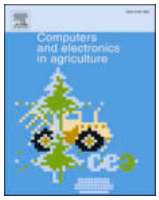
Computers and Electronics in Agriculture 225 (2024) 109277

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag



Original papers

Comparison of different computer vision methods for vineyard canopy detection using UAV multispectral images

Massimo Vincenzo Ferro [a,*], Claus Grøn Sørensen [b], Pietro Catania [a]

[a] University of Palermo, Department of Agricultural, Food and Forest Sciences (SAAF), viale delle Scienze ed. 4, Palermo 90128, Italy
[b] Department of Electrical and Computer Engineering, Aarhus University, Aarhus 8000, Denmark

ARTICLE INFO

Keywords:
Deep learning
OBIA
Vitis vinifera L.
Unmanned aerial vehicles
NDVI

ABSTRACT

In viticulture, the rapid and accurate acquisition of canopy spectral information through ultra-high spatial resolution imagery is increasingly demanded for decision support. The prevalent practice involves creating vigor maps using spectral data obtained from pure vine canopy pixels. Object-Based Image Analysis (OBIA) among conventional methods exhibits a reasonable efficiency in canopy classification due to its feature extraction capabilities. In recent years, deep learning (DL) techniques have demonstrated significant potential in orchard monitoring, leveraging their ability to automatically learn image features. This study assessed the performance of different methodologies, including Mask R-CNN, U-Net, OBIA and unsupervised methods, in identifying pure canopy pixels. The effectiveness of shadow and background detection methods and the impact of misclassified pixels on NDVI were compared. Results were compared with agronomic surveys conducted during the 2021 and 2022 growing seasons, focusing on two distinct phenological stages (BBCH65-BBCH85). Mask R-CNN and U-Net exhibited superior performance in terms of Overall Accuracy (OA), F1-score, and Intersection Over Union (IoU). Among OBIA methods, the Gaussian Mixture Model (GMM) proved to be the most effective classifier for canopy segmentation, and Support Vector Machine (SVM) also demonstrated reasonable stability. Conversely, Random Forest (RF) and K-Means yielded lower accuracy and higher error rates. As a result of the limited accuracy, it is noted for vineyard rows with low vigor canopies that NDVI was overestimated, while for high vigor canopies NDVI was underestimated.

Significantly improved determination coefficients were observed for the comparison between Total Leaf Area (TLA) and NDVI data derived from Mask R-CNN and U-Net. Positive correlations were also found with NDVI data from GMM and SVM algorithms. Regarding leaf chlorophyll (Chl) and NDVI correlations, Mask R-CNN and U-Net methods showed superior performance. Additionally, the relationship between TLA and projected canopy area (PCA) was significantly better represented by U-Net and Mask R-CNN, while PCA was not recommended for estimating chlorophyll content. This investigation establishes that improved vine canopy delimitation contributes to improved vineyard vigour monitoring, providing winegrowers with more accurate and reliable agronomic information for management decisions.

1. Introduction

Viticulture is increasingly focusing on technologies and strategies to enhance vineyard management efficiency, improving yield and grape quality, while also sustaining sustainability and resilience.

Digital technologies and Information and Communication Technology (ICT) applications involve sensor data collection on soil, climate, and plant conditions, enabling subsequent data processing and decision support. Among these technologies, unmanned aerial vehicles (UAV) operating at low altitude acquire high-resolution images (HRI), both in RGB format and in multispectral/hyperspectral or thermal infrared format. These capabilities enhance vineyard monitoring and provide detailed information on vegetation phenotypic traits (Sozzi et al., 2020). Recently, the study of image processing techniques are expanding the application of vineyard monitoring using UAV, enabling the creation of detailed vigor maps that assess the spatial variability of agronomic parameters using vegetation indices (VIs) (Ferro et al., 2023; Moghimi et al., 2020). VIs are computed by considering spectral values of solely

* Corresponding author.
E-mail addresses: massimovincenzo.ferro@unipa.it (M.V. Ferro), cclaus.soerensen@ece.au.dk, claus.soerensen@ece.au.dk (C.G. Sørensen), pietro.catania@unipa.it (P. Catania).

https://doi.org/10.1016/j.compag.2024.109277
Received 7 February 2024; Received in revised form 9 July 2024; Accepted 24 July 2024
Available online 1 August 2024

vineyard canopy pixels. Currently, research is focused on finding new methods of UAV image analysis for vineyards trained with vertical trellis systems (VSP), aiming to isolate vineyard row pixels from the background. The reflectance of the soil is influenced by features such as colour, moisture and roughness (Bannari et al., 1995). The authors affirm that if the soil is very bright, the reflectance in red band of the spectrum increase; consequently, VIs can be affected, especially those based on the red band (e.g. NDVI). Conversely, an increase in soil moisture leads to a general decrease in reflectance. Shadow depends on the density of the vegetation and the geometric characteristics of the canopy, as well as the rows orientation in relation to the sun. In VSP vineyards, shading has a significant effect on canopy NDVI and is dependent on both solar elevation and the horizontal angle of incidence relative to the row orientation (relative azimuth)(Towers and Poblete-Echeverría, 2021). The intensity of shadow is influenced by the intensity of the light source and the amount of blocked light. Several studies have examined how shaded ground pixels and partially shaded canopy pixels can influence pure pixel information due to lower reflectance (Ferreira et al., 2018). This effect is mainly reflected in the digital number (DN) values, where shaded pixels have significantly lower values than sunlit pixels. The image analysis process consists of several steps and can be performed with different methodologies, of which a central process involves canopy segmentation and classification. Vineyard row segmentation methods can be carried out by pixel-based or object-based methods. Within the pixel-based category, threshold-based methods can be used to segment crops by setting a certain threshold value that exploits colour intensity or by considering the reflectance value for pixel (Ferreira et al., 2018; Pádua et al., 2018). This category includes unsupervised clustering algorithms such as K-Means that do not consider the spatial relationships between pixels but provides for assigning them to a cluster based on similarity (Cinat et al., 2019). In a study conducted in a vineyard, it was observed that clustering methods sometimes exhibit inferior performance compared to other machine learning (ML) methods widely employed for object-based image analysis (OBIA) (Poblete-Echeverría et al., 2017). These methods have provided satisfactory results in analysing images from vineyard cultivated using VSP. Aboutalebi et al., (2019) demonstrated that supervised classification outperformed unsupervised methods and highlighted how the presence of shadows significantly influences the calculation of NDVI in a vine canopy.

These mentioned methods are considered to be effective for canopy segmentation, mainly when the edges between regions are clearly delineated; however, they can be sensitive to noise in images, and moreover, it is difficult to choose an appropriate threshold value (Ferro and Catania, 2023). OBIA include several supervised classification algorithms such as Random Forest (RF) and Support vector machines (SVM) that exploit an iterative approach to determine canopies (Modica et al., 2021). Pádua et al., (2022) applied a segmentation approach (OBIA) and achieved accurate classification results for soil, canopies, and shadow pixels. Among the supervised type algorithms, the Gaussian Mixture Model (GMM) provided accurate classification of high-dimensional remote sensing images, using low computation time (Lagrange et al., 2017). This algorithm was used for vegetation classification by improving the accuracy of the model, demonstrating that GMM can effectively group together pixels with similar spectral features (Zhu et al., 2022).

OBIA methods are more complicated than pixel-based methods, but they can generate an improved segmentation result by leveraging different ML classifiers for extraction of spectral and textural features (Jiménez-Brenes et al., 2019). In addition to the methods described above, there are recent applications of classification by deep learning (DL) for UAV images based on semantic and instance segmentation (Osco et al., 2021; Ye et al., 2023). In the first case, object identification and classification are performed by assigning pixels to a predetermined semantic class, conversely, instance segmentation identifies and separates objects in the image based on their identity. Semantic segmentation can identify categories but not individual objects within the same category, constraining canopy image analysis, especially with complex images and specific details. These methods employ the architecture of a convolutional neural network (CNN) trained to perform feature extraction using band combination for canopy segmentation. This process is applied by the Mask R-CNN method, which identifies and delineates the boundary of objects in an image at the pixel level (He et al., 2017). This method has been widely used in remote sensing by UAV, for example, in the context of horticulture, Lucena et al., (2022) applied it to identify and to delineate the canopies of high-density orange tree rows, exploiting canopy height model and RGB images data as input, with optimal overall accuracy. Hao et al., (2021) used different spectral band combinations of UAV images and detect tree crowns and classify trees according to height, with significant positive Intersection over Union (IoU) results. Safonova et al., (2021) have explored the application of Mask R-CNN for the segmenting of olive tree canopies from shadows, aiming to accurately estimate the volume of individual trees. In the context of semantic segmentation, Osco et al., (2021) applied the U-Net neural network in orchard row segmentation studies using multispectral UAV images, obtaining an F1 score of 0.94. Sahin et al., (2023) trained U-Net to perform multi-class segmentation, reporting IoU values for soil, crop, and weed classes of 0.99, 0.90, and 0.75, respectively. U-Net is recognized for its effectiveness in managing class imbalances within the training dataset. In images acquired via UAV, it is common for the canopy class to be under-represented. This is a consequence of the constrained and limited size of the canopy, particularly in vines trained with the VSP system. However, owing to its downsampling and upsampling architecture, U-Net can overcome this issue and achieve significant accuracy. The model was also tested on orthomosaic images for automatic detection and localization of orchard tree canopy under various growing conditions, different growing seasons, and various weed cover levels, obtaining reasonably good analysis accuracy (Anagnostis et al., 2021). Barros et al., (2022) have conducted an interesting study about the application of semantic segmentation methods in vineyard, comparing the variation in performance of DL networks when the input images used change, and define that using RGB+NIR images improve segmentation performance. Similar results were also obtained for OBIA segmentation methods in which higher performance is obtained when using both sensor features (Pádua et al., 2022).

Recent studies have demonstrated that utilizing high-resolution images obtained through UAV enables a more precise evaluation of chlorophyll content (Chl) in vineyards compared to lower-resolution images (Caruso et al., 2023). The more accurate representation of vine physiological activity is achieved by considering the average VIs of solely canopy pixels. Hence, the occurrence of noise factors due to soil reflectance or shaded pixels could affect the accuracy of physiological parameters estimation. Therefore, identifying detection methods that enhance the exclusive segmentation of canopy pixels reduces interference and optimizes the estimation of biochemical parameters (Campos et al., 2021). Zhang et al., (2024) quantitatively evaluated the impact of orchard canopy shading, finding that the removal of canopy shadows significantly improved the accuracy of LAI and leaf chlorophyll content estimation.

Limited research exists on assessing shadows and background components' impact on canopy spectral response, while DL methods' potential in agriculture requires further exploration for vineyard canopy segmentation. Specifically, there has been limited investigation into evaluating the benefits and drawbacks of DL methods compared to traditional OBIA approaches for canopy segmentation and classification process.

In recent years, the use of multispectral cameras by UAVs has been widely applied for vineyard surveys, and in order to provide insight into different methods of high-resolution image segmentation, a comparison of computer vision (CV) methods is proposed. The objective of this study was to compare the efficiency and accuracy of OBIA, K-means and DL

segmentation techniques applied to grapevine canopy at different phenological stages. The effects of segmentation methods on canopy NDVI were evaluated by comparing them with Projected Canopy Area (PCA) and leaf chlorophyll (Chl) parameters, aiming to enhance the estimation of vine biometric parameters.

2. Materials and methods

2.1. Study site

This study was conducted in a vineyard in Mediterranean area (Fig. 1), in the region of Sicily (Italy). The experiment was carried out between 2021 and 2022, in a vineyard established in 2008 with *Vitis vinifera* L. cv. Catarratto. The vineyard has a total area of 8.2 ha and a perimeter of 1162 m. Following the methodology of an on-farm experiment (37°55′12″N; 13°04′27″E), the site was selected as situated within the Bianco Alcamo Protected Designation of Origin (PDO) zone, given that this area is very suitable for the cultivation of grapevine. Vineyard planting layout is 2.4 m between rows and 1.0 m within rows (4,170 plants ha^{-1}). The rows are oriented in the NE-SW direction and the vineyard is located at an average altitude of 350 m above sea level and in an area with a hill-type orography. The elevation variability within the vineyard ranges from a maximum elevation of 367 m a.s.l. to a minimum elevation of 335 m a.s.l., with an average slope value of 9.8 %.

For the shaping of the canopy, the vines are trained through a vertically positioned shoot system (VSP). The pruning system is a double spurred cordon with two buds per spur. The vineyard trellis has an overall height of 1.70 m in which the first supporting wire for the plants is positioned 0.70 m above ground level and the remaining two are positioned with 0.50 m between them. These wires are maintained by galvanized bearing poles placed at 6 m apart from each other.

Vineyard management was conducted through standard agronomic practices applied uniformly across the field. Traditionally, the vineyard was subjected to surface tillage sessions during each year to control weeds and mitigate water evaporation.

2.2. UAV image acquisition

The UAV DJI™ Phantom 4 multispectral (P4, SZ DJI Technology Co., Ltd., Shenzhen, Guangdong, China), equipped with one RGB sensor for imaging on visible light and five monochrome sensors for multispectral image acquisition was used.

Two field survey campaigns were executed from June until the end of August, for two consecutive years, from bloom to harvest, covering the most important phenological stages (Table 1).

In the year 2021, the first flight was performed on June 18 (F1_21) and this timing corresponds to the flowering phenological stage of vineyard (BBCH65), and the second flight was performed on August 26 (F2_21) and this timing corresponds to the berry softening phenological stage (BBCH 85). At the same phenological stages, flights were performed and replicated for the year 2022, the first flight was performed on June 21 (F1_22) and the second flight was performed on August 28 (F2_22). The specific flight parameters are shown in Table 2.

Camera acquired the photos automatically, at an interval of 1 s, and the images were stored in TIFF format. Images with a resolution of 1600 * 1300pixels were captured through this camera. The UAV was programmed to traverse the field following a predetermined path using DJI GO Pro software, with the aim of capturing images of the entire on-farm plot. A total of 258 images were captured per spectral band, for a total of 1290 captures. The UAV images were acquired at a specific time of day, i.e., between 12:00 noon and 1:00p.m., in sunny or non-cloudy weather, considering solar noon for each survey. The flight height was set at 70 m relative to the take-off point to balance the sharpness of the field of view with the total flight duration, which was 19 min for each survey session. Ground conditions during data collection were characterized by the absence of weeds. Furthermore, a reflectance calibration panel (LABSPHERE INC., North Sutton, US) was employed. Using this technique, regulation of the sensors was performed to provide accurate measurements of the crops' reflectance and surrounding surface. This is feasible because the reflectance characteristics of the panel are pre-determined. During UAV flights, the use of calibration panels allows balancing environmental variations such as illumination, ensuring accurate data reflecting surface conditions and enabling instrument stability verification over time.

Table 1
UAV surveys and corresponding phenological stages.

Flight	Phenological stages
F1_21	Full flowering (BBCH65)
F2_21	Softening of berries (BBCH 85)
F1_22	Full flowering (BBCH65)
F2_22	Softening of berries (BBCH 85)

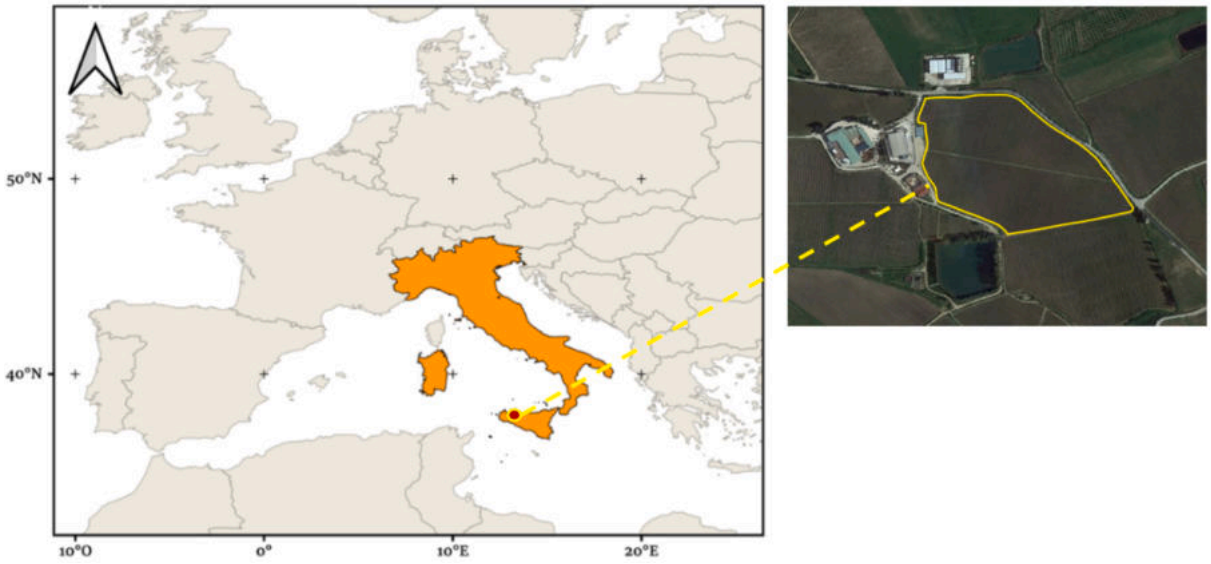


Fig. 1. Location of the experimental vineyard plot in the Mediterranean area.

Table 2
Flight parameter settings.

Parameter	Value/Method
Flight altitude	70 *m*
Flight speed	6,5 m/s
Shooting mode	Timed shooting
Pitch gimbal	−90°
Side overlap rate	60 %
Forward overlap rate	60 %
Ground sampling distance	0.37 *m*
Flight duration	19 min

2.3. UAV image pre-processing

Agisoft Metashape Professional Edition software (Agisoft LLC, St. Petersburg, Russia) version 1.7.2 was used for image pre-processing, where the first step was to apply radiometric corrections for the images, removing any problems such as vignetting, dark pixel offsets, and especially to convert the raw images to radiance and then to reflectance space. Through the software, the point cloud was generated and then the vineyard images were converted into a Digital Orthophoto Map (DOM) (Fig. 2). To increase the accuracy of the DOM, pre-measured ground control points (GCPs) were imported via a GNSS receiver. As the resolution of the DOM is too large, the images were cropped to satisfy the appropriate size for running DL segmentation methods. Using QGis 3.20 software, it was possible to split the DOM into smaller subimages with a standard size of 240 x 240 pixels such dimensions reduce the computational overhead by contributing to improved performance of the segmentation networks. For each survey, the image dataset was divided by a random assignment process into three specific subsets and then a training set (80 %), a validation set (10 %) and a test set (10 %) were generated. For example, through the DOM splitting process, for the survey conducted in F1_21, 835 sub-images were obtained, which were split into: 668 images used for training, 84 for validation and 83 for testing. The annotation was carried out using an open source software Labelme 3.6 (Torralba et al., 2010), which allows for manual image annotation. Class annotation tools were employed, and polygons encompassing the vine canopy, soil, and shadow were generated. At the end of the process, these annotations were saved in a JSON (JavaScript Object Notation) format file.

2.4. Image analysis techniques

This section explains the models employed for canopy image segmentation. In this study, supervised image segmentation techniques such as the OBIA method, unsupervised techniques such as K-means and two DL models were applied.

2.4.1 *Object-based image analysis*.

The OBIA methodology segmented the images into several non-overlapping regions, where each region is called an object; these categories are distinguished as being composed of a group of pixels with homogeneous features. For each object, the relevant features (radiometric, contour, and shape) were extracted. This method consists of three distinct steps involving image segmentation, feature extraction, and classification. The trained models were used to classify the entire vineyard into three distinct classes, such as soil, shadow, and vine canopy. In the dataset of this paper, the vineyard canopy despite having an irregular geometry has a definite shape, and this differs mainly in size from that of the soil and shadow. For spectral features, the average value of each spectral band is calculated. The training set is used to train the model, the testing set is used to evaluate the accuracy of the model on new data, and the validation set is used to adjust the model parameters. The next step was to constitute 390 samples to the training model, divided into 130 samples per each class. An additional 90 samples, divided into 30 for soil, 30 for canopy and 30 for shadow, respectively, were applied for the testing image set and likewise for the 90 samples for the validation images. The number and size of polygons were chosen to maintain balance between classes and to cover a greater diversity of cases. To classify the vineyard objects in the images obtained from DOM subdivision, i.e., subimages that have a spatial resolution of at least 0.37 m per pixel, a segmentation process based on the large-scale mean-shift (LSMS) algorithm developed by (Michel et al., 2014). This algorithm is applied using the OTB library (orfeo toolbox) that provides a set of tools for image analysis, the first step of operations performed by the LSMS algorithm is to apply the LSMS-Smoothing function, in which in a first operation consists of removing noise and then applying a normalization procedure (Grizonnet et al., 2017). Subsequently, a bottom-up approach is adopted, in which objects to be classified are gradually created by aggregating similar pixels, this aggregation is done by exploiting the spatial proximity or spectral similarity of them. The smoothing phase aims to improve the shape and consistency of the objects. The second

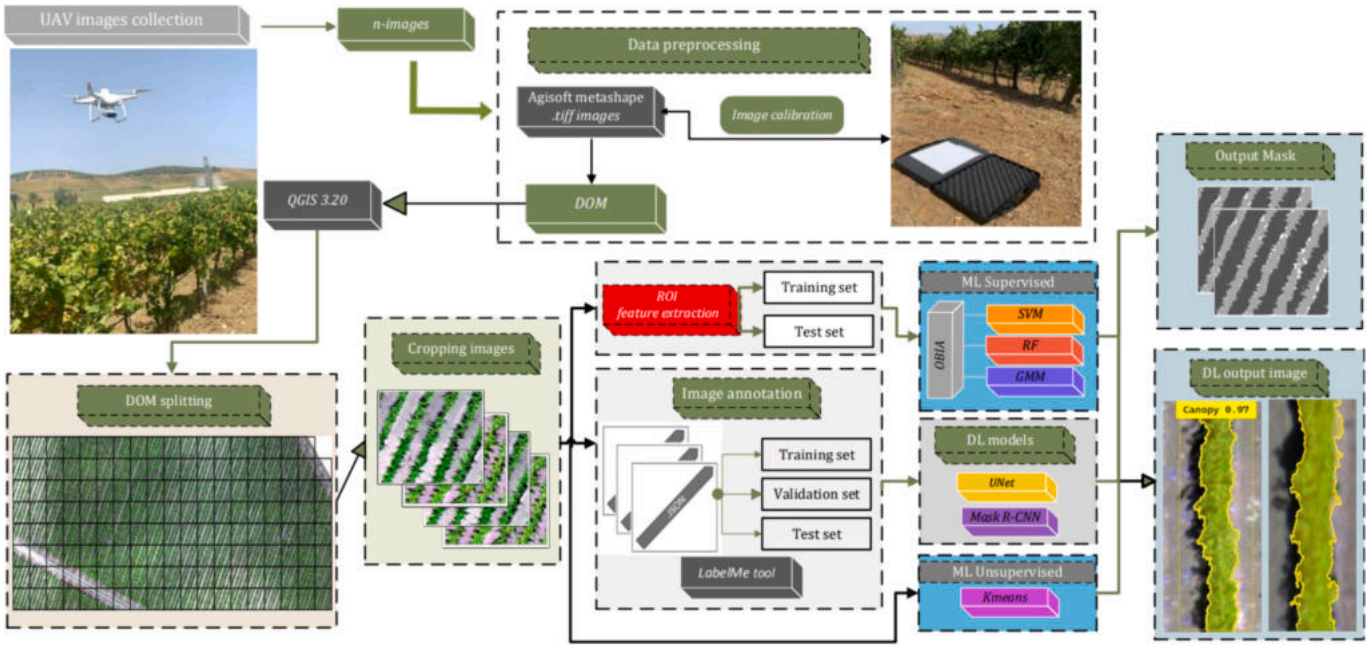


Fig. 2. Process flow involved with the proposed methodology for supervised and unsupervised segmentation methods of UAV images.

step (LSMS-segmentation) is done by setting appropriate parameters, including the range, defined as the minimum Euclidean distance between the spectral signature values of pixels and expressed in radiometric units (De Luca et al., 2019). In this study, the parameter concerning the range radius was set to 25. Conversely, by setting a spatial radius of 15, the distance within which pixels are considered similar was determined. These parameters were selected after conducting multiple tests with different values of range and spatial radius, ultimately yielding the best segmentation (visually assessed). This parameter selection could affect the size of objects and the accuracy of segmentation, potentially resulting in the inclusion of non-targeted objects. The supervised classification models used in this study were chosen according to the reference literature concerning the segmentation of orchards and in general canopies by UAV imaging. In this context, Support Vector Machines (SVM) and Random Forests (RF) machine learning algorithms were selected as having good operational stability and being less affected when applied successively to the operations performed by the OTB library (Modica et al., 2021). Another machine learning algorithm applied for image classification is the Gaussian Mixture Model (GMM), which has provided excellent results in the past for analysis of pixel distribution in agricultural scene images. One of the supervised classifiers that was used is Support Vector Machines, and this is a nonparametric method because it does not assume the distribution of the data, as this algorithm uses an optimization function that tries to find the hyperplane that maximizes the distance between data classes. In the specific case was used SVM type using a linear hyperplane in an ultra-dimensional space, where the SVM classifier determines to which class the pixels nearest to the hyperplane belong in a manner consistent with the training samples (Vapnik, 2006). In this study, a specific linear type of Kernel function with a C-type model that has a cost parameter set to one and gamma to 0.02 were used for the application of the algorithm. The LIBSVM library was used to perform this algorithm; it allows the selection of several parameters to be specified during training, including the kernel type and the C parameter that controls the penalty for classification errors.

The Random Forest algorithm is a supervised machine learning method; it is called *ensemble*, implying that it combines the predictions of multiple models to produce a final prediction. Specifically, the RF algorithm combines predictions from a set of decision trees, this via bootstrap aggregation. For each decision tree, the RF algorithm selects a random subset of features from the training data. The predictions from each decision tree are then combined to produce a final prediction. The number of bootstraps was set to 150 so that different decision trees could be developed and thus increase the robustness of the algorithm, but still was not too high as to affect the accuracy of the classification (Belgiu and Drăguţ, 2016). However, to set some parameters of the algorithm, an estimation was carried out through the out-of-bag (OOB) process to calculate the generalization error, which was set equal to 0.001; these estimates are carried out using the training data that were not used to construct the decision trees. Therefore, the maximum number of decision trees was set as 200, with five maximum tree depths.

The unsupervised Gaussian mixture model (GMM) classification algorithm is a statistical model that assumes the data are generated from a combination of Gaussian distributions. Thus, considering the image in terms of a matrix in which each element represents a pixel, considered a random variable and denoted by the variable x assuming the values of the bands that constitute the image. The classification probability of the elements in the image is represented by the function 1, where represents a weighted sum of Gaussian distributions through:

$$f(x) = \sum_{i=1}^{k} w_i N\left(X|\mu_i, \sigma_i^2\right) \quad (1)$$

where k represents the total number of areas while the set $\{w_1, w_2, \ldots\ldots, w_k\}$ defines weights that satisfy the condition $\sum_{i=1}^{k} w_i = 1$.

In this study, the parameter k in equation (1) was tuned to 3, estimated using the Bayesian Information Criterion (BIC). The number of EM iterations was set to a value of 100, which proved to be sufficient for the convergence of the EM algorithm.

The expression $N(X|\mu_i, \sigma_i^2)$ represents the Gaussian distribution of the i-th region with the mean μ*i* and the standard deviation σ*i* respectively, as indicated by function 2:

$$N(X|\mu_i, \sigma_i^2) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(\frac{-(x-\mu_i)^2}{2\sigma_i^2}\right) \quad (2)$$

the GMM algorithm tries to find the parameters μ, σ and α that maximize this condition and exploits the likelihood function (maximum-likelihood), which represents a system of nonlinear equations. To solve the maximum-likelihood equation in the operations performed by the GMM classifier, an expectation maximization (EM) algorithm is computed, which is an iterative algorithm that updates the parameters μ, σ e α. Therefore, the EM algorithm was terminated when the difference between two consecutive log-likelihood scores was less than 10^{-3}.This algorithm was applied using scikit-learn library (Scikit-Learn, 2022), and then after importing the training images, the number of model components was set. This parameter determines the number of classes into which the data will be clustered; several components that is too low can lead to under-approximation, while a number of components that is too high can lead to an over-approximated model; in this specific case, the number of components were 3. In this study, the GMM classifier was applied through the OTB library, in which Lagrange et al., (2017) developed an efficient implementation.

2.4.1. K-means

In this paper, we chose to use one of the most widely used clustering algorithms for automatic classification of remote sensing images namely K-Means. This learning model is not provided with special labels for the training data, as observed for previous methods, so the method has to learn relationships completely on its own. The K-means algorithm iteratively clusters data points based on their distance, with closer points indicating greater similarity (Sinaga and Yang, 2020). Pixel clustering is performed by calculating the sum of the squared Euclidean distances of the pixel values, compared with the centroids of the clusters. In the case under study, three clustering centers corresponding to the categories of soil, vineyard canopy and shadow were selected. After clustering, the images of the experimental vineyard, are binarized and contain only the three specific categories.

2.4.2. Deep learning image analysis

This paper investigated the neural networks which apply supervised semantic segmentation via U-Net method and a convolutional neural network using the Mask R-CNN method. The architecture of Mask R-CNN performs two separate operations, a first module identifies regions where objects are located, another module performs pixel-level segmentation to delimit object boundaries, this last module is implemented based on the Faster R-CNN structure (Ren et al., 2015). The hyperparameters of the Mask R-CNN model used during the training are shown in Table 3. Mask R-CNN consists of a core network layer (Backbone) comprising the ResNet101 architecture, used in this study. Input images pass through this backbone layer, which performs a set of convolutional and pooling operations. ResNet101 produces an attribute map for each object in the image based on the training information. ResNet101 uses Global Average Pooling at the end of the network, this operation reduces the spatial dimension of the feature maps to a single value for each channel, creating a compact feature representation. In the model structure is the Region Proposed Network (RPN) layer, which is responsible for generating proposals of regions that might contain the image objects, this operation creates positive and negative samples. For each identified region in which the objects are present, Regions of Interest (RoI) are developed, these have different sizes and proportions to each other. Next, using the RoI Align layer, which plays a key role in

Table 3
Values of Mask R-CNN hyperparameters.

Hyperparameters	Mask R-CNN Input
NUM classes	3
Backbone	ResNet101
Pooling layer	Global Average Pooling
Model Structure	RPN – MaskHead
Optimization Algorithm	SGD
Batch size	16
Number of Epochs	50
Learning rate	0.001
Momentum	0.9
Weight decay	0.0005
Image resize mode	Square (240x240)
Image channel count	4 (RGB+NIR)
Train ROIs per image	390

aligning the features extracted from the map to the corresponding (RoI). Another step is the MaskHead, which is responsible for generating the masks associated with the detected objects. These masks indicate which pixels belong to the detected object and are used to perform segmentation. RoIs having the same dimensions are used as input first in the process that indentifies the class and respective probability of each identified object, this process is performed through the Fully Connected Network (FCN). Then the softmax classifier is applied to these RoIs to assign a class label to the object. Generally defined, the output of the model application provides a class label indicating the name of the object class, a bounding box delimiting the object (bounding box), and a mask delimiting the pixels of the object to be detected. This is a binary mask (with the values 0 and 2) in which the value 0 indicates pixels that do not belong to the vine canopy and instead can be traced back to the soil, the value 1 refers to the vine canopy pixels, and finally the value 2 refers to the shadow projected by the canopy. Stochastic Gradient Descent (SGD) optimization algorithm was applied with the purpose of updating the model parameters based on a random subset of data (minibatches), so during each epoch the entire dataset was divided into random mini batches of predetermined dimensions. The experimental process used in this study involved setting a learning rate of 0.001 and a momentum of 0.9, batch size was set to 16 and a weight decay of 0.0005. The neural network was trained with 50 epochs for each spectral band.

U-Net is a model for image segmentation that was first introduced by (Ronneberger et al., 2015). The U-Net architecture consists of two main parts, namely an encoder part and a decoder part. The encoder part is responsible for feature extraction from the input image, while the decoder part is responsible for reconstructing the segmented image. The encoder part of U-Net is a deep convolutional neural network that uses a series of pooling and max pooling layers that both enable feature extraction from the input image. The hyperparameters of the U-Net model used during the training are shown in Table 4. A batch size of 16 with an initial learning rate of 0.001 was used during model training. The encoder stage starts with the definition of a feature map of fixed size, in this study having used training subimages of size 240x240x3.

Table 4
Values of U-Net hyperparameters.

Hyperparameters	U-Net Input
NUM classes	3
Encoder filter size	3x3 (unpadded)
Activation function	ReLU
Decoder upsampling	Deconvolution layers
Loss function	Categorical cross entropy
Final activation function	Softmax
Number of Epochs	50
Batch Size	16
Learning rate	0.001
Weight decay	0.0001
Image resize mode	Square (240x240)
Image channel count	4 (RGB+NIR)

These passing through a series of convolution layers (3 × 3 unpadded) (Long et al., 2015), a batch normalization layer (BatchNorm)(Ioffe and Szegedy, 2015) and pooling layers and to these are added max pooling layers of type 2 × 2 that are used to reduce the image size and number of parameters creating a feature map of size 120x120x16. Each convolution layer then applies a filter to the feature map and progressively produces smaller feature maps until the size is 15x15x64, at which point the feature map is output to the decoder architecture layer. The decoder part of U-Net is an inverted convolutional neural network that starts by applying a series of convolution layers to the feature map defined as upsampling and concatenation to reconstruct the segmented image. The size of the feature map is doubled with each convolution layer providing as output a segmented image of size 240x240x3. The concatenation layers in the decoder part of U-Net combine the features extracted from the encoder part with the features extracted from the decoder part. All convolutional and deconvolutional layers are followed by an activation function ReLU defined Rectified Linear Unit, except for the 1 × 1 convolutional layer at the last, which predicts class probability. In this study, categorical cross entropy was used as a loss function for training the model. This function measures the difference between the predicted probability distribution and the true labels for each pixel. The Softmax function was used to transform the final feature map into probabilities, which allows obtaining a probability distribution over all classes for each pixel.

The investigation were performed using a workstation running laptop with Windows 11 Pro. The DL model were based on PyTorch 1.9.0, while the computing architecture was CUDA 11.1 is used to accelerate the training process, in addition, pakages such as TensorFlow 1.3, Keras 2.0.8 were considered for the application of convolutional network. The deployment of the Mask R-CNN model was carried out following a procedure allowing for processing the TIFF format images that are characterized by more than three bands per image. Specific environment configuration were used as a central processing unit CPU Intel(R) Core$^{(TM)}$i7-12700 K (4.9 GHz), 32.0 GB and a graphics processor unit (GPU) NVIDIA GeForce RTX™ 4090 16 GB GDDR6 with processor CUDA cores.

2.5. Evaluation metrics

In this study, two main performance measures, i.e. overall accuracy (OA) (4), F1-score (5), were used based on a specific confusion matrix with four main factors, such as false negative (FN), false positive (FP), true negative (TN) and true positive (TP), to evaluate the performance of the model for extracting features of soil, canopy and shadow classes from high-resolution UAV images. The OA is specified as the sum of correctly identified pixels divided by the totality of pixels. F1 score, on the other hand, represents the combination of recall and precision metrics. Recall is specified as the percentage of correctly identified pixels among all pixels that were depicted for that class. Precision is calculated as the percentage of pixels identified exactly among the pixels identified in a specific class. An additional parameter calculated to measuring the accuracy of segmentation algorithms was the intersection over union (IoU), also referred to as the Jaccard Index. In Expression (3), this parameter is defined as the ratio of the intersection area, the area identified by the algorithm and the ground truth manually identified by the operator. These are divided with the intersection area between the area observed by the segmentation method and the ground truth.

$$IoU = \frac{areaofoverlap}{areaofunion} = \frac{S \cap S_i}{S \cup S_i} \quad (3)$$

where S represents the area identified by the segmentation algorithm; instead, the term S_i represents the actual area computed manually by segmenting the edges of the category of objects to be identified. Thus, the numerator of the equation indicates the area shared by S and S_i. The denominator represents the total area of S and S_i. A value of IoU greater

than 0.70 generally indicates good agreement between the two components S and S_i.

2.6. Vegetation index

After obtaining the pure vegetation pixels by the segmentation and classification techniques described above, the normalized difference vegetation index was calculated (NDVI) (Rouse et al., 1974). This is one of the most widely considered indices for assessing crop growth and is calculated by considering the near-infrared (NIR) and red spectrum bands.

2.7. Canopy data measurements

Ground sampling was performed to assess two agronomic parameters of the vineyard such as total leaf area (TLA) and leaf chlorophyll (Chl). The ground sampling performed in the on-farm vineyard was carried out in specific areas that were characterized by different growth vigor. In fact, agronomic surveys designed to classify vigor levels for 2021 and 2022 identified two distinct vigor levels. Using field surveys, fourteen sampling blocks were identified divided into areas with low vigor vines (LV) and areas with high vigor vines (HV) (Fig. 3). Each individual sampling block has a length of 6 m and consists of six plants; thus, a total of 84 plants were monitored. Survey activities were preceded by the identification of the 14 ground control points (GCPs), with known coordinates, useful for georeferencing all survey points to a common reference system (WGS84; UTM 32 N).

Evaluation of (TLA) was performed by applying the empirical model for non-destructive estimation of leaf area for primary and lateral shoots as formulated by Lopes and Pinto, (2005). This method includes sampling vine shoots collected randomly within the canopy. Each individual shoot generally consists of a primary shoot axis and secondary shoots that originate from it during the growing season. These shoots were collected from plants placed in the rows in front of the row that was selected as the sampling block in total 60 shoots (one shoot per vine) were collected for each survey session. These shoots were labelled according to the vigor level to which the plants belonged. From each shoot the primary and lateral shoot leaves were separated, on these shoots the number of nodes and the overall length measured from the base to the node where the last leaf counted was inserted were also evaluated. Leaf area was determined in the laboratory with a leaf area meter (LI-3100C area meter Li-COR Biosciences, Lincoln, NE, USA). These surveys were conducted for two years, at the phenological stages of vine flowering (BBCH65) and berry softening (BBCH85). For each primary and lateral shoot, the following variables were measured: sum of primary leaf area and sum of lateral leaf area; these two parameters together constitute the average leaf area (LA) of the shoot. Based on these biometric measurements conducted in the field, it was possible to determine (TLA) on each labelled vine by multiplying the average LA by the total number of shoots per plant. Leaf area assessment method adopted in this study has been applied by several authors because it provides an estimate of vine leaf area regardless of cultivar, year and phenological stage, representing a low-cost, simple and accurate method, however, being a manual sampling method, it is time-consuming and labour intensive for operators.

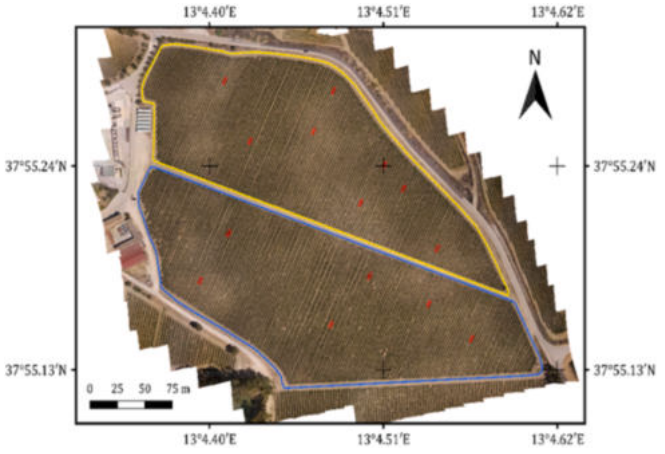


Fig. 3. On-farm vineyard with red blocks placed on the map indicate ground sampling points.

The evaluation of chlorophyll content (Chl) ($\mu g\ cm^{-2}$), done during the growing season in two phenological stages, was measured by proximal sensor, a portable leaf fluorimeter Dualex® Force-A (Orsay, Cedex, Francia). This instrument performs non-destructive, simple and rapid measurements by taking an absorbance measurement of the leaf epidermis, exploiting the mechanism of fluorescence emitted by the chlorophyll pigment (Cerovic et al., 2012). Measurements were performed on the sampling blocks by selecting three leaves per plant. Leaves sampled during the BBCH65 phenological stage were positioned on the 5/6 node of the primary shoot, for BBCH85 phenological survey the sampled leaves were positioned in the 8/10 node of the shoot, this methodology allows for standardization of measurements and sampling of even-aged leaves.

2.8. Statistical analysis

To evaluate the overall performance of canopy segmentation and classification methods, statistical analyses were carried out on the sampling blocks described. In these sampling blocks, the geometric information of the canopy was extracted by evaluating the Projected Canopy Area (m^2) (PCA) parameter, and spectral information derived from NDVI, which represent key parameters needed to properly assess TLA and Chl. This information was obtained using the QGIS plugin (Zonal Statistics), which allows information such as the count, mean, standard deviation, minimum and maximum value of each raster to be extracted. ANOVA test was used after dividing the dataset of sampling blocks into LV and HV to test the equality of the averages of both PCA and NDVI. Comparisons between the averages were performed with Tukey's test ($p \leq 0.05$). In addition to calculating the basic statistic, Gini heterogeneity index was evaluated to appreciate the variability of the dataset (Gini, 1921). This index is a measure of the inequality of a distribution and is a number between 0 and 1. A value of 0 indicates that the frequency distribution is homogeneous, whereas a higher value means that the sample is more heterogeneous. Pearson's correlation coefficient was used to compare the similarity of vineyard rows geometric data (PCA) and NDVI with canopy biometric traits such as TLA, shoot length and chlorophyll content.

3. Experimental results

The on-farm vineyard shows a high degree of vegetative and reproduction variability, and it was observed that one of the factors determining high variability in vine vigour growth was vineyard topography. Field areas show distinct soil profiles based on slope values identified through the Digital Elevation Model (DEM). Fig. 4 illustrates the elevation variation within the plot, showing significant elevation heterogeneity, with elevation values ranging from 365 m at the highest areas to 335 m at the lowest areas of the vineyard. A flat area extends to the highest elevation with elevation values ranging from 362 m a.s.l. at the lowest point to 365 m a.s.l. at the highest point, with an average slope value of 3.55 %. Another section of the DEM corresponds to the slope area, with elevation values ranging from 361 m a.s.l. at the highest point to 338 m a.s.l. at the lowest point, with an average slope value of 14.77 %.

As a result, topographical variability of this specific vineyard, sampling points were identified in correspondence with flat areas and sloping areas that showed different vine canopy growth.

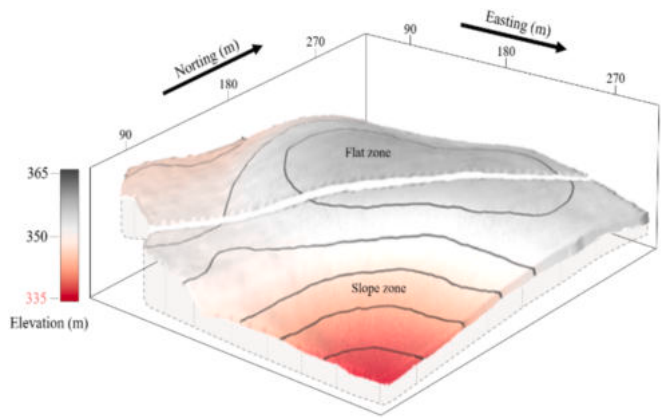


Fig. 4. Representation of Digital Elevation Model (DEM) of the studied on-farm vineyard.

The accuracy indices F1 Score, OA and IoU were calculated both for each phenological stages and for segmentation methods: OBIA, DL and K-means, (Tab. 5). Supervised algorithms relying on the Large-scale mean-shift segmentation (LSMS) method used in OBIA, performed lower than DL methods but still significantly better than the K-means clustering method in two survey epochs. The best results in evaluating the accuracy of segmentation and object detection can be observed through the value of OA, it is evident that U-Net and Mask R-CNN were the most efficient in segmenting and detecting object classes, with accuracy values above 0.85 for both BBCH65 and BBCH85. Nonetheless, it is evident that among the supervised ML methods, the GMM showed the best OA values, showing comparable values to the DL methods.

Specifically, these results indicate that among the supervised classifiers used after the segmentation process, the GMM had the best F1 score performance for the identified classes. Indeed, for the surveys conducted in BBCH65, canopy identification was done with an accuracy index of 0.80 and 0.74 for the surveys conducted at BBCH85. The classifiers RF and SVM obtained lower F1-score values, in fact SVM detects the vineyard canopy with an accuracy of 0.72 and 0.68 for BBCH65 and BBCH85, respectively. RF, on the other hand, obtained the lowest accuracy values among the supervised methods; in fact, the F1 score value for canopy detection was 0.67 and 0.57 for the two phenological periods considered, respectively. The K-Means clustering segmentation method, instead, was the least accurate in identifying the canopy, with an F1 score index of 0.63 and 0.58. A different reasoning can be made if the F1 score data traced back to the U-Net and Mask R-CNN methods are observed, where the identification of the vineyard canopy has an accuracy index of 0.92 and 0.87 for Mask R-CNN and 0.89 and 0.85 for U-Net. The other F1 score values for the remaining classes identified in the images show similar performance, in fact, the U-Net method shows an F1 score value for soil of 0.91 and 0.87 indicating a low level of error, as also found for Mask R-CNN. Among the supervised methods, on the other hand, GMM is the most promising, with values of 0.84 for the first survey epoch and 0.75 for the second. K-means clustering, meanwhile, proves to have low OA values, with a high degree of error, easily inferred from the values of 0.68 and 0.63. Evaluation of the F1 score values when it comes to the shadow highlights that among the supervised methods RF is the least accurate, with values of 0.60 and 0.58 for the two survey periods, respectively. Similarly, K-means clustering exhibits similar F1 score values. The other two supervised methods, instead, while reporting OA values that are not specifically high, still prove to be better than the two previous cases, especially the GMM method, which showed a value of 0.74 for BBCH65 stage. DL methods show higher F1 score values than the previous methods and prove to be more accurate in detecting this category.

Table 5 shows the geometric accuracy ratings of the segmentation and classification methods employed. As can be seen between the two survey periods there are general differences in terms of the accuracy of the segmentation algorithms, it is noticeable that for the surveys performed at BBCH65 both the IoU and accuracy values are better, this indicates a better shape overlap of the surveyed objects. This result could be related to the smaller size of the canopies, and consequently less shadow projected on the ground, which is a disturbance factor in the process of delineating canopy edges. Among the results to be considered, it is noticeable that the K-Means clustering method and the RF supervised classification method show less overlap, especially for IoU results that refer to shadow, both algorithms show values of about 0.50. Instead, among the supervised methods SVM and especially GMM show the best performance, the DL segmentation models show results about

Table 5
F1 score, IoU and OA indices of the segmentation models applied to the object classes Soil, Canopy and Shadow in the two different phenological stages of the vine (BBCH 65 and BBCH 85).

	BBCH 65				BBCH 85		
Method	**Index**	**Soil**	**Canopy**	**Shadow**	**Soil**	**Canopy**	**Shadow**
OBIA (GMM)	IoU	0.77	0.79	0.68	0.74	0.60	0.65
	F1 score	0.84	0.80	0.74	0.75	0.74	0.67
	Tot IoU		**0.76**			**0.72**	
	OA		***0.85***			***0.78***	
OBIA (RF)	IoU	0.58	0.62	0.47	0.62	0.55	0.51
	F1 score	0.71	0.67	0.60	0.64	0.57	0.58
	Tot IoU		**0.58**			**0.55**	
	OA		**0.72**			***0.68***	
OBIA (SVM)	IoU	0.72	0.71	0.64	0.69	0.65	0.55
	F1 score	0.76	0.72	0.66	0.69	0.68	0.63
	Tot IoU		**0.70**			**0.63**	
	OA		**0.77**			***0.71***	
K-Means	IoU	0.64	0.48	0.50	0.60	0.51	0.47
	F1 score	0.68	0.63	0.56	0.63	0.58	0.54
	Tot IoU		**0.55**			**0.53**	
	OA		**0.64**			***0.55***	
DL (Mask R-CNN)	IoU	0.85	0.84	0.79	0.82	0.80	0.78
	F1 score	0.94	0.92	0.90	0.89	0.87	0.85
	Tot IoU		**0.84**			**0.80**	
	OA		**0.92**			***0.89***	
DL (U-NET)	IoU	0.84	0.82	0.76	0.80	0.79	0.75
	F1 score	0.91	0.89	0.87	0.87	0.85	0.83
	Tot IoU		0.82			**0.78**	
	OA		**0.91**			***0.88***	

30 % higher than the clustering algorithm as well as RF and achieved a 10 % increase in accuracy over GMM on all classes considered. These values agree with the results observed for F1 score and OA and further clarify how the U-Net, Mask R-CNN and models were more proficient in classifying image pixels as belonging to the three classes under investigation.

The results for the accuracy indices can be verified with the graphical representations of the images in Fig. 5. This figure shows the masks obtained after applying the segmentation and classification methods. In addition, for each segmentation method, masks referring to two different conditions of canopy growth are represented. It shows a case of areas where vineyards have a larger and more highly developed canopy, attributable to an area of HV, while another case refers to areas of LV, where vineyard canopies are sparser and less developed.

Fig. 6a refers to the phenological period BBCH65 in which the surface area is more covered by soil and the incidence of shadow is lower. In contrast, in the phenological period BBCH85 depicted in Fig. 6b, both the surface area covered by canopy is greater and the surface area covered by shadow is also larger than in the previous period. Among the methods, K-Means was the one that identified the lowest area covered by canopy. The latter and RF were the methods that overestimated the area occupied by the shadow, with values of 22.5 % and 25.8 %, respectively,

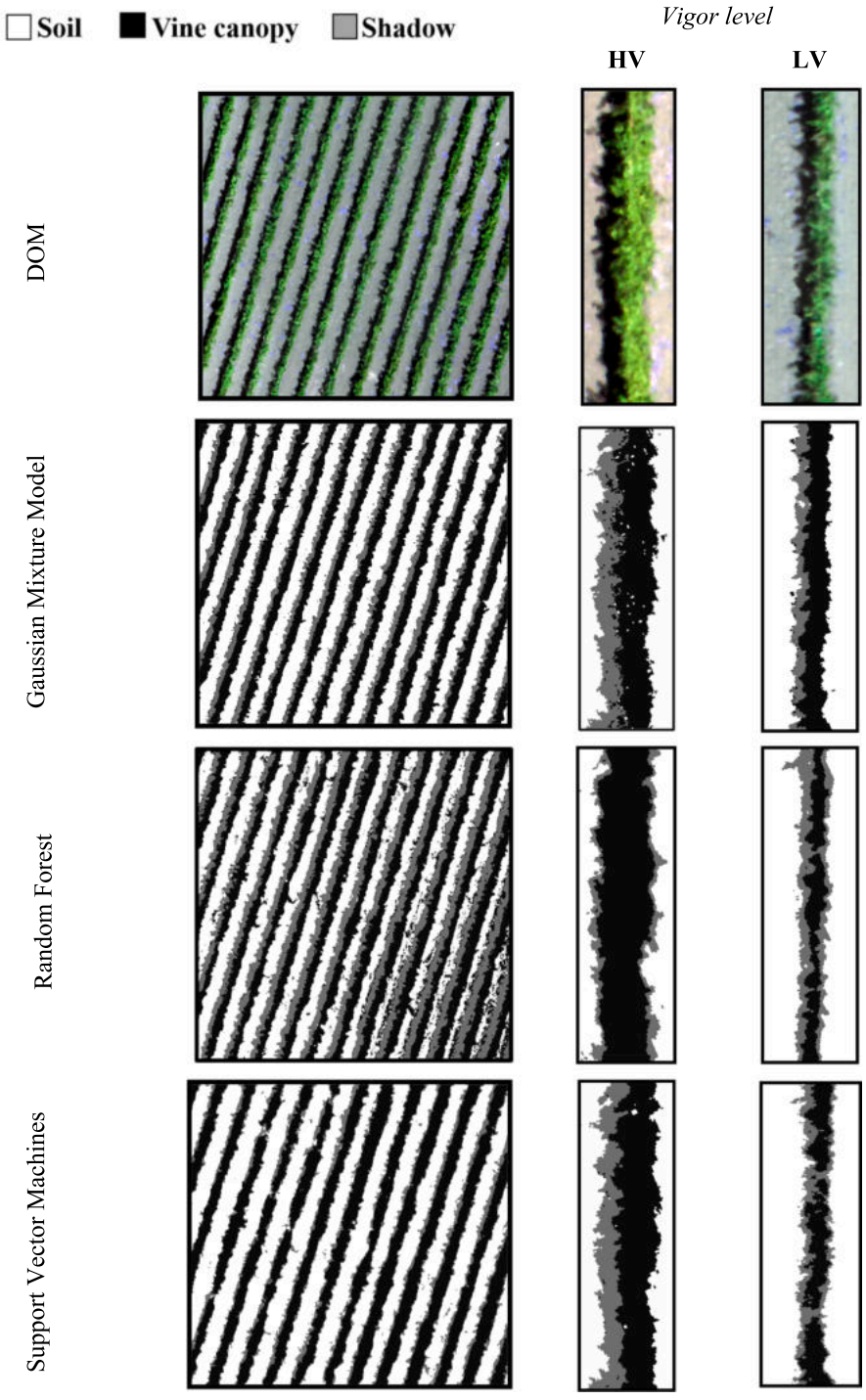


Fig. 5. Graphical representations obtained after the application of the methods to the vineyard sub-images derived from the DOM splitting, and the segmentation masks subsequently obtained. Detailed images of two vineyard vigour conditions and their respective masks are also shown.

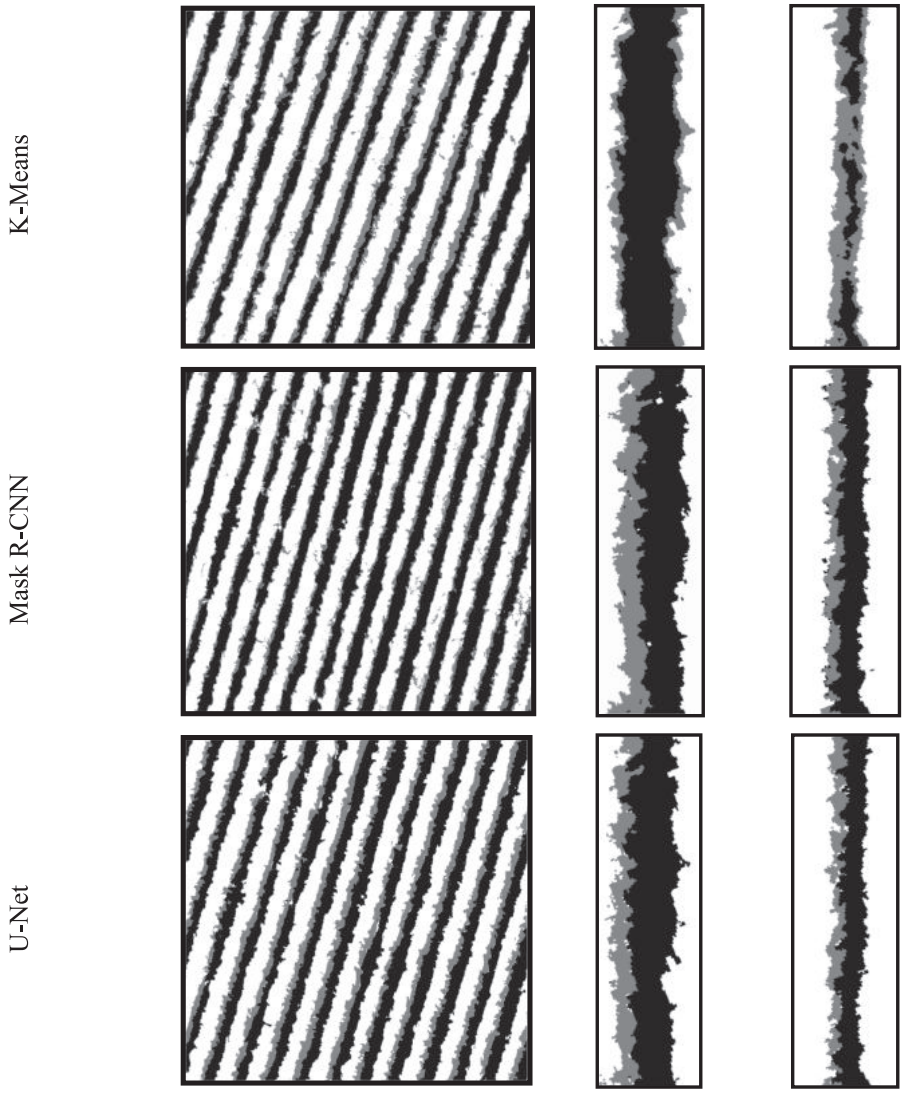


Fig. 5. (*continued*).

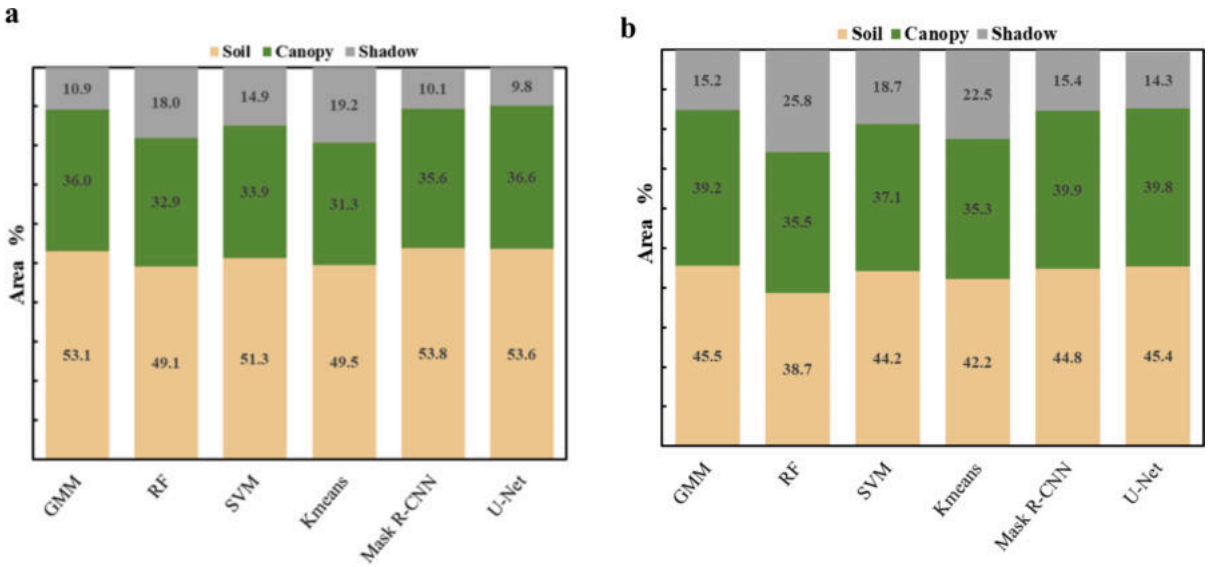


Fig. 6. Distribution of the surface area covered by the three classes composing the DOM images; Figure **a** refers to phenological stage BBCH65; Figure **b** refers to phenological stage BBCH85.

referring to the BBCH85 phenological stage (Fig. 6b). This effect of shadow overestimation is mainly attributable to misclassification errors of canopy pixels that are classified as shadow, which is found for canopies of LV growth. DL methods and the GMM methods reported comparable percentage values with reference to the surface area covered by soil, these values differing significantly from the values reported by RF and K-Means, which instead identified an area of approximately 40 %.

3.1. Assessment of vine canopy delimitation

In order to obtain a more detailed assessment of the impact of the vineyard canopy segmentation methods, the sampling blocks were divided by the vigor level with the aim of identifying differences in the blocks lying in LV zones and those lying in the HV zones. It was observed that on average, the PCA among the different models applied is 4 (m^2). However, some classifiers have underestimated this area, e.g. RF identifies 3.62 (m^2) and even the K-Means reports an average PCA value of 3.51 (m^2). The GMM method reports an average PCA value of 4.21 (m^2), which is comparable to Mask R-CNN and UNet that are 4.12 and 4.07, respectively (m^2).

Through analysis of variance, it was observed that both the vigor levels and image analysis methods applied in this study affect the response of PCA values (*p-value* = 0.008). Boxplots showing how the average PCA figure varied between HV and LV are shown in Fig. 7. Boxplot 7a shows the results of the analysis carried out by Tukey's test related to LV zones, showing how the U-Net and Mask R-CNN methods are statistically different from the K-Means classification model; instead, RF shows a statistically intermediate result between the groups. The graph in Fig. 7b, meanwhile, shows the results of PCA for HV zones. Tukey test did not find differences between GMM, Mask R-CNN and.

U-Net models. This group reported average PCA for the HV canopies with smaller size than the average PCA derived from the SVM and the RF as well as K-Means. The SVM reported the value of PCA in the HV vineyard areas similar to RF, these methods and also the K-Means overestimated this parameter, the latter exhibiting the highest value of canopy overestimation, being statistically different from all other methods.

In the same manner as performed for PCA, the analysis of variance between factors such as vigor levels and segmentation methods has found a significant and positive interaction with the NDVI variable.

The boxplots in Fig. 8 represent the distribution of NDVI corresponding to the canopy segmentation masks in the two vineyard vigor conditions. The boxplot 8a refers to the sampling blocks located in LV zones, it is observed that the classification methods that underestimated the size of the PCA (Fig. 7a), such as RF and K-Means, report higher NDVI than the other methods. In fact, the GMM, Mask R-CNN and U-Net show lower NDVI for LV areas, this is confirmed by the Tukey test. In contrast, Tukey test computed in ANOVA analysis for the RF and SVM indicates that the NDVI are intermediate to the previous groups. Observing the boxplots shows an interesting result about the distribution of NDVI among the models under investigation. Models that tend to overestimate the NDVI of LV canopies are characterised by a narrow boxplot, indicating a narrow dispersion of values within the population and thus less heterogeneity in the data. The boxplot in Fig. 8b refers to NDVI of HV zones, in this case the analysis of Tukey test shows statistically significant differences between DL methods and K-Means, the latter underestimates the NDVI of the HV canopies.

The larger boxplot of K-means represents greater dispersion of the data within the sample and thus a higher data heterogeneity.

Table 6 shows the values for the Gini index, these results allow explaining the observed variability between LV and HV canopies among the image analysis methods. With reference to PCA parameter for vineyard blocks located in LV areas, it is seen that the U-Net and Mask R-CNN methods reported low Gini values. An opposite result was obtained for the RF and K-Means methods, in which the index was 21.44 % and 25.85 %, respectively, which were significantly higher.

On the other hand, observing the distribution of the Gini index for the NDVI, it is noted that this is lower for the K-Means and RF methods and instead is higher for the other methods. Referring to the NDVI, it is noted that the lowest values of the Gini index are reported by the DL methods. Among OBIA methods, GMM and SVM show low Gini values, unlike those reported by K-Means and RF, which exhibit greater heterogeneity.

This performance is determined by the variability of the pixels of the high vigor canopies as a result and where the segmentation method was unable to separate the canopy pixels with those of the shadow and soil.

Fig. 9 shows how segmentation methods that tend to underestimate canopy size exclude edge pixels characterized by lower values than the middle pixels. This phenomenon tends to increase the average NDVI canopies. For HV canopies, segmentation methods that were unable to accurately identify canopy pixels (K-Means; RF), tend to underestimate the NDVI. This underestimation is due to the greater heterogeneity of the

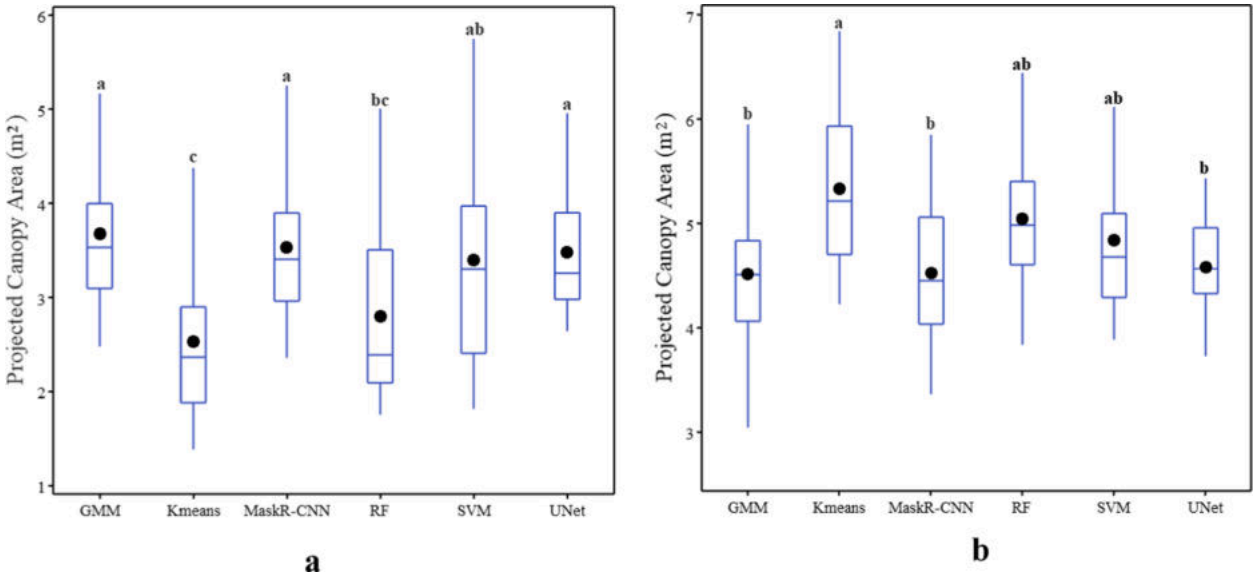


Fig. 7. Boxplot representation of Projected Canopy Area (m^2) values divided among the methods; boxplot **a** pertains to the sampling blocks of low-vigor (LV) canopies, while boxplot **b** pertains to the blocks of high-vigor (HV) canopies.

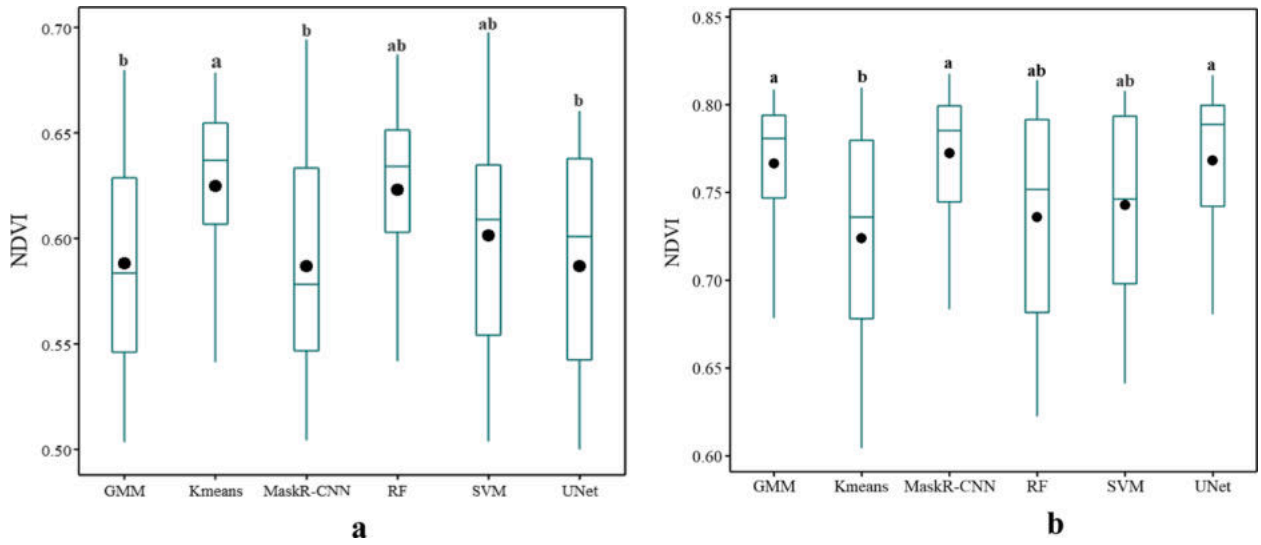


Fig. 8. Boxplot representation of NDVI divided among the classification models; boxplot **a** relates to the sampling blocks of low-vigor zones, while boxplot **b** relates to the blocks of high-vigor zones.

Table 6
Gini heterogeneity index (G%) representing the range of variation for the variables Projected Canopy Area (PCA) and NDVI calculated for the LV and HV sampling blocks and the whole dataset.

Method	*Low vigor*		*High vigor*		*All data set*	
	G% (NDVI)	G% (PCA)	G% (NDVI)	G% (PCA)	G% (NDVI)	G% (PCA)
(OBIA) GMM	6.86	16.07	3.10	6.68	12.57	17.44
K-Means	2.60	25.85	7.87	8.33	9.12	26.45
Mask R-CNN	6.88	15.80	2.13	8.85	11.73	16.38
(OBIA) RF	3.01	21.44	6.73	6.87	8.59	25.70
(OBIA) SVM	4.34	19.72	4.15	6.34	10.04	20.20
U-Net	6.25	12.66	2.74	7.76	11.88	16.84

pixels that characterise the canopy. In particular, the values of shaded pixels that have been misclassified as a canopy, having a lower NDVI, influence the average value of the resulting canopy.

Figs. 9 and 10 show a visual comparison of the results obtained through the different image analysis methods for two vegetative canopy growth conditions. It can be seen clearly through these visual results, which further confirm the analyses of quantitative comparisons performed through the accuracy indices. Fig. 9 refers to LV canopies, in which after removing complex background features such as soil and shadows, only the canopy masks are obtained. Through this visualisation, a comparison can be made between the methods for delineating the canopy edges. Fig. 9 (a), (b) and (d) show the output results of the supervised ML methods, focusing particularly on Fig. 9 (d) it is evident that the RF method is the one that generates the least accurate canopy boundary shape. The K-Means method in Fig. 9 (c) is the worst, with many false detections, this is shown by observing the details on the canopy edges that are completely different from the actual edges. In contrast, in Fig. 9 (e) and (f) canopy segmentation masks are observed to be completely superimposed on the real conditions of LV canopies; in fact, these have the greatest details on the canopy edges. Thus, the classification and segmentation methods that obtained the worst results in LV areas, detect a lower number of pixels and these are mainly the middle pixels of the canopy, and these methods cannot identify the vegetation pixels located at the edges.

Fig. 10, on the other hand, shows the results obtained by the methods for HV canopy, areas of the vineyard where the row vegetation has a larger size. It is noticeable that among the OBIA supervised ML models, GMM (Fig. 10a) was the one that was able to delineate the canopy edges better, compared to the SVM and RF methods, which were instead more affected by shaded pixels. Especially observing Fig. 10 (c) for the K-Means method, the canopy mask does not delineate by vegetation pixels, but rather includes shaded pixels. DL networks were comprehensively better than the other methods for detecting vineyard canopies in HV zones and show greater detail on the canopy edges, which explains their higher accuracy values.

Fig. 11 shows an example of overlapping of two classification masks referring to the methods that accurately classified the canopies, and in this example the U-Net method is mentioned. Furthermore, viewing at the edges of the yellow mask, it can be observed that for the LV vines (Fig. 11a), the canopy pixels are identified, whereas the K-Means (red line) only identifies a part of the canopy. Fig. 11 (c) further clarifies this direct effect on canopy NDVI, as the U-Net also identifies pixels at canopy edges, which usually have a lower value than the middle canopy pixels alone. Consequently, NDVI of the LV canopies identified by DL methods is lower in absolute value and with greater heterogeneity in the dataset. In Fig. 11 (b), meanwhile, it is observed that in HV zones, classifiers such as K-Means (red line), which cannot identify canopies clearly from shaded pixels, report lower NDVI than methods that identify canopies accurately and with acceptable overlap. In Fig. 11 (d), it is observed that in this case the shaded pixels cause an evident interference with the vegetation pixels, and thus a lower NDVI.

The histograms shown in Fig. 12 illustrate the difference in NDVI distribution depending on which method was used to classify the canopies. Histogram 12 (a) refers to LV canopies and three methods are represented, with reference to the U-Net and GMM classifiers, it is noted that both have a broader distribution range compared to K-Means distribution. With reference to this histogram, the relative frequency percentage of NDVI pixel observations with values below 0.60 is significantly higher for U-Net and GMM. For NDVI class of 0.55, the relative frequency percentage of U-Net and GMM is 10 % compared to K-Means, which is 2 %. The histogram in Fig. 12 (b), however, refers to the canopies of the HV zones, and in this case the distribution range of NDVI for the canopy pixels identified by K-Means is greater, in view of the shadowed pixels. For example, for the NDVI value class of 0.70, the relative percentage frequency is 10 % for K-Means and on the other hand is approximately 3 % for NDVI detected by U-Net and GMM methods.

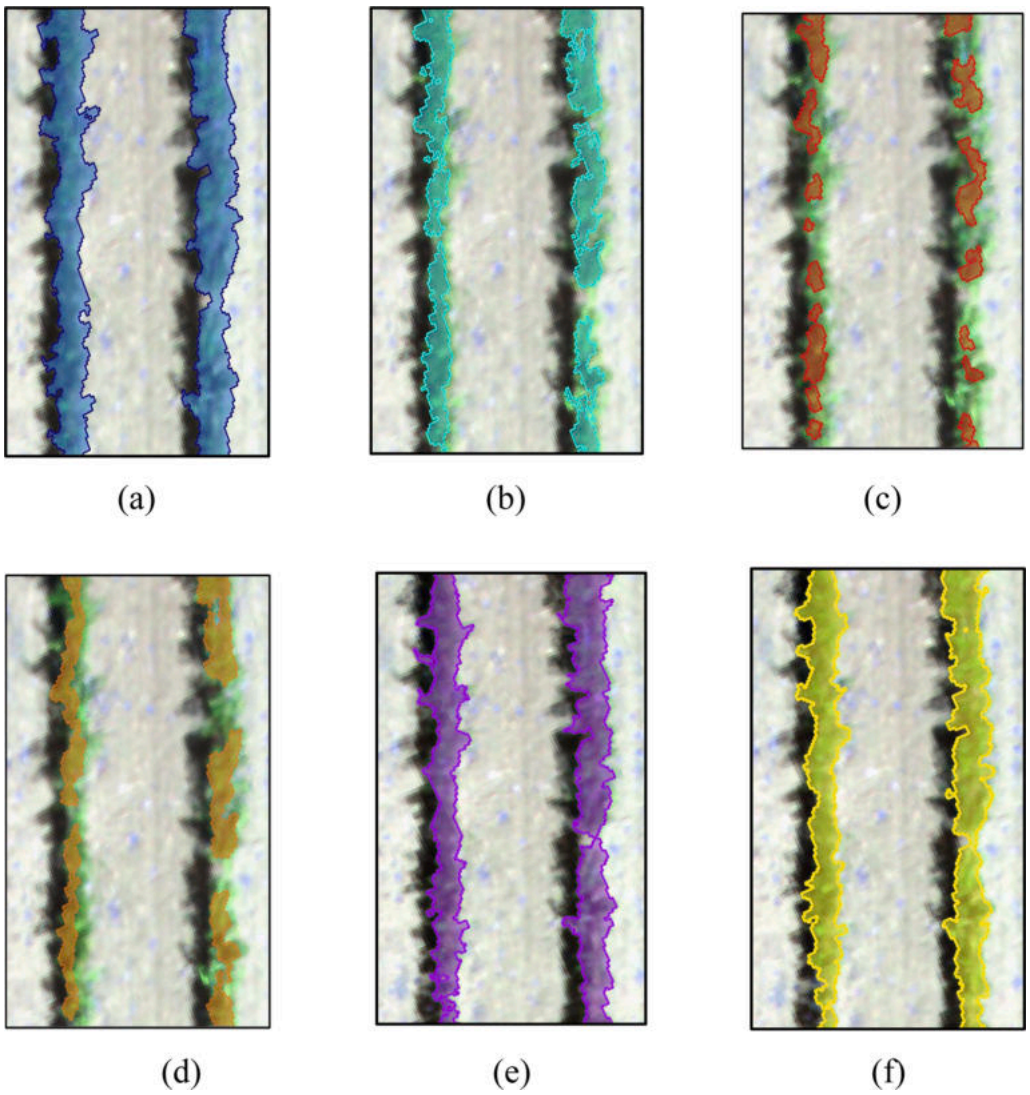


Fig. 9. Results of different vineyard canopy detection methods in low vigor growth conditions. (a) GMM; (b) SVM; (c) K-Means; (d) RF; (e) Mask R-CNN; (f) U-Net.

3.2. Correlation analysis

The results of the spectral and canopy area information obtained through the different segmentation methods were used for correlation analyses combining the vineyard agronomic parameters and evaluated for the two phenological stages (Fig. 13). Observing both correlation matrices, positive Pearson correlation values are noted between the NDVI and the agronomic variables TLA and Chl. In addition, the coefficient values are better for surveys performed during the phenological stage BBCH65. With reference to this phenological stage, the correlation results are shown in Fig. 13 a, considering the canopy NDVI, the highest correlations with the TLA parameter were obtained by U-Net and Mask R-CNN and are respectively 0.89 and 0.87. Among the supervised methods, instead, the correlation between TLA and SVM and GMM is slightly lower than DL networks, in contrast, the correlation between TLA and NDVI values identified by RF classifier is 0.77 and thus lower, a similar value is reported by the correlation between TLA and K-Means and is the worst correlation result ($r = 0.76$). In Fig. 13 b, similar results to those reported for BBCH85 are observed, however, the correlation for DL methods is still higher ($r = 0.78$), compared to methods such as RF and K-Means, which instead report r values of 0.69 and 0.68, respectively. When considering the agronomic parameter Chl conversely, the NDVI values derived by Mask R-CNN and U-Net have the highest correlation coefficients ($r = 0.86$), during the phenological period BBCH65, this value decreases to BBCH85 ($r = 0.80$), remaining the highest result. Thus, the correlation coefficient with leaf Chl increased by 0.10 compared to the NDVI obtained with the RF and K-Means methods. The agronomic parameter S.l was the one that reported lower correlation values and in generally less than 0.50, however even in this case it is observed that DL networks and supervised ML methods report better correlation values. These correlations were also carried out considering the PCA data, and the results obtained are quite remarkable, as shown by the correlation in Fig. 13 a, where TLA and PCA are better represented by U-Net and Mask R-CNN with values of $r = 0.80$ and 0.81, respectively, than the values of the K-Means and RF methods, which are lower ($r = 0.77$). As shown by the correlation coefficients between Chl and PCA, it can be stated that PCA is not a recommended variable for the estimation of Chl content. Correlation coefficients between PCA and Chl were better for phenological stage BBCH65. Referring to the correlation coefficients between the NDVI values of U-Net and Mask R-CNN, are the highest for both phenological periods, in fact with values above 0.95, the clear correspondence between the two methods is evident. Furthermore, DL methods also show a correlation with NDVI of the canopies detected by GMM, further demonstrating their correspondence. Among the results observed in the correlation matrix, it is noteworthy that the lowest correlation values were obtained in the comparison between the U-Net and Mask R-CNN methods with the K-Means method, reporting average values of 0.75 for the BBCH65 and 0.81 for the BBCH85 stage.

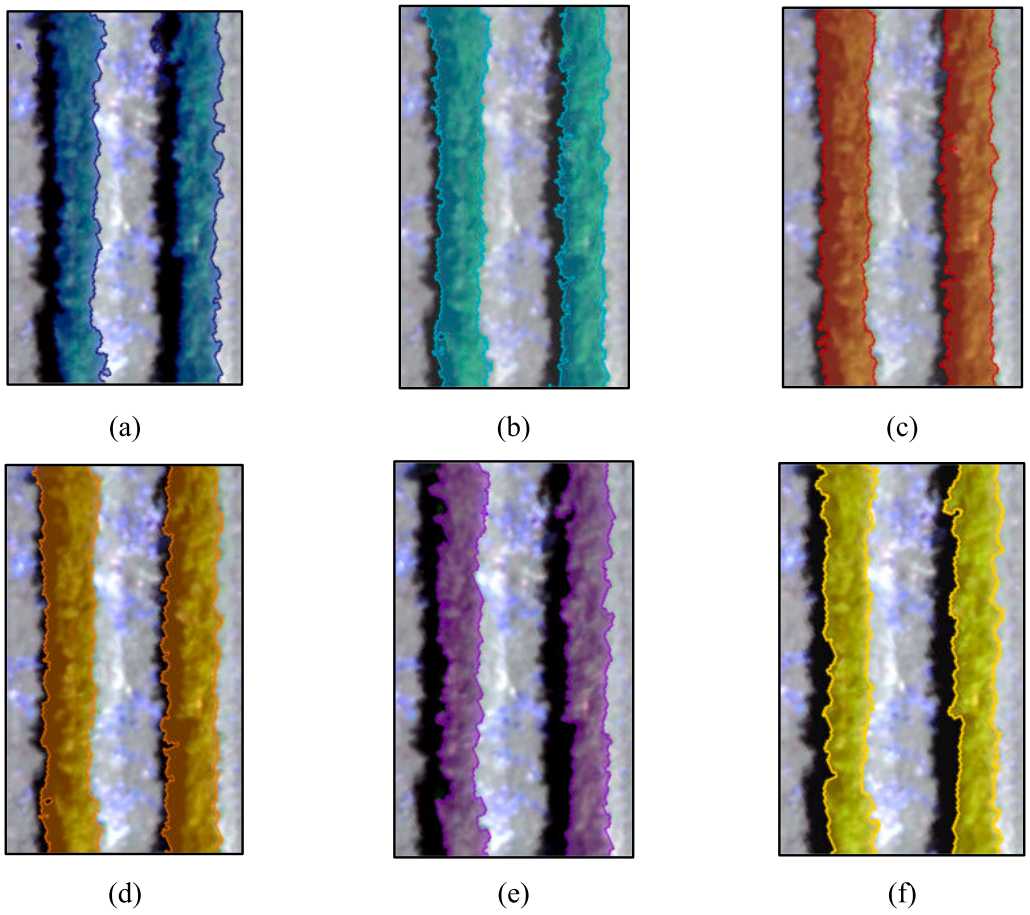


Fig. 10. Results of different vineyard canopy detection methods in high vigor growth conditions. (a) GMM; (b) SVM; (c) K-Means; (d) RF; (e) Mask R-CNN; (f) U-Net.

Additionally, K-Means shows good correlation values when compared to the RF method, both for NDVI and PCA data. Finally, DL methods demonstrate the highest correlation values between NDVI and PCA, compared to the K-Means method, where the correlation between NDVI and PCA is lower, with r values for BBCH65 and BBCH85 of 0.75 and 0.85, respectively.

4. Discussion

4.1. Quantitative comparison of the different methods

The studied vineyard shows a high level of vegetative and reproductive variability, which has been studied in previous investigations at experimental vineyards (Ferro et al., 2023). High variability is found in the heterogeneous canopy growth, which can be attributed to a different canopy thickness due to the variation in shoot number, shoot length and consequently number of leaves. Moreover, this heterogeneity also affects grape yield and quality parameters, which differ significantly between the LV and HV zones. The terrain topography contributes significantly to an uneven change in vineyard vigour. The effects of this phenomenon show that flat areas have higher NDVI values, indicating a higher vegetative density of the canopy. Earlier studies show that in sloping areas there is a decrease in grape yield, together with a lower vegetation density and consequently a low grapevine NDVI is observed (Novara et al., 2018).

This study has considered a complex context, characterized by a vineyard with a wide variability in vine growth, leading to different canopy densities. This parameter is considered one of the main constraints for canopy delimitation (Sarabia et al., 2020). The results of vineyard segmentation and classification vary for different image components and in terms of the size of the obtained segments. However, this was expected since six different pixel-wise methods with varying structural complexity and architecture were employed. In this comparative study between DL networks and classical supervised machine learning methods (GMM; SVM; RF), as well as unsupervised K-Means, the analysis was conducted using a dataset comprising both NIR and RGB bands. This operational mode was adopted because it is known from the literature that the NIR band provides high segmentation performance (Zheng et al., 2020). However, combining NIR with the RGB bands, which are widely used for image segmentation studies, enhances the overall performance (Cinat et al., 2019; Pádua et al., 2022). The image analyses were performed by segmenting each sub-image separately rather than the entire orthomosaic. Accurate detection of vine canopy is crucial for obtaining precise information about vegetation, especially in the presence of elements that can negatively influence canopy spectral response, such as soil, shadows, or weeds. Compared to other traditional machine learning methods, both semantic segmentation and instance segmentation models demonstrated a significant improvement in classification accuracy, with higher (OA) and F1-score. Some other classification methods (e.g., RF; K-Means) misclassified the vine canopy as soil background or erroneously identified shadows as vine canopy. These findings align well with conclusions from other studies that leverage spectral combinations and assess various approaches to canopy segmentation and classification (Huang et al., 2020). Nevertheless, among the supervised ML methods, GMM exhibited the lowest error index and stands out as one of the most stable classifiers. SVM showed some errors in classification, however, it can be considered relatively stable, in contrast to RF, which showed clear errors in determining shadows classified as soil or vegetation. Concerning RF method, lower accuracy in shadow detection in VSP systems has been previously

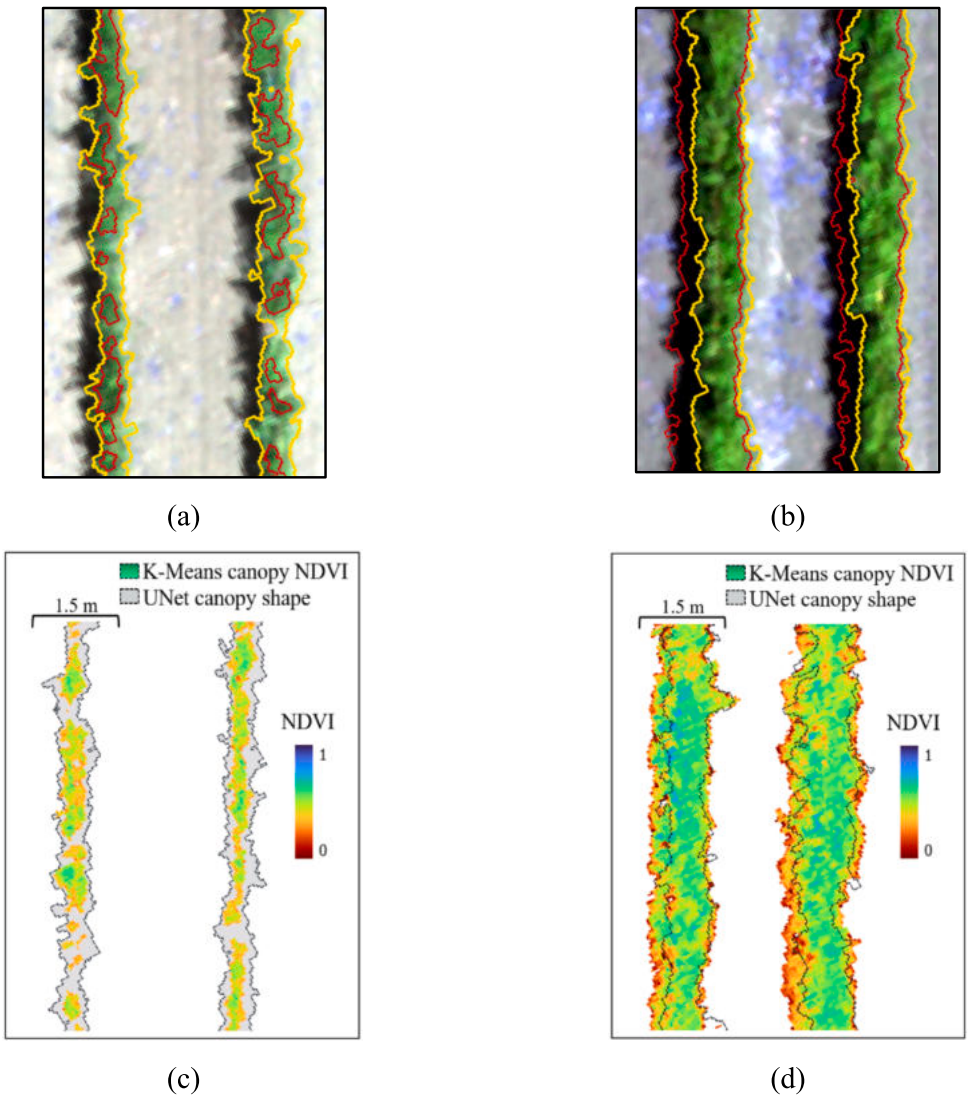


Fig. 11. Illustration of two canopy boundary masks under two different vigor conditions; (a) Overlapping of U-Net (yellow line) and K-Means (red line) canopy masks in LV vineyard zones; (b) overlapping of U-Net (yellow line) and K-Means (red line) canopy masks in HV vineyard zones. (c) graphical representation of LV vineyard canopy NDVI; (d) graphical representation of canopy NDVI in HV zones.

observed as well (Kerkech et al., 2018; Pádua et al., 2022). It is well-known that various factors can influence the accuracy of OBIA methods, including segmentation parameters, feature selection, and detailed algorithm tuning. Therefore, thorough controls are necessary to determine the optimal parameter values (Gonçalves et al., 2019). The division of the image dataset into training, validation, and testing sets was performed after a series of trials with different dataset splits. It was observed that all OBIA methods showed better performance with increasing training dataset size. The dataset splitting procedure used in this study was consistent with that of other studies, where a training dataset portion equal to or greater than 70 % was selected, and the classification accuracy demonstrated robust stability with no overfitting observed for ML models (Ye et al., 2023; Yuba et al., 2021). OBIA methods, tuning the number of elements for the training samples of each image can have a significant impact on the classification accuracy. In this study, the subdivision was made considering the study conducted by (Qian et al., 2014) in which it is observed that the different ML methods become insensitive to an increase in sample size and thus the accuracy of the models does not increase significantly by exceeding the threshold of 125 samples (Zhou et al., 2021). Therefore, the greater stability achieved by GMM and SVM methods could be attributed to their lower sensitivity to changes in the number of samples, as demonstrated by (Qian et al., 2014).

The superior performance of Mask R-CNN observed here can be justified by the use of multi-band images, an approach known to significantly enhance segmentation accuracy compared to models utilizing single-band images (Yu et al., 2022). Moreover, image spatial resolution has a direct impact on the accuracy of detection, which is notably high in this study. It is acknowledged that (CNN) are significantly influenced by spatial resolutions, and their efficiency diminishes as resolution decreases (Hao et al., 2021; Safonova et al., 2021). In line with the findings of this study, a recent comparison was made between the performance of Mask R-CNN and the OBIA-MDTWS algorithm. The experimental results affirmed the effectiveness of Mask R-CNN in segmenting crops, revealing a higher overall average F1-score of 97.63 % compared to the OBIA method (Ye et al., 2023). From vineyard investigations using RGB and NIR images, it is evident that semantic segmentation methods, including U-Net, exhibit overall superior performance compared to unsupervised segmentation methods. This aligns with the results presented in this study. However, the authors demonstrate that some of these methods can have competitive performance under specific conditions (Barros et al., 2022). In this study, U-Net accurately identifies the three classes, particularly in terms of canopy delineation. U-Net has achieved positive outcomes primarily due to the

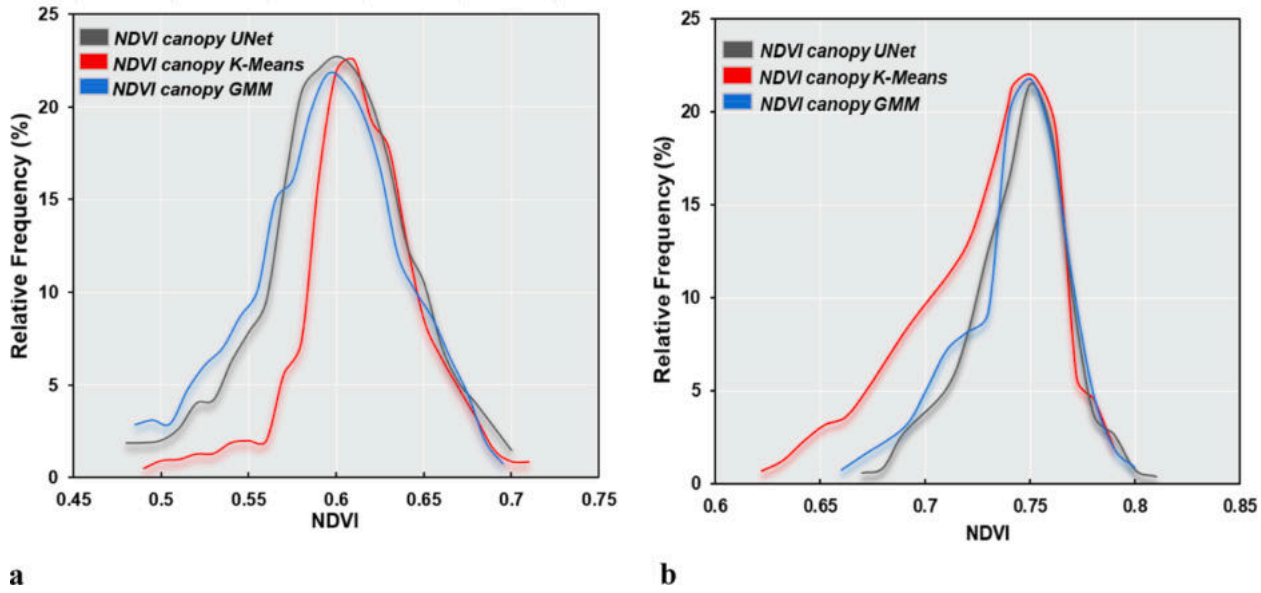


Fig. 12. NDVI histograms of three different canopy detection methods respectively for LV canopy (a) and HV canopy (b) zones of the vineyard.

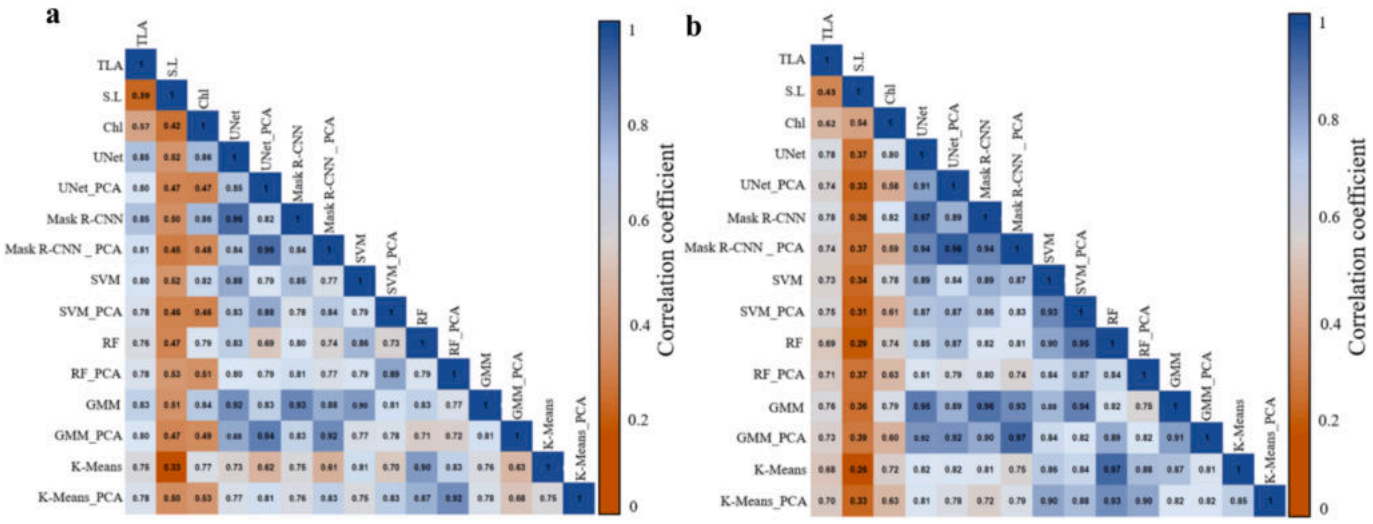


Fig. 13. Correlation matrices for the 15 variables examined. The matrices show the Pearson correlation coefficient (r) values between the agronomic variables and the NDVI and Projected Canopy Area (PCA) data identified using the six image segmentation methods. The colour intensity is proportional to the correlation coefficients, with values close to 1 coloured in blue, and values close to zero in orange. The correlation matrix in figure **a** and **b** refer the phenological period BBCH65 and BBCH85, respectively.

concatenation and upsampling structures it incorporates. The encoder and decoder layers, indeed, allow for the acquisition of detailed information from the features of the training image set (Sun et al., 2022). This result is consistent with the literature, which asserts that the use of semantic segmentation can significantly improve the efficiency of multiclass segmentation (Osco et al., 2021; Sahin et al., 2023).

4.2. Inter-relationship analysis among methods and agronomic parameters

Recent studies have revealed that semantic segmentation significantly improved the accuracy of LAI estimates compared to traditional methods (Shao et al., 2023). This is done by masking multispectral images through U-Net to generate specific canopies segmentations, follows by the computation of VI, together with analyses conducted to understand how specific parts of the canopy influence the spectral characteristics. This study conducted on an on–farm vineyard with VSP system, more accurate canopy delineation results were observed during the vegetative phase BBCH65. During this period, the general conditions of shoot growth did not have a negative impact or make the image analysis process laborious. The canopy exhibited a more distinctive morphological aspect compared to the BBCH85 stage, where the occurrence of shading and other background elements could result in a performance decrease for some methods. The canopy size affects the results of pixel-wise model delineation. Canopies with higher density tend to be oversized, and this effect has been observed for DL networks (Xi et al., 2021). However, in this study, DL networks were not significantly disadvantaged by this effect. On the contrary, traditional segmentation methods were significantly influenced, as also noted by (Ouyang et al., 2020). The variation in canopy sizes plays a crucial role in segmentation and classification processes. In this study, it was observed that differences in canopy size were not solely phenological but were also related to variations in vine vigor within the plot. The accuracy and F1-score results indicate that certain classifiers, both in OBIA methods and unsupervised

approaches, tend to exclude smaller vine canopies in LV areas. HV zones, conversely, vine rows with high canopy thickness and density tend to be oversized. The error is likely related to the presence of shadows in the images, affecting both the K-Means method (Cinat et al., 2019), and OBIA methods, such as the RF classifier (Pádua et al., 2022). The multiclass segmentation analysis allows for the identification of soil and particularly shadow influence on the overall area of pure vegetation pixels. Methods that underestimated canopy sizes in LV zones exhibited an overall higher NDVI compared to more accurate methods. In HV vineyard zones, on the other hand, methods with higher classification errors overestimated canopy pixels, resulting in lower NDVI compared to methods that achieved high accuracy and excluded the spectral response from shadows. The literature regarding this specific result, observed for pure pixels, is limited. However, the phenomenon can be explained by considering the findings from Sozzi et al., (2020) where the Gini index is taken into account. The authors conducted a multi-resolution comparison study, comparing two data acquisition platforms in a vineyard and evaluating their respective NDVI. They report that a high NDVI value is associated with a lower Gini index, and conversely, a lower NDVI is paired with a high Gini index. These results align with those identified in this study. A similar result to that identified in this study was observed by Aboutalebi et al., (2019), where the average NDVI in pure canopy pixels were higher than those in shadowed vegetation pixels. The authors applied both unsupervised classification methods and supervised OBIA techniques, with the latter achieving superior accuracy indices in shadow detection. On the other hand, the results shown by RF and K-Means methods in LV vineyard zones are a consequence of the small and constrained canopy sizes, especially in VSP systems. This phenomenon has also been observed in other studies (Hall et al., 2003; Poblete-Echeverría et al., 2017). Campos et al., 2021, observed that background elements negatively influence the pixels of the vine canopy area. The edge effect on multispectral images significantly impacts the NDVI of the vine canopy, consequently decreasing its correlation with canopy architecture parameters. Correlation matrices demonstrate positive and significant interactions in comparing NDVI data with agronomic parameters such as TLA and Chl. It has been found that the correlations between canopy NDVI and TLA, evaluated through high-resolution UAV images, can be rather low. This is explained by considering the phenomenon that canopy pixels are more influenced by the physiological activity of the vine than canopy structure (Caruso et al., 2023). Conversely, NDVI obtained from lower-resolution images, such as satellite imagery, incorporate reflectance information from both the canopy and the soil. Consequently, it is more influenced by changes in vegetation cover than by plant physiology (Gitelson et al., 2002). In this study, it was demonstrated that the correlation between pure canopy pixel NDVI and vineyard leaf area can be increased by exploiting canopy segmentation and classification methods that remove the negative influence of background effects. These observations are in line with other recent studies, which obtained higher accuracy in the estimation of LAI and leaf Chl after removing shaded canopy pixels in an orchard context (Zhang et al., 2024).

The coefficients of determination obtained between NDVI and leaf chlorophyll content indicate a high correlation, supporting the capability of NDVI to precisely assess chlorophyll content in leaves. These results are in line with the findings reported by Caruso et al., (2023), which highlighted a consistent and stable relationship between these two parameters. Furthermore, correlations related to NDVI data obtained from segmentation methods showed a decline in correlation with leaf chlorophyll content at BBCH85, in agreement with observations by Caruso et al., (2017) in which the estimation of vine leaf chlorophyll content was conducted using UAV images.

Results obtained suggest that it is possible to enhance canopy delineation and achieve an accurate estimation of spectral conditions. Improving vine canopy delineation through the application of these technologies opens new perspectives for vineyard phenotyping and the prediction of key agronomic parameters. The scientific contributions of this study primarily concern the development of more precise and reliable phenotyping systems, enabling the identification of variation in canopy growth in relation to plant vigor. A more realistic monitoring of vigor offers the possibility to accurately identify the onset of abiotic stress. Vine canopy segmentation and vigor monitoring have the potential to provide valuable information for improving grapevine growth simulation models.

4.3. Limitations and future work directions

In this study, the dataset used for training exhibited high variation in vegetation growth conditions, in terms of vigor and overall vegetation growth between low and high vigor canopies. This spatial variability primarily concerns the differing canopy thickness, and consequently also the spectral response evaluated through NDVI between LV and HV vineyard zones. For these reasons, a biennial trial was conducted, in two distinct phenological phases, with the aim of evaluating the impact of canopy growth variability on segmentation and classification processes. In the future, to obtain more accurate measurements of Chl and TLA, it is desirable to retrieve more data on canopy structure. Future studies will include sensors capable of capturing canopy biometric data through direct measurement systems rather than point-based approaches. Manual annotation of ground truth masks for model segmentation training poses a challenge due to its laborious nature in terms of time and effort. For instance, accurately delineating canopy shadows generated by sunlight in aerial photography is one of the main challenges in the labelling process. Possible future directions will be exploration of hybrid annotation approaches, combining semi-automatic methodologies, or using machine learning algorithms capable of generating initial masks. These approaches could drastically reduce the annotation workload required while simultaneously improving the flexibility and overall efficiency of the process. This study did not report the time consumed for model calculations. However, it was observed that the computationally most expensive algorithms, ML ones used in OBIA analysis, demanded more time to complete the analysis, heavily utilizing processor capabilities. ML algorithms utilize the CPU to analyze the multi-dimensional features of each object. In this study, analyzing a very large multispectral image dataset requires multiple computation cycles to segment and classify each class, leading to a significant increase in processing time, which represents an intrinsic limitation of the CPU. On the contrary, the two DL methods leverage the parallel computing power of GPU, designed to handle large amounts of data. These distribute computations across multiple cores, significantly accelerating the segmentation and classification process of classes. This research provides the baseline for future studies on the implementation of UAV-mounted multispectral sensors that could be equipped with GPUs for real-time canopy detections, providing accurate information on vineyard phenotyping or vegetation maps in real time.

5. Conclusions

The utilization of multispectral image analysis technologies captured by UAVs represents an additional asset for the comprehensive monitoring and prediction of vineyard vegetation conditions. By combining these technologies with computational models of OBIA or deep learning, it becomes possible to accurately compute VIs using various spectral bands. This study presents a workflow that includes cutting-edge deep learning methods such as Mask R-CNN and U-Net, supervised machine learning methods such as GMM, SVM, RF, and the unsupervised K-Means method. The analyses were executed employing high-resolution multispectral images acquired from UAV. The methods were employed for vine canopy delineation, aiming to investigate the potential influence of segmentation and classification algorithms on crucial parameters such as the projected canopy area and NDVI. These parameters, derived from image analysis, were subsequently correlated with agronomic data, encompassing total leaf area, shoot length, and leaf

chlorophyll content. In conclusion, the Mask R-CNN and U-Net achieved higher accuracy indices overall compared to OBIA algorithms. However, among the OBIA methods, the GMM algorithm demonstrated the lowest error rate and can be considered one of the valid OBIA methods for extracting canopy margin features. While SVM yielded acceptable accuracy, it faced challenges in efficiently separating shadows from vegetation in certain contexts. On the other hand, the RF and K-Means algorithms reported misclassifications of image components, significantly affecting the spectral data of the canopy. Specifically, for LV vineyard zones, the NDVI was overestimated, while in HV vineyard zones, the NDVI was underestimated. This result has a direct impact on the correlation with canopy agronomic parameters. The NDVI identified in canopies delineated using Mask R-CNN and U-Net methods showed a strong correlation with total leaf area and chlorophyll content. Supervised OBIA methods employing GMM and SVM algorithms exhibit stable and positive determination coefficients with agronomic variables, although marginally lower than those obtained from DL networks. This study represents an innovative approach that elucidates the practical implications of using neural networks and OBIA methods for the detailed assessment of canopy. By expanding this knowledge, vineyard vigor monitoring systems can be enhanced, providing grape growers with increasingly precise and reliable crop information.

CRediT authorship contribution statement

Massimo Vincenzo Ferro: Writing – original draft, Software, Data curation, Conceptualization. **Claus Grøn Sørensen:** Revision and Editing, Validation, Formal analysis. **Pietro Catania:** Writing – review & editing, Validation, Supervision, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by the project “SiciliAn MicronanOTecH Research And Innovation CEnter ”SAMOTHRACE“ (MUR, PNRR-M4C2, ECS_00000022), spoke 3 Università degli Studi di Palermo ”S2-COMMs - Micro and Nanotechnologies for Smart & Sustainable Communities“.

References

Aboutalebi, M., Torres-Rua, A.F., Kustas, W.P., Nieto, H., Coopmans, C., McKee, M., 2019. Assessment of different methods for shadow detection in high-resolution optical imagery and evaluation of shadow impact on calculation of NDVI, and evapotranspiration. Irrig. Sci. 37, 407–429.

Anagnostis, A., Tagarakis, A.C., Kateris, D., Moysiadis, V., Sørensen, C.G., Pearson, S., Bochtis, D., 2021. Orchard mapping with deep learning semantic segmentation. Sensors 21, 3813.

Bannari, A., Morin, D., Bonn, F., Huete, A., 1995. A review of vegetation indices. Remote Sens. Rev. 13, 95–120.

Barros, T., Conde, P., Gonçalves, G., Premebida, C., Monteiro, M., Ferreira, C.S.S., Nunes, U.J., 2022. Multispectral vineyard segmentation: A deep learning comparison study. Comput. Electron. Agric. 195, 106782.

Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 114, 24–31.

Campos, J., García-Ruíz, F., Gil, E., 2021. Assessment of Vineyard Canopy Characteristics from Vigour Maps Obtained Using UAV and Satellite Imagery. Sensors 21, 2363.

Caruso, G., Tozzini, L., Rallo, G., Primicerio, J., Moriondo, M., Palai, G., Gucci, R., 2017. Estimating biophysical and geometrical parameters of grapevine canopies (‘Sangiovese’) by an unmanned aerial vehicle (UAV) and VIS-NIR cameras. Vitis 56, 63–70.

Caruso, G., Palai, G., Tozzini, L., D’Onofrio, C., Gucci, R., 2023. The role of LAI and leaf chlorophyll on NDVI estimated by UAV in grapevine canopies. Sci. Hortic. 322, 112398.

Cerovic, Z.G., Masdoumier, G., Ghozlen, N.B., Latouche, G., 2012. A new optical leaf-clip meter for simultaneous non-destructive assessment of leaf chlorophyll and epidermal flavonoids. Physiol. Plant. 146, 251–260.

Cinat, P., Di Gennaro, S.F., Berton, A., Matese, A., 2019. Comparison of unsupervised algorithms for Vineyard Canopy segmentation from UAV multispectral images. Remote Sens. (Basel) 11, 1023.

De Luca, G., N. Silva, J.M., Cerasoli, S., Araújo, J., Campos, J., Di Fazio, S., Modica, G., 2019. Object-based land cover classification of cork oak woodlands using UAV imagery and Orfeo ToolBox. Remote Sensing 11, 1238.

Ferreira, M.P., Féret, J.-B., Grau, E., Gastellu-Etchegorry, J.-P., Do Amaral, C.H., Shimabukuro, Y.E., de Souza Filho, C.R., 2018. Retrieving structural and chemical properties of individual tree crowns in a highly diverse tropical forest with 3D radiative transfer modeling and imaging spectroscopy. Remote Sens. Environ. 211, 276–291.

Ferro, M.V., Catania, P., Miccichè, D., Pisciotta, A., Vallone, M., Orlando, S., 2023. Assessment of vineyard vigour and yield spatio-temporal variability based on UAV high resolution multispectral images. Biosyst. Eng. 231, 36–56. https://doi.org/10.1016/j.biosystemseng.2023.06.001.

Ferro, M.V., Catania, P., 2023. Technologies and Innovative Methods for Precision Viticulture: A Comprehensive Review. Horticulturae 9, 399.

Gini, C., 1921. Measurement of inequality of incomes. Econ. J. 31, 124–125.

Gitelson, A.A., Kaufman, Y.J., Stark, R., Rundquist, D., 2002. Novel algorithms for remote estimation of vegetation fraction. Remote Sens. Environ. 80, 76–87.

Gonçalves, J., Pôças, I., Marcos, B., Mücher, C.A., Honrado, J.P., 2019. SegOptim—A new R package for optimizing object-based image analyses of high-spatial resolution remotely-sensed data. Int. J. Appl. Earth Obs. Geoinf. 76, 218–230.

Grizonnet, M., Michel, J., Poughon, V., Inglada, J., Savinaud, M., Cresson, R., 2017. Orfeo ToolBox: Open source processing of remote sensing images. Open Geospatial Data, Software and Standards 2, 1–8.

Hall, A., Louis, J., Lamb, D., 2003. Characterising and mapping vineyard canopy using high-spatial-resolution aerial multispectral images. Comput. Geosci. 29, 813–822.

Hao, Z., Lin, L., Post, C.J., Mikhailova, E.A., Li, M., Chen, Y., Yu, K., Liu, J., 2021. Automated tree-crown and height detection in a young forest plantation using mask region-based convolutional neural network (Mask R-CNN). ISPRS J. Photogramm. Remote Sens. 178, 112–123.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn. In: Presented at the Proceedings of the IEEE International Conference on Computer Vision, pp. 2961–2969.

Huang, H., Lan, Y., Yang, A., Zhang, Y., Wen, S., Deng, J., 2020. Deep learning versus Object-based Image Analysis (OBIA) in weed mapping of UAV imagery. Int. J. Remote Sens. 41, 3446–3479.

Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: Presented at the International Conference on Machine Learning, pp. 448–456.

Jiménez-Brenes, F.M., Lopez-Granados, F., Torres-Sánchez, J., Peña, J.M., Ramírez, P., Castillejo-González, I.L., de Castro, A.I., 2019. Automatic UAV-based detection of Cynodon dactylon for site-specific vineyard management. PLoS One 14, e0218132.

Kerkech, M., Hafiane, A., Canals, R., 2018. Deep leaning approach with colorimetric spaces and vegetation indices for vine diseases detection in UAV images. Comput. Electron. Agric. 155, 237–243.

Lagrange, A., Fauvel, M., Grizonnet, M., 2017. Large-scale feature selection with Gaussian mixture models for the classification of high dimensional remote sensing images. IEEE Trans. Comput. Imaging 3, 230–242.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440.

Lopes, C., Pinto, P., 2005. Easy and accurate estimation of grapevine leaf area with simple mathematical models. Vitis 44, 55–61.

Lucena, F., Breunig, F.M., Kux, H., 2022. The Combined Use of UAV-Based RGB and DEM Images for the Detection and Delineation of Orange Tree Crowns with Mask R-CNN: An Approach of Labeling and Unified Framework. Future Internet 14, 275.

Michel, J., Youssefi, D., Grizonnet, M., 2014. Stable mean-shift algorithm and its application to the segmentation of arbitrarily large remote sensing images. IEEE Trans. Geosci. Remote Sens. 53, 952–964.

Modica, G., De Luca, G., Messina, G., Praticò, S., 2021. Comparison and assessment of different object-based classifications using machine learning algorithms and UAVs multispectral imagery: a case study in a citrus orchard and an onion crop. European Journal of Remote Sensing 54, 431–460. https://doi.org/10.1080/22797254.2021.1951623.

Moghimi, A., Pourreza, A., Zuniga-Ramirez, G., Williams, L.E., Fidelibus, M.W., 2020. A Novel Machine Learning Approach to Estimate Grapevine Leaf Nitrogen Concentration Using Aerial Multispectral Imagery. Remote Sens. (Basel) 12, 3515. https://doi.org/10.3390/rs12213515.

Novara, A., Pisciotta, A., Minacapilli, M., Maltese, A., Capodici, F., Cerdà, A., Gristina, L., 2018. The impact of soil erosion on soil fertility and vine vigor. A multidisciplinary approach based on field, laboratory and remote sensing approaches. Sci. Total Environ. 622, 474–480.

Osco, L.P., Nogueira, K., Marques Ramos, A.P., Faita Pinheiro, M.M., Furuya, D.E.G., Gonçalves, W.N., de Castro Jorge, L.A., Marcato Junior, J., dos Santos, J.A., 2021. Semantic segmentation of citrus-orchard using deep neural networks and multispectral UAV-based imagery. Precis. Agric. 22, 1171–1188.

Ouyang, J., De Bei, R., Fuentes, S., Collins, C., 2020. UAV and ground-based imagery analysis detects canopy structure changes after canopy management applications. Oeno One 54, 1093–1103.
Pádua, L., Marques, P., Hruška, J., Adão, T., Bessa, J., Sousa, A., Peres, E., Morais, R., Sousa, J.J., 2018. Vineyard properties extraction combining UAS-based RGB imagery with elevation data. Int. J. Remote Sens. 39, 5377–5401.
Pádua, L., Matese, A., Di Gennaro, S.F., Morais, R., Peres, E., Sousa, J.J., 2022. Vineyard classification using OBIA on UAV-based RGB and multispectral data: A case study in different wine regions. Comput. Electron. Agric. 196, 106905.
Poblete-Echeverría, C., Olmedo, G.F., Ingram, B., Bardeen, M., 2017. Detection and segmentation of vine canopy in ultra-high spatial resolution RGB imagery obtained from unmanned aerial vehicle (UAV): A case study in a commercial vineyard. Remote Sens. (Basel) 9, 268.
Qian, Y., Zhou, W., Yan, J., Li, W., Han, L., 2014. Comparing machine learning classifiers for object-based land cover classification using very high resolution imagery. Remote Sens. (Basel) 7, 153–168.
Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems 28.
Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. Presented at the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, Springer, pp. 234–241.
Rouse Jr, J., Haas, R.H., Deering, D., Schell, J., Harlan, J.C., 1974. Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation.
Safonova, A., Guirado, E., Maglinets, Y., Alcaraz-Segura, D., Tabik, S., 2021. Olive tree biovolume from UAV multi-resolution image segmentation with mask R-CNN. Sensors 21, 1617.
Sahin, H.M., Miftahushudur, T., Grieve, B., Yin, H., 2023. Segmentation of weeds and crops using multispectral imaging and CRF-enhanced U-Net. Comput. Electron. Agric. 211, 107956.
Sarabia, R., Aquino, A., Ponce, J.M., López, G., Andújar, J.M., 2020. Automated identification of crop tree crowns from uav multispectral imagery by means of morphological image analysis. Remote Sens. (Basel) 12, 748.
Scikit-Learn, 2022.
Shao, M., Nie, C., Zhang, A., Shi, L., Zha, Y., Xu, H., Yang, H., Yu, X., Bai, Y., Liu, S., 2023. Quantifying effect of maize tassels on LAI estimation based on multispectral imagery and machine learning methods. Comput. Electron. Agric. 211, 108029.
Sinaga, K.P., Yang, M.-S., 2020. Unsupervised K-Means Clustering Algorithm. IEEE Access 8, 80716–80727.
Sozzi, M., Kayad, A., Marinello, F., Taylor, J., Tisseyre, B., 2020. Comparing vineyard imagery acquired from Sentinel-2 and Unmanned Aerial Vehicle (UAV) platform. Oeno One 54, 189–197.
Sun, Q., Zhang, R., Chen, L., Zhang, L., Zhang, H., Zhao, C., 2022. Semantic segmentation and path planning for orchards based on UAV images. Comput. Electron. Agric. 200, 107222.
Torralba, A., Russell, B.C., Yuen, J., 2010. Labelme: Online image annotation and applications. Proc. IEEE 98, 1467–1484.
Towers, P.C., Poblete-Echeverría, C., 2021. Effect of the illumination angle on NDVI data composed of mixed surface values obtained over vertical-shoot-positioned vineyards. Remote Sens. (Basel) 13, 855.
Vapnik, V., 2006. Estimation of dependences based on empirical data. Springer Science & Business Media.
Xi, X., Xia, K., Yang, Y., Du, X., Feng, H., 2021. Evaluation of dimensionality reduction methods for individual tree crown delineation using instance segmentation network and UAV multispectral imagery in urban forest. Comput. Electron. Agric. 191, 106506.
Ye, Z., Yang, K., Lin, Y., Guo, S., Sun, Y., Chen, X., Lai, R., Zhang, H., 2023. A comparison between Pixel-based deep learning and Object-based image analysis (OBIA) for individual detection of cabbage plants based on UAV Visible-light images. Comput. Electron. Agric. 209, 107822 https://doi.org/10.1016/j.compag.2023.107822.
Yu, K., Hao, Z., Post, C.J., Mikhailova, E.A., Lin, L., Zhao, G., Tian, S., Liu, J., 2022. Comparison of classical methods and mask R-CNN for automatic tree detection and mapping using UAV imagery. Remote Sens. (Basel) 14, 295.
Yuba, N., Kawamura, K., Yasuda, T., Lim, J., Yoshitoshi, R., Watanabe, N., Kurokawa, Y., Maeda, T., 2021. Discriminating Pennisetum alopecuoides plants in a grazed pasture from unmanned aerial vehicles using object-based image analysis and random forest classifier. Grassl. Sci. 67, 73–82.
Zhang, C., Chen, Z., Yang, G., Xu, B., Feng, H., Chen, R., Qi, N., Zhang, W., Zhao, D., Cheng, J., 2024. Removal of canopy shadows improved retrieval accuracy of individual apple tree crowns LAI and chlorophyll content using UAV multispectral imagery and PROSAIL model. Comput. Electron. Agric. 221, 108959.
Zheng, H., Zhou, X., He, J., Yao, X., Cheng, T., Zhu, Y., Cao, W., Tian, Y., 2020. Early season detection of rice plants using RGB, NIR-GB and multispectral images from unmanned aerial vehicle (UAV). Comput. Electron. Agric. 169, 105223.
Zhou, X., Yang, L., Wang, W., Chen, B., 2021. Uav data as an alternative to field sampling to monitor vineyards using machine learning based on uav/sentinel-2 data fusion. Remote Sens. (Basel) 13, 457.
Zhu, C., Ding, J., Zhang, Z., Wang, J., Wang, Z., Chen, X., Wang, J., 2022. SPAD monitoring of saline vegetation based on Gaussian mixture model and UAV hyperspectral image feature classification. Comput. Electron. Agric. 200, 107236.