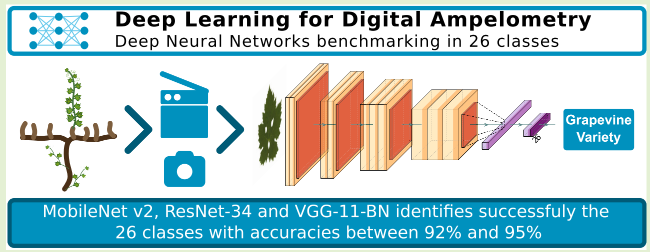
Toward Grapevine Digital Ampelometry Through Vision Deep Learning Models

Sandro Costa Magalhães, Luís Castro, Leandro Rodrigues, Tiago Cerveira Padilha, Frederico de Carvalho, Filipe Neves dos Santos, Tatiana Pinho, Germano Moreira, Jorge Cunha, Mário Cunha, Paulo Silva, and António Paulo Moreira

***Abstract*—Several thousand grapevine varieties exist, with even more naming identifiers. Adequate specialized labor is not available for proper classification or identification of grapevines, making the value of commercial vines uncertain. Traditional methods, such as genetic analysis or ampelometry, are time-consuming, expensive, and often require expert skills that are even rarer. New vision-based systems benefit from advanced and innovative technology and can be used by nonexperts in ampelometry. To this end, deep learning (DL) and machine learning (ML) approaches have been successfully applied for classification purposes. This work extends the state of the art by applying digital ampelometry techniques to larger grapevine varieties. We benchmarked MobileNet v2, ResNet-34, and VGG-11-BN DL classifiers to assess their ability for digital ampelography. In our experiment, all the models could identify the vines' varieties through the leaf with a weighted $F1$ score higher than 92%.**



***Index Terms*— Artificial neural networks (ANN), computer vision, image processing, precision agriculture, vine species identification.**

Manuscript received 17 January 2023; revised 21 March 2023; accepted 21 March 2023. Date of publication 29 March 2023; date of current version 1 May 2023. This work was supported in part by the European Regional Development Fund (ERDF), through the Operational Program for Competitiveness and Internationalization—COMPETE 2020 Program under the Portugal 2020 Partnership Agreement, within Project INCAFO, with Reference POCI-01-0247-FEDER-068922. The work of Sandro Costa Magalhães, Leandro Rodrigues, and Germano Moreira were supported by the Portuguese Foundation for Science and Technology (FCT) under Scholarship Agreements SFRH/BD/147117/2019, 9684/BI-M-ED_B2/2022, and 9688/BI-M-ED_B2/2022, respectively. The associate editor coordinating the review of this article and approving it for publication was Dr. Brajesh Kumar Kaushik. *(Corresponding author: Sandro Costa Magalhães.)*

Sandro Costa Magalhães and António Paulo Moreira are with the INESC TEC—Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência, 4200-465 Porto, Portugal, and also with the Faculty of Engineering, University of Porto, 4200-465 Porto, Portugal (e-mail: sandro.a.magalhaes@inesctec.pt; amoreira@fe.up.pt).

Luís Castro, Filipe Neves dos Santos, and Tatiana Pinho are with the INESC TEC—Instituto de Engenharia de Sistenas, Tecnologia e Ciência, 4200-465 Porto, Portugal (e-mail: luis.r.castro@inesctec.pt; filipe.n.santos@inesctec.pt; tatiana.m.pinho@inesctec.pt).

Leandro Rodrigues, Germano Moreira, and Mário Cunha are with the INESC TEC—Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência, 4200-465 Porto, Portugal, and also with the Faculty of Sciences, University of Porto, 4169-007 Porto, Portugal (e-mail: leandro.a.rodrigues@inesctec.pt; germano.f.moreira@inesctec.pt; mccunha@fc.up.pt).

This article has supplementary downloadable material available at https://doi.org/10.1109/JSEN.2023.3261544, provided by the authors.

Please see the Acknowledgment section of this article for the author affiliations.

Digital Object Identifier 10.1109/JSEN.2023.3261544

I. Introduction

THE choice of adequate grape varieties is critical for successful wine yield and quality in any viticultural system. Therefore, there is a strong demand for accurate and operational identification of grape varieties to improve the efficiency of vineyard decisions, wine quality/ authenticity, and enforcing appellation laws, such as certification of the protected designation of origin (PDO) [1]. Also, the mitigation of climate change scenarios points to the need to reconvert large areas of vineyards for the installation of better-adapted varieties, which must be correctly identified [2].

The widespread of grapevine [3] and the insufficiencies of traditional identification of varieties based on visual phenotyping often lead to multiple local synonyms [4], [5].

Experts estimate the existence between 5000 and 8000 grapevine's varieties around the world, being used between 14 000 and 24 000 different names for identifying them [6]. The *Organization Internationale de la Vigne et du Vin* (OIV) identifies around 4000 varieties [7], which contrasts with the 12 000 varieties names presented by the Vitis International Variety Catalog (VIVC) [8]. The successful identification of the different varieties of fruit can be made using a genetics analysis [9], [10]. However, these methods are expensive, time-consuming, and challenging for the viticulturist to work in the field.

Ampelography is the science of identifying and classifying grape varieties through the unique characteristics linked to morphology detected in the different phases of its phenological development [11]. Inside ampelography, ampelometry ("vine" and "technique of measuring") examines the landmark features of the grapevine's palate leaves (e.g., lobes, sinuses, and vines angles) to identify varieties [11], [12]. It is based on linear and angular measurements, in general, and is used to characterize with precision the type of medium leaf, the shape of berry, and the shape of seed using digital programs, such as SIAMS Mesoplant,[1] imageJ [13], or others. This visual phenotyping-based technique is time-consuming, costly, and demands high-level experts often confined to a wine region. In general, depending on the practice of the Ampelographer, he builds in the memory occurrence characteristic that characterizes the variety, such as, for example, the Cabernet Sauvignon: shown obligatory five lobes in maximum seven, the shape of the leaf is round, and the shape of the base of upper lateral sinuses sinus is U-shaped. So, the accuracy practiced by the human ampelographer varies according to his experience and subjectiveness. Also, many of the studied characteristics are not always uniform due to edaphoclimatic differences across regions or are not simultaneously available at a single phenological stage (i.e., bloom, fruit set, and cluster closure). The traditional ampelometry based on grapevine phenotyping beyond low transferability is not appropriate for the high throughput required to identify large quantities and diversity of grapes varieties. Yet, high-performance and automated identification methods for wide grape varieties are still unavailable. Simpler and low-cost setups are required to allow easy integration into a portable system suitable for field use (e.g., smartphone), with real-time processing capability to assist producers/technicians without deep knowledge in ampelometry. Image processing techniques and machine vision [e.g., machine learning (ML) and deep learning (DL)] have already been widely applied for plant species characterization [14] and also for the within-species (varieties) classification [15].

The automatic classification of 16 grapevine varieties was presented in [16] using two artificial neural network (ANN) based on leaves: one based on leaves' morpho-colourimetric features and the other based on leaf spectroscopy parameters. The first model rendered an accuracy of 94 %, while the other showed an accuracy of 92 %. Similar to this author, others have presented systems based on spectroscopy parameters: Gutiérrez et al. [17] presented an on-the-go hyperspectral imaging and ML algorithms system for the classification of 30 varieties. The authors studied ML solutions based on support vector machine (SVM) and ANN multilayer perceptrons, reaching individual classification accuracies for each variety between 87% and 99%. Fernandes et al. [18] built SVM and convolutional neural network (CNN) to identify one specific variety (Touriga Nacional or Touriga Franca) from 63 others, based on leaves spectra. 81.90 % of the Touriga Nacional spectra was correctly classified by the SVM and 93.82 % for the Touriga Franca spectra by the CNN. Even though the mentioned systems performed well, they imply the acquisition of expensive equipment, which is not within reach of the common user searching for a simple and affordable system.

Nasiri et al. [19] addressed the topic of grapevine varieties classification using a DL model trained with RGB images of the leaves. The developed VGG16-based model, a CNN, identified six Iranian varieties with an average classification accuracy of over 99 %. Other CNNs were applied for the same purpose, and Adão et al. [20] presented three CNNs (VGG-16, ResNet, and Xception) for the detection of six varieties, with Xception being the one with the best performance. Also, AlexNet [21], [22], GoogLeNet [22], DenseNet [22], and MobileNet v2 [23] were applied for grapevine variety classification taking into account leaf features extraction on RGB images.

This work contributes to the state of the art about digital ampelometry by introducing a grapevines' classification system by the following:

1) the creation of a dataset of a larger number of grapevines' varieties (26 classes) than the previous works on the topic;
2) the validation of the grapevines' varieties identification using small size and mobile compatible ANN, i.e., MobileNet v2;
3) benchmarks small size CNN with bigger CNNs ResNet-34 and VGG-11-BN.

The following sections of this article present the development strategies and test results. Section II explains the required materials for the current experiment and the adopted procedures and assumptions. Section III presents the results of the different essays along the various experiments and discusses and compares them with the state of the art. Finally, Section IV concludes the overall experiment and retains the main conclusions and future work.

II. Materials and Methods

A. Data Acquisition

The use of supervised DL models for classification requires the creation of a dataset with a large number of images and with relevant characteristics. The collected dataset comprises scanned and photographed leaves of the selected grapevine varieties. The leaves were harvested from Dois Portos Station Hub—National Wine Station from Instituto Nacional de Investigação Agrária e Veterinária, I.P. (INIAV)[2] at Torres Vedras, Portugal. This station hub is a scientific field for wine study with hundreds of grapevine varieties, the largest Portuguese ampelographic collection established in 1988.

The varieties included in the dataset were selected to contain the most representative grapevine varieties in Portugal and two of the world's top six varieties (Cabernet Sauvignon and Syrah). Fig. 1 contains sample images of all vine varieties in the dataset and the number of leaf samples for each variety. As observed, some varieties are slightly unbalanced,

[1]See SIAMS, 2021. SIAMS Mesoplant Analyzer. URL: https://siams.com/siamsmesoplant/. Last accessed on 11 January, 2023.

[2]See INIAV, I.P., 2022, INIAV, URL: http://www.iniav.pt. Last accessed on 18 August, 2022.

(a) 63 images Alfrocheiro (b) 30 images Alicante Bouschet (c) 30 images Alvarinho (d) 61 images Aragonez (e) 30 images Arinto Açores (f) 62 images Baga (g) 32 images Cabernet Sauvignon (h) 30 images Castelão (i) 33 images Encruzado

(j) 33 images Loureiro (k) 30 images Moreto (l) 66 images Mortágua (m) 30 image Negra Mole (n) 33 images Rufete (o) 30 images Sercial (p) 33 images Syrah (q) 30 images Terrantez Pico (r) 32 images Tinta Antiga Curral

(s) 30 images Tinta Barroca (t) 32 images Tinta Negra (u) 30 images Touriga Franca (v) 30 images Touriga Nacional (w) 70 images Trincadeira (x) 30 images Verdelho (y) 51 images Vinhão (z) 29 images Viosinho

Fig. 1. Samples of leaves images of the vines varieties in the dataset.

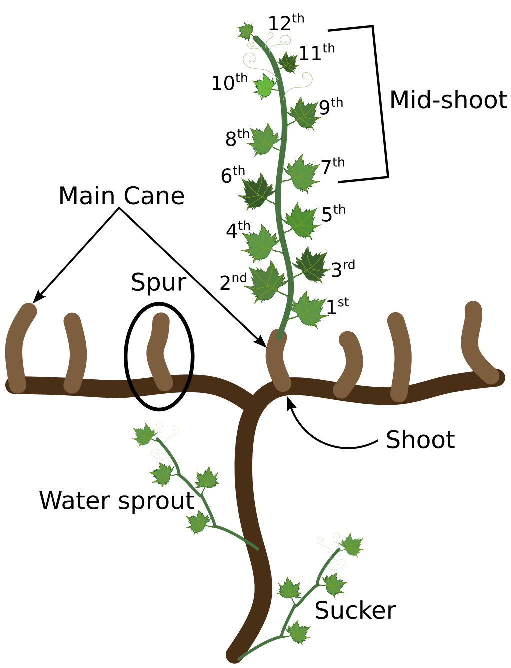
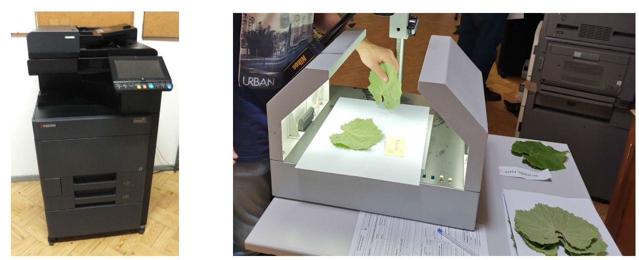
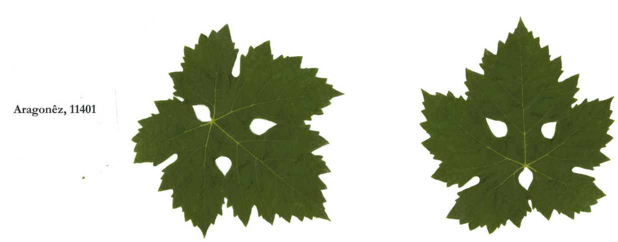


Fig. 2. Description of a vine with the indication of the main canes and the harvested leaves at the mid-shoot.



(a) (b)

Fig. 3. Images acquisition for leaves' dataset creation. (a) Scanner used to acquire leaves images. (b) Support for leafs image acquisition in a controlled environment using a digital camera.



(a) (b)

Fig. 4. Sample of (a) image acquisition and (b) preprocessing of an Aragonês leaf for the dataset.

but increasing the number of samples for each class was more important.

As stated in [19], [24], [25], and [26], the sampled leaves were harvested from the best representative grapevines in the vineyard from the main shoot at the nodes 7th, 9th, and 11th in the mid-shoot (Fig. 2). According to ampelographic experts, the leaves harvested from these nodes are the most representative of the variety's phenotype features [12]. The leaves were harvested on the morning of 23 June 2021, and the image acquisition was performed on the same day. The images were scanned and randomly photographed. The primary source of leaves data for the dataset was the scanner. Using the scanner, we could get uniform images of the leaves with a white background and its flattening. This kind of data made the model easier to train and supports highlighting the main morphologic features for identifying grapevine varieties.

All the harvested leaves were scanned in a Kyocera TASKalfa 2552ci [Fig. 3(a)], using a white background [Fig. 4(a)], helping the segmentation to generate a new image without background [Fig. 4(b)]. The result of the leaf scanning is a high-resolution colored image with 7008 × 7008 px and white background. The background also has the correct variety of the leaf written [Fig. 4(a)].

Some scanned leaves were also randomly sampled for photography shooting [Fig. 3(b)], originating a secondary source of image data for training. The setup for photographing the leaves is composed of a camera support frame with individual and constant illumination and a Nikon D5200 above the scenario [Fig. 3(b)]. This setup maintains a constant distance and position between the images.

B. Dataset Generation and Preprocessing

The input for our DL model was a low-resolution gray-scaled image, 416 × 416 px. We consider that the leaves' color is not significant and can even derange the classification process. So, we intend a square leaf image in gray scale with a clean white background for the network input.

The scans can have more than one leaf and a tag of the grapevine variety. Therefore, the image was segmented to obtain an image for each processed leaf. The Otsu threshold method was used to fully remove the image background, remove elements, such as shadows or variety identification, and separate the different leaves in the scan. We only selected the closed contours from the Otsu threshold with an area bigger than 6 000 000 px^2. Given the high-image resolution of the original images, the selected contours were rescaled and tight-cropped to configure a low-resolution square image with 480 × 480 px without distortion [Fig. 4(b)]. All the images were also rotated to set all the leaves with the same orientation, and the background was replaced by white pixels.

After applying this preprocessing strategy, the dataset ended up with 976 individual leaves belonging to 26 classes. Fig. 1 illustrates the different grapevine varieties in the dataset and the number of leaves for each grapevine variety.

We divided the dataset into train, validation, and test sets using the stratified split method. This method conserves the proportion of each class in all sets, improving dataset balance. A random seed was defined to guarantee consistency in the dataset division, resulting in a more reliable comparison between results. The train set contains 70 % of the images, and the validation and test sets contain 15 % each. This results in 682 samples, 147 samples, and 147 samples for each set. The train set was used to transfer learning the DL model and the validation set to track the evolution of the training process, assessing any overfitting case. The test set is an independent dataset to assess the final performance of the network in the evaluation metrics (Section II-C2).

To improve the ability of the models to generalize and avoid overfitting, we augmented all the images. Data augmentation implies applying small transformations to the original images, such as rotation, scaling, color transformations, or noise additions. To support the data augmentation process, we used the Albumentations library [27], allowing us to change the image information and create unique data easily. We designed a unique augmentation pipeline (see supplementary material) that creates randomly different images. We aimed to create new images without deforming the images too much, keeping the leaves' shape and veins valid, so we avoided affine and blurring transformations and used the last ones with discretion. For each image in the dataset, we generated 40 new images. In addition, we aggregated the generated images with the original image in the same set to reduce the correlation between data between the different sets. For the test set, a less aggressive augmentation pipeline was used. This pipeline aims to get more reliable data similar to the original data. Here, we avoid some transformations, such as color transformations. Therefore, after augmentation, our dataset increased to 40 428 images (28 250 images in the train set, 6069 images in the validation set, and 6109 images in the test set).

TABLE I
NUMBER OF TRAINABLE PARAMETERS

	Number of trainable parameters
MobileNet v2	2 256 602
ResNet 34	21 291 738
VGG-11-BN	128 877 210

C. Data Training and Evaluation

1) Classifiers: We assess the availability of three state-of-the-art CNN feature extractors (MobileNet v2, ResNet-34, and VGG-11-BN) for identifying the studied grape varieties. For all the studied DL models, we used pretrained models in the ImageNet dataset and applied transfer learning in the Pytorch framework. At the end of each feature extractor, we attached a classification head composed of two dense layers of 216 and 26 (number of classes) units. We should be training the different models from scratch, because our task is different from the task of the ImageNet dataset. However, according to [28] and [29], the capability of neural networks to extract features from images is transferable between tasks, and it is better to transfer these networks' weights than to train from scratch. Anyway, some essays in training from scratch were made to corroborate these conclusions.

Visual Geometry Group (VGG) has been widely used for classification problems after winning the first place in the ImageNet contest in 2014. The first study about VGG studied different depth architectures between 11 and 19 layers [30]. The VGG-11 has 11 depth layers composed of eight convolution layers and three fully connected layers. This architecture is less depth in the family but has more trainable parameters than the other DL models in this study (Table I), consuming much of the available memory. Simonyan and Zisserman [30] concluded that the use of normalization layers is not relevant for the accuracy of the trained model to ILSVRC-2012 but could help to lead to increased memory and computational time. In spite of not being state of the art for classification problems, VGGs is still a reference on these problems and should be considered in this work. The use of a less depth architecture supports to lead with overfitting issues due to a small size dataset.

Attempting to improve VGG performance and reduce the training complexity, He et al. [31] studied residual networks (ResNets). This architecture is inspired by plain architectures with shortcuts for residual learning. After VGG, ResNet won

TABLE II
HYPERPARAMETERS FOR TRANSFER LEARNING

	MobileNet v2	ResNet-34	VGG-11-BN
Batch size	32	20	11
Epochs	50	30	50
Learning rate	10^{-4}	10^{-4}	10^{-4}
Optimiser	Adam	Adam	Adam
Loss function	Cross-entropy	Cross-entropy	Cross-entropy

TABLE III
SUMMARY OF THE EVALUATION METRICS FOR THE DIFFERENT DL MODELS

Model	From Scratch			Transfer Learning		
	Prec.	Recall	F1	Prec.	Recall	F1
MobileNet v2	70 %	71 %	70.46 %	95 %	95 %	94.75 %
ResNet-34	83 %	83 %	82.69 %	93 %	93 %	92.98 %
VGG-11-BN	86 %	85 %	84.32 %	93 %	93 %	92.53 %

the ImageNet contest in 2015. This type of architecture reduced the number of trainable parameters (Table I) and the training complexity. Besides, it also improved the model's accuracy and allowed more depth architectures (between 18 and 152 layers). The architectures can vary in the multipliers block usage (ResNet-18 to ResNet-34) or the internal structure block (ResNet-34 to ResNet-50). He et al. [31] concluded that deeper architectures might lead to accuracy degradation and that is not related to overfitting. So, we chose a less deep architecture, ResNet-34.

MobileNet v2 [32] is the most recent architecture considered for this study. The main aim of this architecture is to provide a capable architecture that can execute on mobile devices. In this architecture, Sandler et al. [32] replaced the full convolutional operators with a factorized version that splits the convolution in two layers: depthwise convolution and pointwise convolution. The first convolution is responsible for lightweight filtering, while the pointwise convolution creates new features. This approach reduced severely the training complexity as well as the number of training parameters (Table I). Despite this reduction in the number of parameters, the literature has been proofing that MobileNet V2 is equally capable. A comparative analysis on different DL architectures for identifying COVID-19 in patient lung images proves that MobileNet V2 and ResNet are equally capable, and VGG has a severe accuracy dropping [33]. This comparative work is relevant to us because of its wide analysis in a difficult task similar to ours, where the network has to identify details to distinguish between the different classes.

Given the features of each ANN, Table II states the hyperparameters for transfer learning the models from the ImageNet dataset. The batch size was set to optimize the number of samples being simultaneously trained and that fit in the graphical processing unit (GPU) NVIDIA RTX2060 Super memory. The number of epochs was adjusted to ensure that all the ANN converged.

Since the dataset is not perfectly balanced, the $F1$ score was used to assess the models' performance instead of the accuracy. So, during the training routine, the $F1$ score is computed over the validation and training sets at the end of each epoch. To avoid overfitting, the early stopping technique was applied. So, at the end of the training process, the model stored for inference was the one that optimizes the $F1$ score in the validation set.

2) Models Assessment: To evaluate the models, we used state-of-the-art evaluation metrics: recall (1), which is the model's ability to detect all relevant objects, precision (2), the model's ability to identify only relevant objects, and $F1$ score (3), the first-harmonic mean between recall and precision. As can be concluded, all these metrics depend essentially on the number of true positives (TP) and establish ratios to the total number of detection and ground truths. We also represent the confusion matrix for a better understanding of the results

$$\text{Recall} = \frac{\text{TP}}{\text{All ground-truths}} \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{All detections}} \tag{2}$$

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{3}$$

III. RESULTS AND DISCUSSION

Table III summarizes the metrics used to evaluate the performance of the DL models developed for grapevine classification, and the confusion matrices for the transfer learned models are presented in Figs. 5–7.

The assessed results comprise the performance of the DL models in the test set of the augmented dataset, i.e., in 6109 images. For better engagement with the originally acquired data, we use less aggressively augmented data in the test set. As initially expected, transfer learned models performed better than trained from scratch models (Table III) [28], [29]. Pretrained models on the ImageNet dataset are prepared with tools to identify many features in the images, while our data has a restricted number of features. Therefore, we will be analyzing the transfer learned models only.

Despite the effort to keep reliability in the augmentation process to generate more varied data similar to the original one, additional experiments prove that it is difficult to avoid performance decreasing in the augmented dataset. Table III displays the DL models' performance in the augmented dataset. In addition, we also essayed the same models in the test set without augmentation (i.e., using only the initially acquired and preprocessed images—146 images). The results are more satisfactory, behaving the different DL models with the $F1$ score of 99.33 %, 94.41 %, and 95.26 % for the MobileNet V2, ResNet34, and VGG11 BN, respectively. In this last essay, we used the transfer learned models. As happened in Table III, the performance of the models trained from scratch was worse. Although, the low amount of leaves for each variety cannot offer us reliable conclusions. So, in the following analysis, we only essay the models in the augmented test set.

MobileNet v2 outcome the results with the best $F1$ score, having more confusion in the classification of leaves

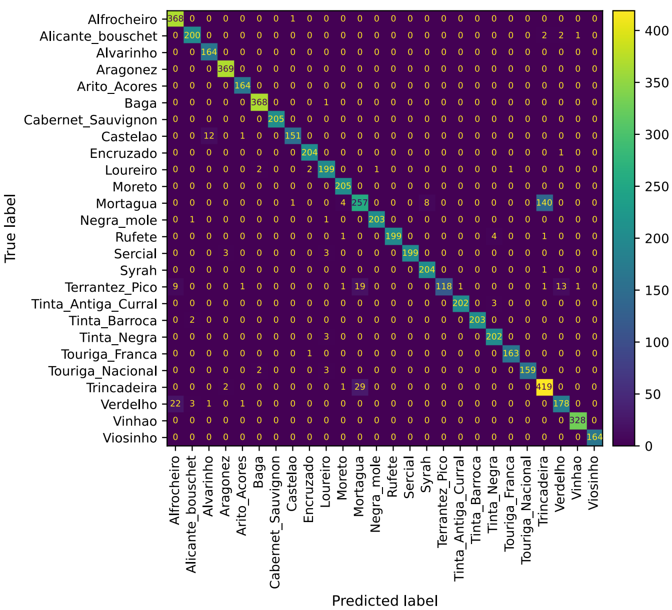


Fig. 5. Confusion matrix for MobileNet v2 classifier.

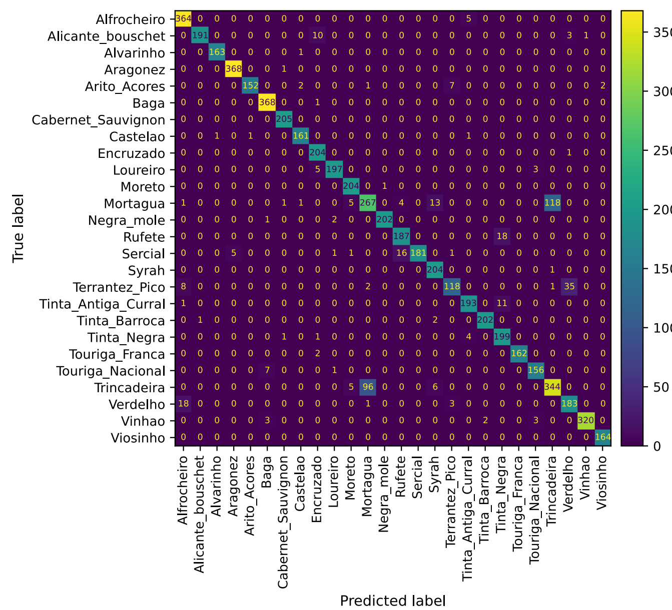


Fig. 7. Confusion matrix for VGG-11-BN classifier.

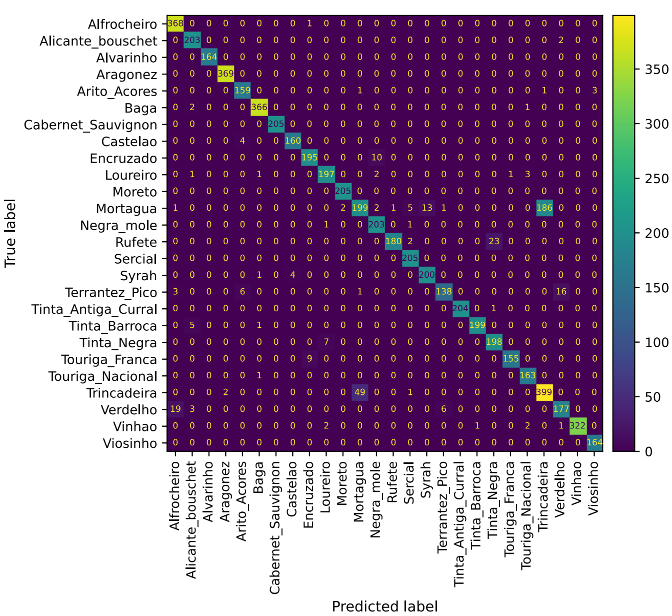


Fig. 6. Confusion matrix for ResNet-34 classifier.

between Mortágua and Trincadeira (Fig. 5). ResNet-34 and VGG-11-BN performed similarly with slightly lower results. Anyway, the performance of the networks for all the tested models is high enough to support digital ampelometry.

A deeper analysis of the confusion matrices in Figs. 5–7 allows stating similarity relationship between grapevine varieties. In a general overview, the selected grapevine varieties have a low similarity ratio, increasing identification success. The CNN could discretize the different varieties' features for successful classification. However, some varieties have similar features, which leads to some classification mistakes between these varieties. These cases can be stated, for instance, between Mortágua and Trincadeira, which are the most confusable varieties. These two varieties are two clones of the same variety with identical genetic information (also, Tinta Antiga do Curral and Tinta Negra are clones). Besides, this dataset also contains parent relationship: Verdelho is a parent of Terrantez do Pico, and Alfrocheiro, Verdelho, and Viosinho are siblings, having the same parent Savagnin Blanc.[3] The essays on the original test set are unreliable due to the small amount of data. However, the augmented test set to 40 times the original one allows for validating the suitability of these models to distinguish and classify between the different grapevine varieties.

Considering the high relationship between Mortágua and Trincadeira, and while they are clones of the same variety, if consider them in the same class, we get that MobileNet v2 reaches an $F1$ score of 98 %, ResNet-34 reaches an $F1$ score of 97 %, and VGG-11-BN reaches an $F1$ score of 96 %. All these results were got on the augmented test set. Therefore, removing the highly correlated classes in this dataset, we got results near 100 %.

From the observed results of the different experiments, we can conclude that all the essayed DL models are suitable for digital ampelometry. However, there are still some concerns in this experiment that should be approached. Despite augmentation improving the network performance, this technique does not offer new relevant information and variability in data. Instead, augmentation provides data with small variations to the original one, making the trained model more robust and prone to overfitting. So, the biggest efforts in future work should concern increasing the amount of data with representative and heterogeneous samples of leaves from the different grapevine varieties.

The essayed DL models are the most common architectures in the literature and widely used for multiple applications,

[3]The grapevine varieties can be reproduced with small changes in their genetic information. As a result, they are clones and have too many similarities. The analysis for synonyms and the relationship between varieties can be supported by the VIVC database [8].

TABLE IV
SUMMARY OF THE PERFORMANCE OF THE ESSAYS IN THE LITERATURE AGAINST OURS

	No. varieties	No. samples	Model	Accuracy
[16]	16	138 – 144	3-layer ANN	63.60 %
[19]	6	300	VGG-16	99.00 %
[20]	6	480	VGG-16 ResNet XCeption	87.5 % 97.10 % 100.00 %
[21]	6	224	AlexNet	71.30 %
[22]	21	5091	AlexNet VGG-16 ResNet-101 ResNet-18 DenseNet GoogLeNet	94.70 % 96.85 % 97.24 % 95.68 % 94.70 % 99.66 %
[23]	5	500	MobileNet v2	97.20 %
ours	26	976 (40 428)	MobileNet v2 ResNet-34 VGG-11-BN	95 % 93 % 93 %

but more recent studies are showing up with relevant results in the benchmarking datasets (ImageNet); this is the case of EfficientNet family and Visual Transformers, that should be essayed. We chose to use small size DL models with a low number of trainable parameters, because we have a limited amount of data that could compromise the training of the DL models and conducting to overfitting. However, increasing the number of samples could provide more varied data and the reinforcement of features, leveraging to bigger and more capable ANN.

The current solution aims to be an offline product that could identify grapevine varieties under request in an external and dedicated server. The user has time to wait for a system's answer, and the application should have a restricted number of requests. So, this study does not aim to analyze the models' performance under real-time conditions nor in restricted systems, such as mobile platforms. So, the future use of deeper DL models and transformations to the models' architecture should approach to improve the networks' results and accommodate more grapevines varieties.

Compared with the literature, the obtained results perform similar to the other works, namely, [22]. Table IV summarizes this comparison of our results with the performance of other works in the literature. While the approach presented in [19] performed better than ours, they used a short sample of six vine varieties. A similar conclusion can be obtained from [23] and [20]. So, against most studies in the literature, our work benefits from an extended dataset with varied grapevine varieties, with some similarities between the leaves' shape and size, grapevine clones (such as between Mortágua and Trincadeira, or Tinta Antiga do Curral and Tinta Negra), and parent proximity. Besides, our work benefits from a curated dataset gathered under ampelographic supervision and a restricted protocol of collecting leaves from the mid-shoot of genetically confirmed vineyards. A fair comparison with our work is found in [22], that do not benefit from the last protocol characteristics. Similar to us, these authors also looked at many grapevine varieties and studied diverse DL models. However, [22] reached a better accuracy than us. The dataset comprises 5091 images with more than 200 images for each class. This aspect highlights future lines to improve our work. The dataset used in [22], despite only comprising images of leaves under controlled environments, also comprises images of leaves acquired in the trees under natural conditions. These conditions have lighting variations, different distances to the leaf sample, and noisy background with other leaves, trunks, and other objects. Despite challenging the model, this varied dataset gave it relevant and varied features for image classification. Against the work present at [22], this study benefits from a structured protocol to gather the samples from the vines. Also, we added four more grapevine varieties and have a high correlation ratio due to kinship closeness, such as between Touriga Nacional and Touriga Franca. Besides, this study benefits from samples of the main grapevine varieties in Portugal.

Analyzing the used protocols in the reviewed literature, only [16] and [19] stated a detailed data acquisition protocol. Both authors gathered leaf samples from the trees' fifth nodes from the apex. Adão et al. [20] and Koklu et al. [23] did not detail their protocol, and the other authors did not gather leaves nor explain how did they acquire the images. Using that leaves, [16] and [19] are avoiding child and undeveloped leaves, contributing with fully developed samples of the tree for the dataset creation. According to best ampelometry practices, the ninth node is most representative of grapevine variety [12]. However, limiting our dataset to the ninth node restricts the amount of data. In this scenario, the 7th and 11th nodes can also be considered with some similar representativeness. Besides, these leaves contribute to polymorphism in the dataset. Most of the time, the 11th node is not fully developed, representing the samples of undeveloped leaves of the vine.

IV. CONCLUSION

We present a benchmark of three DL models for assessing the capability of CNN to identify grapevine varieties through the leaf between an increased dataset with many vine cultivars. A manually acquired dataset was used, containing 976 samples of leaves from 26 different classes.

The dataset benefited from a manually curated dataset using samples of leaves carefully acquired from genetically identified trees and from the 7th, 9th, and 11th nodes. So, all the image samples in the dataset are fully representative of the different grapevine varieties and noise free.

The essay proves that all the assessed DL models performed similarly, but MobileNet v2 and ResNet-34 are the best. However, additional experiments are required with an improved and increased dataset with more varied data and more samples for each variety.

The ability to correctly identify grapevine varieties of the developed DL models justifies their use for improved in situ high-through-output digital phenotyping systems. However, substantial work is still required to improve the models with more samples for each grapevine variety and increase the number of grapevine varieties in the dataset.

Future work concerns to the following.

1) Balance the number of samples between classes.
2) Increase the number of vine varieties and the number of samples for each variety in the dataset.
3) Add more varied data from different sensors and with more varied backgrounds and luminosity.
4) Visualize the class activation maps to understand better the most critical features of the leaves for classification.
5) Essay architectural modifications in the DL models to improve the system's accuracy.

Acknowledgment

Tiago Cerveira Padilha and Paulo Silva are with Vimetrica Agro, 4500-822 Nogueira da Regedoura, Portugal (e-mail: tiago.padilha@vimetrica.pt; paulo.silva@vimetrica.pt).

Frederico de Carvalho is with the School of Engineering, University of Trás-os-Monte e Alto Douro (UTAD), 5000-801 Vila Real, Portugal (e-mail: al77688@alunos.utad.pt).

Jorge Cunha is with the Estação Vitivinícola Nacional, Instituto Nacional de Investigação Agrária e Veterinária (INIAV), 2565-191 Dois Portos, Portugal (e-mail: jorge.cunha@iniav.pt).

References

[1] L. A. Panzone and O. M. Simões, "The importance of regional and local origin in the choice of wine: Hedonic models of Portuguese wines in Portugal," *J. Wine Res.*, vol. 20, no. 1, pp. 27–44, Mar. 2009.

[2] M. Cunha and C. Richter, "The impact of climate change on the winegrape vineyards of the Portuguese Douro region," *Climatic Change*, vol. 138, nos. 1–2, pp. 239–251, Jun. 2016.

[3] (2021). *State of the World Vitivinicultural Sector in 2020*. Organisation Internationale de La Vigne Et Du Vin (OIV). Report. Accessed: Sep. 5, 2022. [Online]. Available: https://www.oiv.int/public/medias/7909/oiv-state-of-the-world-vitivinicultural-sector-in-2020.pdf

[4] J. Cunha et al., "Characterisation of the Portuguese grapevine germplasm with 48 single-nucleotide polymorphisms," *Austral. J. Grape Wine Res.*, vol. 22, no. 3, pp. 504–516, Oct. 2016.

[5] J. Cunha et al., "Genetic relationships among Portuguese cultivated and wild *Vitis vinifera* L. Germplasm," *Frontiers Plant Sci.*, vol. 11, p. 127, Mar. 2020.

[6] A. Schneider, A. Carra, A. Akkak, P. This, V. Laucou, and R. Botta, "Verifying synonymies between grape cultivars from France and Northwestern Italy using molecular markers," *Vitis J. Grapevine Res.*, vol. 40, no. 4, pp. 197–203, 2001.

[7] (2013). *International List of Vine Varieties and Their Synonyms*. Organisation internationale de La Vigne et du Vin, Report. Accessed: Sep. 5, 2022. [Online]. Available: https://www.oiv.int/public/medias/2273/oiv-liste-publication-2013-complete.pdf

[8] E. Maul. (2022). *Vitis International Variety Catalogue (VIVC)*. Accessed: Sep. 5, 2022. [Online]. Available: http://www.vivc.de

[9] B. Işçi, "Genetic relationships of some local and introduced grapes (*Vitis vinifera L.*) by microsatellite markers," *Biotechnol. Biotechnolog. Equip.*, vol. 33, no. 1, pp. 1303–1310, Jan. 2019.

[10] K. M. Sefc, F. Lefort, M. S. Grando, K. D. Scott, H. Steinkellner, and M. R. Thomas, *Microsatellite Markers for Grapevine: A State Art*. Dordrecht, The Netherlands: Springer Netherlands, 2001, pp. 433–463.

[11] D. H. Chitwood et al., "A modern ampelography: A genetic basis for leaf shape and venation patterning in grape," *PLANT Physiol.*, vol. 164, no. 1, pp. 259–272, Jan. 2014.

[12] A. Rodrigues, *Um Método Filométrico de Caracterização Ampelográfica*. Lisbon, Portugal: Portuguese Ministery of Agriculture, Serviço Editorial da Repartição de Estudos, Informação e Propaganda, 1952.

[13] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri, "NIH image to ImageJ: 25 years of image analysis," *Nature Methods*, vol. 9, no. 7, pp. 671–675, Jun. 2012.

[14] J. Huixian, "The analysis of plants image recognition based on deep learning and artificial neural network," *IEEE Access*, vol. 8, pp. 68828–68841, 2020.

[15] S. Dara and P. Tumma, "Feature extraction by using deep learning: A survey," in *Proc. 2nd Int. Conf. Electron., Commun. Aerosp. Technol. (ICECA)*, Mar. 2018, pp. 29–31.

[16] S. Fuentes et al., "Automated grapevine cultivar classification based on machine learning using leaf morpho-colorimetry, fractal dimension and near-infrared spectroscopy parameters," *Comput. Electron. Agricult.*, vol. 151, pp. 311–318, Aug. 2018.

[17] S. Gutiérrez, J. Fernández-Novales, M. P. Diago, and J. Tardaguila, "On-the-go hyperspectral imaging under field conditions and machine learning for the classification of grapevine varieties," *Frontiers Plant Sci.*, vol. 9, p. 1102, Jul. 2018.

[18] A. M. Fernandes, A. B. Utkin, J. Eiras-Dias, J. Cunha, J. Silvestre, and P. Melo-Pinto, "Grapevine variety identification using 'big data' collected with miniaturized spectrometer combined with support vector machines and convolutional neural networks," *Comput. Electron. Agricult.*, vol. 163, Aug. 2019, Art. no. 104855.

[19] A. Nasiri, A. Taheri-Garavand, D. Fanourakis, Y.-D. Zhang, and N. Nikoloudakis, "Automated grapevine cultivar identification via leaf imaging and deep convolutional neural networks: A proof-of-concept study employing primary Iranian varieties," *Plants*, vol. 10, no. 8, p. 1628, Aug. 2021.

[20] T. Adão et al., "Digital ampelographer: A CNN based preliminary approach," in *Progress in Artificial Intelligence*, P. M. Oliveira, P. Novais, and L. P. Reis, Eds. Cham, Switzerland: Springer, 2019, pp. 258–271.

[21] C. S. Pereira, R. Morais, and M. J. C. S. Reis, "Deep learning techniques for grape plant species identification in natural images," *Sensors*, vol. 19, no. 22, p. 4850, Nov. 2019.

[22] Y. Liu et al., "Development of a mobile application for identification of grapevine (*Vitis vinifera* L.) cultivars via deep learning," *Int. J. Agricult. Biol. Eng.*, vol. 14, no. 5, pp. 172–179, 2021.

[23] M. Koklu, M. F. Unlersen, I. A. Ozkan, M. F. Aslan, and K. Sabanci, "A CNN-SVM study based on selected deep features for grapevine leaves classification," *Measurement*, vol. 188, Jan. 2022, Art. no. 110425.

[24] D. Fanourakis, F. Kazakos, and P. A. Nektarios, "Allometric individual leaf area estimation in chrysanthemum," *Agronomy*, vol. 11, no. 4, p. 795, Apr. 2021.

[25] D. H. Chitwood, "The shapes of wine and table grape leaves: An ampelometric study inspired by the methods of Pierre Galet," *Plants, People, Planet*, vol. 3, no. 2, pp. 155–170, Mar. 2021.

[26] C. Konlechner and U. Sauer, "Ultrastructural leaf features of grapevine cultivars (*Vitis vinifera* L. ssp. vinifera)," *Oeno One*, vol. 50, no. 4, pp. 1–18, 2016.

[27] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: Fast and flexible image augmentations," *Information*, vol. 11, no. 2, p. 125, Feb. 2020.

[28] S. J. Pan and Q. Yang, "'A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[29] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger, Eds. Red Hook, NY, USA: Curran Associates, 2014, pp. 1–9.

[30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, May 2015, doi: 10.48550/ARXIV.1409.1556.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[32] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.

[33] Y. Yang et al., "A comparative analysis of eleven neural networks architectures for small datasets of lung images of COVID-19 patients toward improved clinical decisions," *Comput. Biol. Med.*, vol. 139, Dec. 2021, Art. no. 104887.