Early yield prediction in different grapevine varieties using computer vision and machine learning

Fernando Palacios[1,2] · Maria P. Diago[1,2] · Pedro Melo-Pinto[3,4] · Javier Tardaguila[1,2]

Abstract

Yield assessment is a highly relevant task for the wine industry. The goal of this work was to develop a new algorithm for early yield prediction in different grapevine varieties using computer vision and machine learning. Vines from six grapevine (*Vitis vinifera* L.) varieties were photographed using a mobile platform in a commercial vineyard at pea-size berry stage. A SegNet architecture was employed to detect the visible berries and canopy features. All features were used to train support vector regression (SVR) models for predicting number of actual berries and yield. Regarding the berries' detection step, a F1-score average of 0.72 and coefficients of determination (R^2) above 0.92 were achieved for all varieties between the number of estimated and the number of actual visible berries. The method yielded average values for root mean squared error (RMSE) of 195 berries, normalized RMSE (NRMSE) of 23.83% and R^2 of 0.79 between the number of estimated and the number of actual berries per vine using the leave-one-out cross validation method. In terms of yield forecast, the correlation between the actual yield and its estimated value yielded R^2 between 0.54 and 0.87 among different varieties and NRMSE between 16.47% and 39.17% while the global model (including all varieties) had a R^2 equal to 0.83 and NRMSE of 29.77%. The number of actual berries and yield per vine can be predicted up to 60 days prior to harvest in several grapevine varieties using the new algorithm.

Keywords Yield estimation · Non-invasive sensing technologies · Digital viticulture · SegNet architecture

✉ Javier Tardaguila
javier.tardaguila@unirioja.es

1 Televitis Research Group, University of La Rioja, 26006 Logroño, Spain

2 Instituto de Ciencias de la Vid y del Vino (University of La Rioja, CSIC, Gobierno de La Rioja), 26007 Logroño, Spain

3 CITAB—Centre for the Research and Technology of Agro-Environmental and Biological Sciences, Universidade de Trás-os-Montes e Alto Douro, 5000-801 Vila Real, Portugal

4 Departamento de Engenharias, Escola de Ciências e Tecnologia, Universidade de Trás-os-Montes e Alto Douro, 5000-801 Vila Real, Portugal

Introduction

Yield prediction has been recognized as a key subject in agriculture (Klompenburg et al., 2020), and particularly in the grape and wine industry (Carrillo et al., 2016; Clingeleffer et al., 2001; Dunn & Martin, 2003; Laurent et al., 2021; Taylor et al., 2019). Vineyard yield prediction is crucial to achieve the desired fruit quantity and quality (Krstic et al., 1998; Taylor et al., 2019), therefore, the objective and fast estimation of vine yield would be very valuable for grapegrowers (Dunn & Martin, 2000; Laurent et al., 2021; Martin et al., 2003). Several works on yield estimation and forecasting in vineyards have been published in the last decades. An interesting review of the approaches, methods and challenges for vineyard yield estimation, prediction and forecasting was recently published by Laurent et al. (2021). In this review, the most relevant features regarding yield development in an operational scenario that have to be considered by yield assessment methods are summarized and discussed. Moreover, the yield assessment methods reported in the literature are classified and compared in light of the measurement, estimation and modelling approaches. Conventional methods for assessing yield are destructive, labour-demanding and time-consuming (Clingeleffer et al., 2001; Martin et al., 2003), thus accurate, objective and rapid approaches are needed for improving yield assessment in operational contexts (Dunn & Martin, 2004; Laurent et al., 2021; Taylor et al., 2019). Moreover, the new methods should consider the high spatial and temporal variability of yield components (Bramley et al., 2019; Carrillo et al., 2016; Li et al., 2017; Oger et al., 2021a, 2021b; Taylor et al., 2005).

Yield components including the number of clusters per vine, cluster weight, berry number per cluster and berry weight are important in vineyard yield estimation. High annual variation (60 to 70%) of final yield is explained by the number of clusters per vine (Clingeleffer et al., 2001). However, recent trials in several vineyards in southern France (involving different training system, age of vines and grapevine variety) across different seasons revealed that the role of the number of clusters per vine was more variable, explaining from 39 to 99% of the temporal yield variance (Carrillo et al., 2016). Clingeleffer et al. (2001) also reported that the number of berries per bunch, the berry weight and the interaction between the number of clusters per vine and the number of berries per cluster respectively explained 11%, 4% and 20% of the spatial yield variability. Laurent et al. (2021) suggested that the classical values of vineyard yield components variability (60%, 30% and 10% for number of clusters per vine, number of berries per cluster and berry weight, respectively) should not be established for all situations in yield assessment.

Novel sensing technologies enable an efficient acquisition of data and precise forecasts in agriculture. Optical sensors are widely proposed for assessing vineyard spatial variability in precision viticulture (Ballesteros et al., 2020; Hall et al., 2011). Spectral sensors are applied to detect vegetation features, yield and canopy characteristics. A high correlation value was found between vegetation indices and yield using both remote (Hall et al., 2011) and proximal sensing technologies (Sozzi et al., 2019). Moreover, crop models can be used for yield prediction of different vineyard zones (Laurent et al., 2021). Vineyard yield was predicted by combining vegetation indices (using a spectral sensor mounted in UAV) and computer vision to obtain the vegetated fraction cover as an indicator of vine vigour (Ballesteros et al., 2020).

Computer vision systems are powerful tools to automate inspection tasks in agriculture and food processing (Cubero et al., 2011; Tian et al., 2020). Typical target applications of such systems include grading, yield component monitoring (Cubero et al., 2011; Liu et al., 2018; Nuske et al., 2014a, 2014b). With computer vision techniques, a large set of samples

can be evaluated to provide more objective information (Cubero et al., 2011; Grimm et al., 2019; Liu & Whitty, 2015).

Grapevine yield estimation has been widely addressed using computer vision at different phenological stages, such as budbreak (Liu et al., 2017), flowering (Liu et al., 2018; Palacios et al., 2020), pea-size (Aquino et al., 2018; Liu et al., 2020; Palacios et al., 2021, 2022), and harvest (Dunn & Martin, 2004; Liu et al., 2013, 2015; Nuske et al., 2014b; Xin et al., 2020). A recent work (Hacking et al., 2020), concluded that the final stage of berry ripening was the ideal phenological stage for grapevine yield estimation. Information about the number of berries can lead to an adequate and early estimation of the yield in grapevine. Most of the previous works have focused on the detection of visible fruits. Some authors have developed algorithms capable of quantifying the number of visible berries in RGB images acquired under natural conditions, using traditional image analysis methods (Aquino et al., 2018; Nuske et al., 2014b) or newer deep learning techniques (Buayai et al., 2021; Grimm et al., 2019; Klompenburg et al., 2020). These works have suggested a procedure on yield forecasting for defoliated vineyards where the number of visible berries in the images was proportional to the total number of berries. However, the number of visible berries was only a fraction of the actual number of berries in the vine, as the percentage of exposed berries may vary beyond lineal relationships according to canopy conditions in the fruiting zone (Iñiguez et al., 2021). Leaf and berry occlusions are the main challenges for yield forecasting using computer vision based methods in commercial vineyards (Iñiguez et al., 2021; Victorino et al., 2020). As leaf occlusion rate (cluster occlusion by leaves) increased, the relationship between fruit pixels and yield was gradually reduced (Iñiguez et al., 2021). Yield prediction based on computer vision can also be benefited by leaf removal (a common practice in many winegrape regions) in the fruiting zone to decrease the berry occlusion affected by leaves (Iñiguez et al., 2021). Additionally, different grapevine varieties develop canopies of variable vigour, leading to occlusion situations of different extent (Diago et al., 2016; Nuske et al., 2014b), so new algorithms for yield prediction using vine images in different grapevine varieties need to be investigated.

The yield spatial variability within a commercial vineyard is a well-known fact (Bramley et al., 2019; Taylor et al., 2005). For this reason, understanding and monitoring the spatial distribution of yield variation is essential. This cannot be achieved using classical, destructive methods, as a large number of measurements are required for adequate representativeness and yield estimation. Indeed, it is a highly relevant task to perform and to obtain more precise early yield predictions in commercial vineyards. The image acquisition for monitoring commercial vineyards could be carried out using mobile sensing platforms. These moving platforms (e.g. All-terrain vehicles, tractors, agricultural machinery) can be used to improve yield estimation following the suggestions of several authors (Arnó et al., 2017; Oger et al., 2021b; Tisseyre et al., 2018).

Machine learning is an important element of a decision support tool for crop yield prediction in agriculture (Klompenburg et al., 2020). Several machine learning algorithms have been applied for predicting the number of actual berries and final yield at harvest (Zabawa et al., 2020) under different levels of leaf and berry occlusions in grapevines. Similarly, Monga et al. (2018) used convolution neural networks to develop models that can estimate yield using RGB images. A new approach for vineyard yield estimation was recently suggested by Ballesteros et al. (2020) combining remote sensing, computer vision and artificial neural network techniques.

The research idea was guided by a heuristic engineering method proposed by Koen (1988), which involves defining a problem, proposing a solution and testing that solution. Aiming at the task of early yield prediction in commercial vineyards, a method based on

computer vision and machine learning using RGB images acquired on-the-go was developed. To test the solution, the evaluation results were also compared with the reference method and the limitations of the method proposed were analysed. Moreover, the differences in the method performance when prediction modelling was conducted individually for each grapevine variety versus a general model were evaluated.

Materials and methods

The experimental procedure comprised several steps (Fig. 1). It started from on-the-go image acquisition of six grapevine varieties under field conditions at night time. The clusters were then segmented in the vine images, and the output was used for berries' and canopy elements' segmentation. The number of berries, obtained after a second segmentation step, and several features related to elements of the canopy, obtained after canopy elements' segmentation, were then combined into a dataset, and used in regression models to obtain an estimation of the number of actual berries and the yield per vine. Two possible scenarios of applicability were considered for these models: global models, where the training set contained multiple varieties, and local models, where a different model was trained and tested on each variety individually. For global models, three cross validation methods were considered: leave-one-out (LOOCV) and eightfold, where test sets contained
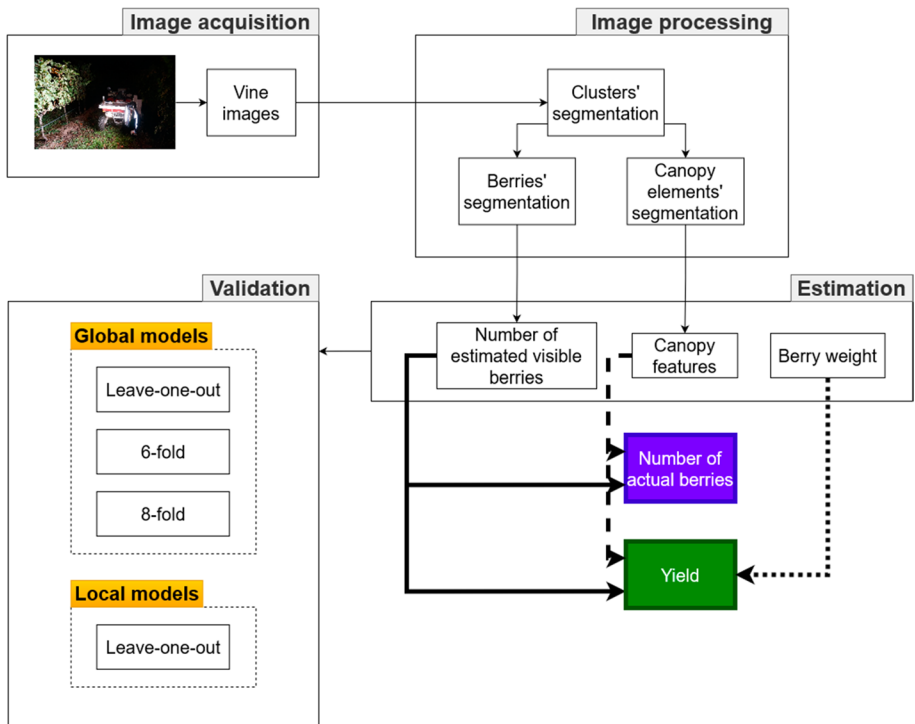

Fig. 1 Flow-chart of the full process for the estimation of the number of actual berries and yield in commercial vineyards

varieties already included in the training set (to analyse the behaviour on varieties previously known by the model), and leave-one-variety-out (LOVOCV), where the test sets included a variety not included in the training set (to analyse the behaviour on varieties unknown by the model). In the case of local models, only LOOCV was performed on each variety individually to test the performance of a model trained only with the variety on which it will be applied.

Experimental layout

A set of six grapevine (*Vitis vinifera* L.) varieties were selected for the experiment in a commercial vineyard during season 2018.Three red (Cabernet Sauvignon, Syrah and Tempranillo) and three white (Malvasia, Muscat of Alexandria and Verdejo) varieties were used in this work (Fig. 2). The vineyard was located in Vergalijo (lat. 42°27′46.0″ N; long. 1°48′13.1″ W; Navarra, Spain). A set of 96 vines (16 per grape variety) were selected and

Fig. 2 Examples of vine canopy images of each grapevine variety: **a** Cabernet Sauvignon, **b** Malvasia, **c** Muscat of Alexandria, **d** Syrah, **e** Tempranillo and **f** Verdejo

vertically delimited using a plastic strip. Each grapevine variety was planted in a single row. In total six rows were monitored (one per variety). A vertical shoot positioned (VSP) trellis system was used for training the vines, with 2 m row spacing and 1 m vine spacing. Partial defoliation, a common viticultural practice in northern Spain, consisting of removal of three leaves per shoot around the fruiting zone was applied to the vines at fruit-set stage (Fig. 2). Partial defoliation in this work was manually applied in order to remove precisely the same number of main basal leaves per each shoot (only three leaves per shoot were removed).

At harvest, for each tagged vine their clusters were collected, weighted and transported to the laboratory for destemming and manual counting of their berries. Yield, berry number and berry weight were determined for each vine.

Image acquisition

The acquisition of RGB images was performed the 3rd July 2018 (66 days before harvest), at the phenological stage of pea size (green berries of around 7 mm diameter, according to Coombe, 1995). A mobile sensing platform developed at the University of La Rioja (Fig. 3a) was used for the night-time image acquisition. The main component of this platform was an all-terrain-vehicle (ATV (Trail Boss 330, Polaris Industries, Minnesota, USA) customized to incorporate a custom-made aluminium structure. This structure included a RGB Canon EOS 5D Mark IV (Canon Inc. Tokyo, Japan) camera, mounting a full-frame CMOS (ccomplementary metal–oxide–semiconductor) sensor (35 mm and 30.4 MP) and equipped with a Canon EF 35 mm F/2 IS USM lens. The camera was positioned at a
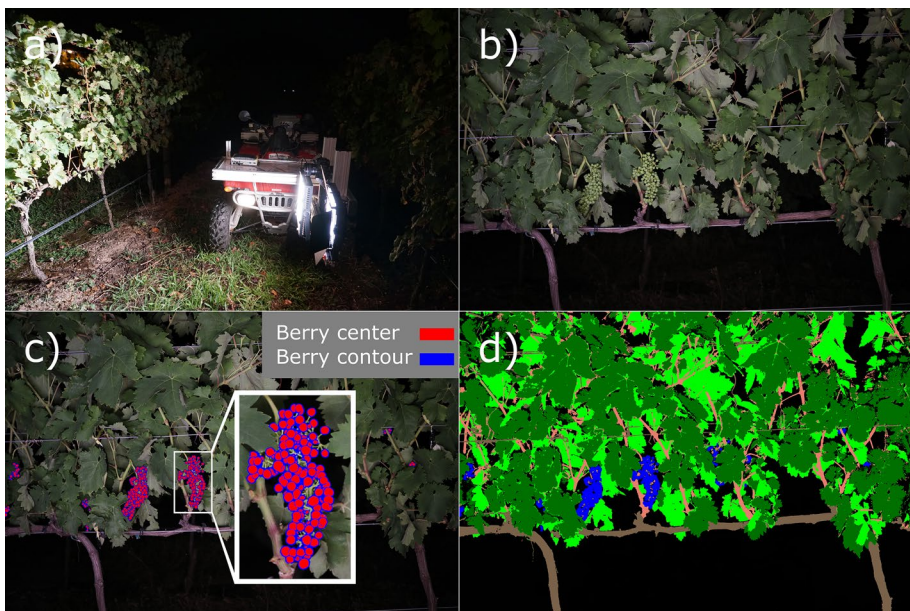

Fig. 3 Mobile imaging platform and image processing stages: **a** an all-terrain-vehicle (ATV) modified to incorporate a RGB camera for on-the-go image acquisition, **b** original vine image, **c** berries' segmentation and **d** canopy features' segmentation

distance of 1.80 m, perpendicular to the canopy and at a 1 m of height from the ground. The camera settings were manually fixed at the beginning of the trials with the following values: exposure time of 1/1600 s, ISO speed of 2000 and f-number of f/5.6. This camera enabled the acquisition of images with a resolution of 6720×4480 pixels. The structure also included a triggering system for automating the RGB image capture and an artificial illumination system for the night-time image acquisition, necessary to obtain vine images with homogeneous illumination, and to differentiate the vine under evaluation from the vines of the opposite row. This illumination system consisted of a 1200 LEDs NanGuang CN-1200CHS panel (72 W, 3681.5 lm) and a 900 LEDs Bestlight panel (54 W, 6480 lm). The mobile platform was described and used in previous works (Diago et al., 2019; Palacios et al., 2020).

Image processing

Deep learning (Deng & Yu, 2014) was used in this work to perform the semantic segmentation of the elements presented in the images (Fig. 3b). Each image pixel was assigned to a predefined class using the SegNet DL architecture (Badrinarayanan et al., 2017). This architecture was designed with an encoder, that performs down-sampling operations in order to extract feature maps from the images, and a decoder, that produces the final pixel-wise labelling by reversing the operations performed by the encoder. The encoder was formed by the layers of a convolutional neural network (CNN) pre-trained model, and the decoder mirrored the encoder. The layers for the encoder used in this work were the ones from VGG16, introduced by Simonyan and Zisserman (2015).

The segmentation of the berries and canopy elements was previously described in Palacios et al. (2022). The first step was to segment the clusters, in order to identify the image regions that contained berries, and to reduce false positives at segmenting them. Also, the segmented clusters were used in canopy segmentation. SegNet was trained with a set of 1646 image patches randomly extracted from 18 full-resolution images from different vines than the 96 presented in the experimental layout section, but from the same vineyard and varieties. Two image scales were considered for these images: the original full-resolution image scale, of which half of them (823 image patches) contained clusters while the remaining ones contained other objects (usually canopy elements). Each pixel of these patches was labelled as "cluster" class or "background" class. The size of the image patches was 1120×1120 but they were resized to 560×560 pixels in order to reduce the training time.

After performing the segmentation of the clusters, SegNet was trained to segment berries (Fig. 3c) using the same image patches with a new pixel labelling. In this step pixels were labelled as "background", representing non-berry pixels, "contour" and "center", representing the contour and the center of the berries, defined as berry pixels completely surrounded by "contour" pixels (Fig. 3c).

Finally, the canopy elements were segmented using six canopy classes (Fig. 3d). These were "gap", "leaf abaxial" (lower side of the leaf), "leaf adaxial" (upper side of the leaf), "shoot", "trunk" and "cluster" (segmented previously in the first step). The labelling of these classes was achieved using a semi-automatic approach. Six colour features, corresponding to the red–green–blue from RGB colour space, and L-a-b from CIELAB, were extracted from the images and a different multinomial logistic regression (MLR) model was trained for each variety using these features. Each training dataset contained 300 pixels (50 per class). Then, these models were used to segment the images of its corresponding

variety and the outputs were used as training masks for SegNet, which enabled performing the semantic segmentation of the canopy elements for all varieties (Fig. 3d).

Models for number of actual berries and yield estimation

The canopy features calculated per vine after the canopy segmentation for the number of actual berries estimation were the following:

- F1: Number of estimated visible berries (calculated as the number of adjacent pixel groups from berry "center" class).
- F2: The average of the ratios between the number of "leaf abaxial" pixels and the total area for each bounding box (rectangular regions of the image) containing a cluster.
- F3: The number of "shoot" class pixels in the fruiting zone.
- F4: The number of "leaf adaxial" class pixels in the fruiting zone.
- F5: The ratio between F4 and the area of the fruiting zone.
- F6: The area of the fruiting zone, expressed in pixels.

These features were included in a regression model to estimate the number of actual berries (Palacios et al., 2022). For yield estimation, in addition to these features, the average berry weight determined at harvest for each grapevine variety was also included in the model. This was done in order to study the hypothetical improvement that could be obtained when the average berry weight per variety (which could be calculated from previous seasons) is included in the model. Descriptive statistics for each feature included in the models and possible outcomes are shown in Tables 1 and 2.

The regression method used in this work for both estimation tasks (number of actual berries and yield) was a support vector regression (Bishop, 2006).

Evaluation metrics

The metrics used in this work to test the performance of the algorithm at detecting individual berries were the following:

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$F_1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{3}$$

Being *TP* the true positives (number of correctly detected berries), *FP* the false positives (number of detected berries that were not actually berries) and *FN* the false negatives (number of actual berries that were not detected). In the context of this work, precision (Eq. 1) represents the proportion of correctly detected berries from the whole set of objects identified as berries by the algorithm. Recall (Eq. 2) is the proportion of actual berries detected by the algorithm from the whole set of berries visible in the images. A perfect performance of the algorithm would be achieved with a precision and recall of 1.0. The

Table 1 Average and standard deviation of each feature included in the model and output parameter, calculated per variety and combining all varieties in the full set

Grapevine variety		Features							Output	
		F1	F2	F3	F4	F5	F6	Berry weight (g)	Number of actual berries	Yield (Kg)
Cabernet Sauvignon	Avg	295	0.188	8.196E+05	4.574E+06	0.454	1.001E+07	1.284	809	1.053
	Std	160	0.064	3.711E+05	1.434E+06	0.063	2.644E+06	0.189	374	0.536
Malvasia	Avg	341	0.268	4.838E+05	5.237E+06	0.551	9.421E+06	2.592	994	2.596
	Std	192	0.104	3.541E+05	1.980E+06	0.076	3.118E+06	0.231	501	1.296
Muscat of Alexandria	Avg	374	0.289	3.228E+05	3.148E+06	0.503	6.372E+06	1.470	857	1.321
	Std	300	0.110	1.906E+05	1.268E+06	0.116	2.766E+06	0.232	712	1.088
Syrah	Avg	473	0.104	1.641E+06	2.308E+06	0.317	7.423E+06	1.606	571	0.900
	Std	196	0.066	5.217E+05	6.481E+05	0.088	1.513E+06	0.129	246	0.350
Tempranillo	Avg	171	0.191	4.000E+05	4.218E+06	0.555	7.630E+06	1.483	588	0.865
	Std	105	0.105	1.534E+05	6.448E+05	0.071	9.275E+05	0.248	243	0.417
Verdejo	Avg	766	0.169	7.206E+05	3.096E+06	0.389	7.973E+06	2.144	1095	2.279
	Std	351	0.067	2.278E+05	1.012E+06	0.072	2.170E+06	0.218	566	1.021
Full set	Avg	403	0.202	7.312E+05	3.763E+06	0.462	8.138E+06	1.764	819	1.502
	Std	293	0.106	5.474E+05	1.580E+06	0.118	2.572E+06	0.526	499	1.087

Avg Average, *Std* Standard deviation, *F1* Number of estimated visible berries, *F2* The average of the ratios between the number of "leaf abaxial" pixels and the total area for each bounding box (rectangular regions of the image) containing a cluster, *F3* The number of "shoot" class pixels in the fruiting zone, *F4* The number of "leaf adaxial" class pixels in the fruiting zone, *F5* The ratio between F4 and the area of the fruiting zone, *F6* The area of the fruiting zone, expressed in pixels.

Table 2 Coefficient of variation for each feature included in the model and output parameter, calculated per variety and combining all varieties in the full set

Grapevine variety	Features							Output	
	F1	F2	F3	F4	F5	F6	Berry weight (g)	Number of actual berries	Yield (Kg)
Cabernet Sauvignon	54.33%	34.04%	45.28%	31.35%	13.90%	26.41%	14.74%	46.18%	50.87%
Malvasia	56.39%	38.79%	73.20%	37.81%	13.74%	33.10%	8.92%	50.38%	49.91%
Muscat of Alexandria	80.04%	38.17%	59.05%	40.30%	23.14%	43.41%	14.79%	83.06%	82.40%
Syrah	41.48%	63.06%	31.80%	28.08%	27.58%	20.38%	8.04%	43.14%	38.86%
Tempranillo	61.59%	55.28%	38.34%	15.29%	12.71%	12.16%	16.73%	41.25%	48.27%
Verdejo	45.80%	39.50%	31.61%	32.68%	18.44%	27.21%	10.17%	51.72%	44.78%
Full set	72.75%	52.65%	74.86%	42.00%	25.63%	31.61%	27.85%	60.89%	72.34%

F1 Number of estimated visible berries, *F2* The average of the ratios between the number of "leaf abaxial" pixels and the total area for each bounding box (rectangular regions of the image) containing a cluster, *F3* The number of "shoot" class pixels in the fruiting zone, *F4* The number of "leaf adaxial" class pixels in the fruiting zone, *F5* The ratio between F4 and the area of the fruiting zone, *F6* The area of the fruiting zone, expressed in pixels.

balance between both metrics is represented by the F1-score (Eq. 3), that is defined as the harmonic mean of precision and recall (Chinchor, 1992).

For the estimation of the number of actual berries and yield, three regression metrics were selected. These include the coefficient of determination (R^2), the root-mean-squared error (RMSE) and the normalized root-mean-squared error (NRMSE), calculated as the ratio of RMSE over the average of the number of actual berries. The RMSE and NRMSE are mathematically defined as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y'_i - y_i)^2}{n}} \tag{4}$$

$$NRMSE = \frac{RMSE}{\frac{\sum_{i=1}^{n} y_i}{n}} \times 100\% \tag{5}$$

where y_i is the observed value from the i-th vine (the number of actual berries or yield weighted), y'_i is the estimation performed on the i-th vine and n is the number of vines.

Support vector regression hyperparameters' tuning

The support vector regression hyperparameter's tuning was performed using a Bayesian Optimization algorithm (Mockus et al., 2014), which aims to minimize an objective function (a performance metric for the model trained with a set of hyperparameters) using a gaussian process model, a bayesian procedure that modifies the Gaussian process model at each new evaluation of the objective function, and an acquisition function that is maximized to find the next point to be evaluated by the objective function (i.e. the next set of hyperparameters). The Bayesian optimization algorithm has probed to outperform other optimization algorithms (Jones, 2001).

The function fitrsvm implemented in Matlab 2018b was used to train the support vector regression and optimize the hyperparameters using the bayesian optimization algorithm. In this algorithm a gaussian process with ARD Matérn 5/2 kernel model and the "expected-improvement-plus" acquisition function were employed, along with the mean-squared-error (MSE) metric as the objective function (calculated using leave-one-out cross validation performed on the full set of 95 vines).

Two sets of hyperparameters for support vector regression were obtained: one for the actual berry number estimation, and another one for yield estimation. These hyperparameters are shown in Table 3 and remained fixed for all cross-validation methods within each estimation task. In addition, these features were standardized before training.

Models' validation

The validation set for the individual berries' segmentation step was formed by a set of 60 full-resolution (6720×4480 pixels) images (ten per grapevine variety). The validation was focused on testing the accuracy of the method at counting the visible berries, rather than testing the quality of the segmentation. Therefore, the centres of the berries were manually checked in the images.

Table 3 Support vector regression hyperparameters sets employed for number of actual berries and yield estimation

Estimation parameter	Kernel function (fixed)	Tuned hyperparameters		
		Box constraint	Kernel scale	ε
Number of actual berries	Linear	0.0010135	0.0028975	83.947
Yield	Linear	852.1	10.254	0.31103

For the estimation model, three cross-validation methods were applied on the 96 vines. These methods were considered to test the applicability and the performance of the support vector regression model as a global model (involving all grapevine varieties) and as a local model (separate models for the different varieties). As a global model, two possible cases of application were considered. In the first one, the model was trained with data acquired from of all varieties and tested on other vines from the same varieties. For this, leave one out cross validation (LOOCV) and eightfold cross validation were tested. In LOOCV, in each iteration the model was trained with 95 vines and tested on one vine, while in the eightfold cross validation, the model was trained with 84 vines and tested with 12 in each iteration, randomly selecting the vines included in the training and test sets. In order to avoid bias in the partitioning of the data into both sets, eightfold was repeated 30 times, and the average of the values obtained for the regression metrics (RMSE, NRMSE and R^2) in the 30 repetitions was selected as the final result for eightfold. In the second case of application as a global model, this was trained with data acquired from five of the six varieties and tested on the remaining variety, in order to verify the generalization capability of the model to estimate the number of actual berries and yield in new varieties previously unknown to the model. Hence, a leave-one-variety-out cross validation was applied, were in each iteration the model was trained with 80 vines from five varieties (all 16 vines from five varieties) and tested in the 16 vines of the remaining variety. As a local model, each variety was considered a different dataset and the model was trained in each variety independently. As the number of vines per variety was low (16 vines), LOOCV was carried out. Likewise, for each iteration the model was trained in 15 vines and tested in the remaining one.

The average berry weight feature for yield estimation was calculated considering only the vines in the training fold for each cross validation method.

Results and discussion

Individual berries' detection

The performance metrics of the individual berries' detection step are shown in Table 4. It can be observed that the algorithm achieved a precision higher than 0.60 for most of the varieties, except for Verdejo (precision = 0.52), indicating a higher false positive rate for this variety. In terms of recall, similar results, all of them above 0.84, were obtained for all varieties. The balance between both metrics is expressed by the F1-Score which shows values over 0.70 for all varieties except for Verdejo (with a F1-Score of 0.65). The poorer results observed for this white variety could be related to the higher size of its clusters, as compared to the other five varieties. Likewise,

Table 4 Performance results of the detection of the number of visible berries

Grapevine variety	Precision	Recall	F1-Score
Cabernet Sauvignon	0.642	0.851	0.730
Malvasia	0.652	0.865	0.741
Muscat of Alexandria	0.689	0.857	0.763
Syrah	0.635	0.885	0.739
Tempranillo	0.617	0.842	0.710
Verdejo	0.520	0.863	0.648
Average	0.626	0.860	0.722

The validation set included 60 images (10 images per variety)

Verdejo clusters, with a bigger number of berries, may have led the algorithm to detect wrong artefacts in the images and identify them as berries. This could be improved by training SegNet with a higher number of images from Verdejo and other varieties with clusters similar to the ones of Verdejo.

The correlation between the number of estimated visible berries (using the detection algorithm) and the actual visible berries per image is shown in Fig. 4. A strong correlation was found for all varieties, with coefficient of determination (R^2) above 0.92, suggesting that the error made by the algorithm is proportional to the number of visible berries, while a fixed deviation still existed for all varieties. Generally speaking, the algorithm overestimated the number of visible berries, but this was more pronounced in Cabernet Sauvignon (Fig. 4a), Malvasia (Fig. 4b) and Verdejo (Fig. 4f).

In the work of Nuske et al. (2014b), the berries' detection algorithm presented by these authors obtained different recall values depending on the illumination conditions and the variety. These values ranged from 0.66 to 0.89. For some varieties the recall values were inferior to the ones presented in Table 4, while for the remaining ones (Traminette, Chardonnay, Flame Seedless and Petite Syrah) the recall values were similar. In good agreement with the present study, Aquino et al. (2018) reported recall values between 0.83 and 0.89 for Cabernet Sauvignon, Syrah and Tempranillo while the precision values ranged from 0.94 to 0.97. This means that Aquino et al. (2018) detected a similar rate of berries potentially due to the lower amount of canopy elements that can be mistaken for berries by the algorithm (mainly leaves), as the fruiting zone was fully defoliated. As in the case of the work of Aquino et al. (2018), the recall values (also the F1-scores) were higher in Grimm et al. (2019) while the precision achieved was very similar to that of the present study. In this case, the difference in colour of the ripe berries versus the canopy leaves (images were acquired after veraison and near harvest) made have played a significant role in the segmentation step, making the berries more distinguishable from the remaining canopy elements, leading to lower false positives' rates. So, either occlusion phenomena (for instance by leaves) or similarities in the green colour of the berries at the pea-size phenological stage with that of leaves, shoots, or other elements of the canopy seem to difficult the proper identification and counting of berries, leading to a considerable number of false positives. In an interesting and recent work, Zabawa et al. (2020) reported high precision and recall values for Riesling berries detection. However, an artificial background was attached to the mobile phenotyping platform, which, from a practical perspective may have more limited application in an operational context.
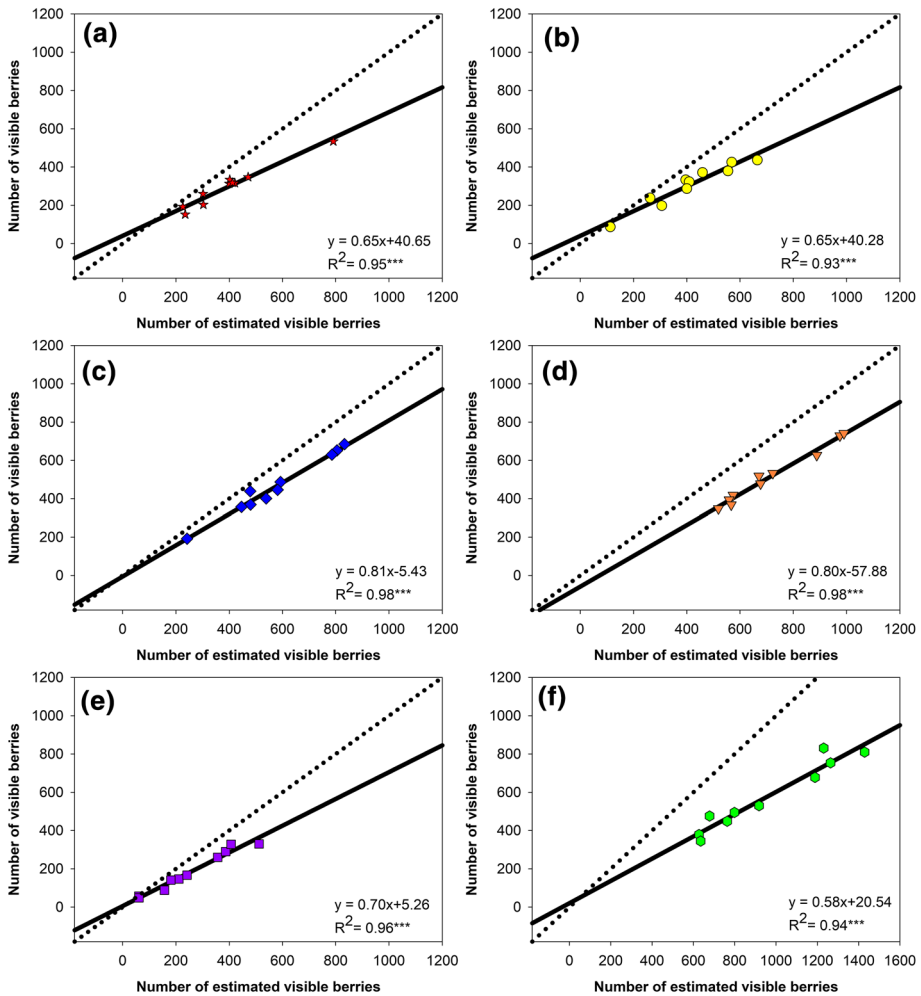
Fig. 4 Correlation between the number of estimated visible berries (F1 feature) and the number of visible berries per image, for each grapevine variety (10 images per variety): **a** Cabernet Sauvignon, **b** Malvasia, **c** Muscat of Alexandria, **d** Syrah, **e** Tempranillo and **f** Verdejo. Determination coefficients (R^2) were significant at $p=0.001$ (***). The solid line represents the correlation line and the dotted line represents the 1:1 line

Estimation of the number of actual berries

The correlations between the number of estimated visible berries and the number of actual berries per vine for each grapevine variety is shown in Fig. 5. These results show high R^2 values for Muscat of Alexandria, Syrah, Tempranillo and Verdejo varieties, while for Cabernet Sauvignon and Malvasia lower R^2 were obtained. These results proved that a linear relationship between the number of visible berries in the images and the number of actual berries in the vines existed for four of the six studied varieties. However, this linear relationship was only achieved within each variety independently, but not combining multiple
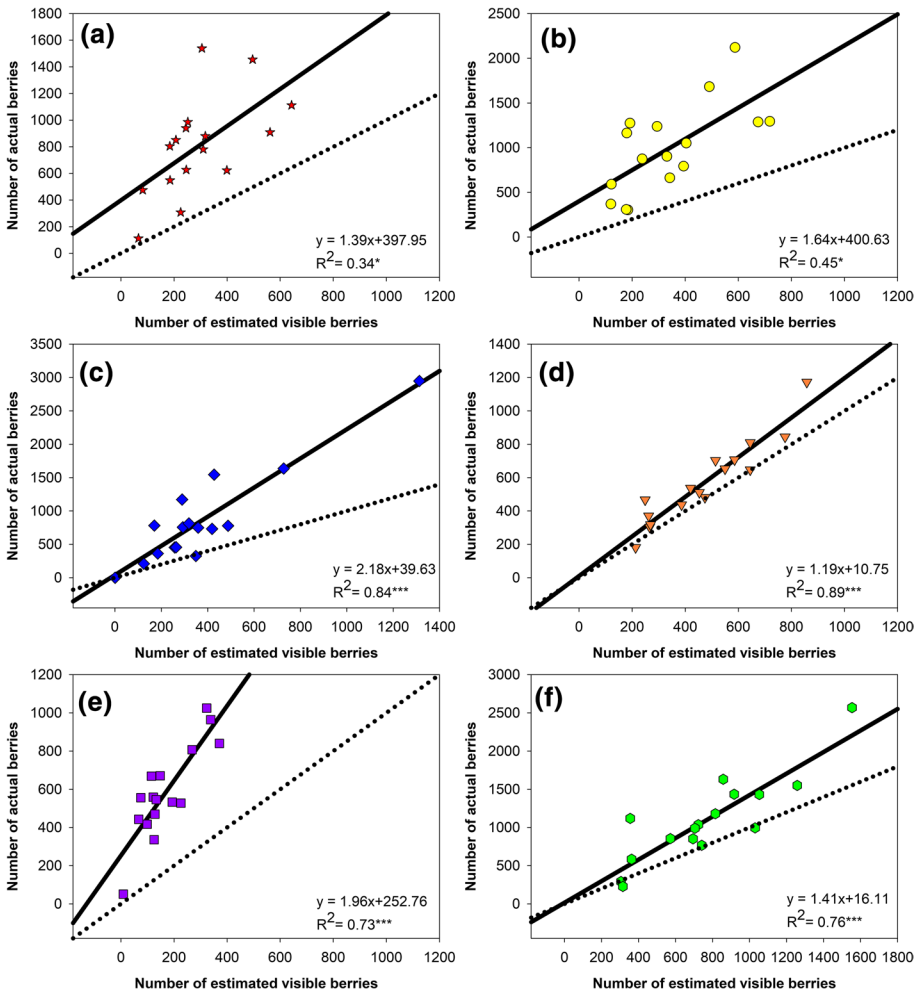
Fig. 5 Linear correlation between the number of estimated visible berries (F1 feature) and the number of actual berries per vine (manually counted in the laboratory), for each grapevine variety: **a** Cabernet Sauvignon, **b** Malvasia, **c** Muscat of Alexandria, **d** Syrah, **e** Tempranillo and **f** Verdejo. Determination coefficients (R^2) were significant at $p=0.05$ (*) and $p=0.001$ (***). The solid line represents the correlation line, and the dotted line represents the 1:1 line

varieties (global model). This can be inferred from the observation of the variability presented in the slopes of the linear functions for the different varieties, as these ranged from 1.19 (Syrah, Fig. 5d) to 2.18 (Muscat of Alexandria, Fig. 5c). Therefore, it was not possible to estimate the number of actual berries per vine using a global linear regression that included only the number of estimated visible berries from different varieties.

The performance of the estimation models that involved the number of estimated visible berries along canopy features is shown in Tables 5 and 6.

The results obtained for the application of a global model, where multiple varieties were used for training, are presented in Table 5. It can be observed that similar results

Table 5 Number of actual berries' estimation performance using a support vector machine (SVM) global model

Grapevine variety	Cross-validation method								
	Leave-one-out			Eightfold			Leave-one-variety-out		
	RMSE	NRMSE (%)	R^2	RMSE	NRMSE (%)	R^2	RMSE	NRMSE (%)	R^2
Cabernet Sauvignon	194	23.94	0.72	199	24.61	0.70	195	24.14	0.72
Malvasia	198	19.88	0.84	207	20.82	0.83	221	22.19	0.82
Muscat of Alexandria	294	34.32	0.86	306	35.71	0.85	310	36.13	0.85
Syrah	102	17.89	0.85	106	18.57	0.84	320	55.96	0.67
Tempranillo	150	25.44	0.63	155	26.34	0.59	149	25.24	0.64
Verdejo	235	21.49	0.84	241	21.97	0.83	274	25.00	0.84
Average	195	23.83	0.79	202	24.67	0.77	245	31.44	0.76

Training and testing with all varieties combined (leave-one-out and eightfold cross validation) and training using five varieties and testing with the other one (leave-one-variety-out cross validation).

Table 6 Number of actual berries estimation performance using a support vector machine (SVM) local model

Grapevine variety	RMSE	NRMSE (%)	R^2
Cabernet Sauvignon	270	33.42	0.44
Malvasia	180	18.13	0.86
Muscat of Alexandria	293	34.17	0.83
Syrah	98	17.07	0.84
Tempranillo	171	29.07	0.48
Verdejo	211	19.25	0.87
Average	204	25.19	0.72

Training and testing were conducted individually for each variety performing leave-one-out cross validation (LOOCV).

were obtained between the two cross validation methods that used all varieties for training and the test set (leave-one-out and eightfold) either for each individual variety (R^2 ranged between 0.59 to 0.86 and NRMSE ranged from 17.9 to 35.7%) and the average for all of them ($R^2 \sim 0.78$; RMSE ~ 200; NRMSE $\sim 24\%$). Regardless the use of LOOCV or eightfold, the estimation of the actual number of berries was more accurate for Syrah and Malvasia, while the poorest estimation was achieved for Muscat of Alexandria. When the LOVOCV was used (five varieties were used for training and the remaining one for testing) the performance metrics did not improve for any variety and particularly the NRMSE notably increased in Syrah (NRMSE = 55.96%) (Table 5).

Table 6 summarizes the performance metrics of the individual models, one per variety, involving the number of estimated visible berries and canopy features, using LOOCV. Overall, the individual or local models did not provide a more accurate estimation than the global models (when LOOCV and eightfold cross validation were used), and this was especially evident for Cabernet Sauvignon and Tempranillo, that showed $R^2 \sim 0.46$ and NRMSE of 33.42% and 29.07%, respectively.

In general terms, the results obtained for all cross-validation methods (Tables 5 and 6) are similar for all varieties excepting for Cabernet Sauvignon, Tempranillo and Syrah. The fact that equally good results were obtained for Syrah in both LOOCV for global and local model (Tables 5 and 6) and eightfold (Table 5) but a poor performance was achieved in LOVOCV (Table 5) suggests that Syrah is highly different from the rest of varieties. In fact, the photographed vine canopies of this variety showed a higher canopy porosity or canopy gaps abundance in the fruiting zone that imaged canopies for all other five varieties (Fig. 2d). From this it can be inferred that the model trained without Syrah data is giving a higher relevance to the canopy features (due to the absence of vines with a high porosity in the training set) and increasing the training set size by including highly defoliated vines should help to improve the performance of the method on more porous vines. For Cabernet Sauvignon and Tempranillo, the results for the local models (Table 6) suggests that either the variability in the canopy conditions of the vines could not be properly captured by a model including only those varieties individually, or that the number of samples in the local models (16 vines) is not enough to estimate the number of actual berries in those varieties, and that a model that included also other varieties or a higher number of vines could improve the estimation on Cabernet Sauvignon and Tempranillo vines.

From a viticultural perspective, the most interesting results are presented in Table 5 for the LOVOCV, where the use of a global model trained with several varieties yielded promising results at estimating a new variety previously unknown for the model. Of the six studied varieties, only Syrah exhibited lower performance using the LOVOCV. This variety, together with Muscat of Alexandria showed the highest values of NRMSE. These results could be related to differential features of the canopy, clusters and berries in Syrah and Muscat as compared to those of the other varieties under study. The absence of Syrah or Muscat of Alexandria samples in the training set prevents to capture their singularity and, as a result, the estimation of the actual number of berries for these two varieties (when they are not included in the training set) is achieved with much higher uncertainty than for the other four cultivars.

The relevance of each feature in the dataset was evaluated (for all three cross validation methods) by testing the performance of the model when that particular feature had been removed. This is shown in Table 7 for leave-one-out cross validation, Table 8 for eightfold cross validation, and Table 9 for LOVOCV, where features were ordered by their relevance to the model, (in decreasing order of the RMSE/NRMSE obtained). As expected, regardless the cross-validation method, the number of estimated visible berries (F1) was the most relevant feature in the model. In fact, when F1 was not considered in the model the average R^2 decreased from 0.77–0.79 to 0.49–0.51 and the NRMSE increased a 17.33% (for LOOCV, Table 7), 16.79% (eightfold, Table 8) and 14.08% (LOVOCV, Table 9). The relevance of the remaining features was found to be very similar for LOOCV (Table 7) and eightfold cross validation (Table 8). Likewise, two close groups of features could be identified in these global model validation methods. The first group formed by F2 and F3 features, whose removal led to models achieving NRMSE between 26 and 27%, and group two including F4, F5 and F6 features, whose removal led to models yielding NRMSE between 24 and 25%. This could imply that the model could be simplified (not including either F4, F5 or F6) with almost no increase in NRMSE and decrease of R^2. For LOVOCV removal of a particular feature led to a greater variability in the NRMSE. Differently to the LOOCV and eightfold methods, when LOVOCV was used (Table 9) the most relevant feature after F1 was not F3 (number of "shoot" class pixels in the fruiting zone) but F2 (average of all cluster bounding boxes ratios between the number of "leaf abaxial" pixels and the total area of the corresponding bounding box). As a matter of fact, F3 was found

Table 7 Results for the estimation of the number of actual berries performing leave-one-out cross validation after removing each feature

Feature removed	Grapevine variety																				
	Cabernet Sauvignon			Malvasia			Muscat of Alexandria			Syrah			Tempranillo			Verdejo			Average		
	RMSE	NRMSE (%)	R²	RMSE	NRMSE (%)	R²	RMSE	NRMSE (%)	R²	RMSE	NRMSE (%)	R²	RMSE	NRMSE (%)	R²	RMSE	NRMSE (%)	R²	RMSE	NRMSE (%)	R²
F1	306	37.87	0.60	332	33.40	0.53	475	55.41	0.73	209	36.51	0.26	234	39.80	0.19	481	43.96	0.65	340	41.16	0.49
F3	225	27.85	0.62	197	19.85	0.84	339	39.59	0.80	134	23.51	0.74	190	32.26	0.38	237	21.69	0.82	221	27.46	0.70
F2	216	26.65	0.66	195	19.57	0.87	344	40.15	0.81	127	22.31	0.76	168	28.62	0.56	269	24.56	0.78	220	26.98	0.74
F4	203	25.08	0.69	225	22.65	0.81	312	36.35	0.83	107	18.77	0.85	152	25.81	0.61	243	22.24	0.83	207	25.15	0.77
F6	205	25.28	0.68	190	19.06	0.85	308	35.89	0.85	120	20.98	0.79	158	26.79	0.59	236	21.57	0.83	203	24.93	0.77
F5	205	25.28	0.69	205	20.62	0.84	311	36.25	0.84	97	17.04	0.87	149	25.36	0.64	242	22.11	0.83	202	24.44	0.78
None	194	23.94	0.72	198	19.88	0.84	294	34.32	0.86	102	17.89	0.85	150	25.44	0.63	235	21.49	0.84	195	23.83	0.79

$F1$ Number of estimated visible berries, $F2$ The average of all cluster bounding boxes ratios between the number of "leaf abaxial" pixels and the total area of the corresponding bounding box, $F3$ The number of "shoot" class pixels in the fruiting zone, $F4$ The number of "leaf adaxial" class pixels in the fruiting zone, $F5$ The ratio between F4 and the area of the fruiting zone, $F6$ The area of the fruiting zone, expressed in pixels.

Table 8 Results for the estimation of the number of actual berries performing eightfold cross validation after removing each feature

Feature removed	Grapevine variety																		Average		
	Cabernet Sauvignon			Malvasia			Muscat of Alexandria			Syrah			Tempranillo			Verdejo					
	RMSE	NRMSE (%)	R²	RMSE	NRMSE (%)	R²	RMSE	NRMS(%)E	R²	RMSE	NRMSE (%)	R²	RMSE	NRMSE (%)	R²	RMSE	NRMSE (%)	R²	RMSE	NRMSE (%)	R²
F1	307	37.88	0.60	339	34.10	0.51	480	55.97	0.72	208	36.42	0.28	236	40.03	0.19	486	44.39	0.65	343	41.46	0.49
F3	225	27.81	0.62	198	19.91	0.84	331	38.61	0.83	133	23.29	0.75	185	31.41	0.41	231	21.08	0.83	217	27.02	0.71
F2	215	26.53	0.66	197	19.82	0.87	345	40.27	0.81	121	21.22	0.79	170	28.89	0.54	267	24.40	0.79	219	26.86	0.74
F4	203	25.07	0.69	224	22.57	0.81	312	36.36	0.84	107	18.77	0.85	156	26.59	0.58	243	22.20	0.82	208	25.26	0.77
F6	206	25.49	0.68	196	19.70	0.84	316	36.89	0.85	122	21.33	0.78	161	27.28	0.56	239	21.83	0.83	207	25.42	0.76
F5	204	25.26	0.69	208	20.94	0.83	313	36.45	0.84	105	18.31	0.84	156	26.44	0.59	243	22.18	0.82	205	24.93	0.77
None	199	24.61	0.70	207	20.82	0.83	306	35.71	0.85	106	18.57	0.84	155	26.34	0.59	241	21.97	0.83	202	24.67	0.77

F1 Number of estimated visible berries, *F2* The average of all cluster bounding boxes ratios between the number of "leaf abaxial" pixels and the total area of the corresponding bounding box, *F3* The number of "shoot" class pixels in the fruiting zone, *F4* The number of "leaf adaxial" class pixels in the fruiting zone, *F5* The ratio between F4 and the area of the fruiting zone, *F6* The area of the fruiting zone, expressed in pixels.

Table 9 Results for the estimation of the number of actual berries performing leave-one-variety-out cross validation after removing each feature

Feature removed	Grapevine variety																				Average		
	Cabernet Sauvignon			Malvasia			Muscat of Alexandria			Syrah			Tempranillo			Verdejo								
	RMSE	NRMSE (%)	R²	RMSE	NRMSE (%)	R²	RMSE	NRMSE (%)	R²	RMSE	NRMSE (%)	R²	RMSE	NRMSE (%)	R²	RMSE	NRMSE (%)	R²				RMSE	NRMSE (%)	R²
F1	371	45.83	0.61	366	36.82	0.47	477	55.68	0.72	250	43.69	0.38	242	41.12	0.21	547	49.97	0.65				376	45.52	0.51
F2	211	26.03	0.68	243	24.41	0.85	341	39.83	0.82	356	62.25	0.59	174	29.63	0.57	312	28.46	0.81				273	35.10	0.72
F4	203	25.13	0.71	241	24.28	0.80	309	36.05	0.84	349	61.05	0.69	150	25.56	0.62	286	26.17	0.84				257	33.04	0.75
F5	198	24.45	0.71	227	22.87	0.83	308	35.94	0.85	290	50.78	0.74	147	24.97	0.65	282	25.77	0.84				242	30.80	0.77
F6	200	24.76	0.70	204	20.50	0.83	315	36.78	0.85	280	48.92	0.65	156	26.58	0.62	277	25.30	0.85				239	30.47	0.75
F3	217	26.78	0.65	218	21.87	0.82	324	37.82	0.84	143	25.03	0.82	184	31.26	0.43	242	22.09	0.85				221	27.47	0.74
None	195	24.14	0.72	221	22.19	0.82	310	36.13	0.85	320	55.96	0.67	149	25.24	0.64	274	25.00	0.84				245	31.44	0.76

F1 Number of estimated visible berries, *F2* The average of all cluster bounding boxes ratios between the number of "leaf abaxial" pixels and the total area of the corresponding bounding box, *F3* The number of "shoot" class pixels in the fruiting zone, *F4* The number of "leaf adaxial" class pixels in the fruiting zone, *F5* The ratio between F4 and the area of the fruiting zone, *F6* The area of the fruiting zone, expressed in pixels.

to be the least relevant feature when LOVOCV was the cross validation method (on average of all varieties). However, when F3 was removed, the three performance metrics of Syrah substantially improved (Table 9), reducing the NRMSE from 55.72% to 25.03%. As observed in Fig. 2d, the presence of shoot pixels in the fruiting zone was noticeably higher for Syrah than for the remaining varieties (Fig. 2a–f), and this could explain these findings. Also, for Syrah in the LOVOCV, the most relevant features were "the average of all cluster bounding boxes ratios between the number of "leaf abaxial" pixels and the total area of the corresponding bounding box" (F2) and "the number of "leaf adaxial" class pixels in the fruiting zone" (F4), in first and second place, respectively, while "the number of estimated visible berries" (F1) was found to be the second least influential factor in the model (Table 9). These findings seem to confirm that the whole fruiting zone, rather than exclusively the number of estimated visible berries are particularly meaningful and determinant in the assessment of the actual number of berries in Syrah, as compared to the other five varieties. As mentioned, this could be explained by the differences in the conditions of the Syrah canopy, where a slight lower occlusion was found (Fig. 2d), compared to the conditions of the canopy for the rest of the varieties (Fig. 2a–f). The characterization of the different canopy elements in the fruiting zone has to do with the occlusion phenomena affecting the berries. Berry occlusion by leaves was recently studied by Iñiguez et al. (2021). In this work, the leaf occlusion rate (berry occlusion caused by leaves) was computed using a pixel segmentation approach. When this rate increased, from low to high, the determination coefficient between the number of cluster pixels and the yield decreased from 0.77 to 0.33.

Victorino et al. (2020) evaluated the use of several canopy features for yield estimation and concluded that counting visible vine organs (as shoots, spurs, inflorescences and clusters) in non-defoliated vines could be used as auxiliary features but may not be sufficient individually for an accurate yield prediction. Their best result was achieved using the cluster projected area as a yield predictor. Moreover, the estimation of the number of bunches is a difficult task under high leaf occlusions conditions using image analysis. Sozzi et al. (2021) detected big bunches, which contribute the most to final grape yield. Other works have also addressed the problem of estimating the number of actual berries using several features extracted from RGB images. Buayai et al. (2021) estimated the number of actual berries using five features extracted from individual clusters' images in which occlusion phenomena were only due to other berries, but not to leaves or other canopy elements. In this work, the authors achieved the best performance using a random forest regression that yielded a mean absolute error of estimation of 3.79 berries per cluster.

Yield prediction

The ultimate goal of identifying and counting the number of visible and actual berries per vine is to be able to predict the grapevine yield. Towards this end, the correlation between the yield, as estimated by the model, and the final yield (actual yield) at harvest, involving all six varieties altogether, is shown in Fig. 6. Significantly strong correlations, with determination coefficients above 0.82, were found for leave-one-out (Fig. 6a) and eightfold (Fig. 6c) cross-validation methods. For LOVOCV (Fig. 6e), a lower ($R^2 = 0.76$), but still strong correlation was found. These correlations were similar to the one obtained between the number of actual berries and the actual yield (weight of the clusters) at harvest per vine, with a R^2 of 0.81 (data not shown). In terms of accuracy, the RMSE per vine achieved
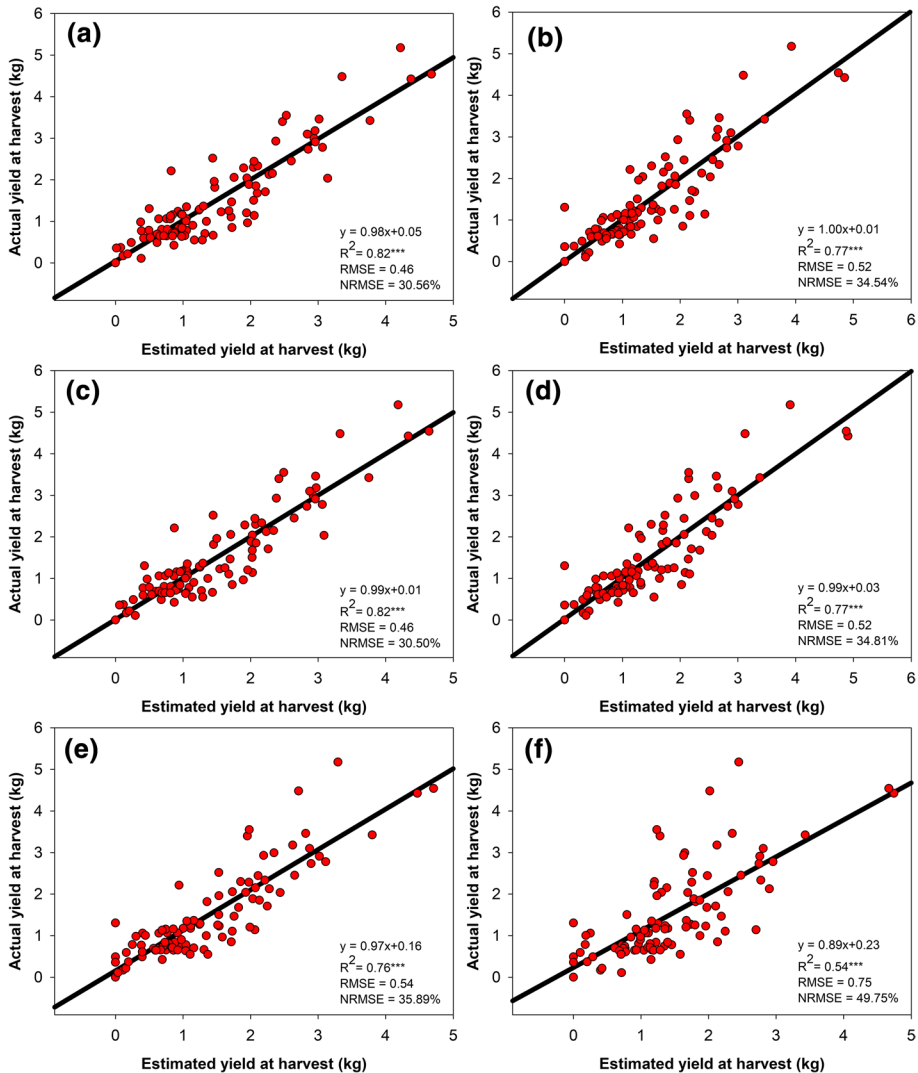
Fig. 6 Linear correlation between the estimated yield and the actual yield per vine for **a** leave-one-out, **c** eightfold and **e** leave-one-variety-out cross-validation methods and **b**, **d** and **f** the same cross-validation methods without including the berry weight as a feature in the model. Determination coefficients (R^2) were significant at $p = 0.001$ (***)

was 0.45 kg for leave-one-out, 0.46 for eightfold, and 0.54 kg for leave-one-variety-out, which corresponded to NRMSE values of 30% (Fig. 6a, c) and 36% (Fig. 6e), respectively.

The inclusion of berry weight in the yield estimation model proved to be highly valuable. Figure 6b, d and f shows the results of the yield estimation using leave-one-out, eightfold and LOVOCV without including the berry weight as a feature, respectively. As it can be observed, poorer results were obtained when the berry weight was removed from the model. For leave-one-out and eightfold results for R^2 worsened from 0.83 and 0.82 (Fig. 6a, c) to 0.77 (Fig. 6b, d) and for RMSE and NRMSE the results worsened from 0.45

and 30% (Fig. 6a, c) to 0.52 and 30% (Fig. 6b, d). The most drastic worsening in the results occurred for LOVOCV (Fig. 6f), changing from previous results (Fig. 6c) for R^2 of 0.76 to 0.54, RMSE of 0.54 to 0.75 and NRMSE of 35.89% to 49.75%. As regards the inclusion of berry weight (harvest values) in the prediction model, it could be argued that intrinsic variability of this yield component could affect yield prediction models. However, as pointed out by Clingeleffer et al. (2001) and Dunn (2010) berry weight (as a yield component) is not the most relevant factor influencing yield variation, which is mainly explained by the number of clusters (around 60–70% of yield variation), and berry number per cluster (Dunn, 2010), rather than by berry weight. Moreover, the work of Pagay and Cheng (2010) reported berry diameter variations smaller than 11% for two different cultivars at harvest time within given vineyards. Furthermore, the use of historical average values of berry weight (from longer periods of time, e.g. 10 years) would also contribute to mitigate inter seasonal fluctuations, providing a more robust indicator to be considered in yield prediction models based on computer vision and machine learning approaches.

For each individual variety, the correlation between the estimated yield and the actual yield at harvest using leave-one-out cross-validation is shown in Fig. 7. For most of the varieties, R^2 values above 0.70 were achieved, except for Syrah (Fig. 7d) and Tempranillo (Fig. 7e), which showed R^2 of 0.65 and 0.54, respectively. RSME values ranged from 0.27 kg (Tempranillo) to 0.72 kg (Malvasia). In terms of NRSME, values around 30% were yielded for four of the six varieties, while Muscat of Alexandria achieved 39.17% and Verdejo 16.47%.

These results prove that the image-based method developed in this work was useful to estimate the final yield in partially defoliated vines of several grapevine varieties in commercial vineyards, nearly two months before harvest. Dunn and Martin (2004) suggested a digital image analysis to predict vineyard yield. These authors observed that the ratio of fruit pixels to total image pixels explained 85% of the variation in yield in progressively de-fruited vines. Following a similar approach, Diago et al. (2012) employed a Mahalanobis colour segmentation method to extract cluster pixels and linear regressions to estimate the yield from the cluster pixels, achieving R^2 values up to 0.73 in progressively de-fruited and defoliated vines, which are similar results to the ones presented in Fig. 6. Liu et al. (2017) attempted yield estimation based on shoot counting, very soon after budbreak, and reported absolute yield estimation errors ranging between 1.18% and 36.02%, which are results near to the NRMSE values presented in Fig. 6 (values near 30%). In another work, Font et al. (2015) using pixel-based segmentation approaches yielded a relative yield error of 16% for 25 non-occluded clusters on images from red grapes acquired using a vehicle and artificial illumination at night-time, these results were lower to the ones presented in this work but also their error metrics were calculated over individual clusters instead of full vines as in this work. On the other hand, Aquino et al. (2018) achieved similar results to the ones presented in this work, with R^2 between the number of detected berries and the actual yield of 0.74 and a R^2 between the predicted visible yield, calculated from the number of detected berries, and the actual yield of 0.78, although in this work the fruiting zone was fully defoliated, avoiding any occlusion phenomenon. Likewise, Nuske et al. (2014b) reported R^2 values in the range between 0.60 and 0.73 (depending on the phenological stage and grapevine variety) for the correlation between the number of detected berries and actual harvest weight. The models in Nuske et al. (2014a, 2014b) were built individually for each variety while the yield estimation model in the present study was built from six different grapevine varieties overall. Nuske et al. (2014a, 2014b) acquired images from intensively defoliated vines in the fruiting zone, where occlusions were mostly due to other berries, but not to leaves or shoots. In the work of Nuske et al. (2014a) overall yield-per
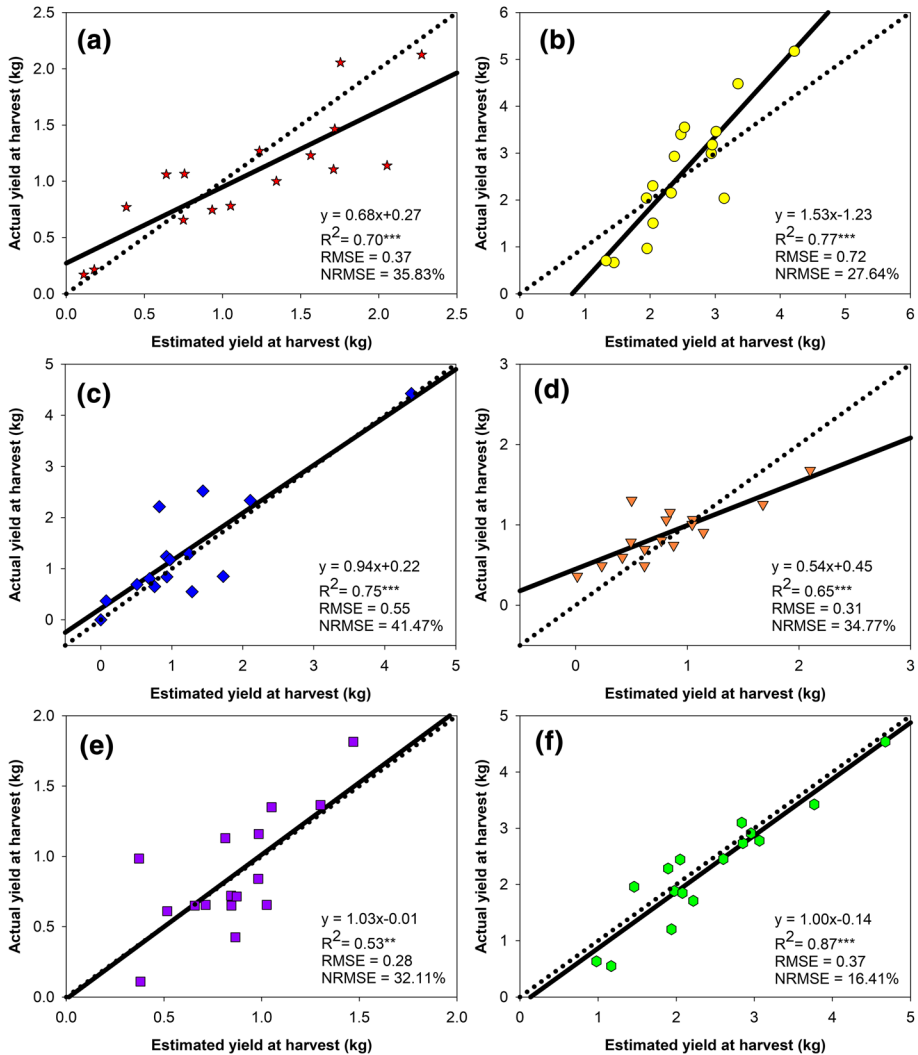
Fig. 7 Linear correlation between the estimated yield and the actual yield per vine including the berry weight as feature in the model and using leave-one-out cross validation for each grapevine variety: **a** Cabernet Sauvignon, **b** Malvasia, **c** Muscat of Alexandria, **d** Syrah, **e** Tempranillo and **f** Verdejo. Determination coefficients (R^2) were significant at $p = 0.01$ (**) and $p = 0.001$ (***)

vine prediction errors based on imaged-based berry counting in Traminette and Riesling varieties ten days prior to harvest varied between 15 and 20%. In a more recent work, Victorino et al. (2020) achieved R^2 values of 0.35, 0.64 and 0.55 (depending on grapevine variety) between the yield and the visible cluster area in non-defoliated vines but no yield estimation errors were provided, in contrast to the results presented in this work in Figs. 6 and 7.

Conventional methods for yield predicting used systematic, manual and labour-demanding procedures under field conditions (Dunn & Martin, 2003; Clingeleffer et al.,

2001). Martin and Dunn (2003) described particular manual protocols to achieve a 15% tolerance in yield and yield components' prediction, but its achievement required intensive manual counting and increased accuracy was expected when estimation was conducted closer to harvest time. In this regard, although prediction errors below 10–15% would be desirable, and certainly very useful for the wine industry, it is expected that the prediction metrics following the vision-approach of the present work would improve should a higher number of data be used to build and validate the models. Moreover, the inclusion of data coming for vineyards of increased canopy vigour variability, would enlarge the range of the data used for model building and validation. This factor could also lead to improved model performance. At this stage, it could be recognized that this study may be considered an initial approach that should deserve further development with substantially increased data from different cultivars and vigour conditions.

On the other hand, comparing the automated image-developed method in this work with conventional methods, it can be pointed out that the method developed in this work was non-destructive and applicable in an automated way to a very large number of vines o even to a whole vineyard using a mobile sensing platform. Thus, the image-based method described in the present study enables the increase of the representativeness of the sampled vines in commercial vineyards.

Finally, it is worth mentioning that the developed method was able to estimate yield near 60 days before harvest. It was applied in partially defoliated vines in commercial vineyards. The potential of addressing the spatial variability of berry number per vine, and its subsequent implication in grapevine yield is very informative and relevant, particularly within a precision viticulture framework. The automated image acquisition performed by a mobile sensing platform enables estimating the berry number and the yield on a large number of vines, which leads to a better modelling of the spatial variability of the vineyard and a more precise prediction of the total yield at harvest.

Conclusions

The results presented in this work prove the capability of the new method to estimate the number of actual berries and yield in different commercial grapevine varieties. The algorithm based on deep learning and computer vision was able to quantify the visible berries and extract canopy features. The cross validation methods proved that the model was able to predict the number of actual berries and yield on not only on varieties included in the model but also in other grapevine varieties.

The use of a mobile sensing platform eases the industrial applicability of this method in commercial vineyards. In addition, the new method allows an early yield estimation two months prior to harvest, which is highly valuable for wine industry, in order to understand and anticipate harvest logistics and pricing.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Aquino, A., Millan, B., Diago, M.-P., & Tardaguila, J. (2018). Automated early yield prediction in vineyards from on-the-go image acquisition. *Computers and Electronics in Agriculture, 144*, 26–36. https://doi.org/10.1016/j.compag.2017.11.026

Arnó, J., Martínez-Casasnovas, J. A., Uribeetxebarria, A., Escolà, A., & Rosell-Polo, J. R. (2017). Comparing efficiency of different sampling schemes to estimate yield and quality parameters in fruit orchards. *Advances in Animal Biosciences, 8*(2), 471–476. https://doi.org/10.1017/S2040470017000978

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(12), 2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615

Ballesteros, R., Intrigliolo, D. S., Ortega, J. F., Ramírez-Cuesta, J. M., Buesa, I., & Moreno, M. A. (2020). Vineyard yield estimation by combining remote sensing, computer vision and artificial neural network techniques. *Precision Agriculture, 21*(6), 1242–1262. https://doi.org/10.1007/s11119-020-09717-3

Bishop, C. M. (2006). *Pattern recognition and machine learning (information science and statistics)*. Springer.

Bramley, R. G. V., Ouzman, J., Trought, M. C. T., Neal, S. M., & Bennett, J. S. (2019). Spatio-temporal variability in vine vigour and yield in a Marlborough Sauvignon Blanc vineyard. *Australian Journal of Grape and Wine Research, 25*(4), 430–438. https://doi.org/10.1111/ajgw.12408

Buayai, P., Saikaew, K. R., & Mao, X. (2021). End-to-end automatic berry counting for table grape thinning. *IEEE Access, 9*, 4829–4842. https://doi.org/10.1109/ACCESS.2020.3048374

Carrillo, E., Matese, A., Rousseau, J., & Tisseyre, B. (2016). Use of multi-spectral airborne imagery to improve yield sampling in viticulture. *Precision Agriculture, 17*(1), 74–92. https://doi.org/10.1007/s11119-015-9407-8

Chinchor, N. (1992). *MUC-4 evaluation metrics*. In *Proceedings of the 4th conference on message understanding* (pp. 22–29). https://doi.org/10.3115/1072064.1072067

Clingeleffer, P. R., Martin, S., Krstic, M., & Dunn, G. M. (2001). Crop development, crop estimation and crop control to secure quality and production of major wine grape varieties. In *A national approach: Final report to grape and wine research & development corporation*. Grape and Wine Research & Development Corporation.

Coombe, B. G. (1995). Growth stages of the grapevine: Adoption of a system for identifying grapevine growth stages. *Australian Journal of Grape and Wine Research, 1*(2), 104–110. https://doi.org/10.1111/j.1755-0238.1995.tb00086.x

Cubero, S., Aleixos, N., Moltó, E., Gómez-Sanchis, J., & Blasco, J. (2011). Advances in machine vision applications for automatic inspection and quality evaluation of fruits and vegetables. *Food and Bioprocess Technology, 4*(4), 487–504. https://doi.org/10.1007/s11947-010-0411-8

Deng, L., & Yu, D. (2014). *Deep learning: Methods and applications (msr-TR-2014-21). Microsoft.* https://www.microsoft.com/en-us/research/publication/deep-learning-methods-and-applications/

Diago, M. P., Aquino, A., Millan, B., Palacios, F., & Tardaguila, J. (2019). On-the-go assessment of vineyard canopy porosity, bunch and leaf exposure by image analysis. *Australian Journal of Grape and Wine Research, 25*(3), 363–374. https://doi.org/10.1111/ajgw.12404

Diago, M. P., Correa, C., Millán, B., Barreiro, P., Valero, C., & Tardaguila, J. (2012). Grapevine yield and leaf area estimation using supervised classification methodology on RGB images taken under field conditions. *Sensors, 12*(12), 16988–17006. https://doi.org/10.3390/s121216988

Diago, M. P., Krasnow, M., Bubola, M., Millan, B., & Tardaguila, J. (2016). Assessment of vineyard canopy porosity using machine vision. *American Journal of Enology and Viticulture, 67*(2), 229–238. https://doi.org/10.5344/ajev.2015.15037

Dunn, G. M. (2010) Yield forecasting. Australian Government: Grape and wine research and development corporation. *Fact sheet. June 2010.* https://www.wineaustralia.com/getmedia/5304c16d-23b3-4a6f-ad53-b3d4419cc979/201006_Yield-Forecasting.pdf

Dunn, G. M., & Martin, S. R. (2000). Spatial and temporal variation in vineyard yields. In *Proceedings of the fifth international symposium on cool climate viticulture & oenology. Precision management workshop* (pp. 1–4). Romsey: Cope Williams Winery.

Dunn, G. M., & Martin, S. R. (2003). The current status of crop forecasting in the Australian wine industry. In *ASVO Seminar Series: Grapegrowing at the Edge* (pp. 4–8).

Dunn, G. M., & Martin, S. R. (2004). Yield prediction from digital image analysis: A technique with potential for vineyard assessments prior to harvest. *Australian Journal of Grape and Wine Research, 10*(3), 196–198. https://doi.org/10.1111/j.1755-0238.2004.tb00022.x

Font, D., Tresanchez, M., Martínez, D., Moreno, J., Clotet, E., & Palacín, J. (2015). Vineyard yield estimation based on the analysis of high resolution images obtained with artificial illumination at night. *Sensors, 15*(4), 8284–8301. https://doi.org/10.3390/s150408284

Grimm, J., Herzog, K., Rist, F., Kicherer, A., Töpfer, R., & Steinhage, V. (2019). An adaptable approach to automated visual detection of plant organs with applications in grapevine breeding. *Biosystems Engineering, 183*, 170–183. https://doi.org/10.1016/j.biosystemseng.2019.04.018

Hacking, C., Poona, N., & Poblete-Echeverria, C. (2020). Vineyard yield estimation using 2-D proximal sensing: A multitemporal approach. *OENO One, 54*(4), 793–812. https://doi.org/10.20870/oeno-one.2020.54.4.3361

Hall, A., Lamb, D. W., Holzapfel, B. P., & Louis, J. P. (2011). Within-season temporal variation in correlations between vineyard canopy and winegrape composition and yield. *Precision Agriculture, 12*(1), 103–117. https://doi.org/10.1007/s11119-010-9159-4

Íñiguez, R., Palacios, F., Barrio, I., Hernández, I., Gutiérrez, S., & Tardaguila, J. (2021). Impact of leaf occlusions on yield assessment by computer vision in commercial vineyards. *Agronomy, 11*(5), 1003. https://doi.org/10.3390/agronomy11051003

Jones, D. R. (2001). A taxonomy of global optimization methods based on response surfaces. *Journal of Global Optimization, 21*(4), 345–383. https://doi.org/10.1023/A:1012771025575

Koen, B. V. (1988). Toward a definition of the engineering method. *European Journal of Engineering Education, 13*(3), 307–315. https://doi.org/10.1080/03043798808939429

Krstic, M. P., Welsh, M. A., & Clingeleffer, P. R. (1998). Variation in Chardonnay yield components between vineyards in a warm irrigated region. In R. J. Blair, A. N. Sas, P. F. Hayes, & P. B. Hoj (Eds.), *Precision agriculture* (pp. 269–270). AWRI.

Laurent, C., Oger, B., Taylor, J. A., Scholasch, T., Metay, A., & Tisseyre, B. (2021). A review of the issues, methods and perspectives for yield estimation, prediction and forecasting in viticulture. *European Journal of Agronomy, 130*, 126339. https://doi.org/10.1016/j.eja.2021.126339

Li, T., Hao, X., Kang, S., & Leng, D. (2017). Spatial variation of winegrape yield and berry composition and their relationships to spatiotemporal distribution of soil water content. *American Journal of Enology and Viticulture, 68*(3), 369–377. https://doi.org/10.5344/ajev.2017.16067

Liu, S., Cossell, S., Tang, J., Dunn, G., & Whitty, M. (2017). A computer vision system for early stage grape yield estimation based on shoot detection. *Computers and Electronics in Agriculture, 137*, 88–101. https://doi.org/10.1016/j.compag.2017.03.013

Liu, S., Li, X., Wu, H., Xin, B., Tang, J., Petrie, P. R., & Whitty, M. (2018). A robust automated flower estimation system for grape vines. *Biosystems Engineering, 172*, 110–123. https://doi.org/10.1016/j.biosystemseng.2018.05.009

Liu, S., Marden, S., & Whitty, M. (2013). Towards automated yield estimation in viticulture. In *Proceedings of the Australasian conference on robotics and automation*, (Vol. 24, pp. 2–6), Sydney, Australia.

Liu, S., & Whitty, M. (2015). Automatic grape bunch detection in vineyards with an SVM classifier. *Journal of Applied Logic, 13*(4), 643–653. https://doi.org/10.1016/j.jal.2015.06.001

Liu, S., Whitty, M., & Cossell, S. (2015). A lightweight method for grape berry counting based on automated 3D bunch reconstruction from a single image. In *ICRA, International conference on robotics and automation (IEEE), workshop on robotics in agriculture* (p. 4).

Liu, S., Zeng, X., & Whitty, M. (2020). A vision-based robust grape berry counting algorithm for fast calibration-free bunch weight estimation in the field. *Computers and Electronics in Agriculture, 173*, 105360. https://doi.org/10.1016/j.compag.2020.105360

Martin, S., Dunstone, R., & Dunn, G. (2003). *How to forecast wine grape deliveries using grape forecaster excel workbook version 7*. Department of Primary Industries.

Mockus, J., Tiesis, V., & Zilinskas, A. (2014). The application of Bayesian methods for seeking the extremum. *Global Optimization, 2*, 117–129.

Monga, T. (2018). Estimating vineyard grape yield from images. In E. Bagheri & J. C. K. Cheung (Eds.), *Advances in artificial intelligence* (pp. 339–343). Springer.

Nuske, S., Gupta, K., Narasimhan, S., & Singh, S. (2014a). Modeling and calibrating visual yield estimates in vineyards. In K. Yoshida & S. Tadokoro (Eds.), *Field and service robotics: Results of the 8th international conference* (pp. 343–356). Springer. https://doi.org/10.1007/978-3-642-40686-7_23

Nuske, S., Wilshusen, K., Achar, S., Yoder, L., Narasimhan, S., & Singh, S. (2014b). Automated visual yield estimation in vineyards. *Journal of Field Robotics, 31*(5), 837–860. https://doi.org/10.1002/rob.21541

Oger, B., Laurent, C., Vismara, P., & Tisseyre, B. (2021a). Is the optimal strategy to decide on sampling route always the same from field to field using the same sampling method to estimate yield? *OENO One, 55*(1), 133–144. https://doi.org/10.20870/oeno-one.2021a.55.1.3334

Oger, B., Vismara, P., & Tisseyre, B. (2021). Combining target sampling with within field route-optimization to optimise on field yield estimation in viticulture. *Precision Agriculture, 22*(2), 432–451. https://doi.org/10.1007/s11119-020-09744-0

Pagay, V., & Cheng, L. (2010). Variability in berry maturation of Concord and Cabernet Franc in a cool climate. *American Journal of Enology and Viticulture, 61*, 61–67.

Palacios, F., Bueno, G., Salido, J., Diago, M. P., Hernández, I., & Tardaguila, J. (2020). Automated grapevine flower detection and quantification method based on computer vision and deep learning from on-the-go imaging using a mobile sensing platform under field conditions. *Computers and Electronics in Agriculture, 178*, 105796. https://doi.org/10.1016/j.compag.2020.105796

Palacios, F., Melo-Pinto, P., Diago, M. P., Iñiguez, R., & Tardaguila, J. (2021) Assessing number of grapevine berries using linear methods and machine learning. In *Proceedings of the 13th European conference on precision agriculture*, ECPA 2021, Budapest, Hungary, 19–22 July 2021.

Palacios, F., Melo-Pinto, P., Diago, M. P., & Tardáguila, J. (2022). Deep learning and computer vision for assessing the number of total berries and yield in commercial vineyards. *Biosystems Engineering, 218*, 175–188. https://doi.org/10.1016/j.biosystemseng.2022.04.015

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *International conference on learning representations*.

Sozzi, M., Cantalamessa, S., Cogato, A., Kayad, A., & Marinello, F. (2021). Grape yield spatial variability assessment using YOLOv4 object detection algorithm. In *Proceedings of the 13th European conference on precision agriculture, ECPA 2021*, (pp. 193–198). Budapest, Hungary, 19–22 July 2021. https://doi.org/10.3920/978-90-8686-916-9_22

Sozzi, M., Kayad, A., Tomasi, D., Lovat, L., Marinello, F., & Sartori, L. (2019). Assessment of grapevine yield and quality using a canopy spectral index in white grape variety. In *Proceedings of the 12th European conference on precision agriculture, ECPA 2019*, (pp. 181–186). Montpellier, France, 8–11 July 2019. https://doi.org/10.3920/978-90-8686-888-9_21

Taylor, J., Tisseyre, B., Bramley, R., & Reid, A. (2005). *A comparison of the spatial variability of vineyard yield in European and Australian production systems* (pp. 907–914). Wageningen Academic Publishers.

Taylor, J. A., Tisseyre, B., & Leroux, C. (2019). A simple index to determine if within-field spatial production variation exhibits potential management effects: Application in vineyards using yield monitor data. *Precision Agriculture, 20*(5), 880–895. https://doi.org/10.1007/s11119-018-9620-3

Tian, H., Wang, T., Liu, Y., Qiao, X., & Li, Y. (2020). Computer vision technology in agricultural automation: A review. *Information Processing in Agriculture, 7*(1), 1–19. https://doi.org/10.1016/j.inpa.2019.09.006

Tisseyre, B., Leroux, C., Pichon, L., Geraudie, V., & Sari, T. (2018). How to define the optimal grid size to map high resolution spatial data? *Precision Agriculture, 19*(5), 957–971. https://doi.org/10.1007/s11119-018-9566-5

van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture, 177*, 105709. https://doi.org/10.1016/j.compag.2020.105709

Victorino, G. F., Braga, R., Santos-Victor, J., & Lopes, C. M. (2020). Yield components detection and image-based indicators for non-invasive grapevine yield prediction at different phenological phases. *OENO One, 54*(4), 833–848. https://doi.org/10.20870/oeno-one.2020.54.4.3616

Xin, B., Liu, S., & Whitty, M. (2020). Three-dimensional reconstruction of *Vitis vinifera* L. cvs Pinot Noir and Merlot grape bunch frameworks using a restricted reconstruction grammar based on the stochastic L-system. *Australian Journal of Grape and Wine Research, 26*(3), 207–219. https://doi.org/10.1111/ajgw.12444

Zabawa, L., Kicherer, A., Klingbeil, L., Töpfer, R., Kuhlmann, H., & Roscher, R. (2020). Counting of grapevine berries in images via semantic segmentation using convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing, 164*, 73–83. https://doi.org/10.1016/j.isprsjprs.2020.04.002