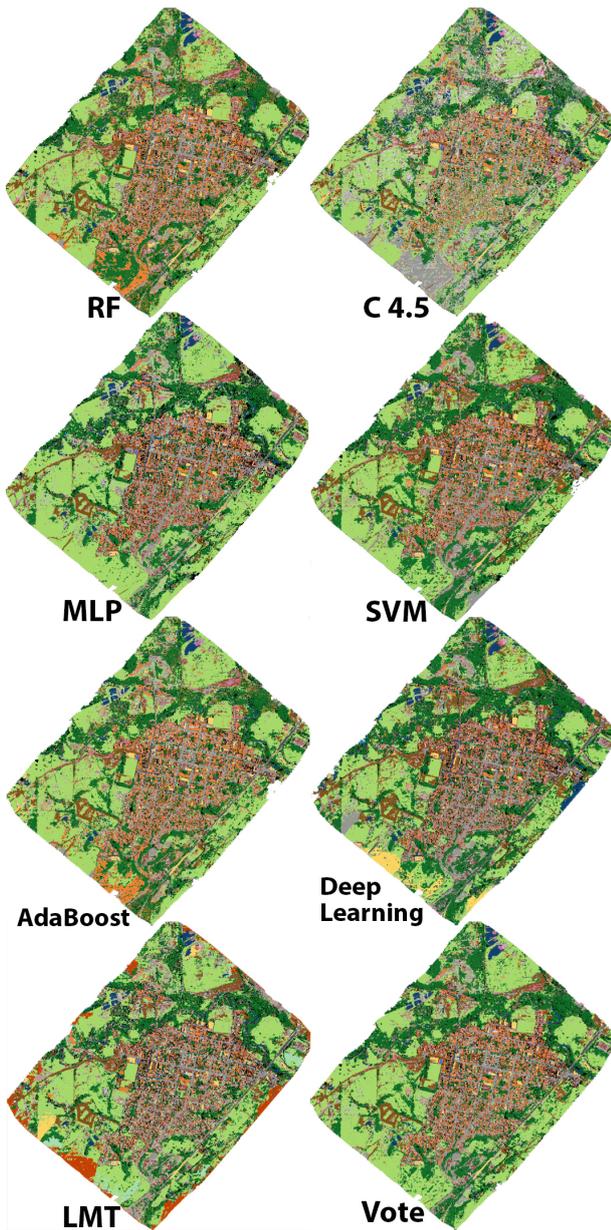


Figure 7. Classified Images

Table 1. Results of Classification.

Alg.	OA	Kappa	Processing Time (seconds)		
			model	classif	Total
RF	70.18%	0.648	1.13	162.69	163.82
C4.5	57.75%	0.510	0.27	132.80	133.07
MLP	76.19%	0.713	36.42	158.00	194.42
SVM	72.48%	0.672	0.78	144.46	145.24
Ada	71.00%	0.656	1.17	149.15	150.32
Deep	68.80%	0.633	6.72	239.16	245.88
LMT	68.00%	0.625	18.39	125.07	143.46
Vote	78.36%	0.740	61.04	482.02	543.06

- According to processing time; and
- According to the relation Accuracy x Processing Time.

In terms of accuracy, the Ensemble by Majority Voting had the best result, with 78.36% of Overall Accuracy and 0.740 of Kappa Index. In a certain way, it was an expected result,

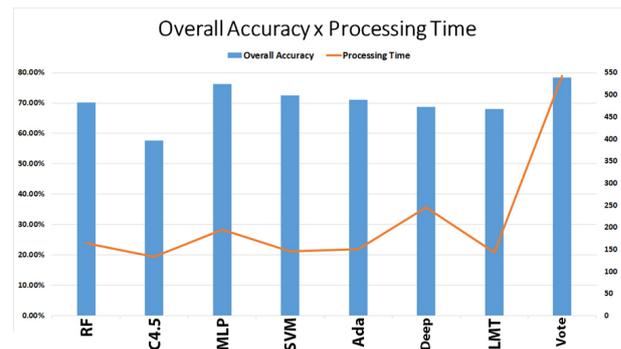


Figure 8. Overall Accuracy x Processing Time

because this algorithm considers the outcome of all other classifiers, decreasing, therefore, the possibility of classification errors.

When it comes to processing time, the decision tree had the best performance (C 4.5), classifying the image in 133.07 seconds. It is worth noting that it took this method only 0.27 seconds to set the classification model, indicating the simplicity and quickness of this method. In terms of processing time, we emphasize that the LMT algorithm took the least time to classify the image, showing that this method has a great capacity to set the classification model, with a certain computational burden, but it has great agility in classifying data a posteriori.

By comparing accuracy and processing time, one may see that the Majority Voting algorithm had the best accuracy (OA = 70.18%). However, it had the biggest processing time (543.06 seconds). On the other hand, the decision tree had the smallest processing time, but it had the worst accuracy (57.75%). Random Forest, SVM, and AdaBoost had very similar accuracy, and their processing time was almost the same.

Therefore, for the orthophoto mosaic analyzed in this paper and taking into account all the eight classifiers, we can conclude that in terms of accuracy the Ensemble by Majority Voting is the best option (without considering the processing time). However, if one seeks to have results quickly, the best method is the Decision Tree. If anyone requires good accuracy, but not willing to spend much time, one can resort to RF, SVM, and AdaBoost.

Another possible analysis is about the differentiation of each class, which is assessed using the confusion matrix. As cited in item 1, the orthophoto mosaic over which this work was done has only the RGB information. So, some classes have similar spectral behavior. But, this analysis was not considered in this work, and this can be addressed in future works. But yet, it is possible to conclude that the Overall Accuracy values obtained substantial agreement considering the number of classes and the difficulty in characterizing them.

REFERENCES

- An, T.-K., Kim, M.-H., 2010. A New Diverse AdaBoost Classifier. *2010 International Conference on Artificial Intelligence and Computational Intelligence*, 1.
- Anjos, C. S. et al., 2015. Identificação de materiais urbanos por meio de métodos inovadores de classificação de imagens. *Anais*

- do XVII Simpósio Brasileiro de Sensoriamento Remoto - SBSR, INPE, 1, p. 4377-4384.
- Anjos, C. S. et al., 2017a. Análise do Nível de Legenda de Classificação de Áreas Urbanas Empregando Imagens Multiespectrais e Hiperspectrais com Métodos Árvore de Decisão C4.5 e Floresta Randômica. *Boletim de Ciências Geodésicas*, 23.
- Anjos, C. S. et al., 2017b. Classification of urban environments using Feature Extraction and Random Forest. *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 1, p. 1205-1208.
- Bartani, T. C., Rossetti, D. F., 2014. Segmentação por diferença espectral para extração de relevos em planícies fluviais amazônicas. *Revista Geonorte, Edição Especial 4*, 10, p. 291-296.
- Bernardes, T., 2006. Caracterização do ambiente agrícola do Complexo Serra Negra por meio de sensoriamento remoto e sistemas de informação geográfica. *Dissertação de Mestrado - Universidade Federal de Lavras*, 1, 119 p.
- Biau, G. et al., 2008. Consistency of Random Forests and other averaging classifiers. *Journal of Machine Learning Research*, 9, p. 2015-2033.
- Breiman, L., 2001. Random Forest. *Machine Learning. Statistics Department - University of California*, 1, 33 p.
- Chaves, B. B., 2012. Estudo do Algoritmo AdaBoost de Aprendizagem de Máquina Aplicado a Sensores e Sistemas Embarcados. *Dissertação de Mestrado - Escola Politécnica USP*, 1.
- Congalton, R. G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37, p. 35-46.
- da Silva Ruiz, P. R. et al., 2018. Classificação da Cobertura do Solo Urbano Usando Árvores de Decisão a partir de uma Cena WorldView-2 para Diferentes Níveis de Legenda. *Geociências - UNESP*, 37 - n.3, p. 597 - 609.
- Definiens, 2007. *Definiens Developer7: Reference Book*. 1, 7.0.2.936 edn, Definiens, Munich, Germany.
- Dietterich, T. G., 2000. Ensemble Methods in Machine Learning. *International Workshop on Multiple Classifier Systems - MCS 2000*, 1857, p. 1 - 15.
- DJI, 2016. *Phantom 4: User Manual*. 1, 1.2 edn, DJI. Disponível em: <http://www.dji.com/product/phantom-4>.
- eCognition, 2019. eCognition. *site: http://www.ecognition.com/*.
- Figueiredo, G. C., Vieira, C. A. O., 2007. Estudo do comportamento dos Índices de Exatidão Global, Kappa e Tau, comumente usados para avaliar a classificação de imagens do sensoriamento remoto. *Anais do XIII Simpósio Brasileiro de Sensoriamento Remoto - SBSR, INPE*, p. 5755-5762.
- Foody, G. M., 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80, p. 185-201.
- Geotecnologias, S., 2010. Classificação Orientada a Objeto - considerações. *site: geotecnologias.wordpress.com/2010/08/19/classificacao-orientada-a-objeto-consideracoes*.
- Gonzalez, R. C., Woods, R. E., 2010. *Processamento Digital de Imagens*. 1, 3 edn, Pearson, São Paulo.
- James, G., 1998. Majority Vote Classifiers: Theory and Applications. *Dissertation to the Department of Statistics and the Committee on Graduate Studies of Stanford University*, 1, 123 p.
- Lacerda, M. G. et al., 2019. Influência dos Parâmetros de Segmentação de Imagem em Ortofotomosaicos confeccionados a partir de fotografias obtidas por Aeronaves Remotamente Pilotadas de Pequeno Porte. *Anais do XIX Simpósio Brasileiro de Sensoriamento Remoto*, 1, 4 p.
- Lang, S. et al., 2019. WekaDeeplearning4j: A deep learning package for Weka based on Deeplearning4j. *Knowledge-Based Systems*, 178, p. 48 - 50.
- Leonardi, F., 2010. Abordagens cognitivas e mineração de dados aplicadas a dados ópticos orbitais e de laser para a classificação de cobertura do solo urbano. *Dissertação de Mestrado - Instituto Nacional de Pesquisas Espaciais, INPE*, 1, 164 p.
- Lin, C.-F., Wang, S.-D., 2002. Fuzzy Support Vector Machines. *IEEE Transactions on Neural Networks*, 13, N° 2, p. 464-471.
- Pix4D, 2019. Pix4D Products. *site: https://www.pix4d.com*.
- QGIS, 2019. QGIS. *site: https://www.qgis.org/en/site/*.
- Revathi, A. R., Kumar, D., 2016. An efficient system for anomaly detection using deep learning classifier. *Springer - Verlag Londo*, 9.
- Ruck, D. W. et al., 1989. Feature Selection Using Multilayer Perceptron. *Journal of Neural Network Computing*, 2 - Number 2, p. 40 - 48.
- Safaviyan, S. R., Landgrebe, D., 1990. *A Survey of Decision Tree: Classifier Methodology*. 1, tr-ee 90-54 edn, School of Electrical Engineering - Purdue University, West Lafayette, Indiana, United State of America.
- Swain, P. H., Hauska, H., 1977. The Decision Tree Classifier: Design and Potential. *IEEE Transactions on Geoscience Electronic*, GE-15, No 2, p. 142 - 147.
- WEKA, 2019. WEKA. *site: https://www.cs.waikato.ac.nz/ml/index.html*.