Image Classification for Precision Agriculture: A Coffee Study Case

Julio Ribeiro*
julioribeiro@alunos.utfpr.edu.br
Federal University of Technology – Paraná (UTFPR)
Apucarana, Paraná, Brasil

Matheus Mileski*
matheuslopes@alunos.utfpr.edu.br
Federal University of Technology – Paraná (UTFPR)
Apucarana, Paraná, Brasil

José Luis Seixas Junior
jlseixasjr@inf.elte.hu
ELTE – Eötvös Loránd University, Faculty of Informatics
Budapest, Hungary

Luiz Fernando Carvalho
luizfcarvalho@utfpr.edu.br
Federal University of Technology – Paraná (UTFPR)
Apucarana, Paraná, Brasil

Rafael Gomes Mantovani
rafaelmantovani@utfpr.edu.br
Federal University of Technology – Paraná (UTFPR)
Apucarana, Paraná, Brasil

ABSTRACT

Precision agriculture has emerged as a transformative approach to optimizing crop production and resource management in the agricultural sector. It leverages advanced technologies like remote sensing, geographic information systems, and decision support tools. In recent years, machine learning has become an integral component of precision agriculture, providing the ability to analyze large volumes of complex data and generate valuable insights for more informed decision-making. This study benchmarked different supervised algorithms in an image classification dataset of coffee and non-coffee areas. Algorithms were comprehensively evaluated and analyzed, with traditional machine learning models trained in three versions of the original dataset, and compared with two Deep Learning algorithms. The experimental results were promising, showing that SVMs and RF algorithms can provide accurate predictions for most images with average accuracy values above 0.8. The RF algorithm, in specific, statistically outperformed all the other traditional algorithms. Deep Learning baselines were slightly more accurate but had a higher computation cost in the training step. VGG16 was the best-evaluated model, followed by the simple CNN and RF, with no statistical difference between the last two. Results suggest that there is still scope for further improvement in using the traditional algorithms.

KEYWORDS

image classification, machine learning, precision agriculture

1 INTRODUCTION

Precision agriculture has emerged as a transformative approach to optimizing crop production and resource management in the agricultural sector [1]. It leverages advanced technologies, such as remote sensing, geographic information systems, and decision support tools, to facilitate the development of site-specific farming practices that maximize yield while minimizing input costs and environmental impacts [2]. In recent years, Machine Learning (ML) has become an integral component of precision agriculture, providing the ability to analyze large volumes of complex data and generate valuable insights for more informed decision-making [3].

Remote sensing is a key aspect of precision agriculture which involves acquiring data about the Earth's surface through satellite or airborne sensors [4]. This technology enables the monitoring of agricultural fields at various spatial, temporal, and spectral resolutions, allowing for the identification of crop types, growth stages, and stress conditions [5]. As a result, remote sensing data can be harnessed to inform crop management strategies, such as irrigation scheduling, nutrient application, and pest control, which are essential for maintaining high productivity and environmental sustainability [6].

In this sense, ML offers powerful techniques for analyzing remote sensing data [7]. By employing algorithms that automatically learn and adapt from data, ML models can identify patterns, make predictions, and uncover hidden relationships within complex datasets [8]. These capabilities have proven particularly useful in image classification, in which these algorithms are trained to recognize and categorize objects within images based on their features [9]. The integration of ML with remote sensing data has the potential to significantly advance the field of precision agriculture [10]. In particular, image classification techniques can distinguish different land cover types, such as crop species and growth stages, from satellite imagery [11]. This information can be utilized by farmers, agronomists, and other stakeholders to optimize their agricultural practices and achieve greater efficiency in resource allocation [12].

Considering the literature scenario, there are two types of studies: "white-box" traditional ML algorithms, requiring a feature engineering step to provide valuable image descriptors, and "black-box" Deep Learning [13]. DL models are well-known and widely used in extracting high-level abstract features. Theoretically, it outperforms the traditional ML algorithms but with a higher computation cost [14]. However, DL performance depends on some assumptions, especially the data amount used to train the models. The bigger the dataset, the better will be the induced DL model.

Thus, in this study, we hypothesized that traditional ML algorithms are good alternatives to the DL. DL algorithms were also considered as baselines to the handcrafted approach. One of the objectives in this work is to compare these algorithms' performance, robustness, and generalizability to classify satellite images as coffee or non-coffee areas. The insights gained from this research are expected to contribute to similar problems and further

*Both authors contributed equally to the paper investigations on the ongoing development of precision agriculture technologies. The remainder of this paper is organized as follows: Section 2 presents some of the necessary concepts about precision agriculture with ML and describes related work. The experimental methodology is presented in Section 3. The results are discussed in Section 4, while the conclusions and final considerations of the study are presented in Section 5.

2 RELATED WORKS

Different ML algorithms have been explored to increase efficiency and productivity in precision agriculture. The growing body of research shows the promising results of such techniques, mainly when applied to specific case studies such as coffee crops. In [15], the authors investigated coffee crops' spectral analysis and classification using Landsat and a topographic-environmental model. They found that satellite mapping of coffee yielded low classification accuracy. However, when a wider spectrum of bands and ancillary data was included, the highest overall accuracy is improved to 65%.

Over a decade later, advances in ML technologies were highlighted by [16]. The authors proposed a method for classifying and detecting nutritional deficiencies in coffee plants using image processing and Convolutional Neural Networks (CNNs). Their model, once detecting a deficiency, suggests the appropriate fertilizers. Their predicted models were trained on images representing eight types of nutritional deficiencies across four coffee varieties and obtained high accuracy in detecting these deficiencies.

In [17], the authors evaluated the potential of combining canopy spectral information with canopy structure features for crop monitoring using satellite/unmanned aerial vehicle (UAV) data fusion and ML. They discovered that combining rich spectral data from satellite imagery with detailed canopy structural features from UAVs significantly improved biomass estimation, leaf area index (LAI), and leaf nitrogen concentration. Similar work was done by [18] in a semiarid area of Morocco. Experimental results demonstrated that the fusion of different source data significantly improved the accuracy of crop-type identification.

Lastly, a comprehensive review by [19] highlighted the role of ML in various aspects of agricultural management, underscoring its potential in handling the challenges of establishing knowledge-based farming systems. They identified the prevalence of models such as Artificial Neural Networks (ANNs), including sub-types like CNNs and Recurrent Neural Networks (RNNs). Ensemble Learning (EL) methods and Support Vector Machines (SVMs) were also widely used. This broad spectrum of methodologies showcases the adaptability and potential of ML in advancing agricultural practices.

All these studies emphasize the increasing role of ML in agriculture. They demonstrate how this technology and high-resolution satellite and UAV data can contribute to more precise crop management and disease detection. They also reveal the potential for ML to tackle the challenges in the establishment of knowledge-based farming systems. Given their focus on ML application in agriculture—particularly in crop management, disease detection, and remote sensing data—they are relevant to our study and provide foundational knowledge that could inform our research. Over the years, the evolution in methodologies and results underlines the continuing improvement in the field and the potential for further development.

3 EXPERIMENTAL METHODOLOGY

This section presents the experimental methodology used in the benchmarking experiments of the ML algorithms. An overview of the flow of experiments, including sub-steps, is shown in Figure 1. The following subsections will explain each step in detail.

3.1 Dataset

The dataset used in this paper was the Brazilian Coffee Scenes [20], which comprises multi-spectral images taken by the SPOT satellites in 2005. The dataset includes images of four cities in the state of Minas Gerais: Arceburgo, Guaranésia, Guaxupé, and Monte Santo. Item 1 in Figure 1 presents a condensed version of the dataset. The satellite captured a mosaic that was split into 2,876 images of 64x64 pixels each (item 2 in the figure). Every image has three bands - green, red, and near-infrared. The dataset also includes images of coffee plants at various ages and states of health. Researchers labelled the images into two categories: "coffee" and "non-coffee". If less than 10% of pixels in an image are composed of coffee plants, it is classified as "noncoffee", while images with more than 85% of pixels with coffee plants are labeled as "coffee". The data used for experimentation consists of 1,438 images in each class, so it has already been balanced.

This dataset was chosen because of its balanced class distribution and low data preprocessing requirements. This aids in limiting biases and potential issues associated with imbalanced data, resulting in accurate and impartial model training, evaluation, and subsequent analysis, as well as reducing the possibility of introducing excessive complexity or potential errors during the preprocessing stage. To extract the band values from the images shown in item 3 of Figure 1, we utilized the `Pillow`[1] package of the Python programming language. The resulting array for each image consists of values for the red, green, and infrared bands of every pixel, generating an array of 12,288 values (64x64x3).

3.2 Data Preprocessing

As the original data are composed of multispectral images, three different versions of datasets were generated:

- **Dataset 1:** includes all pixel values of images in the red, green, and infrared bands;
- **Dataset 2:** includes the pixel values along with each pixel's Normalized Vegetation Index (NDVI), obtained through the equation (1):

$$NDVI = \frac{Infrared - red}{NIR + Infrared} \tag{1}$$

  This index helps evaluate the plant's health by showing the proportion of reflection of near-infrared waves in relation to the red band. The higher the NDVI value, the healthier the plant;
- **Dataset 3:** condenses the existing information for each image by including the average values of green, red, and infrared pixels as well as the average NDVI of the image.

[1] https://pillow.readthedocs.io/en/stable/index.html

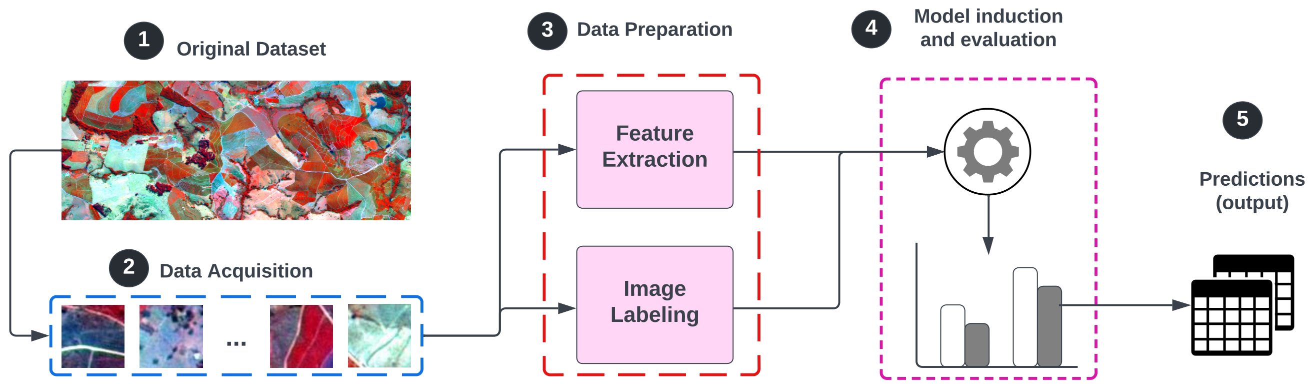


Figure 1: Methodology for comparing algorithms on identifying coffee plantations.

3.3 ML and DL Algorithms

The algorithms used in our experiments were chosen from related works and yielded good predictive performance in similar tasks. We used the following traditional ML algorithms: *k-Nearest Neighbors* (kNN), *Naïve Bayes* (NB), *Decision Trees* (DTs), *Multilayer Perceptron* (MLP), *Support Vector Machines* (SVMs) and *Random Forest* (RF). Each algorithm has a different inductive bias, resulting in different mappings between image characteristics (band values) and corresponding classes.

The kNN algorithm categorizes samples based on how close their descriptors are to the training set. The hyperparameter k decides how many nearby neighbors are taken into account for classification. We used the default value of $k = 5$ to calculate the number of neighbors in the class. The NB algorithm is based on the Bayes Theorem and conditional probabilities [21]. Its Gaussian version was explored in the experiments, since it can handle continuous features, such as the pixel values. Decision Trees (DTs) build hierarchical models iteratively selecting the most information features from the feature space. In a tree, each inner node describes a feature, while the leaves are labelled with the classes. Trees tend to be simpler and interpretable. A MLP feed-forward artificial neural network, consisting of fully connected neurons with a nonlinear activation function, organized in at least three layers. It is widely applied to solve nonlinear problems by mixing different numbers of units (neurons) into convex regions in a hyper-space of features. The SVC is a non-probabilistic linear classifier that separates the data through decision limits defined by a hyperplane, solving an optimization problem based on minimizing a risk structure. SVCs use numerical transformations via *kernels* to transform nonlinear hyperspaces into linear hyperspaces. Finally, RF is an ensemble algorithm which combines several and different DTs generated by a bootstrap resampling strategy, sampling different subsets of instances and features. The ensemble prediction is then computed by the majority voting of the "forest". All the algorithms were implemented in Python using the `Scikit-Learn`[2] library, with their correspondent default hyperparameter values.

We also compared these algorithms with two state-of-the-art Deep Learning (DL) baselines: a simple Convolution Neural Network (CNN) and the VGG16 model [22], extracted from the dataset's original paper [23]. In their study, the authors benchmarked several CNN architectures in Remote Sensing Image datasets. The CNN baseline would be the simplest architecture designed for image recognition (lower baseline). At the same time, the VGG16 was the top-ranked model in several datasets, including the Brazilian Coffee Scenes dataset (higher baseline). These networks' architectures are summarized in Table 1. It is important to mention that the VGG16 model used here was pre-trained in the *ImageNet* dataset, and only the dense layers' weights were trainable in our experiments.

Table 1: DL architectures explored in experiments. Layers' names follow the `Keras` nomenclature.

Model	Layer	Options	f
CNN	Conv2D	32 (3,3)	ReLU
	maxPooling2D	(2,2)	
	Dropout	0.25	
	Conv2D	64 (3, 3)	ReLU
	MaxPooling2D	(2,2)	
	Dropout	0.25	
	Conv2D	64 (3, 3)	ReLU
	Flatten	-	-
	Dense	64	ReLU
	Dropout	0.5	
	Dense	1	Sigmoid
VGG16	VGG16	baseModel	-
	Flatten	-	-
	Dense	4096	ReLU
	Dense	1	Sigmoid

[2] https://scikit-learn.org/stable/

3.4 Evaluation and Reproduction of Experiments

Traditional ML algorithms were trained with preprocessed datasets, while DL was fed with raw images. We split the same images/examples into training and testing partitions in both cases, ensuring that all models evaluate the same data. Due to the large number of examples (2,876), the *holdout* data separation methodology was chosen, where 70% of the data was used for training and the remaining 30% for testing. Furthermore, the experiments were repeated 30 times using different seed generators. The performance metric used to evaluate the results was Balanced Accuracy per Class (BAC). The traditional algorithms were coded in Python using the scikit-learn library.

The DL algorithms were coded with Keras and TensorFlow, but running in a CPU environment. They were trained in 100 epochs, minimizing the binary cross entropy loss function. In a single execution, a total of 30% of the training data is used as the validation data.

We empirically defined a batch size equal to 16 and added an early stop criteria for finishing the training when the validation accuracy could not increase in 10 successive epochs. The chosen optimizer was the default choice in Keras (Adam). The complete experimental setup is detailed in Table 2. The datasets, codes, and experimental results are publicly available on GitHub https://github.com/ribeiro-julio/crop-image-classification

4 RESULTS

Figure 2 depicts the experimental results obtained when executing all the traditional algorithms over all generated datasets. In the figure, the x-axis lists all the algorithms decreasingly ordered according to their balanced accuracy values, while violins show the performance distribution across 30 different repetitions. In all the datasets, the higher accuracies were obtained by Support Vector Machines (SVM) and Random Forest (RF) algorithms. They obtained accuracy values higher than 0.8 in two datasets, an empirical threshold defined in experiments as reference, and identified by the red dotted line.

One may note that the obtained results in datasets 1 and 2 were quite similar. These datasets have quite the same information, with dataset 2 also including each pixel's NDVI values. Results suggest that this inclusion did not significantly improve the induced models. Regarding performance values, we can look closely at the top-ranked algorithms. SVMs decreased its average accuracy from dataset 1 to dataset 2 (0.816 and 0.814), while RF slightly increased (0.811 and 0.823). However, both worsened in dataset 3; they lost information when the average pixel values were explored as image descriptors (0.762 and 0.757), respectively. Even though they were still the best algorithms when compared to the others, since no other algorithm obtained accurate results.

We applied the Wilcoxon paired-test [24] with $\alpha = 0.05$ to compare SVMs and RF performances in datasets 1 and 2. These results indicated that RF induced in dataset 2 is statistically better than any other algorithm in any dataset. Thus, we provided further analysis considering just the predictions obtained in dataset 2.

4.1 Comparison with DL models

Figure 3a summarizes the results of the comparison between DL and traditional ML algorithms. Algorithms are decreasingly ordered on the x-axis according to their BAC average values. VGG16 was the best (0.834), followed by CNN (0.831), RF (0.821) and SVM (0.814). All the algorithms were executed in the same training and testing sets over 30 different repetitions. Thus, we also applied the Wilcoxon paired test to compare their performances. VGG is statistically better than RF and SVM. RF has no statistical difference with the CNN model, while SVM is statistically worse than any other algorithm.

The analysis can be complemented with the Figure 3b. It presents the distribution of epochs required by each DL model to be trained. None required the limit of epochs defined as 100 epochs. On average, CNN trained in 23 epochs, while VGG16 required 13. This behavior is expected since VGG16 was pre-trained in the *ImageNet* dataset and thus acted as a better feature descriptor. However, it does not reflect in terms of performance, which suggests that original images are "simple". The images do not have high-level features in general, as there is no level of object detailing resulted from the images resolution and which consequently does not provide class distinction. It would be reasonable to argue that no shapes, textures, or complex objects are being identified; just the pixel values are used to make predictions. In the same direction, it is a positive aspect in favor of the RF, achieving accurate values and requiring much less time to obtain predictions.

4.2 Obtained predictions

Figure 4 details the algorithms' predictions in dataset 2. All the algorithms are listed on the y-axis, while the individual images are projected on the x-axis. The "Y" row shows the actual target values (classes). Whenever a row contains a cell with a color different than this real class (Y row), it shows a misclassification.

Different algorithms presented different patterns, which is expected since they follow different learning biases. All the algorithms performed quite similarly, predicting the "non-coffee" regions (gray predictions), alternating the misclassifications between different images. Conversely, DL algorithms (CNN, VGG16) were the most accurate models in the coffee class. We identified 97 images by all the top algorithms: 34 images of non-coffee regions and 63 of coffee. Figure 5 shows these misclassified images: the first line contains examples of "*coffee*" images, while the second one depicts examples of "*non-coffee*" images.

Analyzing the coffee images, it is expected they present more green pixels, indicating healthy plants. Figure 5 a) and b) show two examples following this property, and it might be the case that algorithms did not perform well in recognizing healthy regions/plants. However, Sub-figures 5 c) and d) are '*confusing*' because there have more red pixels than green pixels. A possibility is that they contain some satellite capture error due to noise or atmospheric effects, which justify misclassifications. In the non-coffee images, the opposite is desired: images should present lower green values than red values. It can be identified in sub-figures 5 e) to h). On the other hand, some of them (g and h) seem to present a certain level of green values that might confuse the algorithm-defined decision boundary.

Table 2: Complete experimental setup.

Element	Option	R/Python package
Traditional ML algorithms	k-Nearest Neighbors (kNN) Naïve Bayes (NB) Decision Trees (DT) Support Vector Machines (SVMs) Random Forest (RF) Multilayer Perceptron (MLP)	`Python: scikit-learn`
Resampling	Holdout Training set = 70% Testing set = 30%	`Python:scikit-learn`
Evaluation measure	Balanced per Class Accuracy	`Python: scikit-learn, Keras`
DL algorithms	Convolutional Neural Network (CNN) VGG16	`Python: Keras`
DL setup	epochs = 100 validation split = 0.3 batch size = 16 optimizer = Adam optimized measure = Binary cross entropy early stop criteria = 10 epochs (val accuracy)	`Python: Keras`
Repetitions	30 times with different seeds seeds = $\{0, \ldots, 29\}$	-
Statistical Evaluation	Wilcoxon α = 0.05	`R: stats`

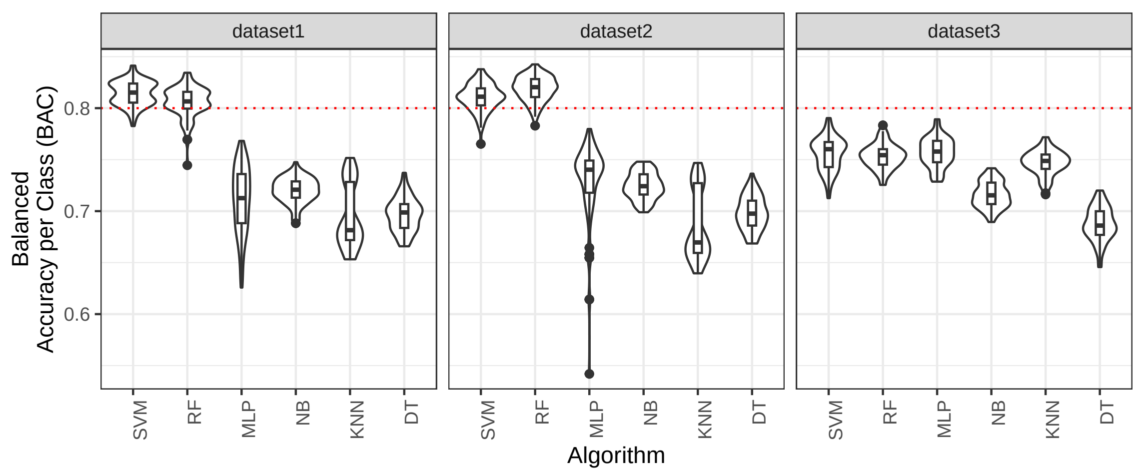
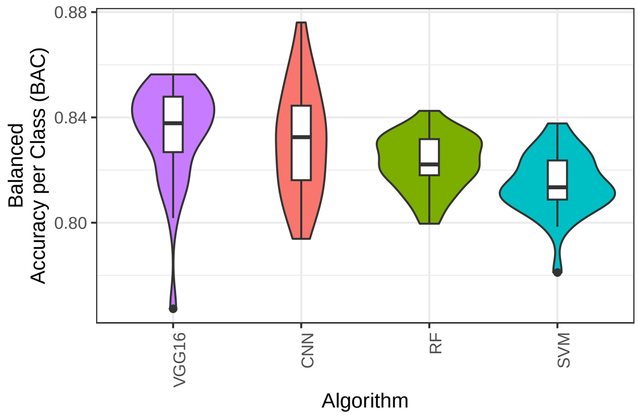
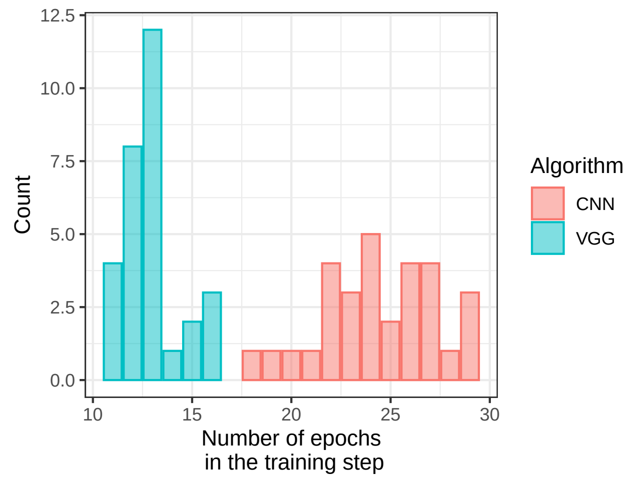


Figure 2: Overall results obtained using three different datasets for the coffee x non-coffee problem.

Another hypothesis to explain the misclassifications is that these images present the same pixel distribution in the three channels; thus, algorithms cannot differentiate them. Figure 6 presents the pixel distribution of the three image channels (red, green, near-infrared) for the misclassified images. The figure shows that green and near-infrared distributions differ between classes: coffee images present higher values of green and near-infrared pixels and non-coffee lower ones. However, red values present a similar distribution, which makes our hypothesis partially true. It may be the case that these values are making it difficult to separate the classes and emphasize the need for more robust feature descriptors would have to solve such misclassification and issues.



(a) Violin plots of top-ranked ML and DL and models.



(b) Number of epochs required to train CNN and VGG models.

Figure 3: Comparison of DL models with top-ranked traditional ML algorithms.

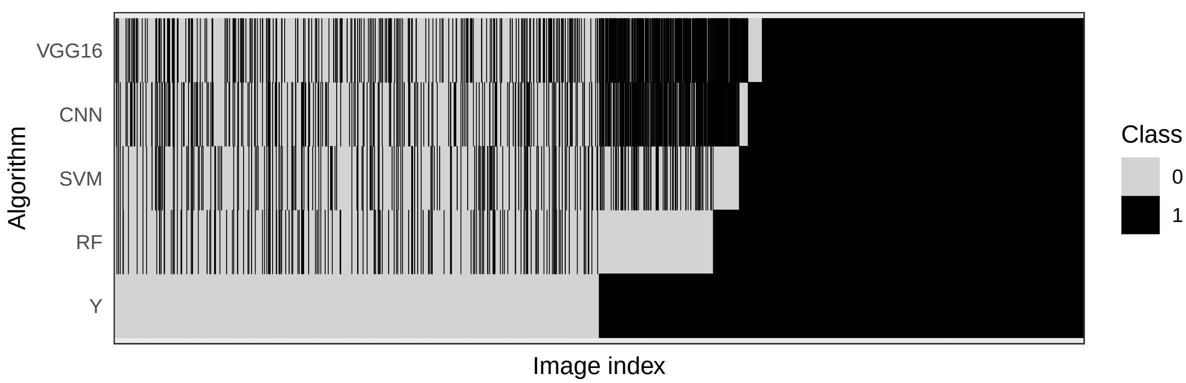


Figure 4: Obtained predictions by the top-ranked learners (DL and traditional ML algorithms).

5 CONCLUSIONS

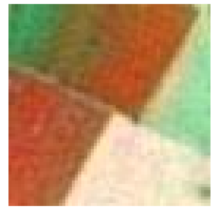
This study compares different ML and DL algorithms in a coffee and non-coffee image classification dataset. The comparison among traditional ML algorithms, including SVMs, RF, NB, kNN, and DTs revealed that SVMs and RF algorithms performed better, with an average balanced accuracy greater than 0.8. The Wilcoxon test corroborated this finding, demonstrating that RF is statistically better to the other algorithms. The second preprocessed dataset produced the best results, demonstrating that the NDVI index affected the classification.

DL baselines (VGG16 and CNN) were slightly better to RF and SVMs. Both obtained BAC values higher than 0.83. The VGG16 model required fewer epochs to train but spent more computational time on this task due to the model's higher number of parameters (network weights). However, there was no statistical difference between them. Compared to the traditional ML algorithms, VGG was statistically better to both (RF and SVMs), but there was no statistical difference between RF and CNN. VGG not being significantly better than a shallower CNN which reinforces that no high-level features are being discernible for class distinction, meaning that, increasing the complexity of the network or another algorithm boundary will not necessarily increase the ability of these algorithms to solve the problem.

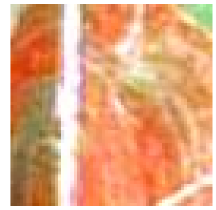
Analyzing their predictions, it was possible to identify that coffee images were more challenging to categorize than non-coffee images. Coffee images should have more green pixels, indicating healthy plants, and non-coffee images should have more red pixels, as the channels used are around reddish channels, it is expected that the non-coffee class would be able to better utilize channel variations. Some coffee images had a high proportion of red pixels, and some non-coffee images had a high proportion of green pixels, resulting in misclassifications. This could have happened due to noise or atmospheric effects during the image extraction. Additionally, it was found that the red pixel distribution on these images is comparable

(a) Arceburgo-T1536-B1600-L4992-R5056

(b) Arceburgo-T2240-B2304-L4992-R5056



(c) Guaxupe-T960-B1024-L1536-R1600



(d) Montesanto-T3392-B3456-L10496-R10560

(e) Guaranesia-T5632-B5696-L3200-R3264

(f) Guaxupe-T7424-B7488-L2240-R2304

(g) Guaranesia-T3200-B3264-L4288-R4352

(h) Guaranesia-T4288-B4352-L1664-R1728

Figure 5: Examples of misclassified images by all the algorithms in experiments. The first row presents images with the label value as "*coffee*", while the second presents "*non-coffee*" images.

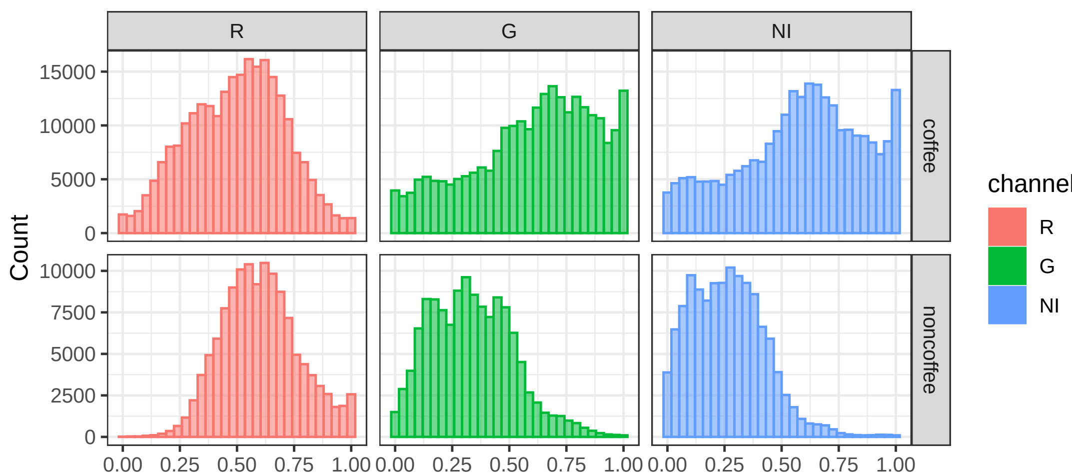


Figure 6: Channels histogram of misclassified images.

in coffee and non-coffee classes, making it challenging for the algorithms to categorize them.

The dataset has only images with the three bands (red, green, and near-infrared) available. Multi-spectral images typically include additional bands, such as red edge and near-infrared, in addition to the primary bands (red, green, and blue), not only in the sense of having more channel options, but also by increasing the number, not just three like RGB images or those used in this work. This work may have been enhanced if the images had more bands than provided. These additional bands would enable further analysis of the inaccurately classified images. Furthermore, this research can be extended in four different ways: (1) increasing the amount of data by extracting additional image characteristics, such as texture; (2) using filters to pre-process the images before feeding the algorithms; (3) using other agricultural indexes to test their impact on classification; and (4) experimenting with different deep learning algorithms.

REFERENCES

[1] Chunhua Zhang and John M. Kovacs. The application of small unmanned aerial systems for precision agriculture: a review. *Precision Agriculture*, 2012. doi: 10.1007/s11119-012-9274-5.

[2] G.A.S. Ginigaddara and A.N. Kodithuwakku. Chapter 23 - legumes for energy efficiency in agricultural systems. In Ram Swaroop Meena and Sandeep Kumar, editors, *Advances in Legumes for Sustainable Intensification*, pages 441–460. Academic Press, 2022. ISBN 978-0-323-85797-0. doi: https://doi.org/10.1016/B978-0-323-85797-0.00019-7. URL https://www.sciencedirect.com/science/article/pii/B9780323857970000197.

[3] Andreas Kamilaris and Francesc X. Prenafeta-Boldú. Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 2018. doi: 10.1016/j.compag.2018.02.016.

[4] Prasad S Thenkabail, Murali Krishna Gumma, Pardhasaradhi Teluguntla, and AM Irshad. Hyperspectral remote sensing of vegetation and agricultural crops. *Photogrammetric Engineering & Remote Sensing (TSI)*, 80(8):695–723, 2014.

[5] David J. Mulla. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering*, 114(4):358–371, 2013. ISSN 1537-5110. doi: https://doi.org/10.1016/j.biosystemseng.2012.08.009. URL https://www.sciencedirect.com/science/article/pii/S1537511012001419. Special Issue: Sensing Technologies for Sustainable Agriculture.

[6] H. Gaznayee, Sara H. Zaki, Ayad M. Fadhil Al-Quraishi, Payman Hussein Aliehsan, Kawa K. Hakzi, H. Razvanchy, M. Riksen, and Karrar Mahdi. Integrating remote sensing techniques and meteorological data to assess the ideal irrigation system performance scenarios for improving crop productivity. *Water*, 2023. doi: 10.3390/w15081605.

[7] Lei Ma, Yu Liu, Xueliang Zhang, Yuanxin Ye, Gaofei Yin, and Brian Alan Johnson. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152:166–177, 2019. ISSN 0924-2716. doi: https://doi.org/10.1016/j.isprsjprs.2019.04.015. URL https://www.sciencedirect.com/science/article/pii/S0924271619301108.

[8] M. I. Jordan and T. M. Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260, 2015. doi: 10.1126/science.aaa8415. URL https://www.science.org/doi/abs/10.1126/science.aaa8415.

[9] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012. URL https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

[10] Konstantinos G. Liakos, Patrizia Busato, Dimitrios Moshou, Simon Pearson, and Dionysis Bochtis. Machine learning in agriculture: A review. *Sensors*, 18(8), 2018. ISSN 1424-8220. doi: 10.3390/s18082674. URL https://www.mdpi.com/1424-8220/18/8/2674.

[11] Jingzheng Zhao, Liyuan Wang, Hui Yang, Penghai Wu, Biao Wang, Chengrong Pan, and Yanlan Wu. A land cover classification method for high-resolution remote sensing images based on ndvi deep learning fusion network. *Remote Sensing*, 14(21), 2022. ISSN 2072-4292. doi: 10.3390/rs14215455. URL https://www.mdpi.com/2072-4292/14/21/5455.

[12] Jorge Torres-Sánchez, Francisca López-Granados, N. Serrano, O. Arquero, José M. Peña, José M. Peña, José M. Peña, and José Manuel Sánchez Pena. High-throughput 3-d monitoring of agricultural-tree plantations with unmanned aerial vehicle (uav) technology. *PLOS ONE*, 2015. doi: 10.1371/journal.pone.0130479.

[13] Charu C. Aggarwal. *Neural Networks and Deep Learning*. Springer International Publishing, 2018. doi: 10.1007/978-3-319-94463-0.

[14] Rodrigo Fernandes de Mello and Moacir Antonelli Ponti. *Machine learning: a practical approach on the statistical learning theory*. Springer, 2018. doi: 10.1007/978-3-319-94989-5.

[15] S. Cordero-Sancho and S. A. Sader. Spectral analysis and classification accuracy of coffee crops using landsat and a topographic-environmental model. *International Journal of Remote Sensing*, 28(7):1577–1593, 2007. doi: 10.1080/01431160600887680. URL https://doi.org/10.1080/01431160600887680.

[16] Khenilyn P. Lewis and Juancho Espineli. Classification and detection of nutritional deficiencies in coffee plants using image processing and convolutional neural network (cnn). 2020.

[17] Maitiniyazi Maimaitijiang, Vasit Sagan, Paheding Sidike, Ahmad M. Daloye, Hasanjan Erkbol, and Felix B. Fritschi. Crop monitoring using satellite/uav data fusion and machine learning. *Remote Sensing*, 12(9), 2020. ISSN 2072-4292. doi: 10.3390/rs12091357. URL https://www.mdpi.com/2072-4292/12/9/1357.

[18] Aicha Moumni and Abderrahman Lahrouni. Machine learning-based classification for crop-type mapping using the fusion of high-resolution satellite imagery in a semiarid area. *Scientifica*, 2021:8810279, Apr 2021. ISSN 2090-908X. doi: 10.1155/2021/8810279. URL https://doi.org/10.1155/2021/8810279.

[19] Lefteris Benos, Aristotelis C. Tagarakis, Georgios Dolias, Remigio Berruto, Dimitrios Kateris, and Dionysis Bochtis. Machine learning in agriculture: A comprehensive updated review. *Sensors*, 2021. doi: 10.3390/s21113758.

[20] Otávio AB Penatti, Keiller Nogueira, and Jefersson A Dos Santos. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 44–51, 2015.

[21] T.M. Mitchell. *Machine Learning*. McGraw-Hill International Editions. McGraw-Hill, 1997. ISBN 9780071154673.

[22] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*, San Diego, CA, USA, 2015. CoRR.

[23] Keiller Nogueira, Otávio A.B. Penatti, and Jefersson A. dos Santos. Towards better exploiting convolutional neural networks for remote sensing scene classification. *Pattern Recognition*, 61:539–556, 2017. ISSN 0031-3203. doi: https://doi.org/10.1016/j.patcog.2016.07.001.

[24] G. Santafe, I. Inza, and J.A. Lozano. Dealing with the evaluation of supervised classification algorithms. *Artif Intell Rev*, 44:467—-508, 2015. doi: https://doi.org/10.1007/s10462-015-9433-y.