Automatic Detection of Abandoned Vineyards Using Aerial Imagery

Igor Teixeira [1]
igor_teixeira@hotmail.com
Danilo Leite [1]
danilol@utad.pt
Joaquim J. Sousa [1,2]
jjsousa@utad.pt
António Cunha[1,2]
acunha@utad.pt

[1]Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal

[2] INESC Technology and Science (INESC-TEC), Porto, Portugal

Abstract

The European Union (EU) has established, through the Common Agricultural Policy (CAP), a system of aid and subsidies for farmers cultivating vineyards. Eligible areas must be monitored and registered in Geographic Information Systems. The agencies providing this support must verify that the parcels are engaged in agricultural activity through on-site checks or the analysis of aerial or satellite images. Abandonment situations lead to the cancellation of aid payments. In the Douro Demarcated Region of Portugal, inspections are conducted according to methods defined by the EU. However, due to the vast size of the region, the time required for analysis and the specialized human resources needed for these inspections are significant. In this study, a dataset was created to train convolutional neural networks (CNNs), and pre-trained VGG models were fine-tuned to classify vineyards as abandoned or non-abandoned. The model achieved an accuracy of 95.1% on the test dataset, while the top-performing model achieved an impressive overall accuracy and F1 score of 99% for both classes.

1 Introduction

The European Union (EU) has implemented a system of aid and subsidies for farmers engaged in vineyard cultivation through the Common Agricultural Policy (CAP). To qualify for these benefits, vineyard areas need to undergo control and registration processes, which rely on cartographic and cadastral information managed in Geographic Information Systems (GIS) [1]. The agencies responsible for disbursing agricultural support must ensure that agricultural activity is present in the parcels through on-site checks or analysis of ortho-rectified aerial or satellite images. Aid payments may be cancelled in cases of abandonment. In the Douro Demarcated Region (DDR), the Port and Douro Wines Institute (IVDP) is the government agency tasked with overseeing not only the production quotas for individual farmers but also the updating of cadastral information. This is crucial to prevent situations in which farmers falsely declare wine production on parcels where vineyards have been abandoned and are no longer in operation.

Currently, field inspections only cover a small percentage (5%) of farmers with crop declarations [2]. As a result, the EU has mandated that member states use automatic methods, such as machine learning techniques, using satellite data from Sentinel-1 and Sentinel-2, for analyzing cultivated areas [2]. In the DDR, inspections are conducted using the same methods defined by the EU. However, due to the vast size of the region, approximately 250,000 hectares, with vineyard cultivation occupying 43,843 hectares, the process requires significant analysis time and specialized human resources. Manual detection of abandonment situations is expensive in the DDR, which is a mountainous viticulture region, and analyzing vineyard images is particularly challenging due to steep slopes compared to other regions.

Deep Learning (DL) algorithms have been employed in various tasks, including the identification and classification of crop status [3], as well as land abandonment detection. Convolutional Neural Networks (CNNs), a type of DL method, have shown outstanding performance in detecting patterns, making them widely used in image classification problems.

In the context of abandonment detection, noteworthy studies include [4]. The first study utilized a combination of CNNs, very high-resolution aerial imagery, and Sentinel-2 data to classify crops, resulting in a 93% accuracy in crop classification and an 88% accuracy in detecting permanent crop abandonment. In the second study, the authors employed LSTM and Bi-LSTM networks along with five spectral indices to detect abandoned parcels, achieving an overall accuracy of 94.6%, outperforming other Machine Learning models.

In this research, we propose a methodology for classifying abandoned and non-abandoned vineyard parcels in the Douro Demarcated Region using DL models on high-resolution aerial imagery. Specifically, we used a private set of aerial images obtained from photogrammetric flights conducted over six different years, which offer superior spatial resolution compared to satellite images.

2 Methods and materials

We curated a novel dataset to train Convolutional Neural Networks (CNNs) for classifying abandoned and non-abandoned vineyards in the DDR. This dataset, to the best of our knowledge, is unique in its use of high-resolution aerial imagery for abandoned vineyard detection. The dataset was created using ArcGIS Pro software, where a mosaic dataset of the entire target region was generated. The Extract by Mask geoprocessing tool was then utilized to extract a total of 7,779 images, representing productive (7,300) and abandoned (479) vineyards, using GIS polygons as masks. Data labeling was done based on cadastral and GIS information, and subsequently verified manually. Given the imbalanced nature of the dataset and the context of the problem, several experiments were conducted to address potential overfitting. This included oversampling of the least represented class through random left and right rotations, horizontal flips, as well as the inclusion of Dropout and L2 regularization techniques. A pipeline was defined using the Augmentor library to generate new samples, involving left and right rotations of up to fifteen degrees with a 50% probability, and horizontal flips with the same probability. As a result, 7,279 new images of abandoned vineyard parcels were obtained.

All images in the dataset were resized to 224x224 pixels and then shuffled. The dataset was divided into training, validation, and test sets in the proportion of 80-10-10, respectively. The balanced dataset consisted of 11,663 images for training, 1,458 images for validation, and 1,458 images for testing. The architectures for computer vision, were explored for the purpose of classifying the two classes of images. Initial experiments were conducted as a baseline to investigate the behaviour of the neural networks when using a completely new dataset. The experimental configurations for all the models are summarized in Table 1.

Exp.	Model	Epochs
E1	Convnet	50
E2	Convnet + Oversampling	50
E3	Convnet + Oversampling + Regularizer	100
E4	VGG16 + Oversampling + Regularizer	100
E5	VGG19 + Oversampling + Regularizer	100

Table 1. – Experiments configurations.

The CNN model utilized in Experiment 1 consists of eleven sequential layers. The initial layers are Conv2D layers with 32 filters, a kernel size of 3x3, and a ReLU activation function. Each Conv2D layer is followed by a MaxPooling2D layer with a pool size of 2x2. After the third MaxPooling2D layer, there is a 25% Dropout layer. This is followed by a Flatten layer, a Dense layer with 128 neurons and a ReLU activation function, and a 50% Dropout layer. The final output layer is a Dense layer with two neurons and a softmax activation function. It's worth noting that this model was trained on an unbalanced dataset, while the remaining were trained on a balanced version of the dataset.

Experiments 2 and 3 were configured with the same architecture as Experiment 1, but the Experiment 3 model uses kernel regularization in one of the dense layers with a weight of 0.01. In Experiments 4 and 5, pre-trained Visual Geometry Group networks, VGG16 and VGG19 respectively, were employed. All VGG layers were frozen to prevent weight adjustment. These models include two additional dense layers stacked on top of the VGG network. The first dense layer has 128 neurons, uses a ReLU activation function, and has an L2 regularizer of 0.01. The second layer is an output layer with two neurons, preceded by a 50% dropout layer, and uses the softmax activation function for predicting classification into two categories. All models were compiled with the categorical cross-entropy loss function and used the Adaptive Moment Estimation (Adam) optimizer. The performance of the models was assessed using various methods. Firstly, the training and validation accuracy and loss were plotted and analysed. Additionally, the confusion matrix and classification report of the model's predictions were evaluated to assess precision, recall, overall accuracy, and F1-score.

In conclusion, the Grad-CAM technique was employed to visually evaluate the results of the predictions and identify the regions of the images that the model focused on. This was done to verify that the model effectively learned the distinguishing characteristics between abandoned and non-abandoned plots

3 Experiments and Results

The outcomes of the trained models are summarized in Table 2. Despite the requirement for ample data by CNN models, the baseline model achieved an overall accuracy of 95.4% and an F1-score of 56.1%. The lower F1-score can be attributed to the imbalanced distribution of data in the test set, with significantly fewer images in the abandoned class (47) compared to the non-abandoned class (730). Among the experimental models, E5 demonstrated the best performance, with an overall accuracy and F1-score of 98.6%, precision of 99.6%, and recall of 97.7%.

Exp.	Acc.	Pre.	Rec.	F1 Score
E1	0.954	0.489	0.657	0.561
E2	0.977	0.988	0.966	0.977
E3	0.979	0.992	0.968	0.980
E4	0.975	**0.996**	0.957	0.976
E5	**0.986**	**0.996**	**0.977**	**0.986**

Table 2. – Experiments results

The results clearly demonstrate that oversampling the dataset had a substantial positive impact on the performance, highlighting the importance of training CNN models with balanced data. Experiment E2, compared to E1, showed significant improvement with an accuracy of 97.7% in the test set. Notably, precision and F1-score metrics also showed substantial improvement, reaching 98.8% and 97.7% respectively. Introduction of the L2 regularizer in experiment E3 resulted in a slight increase in all evaluated metrics. Unexpectedly, the use of the pre-trained VGG16 model only improved the precision metric, indicating the presence of bias that may require additional training as reported in [4], [5]. Further experimentation with different settings on stacked layers could potentially enhance performance. Experiment E5 achieved the best results across all evaluated metrics, matching only by E4 in terms of precision. In terms of model size, the experiment E4 model was significantly larger than the E3 model, with a magnitude of six times greater. However, this increase in model size did not lead to an overall improvement in performance, except for the Precision metric. This is because the transfer learning technique was employed in experiment E4 to reduce the training time.

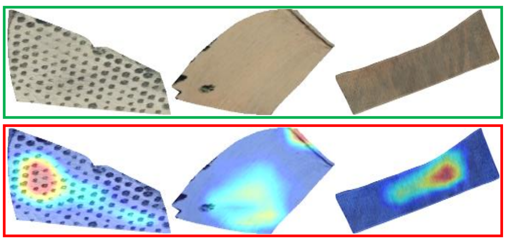

Fig. 1 - Wrong predictions. Green: ground truth; Red: Grad-CAM.

In experiment E5, a deeper and more complex network was utilized, resulting in 5,309,826 trainable parameters out of a total of 20,614,594. This approach resulted in the best performance across all evaluated metrics. In contrast, experiment E3 had 2,788,674 trainable parameters, almost five times more than experiment E5. This suggests that increasing the number of trainable parameters and the complexity of the network allowed for the best results in the context of this work.

The Grad-CAM visualization for the top plot in Fig. 1 reveals that resizing reduced the visibility of the row spacing between vines, which may have contributed to prediction errors. Additionally, the vineyard being relatively young could have also impacted the accuracy due to limited vegetation, as seen in the middle plot. It has been suggested in previous studies [6] that an additional class for bare soil could be considered, as shown in the middle plot, to prevent incorrect predictions. Similarly, the inclusion of a separate class for trees, as demonstrated in the bottom plot, could potentially improve prediction accuracy.

4 Conclusions

This study introduces a new dataset of aerial imagery for abandoned vineyard detection, aiming to classify vineyards as abandoned or non-abandoned using CNN models. To address the challenge of dataset imbalance, several experiments were conducted, including oversampling, regularization techniques, and the use of pre-trained models, which significantly improved performance. Experiment E5 achieved the best results with 98.6% overall accuracy and F1-score, underscoring the importance of balanced data in training CNN models for image classification tasks.

Future work will involve incorporating new images obtained from updated geometries registered in the GIS. The models will be retrained, and results evaluated. The goal is to create a tool, similar to the one presented in [7], to aid governmental agencies in automatic abandonment detection, complementing manual verification. To enhance prediction accuracy, it is necessary to create classes that differentiate between young vineyards, adult vineyards, bare soil, trees, and vineyards grown intercropped with trees. Additionally, using data from all four spectral bands (RGB and NIR) instead of just RGB, as done in this study, may provide additional relevant information about the land's condition, potentially improving accuracy.

Acknowledgements

Authors would like to acknowledge the Vine&Wine Portugal Project, co-financed by the RRP - Recovery and Resilience Plan and the European NextGeneration EU Funds, within the scope of the Mobilizing Agendas for Reindustrialization, under reference C644866286-00000011.

References

[1] European Commission, Directorate-General for Agriculture and Rural Development, Commission Delegated Regulation (EU) 2018/273. 2017.

[2] European Court of Auditors, Using new imaging technologies to monitor the Common Agricultural Policy: steady progress overall, but slower for climate and environment monitoring. Special report No 04, 2020. Publications Office, 2021.

[3] L. Zhong, L. Hu, and H. Zhou, 'Deep learning based multi-temporal crop classification', Remote Sensing of Environment, vol. 221, pp. 430–443, 2019.

[4] L. A. Ruiz, J. Almonacid-Caballer, P. Crespo-Peremarch, J. A. Recio, J. E. Pardo-Pascual, and E. Sánchez-García, 'Automated Classification of Crop Types and Condition in a Mediterranean Area Using a Fine-Tuned Convolutional Neural Network', in The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Copernicus Publications, pp. 1061–1068, 2020.

[5] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, 'Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization', Int J Comput Vis, vol. 128, no. 2, pp. 336–359, 2020.

[6] L. Agilandeeswari, M. Prabukumar, V. Radhesyam, K. L. N. B. Phaneendra, and A. Farhan, 'Crop Classification for Agricultural Applications in Hyperspectral Remote Sensing Images', Applied Sciences, vol. 12, no. 3, 2022.

[7] A. Lozano-Tello et al., 'Crop identification by massive processing of multiannual satellite imagery for EU common agriculture policy subsidy control', European Journal of Remote Sensing, vol. 54, no. 1, pp. 1–12, 2021