UAV4TREE: DEEP LEARNING-BASED SYSTEM FOR AUTOMATIC CLASSIFICATION OF TREE SPECIES USING RGB OPTICAL IMAGES OBTAINED BY AN UNMANNED AERIAL VEHICLE

Roberto Pierdicca[a,*], Lindo Nepi[b], Adriano Mancini[b], Eva Savina Malinverni[a], Mattia Balestra[c]

[a]Department of Civil, Building Engineering and Architecture (DICEA), Via Brecce Bianche, Ancona, 60131, Italy, IT
[b]Department of Information Engineering (DII), Via Brecce Bianche, Ancona, 60131, Italy, IT;
[c]Department of Agricultural, Food and Environmental Sciences (D3A), Via Brecce Bianche, Ancona, 60131, Italy, IT ;

KEY WORDS: Deep Learning, Artificial Intelligence, RGB optical images, Unmanned Aerial Vehicle, Remote Sensing.

ABSTRACT:

Automated tree classification from unmanned aerial vehicle (UAV) images is a challenging task with several applications in forest management and conservation. In this study, we propose UAV4Tree a Deep Learning based system that automatically classifies RGB optical images obtained by the UAV. In particular, we explore the use of augmented datasets and various deep learning models, including ResNet, DenseNet, InceptionV3, and Vision Transformer, for the classification of tree images obtained from UAVs. Our experiments show that the use of an augmented dataset can significantly improve the accuracy of the classification by approximately 10 points compared to the use of a non-augmented dataset. We also found that fine-tuning and the introduction of dropout were essential for improving the generalization ability of the models on the augmented dataset. Furthermore, the use of Super Resolution Generative Adversarial Network (SR-GAN) in the original dataset allowed us to increase the performance of some models. Our findings provide valuable insights into the use of deep learning models for automated tree classification from UAV imagery, which has significant implications for sustainable forest management and conservation.

1. INTRODUCTION

Trees play an important role in forest science because they are involved in many aspects of the ecosystem and human life (Salleh et al., 2023). Trees remove carbon dioxide from the air and store it in their roots, trunks, and leaves. They are one of the most effective ways to combat climate change by reducing the amount of carbon emissions in the atmosphere (Palmer, 2021). They provide a habitat for many species of animals, birds, insects, and plants and are an essential component of the forest ecosystem and support biodiversity (Andreas et al., 2023). Moreover, trees invest a huge importance in soil conservation by reducing erosion and improving fertility. Their roots help to bind the soil together, preventing it from being washed away during heavy rainfall. Trees are also a valuable resource for human beings. They provide timber, fuelwood, medicines, and fruits for consumption and forests also provide recreational and cultural opportunities. Tree species mixing is a common occurrence in natural forests. As per the 2018 MODIS Collection 6 MCD12C1 IGBP land cover classification, roughly 42% of forests in Canada and 28% around the world consist of combined establishments (Sulla-Menashe and Friedl, 2018). Identifying the species compositions and interactions in the ecological setting is a significant guide for forest management techniques like reforestation, harvesting, and selective thinning (White et al., 2016).

In the dimension of the modern practices used for the detection of tree species, the adoption of airborne remote sensing technologies has remarkably increased in recent times. Such techniques have enabled the swift measurement and classification of a multitude of trees (Quan et al., 2023). Remote sensing with the use of satellites can capture extensive regions of urban forests, where cloud cover and fog can greatly impact on data collection. Unmanned aerial vehicle (UAV)-based remote sensing, being an effective method for obtaining high-quality images, has gained wide adoption in several domains, especially in forestry tree classification due to its affordability, rapid data collection, and adaptable functionality (Pereira et al., 2023).

Knowledge of tree species is crucial for effective forest management, planning, environmental protection, and statistics pertaining to forest resources. Every year, companies and governments conduct forestry investigations, which absorb a large amount of energies and financial resources (Zou et al., 2017). Forest companies desire species-specific size distributions of trees for optimal output, since traditional methods based on field inventory are laborious, time-consuming and limited by spatial extent. Consequently, remote sensing methods using large-scale aerial color or infra-red images have been introduced (Jutras-Perreault et al., 2023). While such techniques have found widespread use in forest applications, traditional optical remote sensing methods suffer from a lack of the ability to capture the complex three-dimensional structures of mixed-species forests with multiple canopy layers, particularly those that are unevenly-aged (Lovell et al., 2003). However, active remote sensing techniques (particularly those adopting laser scanning) have recently emerged as a promising alternative for forest mapping and other applications due to their ability to provide comprehensive 3D forest information.

Diverse tree species display distinct qualities in both spectral and textural aspects. Currently, some academics implement hyperspectral and LiDAR data to classify various tree species (Dian et al., 2016). Maschler et al. (Maschler et al., 2018) accomplished the automatic classification of 13 tree species with the use of hyperspectral data, revealing highly dependable results from hyperspectral data containing the visible and near-infrared

[*] r.pierdicca@staff.univpm.it

https://doi.org/10.5194/isprs-annals-X-1-W1-2023-1089-2023 | © Author(s) 2023. CC BY 4.0 License.

(NIR) spectral features. In (Liu et al., 2017), the authors identified that the variables extracted from LiDAR data outmatched spectral features for species prediction. Nonetheless, the procedure for extricating hyperspectral data is elaborate, and the spectral interval information may be redundant among different tree species, leading to inaccuracies whilst classifying. An array of weather phenomena including rain, snowfall, and fog reduce the quality of LiDAR data. In contrast, it is convenient to acquire RGB optical images. Therefore, investigating whether RGB optical images gathered through an UAV can be valuable information for tree-species classification.

Developing appropriate algorithms is crucial for the classification of tree species. Traditionally, feature selection is often mixed with the selection of a proper classifier. The classification process encompasses choosing the optimal classifier, fine-tuning and training it, testing its repeatability, and investigating its implementation. A wide range of supervised and unsupervised methods can be employed to tackle trees classification, including logistic regression (LR), probabilistic graphical models (PGM), decision tree (DT) classifiers, support vector machines (SVM), nearest-neighbor (NN) classifiers, clustering techniques like k-means, and deep learning approaches such as multi-layered perceptron and convolutional neural networks (CNNs). Each of these classification methods exhibits distinct functionality, benefits, and limitations which have been extensively explored by researchers in the field of tree species classification (Xi et al., 2020). Recent studies have shown that Deep Learning (DL) models outperform more conventional approaches such as SVM, KNN, or redclassical machine learning based methods for images classification tasks (Wang et al., 2021), (Nezami et al., 2020). In particular, the work of Wang et al. (Wang et al., 2021) compares classical machine learning techniques (e.g., SVM) and deep learning techniques (e.g., CNN). In their study, the DL-based models (e.g., CNN) outperformed classical machine learning models (e.g., SVM) in terms of accuracy for image classification on large-scale datasets. For large datasets with high spatial and spectral resolution, Deep Neural Networks (DNNs) have been shown to be particularly effective classifiers, which exhibit unique capabilities for handling high-dimensional classification challenges that other techniques may struggle with. However, there has been limited efficient assessment of tree species classifications from complex forest scans using DL models.

Considering the above, in this paper it is proposed UAV4Tree a DL-based system for tree species classification. The aim of UAV4Tree is the automatic analysis and processing of huge amount of RGB optical images obtained by the UAV. This problem has not been investigated by the computer vision community properly yet due to weather conditions, camera Calibration (to obtain accurate RGB images, the camera mounted on the UAV must be calibrated correctly, any minor errors during calibration can result in inaccurate color reproduction and distortion in the images), image Resolution, flight path and altitude (this impacts the quality of the RGB images produced), and processing technique (obstructions or changes in lighting conditions during the flight can impact the accuracy of the data extracted). The proposed approach aims at reducing hand-operated analysis and at the same time using manual annotation as a form of continuous learning. The whole system needs manual tagging of large training data. Up until now, large datasets have been necessary to boost the performance of DL models and all manually verified data will be used as continuous learning and will be maintained as training datasets. Tree4UAV comprises three important phases: images classification by using Resnet (He et al., 2016), Densenet (Huang et al., 2017), InceptionV3 (Szegedy et al., 2016), Vision Transformer (ViT) (Dosovitskiy et al., 2020) on a dataset composed by tree images collected from iNaturalist[1]; testing with UAV images; performance improvement by using the Super Resolution technique to increase the resolution of the images obtained from UAVs trhrough SR-GAN (Ledig et al., 2017).

The paper is organized as follows. Section 2 provides a description of AI approaches that were adopted for tree images classification. Section 3 describes the proposed DL-based pipeline. In Sections 4 and 5, an evaluation of our approach is offered, as well as a detailed analysis of each component of our DL system. Finally, in Section 6, conclusions and discussion about future directions for this field of research are drawn.

2. RELATED WORKS

Tree species information is crucial in order to estimate stem size and amount of biomass accurately, and is consequently imperative in making informed decisions for effective management. Remote sensing can provide information on forest structure and health, which can be used to identify changes in forest composition and assist in the conservation and management of forest resources. This information can help to inform decisions related to forest management, such as the identification of areas of high conservation value, monitoring of forest growth and restoration programs, and assessment of the impact of natural and human-induced disturbances on forest ecosystems (Yu et al., 2017). In recent years, DNNs have been applied to recognize tree species since they automatically learn features and patterns that are difficult for humans to detect, such as subtle differences in leaf shape, texture, and color. DL models can also handle large and complex datasets, allowing for better representation of the diversity and variability of tree species across different regions and environments (Li et al., 2016).

He et al. (He et al., 2023) employed a deep learning method for classifying forest tree species. They compared the performance of DenseNet, EfficientNet, MobileNet, ResNet, and ShuffleNet. The experiments were assessed on remote sensing classification satellite imagery dataset, NWPU RESISC-45 was also trained and validated in the paper.

In (Anagnostis et al., 2021), the authors proposed a DL approach for the detection of leaves with disease. They developed an object detection system that identified anthracnose-infected leaves on walnut trees. Another study that adopted DL approaches for leaves classification is the work proposed by Minowa et al. (Minowa et al., 2022). The goal of this paper was to verify the accuracy of tree species identification by DL with leaf images of broadleaf and coniferous trees in outdoor photographs. They used the DL framework Caffe and AlexNet and GoogLeNet as DNNs.

In literature, the classification of tree species using 3D point clouds has drawn wide attention in surveys and forestry investigations. There are research papers that use 3d point cloud data for recognizing trees. Zou et al. (Zou et al., 2017) proposed voxel-based deep learning method to classify tree species in 3-D point clouds collected from complex forest scenes. Their method comprised three stages: 1) individual tree extraction

[1] https://www.inaturalist.org

https://doi.org/10.5194/isprs-annals-X-1-W1-2023-1089-2023 | © Author(s) 2023. CC BY 4.0 License.

considering the point clouds density; 2) low-level feature representation through voxel-based rasterization; and 3) tree species classification using deep learning model.

In (Martins et al., 2021), the authors proposed a multi-task CNN to map tree species in a highly diverse neighbourhood in Rio de Janeiro, Brazil. The network described architecture took an aerial image and had two outputs: a semantically segmented image and a distance map transform. The post-processing approach aimed to produce realistic tree species composition map by labelling only pixels of the target species with high class membership probabilities.

Considering the development of AI technologies, and in particular DL, in this paper it is proposed UAV4Tree a DL-based system for the automatic establishment of RGB optical images obtained by the UAV. The main contributions could be summarized as follows:

- the design of an efficient DL-based system to realize the automatic learning and analysis of RGB optical images obtained by the UAV. Moreover, it includes the performance comparison of state-of-art deep learning models, including ResNet, DenseNet, InceptionV3, and Vision Transformer, for tree classification.
- the automatic design method of DNN combined with SRGAN for improving the recognition accuracy.
- investigation of the use of augmented datasets for automated tree classification from UAV imagery.
- Identification of fine-tuning and dropout as essential techniques for improving the generalization ability of the models on augmented datasets.
- valuable insights into the development of accurate and reliable automated tree classification systems from UAV imagery, which has significant implications for sustainable forest management and conservation.

3. METHODOLOGY

In this section, UAV4Tree as well as the study area where the dataset for evaluation is collected, are introduced. In particular, it comprises three phases: *Data collection*, use of *Deep Learning Pipeline* and *Performance Evaluation*. The details are given in the following Subsections.

3.1 Study Area

The study area is located in the Marche Region, in the center of Italy, in particular, some species present in the province of Ascoli Piceno (42°51’17” N, 13° 34’31” E) were studied (in Figure 1 is reported the Geographical Localization of the area). As far as the climate of this area is concerned: in the lowland and hill areas a rather subcontinental climate reigns, with very sultry summers and cold winters. In the mountainous and high hill areas there are cool summers and cold winters with a large possibility of snow; winters are also harsh in the inland hilly areas where low temperatures can occur. Among the species present in the Marche Apennines we find *Acer opalus*, found in abundance in the mountainous area of the territory of Ascoli Piceno and in other areas of the region, including the Metauro valley in the coastal area of Monte Conero (Ancona), together with other species including Quercus pubescens. In the woods near Ascoli Piceno (Monte Ascensione), it is possible to find *Acer opalus* and mixed woods of downy oak (*Quercus pubescens*) and chestnut (*Castanea sativa*). In the mountainous area of the province of Ascoli Piceno (Monti Sibillini and Monti della Laga), it is possible to find vast woods of chestnut (*Castanea sativa*) associated with *Acer opalus* and other species. In these areas, chestnut woods are of particular importance from an economic point of view (the area occupied by chestnut groves corresponds to about 2300 ha, which is almost 100% of the chestnut area in the entire region). In the Ascoli Piceno area and in other areas of the region, the cultivation of the *Olea europaea* species is widespread, particularly in the “Tenera ascolana” variety, which is particularly important for the economy of these territories. Cultivation takes place in both hilly and coastal areas. The cultivation of *Olea europaea* in these territories has historical origins dating back to before the Romans. It was in fact introduced by the Phoenicians and Greeks. The images of the test dataset, taken by UAV, were collected in this geographical area.

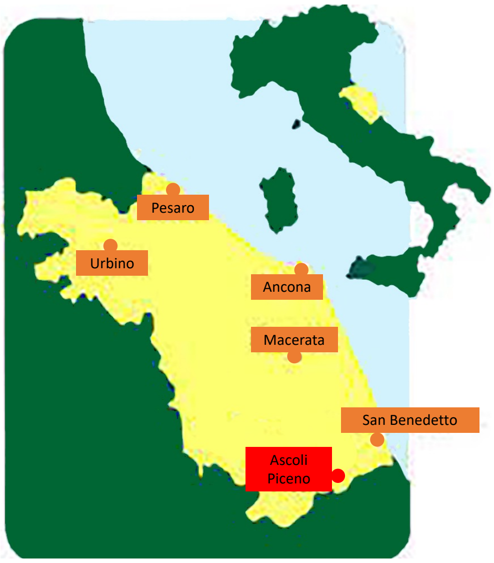

Figure 1. Geolocalization of the area of study, where the dataset have been collected.

3.2 Dataset

In the above-described geographical area, images from the test dataset were collected using UAVs in different weather conditions and lighting. As previously explained, a significantly different set of images from the training and validation dataset is used, obtained through the iNaturalist website, depicting the specimens of the four species in different phenological states, captured from the ground, framing the subject from an ”at-eye level” perspective. The images were divided as follows: 2690 images in the training dataset (Figure 2), 560 in the validation dataset, and 2650 in the testing dataset (Figure 3). Manual selection operations were carried out on the training dataset to eliminate potentially noisy images (presence of blurs, presence of plants belonging to other species). The class IDs present in the datasets are as follows: 0 for the *Acer opalus* class, 1 for the *Castanea sativa* class, 2 for the *Olea europaea* class, and 3 for the *Quercus pubescens* class. The RGB images, in JPG format with various resolutions, are processed and transformed using the `RandomResizedCrop()` function present in the PyTorch module, to obtain images of appropriate size for the different network models used. The test dataset contains images obtained through frame grabs from films made using UAVs in the geographic areas of interest for this study.

The images were partly captured using a DJI drone equipped with an FC7303 camera, with a focal length of 4.5 mm, and at an altitude ranging from 20 to 40 meters (above ground level). However, it should be noted that in most cases, the images were extracted from videos captured by external sources, and therefore, such information is not available.

The objective of our work was to compare the results of different deep learning models for image classification obtained from UAV (Unmanned Aerial Vehicle) captured images. These models were trained on a dataset of tree images captured at close range and already accurately annotated by participants in the iNaturalist project. These models were used to recognize images captured by UAV, which show trees from a different perspective and with less detail compared to the images used for training and validation of the models. The use of highly detailed images for the training/validation dataset allows capturing the details of different species (textures, shapes, leaf colors) with higher resolution than what can be achieved with a UAV device. The process of selecting training and validation images in our experiments was facilitated by the fact that the images were already annotated by iNaturalist. Therefore, we only needed to crop some images to remove any extraneous elements unrelated to the main subject. The images acquired with UAVs that we used in our tests sometimes exhibited inadequate quality and noise levels, which could compromise accurate annotation for use in the training dataset. This could potentially lead to a degradation in the precision of the models. To utilize these images correctly and avoid the risk of mislabeling, it would have been necessary to supplement them with ground truth information that we did not possess. Obtaining such information often requires additional time-consuming human intervention. At the same time, we chose to use UAV-acquired images, which potentially could have mislabeling issues due to the difficulty of precisely identifying the plant classes, only for the test dataset. This approach allowed us to achieve good results during prediction and subsequently manually analyze only the images that were not correctly recognized by the model. Including ”mislabelled” images in the training dataset could have significantly degraded the model’s performance, as we inferred from the results obtained by reversing the training/validation dataset with the test dataset. We adopted this approach specifically to study the behavior of different deep learning models when the images in the training/validation dataset and those in the test dataset, despite representing the same subject, vary in viewpoints and resolutions for various reasons. Certainly, in the future, this work can be expanded by incorporating other deep learning models or by performing object detection tasks.

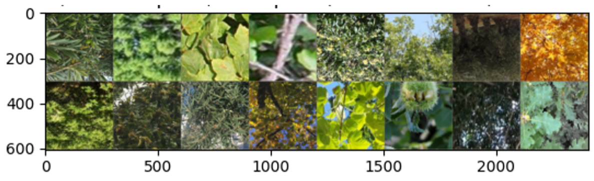

Figure 2. The image shows few images, provided as example, of the pictures composing the Training Dataset.

3.3 Deep Learning pipeline

The following tasks were executed: first, various DL models (namely Resnet, Densenet, InceptionV3, Vision Transformer) were tested on datasets; afterwards, fine-tuning (running adjustment of drop-out, learning rate, weight decay) where necessary to achieve better performance and avoid overfitting. The various models used are pre-trained (Imagenet1K), and the number of training epochs for each model is less than 300 epochs. Despite achieving high accuracy values during the train/validation phase, the Resnet and InceptionV3 networks achieved unsatisfactory accuracy results during testing. This is likely due to the limited number of images in the datasets, as well as the significant difference, especially for some classes, between the images in the train/validation dataset, which were taken from ground level, and the test dataset containing images taken via UAV. Fine-tuning, for some network models, has allowed for the avoidance of overfitting and improvement in performance, especially through the introduction of drop-out and modification of weight-decay and learning rate. Improved performance has been noticed when using the SGD optimizer, as opposed to Adam. The model with the best performance is Densenet161 (Huang et al., 2017) (Fig. 5) (Fig. 4) with LR=0.0004 and SGD optimizer. Super Resolution is then used to evaluate possible improvements in performance achieved in the previous tests, applying high resolution to the images obtained via UAV in the test dataset. The test dataset images are processed via an SR-GAN (Ledig et al., 2017). The images are obtained by applying Super Resolution (with a multiplier 4×) via GAN to the images in the original test dataset, previously subjected to manual crop operations. For some tests (InceptionV3), an improvement in performance in terms of accuracy of up to 10-15 points is detected, compared to the inference on the dataset that does not use images generated by SR-GAN. The results obtained are summarized in Table 1, where they are compared to the results of the previous test on the original dataset, and the values obtained via the ViT B16 model are shown in Table 3. By reversing the train/validation dataset with the test dataset, some tests are carried out using the aforementioned models, resulting in an accuracy lower than 0.60 in all cases. The train/validation dataset used so far was then modified using augmentation techniques to verify the increase in generalization capabilities of the various models. Following what is reported in (Dosovitskiy et al., 2020), the network Densenet161 is chosen to test the performance, having shown the best performance in the test on the original dataset, with the Vision Transformer B16 model, generating an augmented dataset from the original train dataset. The images in the train/validation dataset have been transformed with augmentation techniques like: flip (random flip, flip top/bottom, and flip left/right), skew, and ”random erasing” (Zhong et al., 2020). For the use of Vision Transformer B16, the images are resized via `RandomResizedCrop` to the size of 224×244, necessary to manage the typical 16×16 patches of this model. The test dataset is not modified. The results are presented in Table 4. The Vision Transformer B16 model shows a precision of 0.75 for the *Acer opalus* class, 0.87 for the *Castanea sativa* class, 0.89 for the *Olea europaea* class, and

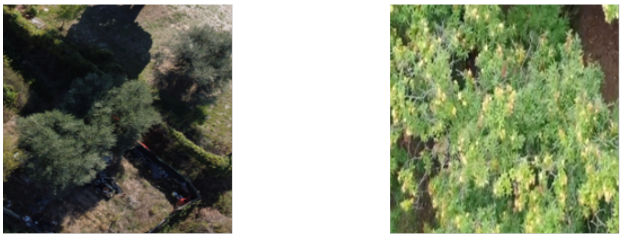

Figure 3. Few images taken from the Test Dataset, respectively Class ID 2 e Class ID 1

0.60 for the *Quercus pubescens* class, with an overall accuracy of 0.81 and F1-score of 0.81. Even for the augmented dataset, a test is carried out by running some models after reversing the train/validation dataset with the test dataset. The overall accuracy of the Vision Transformer B16 model is illustrated in Table 5, showing an accuracy of 0.48.

The purpose of the reversing process was to evaluate the ability of various neural network models to correctly classify close-range images of trees (from an ”eye level” viewpoint) based on images captured by a drone. The low level of accuracy achieved by the models is certainly attributed to the difficulty of manually annotating images captured by a drone, where different tree species are often present in the same photo. In such cases, the photo editing work required to avoid mislabeling issues becomes complex and time-consuming. However, in UAV images where only one tree species is present, the model's performance is satisfactory.

4. RESULTS

In this section, we present the results of our experimental study, which was conducted following the previously described methods. The purpose of this study was to automatically classify tree images obtained from UAV. As already stated, the classification of these images is critical for various applications, including forest management and conservation. To achieve this goal, we designed and implemented deep learning algorithms that can accurately classify the images based on various features, such as leaf size, color, and shape. Specifically, we evaluated the performance of state-of-art deep learning models, including ResNet, DenseNet, InceptionV3, and Vision Transformer. These models were trained using a dataset of labeled images, and their performance was evaluated using various metrics, including accuracy, precision, recall, and F1-score. These experiments were conducted to compare the performance of each algorithm and identify the best-performing model for tree image classification.

Firstly, we introduce the table regarding the starting dataset and on the test set, through SR-GAN (Table 1).

For the classes examined in this experiment, we report the results of Precision, Recall and F1-Score obtained with the Densenet161 (Table 2).

Besides, for the sake of completeness, the confusion matrix obtained with the Densenet161 is reported in Figure 4, while the Accuracy and Loss for both training and validation are depicted in Figure 5.

As stated in the methodology section, the Vision Transformer B16 method has been used on the test set SR-GAN; here below the results achieved (Table 3).

In order to evaluate the performances on the dataset after the augmentation, we report the results in the following table (Table 4).

We finally report the results on the dataset with augmentation, but exchanging the training set with the test set.

The results of our experiments provide valuable insights into the use of deep learning for automated tree classification from UAV imagery.

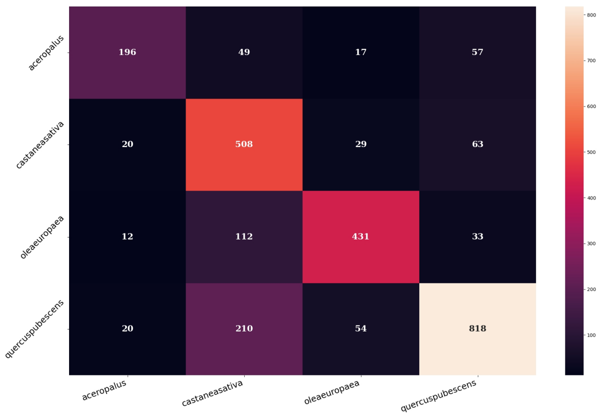

Figure 4. Confusion Matrix

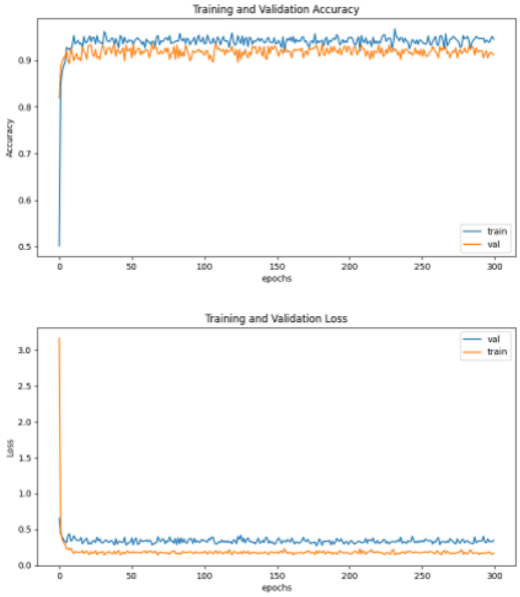

Figure 5. Accuracy and Loss curves for both the training and validation set

5. DISCUSSIONS

Starting from the initial test (that involved the execution of several models on the starting dataset characterized by a few hundred training images for each class), the Inception V3 network shows (Table 1) superior performance compared to other models, as is typical of this network in situations where datasets have a low number of samples. It is worth noting how in the test using SR-GAN for the Inception V3 model, an improvement in performance (Table 1) in terms of accuracy was detected, with an increase of up to 10-15 points compared to what was obtained from inference on the dataset that did not use images generated by SR-GAN. Despite the worse performance of the Densenet161 model when using an augmented dataset (Table 1) compared to the test previously performed on the initial dataset, an increase in performance of the ViT-B16 model is noted using the augmented train/validation dataset, confirming the characteristic of this model to improve its performance when the dataset has larger dimensions and the model improves its generalization ability. The ”augmented” dataset consists of 63,000 images in the training dataset, 3,000 images in the validation dataset, and 2,640 images in the test dataset.

Table 1. Results on the original dataset

Model	Accuracy (validation)	Accuracy (original test dataset)	F1-score (original test dataset)	Accuracy (SR-GAN test dataset)	F1-score (SR-GAN test dataset)
Resnet152	0.92	0.53	0.54	Not Verified	Not verified
Densenet161	0.93	0.72	0.74	0.69	0.72
VGG19	0.91	0.57	0.59	Not verified	Not Verified
InceptionV3	0.98	0.37	0.34	0.54	0.55
Vision Transformer B16	0.96	0.68	0.68	0.75	0.76

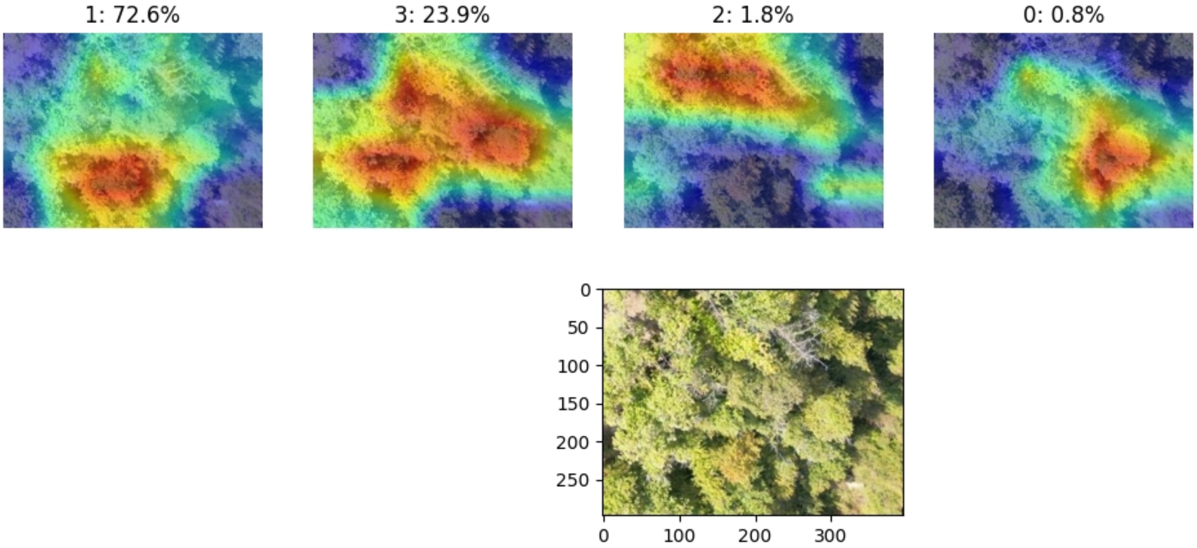

Figure 6. Grad-Cam analysis, performed on the ground truth Class 1 (*Castanea sativa*)

Table 2. Results obtained with the Densenet161

Class-ID	Precision	Recall	F1-score
0	0.75	0.79	0.77
1	0.76	0.74	0.75
2	0.78	0.70	0.74
3	0.51	0.62	0.56

Table 3. Results of the Vision Transformer B16 model on the dataset test SR-GAN.

Class-ID	Precision	Recall	F1-score
0	0.98	0.42	0.59
1	0.69	0.61	0.65
2	1.00	0.71	0.83
3	0.69	1.00	0.82

Table 4. Results on test dataset with augmentation.

Model	Epoch	Accuracy (validation)	Accuracy (augmented)	F1-score (augmented)
Densenet161	50	0.90	0.67	0.69
Vision Transformer B16	30	0.87	0.81	0.81

Table 5. Results on the dataset by exchanging training and test set

Class-ID	Precision	Recall	F1-score
0	0.75	0.27	0.39
1	0.31	0.16	0.21
2	0.90	0.43	0.58
3	0.41	0.86	0.55

Analyzing the metrics reported in the various tests (Tables 2 and 3, lower results can be noticed regarding class 3 *Quercus pubescens*, probably due to the presence of ”noise” in the test dataset, caused by the presence of plants of other species in the images. It should also be noted that, as reported by (Zhang et al., 2021), variations during different periods of the year, such as the shape, color, and consistency of the tree crowns in the dataset, can reduce classification accuracy. By reducing the classification to only classes with ID 0, 1, and 2, an overall accuracy of 0.89 is obtained with the Vision Transformer L16 model and an overall accuracy of 0.91 with the Vision Transformer DeiT model. From the tests performed, using the Vision Transformer B16 model, the importance of using an appropriate drop-out value to avoid overfitting has emerged. The model, in fact, with a 0.3 drop-out, reaches an average accuracy of 0.81 (using the dataset with 4 classes), while the same model, without the use of drop-out, achieves an accuracy of about 0.57. Analyzing the results of the tests performed by reversing the train/validation dataset with the test dataset, better performance is highlighted regarding classes 0 *Acer opalus* and 2 *Olea europaea*. This is probably due to the fact that in the images present in the dataset used for training (i.e., the test dataset used in the initial task), there are images of forests with species of classes 1 and 3 in which trees of other species are also present, while in olive cultivations and maple forests, it is easier to identify and isolate only the plants of the species under study in the images. This difficulty is further compounded by high variability in the phenological state in the images of the train and test dataset, which is more marked for some of the classes under study.

With the aim of improving the results comprehension and inter-

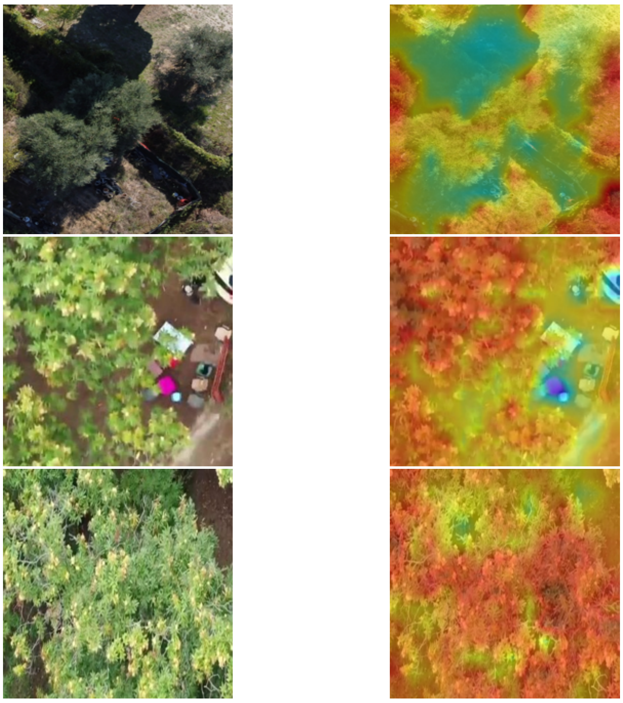

Figure 7. Original images and corresponding Rollout Attention for the classes 1, 2 and 3

pretability, we performed few test with visualization methods, helping in the visualization of the attention models and, consequently, the features learned by the models. Class Activation Maps (CAM) (Selvaraju et al., 2016) are visualization methods used to explain various deep learning models, identifying which features of the image motivated the selection of our classification model. Grad-Cam (Figure 6) is used for CNN models (Selvaraju et al., 2017) because Gradient-Weighted Class Activation Mapping is highly class-discriminative and therefore targeted at identifying which features and regions of the image exclusively influence the choice of the indicated class (Class ID = 1 in the case of the Figure 6 image). For the Vision Transformer model, Visual Attention Maps are used: matrices representing the importance of different parts of an input image with respect to different parts of the model's learned representations. In particular, the Rollout Attention technique (Figure 7) is used to visualize the regions of the image that have been most relevant for the model's prediction and how these regions have been weighted during the attention process.

6. CONCLUSION AND FUTURE WORKS

Automated tree classification from UAV images is a critical task that has numerous applications in forest management and conservation. Accurate classification of tree species and individual trees can provide valuable information for monitoring forest health, predicting forest growth and yield, and identifying areas that require intervention or protection. With the growing demand for sustainable forest management practices, the need for accurate and efficient automated tree classification systems has become increasingly important. UAVs offer a cost-effective and efficient way to collect high-resolution imagery of forests, making it possible to accurately classify trees at a large scale. Automated tree classification from UAV images can also help reduce the need for manual labour and costly field surveys, making it an attractive option for forestry management and research. Overall, the development of accurate and reliable automated tree classification systems from UAV imagery has significant implications for the sustainable management and conservation of forests around the world. In this paper, we investigated the use of augmented datasets and we have applied state-of-art deep learning models for the classification of tree images obtained from UAV. Our results show that the use of an augmented dataset can significantly improve the average accuracy by approximately 10 points compared to the use of a non-augmented dataset. We found that the Densenet model performed reasonably well on the non-augmented dataset, but not on the augmented dataset, where it showed lower accuracy compared to the Vision Transformer (VIT) model. Furthermore, we found that fine-tuning and specifically the introduction of dropout were essential in the tests on the augmented dataset to improve the generalization ability and avoid overfitting. Finally, we explored the use of SR-GAN in the original dataset and found that it allowed us to increase the performance of some models. Our study provides valuable insights into the use of augmented datasets and deep learning models for automated tree classification from UAV imagery. Our findings can inform future research in this area and contribute to the development of more accurate and reliable automated tree classification systems.

REFERENCES

Anagnostis, A., Tagarakis, A. C., Asiminari, G., Papageorgiou, E., Kateris, D., Moshou, D., Bochtis, D., 2021. A deep learning approach for anthracnose infected trees classification in walnut orchards. *Computers and Electronics in Agriculture*, 182, 105998.

Andreas, M., Prausová, R., Brestovanská, T., Hostinská, L., Kalábová, M., Bogusch, P., Halda, J. P., Rada, P., Štěrba, L., Čížek, M. et al., 2023. Tree species-rich open oak woodlands within scattered urban landscapes promote biodiversity. *Urban Forestry & Urban Greening*, 83, 127914.

Dian, Y., Pang, Y., Dong, Y., Li, Z., 2016. Urban tree species mapping using airborne LiDAR and hyperspectral data. *Journal of the Indian Society of Remote Sensing*, 44, 595–603.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S. et al., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

He, T., Zhou, H., Xu, C., Hu, J., Xue, X., Xu, L., Lou, X., Zeng, K., Wang, Q., 2023. Deep Learning in Forest Tree Species Classification Using Sentinel-2 on Google Earth Engine: A Case Study of Qingyuan County. *Sustainability*, 15(3), 2741.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K. Q., 2017. Densely connected convolutional networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4700–4708.

Jutras-Perreault, M.-C., Gobakken, T., Næsset, E., Ørka, H. O., 2023. Comparison of Different Remotely Sensed Data Sources for Detection of Presence of Standing Dead Trees Using a Tree-Based Approach. *Remote Sensing*, 15(9), 2223.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z. et al., 2017. Photo-realistic single image super-resolution using a generative adversarial network. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4681–4690.

Li, W., Fu, H., Yu, L., Cracknell, A., 2016. Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. *Remote sensing*, 9(1), 22.

Liu, L., Coops, N. C., Aven, N. W., Pang, Y., 2017. Mapping urban tree species using integrated airborne hyperspectral and LiDAR remote sensing data. *Remote Sensing of Environment*, 200, 170–182.

Lovell, J., Jupp, D. L., Culvenor, D., Coops, N., 2003. Using airborne and ground-based ranging lidar to measure canopy structure in Australian forests. *Canadian Journal of Remote Sensing*, 29(5), 607–622.

Martins, G. B., La Rosa, L. E. C., Happ, P. N., Coelho Filho, L. C. T., Santos, C. J. F., Feitosa, R. Q., Ferreira, M. P., 2021. Deep learning-based tree species mapping in a highly diverse tropical urban setting. *Urban Forestry & Urban Greening*, 64, 127241.

Maschler, J., Atzberger, C., Immitzer, M., 2018. Individual tree crown segmentation and classification of 13 tree species using airborne hyperspectral data. *Remote Sensing*, 10(8), 1218.

Minowa, Y., Kubota, Y., Nakatsukasa, S., 2022. Verification of a deep learning-based tree species identification model using images of broadleaf and coniferous tree leaves. *Forests*, 13(6), 943.

Nezami, S., Khoramshahi, E., Nevalainen, O., Pölönen, I., Honkavaara, E., 2020. Tree species classification of drone hyperspectral and RGB imagery with deep learning convolutional neural networks. *Remote Sensing*, 12(7), 1070.

Palmer, L., 2021. How trees and forests reduce risks from climate change. *Nature Climate Change*, 11(5), 374–377.

Pereira, J. A., Varela, D., Scarpa, L. J., Frutos, A. E., Fracassi, N. G., Lartigau, B. V., Piña, C. I., 2023. Unmanned aerial vehicle surveys reveal unexpectedly high density of a threatened deer in a plantation forestry landscape. *Oryx*, 57(1), 89–97.

Quan, Y., Li, M., Hao, Y., Liu, J., Wang, B., 2023. Tree species classification in a typical natural secondary forest using UAV-borne LiDAR and hyperspectral data. *GIScience & Remote Sensing*, 60(1), 2171706.

Salleh, S. A., Abd. Latif, Z., Pardi, F., Mushtaha, E., Ahmad, Y., 2023. Conceptualising the citizen-driven urban forest framework to improve local climate condition: Geospatial data fusion and numerical simulation. *Concepts and Applications of Remote Sensing in Forestry*, Springer, 337–353.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE international conference on computer vision*, 618–626.

Selvaraju, R. R., Das, A., Vedantam, R., Cogswell, M., Parikh, D., Batra, D., 2016. Grad-CAM: Why did you say that? *arXiv preprint arXiv:1611.07450*.

Sulla-Menashe, D., Friedl, M. A., 2018. User guide to collection 6 MODIS land cover (MCD12Q1 and MCD12C1) product. *Usgs: Reston, Va, Usa*, 1, 18.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2818–2826.

Wang, P., Fan, E., Wang, P., 2021. Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. *Pattern Recognition Letters*, 141, 61–67.

White, J. C., Coops, N. C., Wulder, M. A., Vastaranta, M., Hilker, T., Tompalski, P., 2016. Remote sensing technologies for enhancing forest inventories: A review. *Canadian Journal of Remote Sensing*, 42(5), 619–641.

Xi, Z., Hopkinson, C., Rood, S. B., Peddle, D. R., 2020. See the forest and the trees: Effective machine and deep learning algorithms for wood filtering and tree species classification from terrestrial laser scanning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 168, 1–16.

Yu, X., Hyyppä, J., Litkey, P., Kaartinen, H., Vastaranta, M., Holopainen, M., 2017. Single-sensor solution to tree species classification using multispectral airborne laser scanning. *Remote Sensing*, 9(2), 108.

Zhang, C., Xia, K., Feng, H., Yang, Y., Du, X., 2021. Tree species classification using deep learning and RGB optical images obtained by an unmanned aerial vehicle. *Journal of Forestry Research*, 32(5), 1879–1888.

Zhong, Z., Zheng, L., Kang, G., Li, S., Yang, Y., 2020. Random erasing data augmentation. *Proceedings of the AAAI conference on artificial intelligence*, 34number 07, 13001–13008.

Zou, X., Cheng, M., Wang, C., Xia, Y., Li, J., 2017. Tree classification in complex forest point clouds based on deep learning. *IEEE Geoscience and Remote Sensing Letters*, 14(12), 2360–2364.