Preprints.org

Review

Not peer-reviewed version

Deep Learning for Automatic Grapevine Varieties Identification: A Brief Review

Gabriel A. Carneiro * , António Cunha , Joaquim Sousa

Posted Date: 8 March 2024

doi: 10.20944/preprints202403.0484.v1

Keywords: Deep Learning; Computer Vision; Grape Variety Classification; Grape Variety Identification; Artificial Intelligence; Explainable Artificial Intelligence; Precision Viticulture

Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 March 2024 doi:10.20944/preprints202403.0484.v1

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Deep Learning for Automatic Grapevine Varieties Identification: A Brief Review

Gabriel A. Carneiro [1,2,*], António Cunha [1,2] and Joaquim Sousa [1,2]

[1] University of Trás-os-Montes and Alto Douro, 5000-801, Vila Real, Portugal
[2] Instituto de Engenharia de Sistemas e computadores, Tecnologia e Ciência, 3200-465, Porto, Portugal
[*] Correspondence: gabrielc@utad.pt

Abstract: The Eurasian grapevine (*Vitis vinifera L.*) is the most widely grown horticultural crop in the world and is important for the economy of many countries. In the wine production chain, grape varieties play an important role, as they directly influence the authenticity and classification of the product. Identifying the different grape varieties is therefore fundamental for quality control and control activities, as well as for regulating production. Currently, ampelography and molecular analysis are the main approaches to identifying grape varieties. However, both methods have limitations. Ampelography is subjective and prone to errors and is experiencing enormous difficulties as ampelographers are increasingly scarce. On the other hand, molecular analyses are very demanding in terms of cost and time. In this scenario, Deep Learning (DL) methods have emerged as a classification alternative to deal with the scarcity of ampelographers and avoid molecular analyses. In this study, the most recent and current methods for identifying grapevine varieties using DL classification-based approaches are presented through a systematic literature review. The steps of the standard DL-based classification pipeline were described for the 18 most relevant studies found in the literature, highlighting their pros and cons. Potential directions for improving this field of research were also presented.

Keywords: deep learning; computer vision; grape variety classification; grape variety identification; artificial intelligence; explainable artificial intelligence; precision viticulture

---

1. Introduction

The Eurasian grapevine (*Vitis vinifera L.*) holds the title of being the most extensively cultivated and economically significant horticultural crop globally, being cultivate since the ancient times [1]. Due to its substantial production, this crop plays a crucial part in the economies of many countries [2]. The fruit is important because it can be used to consumption and also to the production of wine. The number grape varieties present in the world is unknown, but specialists estimate it to be around 5000 to 8000, under 14000 to 24000 different names [3–5]. Despite this huge number, only 300 or 400 varieties account most of the grape plantings in the world [4]. The most common varieties of grapes in the world are Kyoho, Carbenet Sauvignon, Sultanina, Merlot, Temparillo, Airen, Chardonnay, Syrah, Red Globe, Grenache Noir, Pinot Noir, Trebbiano Toscano [6].

The grape variety plays an important role in the wine production chain and in the leave consumption, since in some cases the they can be more costly than the fruit [7,8]. Wine is one of the most popular agri-foods in the four corners of the world [9]. In 2019, the European Union accounted for 48% of world consumption and 63% of world production [10]. In terms of value, the wine market share totalled almost 29.6 billion euros in 2020, despite the Covid-19 pandemic crisis [10]. The varieties used on the production of the drink directly influences its authenticity and classification, and due to its socioeconomic importance, identifying grape varieties became an important part of the production regulation. Furthermore, recent results achieved by Jones and Alves [11] highlighted that some varieties can be prone to wamer environments, in the context of climate changes, accentuating the need of tools for grapevine variety identification.

Nowadays the identification of grapevine varieties is carried out mostly using ampelography or molecular analysis. Ampelography, defined by Chitwood et al. [12] as "the science of phenotypic distinction of vines", is one of the most accurate ways of identifying grape varieties through visual

© 2024 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 March 2024 doi:10.20944/preprints202403.0484.v1

analysis. Its authorised reference is Precis D'Ampelographie Pratique [13]; however, it uses well-defined official descriptors provided in the identity of the plant material for grape identification [14,15]. Despite it wide utilisation, Ampelography depends on the person carrying it out, as with any visual analysis task, making the process subjective. It can be exposed to interference from environmental, cultural and genetic conditions, introducing uncertainty into the identification process [14,16]. It can be time-consuming and error-prone, just like any other human-based task, and ampelographers are becoming scarce [17].

Molecular markers is another technique that has been used to identify grape varieties [17]. Among the used markers, random amplified polymorphic DNA, amplified fragment length polymorphism and microsatellite markers have been used in the grape variety identification [17]. This technique makes it possible to deal with subjectivity and environmental influence. However, it must be complemented by ampelography due to leaf characteristics that can only be assessed in the field [3,18,19]. In addition, the identification of grape varieties with a focus on production control and regulation would involve several molecular analyses, increasing the costs and time required.

With the advance of computer vision techniques and easier access to data, several studies have emerged with the aim of automatically identifying grapevine varieties. Initially, they were based on classic machine learning classifiers, e.g. Support Vector Machines, Artificial Neural Networks [20], Nearest Neighbour algorithm [21], Partial least squares regression [22], and using manually or statically extracted characteristics, e.g. indices, or the data directly. However, in 2012, with the advent of Deep Learning (DL), more specifically the study by Krizhevsky et al. [23], computer vision classifiers became capable of reaching or, in some cases, surpassing human capacity. Lately, transfer learning and fine-tuning approaches have allowed these models to be applied to many general computer vision tasks, such as object detection, semantic segmentation and instance segmentation, and in other research domains, for example precision agriculture and medical image analysis. The automatic identification of grapevine varieties has followed this lead, and most studies now use DL-based classifiers in their approaches.

In this study, recent literature on the identification of grapevine varieties using DL-based classification approaches was reviewed. The steps of the DL-based classification process (data preparation, choice of architecture, training and model evaluation) were described for the 18 most relevant studies found in the literature, highlighting their pros and cons. Possible directions for improving this field of research are also presented. To the best of our knowledge, there are no studies in the literature with the same objective. However, this study may have some intersection with Chen et al. [24], which aimed to review studies that used deep learning for plant image identification. Besides, Mohimont et al. [25] reviewed studies that used computer vision and DL for yield-related precision viticulture tasks, e.g. flower counting, grape detection, berry counting and yield estimation, while Ferro and Catania [26] surveyed the technologies employed in precision viticulture, covering topics ranging from sensors to computer vision algorithms for data processing. It is important to emphasise that the explanation of computer vision algorithms is already widespread in the literature and will not be covered in this study. One can refer to Chai et al. [27] and Khan et al. [28] for advances in the field of natural scenes, or Dhanya et al. [29] for developments in the field of agriculture.

The remainder of this article is organised as follows. In Section 2, the research questions, inclusion criteria, search strategy and extraction of the characteristics of the selected studies are described. Then, in Section 3, the results are presented, highlighting the approach used in the stage of creating the DL-based classifier. In Section 4, a discussion around the selected studies is presented, focussing on the pros and cons of the approaches used and also introducing techniques that can still be explored in the context of identifying grapevine varieties using DL-based methods. Finally, in Section 5, the main conclusions are presented.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 March 2024 doi:10.20944/preprints202403.0484.v1

2. Methods

2.1. Research Questions

The following research questions (RQ) were used to substantiate this systematic review:

- (RQ1) How Deep-Learning techniques have been used for automatic identification of grapevine varieties?
- (RQ2) Which are the best DL-based models for Automatic Grapevine Verities Identification?
- (RQ3) What are the main challenges and future development trends in identifying grape varieties using DL-based models?

2.2. Inclusion Criteria

The following three inclusion criteria were used to conduct the study: (1) studies written in English; (2) studies that used a DL-based approach to identify different vine varieties; and (3) studies published between 2019 and 2024. The search was limited to 2019 because the field of DL research has shown a considerable improvement in performance over the last five years.

2.3. Search Strategy

Studies that met the inclusion criteria in the following electronic databases were considered: IEEE Xplore, MDPI, ScienceDirect and Springer. The International Journal of Academic Engineering Research (IJAER) was also included. The expression *(("grape variety" OR "grapevine") AND ("classification" OR "identification") AND "deep learning")* was used to identify the articles in these databases. Finally, the articles identified were evaluated by their title and abstract, and those considered to be outside the scope of this work were discarded. This consultation was carried out on 1 February 2024.

2.4. Extraction of Characteristics

In order to better understand the context in which the studies were carried out, several characteristics were surveyed and organised in Table 1: year of publication, location, description of the dataset used, part of the vine focused on, architecture used and results. The results are evaluated using the F1 Score, Accuracy (Acc) or Area-Under-The-Curve (AUC) metrics, in this order of preference. Only the best result obtained is presented. It is important to note that the description can include the number total of images, considering generated samples.

Table 1. Overview about the selected studies. The year, dataset location, description, the part of plant where the images were focused, used architectures and the results are presented.

Study	Year	Dataset Location	Dataset Description	Focus	Architecture	Results
Doğan et al. [30]	2024	Turkey	7000 RGB images distributed between 5 classes acquired in controlled environment	Leaves	Fused Deep Features + SVM	1.00 (F1)
Magalhães et al. [31]	2023	Portugal	40428 RGB images distributed between 26 classes acquired in a controlled environment	Leaves	MobileNetV2, ResNet-34 and VGG-11	94.75 (F1)
Carneiro et al. [32]	2023	Portugal	6216 RGB images distributed between 14 classes acquired in field	Leaves	EfficientNetV2S	0.89 (F1)
Carneiro et al. [33]	2023	Portugal	675 RGB images distributed between 12 classes; 4354 RGB images distributed between 14 classes; both acquired in field	Leaves	EfficientNetV2S	0.88 (F1)
Gupta and Gill [34]	2023	Turkey	2500 RGB images distributed between 5 classes acquired in controlled environment	Leaves	EfficientNetB5	0.86 (Acc)
Carneiro et al. [35]	2022	Portugal	28427 RGB images distributed between 6 classes acquired in field	Leaves	Xception	0.92 (F1)
Carneiro et al. [36]	2022	Portugal	6922 RGB images distributed between 12 classes acquired in-field	Leaves	Xception	0.92 (F1)
Carneiro et al. [37]	2022	Portugal	6922 RGB images distributed between 12 classes acquired in-field	Leaves	Vision Transformer (ViT_B)	0.96 (F1)
Koklu et al. [7]	2022	Turkey	2500 RGB images distributed between 5 classes acquired in controlled environment	Leaves	MobileNetV2 + SVM	97.60 (Acc)
Carneiro et al. [38]	2021	Portugal	6922 RGB images distributed between 12 classes acquired in-field	Leaves	Xception	0.93 (F1)
Liu et al. [39]	2021	China	5091 RGB images distributed between 21 classes acquired in-field	Leaves	GoogLeNet	99.91 (Acc)
Škrabánek et al. [40]	2021	Czech Republic	7200 RGB images distributed between 7 classes acquired in-field	Fruits	DenseNet	98.00 (Acc)
Nasiri et al. [41]	2021	Iran	300 RGB images distributed between 6 classes acquired in a controlled environment	Leaves	VGG-16	99.00 (Acc)
Peng et al.[42]	2021	Brazil	300 RGB images distributed between 6 classes acquired in-field	Fruits	Fused Deep Features	96.80 (F1)
Franczyk et al. [43]	2020	Brazil	3957 RGB images distributed between 5 classes acquired in-field	Fruits	ResNet	99.00 (Acc)
Fernandes et al. [44]	2019	Portugal	35933 Spectra distributed between 64 classes in-field	Leaves	Handcraft	0.68 (AUC) 0.98 (AUC)
Adão et al. [45]	2019	Portugal	3120 RGB images distributed between 6 classes acquired in controlled environment	Leaves	Xception	100.00 (Acc)
Pereira et al. [46]	2019	Portugal	224 RGB images distributed between 6 classes acquired in controlled environment	Leaves	AlexNet	77.30 (Acc)

3. Results

As shown in Table 1, 18 studies were identified from the selected sources. Figure 1 shows a graph comprising the countries of origin of the datasets, the years and the focus of the selected studies. The majority of the studies were published in 2021 and 2023, and most of the datasets used have Portugal as their source of localisation. Therefore, this field of research has been active these days, especially in countries where the grape cultivar is economically relevant. Furthermore, most studies have focused on the leaves to identify grape varieties.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 March 2024 doi:10.20944/preprints202403.0484.v1

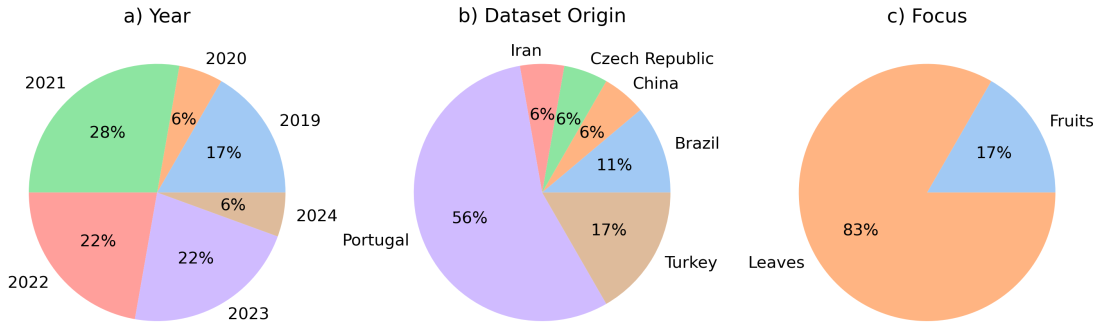


Figure 1. Years, countries of origin of the datasets and focus of the selected studies. Most of the studies were published in 2021. Portugal was the biggest source of datasets. The studies focus on leaves rather than fruit to identify grape varieties.

All the selected studies followed the classic process for training DL models in their method. This process can be seen in Figure 2. First, the data is acquired and prepared for training DL models. Next, pre-processing steps are applied to the data in order to increase the quality of the classification. Then, architectures are selected or created, and subsequently trained on the data. In the final step, these models are evaluated. In order to better understand the different approaches used in the different steps, the pipeline will be followed to guide this discussion.

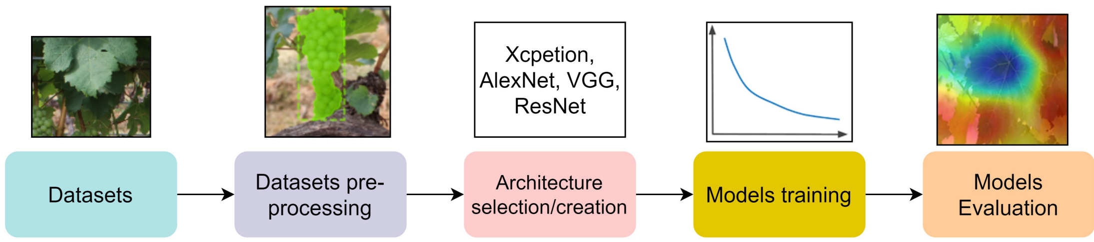


Figure 2. Classic pipeline for training DL-based models. Starting with data acquisition and ending with model evaluation.

Firstly, the Datasets and Benchmarks used in the studies will be presented, and then the approaches used in the pre-processing stage will be detailed. Next, the architecture and training process adopted will be discussed. Finally, the metrics and explanation techniques used to evaluate the studies will be discussed.

3.1. Datasets and Benchmarks

Details of the datasets used by the studies included in this review are presented in Table 2.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 March 2024 doi:10.20944/preprints202403.0484.v1

Table 2. Datasets characteristics for each selected study.

Study	Acquisition Device	Publicly Available	Acquisition Period	D. Augmentation	Acquisition Environment
Doğan et al. [30]	Camera - Prosilica GT2000C	Yes	-	Static augmentations and artificially generated images	Special illumination box
Magalhães et al. [31]	Kyocera TASKalfa 2552ci	No	1 day in June 2021	Blur, rotations, variations in brightness, horizontal flips, and gaussian noise	Special illumination box
Carneiro et al. [32]	Smartphones	No	Seasons 2021 and 2020	Static augmentations, CutMix and RandAugment	In-field
Carneiro et al. [33]	Camera and Smarthphone	No	1 Season and 2 Seasons	Rotations, Flips and Zoom	In-field
Gupta and Gill [34]	Camera - Prosilica GT2000C	Yes	-	Angle, scaling factor, translation	Special illumination box
Carneiro et al. [35]	Mixed Camera and Smartphone	No	Seasons 2017 and 2020	Rotation, shift, flip, and brightness changes	In-field
Carneiro et al. [36]	Camera - Canon EOS 600D	No	-	Rotations, shifts, variations in brightness and flips	In-field
Carneiro et al. [37]	Camera - Canon EOS 600D	No	1 Season	Rotations, shifts, variations in brightness and flips	In-field
Koklu et al. [7]	Camera - Prosilica GT2000C	Yes	-	Angle, scaling factor, translation	Special illumination box
Carneiro et al. [38]	Camera - Canon EOS 600D	No	1 Season	Rotations, shifts, variations in brightness and flips	In-field
Liu et al. [39]	Camera - Canon EOS 70D	No	-	Scaling, transposing, rotation and flips	In-field
Škrabánek et al. [40]	Camera – Canon EOS 100D and Canon EOS 1100D	No	2 days in August 2015	-	In-field
Nasiri et al. [41]	Camera – Canon SX260 HS	On request	1 day in July 2018	Rotation, height and width shift	Capture station with artificial light
Peng et al.[42]	Mixed Camera and Smartphone	Yes	1 day in April 2017 and 1 day in April 2018	-	In-field
Franczyk et al. [43]	Mixed Camera and Smartphone	Yes	1 day in April 2017 and 1 day in April 2018	-	In-field
Fernandes et al. [44]	Spectrometer – OceanOptics Flame-S	No	4 days in July 2017	-	In-field
Adão et al. [45]	Camera - Canon EOS 600D	No	Season 2017	Rotations, contrasts/brightness, vertical/horizontal mirroring, and scale variations	Controlled environment with white background
Pereira et al. [46]	Mixed Camera and Smartphone	No	Seasons 2016 and 2019	Translation, reflection, rotation	In-field

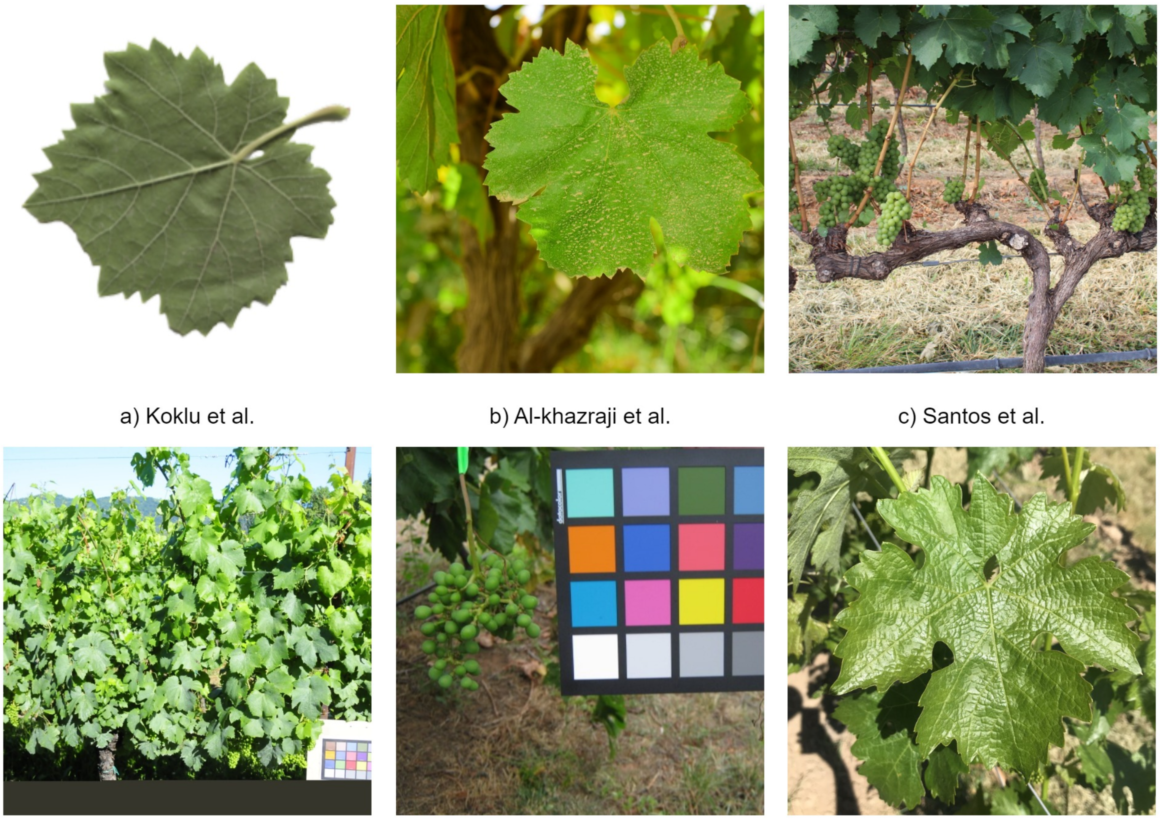
Images are the main way of identifying grapevine varieties in DL-based classification studies, although some of them have focused on leaves and others on fruit. Most of the studies used datasets composed of images acquired in the field with a camera. This is justified because classifiers trained with images acquired in a controlled environment can have limited usability. In addition, most controlled environment-based techniques are invasive, requiring the leaf to be removed from the plant.

Similarly to other fields of research, only a few studies provided their datasets. Peng et al. [42] and Franczyk et al. [43] used the grape instance segmentation dataset from Embrapa Vinho (Embrapa WGISD) [47]. This dataset is composed of 300 images belonging to 6 grape varieties. Koklu et al. [7] provided the dataset used and, more recently, studies by Doğan et al.[30] and Gupta and Gill [34] followed and explored the same dataset. This dataset is comprised of 5 classes and was acquired in a controlled environment, resulting in 500 images. In addition, other datasets have been proposed in the literature. Al-khazraji et al. [48] proposed a dataset with 8 different grape varieties acquired in the field. Vlah [49] organised a dataset composed of 1009 images distributed over 11 varieties. Sozzi et al. [50] proposed a dataset for the detection of bunches, but it can be used to identify 3 different varieties. Along the same lines, Seng et al. [51] presented a dataset with images of fruit at different stages of development that is composed of 15 different varieties. Table 3 summarises all the publicly available datasets that, as far as we know, can be used to train and evaluate DL models with the aim of classifying different grape varieties. Figure 3 shows examples of images for each publicly available dataset.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 March 2024 doi:10.20944/preprints202403.0484.v1

Table 3. List of publicly available datasets which can be used for grapevine variety classification.

Dataset	Number of Classes	Number of images	Classes	Balanced
Koklu et al. [7]	5	500	Ak, Ala Idris, Buzgulu, Dimnit, Nazli	Yes, 100 images per class
Al-khazraji et al. [48]	8	8000	deas al-annz, kamali, halawani, thompson seedless, aswud balad, riasi, frinsi, shdah	Yes, 1000 images per class
Santos et al. [47]	6	300	Chardonnay, Cabernet Franc, Cabernet Sauvignon, Sauvignon Blanc, Syrah	No
Sozzi et al. [50]	3	312	Glera, Chardonnay, Trebbiano	No
Seng et al. [51]	15	2078	Merlot, Cabernet Sauvignon, Saint Macaire, Flame Seedless, Viognier, Ruby Seedless, Riesling, Muscat Hamburg, Purple Cornichon, Sultana, Sauvignon Blanc, Chardonnay	No
Vlah [49]	11	1009	Auxerrois, Cabernet Franc, Cabernet Sauvignon, Chardonnay, Merlot, Müller Thurgau, Pinot Noir, Riesling, Sauvignon Blanc, Syrah, Tempranillo	No



a) Koklu et al. b) Al-khazraji et al. c) Santos et al.

d) Sozzi et al. e) Seng et al. e) Vlah

Figure 3. Example images for each publicly available dataset. The images have been cropped into squares, so they cannot represent the actual size of the images in the dataset.

In the studies that provided the acquisition period, most used data obtained over a short period of time (less than a month). Carneiro et al. [32] and Carneiro et al. [33] used the most representative datasets, in terms of time, to identify grape varieties. It should be noted that since grapes are seasonal

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 March 2024 doi:10.20944/preprints202403.0484.v1

plants, it is very important to represent different periods of the season in the dataset in order to capture the different phenotypic characteristics of the leaves over time. Seasonal representation in the dataset directly implies the classifier's ability to generalise.

To the best of our knowledge, Fernandes et al. [44] is the only study that did not use images, using spectral data acquired in the field to identify grape varieties. In total, 35,933 spectra belonging to 64 different species were used. However, the aim of the study was only to separate two varieties from the others by creating two binary classifiers for each of them.

Furthermore, Magalhães et al. [31] was the only study that was concerned with the position of the leaves used in the classification. The authors argue that leaves from nodes between the 7th and 8th should be used, as they are the most representative in terms of phenotypic characteristics [52].

Given that DL-based classification models are prone to overfitting, most models have used data augmentation techniques to improve the quality of the data used during training. Rotations, reflections and translations are the main modifications applied to images. Furthermore, these modifications are generally only applied to the training subset. Carneiro et al. [32] was the only study that tested different techniques for data augmentation, however, they concluded that offline geometric augmentations (zoom, rotation and flips) led the models to better classification.

Furthermore, it is evident that the datasets primarily consisted of a limited number of varieties, in contrast to the estimation conducted by specialists, which involved at least 5000 varieties.

3.2. Pre-Processing

Few studies have presented pre-processing techniques to improve classification results. Fernandes et al.[44] after calculating the reflectance using the acquired spectra, applied the Savitzky-Golay (SG) filter, logarithm, multiplicative scatter correction (MSC), standard normal variant (SNV), first derivative and second derivative to the data, comparing the results for each approach.

In the image context, Liu et al. [39] used the complemented images in training, so that each colour channel in the resulting image was the complement of the corresponding channel in the original image. Pereira et al. [46] tested many types of pre-processing: fixed-point FastICA algorithm [53], canny edge detector [54], greyscale morphology processing [55], background removal with the segmentation method proposed by Pereira et al. [56], and the proposed four-corners-in-one method. The FastICA algorithm is an independent component analysis method based on kurtosis maximisation that was applied to blind source separation. The idea behind applying independent component analysis (ICA) to images is that each image can be understood as a linear superposition of weighted features and then ICA decomposes them into a statistically independent source base with minimal loss of information content to achieve detection and classification [57,58]. Unlike ICA, grey-scale morphological processing is a method of extracting vine leaves based on classical image processing. Firstly, the image is transformed into greyscale, based on its tonality and intensity information. Next, morphological greyscale processing is applied to remove colour overlap in the leaf vein and background. Linear intensity adjustment is then used to increase the difference in grey values between the leaf vein and its background. Finally, the Otsu threshold [59] is calculated to separate the veins from the background, and detailed processing is carried out to connect lines and remove isolated points [55].

The next method used by Pereira et al. [46] was proposed in Pereira et al. [56] and is also based on classical image processing. This study presents a method for segmenting grape leaves from fruit and background in images acquired in field. This approach is based on growing regions using a colour model and thresholding techniques and can be separated into three stages: pre-processing; segmentation; and post-processing. In pre-processing, the image is resized, the histogram is adjusted to increase contrast and then the resulting image and the raw image are converted to the hue, saturation and intensity (HSI) colour model, and the original image is also converted to the CIELAB (L*a*b*) colour model. In the segmentation phase, the luminance component of the raw image (L*) is used to detect the shadow regions, then the shadow and non-shadow regions are processed, removing

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 March 2024 doi:10.20944/preprints202403.0484.v1

the marks and the background with different approaches for each. Finally, in the post-processing step, the method fills in small holes using morphological operations. These holes are usually due to the presence of diseases, pests, insects, sunburn and dust on the leaves. The method achieved 94.80% average accuracy. Finally, the same authors also propose a new pre-processing method called four-corners-in-one. The idea is to concentrate all the non-removed pixels in the north-west corner of the image, after segmenting the leaves of the vines in an image; then a left-shift sequence is performed followed by a sequence of up-shift operations on the coloured pixels in the image. This algorithm is replicated for the other three corners. According to the authors, this method obtained the best classification accuracy in the set of experiments carried out.

Carneiro et al. [35] and Carneiro et al. [33] evaluated the use of segmentation models to remove the background from images acquired in field before classification. Both studies applied a U-Net [60], and Carneiro et al. [35] also tested SegNet [61] to segment the data before classification. The results show that performance can be reduced if secondary leaves are removed and the model trained with the segmented leaves paid more attention to the centre leaves.

Doğan et al. [30] used Enhanced Super Resolution Generative Adversarial Networks (ESRGAN) [62] to increase the resolution of the images, after decreasing it, so that the method could work as a generator of new samples. The idea is to apply a Generative Adversarial Network [63] to recover a high-resolution image from a low-resolution one. The authors applied this approach as a data augmentation technique, decreasing the resolution of the images in the dataset and increasing it again using ESRGAN, so that these new images were considered new samples.

3.3. Architecture and Training

The main approach used in the architecture was an ensemble of transfer learning and tuning. However, a few different techniques were also employed: hand-crafted architectures, Fused Deep Features and extraction using DL-based models plus Support Vector Machines (SVM) classifiers. It is important to note that Fernandes et al. [44] were the only authors who used a handcraft architecture, since their dataset was composed of spectra samples.

AlexNet [23], VGG-16 [64], ResNet [65], DenseNet [66], Xception [67] and MobileNetV2 [68] were the Convolutional Neural Networks (CNNs) architectures employed in the image-based studies included in this review. These networks were first trained on ImageNet and then transfer learning and fine-tuning were employed in two stages. In the first step, their classifiers are replaced by a new one and the convolutional weights are frozen so that only the new classifier is trained and has its weights updated (transfer learning). In the second step, all the weights are unfrozen and the entire architecture is retrained (tuning). Detailed information on each architecture used can be found in Alzubaidi et al. [69]. In a different way, Carneiro et al. [35] used a Vision Transformer (ViT)[70], but followed the same learning strategy.

Unlike other studies based on image classification, Peng et al. [42] and Doğan et al. [30] used Fused Deep Features to identify grapevine varieties. This approach consists of extracting features from images from more than one source, concatenating all the extracted features and then classifying them. Peng et al. [42] extracted features from AlexNet, ResNet and GoogLeNet, then fused the features using the Canonical Correlation Analysis algorithm [71] and then classified the vine varieties using an SVM classifier. In addition, the authors trained the aforementioned architectures with fully connected classifiers, which resulted in worse performance than the proposed method. They argued that the small size of the dataset is the main reason why it is difficult to obtain better results using CNN directly. Doğan et al. [30] merged attributes from VGG-19 and MobileNetV2. The difference is that the authors used a Genetic Based Support Vector Machine to select the best features for classification, improving the results by 3 percentage points. The final classification of the selected feature was done using an SVM. Koklu et al. [7] also used the features extracted from a pre-trained CNN architecture plus an SVM classifier. The idea was to extract features from the logits of the first fully connected layer of MobileNetV2 and use them to test the performance of four different SVM kernels: Linear, Quadratic,

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 March 2024 doi:10.20944/preprints202403.0484.v1

Cubic and Gaussian. In addition, the authors carried out experiments using the Chi-Square test to select the 250 most representative features in the logits.

Some optimisers, global pooling techniques and losses were used for training. Among the optimisers available in the literature for training machine learning models, Stochastic Descent Gradient (SGD) and Adam [72] were used in the selected studies. In addition to SGD, it was possible to use an adaptive learning rate scaler or a momentum technique to improve the training process.

All the image-based studies that have used a global pooling method have opted for Global Average Pooling, with the aim of reducing the CNN activation maps before classification. The losses used were Cross Entropy loss (CE) and Focal Loss (FL) [73]. Focal Loss is a modification of the CE loss that reduces the weight of easy examples and thus concentrates training on difficult cases. It was first used in object detection studies, due to its huge imbalance between detected bins of "objects" and "non-objects", however Mukhoti et al. [74] concluded that it can also be used to deal with calibration errors of multi-class classification models, in the sense that the probability values they associate with the labels of the classes they predict overestimate the probabilities of those labels being correct in the real world. Carneiro et al. [38] used Focal Loss to mitigate the imbalance in the dataset used.

3.4. Evaluation

Aiming to quantitatively evaluate trained models, accuracy is the most used metric, followed by the F1 Score. Some studies also use precision, recall, Area-Under-The-Curve (AUC), specificity, or the Matthews correlation coefficient (MCC).

On the other hand, as in other areas of research, some studies use Explainable Artificial Intelligence (XAI) to qualitatively evaluate their models. XAI is a set of processes and methods aimed at enabling humans to understand, adequately trust and effectively manage the emerging generation of artificially intelligent models [75]. The techniques employed by the selected studies are model-agnostic for post-hoc explainability,which means that no modification to the architecture was necessary in order to apply them.

Nasiri et al. [41] and Pereira et al.[46] extracted the filters learnt by their models. In addition, Nasiri et al. [41] also produced Saliency Maps. Carneiro et al. [37] and Liu et al. [39] used Grad-CAM [76] to obtain heatmaps focused on the pixel's contribution to a specific class. Carneiro et al. [32] also used Grad-CAM to evaluate models, but instead of analysing the heatmaps generated, they used them to calculate the classification similarity between pairs of trained models. The authors calculated the heatmaps for the test subset for the models and calculated the cosine similarity between these heatmaps for the pairs of models. The authors concluded that, among the data augmentation approaches used, static geometric transformations generate representations more similar to RandAugment than to CutMix.

Carneiro et al. [38] used Local Interpretable Model-Agnostic Explanations (LIME) [77] for the same purpose. Furthermore, Carneiro et al. [35] extracted attention maps from ViT and checked the impact of sample rotation using them.

To generate saliency maps, Nasiri et al. [41] began by calculating the derivative of a class score function, which can be approximated by a first-order Taylor expansion. The elements of the calculated derivative were then rearranged. Grad-CAM is a technique proposed by Selvaraju et al. [76] that aims to explain how a model concludes that an image belongs to a certain class. The idea is to use the gradient of the score of the predicted class in relation to the activation maps of a selected convolutional layer. The selection of the convolutional layer is arbitrary. As a result, heat maps are obtained containing the regions that contribute positively to image classification. According to the authors, obtaining explanations of the predictions using Grad-CAM makes it possible to increase human confidence in the model and, at the same time, to understand classification errors. Like Grad-CAM, LIME [77] is an explainability approach used to explain individual machine learning model predictions for a specific class. Unlike Grad-CAM, it is not restricted to CNNs, so it is applicable to any machine learning classifier. The idea behind LIME is to train an explainable surrogate model with a

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 March 2024 doi:10.20944/preprints202403.0484.v1

new dataset composed of perturbed samples (e.g. hiding parts of the image) derived from the target data, so that it becomes a good approximation of the original model locally (in the neighbourhood of the target data). Then, from the surrogate interpretative model, it is possible to obtain the regions that have contributed to the classification, both positively and negatively.

4. Discussion and Future Directions

Considering that the studies presented in this research represent the current state of the art on the use of DL-based models to identify grapevine varieties, it can be concluded that there is still much to be done. Our idea here is to provide a discussion of the techniques used and a guide for future work.

4.1. Looking into the Grapevine Varieties Identification Problem

Before starting to discuss the solutions found in the literature, it is important to define the characteristics that the classification of grapevine varieties using Deep Learning encompasses.

- Grapevines are seasonal plants. This means that there are periods when the plants will have leaves, and others when they won't. This feature has a direct impact on the preparation of the dataset, which ideally should cover different phases of leaf growth. In addition, this feature limits the use of fruits in identification, as they take longer to grow than leaves;
- The presence of some grape varieties (e.g. Syrah, Chardonnay) is more common than others (e.g. Alvarinho), so the datasets are naturally unbalanced and can be treated as a *long-tailed data distribution classification* (some classes represent the majority of the data, while most classes are under-represented [78]);
- The classification of varieties within a species has a high inter-class similarity and high intra-class variations, placing the task in the *fine-grained recognition problems* family;
- There will be a high presence of unrelated information in the images acquired in the field, which could contribute to classification errors;
- There is a large amount of publicly available leaf and fruit images that are not annotated for variety identification;

4.2. Datasets

Analysing the datasets used, one can see that they are composed of few varieties and are mostly based on leaves. The largest dataset used is made up of images of 26 different grape varieties [31], so there is a big difference between the number of existing varieties and the number of annotated data available. In a practical scenario, taking Portugal as an example, 285 grape varieties are permitted nationwide. In the Douro Demarcated Region (DDR), the region responsible for the "Douro" and "Porto" wine denominations in Portugal, 115 grape varieties are allowed. If a classifier is created to improve production regulations in this region, all 115 grape varieties should be covered. Therefore, a classifier intended for practical application in the DDR should be trained with more grape varieties. The widespread use of leaf images in studies is probably due to the fact that fruit takes longer to grow.

Another problem related to datasets is the acquisition period. Despite the fact that Carneiro et al. [32] and Carneiro et al. [33] acquired data over two complete seasons, the other studies that specified the acquisition period did so in short order. Considering that grapes are plants with large seasonal changes, acquiring data for the entire season is mandatory to improve the usability of the classifier, since this will directly affect its generalisation capacity and practical applications. These facts emphasise the need for an open-source annotated datasets with more varieties, acquired over a long period.

Most of the studies identified used simple data augmentation strategies to increase the number of samples in the training stage. Carneiro et al. [32] evaluated the use of CutMix, RandAugmentation and static geometric augmentations and concluded that geometric transformations remain the best approach for augmenting data in varietal identification. On the other hand, other alternative methods can still be employed to increase the number of samples with higher quality, such as generating

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 March 2024 doi:10.20944/preprints202403.0484.v1

synthetic data with General Adversarial Networks, AutoAugment, MixUp, or increasing the feature space.

It is widely discussed in computer vision literature that synthetic data can be used to increase the number of samples in datasets, improving classification results. In the context of vine identification, Generative Adversarial Networks (GANs) can be used to produce images of vines. GANs were proposed by Goodfellow et al. [63] and their general operation consists of training two different models: a generative one, which captures the distribution of the data, and a discriminative one, which estimates the probability of a sample being from a training set or originating from the generative model. The aim of training is to make the generator so good at generating samples that the discriminator can't see any differences between them and the samples in the training set. Doğan et al. [30] used GANs in their study, however, they were used to improve the resolution of the images, rather than to generate entirely new samples.

Although it has achieved good performance in other fields of research, synthesising data using GAN should be used with attention. According to the results of some studies [79–81], there is a mismatch between the data generated by GAN and reality, which can lead to increased misclassification by models trained with synthetic data. To the best of our knowledge, there are no studies that have used GANs to augment grapevine variety identification datasets, however, in the plant context, there are a few that have aimed to identify diseases [82–86].

AutoAugment, proposed by Cubuk et al. [87], can also be used to deal with the lack of data. This approach consists of searching for the best policy to do data augmentation in an image. The policy consists of sub-policies; these sub-policies can be translation, rotation and shear, and the probability and magnitude of each. A search algorithm is then used to find the best policy so that the model produces the best result on a target dataset. In addition, searched policies for big agricultural datasets, e.g. iNaturalist, can be transferred to small data sets. MixUp [88] can also be used to generate new images by combining images and their labels in pairs. According to the authors, the combination regularises the models, improving their ability to generalise.

Another strategy consists of inserting noise, interpolating or extrapolating the learnt feature space to modify features already generated by a model [89]. Chu et al. [90] presented a technique for augmenting the data in the feature space in order to deal with the problem of high data imbalance present in datasets. The authors inserted a unit of attention with the help of the class activation map to obtain class-specific and class-generic features for each class and then, for the class with few samples, mixed the class-specific features with class-generic features from other classes.

Furthermore, most of the studies were based on data acquired by handheld devices such as cameras and smartphones, and no studies were found that used remote sensing technology, for example images from unmanned aerial vehicles (UAVs) or satellites. Similarly, all the studies that used images as input centred on RGB images, so there is space for the exploration of multispectral and hyperspectral images.

4.3. Pre-Processing

Segmentation methods based on classical image processing and DL-based architectures are applied in the pre-processing steps. The results obtained by Carneiro et al. [38], Carneiro et al. [36] and Ferentinos [91] show that leaf classification using images acquired in the field can be prone to interference from unrelated information. However, the results of Carneiro et al. [35] and Carneiro et al. [33] showed that the application of DL-based segmentation architectures (U-Net and SegNet) did not improve the results in terms of metrics. Indeed, in Carneiro et al. [35] performance decreased by 0.1 percentage points. However, in both studies the dataset used to train the models is small or the annotation was rough, and the results obtained in other studies indicate that removing the background can lead the models to faster convergence and better results in classifying leaves [46,92,93], suggesting that more research can be conducted in this direction.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 March 2024 doi:10.20944/preprints202403.0484.v1

4.4. Architectures and Training

In architectural terms, CNN and Transformer architectures pre-trained on ImageNet were used, individually and also through feature fusion. The use of such architectures implies the application of various techniques in caste classification: residual connections (ResNet, Xception, ViT), Depth Separable Convolutions (Xception, MobileNetV2), Dense Blocks (DenseNet) and Linear Bottlenecks (MobileNetV2), self-attention mechanism (ViT) are examples. These techniques have led to major performance improvements in various image classifications over time.

Focusing on the studies by Pereira et al. [46] and Nasiri et al. [41], they used the pre-trained AlexNet and VGG-16 models, respectively. Despite good results in the past, ImageNet classification results show that there is a high probability of improving results if more recent pre-trained CNN architectures are used, as in Carneiro et al. [33], or Magalhães et al. [31]. In addition, there are newer CNN architectures that remain untested, for example MobileNetV3 [94], larger members of the EfficientNet family [95,96], ConvNext [97,98].

Although CNNs represent the state of the art in most computer vision tasks, they do have some limitations. Sabour et al. [99], in their introduction to Capsule Networks (CapsNet), cited the inability of CNNs to recognise pose, texture and deformation of an image or its parts, and the lack of use of spatial information. These characteristics can be very useful in images acquired in the field, since leaves are not always placed in the same position, or at the same angle, they are subject to oscillations due to wind, and there is a large amount of unrelated information. According to Patrick et al. [100] the invariance of the characteristics learnt by CNNs comes from the pooling operation, which leads the model to lose information. CapsNets have been proposed to solve these problems and, in this context, can be applied to identifying grape varieties. The authors proposed replacing max-pooling with a "routing-by-agreement" process and convolutional filters with Capsules. Unlike a convolutional filter, a capsule is a group of neurons that learn a vector of features (e.g. angle, scale, pose). Capsules make CapsNets utilise space when extracting features and routing by agreement avoids information loss. In addition, Andrushia et al. [101] successfully applied CapsNet to the detection of diseases in grapevine leaves and concluded that this model is able to extract more useful features than CNNs due to its ability to map hierarchical pose relationships.

Carneiro et al. [35] was the only study to explore transformers in the context of identifying grapevine varieties. Dosovitskiy et al. [70] introduced the concept of self-attention in image classification with ViT models, with the aim of repeating the good results obtained by Transformer Networks in natural language processing. Raghu et al. [102] compared the representations generated by ViTs and CNNs. As a result, they were able to show that ViTs make better use of global information, strongly propagate information between layers and preserve more spatial information. However, these models have to be trained with a large amount of data to maintain similar performance or outperform CNNs in computer vision tasks. Thus, the fine-tuning strategy becomes mandatory to apply it to grape variety identification, since it is a task based on small data sets [103]. More details on ViTs can be found in Khan et al. [104]. In addition, new attention-based architectures can also be evaluated. Among them, the Swin Transform [105] was used for the detection of diseases in leaf datasets [106–108] and the results obtained indicate that it is a good candidate for the identification of grape varieties, since the data is similar.

An important factor in the results is that all the studies that used pre-trained architectures explored the transfer-learning plus fine-tuning configuration, using the pre-trained weights to initialise the models. Recent studies [109–111] have pointed out that, despite the excellent results brought by this arrangement, it introduces two inconsistencies in training in the agricultural context: *domain shift* and *supervision collapse*. The domain shift refers to the fact that the classification of natural scenes is very different from the classification of agricultural images, while the supervision collapse may occur because the pre-training was carried out using a fixed number of labels, and the model tends to concentrate on the information relevant to mapping the input data onto these labels, then discarding the information relevant to the classification of agricultural images. Another impediment caused by

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 March 2024 doi:10.20944/preprints202403.0484.v1

using pre-trained weights from generalised datasets is the fixed type of inputs. Since these models were trained with RGB images, multispectral and hyperspectral images cannot be used as input for the applications. Other alternatives to overcome transfer learning are to replace supervised training with different learning strategies in the agricultural domain, such as unsupervised learning [112], semi-supervised learning [113] or self-supervised learning [114]. As mentioned above, despite the lack of annotated data, there are many images of the context of grape cultivars publicly available on the web (e.g. iNaturalist collection [1]), so these approaches can be used to pre-train DL-large models using the unlabelled data.

Another factor to pay attention to is the natural existence of tailed classes in datasets, since this task can treated as a long-tailed data distribution classification. Imbalance can be problematic because the model tends to overfit the classes with more samples [115]. In this context, few-shot learning approaches [116] can be applied to minimise the lack of samples for some classes. Carneiro et al. [38] tested the use of Focal Loss to deal with unbalance in the dataset, even if there were no tail classes in the dataset (between 60 and 75 images per class), but other balanced losses [78,117] can still be tested. Cui et al. [78] introduced the Softmax Class-Balanced cross entropy loss, a modification of the cross entropy loss that introduces a weighting factor inversely proportional to the effective number of samples in a class. Similarly, Park et al. [117] created a loss function using an influence measure to identify how each sample affects biased decisions that cause the model to overfit. They assigned weights to each sample accordingly.

Furthermore, no study has analysed the classification of grape varieties as a fine-grained recognition problem. This family of classifiers aims to recover and recognise images belonging to multiple subordinate categories of a supercategory [118]. Among the techniques that can be applied are Bilinear CNNs [119,120] and Multi-Objective Matrix Normalisation [121]. The idea behind these techniques is to insert coded high-order statistics into the final features of the models. Both models use the covariance matrix-based representation with the aim of improving results for fine-grained tasks. Other tools that can be explored are the loss functions built specifically for fine-grained recognition tasks [122–124].

4.5. Evaluation

Grad-CAM, LIME, visualization of learned features, attention maps, and saliency maps were the XAI techniques employed by some of the selected studies. These approaches allowed to understand what the models are seeing when they make decisions. This understanding is very important when it comes to practical applications. Taking DDR as an example, if a classifier based on DL were to be used to help regulate production, it should be ensured that the decisions are using information related to the vine, for example, leaf, stem and fruit. In addition, these techniques can be used to check the impact of pre-processing steps, e.g. Liu et al. [39]. Thus, generating explanations for predictions increases human confidence in the classification process. Given that most of the selected studies are centred on images, the rest of the analysis focused on this type of data.

There are problems with Grad-CAM and LIME. The usability of Grad-CAM is restricted to CNNs, and it is mandatory to choose a layer for the calculations. There is a consensus in the literature that more top layers are chosen because they extract features at a higher level. Furthermore, Subramanya et al. [125] showed with adversarial attacks that this method cannot necessarily show the reason for misclassification.

On the other hand, LIME can be used in any model that makes a prediction, since it is based on surrogate models. However, some studies have tested the stability of the model when generating explanations. Stability is referred to as the ability of the method to generate the same explanation when it is applied several times to the same sample. When applying LIME to explain image-based

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

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 March 2024 doi:10.20944/preprints202403.0484.v1

classifiers, small perturbations can generate large differences in explanations, but if fixed parameters are set, the method can be stable, even if the surrogate model is not faithful [126–128].

Moreover, more recent techniques such as Guided Integrated Gradients [129], XRAI [130] and SmoothGrad [131] can also be used to obtain explanations for the classification.

4.6. Comparison with other Subfields of Precision Viticulture

Precision viticulture includes several tasks that can be improved with the use of computer vision. Flower counting, grape detection, berry counting, yield estimation, disease identification and variety identification are examples of such tasks. Mohimont et al. [25] analysed recent precision viticulture studies that applied DL techniques to all of the above yield-related tasks, excluding variety identification and disease detection. The number of studies the authors found for each task individually does not differ from the number of studies found in the present study. Furthermore, the literature pays more attention to disease detection. To the best of our knowledge, no study has analysed methods for identifying grapevine diseases. However, a simple search of the literature reveals studies that have explored self-supervised learning [132], unsupervised learning [133], the use of multispectral data acquired with UAVs [134], and other more recent techniques [135]. This behaviour is repeated in other plant species, since diseases affect food safety, thus there are more studies aimed at identifying diseases, although, at the same time, this implies that research into methods for the automatic identification of grapevine varieties through deep learning is still moving at a small pace and has a lot to explore.

5. Conclusion

This study presents a review of studies aimed at identifying grapevine varieties using Deep Learning approaches. Eighteen articles were analysed, taking into account the inclusion criteria. The results indicate that the automatic identification of grapevines using DL models is still taking small steps, with enormous room for improvement. Most of the studies focused on identification using images, applying transfer learning and fine-tuning to pre-trained models on large datasets for general scene classification. From the eighteen studies selected in this review, it can be concluded that:

- (RQ1) How Deep-Learning techniques have been used for automatic identification of grapevine varieties? Pre-trained architectures for image classification are the way in which DL has been most widely applied to the identification of grapevine varieties.
- (RQ2) Which are the best DL-based models for Automatic Grapevine Verities Identification? Since the datasets are small, tuning and transfer learning techniques have been used in all image-based classification works. The most commonly used pre-trained architecture was Xception, accompanied by cross entropy loss and static geometric augmentation strategies. In the Evaluation stage, in addition to the popular metrics, Grad-CAM was the most commonly used XAI method.
- (RQ3) What are the main challenges and future development trends in identifying grape varieties using DL-based models? There is still room to evaluate the removal of the complex background; the generation of new samples through GANs can be explored; new architectures can be tested, e.g. Swin Transformers, Capsules Networks, or Bilinear CNNs; different losses still can be explored [78,117,122–124]; and other XAI approaches can also be employed, e.g. Guided Integrated Gradients, XRAI, and SmoothGrads.

Author Contributions: Conceptualization, G.A.C., A.C., and J.S.; methodology, G.A.C.; validation, J.S. and A.C.; formal analysis, G.A.C., A.C., and J.S.; investigation, G.A.C.; resources, G.A.C., A.C., and J.S.; data curation, G.A.C.; writing—original draft preparation, G.A.C and J.S; writing—review and editing, G.A.C, A.C., and J.S.; visualization, G.A.C.; supervision, J.S and A.C..; project administration, J.S.; funding acquisition, J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research activity was supported by –"DATI - Digital Agriculture Technologies for Irrigation efficiency" project. PRIMA – Partnership for Research and Innovation in the Mediterranean Area, (Research and Innovation activities), financed by the states participating in the PRIMA partnership and by the European Union, through Horizon 2020 and by FCT - Portuguese Foundation for Science and Technology, under the project UIDB/04033/2020.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 March 2024 doi:10.20944/preprints202403.0484.v1

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: The data presented in this study are available on request from the corresponding author

Conflicts of Interest: The authors declare that they have no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

Acc	Accuracy
AUC	Area-Under-The-Curve
CapsNet	Capsule Networks
CE	Cross Entropy
CNN	Convolutional Neural Network
DNA	Deoxyribonucleic acid
DDR	Douro Demarcated Region
DL	Deep Learning
ESRGAN	Enhanced Super Resolution Generative Adversarial Networks
FL	Focal Loss
GAN	Generative Adversarial Network
Grad-CAM	Gradient-weighted Class Activation Mapping
HSI	Hue, Saturation and Intensity
ICA	Independent Component Analysis
MSC	Multiplicative Scatter Correction
MCC	Matthews correlation coefficient
LIME	Local Interpretable Model-Agnostic Explanations
RGB	Red, Green, Blue
RQ	Research Questions
SG	Savitzky-Golay
SNV	Standard Normal Variant
SVM	Support Vector Machines
SGD	Stochastic Descent Gradient
UAV	Unmanned Aerial Vehicles
ViT	Vision Transformer
XAI	Explainable Artificial Intelligence

References

1. Eyduran, S.P.; Akin, M.; Ercisli, S.; Eyduran, E.; Maghradze, D. Sugars, organic acids, and phenolic compounds of ancient grape cultivars (Vitis vinifera L.) from Igdir province of Eastern Turkey. *Biological Research* **2015**, *48*, 2. https://doi.org/10.1186/0717-6287-48-2.
2. Nascimento, R.; Maia, M.; Ferreira, A.E.N.; Silva, A.B.; Freire, A.P.; Cordeiro, C.; Silva, M.S.; Figueiredo, A. Early stage metabolic events associated with the establishment of *Vitis vinifera – Plasmopara viticola* compatible interaction. *Plant Physiology and Biochemistry* **2019**, *137*, 1–13. https://doi.org/10.1016/j.plaphy.2019.01.026.
3. Cunha, J.; Santos, M.T.; Carneiro, L.C.; Fevereiro, P.; Eiras-Dias, J.E. Portuguese traditional grapevine cultivars and wild vines (Vitis vinifera L.) share morphological and genetic traits. *Genetic Resources and Crop Evolution* **2009**, *56*, 975–989. https://doi.org/10.1007/S10722-009-9416-4.
4. Schneider, A.; Carra, A.; Akkak, A.; This, P.; Laucou, V.; Botta, R. Verifying synonymies between grape cultivars from France and Northwestern Italy using molecular markers. *VITIS - Journal of Grapevine Research* **2001**, *40*, 197–197. Number: 4, https://doi.org/10.5073/vitis.2001.40.197-203.
5. Lacombe, T. Contribution à l'étude de l'histoire évolutive de la vigne cultivée (Vitis vinifera L.) par l'analyse de la diversité génétique neutre et de gènes d'intérêt. Theses, Institut National d'Etudes Supérieures Agronomiques de Montpellier, 2012. MR AGAP - équipe DAAV - Diversité, adaptation et amélioration de la vigne.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 March 2024 doi:10.20944/preprints202403.0484.v1

6. Distribution of the world's grapevine varieties. Technical Report 979-10-91799-89-8, International Organisation of Vine and Wine.
7. Koklu, M.; Unlersen, M.F.; Ozkan, I.A.; Aslan, M.F.; Sabanci, K. A CNN-SVM study based on selected deep features for grapevine leaves classification. *Measurement* **2022**, *188*, 110425. https://doi.org/10.1016/J.MEASUREMENT.2021.110425.
8. Moncayo, S.; Rosales, J.D.; Izquierdo-Hornillos, R.; Anzano, J.; Caceres, J.O. Classification of red wine based on its protected designation of origin (PDO) using Laser-induced Breakdown Spectroscopy (LIBS). *Talanta* **2016**, *158*, 185–191. https://doi.org/https://doi.org/10.1016/j.talanta.2016.05.059.
9. Giacosa, E. Wine Consumption in a Certain Territory. Which Factors May Have Impact on It? In *Production and Management of Beverages*; Woodhead Publishing, 2019; pp. 361–380. https://doi.org/10.1016/b978-0-12-815260-7.00012-2.
10. of Vine, T.I.O.; Wine. STATE OF THE WORLD VITIVINICULTURAL SECTOR IN 2020 **2020**.
11. Jones, G.; Alves, F. Impact of climate change on wine production: a global overview and regional assessment in the Douro Valley of Portugal. *Int. J. of Global Warming* **2012**, *4*, 383–406. https://doi.org/10.1504/IJGW.2012.049448.
12. Chitwood, D.H.; Ranjan, A.; Martinez, C.C.; Headland, L.R.; Thiem, T.; Kumar, R.; Covington, M.F.; Hatcher, T.; Naylor, D.T.; Zimmerman, S.; et al. A Modern Ampelography: A Genetic Basis for Leaf Shape and Venation Patterning in Grape. *Plant Physiology* **2014**, *164*, 259–272. https://doi.org/10.1104/PP.113.229708.
13. Galet, P. *Précis d'ampélographie pratique*; 1952.
14. Garcia-Muñoz, S.; Muñoz-Organero, G.; de Andrés, M.; Cabello, F. Ampelography - An old technique with future uses: the case of minor varieties of Vitis vinifera L. from the Balearic Islands. *Journal International des Sciences de la Vigne et du Vin* **2011**, *45*, 125–137. https://doi.org/10.20870/oeno-one.2011.45.3.1497.
15. Pavek, D.S.; Lamboy, W.F.; Garvey, E.J. Selecting in situ conservation sites for grape genetic resources in the USA. *Genetic Resources and Crop Evolution 2003 50:2* **2003**, *50*, 165–173. https://doi.org/10.1023/A:1022947605916.
16. Tassie, L. Vine identification – knowing what you have. *The Grape and Wine Research and Development Corporation (GWRDC) Innovators network* **2010**.
17. This, P.; Jung, A.; Boccacci, P.; Borrego, J.; Botta, R.; Costantini, L.; Crespan, M.; Dangl, G.S.; Eisenheld, C.; Ferreira-Monteiro, F.; et al. Development of a standard set of microsatellite reference alleles for identification of grape cultivars. *Theoretical and Applied Genetics* **2004**, *109*, 1448–1458. https://doi.org/10.1007/s00122-004-1760-3.
18. Calo, A.; Tomasi, D.; Crespan, M.; Costacurta, A. Relationship between environmental factors and the dynamics of growth and composition of the grapevine. *Acta Horticulturae* **1996**, *427*, 217–231. https://doi.org/10.17660/actahortic.1996.427.27.
19. Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Garcia-Rodriguez, J. A Review on Deep Learning Techniques Applied to Semantic Segmentation **2017**.
20. Gutiérrez, S.; Tardaguila, J.; Fernández-Novales, J.; Diago, M.P. Support Vector Machine and Artificial Neural Network Models for the Classification of Grapevine Varieties Using a Portable NIR Spectrophotometer. *PLOS ONE* **2015**, *10*, e0143197. Publisher: Public Library of Science, https://doi.org/10.1371/journal.pone.0143197.
21. Karakizi, C.; Oikonomou, M.; Karantzalos, K. Vineyard Detection and Vine Variety Discrimination from Very High Resolution Satellite Data. *Remote Sensing* **2016**, *8*, 235. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute, https://doi.org/10.3390/rs8030235.
22. Diago, M.P.; Fernandes, A.M.; Millan, B.; Tardaguila, J.; Melo-Pinto, P. Identification of grapevine varieties using leaf spectroscopy and partial least squares. *Computers and Electronics in Agriculture* **2013**, *99*, 7–13. https://doi.org/10.1016/j.compag.2013.08.021.
23. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. Curran Associates, Inc., 2012, Vol. 25.
24. Chen, Y.; Huang, Y.; Zhang, Z.; Wang, Z.; Liu, B.; Liu, C.; Huang, C.; Dong, S.; Pu, X.; Wan, F.; et al. Plant image recognition with deep learning: A review. *Computers and Electronics in Agriculture* **2023**, *212*, 108072. https://doi.org/10.1016/j.compag.2023.108072.
25. Mohimont, L.; Alin, F.; Rondeau, M.; Gaveau, N.; Steffenel, L.A. Computer Vision and Deep Learning for Precision Viticulture. *Agronomy* **2022**, *12*, 2463. Number: 10 Publisher: Multidisciplinary Digital Publishing Institute, https://doi.org/10.3390/agronomy12102463.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 March 2024 doi:10.20944/preprints202403.0484.v1

26. Ferro, M.V.; Catania, P. Technologies and Innovative Methods for Precision Viticulture: A Comprehensive Review. *Horticulturae* **2023**, *9*, 399. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute, https://doi.org/10.3390/horticulturae9030399.
27. Chai, J.; Zeng, H.; Li, A.; Ngai, E.W.T. Deep learning in computer vision: A critical review of emerging techniques and application scenarios. *Machine Learning with Applications* **2021**, *6*, 100134. https://doi.org/10.1016/j.mlwa.2021.100134.
28. Khan, S.; Naseer, M.; Hayat, M.; Zamir, S.W.; Khan, F.S.; Shah, M. Transformers in Vision: A Survey. *ACM Computing Surveys* **2022**, *54*, 1–41. arXiv:2101.01169 [cs], https://doi.org/10.1145/3505244.
29. Dhanya, V.G.; Subeesh, A.; Kushwaha, N.L.; Vishwakarma, D.K.; Nagesh Kumar, T.; Ritika, G.; Singh, A.N. Deep learning based computer vision approaches for smart agricultural applications. *Artificial Intelligence in Agriculture* **2022**, *6*, 211–229. https://doi.org/10.1016/j.aiia.2022.09.007.
30. Doğan, G.; Imak, A.; Ergen, B.; Sengur, A. A new hybrid approach for grapevine leaves recognition based on ESRGAN data augmentation and GASVM feature selection. *Neural Computing and Applications* **2024**. https://doi.org/10.1007/s00521-024-09488-2.
31. Magalhaes, S.C.; Castro, L.; Rodrigues, L.; Padilha, T.C.; Carvalho, F.D.; Santos, F.N.D.; Pinho, T.; Moreira, G.; Cunha, J.; Cunha, M.; et al. Toward Grapevine Digital Ampelometry Through Vision Deep Learning Models. *IEEE Sensors Journal* **2023**, *23*, 10132–10139. https://doi.org/10.1109/JSEN.2023.3261544.
32. Carneiro, G.; Neto, A.; Teixeira, A.; Cunha, A.; Sousa, J. Evaluating Data Augmentation for Grapevine Varieties Identification **2023**. pp. 3566–3569. https://doi.org/10.1109/IGARSS52108.2023.10283128.
33. Carneiro, G.A.; Texeira, A.; Morais, R.; Sousa, J.J.; Cunha, A. Can the Segmentation Improve the Grape Varieties' Identification Through Images Acquired On-Field? *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2023**, *14116 LNAI*, 351–363. https://doi.org/10.1007/978-3-031-49011-8_28/COVER.
34. Gupta, R.; Gill, K.S. Grapevine Augmentation and Classification using Enhanced EfficientNetB5 Model. *2023 IEEE Renewable Energy and Sustainable E-Mobility Conference, RESEM 2023* **2023**. https://doi.org/10.1109/RESEM57584.2023.10236406.
35. Carneiro, G.A.; Padua, L.; Peres, E.; Morais, R.; Sousa, J.J.; Cunha, A. Segmentation as a Preprocessing Tool for Automatic Grapevine Classification. *International Geoscience and Remote Sensing Symposium (IGARSS)* **2022**, *2022-July*, 6053–6056. https://doi.org/10.1109/IGARSS46834.2022.9884946.
36. Carneiro, G.S.; Ferreira, A.; Morais, R.; Sousa, J.J.; Cunha, A. Analyzing the Fine Tuning's impact in Grapevine Classification. *Procedia Computer Science* **2022**, *196*, 364–370. https://doi.org/10.1016/J.PROCS.2021.12.025.
37. Carneiro, G.A.; Pádua, L.; Peres, E.; Morais, R.; Sousa, J.J.; Cunha, A. Grapevine Varieties Identification Using Vision Transformers. In Proceedings of the IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium, 2022, pp. 5866–5869. https://doi.org/10.1109/IGARSS46834.2022.9883286.
38. Carneiro, G.; Padua, L.; Sousa, J.J.; Peres, E.; Morais, R.; Cunha, A. Grapevine Variety Identification Through Grapevine Leaf Images Acquired in Natural Environment **2021**. pp. 7055–7058. https://doi.org/10.1109/IGARSS47720.2021.9555141.
39. Liu, Y.; Su, J.; Shen, L.; Lu, N.; Fang, Y.; Liu, F.; Song, Y.; Su, B. Development of a mobile application for identification of grapevine (Vitis vinifera l.) cultivars via deep learning. *International Journal of Agricultural and Biological Engineering* **2021**, *14*, 172–179. https://doi.org/10.25165/j.ijabe.20211405.6593.
40. Škrabánek, P.; Doležel, P.; Matoušek, R.; Junek, P. RGB Images Driven Recognition of Grapevine Varieties. Springer, 9 2021, Vol. 1268 AISC, *Advances in Intelligent Systems and Computing*, pp. 216–225. https://doi.org/10.1007/978-3-030-57802-2_21.
41. Nasiri, A.; Taheri-Garavand, A.; Fanourakis, D.; Zhang, Y.D.; Nikoloudakis, N. Automated grapevine cultivar identification via leaf imaging and deep convolutional neural networks: A proof-of-concept study employing primary iranian varieties. *Plants* **2021**, *10*. https://doi.org/10.3390/plants10081628.
42. Peng, Y.; Zhao, S.; Liu, J.; Peng, Y..; Zhao, S..; Liu, J. Fused Deep Features-Based Grape Varieties Identification Using Support Vector Machine. *Agriculture 2021, Vol. 11, Page 869* **2021**, *11*, 869. https://doi.org/10.3390/AGRICULTURE11090869.
43. Franczyk, B.; Hernes, M.; Kozierkiewicz, A.; Kozina, A.; Pietranik, M.; Roemer, I.; Schieck, M. Deep learning for grape variety recognition. *Procedia Computer Science* **2020**, *176*, 1211–1220. https://doi.org/10.1016/J.PROCS.2020.09.117.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 March 2024 doi:10.20944/preprints202403.0484.v1

44. Fernandes, A.M.; Utkin, A.B.; Eiras-Dias, J.; Cunha, J.; Silvestre, J.; Melo-Pinto, P. Grapevine variety identification using "Big Data" collected with miniaturized spectrometer combined with support vector machines and convolutional neural networks. *Computers and Electronics in Agriculture* **2019**, *163*, 104855. https://doi.org/10.1016/j.compag.2019.104855.
45. Adão, T.; Pinho, T.M.; Ferreira, A.; Sousa, A.; Pádua, L.; Sousa, J.; Sousa, J.J.; Peres, E.; Morais, R. Digital ampelographer: A CNN based preliminary approach, Cham, 2019; Vol. 11804 LNAI, pp. 258–271. https://doi.org/10.1007/978-3-030-30241-2_23.
46. Pereira, C.S.; Morais, R.; Reis, M.J.C.S. Deep learning techniques for grape plant species identification in natural images. *Sensors (Switzerland)* **2019**, *19*, 4850. https://doi.org/10.3390/s19224850.
47. Santos, T.; de Souza, L.; Andreza, d.S.; Avila, S. Embrapa Wine Grape Instance Segmentation Dataset – Embrapa WGISD **2019**. https://doi.org/10.5281/ZENODO.3361736.
48. Al-khazraji, L.R.; Mohammed, M.A.; Abd, D.H.; Khan, W.; Khan, B.; Hussain, A.J. Image dataset of important grape varieties in the commercial and consumer market. *Data in Brief* **2023**, *47*, 108906. https://doi.org/10.1016/J.DIB.2023.108906.
49. Vlah, M. Grapevine Leaves, 2021. https://doi.org/10.34740/KAGGLE/DS/1248678.
50. Sozzi, M.; Cantalamessa, S.; Cogato, A.; Kayad, A.; Marinello, F. wGrapeUNIPD-DL: An open dataset for white grape bunch detection. *Data in Brief* **2022**, *43*, 108466. https://doi.org/10.1016/J.DIB.2022.108466.
51. Seng, K.P.; Ang, L.M.; Schmidtke, L.M.; Rogiers, S.Y. Computer vision and machine learning for viticulture technology. *IEEE Access* **2018**, *6*, 67494–67510. https://doi.org/10.1109/ACCESS.2018.2875862.
52. Rodrigues, A. Um método filométrico de caracterização ampelográfica **1952**.
53. Hyvärinen, A.; Oja, E. A Fast Fixed-Point Algorithm for Independent Component Analysis. *Neural Computation* **1997**, *9*, 1483–1492. https://doi.org/10.1162/neco.1997.9.7.1483.
54. Canny, J. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **1986**, *PAMI-8*, 679–698. https://doi.org/10.1109/TPAMI.1986.4767851.
55. Zheng, X.; Wang, X. Leaf Vein Extraction Based on Gray-scale Morphology. *International Journal of Image, Graphics and Signal Processing* **2010**, *2*, 25–31. https://doi.org/10.5815/ijigsp.2010.02.04.
56. Pereira, C.S.; Morais, R.; Reis, M.J. Pixel-Based Leaf Segmentation from Natural Vineyard Images Using Color Model and Threshold Techniques. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2018**, *10882 LNCS*, 96–106. https://doi.org/10.1007/978-3-319-93000-8_12.
57. Du, Q.; Kopriva, I.; Szu, H. Independent-component analysis for hyperspectral remote sensing imagery classification. *Optical Engineering - OPT ENG* **2006**, *45*. https://doi.org/10.1117/1.2151172.
58. Vaseghi, S.; Jetelova, H. Principal and independent component analysis in image processing. *Proceeding of the 14th ACM International Conference on Mobile Computing and Networking* **2006**.
59. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics* **1979**, *9*, 62–66. https://doi.org/10.1109/TSMC.1979.4310076.
60. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *CoRR* **2015**, *abs/1505.0*.
61. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2017**, *39*, 2481–2495. arXiv: 1511.00561, https://doi.org/10.1109/TPAMI.2016.2644615.
62. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Loy, C.C.; Qiao, Y.; Tang, X. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks, 2018. arXiv:1809.00219 [cs], https://doi.org/10.48550/arXiv.1809.00219.
63. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. *Communications of the ACM* **2014**, *63*, 139–144. https://doi.org/10.1145/3422622.
64. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* **2014**.
65. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **2015**, *2016-December*, 770–778. https://doi.org/10.1109/CVPR.2016.90.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 March 2024 doi:10.20944/preprints202403.0484.v1

66. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* **2016**, *2017-January*, 2261–2269. https://doi.org/10.1109/CVPR.2017.243.
67. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. 2017, pp. 1800–1807. https://doi.org/10.1109/CVPR.2017.195.
68. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **2018**, pp. 4510–4520. https://doi.org/10.1109/CVPR.2018.00474.
69. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data 2021 8:1* **2021**, *8*, 1–74. https://doi.org/10.1186/S40537-021-00444-8.
70. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale **2020**.
71. Sun, S.Q.; Zeng, S.G.; Liu, Y.; Heng, P.A.; Xia, D.S. A new method of feature fusion and its application in image recognition. *Pattern Recognition* **2005**, *38*, 2437–2448. https://doi.org/10.1016/J.PATCOG.2004.12.013.
72. Kingma, D.P.; Ba, J.L. Adam: A method for stochastic optimization, 2015.
73. Lin, T.Y.; Goyal, P.; Girshick, R.B.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. *CoRR* **2017**, *abs/1708.0*.
74. Mukhoti, J.; Kulharia, V.; Sanyal, A.; Golodetz, S.; Torr, P.H.; Dokania, P.K. Calibrating deep neural networks using focal loss. Neural information processing systems foundation, 2 2020, Vol. 2020-Decem.
75. Barredo Arrieta, A.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; Garcia, S.; Gil-Lopez, S.; Molina, D.; Benjamins, R.; et al. Explainable Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* **2020**, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012.
76. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *International Journal of Computer Vision* **2016**, *128*, 336–359. https://doi.org/10.1007/s11263-019-01228-7.
77. Ribeiro, M.T.; Singh, S.; Guestrin, C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. Association for Computing Machinery, 2 2016, Vol. 13-17-Augu, pp. 97–101. https://doi.org/10.18653/v1/n16-3020.
78. Cui, Y.; Jia, M.; Lin, T.Y.; Song, Y.; Belongie, S. Class-Balanced Loss Based on Effective Number of Samples. 2019, pp. 9268–9277.
79. Barratt, S.; Sharma, R. A Note on the Inception Score **2018**. https://doi.org/10.48550/arxiv.1801.01973.
80. Ravuri, S.; Vinyals, O. Seeing is Not Necessarily Believing: Limitations of BigGANs for Data Augmentation. 2019, pp. 1–5.
81. Shmelkov, K.; Schmid, C.; Alahari, K. How good is my GAN? *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2018**, *11206 LNCS*, 218–234. https://doi.org/10.48550/arxiv.1807.09499.
82. Gomaa, A.A.; El-Latif, Y.M. Early Prediction of Plant Diseases using CNN and GANs. *International Journal of Advanced Computer Science and Applications* **2021**, *12*, 514–519. https://doi.org/10.14569/IJACSA.2021.0120563.
83. Nazki, H.; Lee, J.; Yoon, S.; Park, D.S. Image-to-Image Translation with GAN for Synthetic Data Augmentation in Plant Disease Datasets. *Smart Media Journal* **2019**, *8*, 46–57. https://doi.org/10.30693/SMJ.2019.8.2.46.
84. Talukdar, B. Handling of Class Imbalance for Plant Disease Classification with Variants of GANs. 2020, pp. 466–471. https://doi.org/10.1109/ICIIS51140.2020.9342728.
85. Yilma, G.; Belay, S.; Qin, Z.; Gedamu, K.; Ayalew, M. Plant Disease Classification Using Two Pathway Encoder GAN Data Generation. *2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing, ICCWAMTIP 2020* **2020**, pp. 67–72. https://doi.org/10.1109/ICCWAMTIP51612.2020.9317494.
86. Zeng, Q.; Ma, X.; Cheng, B.; Zhou, E.; Pang, W. GANS-based data augmentation for citrus disease severity detection using deep learning. *IEEE Access* **2020**, *8*, 172882–172891. https://doi.org/10.1109/ACCESS.2020.3025196.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 March 2024 doi:10.20944/preprints202403.0484.v1

87. Cubuk, E.D.; Zoph, B.; Mane, D.; Vasudevan, V.; Le, V.Q. AutoAugment: Learning Augmentation Policies from Data. *Cvpr 2019* **2018**, pp. 113–123. https://doi.org/10.48550/arxiv.1805.09501.
88. Zhang, H.; Cisse, M.; Dauphin, Y.N.; Lopez-Paz, D. mixup: Beyond Empirical Risk Minimization, 2018. arXiv:1710.09412 [cs, stat], https://doi.org/10.48550/arXiv.1710.09412.
89. DeVries, T.; Taylor, G.W. Dataset Augmentation in Feature Space, 2017, [arXiv:stat.ML/1702.05538].
90. Chu, P.; Bian, X.; Liu, S.; Ling, H. Feature Space Augmentation for Long-Tailed Data, 2020. arXiv:2008.03673 [cs], https://doi.org/10.48550/arXiv.2008.03673.
91. Ferentinos, K.P. Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture* **2018**, *145*, 311–318. https://doi.org/10.1016/J.COMPAG.2018.01.009.
92. Kc, K.; Yin, Z.; Li, D.; Wu, Z. Impacts of Background Removal on Convolutional Neural Networks for Plant Disease Classification In-Situ. *Agriculture* **2021**, *11*, 827. Number: 9 Publisher: Multidisciplinary Digital Publishing Institute, https://doi.org/10.3390/agriculture11090827.
93. Wu, Y.J.; Tsai, C.M.; Shih, F. Improving Leaf Classification Rate via Background Removal and ROI Extraction. *Journal of Image and Graphics* **2016**, *4*, 93–98. https://doi.org/10.18178/joig.4.2.93-98.
94. Howard, A.; Sandler, M.; Chen, B.; Wang, W.; Chen, L.C.; Tan, M.; Chu, G.; Vasudevan, V.; Zhu, Y.; Pang, R.; et al. Searching for mobileNetV3. Institute of Electrical and Electronics Engineers Inc., 5 2019, Vol. 2019-Octob, pp. 1314–1324. https://doi.org/10.1109/ICCV.2019.00140.
95. Tan, M.; Le, V.Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *36th International Conference on Machine Learning, ICML 2019* **2019**, *2019-June*, 10691–10700.
96. Tan, M.; Le, Q.V. EfficientNetV2: Smaller Models and Faster Training, 2021. arXiv:2104.00298 [cs], https://doi.org/10.48550/arXiv.2104.00298.
97. Liu, Z.; Mao, H.; Wu, C.Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A ConvNet for the 2020s, 2022. arXiv:2201.03545 [cs], https://doi.org/10.48550/arXiv.2201.03545.
98. Woo, S.; Debnath, S.; Hu, R.; Chen, X.; Liu, Z.; Kweon, I.S.; Xie, S. ConvNeXt V2: Co-designing and Scaling ConvNets with Masked Autoencoders, 2023. arXiv:2301.00808 [cs], https://doi.org/10.48550/arXiv.2301.00808.
99. Sabour, S.; Frosst, N.; Hinton, G.E. Dynamic Routing Between Capsules. *Advances in Neural Information Processing Systems* **2017**, *2017-December*, 3857–3867.
100. Kwabena Patrick, M.; Felix Adekoya, A.; Abra Mighty, A.; Edward, B.Y. Capsule Networks – A survey. *Journal of King Saud University - Computer and Information Sciences* **2022**, *34*, 1295–1310. https://doi.org/10.1016/j.jksuci.2019.09.014.
101. Andrushia, A.D.; Neebha, T.M.; Patricia, A.T.; Sagayam, K.M.; Pramanik, S. Capsule network-based disease classification for Vitis Vinifera leaves. *Neural Computing and Applications* **2024**, *36*, 757–772. https://doi.org/10.1007/s00521-023-09058-y.
102. Raghu, M.; Unterthiner, T.; Kornblith, S.; Zhang, C.; Dosovitskiy, A. Do Vision Transformers See Like Convolutional Neural Networks? **2021**.
103. Steiner, A.; Kolesnikov, A.; Zhai, X.; Wightman, R.; Uszkoreit, J.; Beyer, L. How to train your ViT? Data, Augmentation, and Regularization in Vision Transformers **2021**.
104. Khan, S.; Naseer, M.; Hayat, M.; Zamir, S.W.; Khan, F.S.; Shah, M. Transformers in Vision: A Survey. *ACM Computing Surveys* **2021**. https://doi.org/10.1145/3505244.
105. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows, 2021. arXiv:2103.14030 [cs], https://doi.org/10.48550/arXiv.2103.14030.
106. Guo, Y.; Lan, Y.; Chen, X. CST: Convolutional Swin Transformer for detecting the degree and types of plant diseases. *Computers and Electronics in Agriculture* **2022**, *202*, 107407. https://doi.org/10.1016/j.compag.2022.107407.
107. Chang, B.; Wang, Y.; Zhao, X.; Li, G.; Yuan, P. A general-purpose edge-feature guidance module to enhance vision transformers for plant disease identification. *Expert Systems with Applications* **2024**, *237*, 121638. https://doi.org/10.1016/j.eswa.2023.121638.
108. Wang, F.; Rao, Y.; Luo, Q.; Jin, X.; Jiang, Z.; Zhang, W.; Li, S. Practical cucumber leaf disease recognition using improved Swin Transformer and small sample size. *Computers and Electronics in Agriculture* **2022**, *199*, 107163. https://doi.org/10.1016/j.compag.2022.107163.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 March 2024 doi:10.20944/preprints202403.0484.v1

109. El-Nouby, A.; Izacard, G.; Touvron, H.; Laptev, I.; Jegou, H.; Grave, E. Are Large-scale Datasets Necessary for Self-Supervised Pre-training?, 2021. arXiv:2112.10740 [cs], https://doi.org/10.48550/arXiv.2112.10740.
110. Doersch, C.; Gupta, A.; Zisserman, A. CrossTransformers: spatially-aware few-shot transfer, 2021. arXiv:2007.11498 [cs], https://doi.org/10.48550/arXiv.2007.11498.
111. Lu, Y.; Young, S. A survey of public datasets for computer vision tasks in precision agriculture. *Computers and Electronics in Agriculture* **2020**, *178*, 105760. https://doi.org/10.1016/j.compag.2020.105760.
112. Nazki, H.; Yoon, S.; Fuentes, A.; Park, D.S. Unsupervised image translation using adversarial networks for improved plant disease recognition. *Computers and Electronics in Agriculture* **2020**, *168*, 105117. https://doi.org/10.1016/j.compag.2019.105117.
113. Homan, D.; du Preez, J.A. Automated feature-specific tree species identification from natural images using deep semi-supervised learning. *Ecological Informatics* **2021**, *66*, 101475. https://doi.org/10.1016/j.ecoinf.2021.101475.
114. Güldenring, R.; Nalpantidis, L. Self-supervised contrastive learning on agricultural images. *Computers and Electronics in Agriculture* **2021**, *191*. Type: Article, https://doi.org/10.1016/j.compag.2021.106510.
115. Buda, M.; Maki, A.; Mazurowski, M.A. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks* **2018**, *106*, 249–259. arXiv:1710.05381 [cs, stat], https://doi.org/10.1016/j.neunet.2018.07.011.
116. Argüeso, D.; Picon, A.; Irusta, U.; Medela, A.; San-Emeterio, M.G.; Bereciartua, A.; Alvarez-Gila, A. Few-Shot Learning approach for plant disease classification using images taken in the field. *Computers and Electronics in Agriculture* **2020**, *175*, 105542. https://doi.org/10.1016/j.compag.2020.105542.
117. Park, S.; Lim, J.; Jeon, Y.; Choi, J.Y. Influence-Balanced Loss for Imbalanced Visual Classification. 2021, pp. 735–744.
118. Wei, X.S.; Song, Y.Z.; Mac Aodha, O.; Wu, J.; Peng, Y.; Tang, J.; Yang, J.; Belongie, S. Fine-Grained Image Analysis with Deep Learning: A Survey, 2021. arXiv:2111.06119 [cs], https://doi.org/10.48550/arXiv.2111.06119.
119. Lin, T.Y.; RoyChowdhury, A.; Maji, S. Bilinear CNNs for Fine-grained Visual Recognition, 2017. arXiv:1504.07889 [cs], https://doi.org/10.48550/arXiv.1504.07889.
120. Gao, Y.; Beijbom, O.; Zhang, N.; Darrell, T. Compact Bilinear Pooling. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 317–326. ISSN: 1063-6919, https://doi.org/10.1109/CVPR.2016.41.
121. Min, S.; Yao, H.; Xie, H.; Zha, Z.J.; Zhang, Y. Multi-Objective Matrix Normalization for Fine-grained Visual Recognition. *IEEE Transactions on Image Processing* **2020**, *29*, 4996–5009. arXiv:2003.13272 [cs], https://doi.org/10.1109/TIP.2020.2977457.
122. Dubey, A.; Gupta, O.; Guo, P.; Raskar, R.; Farrell, R.; Naik, N. Pairwise Confusion for Fine-Grained Visual Classification, 2018. arXiv:1705.08016 [cs], https://doi.org/10.48550/arXiv.1705.08016.
123. Sun, G.; Cholakkal, H.; Khan, S.; Khan, F.S.; Shao, L. Fine-grained Recognition: Accounting for Subtle Differences between Similar Classes, 2019. arXiv:1912.06842 [cs], https://doi.org/10.48550/arXiv.1912.06842.
124. Chang, D.; Ding, Y.; Xie, J.; Bhunia, A.K.; Li, X.; Ma, Z.; Wu, M.; Guo, J.; Song, Y.Z. The Devil is in the Channels: Mutual-Channel Loss for Fine-Grained Image Classification. *IEEE Transactions on Image Processing* **2020**, *29*, 4683–4695. arXiv:2002.04264 [cs], https://doi.org/10.1109/TIP.2020.2973812.
125. Subramanya, A.; Pillai, V.; Pirsiavash, H. Fooling network interpretation in image classification. *Proceedings of the IEEE International Conference on Computer Vision* **2019**, *2019-Octob*, 2020–2029. https://doi.org/10.1109/ICCV.2019.00211.
126. Alvarez-Melis, D.; Jaakkola, T.S. On the Robustness of Interpretability Methods **2018**.
127. Garreau, D.; von Luxburg, U. Explaining the Explainer: A First Theoretical Analysis of LIME **2020**.
128. Stiffler, M.; Hudler, A.; Lee, E.; Braines, D.; Mott, D.; Harborne, D. An Analysis of Reliability Using LIME with Deep Learning Models **2018**.
129. Kapishnikov, A.; Venugopalan, S.; Avci, B.; Wedin, B.; Terry, M.; Bolukbasi, T. Guided Integrated Gradients: An Adaptive Path Method for Removing Noise. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **2021**, *abs/2106.0*, 5048–5056. https://doi.org/10.1109/CVPR46437.2021.00501.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 March 2024
doi:10.20944/preprints202403.0484.v1

130. Kapishnikov, A.; Bolukbasi, T.; Viegas, F.; Terry, M. XRAI: Better Attributions Through Regions. *Proceedings of the IEEE International Conference on Computer Vision* **2019**, *2019-October*, 4947–4956. https://doi.org/10.48550/arxiv.1906.02825.
131. Smilkov, D.; Thorat, N.; Kim, B.; Viégas, F.; Wattenberg, M. SmoothGrad: removing noise by adding noise **2017**. https://doi.org/10.48550/arxiv.1706.03825.
132. Jin, H.; Chu, X.; Qi, J.; Zhang, X.; Mu, W. CWAN: Self-supervised learning for deep grape disease image composition. *Engineering Applications of Artificial Intelligence* **2023**, *123*, 106458. https://doi.org/10.1016/j.engappai.2023.106458.
133. Jin, H.; Li, Y.; Qi, J.; Feng, J.; Tian, D.; Mu, W. GrapeGAN: Unsupervised image enhancement for improved grape leaf disease recognition. *Computers and Electronics in Agriculture* **2022**, *198*, 107055. https://doi.org/10.1016/j.compag.2022.107055.
134. Kerkech, M.; Hafiane, A.; Canals, R. Vine disease detection in UAV multispectral images using optimized image registration and deep learning segmentation approach. *Computers and Electronics in Agriculture* **2020**, *174*, 105446. https://doi.org/10.1016/j.compag.2020.105446.
135. Jin, H.; Chu, X.; Qi, J.; Feng, J.; Mu, W. Learning multiple attention transformer super-resolution method for grape disease recognition. *Expert Systems with Applications* **2024**, *241*, 122717. https://doi.org/10.1016/j.eswa.2023.122717.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.