

Automated urban tree survey using remote sensing data, Google street view images, and plant species recognition apps

Irene Capecchi, Tommaso Borghini & Iacopo Bernetti

To cite this article: Irene Capecchi, Tommaso Borghini & Iacopo Bernetti (2023) Automated urban tree survey using remote sensing data, Google street view images, and plant species recognition apps, European Journal of Remote Sensing, 56:1, 2162441, DOI: [10.1080/22797254.2022.2162441](https://doi.org/10.1080/22797254.2022.2162441)

To link to this article: <https://doi.org/10.1080/22797254.2022.2162441>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 19 Jan 2023.



Submit your article to this journal



Article views: 2189



View related articles



CrossMark

View Crossmark data

Automated urban tree survey using remote sensing data, Google street view images, and plant species recognition apps

Irene Capecchia^a, Tommaso Borghini^b and Iacopo Bernetti^a

^aDepartment of Agriculture, Food, Environment and Forestry (DAGRI), University of Florence, Florence, Italy; ^bDepartment of Architecture (DIDA), University of Florence, Florence, Italy

ABSTRACT

Urban tree inventories have mostly focused on the information of individual trees because these allows city authorities to efficiently plan urban forestation. However, single-tree urban tree inventories are expensive for municipalities, so the inventories lack detail and are often out of date. In this work, we aim to integrate the possibility of using online applications for automatic species identification with worldwide coverage Pl@ntNet and Plant.Id on Google Street View (GSV) images in order to perform cost-effective urban tree inventories at the single-tree level and evaluate the performance of the two applications through comparison with a locally trained neural network using an appropriate set of metrics. Our work showed that the Plant.Id application gave the best performance by correctly identifying plants in the city of Prato with a median accuracy of 0.73 and better performance for the most common plants: *Pinus pinea* 0.87, *Tilia aeuropea* 0.87, *Platanus hybrida* 0.89. The proposed method also has a limitation. Trees within parks, walking paths and private green areas cannot be photographed and identified because Google cars cannot access them. The solution to this limitation is to combine GSV images with spherical photos taken via light unmanned aircraft.

ARTICLE HISTORY

Received 7 September 2022
Revised 17 December 2022
Accepted 21 December 2022

KEYWORDS

Pl@ntnet; Plant.Id;
GoogLenet; urban green;
Google street view;
convolutionary neural
network

Introduction

Urban tree inventories are a key tool for urban planning especially in the context of global climate change trends. (Padayachee et al., 2017). Urban tree inventories have mostly focused on single-tree information rather than surveying forest stands Östberg et al. (2013) because having information on individual trees allows city authorities regarding efficient urban forestation planning based on species selection, risk management, and subsequent replanting decisions (Keller & Konijnendijk, 2012). However urban single-tree inventories are generally expensive for municipalities so, despite their importance, inventories lack detail and are often out of date due to the costs associated with mapping and monitoring trees over time and over large areas (Nielsen et al., 2014).

Nielsen et al. (2014) distinguish four types of inventories for data collection at the individual tree level: 1) satellite remote sensing, 2) aircraft remote sensing, 3) field scanning or digital photography, and 4) field surveys with direct hand measurements and/or visual assessment.

Satellite or airborne remote sensing-based methods can cost-effectively collect information over very large areas (Cook & Iverson, 1991; Small & Lu, 2006). Very high-resolution multispectral imagery can also be used to collect information at the individual tree level

(Jansen et al. 2006). Combining multispectral and LiDAR data can also be used to segment individual tree crowns (Alonzo et al., 2014; Wallace et al., 2021).

Compared to remote sensing methods, data collection and processing from digital scanning (Patterson et al., 2011) or ground-based photography (proximate sensing) is limited to small areas because each scan/photo image is limited to a single tree or small group of trees. Although this technology is developing rapidly, it is still time consuming. Recently, methods based on images from Google Street View™ (GSV) have been developed to conduct a virtual inventory of street trees (Berland & Lange, 2017, Barbierato et al. 2021).

Field survey methods include dendrometric and/or phytopathological surveys on individual trees by volunteers or professional staff. (Adkins et al., 1997; Martin, 2011; Östberg et al., 2012). Although field surveys are labour-intensive and time-consuming, this inventory method is the most adopted (Nielsen et al., 2014).

Species information has been identified as the most important data parameter in single tree level inventories (Östberg et al., 2013). Tree species classification from remote sensing data uses satellite multispectral aerial imagery (Leckie et al., 2005; Waser et al., 2010; Pu and Landry, 2012), hyperspectral data (Clark et al., 2005; Roth et al., 2015), dense LiDAR point clouds (;

CONTACT Iacopo Bernetti  iacopo.bernetti@unifi.it  Department of Agriculture, Food, Environment and Forestry (DAGRI), University of Florence, piazzale delle Cascine 18, 50144 Florence, Italy

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Yao & Wei, 2013), or a combination of LiDAR and multispectral data (Heikkinen et al., 2011; Korpela et al., 2010; Heinzel & Koch, 2011).

Most of the more recent works follows the classic processing method based on artificial intelligence techniques: extract a small set of structure and shape features from the images and/or Aerial Light Detection and Ranging (LiDAR) data, and train a classifier (e.g. Linear Discriminant Analysis, Support Vector Machines, or more recently Deep Learning) to distinguish among a small number of species 3 in (Korpela et al., 2010); Leckie et al., (2005); (Heikkinen et al., 2011), 4 in (Heinzel & Koch, 2011), 7 in (Waser et al., 2010; Pu and Landry, 2012). Recently, some authors (Branson et al., 2018; Ringland et al. 2021) applied tree detection and species recognition methods using publicly available Google Maps(TM) aerial and street view images by applying convolutional neural networks (CNNs). However, no studies have been conducted on the transferability of the methodology to other cities worldwide.

Over the past 10 years, much research of deep learning image recognition approaches for plant identification has been published. (see e.g. Wäldchen & Mäder, 2018). The development of many trained convolutional neural network stems from the Cross Language Evaluation Forum (CLEF) initiative (<http://www.clef-initiative.eu/association>), which since 2013 has included the LifeCLEF challenge to develop automatic identification systems for living organisms. The PlantCLEF sub-project has focused on plant identification (Goëau et al. 2013; Cappellato et al. 2017), with experiments aimed at comparing the performance of “experts” with that of the best deep learning algorithms (Bonnet et al. 2018).

In this paper, we aim to integrate the possibility of using automatic species identification apps with worldwide coverage on Google Street View images to be able to perform cost-effective urban green inventories at the individual tree level. In detail, the objectives of the research are:

- (1) identify a methodology to extract images of urban trees from GSV using LiDAR and multispectral data, without the need for other ancillary data;
- (2) to classify plant species in GSV images through the two globally available applications that can be queried through a programming interface: Pl@ntNet and Plant.Id;
- (3) compare the classification performances of the two apps with those of a Convolutional Neural Network (CNN) trained on the study area using an appropriate set of metrics;
- (4) evaluate the benefits and limitations of an automated urban green inventory integrating LiDAR data, GSV images, and tree species classification apps.

Materials

Study area

Prato (Figure 1) is a city in Tuscany (Italy) with 200,647 inhabitants and an area of 44.37 square kilometres placed at coordinates 43°52'50.93"N 11°05'47.62"E. Prato is the third largest city in central Italy after Rome and Florence, thanks to the immigrants that arrived first from the countryside, then from southern Italy. The city's climate is characterized by rather cold and moderately dry winters and hot and sometimes sultry summers.

The city of Prato has an urban greenspace of 62.56 square kilometers, of which 13.77hectares are public urban trees. There are 3 protected natural areas in the city totalling 11.63 square kilometres (ISTAT, 2019). According to official statistics (ISTAT, 2019) Prato is one of the greenest cities in Italy with 14.2% urban green area, compared to 13.2% in Florence and 9.6% in Rome. According to the city's public green census (Figures 2 and 3), there are 147 different species in the city of Prato, and 27 species are present with at least 50 plants. The predominant species (Figure 2) is *Pinus pinea*, with almost 27% of the total green area covered, followed by *Tilia x europaea* with 16% and *Platanus x acerifolia* with almost 10%.

The city of Prato is one of the most active Italian cities in urban green planning. Since 2021, Prato has adopted the “Prato Urban Jungle” reforestation plan. The “Prato Urban Jungle” project will redesign the city's neighbourhoods in a sustainable and inclusive way, developing high-density green areas that will be incorporated into the urban landscape, multiplying the natural ability of plants to abate pollutants, transforming areas of urban marginality into green places of well-being within the city. Urban jungles will be co-designed with the help of citizens, through shared urban planning facilitated by the use of digital platforms. Implementation of the plan will involve the planting of 190,000 trees in highly urbanized areas to create multifunctional ecological spaces and corridors that generate urban renaturalization processes.

Automated image recognition apps for plant identification

Over the past 20 years, much progress has been made in the development of image recognition/AI approaches for plant identification. Much effort has been focused on the Cross Language Evaluation Forum (CLEF) initiative, which with the PlantCLEF subproject has focused on plant identification (Goëau et al., 2021), with many different research groups contributing their models from 2011 to 2022, all with the goal of comparing the performance of “experts” with that of the best deep learning algorithms (Bonnet et al. 2018). Image recognition technology is maturing

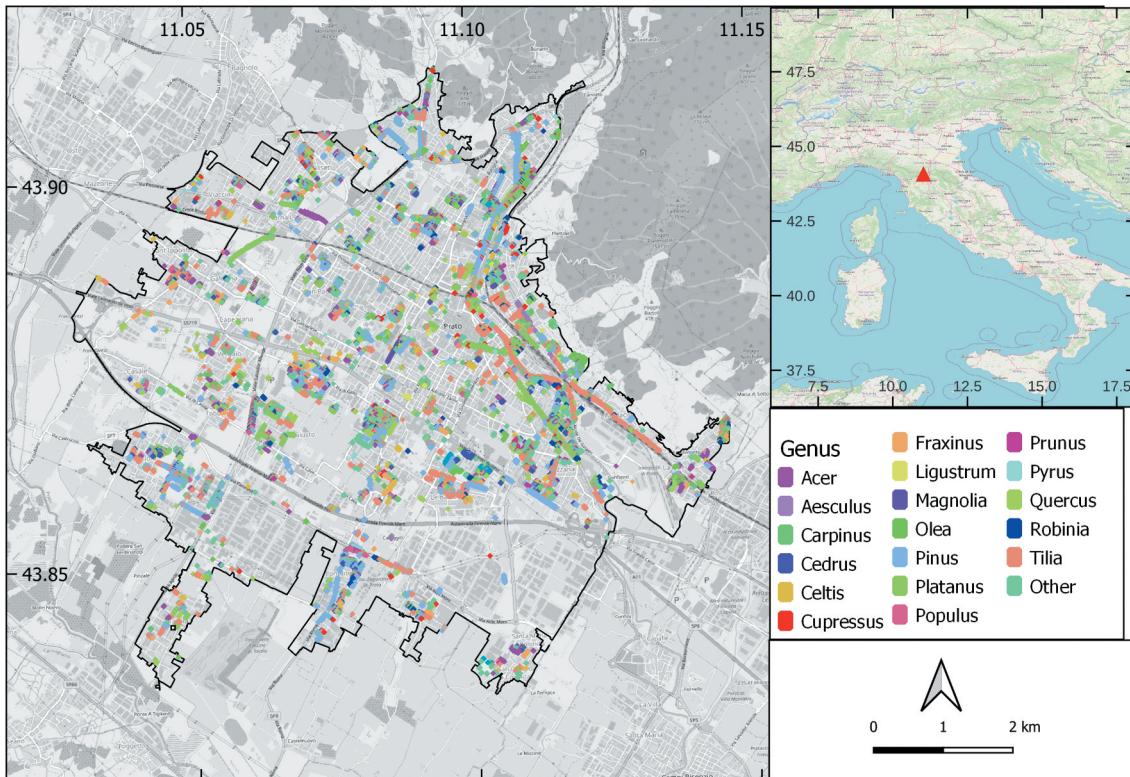


Figure 1. The city of prato (Points data: Census of public urban trees fo municipality of prato; basemap: OpenStreetmap).

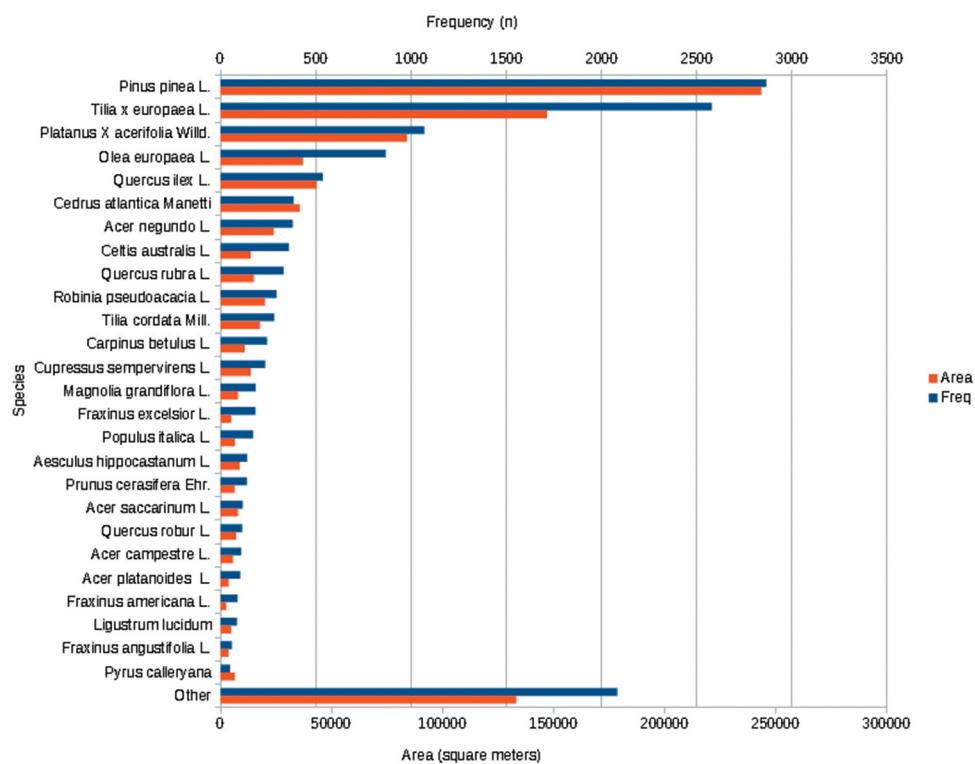


Figure 2. Frequency distribution and total area covered by major tree species (species with at least 50 plants).

so rapidly that numerous automatic plant identification apps are now available for Personal Computers, smartphones, and tablets, so it is worth considering the state of the art of this technology (Jakuschona et al., 2022).

The two apps chosen for the present research, Plant.Id and Pl@ntNet had the following advantages: (a) availability of an Application Programming Interface (API) for personal computers; (b) good performance in recognizing photographic images of

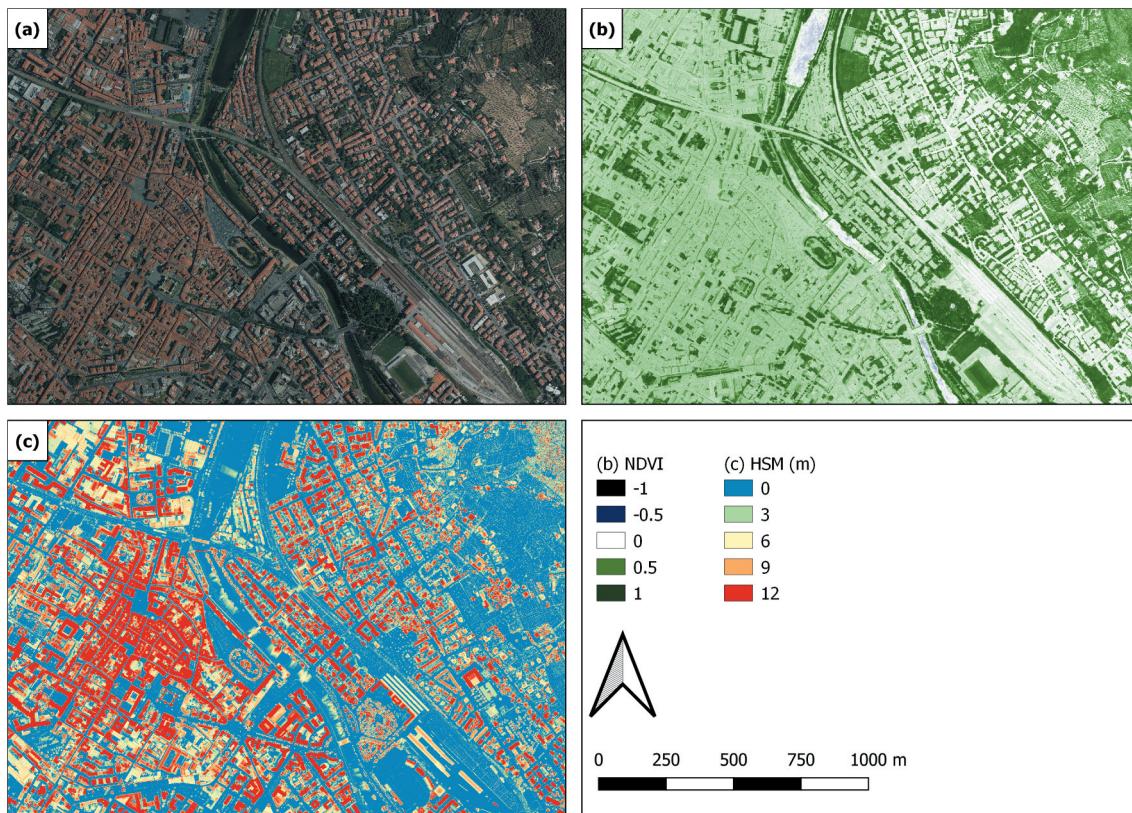


Figure 3. Small examples of the raster layers used: (a) RGB orthophoto from UltraCam Xp camera, (b) Normalized difference vegetation index (NDVI) from UltraCam Xp data and (c) LiDAR layer.

plants from standardized datasets (Jones, 2020; Jakuschona et al., 2022).

The Pl@ntNet project is an app made by a consortium including CIRAD, INRA, INRIA, IRD and the Agropolis Foundation, is a tool that supports image-based plant identification for both amateurs but especially professionals. The model behind the API is updated monthly both in terms of training data and new training architecture. Pl@ntNet's identification service is a RESTful JSON API, which returns the list of species corresponding to the query, each associated with the classification score emitted by the deep learning model. For each species the scientific name, common name, and genus and family name of the identified plant is provided.

Plant.id is a project developed by the FlowerChecker team, whose main goal is to facilitate the monitoring of invasive and endangered species for a wide range of use scenarios, from businesses to private use. The API is based on TensorFlow, Python and AWS technologies. For matching images, the API returns results with predictions about the species represented in the image and additional information about the species, such as potential plant diseases. For Plant.id, it is also possible to specify geographic coordinates of plant location, as it significantly improves the accuracy of classification. Again, to increase the efficiency Plant.Id API allows multiple images of the same plant to be uploaded. API returns the scientific

and common name. Plant.Id also associates each proposed prediction with a prediction score. Plant.id can also identify whether the plant is affected by a disease and provides additional details about the disease.

Remote sensing and other spatial data

The spatial data used in the research are derived from remote sensing (multispectral orthophotos and data derived from LiDAR survey) and field survey, regarding the public urban green census of the municipality of Prato.

The remotely sensed data, both of which were downloaded from the Tuscany Region's mapping portal. The multispectral aerial frames have four bands: red, green, and blue + near-infrared (NIR) regions and were acquired in 2019 by means of a digital metric camera UltraCam Xp (Vexcel), with a resolution of 0.2×0.2 m (Table 1). The spectral sensitivity of red, green, blue and near infrared and the panchromatic channel PAN from 410 nm to 690 nm, RED from 580 nm to 700 nm, GREEN from 480 nm to 630 nm, BLUE from 410 nm to 570 nm and NIR from 690 nm to 1000 nm.

The LiDAR data were provided by the Italian Ministry of the Environment, Land and Sea. The points acquired from this survey have an altimetric accuracy of ± 15 cm and a planimetric accuracy is ± 30 cm. In this work the

Table 1. Features of UltraCam Xp (Vexcel) camera.

Channel	From (nm)	To (nm)
PAN	410	690
RED	580	700
GREEN	480	630
BLUE	410	570
Near Infrared	690	1000

Large image format of 17,310 pixels cross track and 11,310 pixels along track.

Optical system with 100 mm focal length for the panchromatic camera heads and 33 mm for the multi spectral camera heads.

data available by the geographical portal of the Tuscan region with a resolution of 1×1 meter were used. That resolution was considered satisfactory for the objectives of the work. If necessary, however, the proposed method could therefore also be applied to more detailed scales. The LiDAR survey data consist of two datasets: Digital Surface Model (DSM) and Digital Terrain Model (DTM). All input raster were aligned at 1×1 m resolution using polynomial warp algorithm.

The census of public urban trees was conducted by GPS survey in 2017 and is accessible for download through the open data network of the municipality of Prato. The dataset primitive is pointwise, and each point is associated with the scientific name of the plant and the coordinates of the trunk with reference system EPSG 3003. Figure 3 shows some small examples of the raster layers used.

Methods

The workflow followed in the research is as follows (see Figure 4):

1) Data Production

a) Identification of tree morphological parameters:

- (i) Segmentation of tree crowns using a watershed segmentation algorithm for automatic extraction of tree crowns from LiDAR and Normalized Difference Vegetation Index (NDVI) data.
- (ii) Extraction of tree geometric features: (i) height, (ii) crown diameter, (iii) x and y coordinates of the crown center of gravity.
- (iii) Ground identification of genus and species for trees visible from public roads.

b) The harvesting of canopy images from Google Street View (GSV):

- (i) downloading of all GSV photo codes in the city of Prato;
- (ii) identification for each tree of the GSV codes of the nearest spherical photos.
- (iii) calculation of the zenith and elevation angle of the canopy with respect to each GSV photo point;

- (iv) download of ground images of the crowns from GSV;

2) Processing

- (a) Identification of tree genus and species with Plant.Id and Pl@ntNet apps;
- (b) Identification of tree genus and species with a GoogLeNet CNN.

3) Validation

- (a) Evaluation of classification efficiency by comparing app and CNN results with the tree census of Prato municipality using a set of classification performances metrics.

Crown model segmentation

The basis for modern LiDAR-based forest measurements is based on the acquisition of three surfaces, namely the crown height model (CHM), digital terrain model (DTM), and digital surface model (DSM) (Hyppä et al., 2017). To segment the crowns, we applied the geographic watershed algorithm to the CHM. The watershed algorithm is the typical procedure applied to CHMs to extract individual tree crowns of woody vegetation (Chen et al., 2006; Hyppä et al., 2001). Based on the similarity between geographic reliefs and tree crown surfaces, the watershed segmentation approach (Jing et al., 2012; Silván-Cárdenas, 2012) is widely used to segment images for tree crown delineation. In applying the procedure to the canopy, it is first necessary to invert the filtered CHM such that the highest value becomes the lowest and vice versa. In order to apply the algorithm in urbanized area, it was necessary to create a green mask to remove the artificialized areas from the analysis. We created a mask by calculating the NDVI by adopting a threshold value of 0.6 identified from the literature related to the sensor used for imaging (Alvarez et al., 2010). In order to separate the areas of shrubs and grasslands, we included a condition in the mask that the Digital Height Model (DHM), calculated as the difference between DSM and DTM, should be greater than 3 meters. Formally:

$$M = \begin{cases} 1, & \frac{NIR-Red}{NIR+Red} \geq 0.6 \cap DHM \geq 3 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

with M binary mask, NIR infrared band, Red band. Finally

$$CHM = DHM \cdot M \quad (2)$$

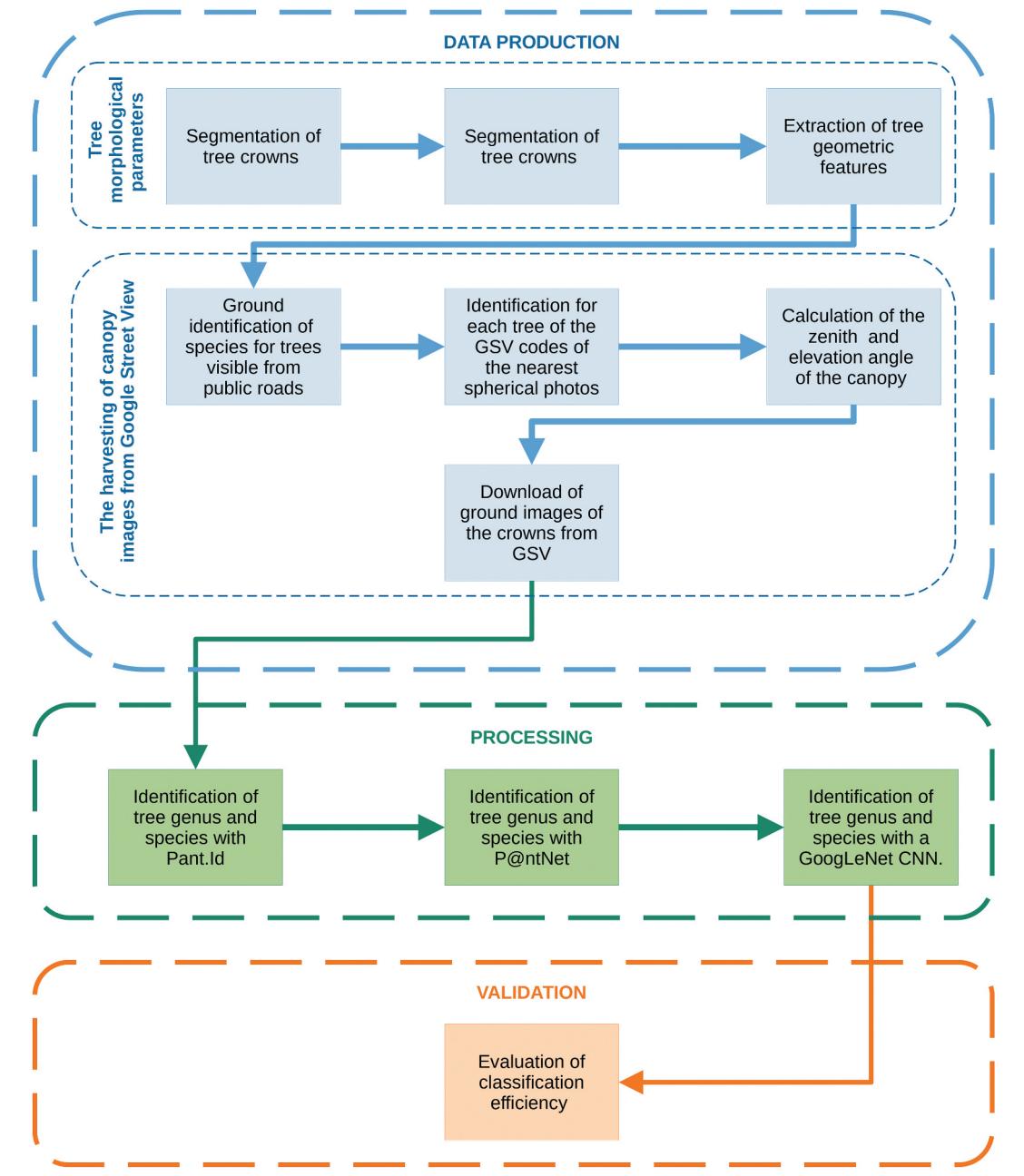


Figure 4. Flow-chart of methodology.

The CHM obtained by mask analysis was subjected to a two-dimensional Gaussian filter (2D):

$$G = \frac{1}{2\pi\sigma} e^{\frac{r^2}{2\sigma^2}} \quad (3)$$

$$CHM_G = CHM \cdot G \quad (4)$$

With G Gaussian filter, σ standard deviation, r window radius, fixed at a value of 2 pixels, like the methods proposed by Persson et al. (2002), Brandtberg et al. (2003) and Falkowski et al. (2006). By applying such Gaussian filtering to the CHM we made the tree crowns with smaller sizes were better outlined; at the same time those with larger sizes became more regular (Falkowski et al.,

2006). For watershed segmentation method we used SAGA software (Conrad et al., 2015). To identify the canopies of public green areas, we performed a map overlay operation with urban public areas. Through the canopy geodatabase maximum diameter of the canopy (maxCanopy) was calculated for each Tree element based on the coordinates of the bounding boxes. Finally, through a map overlay operation with the CHM we assigned the tree height (Htree). Thus, the final structure of the database is as follows:

$$\text{Tree}(\text{TreeId}, \text{Species}, \text{Tlat}, \text{Tlon}, \text{maxCanopy}, \text{Htree}) \quad (5)$$

with: TreeId, tree identifier; Species, scientific name of the tree; Tlat and Tlon, longitude and latitude of the center of gravity of the canopy in EPSG:32632; maxCanopy, maximum diameter of the canopy; Htree, height of the tree.

Street level imagery

GSV is a Google service (Google, 2014) that provides 360° horizontal and 180° vertical panoramic views along streets, 10–20 meters apart, and is available to most nations around the world. Through a specific Application Programming Interface, Street View Static API (SWSAPI), square portions of the panoramic images can be downloaded. By specifying different parameters in the SWSAPI, users can download GSV images of different locations, direction angles, and pitch angles. Figure 5 shows the parameters needed to locate a specific portion of the panoramic views: heading indicates the azimuth angle (heading values range from 0 to 360), pitch specifies the elevation angle relative to the ground plane, and FOV determines the horizontal field of view of the image.

To identify the parameters of the SWSAPI related to urban trees in the study area, we used the following procedure.

By querying the SWSAPI with the metadata option, it was possible to obtain a points geodatabase with the Id of the panoramic photo and the geographic coordinates of the photo's shooting point closest to each tree surveyed on the ground.

To avoid obstacles standing between GSV's car and the tree, we performed a viewshed analysis using the digital height model obtained from the LiDAR data as the DEM and GSV's shot points as the observation points. The analysis was performed with the "Viewshed" module of QGIS software by setting a search radius of 30 meters (considered appropriate to obtain images with satisfactory detail) and a GSV

car height of 3 meters. Through the intervisibility layer, only trees visible from the shot points were selected. A GSV snapshot point database was then formed with the following structure:

$$\text{GSV}(\text{TreeId}, \text{PanoId}, \text{Glat}, \text{Glon}) \quad (6)$$

with *PanoId* identifying the GSV panoramic image, *Glat* and *Glon* coordinates of the image shot point in EPSG:32632 - WGS 84/UTM zone 32N reference system.

The two databases *Tree* and *GSV* were merged via the *TreeId* field obtaining the database *TreeGSV*:

$$\text{TreeGSV}(\text{TreeId}, \text{Tlat}, \text{Tlon}, \text{PanoId}, \text{Glat}, \text{Glon}, \text{maxCanopy}, \text{Htree}) \quad (7)$$

For each *TreeGSV* record, the parameters of SWSAPI were identified as shown in Figure 6:

$$\text{pitch} = \arcsin\left(\frac{\frac{h_{\text{tree}} - h_{\text{GSV}}}{2}}{\sqrt{(Glat - Tlat)^2 + (Glon - Tlon)^2}}\right); \quad (8)$$

with $h_{\text{GSV}} = 2.5$ meters panoramic camera height of GSV;

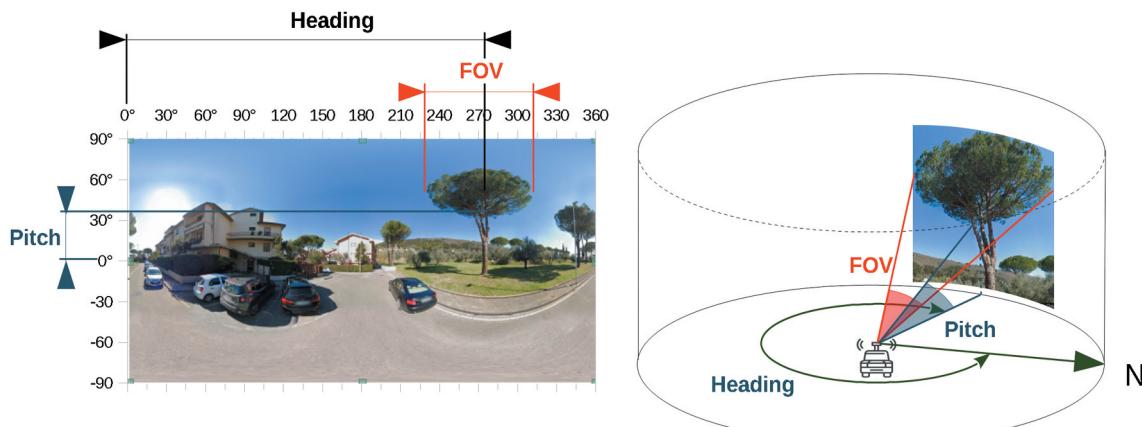
$$\text{FOV} = \arcsin\left(\frac{\text{maxCanopy}/2}{\sqrt{(Glat - Tlat)^2 + (Glon - Tlon)^2}}\right) \cdot 2; \quad (9)$$

$$\text{heading} = \arctan\left(\frac{Glat - Tlat}{Glon - Tlon}\right). \quad (10)$$

Thus, the final database has the following structure:

$$\text{TreeGSV}(\text{TreeId}, \text{Tlat}, \text{Tlon}, \text{PanoId}, \text{Glat}, \text{Glon}, \text{maxCanopy}, \text{Htree}, \text{pitch}, \text{FOV}, \text{heading}). \quad (11)$$

The Tree database data were used to download tree crown images from GSV via the API. In the case of



panoId=OlkaF8xknk5dXC9OR_f3Cg; lat=43.875160353942903; lon= 11.118892430225699

Figure 5. Google street view static API parameters.

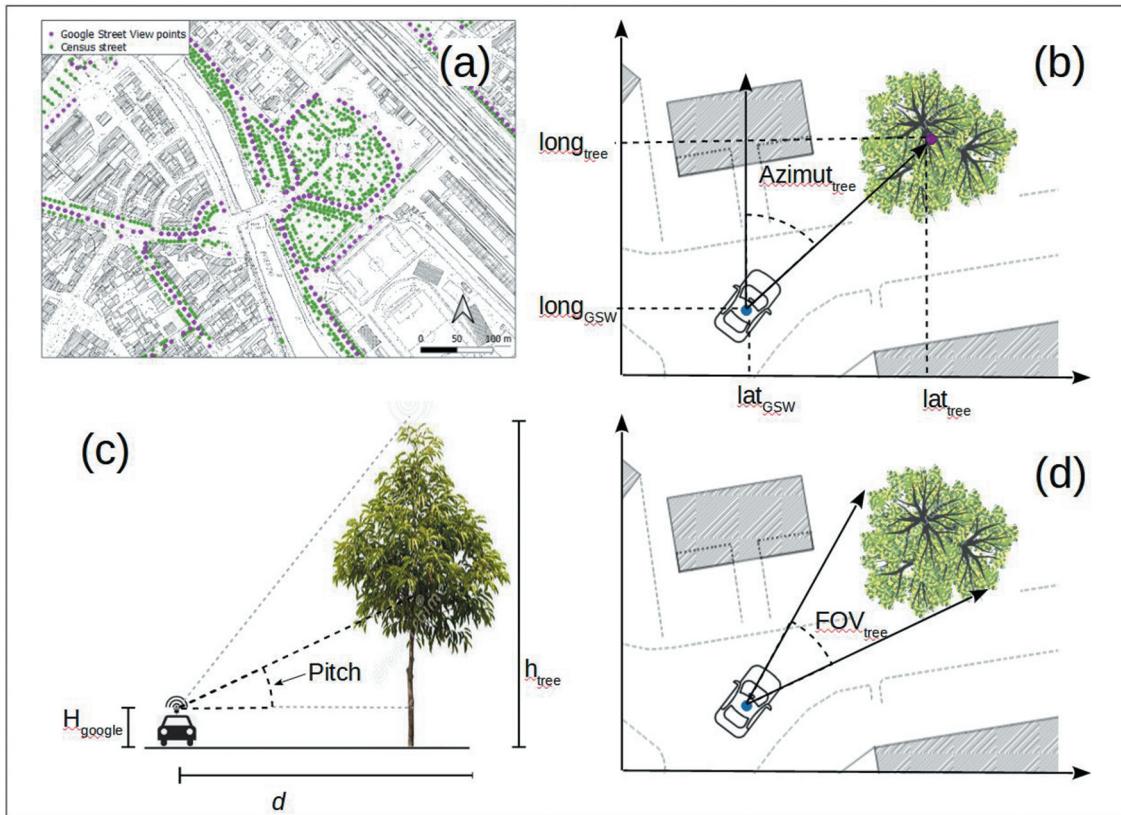


Figure 6. Google street view static API parameter calculation procedures. (a) example map; (b) azimuth angle; (c) pitch angle; (d) field of view (FOV).

Plant.Id, since the model accepted up to 5 images of the same tree are also downloaded for each tree a second image by setting $FOV = 20$ and leaving pitch and azimuth unchanged, so as to have also a more detailed detail of the foliage in the center of the canopy. To access the API we used a procedure based on the googleway library of the software R.

To provide classification apps with images of sufficient quality, we examined the actual presence of greenery in the photos collected by GSV. We estimated the total area covered by trees in each image by applying semantic segmentation with deep learning (for more details on the procedure, see Barbierato et al., 2020). A semantic segmentation network classifies each pixel in an image, resulting in an image segmented by class. In this phase of the work, we used the MatLab software-based Deeplabv3+ pre-trained network, which is a type of convolutional neural network (CNN) designed for semantic image segmentation (Brostow et al., 2009), with weights initialized from a pre-trained Resnet-18 network. ResNet-18 is an efficient network, suitable for applications with limited computational resources. The network is from MatLab software was trained using the University of Cambridge's CamVid dataset (Zhang et al., 2010), a collection of images containing street-level views. There are 11 image segmentation classes: "Sky", "Building", "Pole", "Road", "Sidewalk", "Tree", "SignSymbol", "Fence", "Car", "Pedestrian", and

"Cyclist". Once all images were classified, we then filtered those with at least 50% pixels classified as "Tree", as on the basis of pre-processing, this limit was considered satisfactory to achieve efficient classification.

Tree species classification by Pl@ntNet and Plant.Id

Images of tree crowns were sent to the Pl@ntNet and Plant ID APIs using R software with the rjson, httr, and base64 libraries. In the case of Pl@ntNet, one image was used for each API access, while in the case of Plant.Id for each query, the two images with $FOV=FOV$ and $FOV = 20$ along with the coordinates of the crown center of gravity were sent.

For each tree in the Tree database, the result was obtained in json format. The models return a list of suggested species, but we considered only the result with the best score to evaluate the accuracy of classification.

Tree species classification by convolutional neural networks

Classification of tree species was carried out using the GoogLeNet CNN. As pioneered by Wegner et al. (2016), the GoogLeNet CNN model offered the best

trade-off in terms of recognition performance, execution time and memory consumption.

We resized the GSV images to 400×400 pixels, and the data were divided into a training set and a validation set with a ratio of 90 to 10. In the training parameter settings, to avoid overfitting, the maximum epoch value was set to 500 and the initial learning rate to 0.0003.

To expand the size of the GSV image set, the data augmentation method was performed. We employed four geometric transformations of GSV images: horizontal inversion and vertical inversion rotation of 90, 180, and 270 degrees.

For training the CNN for each species, we used 90% of the photos for the training set and 10% of the photos for the testing set. To have an adequate number of photos in the testing set for calculating performance metrics, we selected only species with at least 200 GSV images, for a total of 26 species.

Tree species classification metrics

The confusion matrix

The goal of the Machine Learning models used in our research is to classify trees according to their genus and species, so the problem falls under “multiclass classification”. The metrics that we employed are based on the concept of multiclass confusion matrix. The confusion matrix is a cross table that records the number of occurrences between detected classification and predicted classification. By custom, the columns represent the model prediction, while the rows show the actual classification. In binary confusion matrices (Figure 7a) The main diagonal reports the correct answers (true positives TP), the position above the main diagonal reports the number of false negatives (FN) finally the position below the diagonal reports the false positives (FP). Multiclass confusion matrices can be brought back to the binary case by performing a separate analysis for each row (C_i class) as shown in Figure 7b.

The multiclass metrics for unbalanced data are derived from the two main indices of confusion matrix: specificity and sensitivity:

		Predicted Class	
		Positive	Negative
Actual Class	Positive	TP	FN
	Negative	FP	TN

(a)

		Predicted Class			
		C_1	C_2	...	C_n
Actual Class	C_1		FP		
	C_2	FN	TP	...	FN
		
	C_n		FP		

(b)

Figure 7. (a) Binary classification problem confusion matrix. (b) Multi-class classification problem confusion matrix.

$$sensitivity_{C_i} = \frac{TP_{C_i}}{TP_{C_i} + \sum FN_{C_i}} \quad (12)$$

$$specificity_{C_i} = \frac{\sum TN_{C_i}}{\sum TN_{C_i} + \sum FP_{C_i}} \quad (13)$$

Sensitivity (also called recall) is the percentage of correct positive predictions (TP) out of the total number of items actually in class C_i and ranges from 0 (worst) to 1 (best). Specificity is the percentage of negative corrected predictions (TN) out of the total predictions misclassified by the model in the C_i class, again varies from 0 (worst) to 1 (best).

The evaluation metrics.

Balanced accuracy is a metric used in remote sensing applications to evaluate classification efficiency Gibson et al. (2020); Simoniello et al. (2022). It is calculated as the average of sensitivity and specificity, which can also be defined as the mean accuracy obtained over one of the two classes.

$$BA = \frac{sensitivity_{C_i} + specificity_{C_i}}{2} \quad (14)$$

Another metric applied to the evaluation of classification performance using remotely sensed data is the geometric mean (Gmean) between sensitivity and specificity (Silva et al., 2017). Gmean was proposed in Burez and Van den Poel (2009) by combining the prediction accuracies, i.e. sensitivity as accuracy in positive classifications and specificity as accuracy in negative classifications. Poor performance in identifying true positives will lead to a low G-mean value, even if negative examples are classified correctly by the model (Hido et al., 2009).

$$Gmean = \sqrt{sensitivity_{C_i} \cdot specificity_{C_i}} \quad (15)$$

The likelihood ratio (Bekkar et al., 2013; Daboor & Shokr, 2013) is divided into positive and negative. The positive likelihood ratio LR(+) represents the ratio of the probability of classifying a case correctly to the probability of classifying incorrectly.

Table 2. Thresholds for positive likelihood ratio interpretation.

L value	Model contribution	
	LR(+)	LR(-)
1	Negligible	Good
1-5	Poor	Fair
5-10	Fair	Poor
>10	Good	Negligible

$$\begin{aligned} LR(+) &= \frac{P(\text{positive}|\text{positive})}{P(\text{positive}|\text{negative})} \\ &= \frac{TP_{C_i}/(TP_{C_i} + \sum FN_{C_i})}{\sum FP_{C_i}/(\sum TN_{C_i} + \sum FP_{C_i})} \\ &= \frac{\text{sensitivity}_{C_i}}{1 - \text{specificity}_{C_i}} \end{aligned} \quad (16)$$

While the negative likelihood ratio LR(-) is the ratio of the probability of predicting an example as negative when it is actually positive to the probability of correctly classifying it as negative.

$$\begin{aligned} LR(-) &= \frac{P(\text{positive}|\text{positive})}{P(\text{positive}|\text{negative})} \\ &= \frac{TP_{C_i}/(TP_{C_i} + \sum FN_{C_i})}{\sum FP_{C_i}/(\sum TN_{C_i} + \sum FP_{C_i})} \\ &= \frac{\text{sensitivity}_{C_i}}{1 - \text{specificity}_{C_i}} \end{aligned} \quad (17)$$

A higher positive likelihood ratio detects better performance on positive classes instead a lower negative likelihood ratio better performance on negative classes. For example, a positive likelihood ratio of 50 means that the probability of correctly classifying a lime tree is 50 times greater than the probability of incorrectly classifying the species under consideration; a negative likelihood ratio of 0.01 means that the probability of classifying a plane tree as a lime tree is 100 times lower ($1/0.01 = 100$) than it is to classify the plane tree correctly.

To interpret the results of LR(+) and LR(-) we can refer to the limits proposed by Bekkar et al. (2013), shown in Table 2.

Results

Given the complexity of the methodology, the exposition of results will systematically follow the workflow previously illustrated and depicted in Figure 4, and will therefore be divided into the following subsections: data production, processing and validation.

Data production

Identification of tree morphological parameters Canopy segmentation in the urban perimeter of the city of Prato resulted in the identification of

329,026 canopies for an area covered by urban forest of 782.55 hectares (Figure 8). The result of the map overlay operation resulted in the identification of 14,057 “public urban trees” with an area of 91.38 hectares; the category “other urban green”, including private green spaces and abandoned areas with shrub/herbaceous species covers an area of 691.17 hectares. Comparison with data from official national statistics leads to a greater extent of green areas by almost 25%; this figure is plausible because the procedure adopted also leads to segmenting areas that are not officially classified as urban green, mainly abandoned urban areas under natural succession (so-called brownfields) or small residual agricultural areas totally included in the urban perimeter. According to many authors (Mathey et al., 2015; Pueffel et al., 2018; Sikorski et al., 2021) even these unofficially recognized green areas contribute to the provision of ecosystem services and it would be correct for them to be officially surveyed.

The harvesting of canopy images from Google Street View GSV Within a 30-meter radius of GSV, 12659 urban public trees were found to be visible out of a total of 14,057 trees taller than 3 meters surveyed by the City of Prato, accounting for about 90%. Trees not visible can be attributed to two main reasons: plants covered by other plants in the perspective of GSV’s camera point or plants within urban parks or other areas not accessible to Google machines. In the latter case, however, trees at the edges were often visible. Therefore, 12659 photos were downloaded from GSV taken during the period from June 2021 to September 2020.

Based on classification through Resnet18 we filtered 11,552 shot points (91%) with green index above 50%. For these images, photos with FOV = 20 were also downloaded. An example of the downloaded images for the most important species is shown in Figure 9.

New data production

Classification procedures using Plant.Id, Pl@ntNet, and GoogLeNet CNN were applied to classify canopy images downloaded from GSV. The result was a new geographic database containing the vector geometry of the canopy and the following features: coordinates of the centre of gravity of the canopy, height of the tree, code of the GSV images collected, results of the classification in genus and species of the ground survey of the census, results of the classification in genus and species of the classification by Plant.Id, Pl@ntNet and GoogLeNet CNN. Figure 10 shows a sample of the result of the three classifications compared with ground observation.

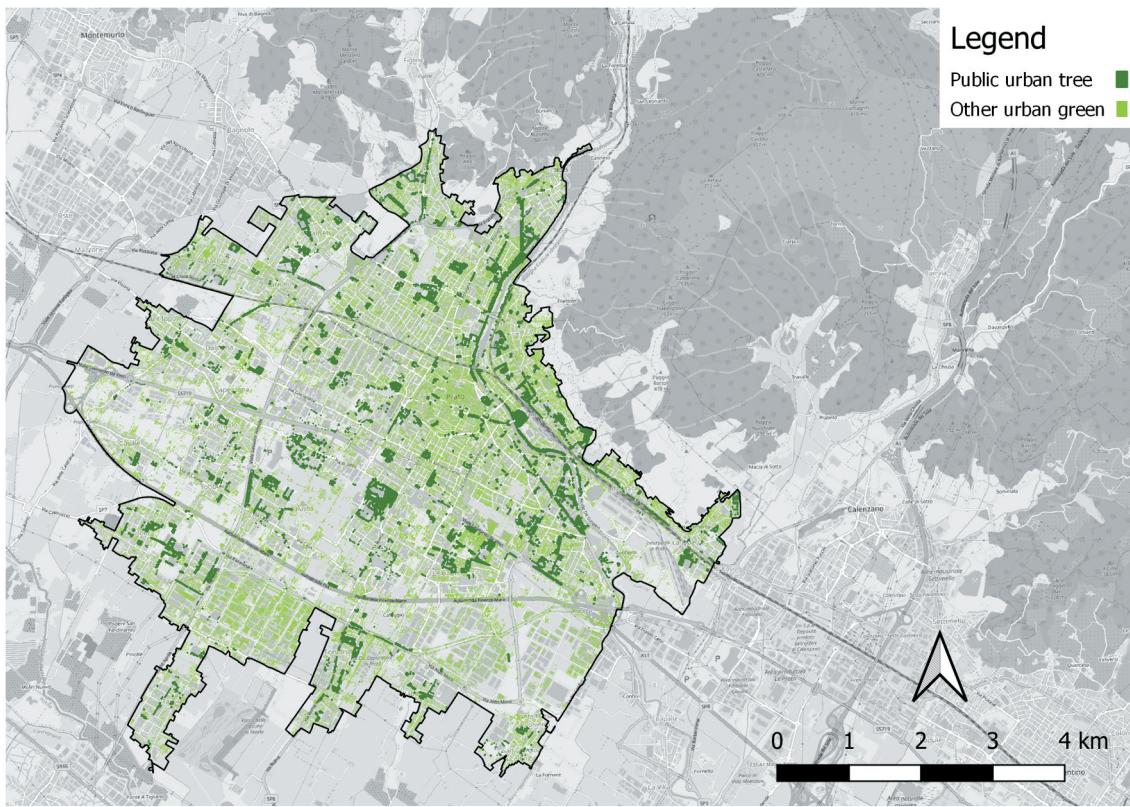


Figure 8. Urban greenery and public urban trees in the perimeter of the city of Prato.

Validation

In accordance with the literature, we evaluated the performance of the two applications Pl@ntNet and Plant.Id and the CNN GoogLeNet classifier at the genus and species level. Tables 3–8 report performance metrics for the most frequent species and genera as well as descriptive statistics for all species and genera in the study area.

The Pl@ntNet classification (Table 3) at the species level has a barely acceptable median balanced accuracy (0.60), with a rather narrow frequency distribution (first and third quartiles at 0.64 and 0.55, respectively). Relative to the three most frequent species in the study area *Pinus pinea* and *Tilia europea* have slightly better performances (0.75 and 0.62, respectively), while *Platanus hybrida* is below the median (0.54). Among species with at least 50 trees in the study area, the best ranking performances are for *Pinus* and the worst for *Robinia*. The geometric mean shows worse performance and greater dispersion of descriptive statistics, (median 0.48, first quartile 0.33 and third quartile 0.57). Classification of the top three species has acceptable results for *Pinus pinea* (0.71), but *Tilia europaea* and *Platanus Hybrida* have poor performance (0.58 and 0.33, respectively). The best performance is for *Pinus pinea* and the worst for *Robinia pseudoacacia* (just 0.13). LR(+) shows good performance in identifying true positive species: the first quartile is above the threshold of 5 and the median is above 10. For the

top three species the correct classifications are also good for this metric only *Pinus pinea*, but barely acceptable for *Platanus hybrida* and poor for *Tilia europaea*. For other species with more than 50 plants there are poor performances ($LR(+)$ < 5) for some important species in the study area: *Acer campestre*, *Acer platanoides*, *Fraxinus excelsior* and *Populus nigra*. $LR(-)$, on the other hand, has values always below the limit of 1 with thus low probability of false positives.

The performance of Pl@ntNet at the genus level is slightly better than the classification at the species level (Table 4). At the genus level, balanced accuracy has an acceptable average of 0.62, with first and third quartiles at 0.69 and 0.56, respectively. Relative to the three most frequent genera in the study area, *Pinus* and *Platanus* have significantly higher performances (0.81 and 0.78 respectively), while *Tilia* is only slightly above the median value (0.68). Among the genera with at least 50 occurrences, the best rankings are for *Pinus*, and the worst for *Robinia*. The geometric mean shows a median of 0.48, with first quartile 0.35 and third quartile 0.63. However, the classification of *Pinus*, *Tilia* and *Platanus* has results above the median (0.80, 0.77 and 0.63, respectively). The best result is for *Cedrus* and the worst for *Robinia* (just 0.13). $LR(+)$ already has the first quartile above the threshold of 10 (10.2). For the three main species, the correct classification performances are good for *Pinus* and *Platanus*, but only acceptable for *Tilia*. For the set of genera with

	FOV=crown	FOV=20
Pinus Pinea		
Tilia x Europaea		
Platanus x Orientalis		

Figure 9. An example of the downloaded images for the most important species (FOV= Field of View).

more than 50 plants, there are poor performances ($LR(+)$ < 5) for *Quercus* and *Acer*. Even at the genus level $LR(-)$, on the other hand, has values always below the limit of 1 with thus low probability of false positives.

As shown in Table 5, at the level of balanced accuracy, the performances of Plant.Id classification by species are quite good, with a median value of 0.73, first quartile 0.67 and third quartile 0.8. The top three species have definitely good results: *Pinus pinea* 0.87, *Tilia aeuropea* 0.87, *Platanus hybrida* 0.89. The species with the lowest value is *Ligustrum lucidum* with 0.59. Again the geometric mean has lower values: median 0.69 first and third quartiles 0.58 and 0.78. The values of the three most important species remain high: *Pinus pinea* 0.87, *Tilia europea* 0.87 and *Platanus hybrida* 0.89. $LR(+)$ is very good in descriptive statistics and in the most representative species with the only exception of *Tilia europea* which has a value of 8.92, still acceptable. $LR(-)$ is always less than 1.

The results at the genus level are also slightly better than the classification by species (Table 6). The balanced accuracy at median 0.75, first and third quartiles of at 0.70 and 0.82; the geometric mean has slightly lower values: median 0.70, first and third

quartiles 0.64 and 0.80. Good performances are maintained for the most important genera with $LR(+)$ and $LR(-)$ values all in the “good” limit. *Pinus*, *Tilia* and *Platanus* have balanced accuracy and geometric mean values all above 0.9.

GoogLeNet’s classification performance is comparable to that of Plant.Id and higher than that of Pl@ntNet (see Tables 7 and 8). The median balanced accuracy value of 0.71 is slightly lower than that of Plant.Id, a trend confirmed by the geometric mean (0.65 for GoogLeNet and 0.69 for Plant.Id). The top three species are ranked efficiently: *Pinus Pinea* 0.93 for balanced accuracy and 0.93 for geometric mean, *Tilia aeuropea* 0.847 and 0.846 respectively, and *Platanus hybrida* 0.82 and 0.81.

Examining the frequency distribution statistics, GoogLeNet has a wider distribution especially for geometric mean. The first and third quartiles of the balanced accuracy are 0.52 and 0.85, respectively, compared with 0.67 and 0.80 for Plant.Id; for the geometric mean we even have 0.22 and 0.84 compared with 0.57 and 0.78. This phenomenon can be explained by the fact that for less frequent species GoogLeNet does not have many images to process in

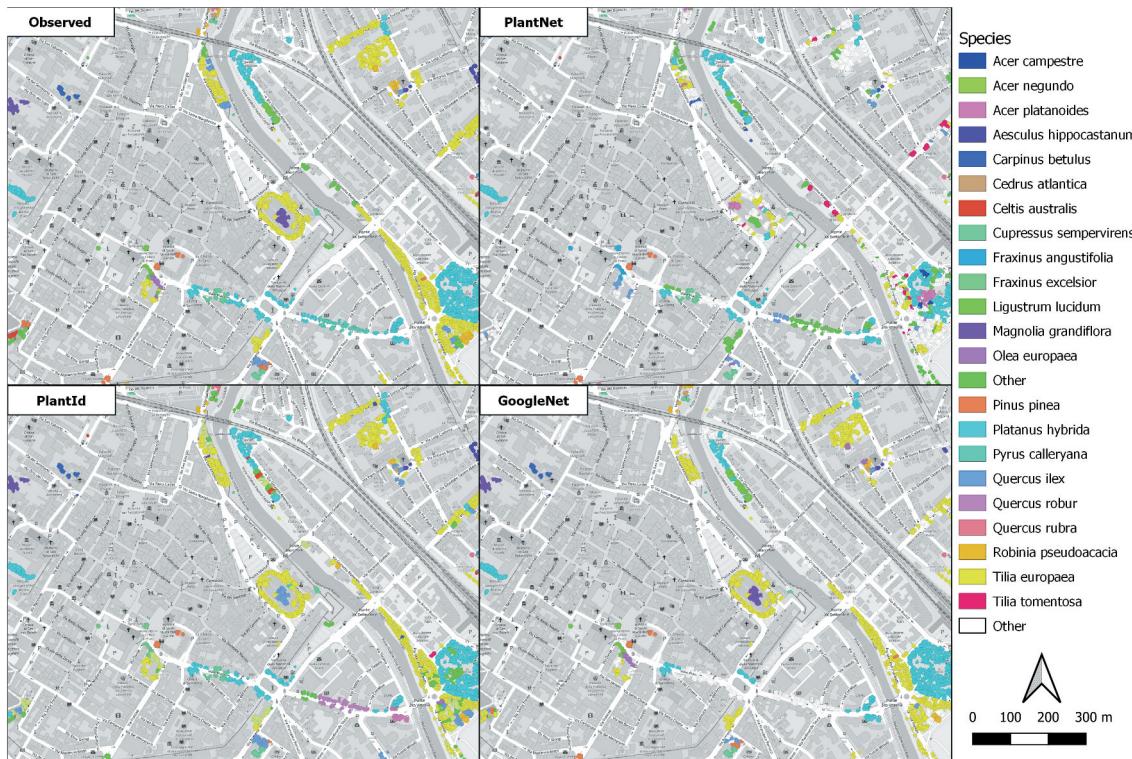


Figure 10. Sample of the result of the three classifications compared with ground observation.

Table 3. Performance indicators of the Pl@ntNet app by species.

	Species	Balanced accuracy	LR+	LR-	Geometric mean
Species with more than 50 plants	Acer campestre	0.543	2.511	0.060	0.367
	Acer negundo	0.524	15.287	0.003	0.224
	Acer platanoides	0.500	1.022	0.026	0.160
	Aesculus hippocastanum	0.569	32.446	0.004	0.377
	Carpinus betulus	0.608	8.628	0.029	0.487
	Cedrus atlantica	0.739	66.465	0.007	0.694
	Celtis australis	0.598	27.381	0.007	0.449
	Cupressus sempervirens	0.671	19.724	0.019	0.594
	Fraxinus angustifolia	0.553	10.782	0.011	0.341
	Fraxinus excelsior	0.611	3.170	0.114	0.540
	Ligustrum lucidum	0.533	40.378	0.002	0.258
	Magnolia grandiflora	0.641	255.283	0.001	0.532
	Olea europaea	0.561	32.179	0.004	0.353
	Pinus pinea	0.749	18.079	0.030	0.715
	Platanus hybrida	0.546	5.082	0.023	0.333
	Populus nigra	0.612	2.055	0.268	0.585
	Prunus cerasifera	0.623	75.708	0.003	0.499
	Pyrus calleryana	0.532	117.380	0.001	0.255
	Quercus ilex	0.663	7.011	0.057	0.599
	Quercus robur	0.613	4.610	0.067	0.521
	Quercus rubra	0.583	18.684	0.009	0.417
	Robinia pseudoacacia	0.509	-	0.000	0.134
	Tilia europaea	0.624	2.851	0.155	0.576
All species	Minimum	0.500	1.022	0.000	0.134
	First quartile	0.546	6.895	0.710	0.332
	Median	0.609	19.316	0.778	0.476
	Third quartile	0.641	90.900	0.906	0.576
	Maximum	1.000	255.000	0.999	1.000

the training set, so it gets poor results, while for species with more images it gets satisfactory results. This hypothesis is confirmed by the LR(+) values, which are definitely unsatisfactory for Acer campestre, Acer saccharinum, Fraxinus angustifolia, Prunus cerasifera, and Quercus robur and only fair for Aesculus hippocastanum, Quercus rubra, and Tilia europaea. LR(-) is

always less than 1 thus satisfactory. By grouping species by genus GoogLeNet succeeds in making a better classification because even less frequent plants refer to a larger training set. In fact at the genus level GoogLeNet gets better results than Plant.Id. The median balanced accuracy is 0.78 with first and third quartiles at 0.60 and 0.87, compared to 0.75, 0.70 and

Table 4. Performance indicators of the Pl@ntNet app by genus.

	Genus	Balanced accuracy	LR+	LR-	Geometric mean
Genera with more than 50 plants	Acer	0.564	2.792	0.863	0.429
	Aesculus	0.569	42.433	0.863	0.374
	Carpinus	0.613	12.867	0.771	0.489
	Cedrus	0.870	54.106	0.250	0.862
	Celtis	0.599	33.771	0.801	0.450
	Cupressus	0.678	24.630	0.639	0.604
	Fraxinus	0.637	5.267	0.708	0.562
	Ligustrum	0.533	60.444	0.934	0.258
	Magnolia	0.641	383.066	0.718	0.532
	Olea	0.561	48.304	0.877	0.353
	Pinus	0.815	11.816	0.331	0.805
	Platanus	0.787	10.424	0.389	0.772
	Populus	0.689	3.266	0.546	0.674
	Prunus	0.570	16.889	0.859	0.383
	Pyrus	0.530	33.089	0.941	0.246
	Quercus	0.650	3.878	0.665	0.602
	Robinia	0.509	-	0.982	0.134
	Tilia	0.683	6.058	0.606	0.637
All species	Minimum	0.509	2.792	0.000	0.134
	First quartile	0.562	10.242	0.605	0.358
	Median	0.612	34.424	0.774	0.480
	Third quartile	0.688	57.602	0.874	0.635
	Maximum	1.000	1360.000	0.982	1.000

Table 5. Performance indicators of the Plant.Id app by species.

	Species	Balanced accuracy	LR+	LR-	Geometric mean
Species with more than 50 plants	Acer campestre	0.616	23.011	0.765	0.490
	Acer negundo	0.800	65.237	0.395	0.777
	Acer platanoides	0.680	47.895	0.636	0.605
	Acer saccharinum	0.615	46.599	0.769	0.484
	Aesculus hippocastanum	0.840	141.050	0.317	0.825
	Carpinus betulus	0.781	48.307	0.430	0.754
	Cedrus atlantica	0.764	81.485	0.468	0.729
	Celtis australis	0.777	42.417	0.439	0.748
	Cupressus sempervirens	0.825	69.482	0.344	0.808
	Fraxinus angustifolia	0.640	21.615	0.716	0.539
	Fraxinus excelsior	0.734	15.743	0.516	0.696
	Ligustrum lucidum	0.588	122.667	0.823	0.421
	Magnolia grandiflora	0.810	320.427	0.378	0.788
	Olea europaea	0.743	65.154	0.510	0.700
	Pinus pinea	0.872	32.460	0.238	0.866
	Platanus hybrida	0.894	30.084	0.189	0.891
	Populus nigra	0.704	16.475	0.581	0.651
	Prunus cerasifera	0.687	221.786	0.626	0.612
	Pyrus calleryana	0.717	223.196	0.566	0.659
	Quercus ilex	0.875	30.953	0.231	0.869
	Quercus robur	0.793	29.936	0.403	0.770
	Quercus rubra	0.738	139.978	0.522	0.691
	Robinia pseudoacacia	0.728	61.720	0.540	0.679
	Tilia europaea	0.874	8.920	0.174	0.874
All species	Minimum	0.523	6.007	0.000	0.213
	First quartile	0.666	34.227	0.397	0.577
	Median	0.734	84.602	0.522	0.691
	Third quartile	0.800	256.479	0.668	0.776
	Maximum	0.894	320.427	0.823	0.891

0.82 of Plant.Id, respectively. The geometric mean also shows similar performances. LR(+) values improve markedly from the species-level classification and are always above 10 with the exception of *Fraxinus* and *Tilia* often confused with each other in classification.

Discussion

The objective of our study was to identify a framework to automatically create the urban forest inventory at the individual tree level through the integrating LiDAR data and publicly open GSV 360 images by applying online

automatic species classification apps. Our work demonstrated that it is possible to extract images of urban tree canopies from GSV by segmenting LiDAR images filtered through NDVI computed using multispectral data from high-resolution remote sensing.

GSV images were classified by two artificial intelligence applications (Pl@ntNet and Plant.Id) queried through an API and a classifier trained through the GoogLeNet CNN. Despite the low resolution of GSV images, Plant.Id and GoogLeNet achieved satisfactory classification efficiency especially for the most frequent plants in the study area. The good performance

Table 6. Performance indicators of the Plant.Id app by genus.

	Genus	Balanced accuracy	LR+	LR-	Geometric mean
Genera with more than 50 plants	acer	0.795	16.939	0.387	0.777
	aesculus	0.843	171.543	0.312	0.829
	carpinus	0.782	55.404	0.430	0.754
	cedrus	0.829	97.344	0.338	0.812
	celtis	0.777	48.883	0.439	0.748
	cupressus	0.820	78.631	0.355	0.802
	fraxinus	0.817	31.563	0.352	0.801
	ligustrum	0.588	120.533	0.823	0.421
	magnolia	0.810	367.514	0.378	0.788
	olea	0.744	74.802	0.509	0.701
	pinus	0.926	29.264	0.122	0.925
	platanus	0.899	35.460	0.183	0.896
	populus	0.786	31.246	0.417	0.761
	prunus	0.630	15.742	0.735	0.522
	pyrus	0.730	270.876	0.540	0.678
	quercus	0.845	18.424	0.282	0.837
	robinia	0.729	70.781	0.539	0.679
	tilia	0.905	20.528	0.156	0.903
All species	Minimum	0.523	15.742	0.000	0.213
	First quartile	0.705	38.581	0.352	0.641
	Median	0.750	123.589	0.500	0.707
	Third quartile	0.820	380.000	0.589	0.802
	Maximum	1.000	2373.500	0.955	1.000

Table 7. Performance indicators of the GoogleNet network by species.

	Genus	Balanced accuracy	LR+	LR-	Geometric mean
Species with more than 200 plants	Acer campestre	0.499	0.000	1.002	0.000
	Acer negundo	0.747	23.069	0.495	0.710
	Acer platanoides	0.599	73.800	0.802	0.447
	Acer saccharinum	0.500	0.000	1.001	0.000
	Aesculus hippocastanum	0.520	8.415	0.960	0.213
	Carpinus betulus	0.982	78.670	0.023	0.982
	Cedrus atlantica	0.932	85.225	0.126	0.931
	Celtis australis	0.757	41.459	0.479	0.722
	Cupressus sempervirens	0.916	44.791	0.152	0.914
	Fraxinus angustifolia	0.499	0.000	1.001	0.000
	Fraxinus excelsior	0.673	57.686	0.651	0.592
	Ligustrum lucidum	0.529	11.570	0.943	0.249
	Magnolia grandiflora	0.569	35.190	0.861	0.377
	Olea europaea	0.661	30.708	0.674	0.574
	Pinus pinea	0.934	13.756	0.068	0.934
	Platanus hybrida	0.818	13.659	0.330	0.807
	Populus nigra	0.789	122.250	0.419	0.762
	Prunus cerasifera	0.500	0.000	1.001	0.000
	Pyrus calleryana	0.847	113.167	0.302	0.834
	Quercus ilex	0.807	34.688	0.375	0.787
	Quercus robur	0.499	0.000	1.003	0.000
	Quercus rubra	0.607	6.882	0.778	0.491
	Robinia pseudoacacia	0.889	67.848	0.213	0.883
	Tilia europaea	0.847	5.306	0.173	0.847
All species	Minimum	0.499	0.000	0.023	0.000
	First quartile	0.522	5.700	0.235	0.222
	Median	0.710	26.889	0.573	0.651
	Third quartile	0.847	65.308	0.955	0.844
	Maximum	0.982	122.250	1.003	0.982

of Plant.Id probably stems from the ability to also use geographic location as input data in the query. GoogLeNet CNN, on the other hand, had the advantage of training the AI with images taken from GSV while the global apps are trained with images from very diverse sources, mainly provided by users through smartphones. Pl@ntNet, on the other hand, yielded less satisfactory results, partly because this AI performs best when using images of plant organs (leaves, flowers, and bark) of the plant as input data, whereas GSV images were only crowns.

The differences from other studies using GSV 360 imagery our model does not require prior sampling to train a deep learning classifier as one of the two apps tested, Plant.Id, demonstrated similar performance to a Deep Learning classifier trained with species in the study area. With the Plant.Id app it is possible to classify even species with very few plants in the study area (at the limit even species with only one plant), while in training a CNN classifier it is possible to identify only species with a sufficient number of observations. These advantages further reduce costs and

Table 8. Performance indicators of the GoogleNet network by genus.

	Genus	Balanced accuracy	LR+	LR-	Geometric mean
Genera with more than 50 plants	Acer	0.578	10.580	0.483	0.717
	Aesculus	0.500	201.409	0.864	0.369
	Carpinus	0.890	68.063	0.115	0.935
	Cedrus	0.930	85.517	0.126	0.931
	Celtis	0.720	18.262	0.404	0.768
	Cupressus	0.886	30.660	0.197	0.887
	Fraxinus	0.565	6.959	0.858	0.398
	Ligustrum	0.500	11.609	0.943	0.249
	Magnolia	0.500	63.471	0.788	0.462
	Olea	0.693	32.412	0.591	0.641
	Pinus	0.928	17.233	0.055	0.946
	Platanus	0.810	14.167	0.413	0.761
	Populus	0.582	115.000	0.377	0.788
	Prunus	0.500	-	1.000	0.000
	Pyrus	0.759	44.220	0.254	0.859
	Quercus	0.579	10.125	0.616	0.626
All species	Robinia	0.524	60.135	0.266	0.853
	Tilia	0.837	6.679	0.203	0.849
	Minimum	0.500	6.679	0.055	0.000
	First quartile	0.596	11.095	0.201	0.446
	Median	0.785	30.660	0.408	0.764
	Third quartile	0.873	65.767	0.806	0.866
	Maximum	0.946	201.409	1.000	0.946

make it possible to transfer the method to even small cities that would not allow for a large enough training set to train the classifier.

The results are consistent with those reported in the literature (Ringland et al., 2021, Lumnitz et al., 2021; Jakuschona et al., 2022). Plant.Id and GoogLeNet demonstrated good performance in detecting and classifying major urban street trees and tree species from GSV 360 images using DL-based techniques, showing Balanced accuracy at the genus level of 0.75 and 0.78, respectively, and at the species level of 0.73 and 0.71.

This performance is lower than that of Zarrin (2019), who reported a tree classification performance of 0.96 for all trees at the genus level. The authors collect the images in situ via smartphone camera, thus with optimal image quality compared to those detectable by GSV.

Comparing our results with in work that used GSV images, in Berland and Lange's (2017) research, genus identification was in agreement between field and virtual surveys for 90% of trees ($\kappa = 0.88$, $p < 0.001$), and at the species level, the level of agreement for tree identification was 66% of trees ($\kappa = 0.64$, $p < 0.001$). GSV Branson et al. (2018), achieved slightly better classification performances at the species level, with an average class precision of 0.83 for 30 different species; finally, recently Choi et al. (2022) performed worse, with a mean precision accuracy of only 0.54.

In agreement with Choi et al. (2022) we believe that the performance of CNN-based species classification systems is greatly influenced by the morphological and phenological characteristics of each tree species. According to our results, tree species with distinct morphological characteristics showed better classification accuracy. *Pinus pinea*

showed the highest classification accuracy among the main species in our study because of the characteristic umbrella canopy clearly different from other species.

For the evaluation of app performance, we used a set of metrics that provided comprehensive information on strengths and weaknesses. In particular, balanced accuracy and geometric mean allowed us to evaluate the efficiency in classifying the most frequent species, LR(+) the probability of correct positive classification. Finally, the low values of LR(-) showed that neither Pl@ntNet nor Plant.Id are prone to systematic misclassification for trees in the study area.

Contrary to the findings of Zarrin et al. (2019), work has also shown that Plant.Id performance does not vary significantly by classifying trees at the genus rather than species level .

Related to the research question, "Does the integration of multispectral data, LiDAR, GSV imagery and AI for species recognition enable automatic censuses of urban greenery at the individual tree level?" the answer is only partially positive.

Google machines do not penetrate inside parks and pedestrian paths, so trees inside cannot be photographed and identified. It may also be the case that even for street trees a tree is masked by another object than where the GSV photo is taken. Finally, with GSV images it is not possible to survey private urban greenery. This aspect represents the main limitation of the present work.

Relative to errors in the acquisition of canopy images the main causes of failure of the methodology were:

- (1) the GSV images were not all taken on the same date, and it may be the case that they were taken

- at very different times, and thus are not temporally aligned with the remotely sensed data.
- (2) GSV images were taken in seasons when leaves were not present.

The solution to this limitation is to integrate unsatisfactory GSV images using spherical photos taken on foot or with small electric vehicles that can be used in urban parks via backpack-mounted cameras. A planned development of the research will be to apply our methodology with this modality by assessing whether there are significant differences in performance between professional or consumer cameras. A further development involves integrating spherical photos with an aerial survey using unmanned aerial vehicles (UAV).

Another limitation of the work is that the canopy segmentation procedure was not subjected to statistical validation, since the purpose of the work was to evaluate the efficiency of classification using Pl@ntNet and Plant.Id, and thus the canopy segmentation was only functional for GSV image acquisition. From non-systematic recognition, however, the quality of segmentation appeared good at different scales, as shown in the example in Figure 11.

However, it will be necessary in a future work to investigate the most efficient methodologies for segmenting canopies in urban areas efficiently for both automatic identification and map rendering.

Finally, the economic costs of automated urban forest inventories by integrating GSV images, spherical images taken from the ground and UAVs, and LiDAR data classified with Plant.Id should be carefully evaluated. Currently, the base price of Plant.Id is 0.05 euros per request, but significant discounts are expected for larger volumes of identifications. It will be necessary to conduct cost simulations between our methodology and other methods of conducting an urban forest inventory at the individual tree level.

Conclusion

Our work demonstrated that it is possible to combine the Plant.Id application with photos downloaded from GSV and LiDAR and multispectral data to make single-tree public green inventories.

Through a crown segmentation procedure based on the capture area method, we were able to calculate the vertical projection parameters of tree crowns and extract crown images from spherical GSV photos. The Plant.Id application correctly identified plants in the city of Prato with a median accuracy of 0.73 and with better performance for the most common plants: *Pinus pinea* 0.87, *Tilia aeuropea* 0.87, *Platanus hybrida* 0.89.

As emerged in the discussion section there are many possible research developments. The most promising in our opinion are the following.

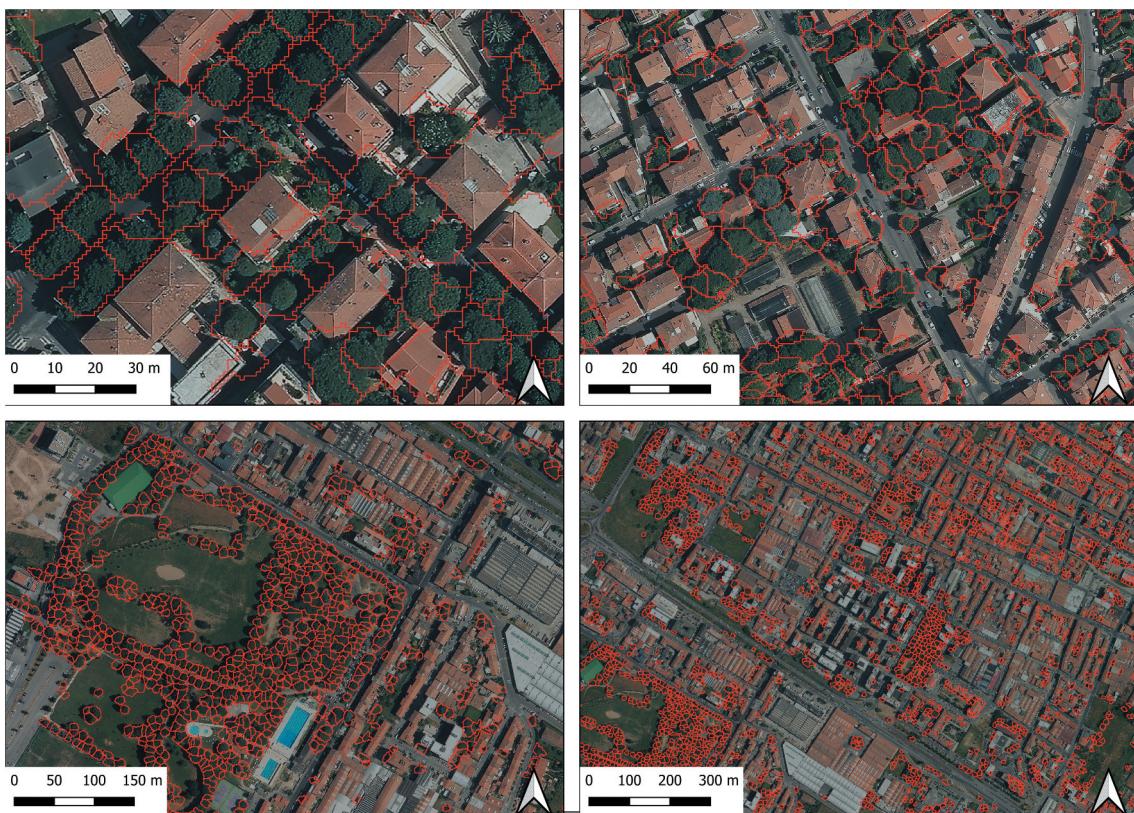


Figure 11. Sample of crown segmentation at different scale.

- The use of higher detail azimuth spherical images obtained from UAVs.
- The use of more advanced machine-learning-based crown segmentation procedures applied to aerial zenith images, again from UAVs.
- The evaluation of the efficiency of the new Plant.Id feature dedicated to plant disease classification for health monitoring of urban greenery.

GSVGSVWe believe that our procedure can be useful for city administrators to update urban green censuses in order to set the correct maintenance and management actions.

We hope that this work will contribute to the dissemination of single-tree urban green inventories even in small cities.

Acknowledgements

We would like to thank the Pl@ntNet and Plant.Id teams for opening their software platforms essential for us to carry out this research.

Disclosure statement

No potential conflict of interest was reported by the authors.

Data availability statement

Data supporting the results of this study are openly available on the Internet (<http://odn.comune.prato.it/dataset/alberi-prato>; <https://www502.regione.toscana.it/geoscopio/carta-teca.html>).

References

- Adkins, R.V. -C., Michael, R. K., Blahna, D. J., & Blood, M. W. (1997). Urban forest resource management at hill air force base, Ogden, Utah. *Journal of Arboriculture*, 23(4), 136. <https://doi.org/10.48044/jauf.1997.021>
- Alonso, M., Bookhagen, B., & Roberts, D. A. (2014). Urban tree species mapping using hyperspectral and lidar data fusion. *Remote sensing of environment*, 148, 70–83. <https://doi.org/10.1016/j.rse.2014.03.018>
- Alvarez, F., Catanzarite, T., Castellanos, J., & Blanco-Medina, V. (2010). Biomass estimation using digital photogrammetric cameras. In Presented at the International Calibration and Orientation Workshop EuroCOW, Castelldefels, Spain, (pp. 10–12).
- Barbiero, E., Bernetti, I., Capecchi, I., & Saragosa, C. (2020). Integrating remote sensing and street view images to quantify urban forest ecosystem services. *Remote Sensing*, 12(2), 329. <https://doi.org/10.3390/rs12020329>
- Bekkar, M., Djemaa, H. K., & Alitouche, T. A. (2013). Evaluation measures for models assessment over imbalanced data sets. *Journal of Information Engineering and Applications*, 3(4), 15–33. <https://doi.org/10.5121/ijdkp.2013.3402>
- Berland, A., & Lange, D. A. (2017). Google street view shows promise for virtual street tree surveys. *Urban Forestry & Urban Greening*, 21, 11–15. <https://doi.org/10.1016/j.ufug.2016.11.006>
- Bonnet, P., Goëau, H., Hang, S. T., Lasseck, M., Šulc, M., Malécot, V., & Joly, A. (2018). Plant identification: experts vs. machines in the era of deep learning. In *Multimedia tools and applications for environmental & biodiversity informatics* (pp. 131–149). Cham: Springer.
- Brandtberg, T., Warner, T. A., Landenberger, R. E., & McGraw, J. B. (2003). Detection and analysis of individual leaf-off tree crowns in small footprint, high sampling density lidar data from the eastern deciduous forest in North America. *Remote Sensing of Environment*, 85(3), 290–303. [https://doi.org/10.1016/S0034-4257\(03\)00008-7](https://doi.org/10.1016/S0034-4257(03)00008-7)
- Branson, S., Wegner, J. D., Hall, D., Lang, N., Schindler, K., & Perona, P. (2018). From Google Maps to a fine-grained catalog of street trees. *Isprs Journal of Photogrammetry and Remote Sensing*, 135, 13–30. <https://doi.org/10.1016/j.isprsjprs.2017.11.008>
- Brostow, G. J., Fauqueur, J., & Cipolla, R. (2009). Semantic object classes in video: A high-definition ground truth database. *Pattern recognition letters*, 30(2), 88–97.
- Burez, J., & Van den Poel, D. (2009). Handling class imbalance in customer churn prediction. *Expert Systems with Applications*, 36(3), 4626–4636. <https://doi.org/10.1016/j.eswa.2008.05.027>
- Cappellato, L., Ferro, N., Jones, G. J., Kamps, J., Mothe, J., Pinel-Sauvagnat, K., & Savoy, J. (2016, January). Report on CLEF 2015: Experimental IR Meets Multilinguality, Multimodality, and Interaction. In *ACM SIGIR Forum* (Vol. 49, No. (2), pp. 47–56). New York, NY, USA: ACM.
- Chen, Q., Baldocchi, D., Gong, P., & Kelly, M. (2006). Isolating individual trees in a savanna woodland using small footprint lidar data. *Photogrammetric Engineering & Remote Sensing*, 72(8), 923–932. <https://doi.org/10.14358/PERS.72.8.923>
- Choi, K., Lim, W., Chang, B., Jeong, J., Kim, I., Park, C. R., & Ko, D. W. (2022). An automatic approach for tree species detection and profile estimation of urban street trees using deep learning and Google street view images. *Isprs Journal of Photogrammetry and Remote Sensing*, 190, 165–180. <https://doi.org/10.1016/j.isprsjprs.2022.06.004>
- Clark, M. L., Roberts, D. A., & Clark, D. B. (2005). Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sensing of Environment*, 96(3–4), 375–398. <https://doi.org/10.1016/j.rse.2005.03.009>
- Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J. (2015). System for automated geoscientific analyses (SAGA) v. 2.1. 4. *Geoscientific Model Development*, 8(7), 1991–2007. <https://doi.org/10.5194/gmd-8-1991-2015>
- Cook, E. A., & Iverson, L. R. (1991). Inventory and change detection of urban land cover in illinois using landsat thematic mapper data. In *Technical papers ACSM-ASPRS annual convention, Balti- more, 1991. Vol. 3: Remote sensing* (pp. 83–92).
- Dabboor, M., & Shokr, M. (2013). A new likelihood ratio for supervised classification of fully polarimetric SAR data: An application for sea ice type mapping. *Isprs Journal of Photogrammetry and Remote Sensing*, 84, 1–11. <https://doi.org/10.1016/j.isprsjprs.2013.06.010>
- Falkowski, M. J., Smith, A. M., Hudak, A. T., Gessler, P. E., Vierling, L. A., & Crookston, N. L. (2006). Automated estimation of individual conifer tree height and crown diameter via two-dimensional spatial wavelet analysis of lidar data. *Canadian Journal of Remote Sensing*, 32(2), 153–161. <https://doi.org/10.5589/m06-005>

- Gibson, R., Danaher, T., Hehir, W., & Collins, L. (2020). A remote sensing approach to mapping fire severity in south-eastern Australia using sentinel 2 and random forest. *Remote sensing of environment*, 240, 111702. <https://doi.org/10.1016/j.rse.2020.111702>
- Goëau, H., Bonnet, P., & Joly, A. (2021, September). Overview of PlantCLEF 2021: Cross-domain plant identification. In *Working Notes of CLEF 2021-Conference and Labs of the Evaluation Forum*. (Vol. 2936), 1422–1436, Sierre, Switzerland
- Goëau, H., Bonnet, P., & Joly, A. (2021, September). Overview of PlantCLEF 2021: Cross-domain plant identification. In *Working Notes of CLEF 2021-Conference and Labs of the Evaluation Forum* (Vol. 2936, pp. 1422–1436).
- Heikkinen, V., Korpela, I., Tokola, T., Honkavaara, E., & Parkkinen, J. (2011). An SVM classification of tree species radiometric signatures based on the Leica ADS40 sensor. *IEEE Transactions on Geoscience and Remote Sensing*, 49 (11), 4539–4551.
- Heinzel, J., & Koch, B. (2011). Exploring full-waveform LiDAR parameters for tree species classification. *International Journal of Applied Earth Observation and Geoinformation*, 13(1), 152–160. <https://doi.org/10.1016/j.jag.2010.09.010>
- Hido, S., Kashima, H., & Takahashi, Y. (2009). Roughly balanced bagging for imbalanced data. Statistical analysis and data mining. *The ASA Data Science Journal*, 2(5□6), 412–426. <https://doi.org/10.1002/sam.10061>
- Hyppä, J. U. H. A., Hyppä, H., Yu, X., Kaartinen, H. A. R. R. I., Kukko, A. N. T. E. R. O., & Holopainen, M. (2017). Forest inventory using small-footprint airborne lidar. In *Topographic laser ranging and scanning* (pp. 335–370). CRC Press.
- Hyppä, J., Kelle, O., Lehtikoinen, M., & Inkinen, M. (2001). A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. *IEEE Transactions on Geoscience and Remote Sensing*, 39(5), 969–975, Taylor & Francis Group. <https://doi.org/10.1109/36.921414>
- Jakuschona, N., Niers, T., Stenkamp, J., Bartoschek, T., Schade, S., & Cardoso, A. C. (2022). Evaluating image-based species recognition models suitable for citizen science application to support european invasive alien species policy (No. JRC128240. *Joint Research Centre (Seville Site)*.
- Jansen, L. J., Carrai, G., Morandini, L., Cerutti, P. O., & Spisni, A. (2006). Analysis of the spatio-temporal and semantic aspects of land-cover/use change dynamics 1991–2001 in Albania at national and district levels. *Environmental Monitoring and Assessment*, 119(1), 107–136.
- Jing, L., Hu, B., Noland, T., & Li, J. (2012). An individual tree crown delineation method based on multi-scale segmentation of imagery. *Isprs Journal of Photogrammetry and Remote Sensing*, 70, 88–98. <https://doi.org/10.1016/j.isprsjprs.2012.04.003>
- Jones, H. G. (2020). What plant is that? Tests of automated image recognition apps for plant identification on plants from the British flora. *AoB Plants*, 12(6), laa052. <https://doi.org/10.1093/aobpla/plaa052>
- Keller, J. K. K., & Konijnendijk, C. C. (2012). A comparative analysis of municipal urban tree inventories of selected major cities in North America and Europe. *Arboriculture & Urban Forestry*, 38(1), 24–30. <https://doi.org/10.48044/jauf.2012.005>
- Korpela, I., Ørka, H. O., Maltamo, M., Tokola, T., & Hyppä, J. (2010). Tree species classification using airborne LiDar—effects of stand and tree parameters, down-sizing of training set, intensity normalization, and sensor type. *Silva Fennica*, 44(2), 319–339. <https://doi.org/10.14214/sf.156>
- Leckie, D. G., Tinis, S., Nelson, T., Burnett, C., Gougeon, F. A., Cloney, E., & Paradine, D. (2005). Issues in species classification of trees in old growth conifer stands. *Canadian Journal of Remote Sensing*, 31 (2), 175–190. <https://doi.org/10.5589/m05-004>
- Lumnitz, S., Devisscher, T., Mayaud, J. R., Radic, V., Coops, N. C., & Griess, V. C. (2021). Mapping trees along urban street networks with deep learning and street-level imagery. *Isprs Journal of Photogrammetry and Remote Sensing*, 175, 144–157. <https://doi.org/10.1016/j.isprsjprs.2021.01.016>
- Martin, N. A. (2011). *A 100% tree inventory using i-Tree Eco protocol: A case study at Auburn University, Alabama* [Doctoral dissertation].
- Mathey, J., Rößler, S., Banse, J., Lehmann, I., & Bräuer, A. (2015). Brownfields as an element of green infrastructure for implementing ecosystem services into urban areas. *Journal of Urban Planning and Development*, 141(3), A4015001. [https://doi.org/10.1061/\(ASCE\)UP.1943-5444.0000275](https://doi.org/10.1061/(ASCE)UP.1943-5444.0000275)
- Nielsen, A. B., Östberg, J., & Delshammar, T. (2014). Review of urban tree inventory methods used to collect data at single-tree level. *Arboriculture & Urban Forestry*, 40(2), 96–111. <https://doi.org/10.48044/jauf.2014.011>
- Östberg, J., Delshammar, T., Wiström, B., & Nielsen, A. B. (2013). Grading of parameters for urban tree inventories by city officials, arborists, and academics using the Delphi method. *Environmental management*, 51(3), 694–708. <https://doi.org/10.1007/s00267-012-9973-8>
- Östberg, J., Martinsson, M., Stål, Ö., & Fransson, A. M. (2012). Risk of root intrusion by tree and shrub species into sewer pipes in Swedish urban areas. *Urban Forestry & Urban Greening*, 11(1), 65–71. <https://doi.org/10.1016/j.ufug.2011.11.001>
- Padayachee, A. L., Irlich, U. M., Faulkner, K. T., Gaertner, M., Proches, S., Wilson, J. R., & Rouget, M. (2017). How do invasive species travel to and through urban environments?. *Biological invasions*, 19(12), 3557–3570.
- Patterson, M. F., Wiseman, P. E., Winn, M. F., Lee, S. M., & Araman, P. A. (2011). Effects of photographic distance on tree crown attributes calculated using urbancrowns image analysis software. *Arboriculture & Urban Forestry*, 37(4), 173–179. <https://doi.org/10.48044/jauf.2011.023>
- Persson, A., Holmgren, J., & Soderman, U. (2002). Detecting and measuring individual trees using an airborne laser scanner. *Photogrammetric Engineering and Remote Sensing*, 68(9), 925–932.
- Pueffel, C., Haase, D., & Priess, J. A. (2018). Mapping ecosystem services on brownfields in Leipzig, Germany. *Ecosystem Services*, 30, 73–85. <https://doi.org/10.1016/j.ecoser.2018.01.011>
- Pu, R., & Landry, S. (2012). A comparative analysis of high spatial resolution IKONOS and WorldView-2 imagery for mapping urban tree species. *Remote Sensing of Environment*, 124, 516–533. <https://doi.org/10.1016/j.rse.2012.06.011>
- Ringland, J., Bohm, M., Baek, S. R., & Eichhorn, M. (2021). Automated survey of selected common plant species in Thai homegardens using Google Street View imagery and a deep neural network. *Earth Science Informatics*, 14(1), 179–191. <https://doi.org/10.1007/s12145-020-00557-3>

- Roth, K. L., Roberts, D. A., Dennison, P. E., Alonso, M., Peterson, S. H., & Beland, M. (2015). Differentiating plant species within and across diverse ecosystems with imaging spectroscopy. *Remote Sensing of Environment*, 167, 135–151. <https://doi.org/10.1016/j.rse.2015.05.007>
- Sikorski, P., Gawryszewska, B., Sikorska, D., Chormański, J., Schwerk, A., Jojczyk, A., Cieżkowski, W., Archiciński, P., Łepkowski, M., Dymitryszyn, I., Przybysz, A., Wińska-Krysiak, M., Zajdel, B., Matusiak, J., Łaszkiewicz, E. (2021). The value of doing nothing—how informal green spaces can provide comparable ecosystem services to cultivated urban parks. *Ecosystem Services*, 50, 101339. <https://doi.org/10.1016/j.ecoser.2021.101339>
- Silva, J., Bacao, F., Dieng, M., Foody, G. M., & Caetano, M. (2017). Improving specific class mapping from remotely sensed data by cost-sensitive learning. *International Journal of Remote Sensing*, 38(11), 3294–3316. <https://doi.org/10.1080/01431161.2017.1292073>
- Silván-Cárdenas, J. L. (2012, June). A segmentation method for tree crown detection and modelling from LiDAR measurements. In Mexican Conference on Pattern Recognition, (pp. 65–74). Springer, Berlin, Heidelberg.
- Simoniello, T., Coluzzi, R., Guariglia, A., Imbrenda, V., Lanfredi, M., & Samela, C. (2022). Automatic filtering and classification of low-density airborne laser scanner clouds in shrubland environments. *Remote Sensing*, 14 (20), 5127. <https://doi.org/10.3390/rs14205127>
- Small, C., & Lu, J. W. (2006). Estimation and vicarious validation of urban vegetation abundance by spectral mixture analysis. *Remote Sensing of Environment*, 100 (4), 441–456. <https://doi.org/10.1016/j.rse.2005.10.023>
- Sottini, V. A., Barbierato, E., Capecchi, I., Borghini, T., & Saragosa, C. (2021). Assessing the perception of urban visual quality: An approach integrating big data and geostatistical techniques. *Aestimum*, 79, 75–102.
- Wäldchen, J., & Mäder, P. (2018). Machine learning for image based species identification. *Methods in Ecology and Evolution*, 9(11), 2216–2225. <https://doi.org/10.1111/2041-210X.13075>
- Wallace, L., Sun, Q. C., Hally, B., Hillman, S., Both, A., Hurley, J., & Saldias, D. S. M. (2021). Linking urban tree inventories to remote sensing data for individual tree mapping. *Urban Forestry & Urban Greening*, 61, 127106. <https://doi.org/10.1016/j.ufug.2021.127106>
- Waser, L. T., Klonus, S., Ehlers, M., Küchler, M., & Jung, A. (2010). Potential of digital sensors for land cover and tree species classifications—a case study in the framework of the DGPF-project. *Photogramm Fernerkun*, 2010, 141–156. <https://doi.org/10.1127/1432-8364/2010/0046>
- Wegner, J. D., Branson, S., Hall, D., Schindler, K., & Perona, P. (2016). Cataloging public objects using aerial and street-level images—urban trees. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. (pp. 6014–6023), San Juan, PR, USA.
- Yao, W., & Wei, Y. (2013). Detection of 3-D individual trees in urban areas by combining airborne LiDAR data and imagery. *IEEE Geoscience and Remote Sensing Letters*, 10 (6), 1355–1359. <https://doi.org/10.1109/LGRS.2013.2241390>
- Zarrin, I. (2019, March). Leaf based trees identification using convolutional neural network. In 2019 IEEE 5th International Conference for Convergence in Technology (I2CT). (pp. 1–4), IEEE, Pune, India
- Zhang, C., Wang, L., & Yang, R. (2010, September). Semantic segmentation of urban scenes using dense depth maps. In European Conference on Computer Vision, (pp. 708–721). Springer, Berlin, Heidelberg.