sustainability

MDPI

Article

Use of Machine Learning Techniques on Aerial Imagery for the Extraction of Photovoltaic Data within the Urban Morphology

Fabio Giussani [1,2,*], Eric Wilczynski [1], Claudio Zandonella Callegher [1], Giovanni Dalle Nogare [1], Cristian Pozza [1], Antonio Novelli [3] and Simon Pezzutto [1]

[1] Eurac Research, Institute for Renewable Energy, Viale Druso 1, 39100 Bolzano, Italy; eric.wilczynski@eurac.edu (E.W.); claudio.zandonellacallegher@eurac.edu (C.Z.C.); giovanni.dallenogare@eurac.edu (G.D.N.); cristian.pozza@eurac.edu (C.P.); simon.pezzutto@eurac.edu (S.P.)
[2] Laboratorio di Simulazione Urbana Fausto Curti, Department of Architecture and Urban Studies (DASTU), Politecnico of Milan, Via Bonardi, 3, 20133 Milan, Italy
[3] RHEA Group, Via di Grotte Portella 28, Edificio Clorofilla, Scala C, Piano 3, 00044 Frascati, Italy; a.novelli@rheagroup.com
* Correspondence: fabio.giussani@eurac.edu; Tel.: +39-0471055166

Abstract: Locating and quantifying photovoltaic (PV) installations is a time-consuming and labor-intensive process, but it is necessary for monitoring their distribution. In the absence of existing data, the use of aerial imagery and automated detection algorithms can improve the efficiency and accuracy of the data collection process. This study presents a machine learning approach for the analysis of PV installations in urban areas based on less complex and resource-intensive models to target the challenge of data scarcity. The first objective of this work is to develop a model that can automatically detect PV installations from aerial imagery and test it based on the case study of Crevillent, Spain. Subsequently, the work estimates the PV capacity in Crevillent, and it compares the distribution of PV installations between residential and industrial areas. The analysis utilizes machine learning techniques and existing bottom-up data to assess land use and building typology for PV installations, identifying deployment patterns across the town. The proposed approach achieves an accuracy of 67% in detecting existing PV installations. These findings demonstrate that simple machine learning models still provide a reliable and cost-effective way to obtain data for decision-making in the fields of energy and urban planning, particularly in areas with limited access to existing data. Combining this technology with bottom-up data can lead to more comprehensive insights and better outcomes for urban areas seeking to optimize and decarbonize their energy supply while minimizing economic resources.

Keywords: photovoltaic; installed capacity; machine learning; object detection; aerial imagery; urban morphology; building stock analysis

check for updates

Citation: Giussani, F.; Wilczynski, E.; Zandonella Callegher, C.; Dalle Nogare, G.; Pozza, C.; Novelli, A.; Pezzutto, S. Use of Machine Learning Techniques on Aerial Imagery for the Extraction of Photovoltaic Data within the Urban Morphology. *Sustainability* **2024**, *16*, 2020. https://doi.org/10.3390/su16052020

Received: 8 January 2024
Revised: 12 February 2024
Accepted: 18 February 2024
Published: 29 February 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The use of photovoltaic (PV) is crucial for reducing carbon emissions in our energy system; it is gaining more government support globally as a means of achieving sustainable energy transformation and addressing climate change. During 2021, the solar market in the European Union (EU) experienced an expansion of 25.7 GW of capacity installed for a total installed capacity of 162 GW, making solar energy the source of approximately 5.7% of the EU's overall electricity production [1]. Solar power is an economical, environmentally friendly, and adaptable energy source. In the past 10 years, the cost of solar energy panels has decreased by about 82%, making it the most cost-effective electricity option in several regions of the EU [1]. Over time, the development of solar PV technology has significantly improved thanks to technological advancements, reduced material costs, and the global push for electricity generation using renewable sources. As a result, the utilization of solar PV technology has substantially increased.

Obtaining geospatially precise information on rooftop PV for localized regions like towns and counties presents a significant challenge. Monitoring PV installations is important to a variety of stakeholders and applications. For instance, utility enterprises need PV databases to execute long-term capacity planning for electricity networks. Governments and policymakers depend on current and reliable PV databases to formulate, monitor, and evaluate energy policies related to PV adoption. Knowledge of accurate PV data is an asset to making educated choices when determining energy policies and regulations, planning for capacity expansion, upgrading transmission and distribution systems, and making operational decisions to maintain grid reliability and resilience. Researchers also benefit from updated data when developing innovative solutions in the PV field [2–4].

Another value of PV monitoring is the use of the data for the purpose of urban planning. The study of the distribution of PV in a city can provide valuable insights to urban planners. By analyzing the location and density of PV installations, they can identify areas with high potential for solar energy generation and incorporate this information into their planning decisions. For example, they can prioritize the development of solar-friendly building codes and zoning regulations, encourage the installation of PV systems in public buildings and spaces, and promote the use of solar energy in transportation systems. This can help to reduce the city's carbon footprint, increase energy efficiency, and improve the overall sustainability of the urban environment. Additionally, the study of PV distribution can help urban planners to identify areas that are vulnerable to power outages and develop strategies to improve grid resilience and reliability [5,6].

PV systems are also a cornerstone for the development of positive energy districts (PEDs) in dense urban contexts because other sources of energy such as wind turbine systems or biomass are less frequently found within a dense setting and may require extensive transportation, depending on the distance from the energy generation system to the consumers [7]. In dense urban environments, the presence of tall buildings can make solar availability and urban daylight scarce, creating a complex settlement scenario. Photovoltaic (PV) panels on the roof of multifamily residential buildings can provide less energy per capita to the inhabitants of the building. This is due to the limited roof area; multifamily residential buildings often have limited roof area compared to single-family homes, which can limit the amount of PV panels that can be installed.

Analyzing the distribution of PV installations in relation to the urban morphology in an urban context is, therefore, necessary for the development of solar planning and energy policymaking.

One area for improvement with the quick growth of solar PV, however, is the monitoring and mapping of their installations. PV systems are primarily smaller in scale and installed on rooftops by individual owners, in contrast to other electricity generation technologies such as coal, gas, or wind power plants. The decreasing cost of PV systems and the subsequent rapid increase in the use of this technology have presented a challenge for monitoring all the installations due to their decentralized nature and large quantity [8].

According to various sources, solar PV and the energy they produce are not appropriately recorded or tracked in a centralized database; the information about the distribution of solar PV installations can vary considerably from country to country. Traditional methods, such as surveys and utility company data logs, are not suitable for this purpose as they are burdensome and yield insufficient data for the desired level of precision [3]. Additionally, data estimated using these methods quickly become outdated due to the rapid growth of rooftop PV; costly and periodic data collections are necessary to keep the data current.

One way to keep track of installed PV systems is by gathering data through self-reports like the "Tracking the Sun" project [9] and Germany's official PV registry [10]. These registries can be customized to collect specific information about the PV systems, including ownership. However, manual data collection can be time-consuming and prone to human error. Moreover, PV systems are often registered by street address, which may not always accurately reflect the system's actual location. There are situations in which network operators keep records of the precise locations and capacities of all solar PV systems that

are connected to their networks. Despite this, there is a possibility that a portion of systems are not recorded or are inadequately logged or that omissions in the registry may occur. Also, in certain cases, network operators may decline or be unable to disclose these data to external parties [2,11]. A Dutch study reported a discrepancy of about 25% between registered and unregistered PV installations in several residential areas of the country [12].

The adoption of solar PV in off-grid settings is also increasing significantly. Such systems are unlikely to be recorded with a centralized entity, and yet it would be beneficial to acknowledge their existence to monitor the overall expansion of solar PV in a given area.

Object detection from aerial imagery has been investigated extensively over the last decade, however, the detection of solar panels has been unexplored until very recent years. Several promising methods have emerged so far, although the use of satellite and aerial imagery for identifying solar PV systems is still a developing field of research. Malof et al. 2016 [11] have made significant advancements in two types of approach. Initially, they employed a support vector machine for object detection using standard pixel-wise feature extraction and classification. Later, they replaced the support vector machine with a random forest classifier to assign probabilities to each pixel, enabling segmentation. Their research progressed to the use of a convolutional neural network with convolutional and max-pooling layers for further development [3].

Overall, the literature on this subject has focused primarily on two detection algorithms, random forest (RF) classifiers [13] and convolutional neural networks (CNN) [11], both of which have been effective in image recognition tasks. While both approaches have proven successful, CNNs have outperformed other methods on significant image recognition benchmarks in recent years. Nevertheless, the time required to train CNNs is often substantially greater than that of other competing algorithms, such as RF. Furthermore, the process of designing CNNs is challenging, and complex models may require extensive testing [14]. DeepSolar is a noteworthy study that made significant progress in this field by training on more than 350,000 images, achieving a 90% precision and recall for detecting solar panels and a mean relative error of 2.1% for estimating their size. However, their methodology demands a substantial quantity of image data and computational power, making it only accessible to organizations with sufficient resources [8].

The aforementioned issues raise a need for a new approach to obtaining rooftop PV information in contexts of data scarcity, namely, utilizing algorithms that require a low level of training.

The ultimate objective of this case study is to investigate the relationship between urban residential and industrial areas concerning the presence of PV installations. Data elaborated with machine learning technology and existing bottom-up data are used together to analyze land use and typology of the buildings for PV installations, as well as to identify tendencies and patterns in the use of PV technology across the territory.

In the Materials and Methods section, the data sources used in the analysis are listed, and the steps followed during the process of data elaboration are described. The results of the different steps of the work are presented in the Results section, and their meanings and implications are further elaborated in the Discussion section.

2. Materials and Methods

The process of detection of PV installations is conducted in a geographic information system (GIS) environment. The software of choice is GRASS GIS Version 8.2, in which a variety of works of object-based classification and image segmentation has been conducted and properly documented [15,16]. The workflow of the machine learning model used in this work is shown in Figure 1.

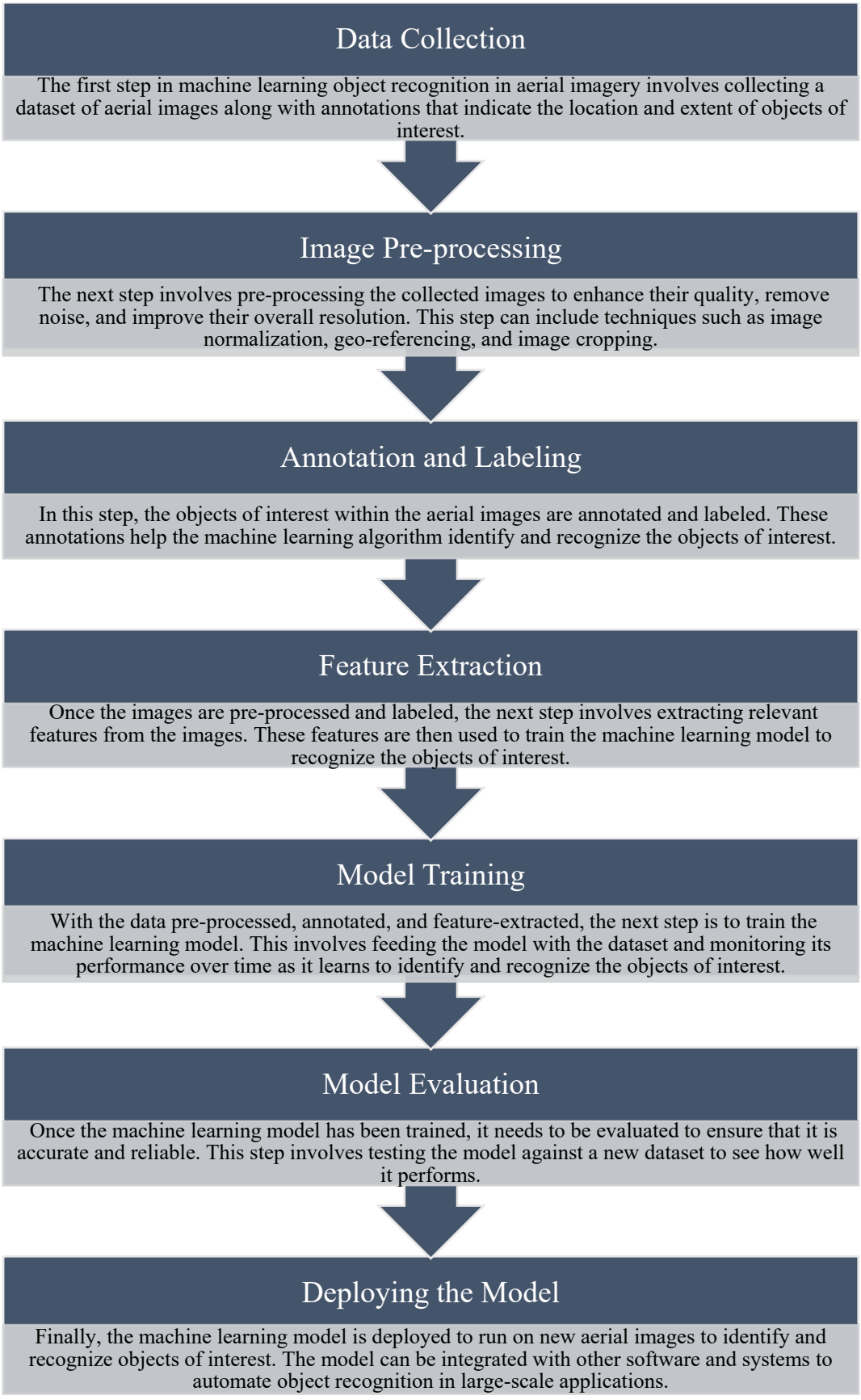


Figure 1. Steps for the application of a machine learning model on aerial imagery.

Crevillent is a Spanish town with around 29,717 inhabitants as of 2021, according to the Spanish National Institute of Statistics—INE [17]. It is located in the southeast of the country, in the Alicante province, and it is part of the administrative division "Comunitat Valenciana".

Crevillent was chosen as a representative case study thanks to the access of bottom-up data provided by Istituto Valenciano de la Edificaciòn (IVE), which served as faithful ground truth data against which the result of the work could be compared and evaluated.

Furthermore, Crevillent presents a clear distinction between a historic city center, mostly serving the function of residential area, and new industrial districts developed in the last decades and a large extent of agricultural land plots [18].

The aerial imagery used in this project is freely available as open data and it has been retrieved from the portal of Infraestructura Valenciana de Dades Espacials (IDEV) [19].

The imagery consists of orthophotos in a true color representation (red, green, and blue (RGB)) and false infrared color (IRG) covering the extent of the Comunidad Valenciana at a resolution of 25 cm (centimeters) per pixel, with 8-bit color depth per band. The images are based on an RGBI digital photogrammetric flight and were taken from 8 May 2022 to 11 June 2022. The data were downloaded in the form of six 1:5000 sheets in a TIFF format, which were then patched as a unique image in a GIS environment.

Bottom-up data on PV installations in Crevillent were provided by Enercoop, the energy cooperative managing the electricity network of the town. The data were in the form of a list of 98 addresses with photovoltaic installations as of 2022 and the relative installed inverter capacity expressed in Watts.

The list of addresses with PV installations was used as a source to manually locate the buildings within the aerial imagery. It has not been possible to execute a work of automated geocoding as the addresses have been provided in different formats that are not recognized by web mapping platforms. Instead, addresses have been manually located using the Spanish cadaster web service [20]. Photovoltaic (PV) installations have been manually annotated by drawing vector polygons within QGIS. Polygons were drawn over groups of PV panels adjacent one to the other. As there are no plans to address individual panels or treat each panel separately, the decision was made not to draw polygons over singular panels that are installed in proximity to one another. In total, 416 polygons have been labeled within the municipality of Crevillent. The total area of PV surface annotated amounts to 20,223.8 m^2. These annotations will then be used as ground truth for the evaluation of the effectiveness of the detection model. Please see Figures 2 and 3 for examples of manual annotation of PV installations. Figures 2 and 3 adapted by the authors on QGIS, are examples of manual annotation of PV installations. In Figure 2a, the aerial image of the municipal stadium of Crevillent shows the presence of PV installations over the roof of the seating area. Figure 2b represents the manual annotations of the aforementioned PV installations with the color purple. Similarly, Figure 3a represents an industrial building in Crevillent with PV installations mounted on the roof, while Figure 3b includes the manual annotations, in purple, of the PV installations.

A principal component analysis (PCA) has been performed in GRASS GIS to obtain the definition of band statistics that best highlights the presence of PV panels. In image processing, principal component analysis (PCA) is a linear technique used to look for a new representation domain of the multidimensional information enclosed in the multichannel image data while preserving most of the relevant information. It computes the covariance matrix of the image pixels and finds the eigenvectors and eigenvalues of the covariance matrix. The eigenvectors represent the principal components of the image (i.e., the new representation domain), which are linear combinations of the original pixel values. The eigenvalues indicate the amount of variance in the image explained by each principal component [21].



(**a**) (**b**)

Figure 2. (**a**) Aerial image of the municipal stadium of Crevillent. (**b**) Manual annotation (purple) of PV installations at the municipal stadium.



(**a**)

(**b**)

Figure 3. (**a**) Aerial image of an industrial building in Crevillent. (**b**) Manual annotation (purple) of PV installations on the roof of an industrial building.

To obtain a better result in the automatic recognition of PV panels, it was decided to operate only on roof mounted PV installations, avoiding a large number of ground elements that could be recognized by the algorithm as false positives. This approach has been adopted in previous works for the same operative reasons; rooftop-mounted installation is preferable in the vast majority in cities and towns due to the scarce amount of ground space available [14,22]. This step also allowed us to drastically reduce the computational resources needed by the software to run the machine learning algorithms. This choice was made after researching and verifying the absence of large-scale ground-mounted PV plants within the territory of Crevillent. Data about the footprint of the buildings within the municipality of Crevillent were downloaded from the spatial database of the Generalitat Valenciana IDEV. The downloaded data are at a scale of 1:5000 and were last updated in January 2023 [23].

Due to a tilt in the aerial imagery at disposal, a portion of the facades of buildings was visible and represented an obstacle for the automatic detection of objects, as it was included in the vectorial polygons recognized as buildings. To overcome this issue, an operation of orthorectification was conducted. Orthorectification is a process of removing the distortions present in remotely sensed imagery and converting it into an accurate, corrected, and georeferenced image that can be used for mapping and analysis purposes [24].

The process involved the identification of the exact location of the three cameras from which the three different orthophotos composing the municipality of Crevillent had been taken. Then, three buildings were selected, serving as ground control points. The three different orthophotos were then overlapped. This allowed for the manual reconstruction of the seamlines composing the mosaic of the pictures by finding areas where there were large differences between the adjacent images. Seamlines are the boundaries between adjacent image strips that are mosaicked together to create a larger, seamless image.

Once the seamlines dividing the three pictures had been delineated, it was possible to exclude the facades from the vectorial footprints of buildings, as shown in Figure 4, adapted by the authors on QGIS.

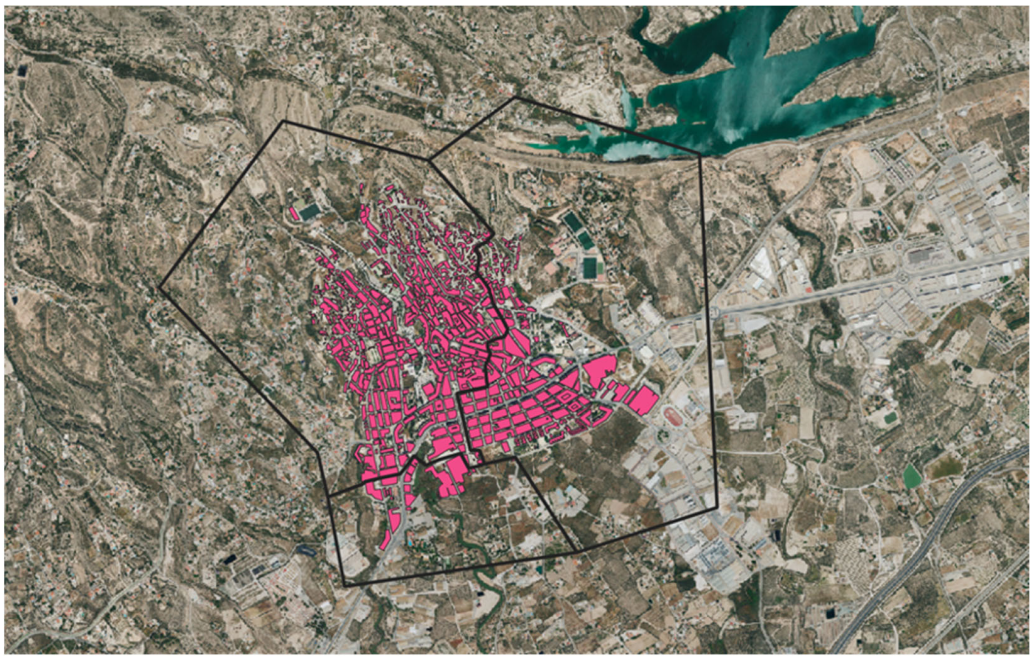


Figure 4. Seamlines (black) detected within the built center of Crevillent (pink).

The detection of PV panels was performed through techniques of machine learning using the function r.learn.ml in GRASS GIS. This function allowed us to perform supervised classification and regression of GRASS raster data using the Python scikit-learn package [25].

To train the model, two classes of images were used: one with PV installations and the other one with roofs. A total of 60 images were randomly selected for each of these two classes. Training images of PV panels were created by randomly selecting PV installations and drawing polygons over them, as well as by selecting smaller sections from various PV installations. A total of 40 images were used as test images, creating a ratio of 75:25 between the training set and the test set.

The machine learning model used for the detection of PV installation in this context of scarcity of raw data is the logistic regression algorithm. The choice to use traditional machine learning methods rather than deep learning methods (e.g., CNN) was made due to the extensive training required for the latter, in favor of a more immediate approach in line with the scope and resources of the project. Moreover, the amount of collected ground truth could not support the implementation of CNN training task. Notable examples of application of CNN are the works of Xia et al. (2023), in which CNN was used to identify and measure the area of precipitate in novel chromium-based alloys [26]. Overall, many research works have sought to test and deploy models that are not computationally expensive, as in the work of Cheng et al. (2022), who developed a ML model for the prediction of wildfires using satellite imagery [27].

Several classification and regression methods are available within the r.learn.ml function. Due to the relatively small number of training images and modest size of the area to be analyzed, the technique of logistic regression was chosen over other methods like random forest classifier. Logistic regression is part of the generalized linear model (GLM) family and is used specifically to model the probability of a binary or categorical outcome based on one or more predictor variables. It assumes that the relationship between the predictor variables and the outcome is linear in the transformed space and can be represented by a logistic function. Logistic regression is a simple and interpretable algorithm that can be used for both small and large datasets, and it works well when the relationship between the predictors and the outcome is linear or when the data are not too complex. However, it may not work well when the relationship between the predictors and the outcome is nonlinear or when there are interactions between the predictors [28].

Random forest classifier, on the other hand, is a nonlinear ensemble model that combines multiple decision trees to classify data. It works by randomly selecting subsets of the data and the predictors to build each decision tree and then aggregating the results of the trees to make a final prediction. Random forest is a powerful algorithm that can handle nonlinear relationships, interactions between predictors, and high-dimensional data. It is also less prone to overfitting than a single decision tree. However, random forest can be more complex and harder to interpret than logistic regression, and it may perform poorly on small datasets or datasets with imbalanced classes [29].

The next step consisted of using the r.neighbors function within GRASS GIS to obtain more homogeneous results, filling gaps in areas markedly recognized as PV panels, as well as filtering out single pixels or very small patches that can be considered as outliers. In GRASS GIS, the r.neighbors function is used to perform a neighborhood operation on a raster map layer. The function calculates a new value for each cell in a raster map based on the values of its neighboring cells within a specified window size [30].

The performance of the model is measured using mean intersection over union (mIoU), which measures the number of pixels common between the ground truth and prediction masks divided by the total number of pixels present across both masks. Intersection over union (IoU) is a number that quantifies the degree of overlap between two boxes. In the case of object detection and segmentation, IoU evaluates the overlap of the ground truth and prediction region [31]. This evaluation method was preferred over other methods (e.g., average precision, F1 score) because of the connotation of this work. The main objective is to estimate the PV capacity installed, which is directly linked and proportional to the area of PV panels detected. Therefore, a metric that informs of the surface detected, rather than the number of panels, is preferred.

The capacity of a PV installation can be calculated when the surface area of the installation is known.

As of 2022, on average, PV modules on the market have an efficiency ranging between 15% and 20% [10] Therefore, the PV efficiency value will be between 0.15 to 0.20.

The capacity, expressed in kWp, is then computed according to Equation (1), as follows:

$$Capacity\ (\text{kWp}) = PV\ Area \times PV\ Efficiency\ value \quad (1)$$

The result is presented as a range, depending on the efficiency of the module. This accounts for the impossibility of having data on the efficiency of single installed PV modules relying solely on aerial imagery.

The slope of the examined PV installations poses another variable in the calculation of its exact surface area. This can be solved through the use of a digital surface model (DSM) to verify the slope of the surface on which the PV modules are installed or the tilt of the PV module itself. For the conducted case study, however, no DSM matching the aerial imagery resolution of 0.25 m was available. A DSM with a 1 m resolution did not allow the exact determination of the slope of the modules except for a few extensive installations. This variable can be solved by accounting for those modules for which the slope cannot be identified, for a default slope value between 0 and 60 degrees. The optimal tilt of a PV panel depends on several factors, including location and season of the year [32]. For the purpose of this work, an average value of 30 degrees will be considered. The influence of this value does not drastically change the results of the capacity estimation, as proven by So et al. 2017 [33].

The equation for calculating the true area of an inclined PV panel, given the area detected from above, relies on trigonometry and it is computed according to Equation (2). It accounts for the detected area, divided by the COSIN of the angle at which the PV panel is inclined, which for the purpose of this work is 30 degrees.

$$True\ Area = Detected\ Area \times \frac{1}{\cos(30)} \quad (2)$$

Concerning the estimation of the capacity of each PV installation address, the meter of comparison of the success of the calculations was the list of addresses reporting the PV capacity expressed in kWp for 98 addresses in Crevillent. The list was provided by Enercoop to serve as ground truth data against the model.

For the purpose of this work, the detection of buildings is conducted using the Mapflow 2.5.0 software. The software is run as a plugin within QGIS, version 3.24.1. Mapflow is an artificial intelligence (AI) mapping platform that uses machine learning models to detect and extract features from satellite and aerial images. It can extract the roof prints of buildings from high-resolution imagery with reportedly high accuracy [34,35].

A normalized digital surface model (nDSM) of the province of Alicante from the year 2016 was downloaded from the Geoportal of the Comunitat Valenciana [36]. The nDSM was then joined with the vector layer of the buildings recognized with the Mapflow service. This passage linked the footprint of buildings with their height. The height of the building was calculated as an average of the measure of all the points within the building shape.

A spatial join command finally allowed us to link the buildings with the PV installations.

Information concerning the land use of the area was obtained from the municipal plan of the city by dividing the territory in three main categories: residential, industrial, and agricultural/rural. To operate within an urban context and calculate plausible and comparable measures of urban density and form, nonurban land (e.g., agricultural areas, big parks, rivers) was discarded from calculation, leaving residential and industrial areas as the two typologies urban form to study. In the context of Crevillent, residential areas present a dense fabric, tall buildings, and few open spaces. On the contrary, industrial areas present low-density urbanization characterized by two-story-tall buildings, on average,

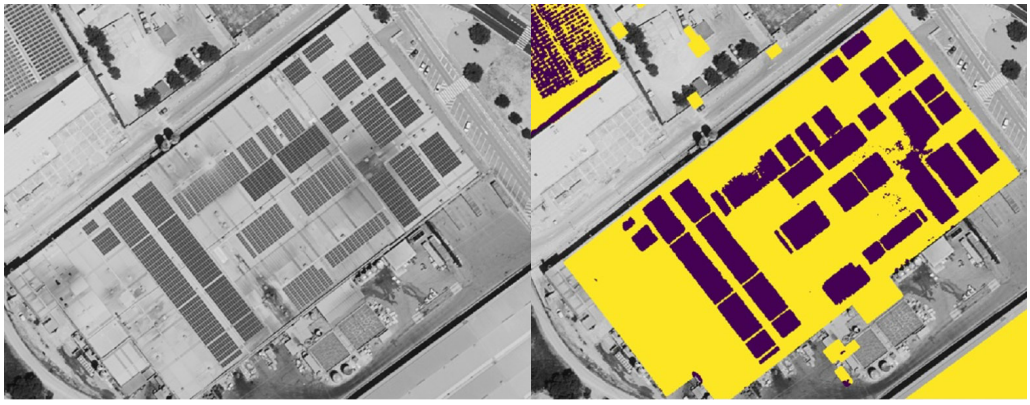
and generous open spaces. The distribution of PV installations within the urban tissue of Crevillent was analyzed based on the relation with the following variables:

- Land use of the area, either residential or industrial
- Number of buildings with PV installations
- Total area of PV installations
- Height of buildings with PV installations
- Area of roofs hosting PV installations

For the purpose of this work, a list of addresses with PV installations in the city of Crevillent, Spain, was used. The PV installations were manually located, and their area was calculated.

3. Results

The results of the logistic regression were compared against the manual annotations of PV installations used as ground truth. Figure 5, which was adapted by the authors on GRASS GIS, shows a specific example of how the model with logistic regression performed in detecting PV installations, compared to the ground truth. In particular, Figure 5a shows an industrial roof in Crevillent with PV installations, while Figure 5b shows the results of the trained model in detecting PV installations.



(**a**) (**b**)

Figure 5. (**a**) Aerial image of an industrial building in Crevillent. (**b**) Examples of results of logistic regression in detecting PV installations (purple). Source: Adapted by the author based on GRASS GIS.

The model used for identifying PV installations achieved a mIoU of 0.67. The result obtained with a logistic regression in this case study does not achieve the accuracy of more complex object detection models, which scored over 0.9090. However, it can still be considered satisfactory for the purpose of this work, considering the relatively low training required and low amount of input data needed. For the purpose of this work, different algorithms provided within the GRASS GIS r.learnml calculation module have been tested.

While logistic regression achieved the best score in terms of mean intersection over union, random forest classifier and gradient boosting classifier achieved lower scores.

By tuning the number of estimators, in the context of Crevillent, the random forest classifier achieved a MIoU between 38 and 47%, while the gradient boosting classifier did not reach 25%.

The results of the calculations concerning the installed photovoltaic capacity consist of a range, considering a PV module efficiency between 0.15 and 0.20.

Overall, 58 out of the 98 addresses fit within the calculated kWp range. An example of the calculation process for one address is reported in Figure 6, which was adapted by the authors on GRASS GIS.



Figure 6. Example of detection of PV installations (blue). Source: Adapted by the author based on GRASS GIS.

The logistic regression found the two separated PV installations and attributed to them surface areas of 29.1 m^2 and 21.3 m^2.

This address presents a total surface area of 50.4 m^2.

The slope of the roof, and therefore the slope of the PV installations, cannot be determined using the digital surface model. As a consequence, a range between 0 and 30 degrees of slope is attributed to the surface. In case the surface was at an angle of 30 degrees Equation (2) is used:

The value of COSIN(30) is 0.866. Therefore, in the example, the real area of the panel tilted at an angle of 30 degrees would be 58.19 m^2m^2.

The limits of the range of capacity is then calculated according to Equation (3) and Equation (4):

$$Minimum\ Value: \text{ True PV Area } \times\ 0.15\ =\ \text{PV capacity} \quad (3)$$

$$Maximum\ Value: \text{True PV Area } \times\ 0.20\ =\ \text{PV capacity} \quad (4)$$

The correct value of 10,000 Watts that was provided falls within the calculated range of 8.72 kWp to 11.64 kWp.

The analysis of the distribution of PV installations within the city of Crevillent highlighted differences between the residential and industrial areas of the city. The main residential area hosting rooftop mounted PV installations is the historical city center of Crevillent. The more recent residential area of San Felipe Neri also presents four buildings with PV installations.

The center presents the largest number of buildings with PV installations, 15, but it also hosts one of the lowest areas of PV installations. This element indicates a high degree of fragmentation of roofs in the city center as opposed to industrial areas.

The surface of PV installations in residential areas is 520 m^2m^2, distributed over 19 different buildings. This value is extremely small if compared with the surface of PV installations present in industrial areas, which is 18,424 m^2m^2, distributed over 35 different buildings.

In fact, the average area of a residential roof hosting PV installations is 22,017 m^2, which is lower than the 3953.6 m^2m^2 characterizing the average industrial roof with PV presence.

The height of the buildings hosting PV installation reflects the characteristic forms of the land use of interest. The average height of buildings with PV installations in residential areas, which are notoriously more dense, is 9.193 mm, while industrial buildings average 7.755 mm.

4. Discussion

The main advantage of the adopted methodology is that it requires relatively low effort in training. This is particularly useful in situations with a limited availability of data or resources. Table 1 shows how the results of the logistic regression compare to previous studies that employed a higher number of training images. It needs to be mentioned that the other studies were also conducted on a bigger spatial scale. It can also be considered as an aid for manual data collection concerning PV installations, which remains costly and time consuming.

Table 1. Comparison of results with previous studies on the topic of PV detection.

Study	Methodology	Size of Training Set	MIoU	Spatial Extent
Mayer et al., 2022 [37]	Deep Neural Network + 3D Spatial Data Processing Techniques	38,604	74.1%	Four Counties in Germany
SolarNet—Hou et al., 2019 [38]	Deep Learning	819	94%	439 solar farms in China
Zhuang et al., 2020 [39]	U-Net	921	73–74%	City of Fresno, USA
Roofpedia—Wu and Biljecki, 2021 [40]	Fully Convolutional Neural Network	2317	78.4%	Developed on 8 cities
Crevillent PV Detection	Logistic Regression	60	67%	Village of Crevillent

Although aerial imagery has proven to be effective in accurately identifying buildings, vehicles, and roads, solar PV systems present a unique set of characteristics that pose challenges to their identification. This section outlines the main challenges that have been encountered in the work and that have been common in related research on the topic.

PV systems can be easily mistaken for various objects that share similar visual characteristics. These objects may include solar hot water systems, skylights, edges of houses, cables, glass greenhouses, and even swimming pools. Performing the image detection only on roofs limits possible error sources, such as vehicle windshields and swimming pools, but the issue still persists to a certain extent [2].

False negatives present another challenge in the development of this typology of project. Solar PV systems can present a significant identification challenge, even for experienced human annotators, in various situations. For instance, this may occur when black panels are installed on black rooftops, especially when the image resolution is inadequate. Difficulties in identifying such systems even by manual annotators can result in lower quality of training sets, leading to less precise classification or segmentation. Additionally, small, or atypically configured solar PV systems, which are common on residential rooftops,

can be particularly challenging to identify. A similar issue is presented by new technologies that integrate PV installations in materials and styles developed to be less visible as possible, rejecting the common style of PV panels in favor of tiles and wall mounted installations [2].

One other aspect that can pose a challenge to the immediate replicability of the work is the heterogeneity in the availability of high-resolution aerial imagery depending on the local operational context.

The estimation of the capacity installed paves the way for a calculation of the actual energy output per PV installation, considering the azimuth of the panels and the hours of sun. These steps can be developed in future research.

The results obtained in the case study reflect the relationship between urban morphology, in particular urban density, and PV installations that emerge from the literature on the topic. The vast majority of PV surface is present in industrial areas connotated by lower building density. It can be assumed that this is caused by the favorable conditions that this typology of urban tissue offers, including less building density and extensive and unfragmented roof surfaces.

Poon et al. 2020 [41] also found that urban morphology on the neighborhood scale has a significant impact on electricity generation by PV panels that are installed on building rooftops and facades. However, the study also noted that morphological studies cannot provide replicable, transposable model types as each city is different.

The most effective way to evaluate the impact of urban morphology on rooftop solar potential is through a city-scale approach that takes into account the building footprint and shading patterns. Boccalatte et al. 2022 [42] used Geneva GIS data to evaluate the impact of urban morphology on rooftop solar radiation. The study mapped the proportion of the tessellation cell covered by the building footprint, which allowed for a precise mapping of the densest areas within the urban fabric.

If the development of solar energy is important in cities, which are the main consumers of energy, dense areas limit the incoming sunlight and the deployment of urban solar power plants. City centers present a more difficult environment for the expansion of PV, as building roofs are often of complex structure and split into numerous roof sections [43]. The largest roofs within dense urban tissues can often be found on public and service buildings.

While the compact city model is preferrable to a less dense setting because it permits synergies and green policies that require minimum population densities, it has been demonstrated how density impacts potential for passive solar architecture [43].

Because of its focus on the urban environment, the case study has not included rural areas of the territory; therefore, a share of PV installations was neglected. A future study may want to encompass rural areas, as well, for a more complete analysis.

5. Conclusions

PV installations are a key component of the decarbonization of the energy sector, and data about their presence and distribution are crucial for developing renewable energy policies and for energy planning. This study investigated possibilities in the use of machine learning technology and its efficacy as a solution to the challenge of data scarcity in the analysis of PV installations in urban areas. The town of Crevillent in Spain was used. It served as a sample of a delimited area where new data acquired by satellite imagery could be compared with existing data collected through traditional bottom-up methodology.

The first objective of this work was to develop a machine learning model that could automatically detect PV installations from aerial imagery and that required a relatively low level of training compared to more sophisticated algorithms. The model was used to estimate the PV capacity installed in Crevillent. The study also aimed to compare the distribution of PV installations between residential and industrial urban areas in the town.

Data acquired with machine learning technology were compared and integrated with existing bottom-up data to analyze land use and typology of the buildings for PV installations, as well as to identify patterns and trends in PV deployment across different

parts of town. Additionally, an estimate of the PV energy production was calculated, which provided information useful for future energy planning.

The first part of the work focused on the automatic detection of rooftop-mounted PV installations using logistic regression within GRASS GIS. This method was chosen for its low training requirements compared to more complex algorithms. The validation method of intersection over union achieved a score of 67% based on a list of addresses with PV installations serving as ground truth data. Although the method does not provide as high accuracy as other algorithms, the low amount of training required to obtain the data justifies the choice of this method.

In the second part of the project, a methodology was developed for calculating the installed PV capacity using the area of the detected PV installations. To account for the tilt of the PV panels, the first step consisted of finding the true area of the installations, and then an efficiency value of 0.15 to 0.20 was used to create a range of results. A list of addresses with reported installed capacity served as ground truth data; 58 of 98 of these addresses fit within the calculated range.

The third part of the work focused on an analysis of the distribution of PV installations in industrial and residential areas. First, footprints of buildings were detected using a machine using a GIS plugin and integrated with the previously detected PV installations. Then, the vectorial data were merged with a normalized digital surface model and with bottom-up information on the land use. This led to the analysis of different variables: land use of the area, either residential or industrial setting; the number of buildings with PV installations; the total area of PV installations; the heights of buildings with PV installations; and the square footage of roofs hosting PV installations. The findings show that the majority of PV installations are located on the roofs in industrial, less dense urban areas.

The findings demonstrate that machine learning technology provides a reliable and cost-effective way to obtain data for decision-making in the fields of energy and urban planning, particularly in areas with limited access to existing data. Moreover, combining this technology with bottom-up data can lead to more comprehensive insights and better outcomes for urban areas seeking to optimize their renewable energy supply while minimizing economic resources.

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

Funding: This research was funded by the Horizon Europe project "MODERATE", grant agreement number 101069834.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available at https://github.com/MODERATE-Project/building-stock-analysis (accessed on 8 February 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. European Commission. Solar Energy. 2023. Available online: https://energy.ec.europa.eu/topics/renewable-energy/solar-energy_en (accessed on 8 February 2024).
2. Hoog, J.; Maetschke, S.; Ilfrich, P.; Kolluri, R.R. Using satellite and aerial imagery for identification of solar PV. In Proceedings of the Eleventh ACM International Conference on Future Energy Systems, Virtual Event, 22–26 June 2020. [CrossRef]
3. Malof, J.M.; Hou, R.; Collins, L.M.; Bradbury, K.; Newell, R. Automatic solar photovoltaic panel detection in satellite imagery. In Proceedings of the 2015 International Conference on Renewable Energy Research and Applications (ICRERA), Palermo, Italy, 22–25 November 2015. [CrossRef]

4. Stowell, D.; Kelly, J.; Tanner, D.; Taylor, J.; Jones, E.; Geddes, J.; Chalstrey, E. A harmonised, high-coverage, open dataset of solar photovoltaic installations in the UK. *Sci. Data* **2020**, *7*, 394. [CrossRef]
5. Akrofi, M.M.; Okitasari, M. Integration of Solar Energy Considerations into Urban Planning/Design is Necessary to Ensure that Future Cities do not only Consume But Also Produce Energy Locally through Solar. *Urban Gov.* **2022**, *2*, 157–172. Available online: https://www.sciencedirect.com/science/article/pii/S2664328622000171 (accessed on 8 February 2024). [CrossRef]
6. Formolli, M.; Croce, S.; Vettorato, D.; Paparella, R.; Scognamiglio, A.; Mainini, A.G.; Lobaccaro, G. Solar Energy in Urban Planning: Lesson Learned and Recommendations from Six Italian Case Studies. 14 March 2022. Available online: https://www.mdpi.com/2076-3417/12/6/2950 (accessed on 8 February 2024).
7. Morello, E.; Bignardi, M.; Rudini, M.A. Proposal for a spatial planning support system to estimate the urban energy demand and potential renewable energy scenarios. In Proceedings of the International Conference CISBAT 2015 "Future Buildings and Districts—Sustainability from Nano to Urban Scale", Lausanne, Switzerland, 9–11 September 2015; pp. 603–608. [CrossRef]
8. Yu, J.; Wang, Z.; Majumdar, A.; Rajagopal, R. DeepSolar: A machine learning framework to efficiently construct a solar deployment database in the United States. *Joule* **2018**, *2*, 2605–2617. [CrossRef]
9. Barbose, G.; Darghouth, N.R. *Tracking the Sun: Pricing and Design Trends for Distributed Photovoltaic Systems in the United States*; USDOE Office of Energy Efficiency and Renewable Energy (EERE), Renewable Power Office. Solar Energy Technologies Office: Washington, DC, USA, 2019. [CrossRef]
10. Fraunhofer ISE. Recent Facts about Photovoltaics in Germany-Fraunhofer. 2021. Available online: https://www.ise.fraunhofer.de/content/dam/ise/en/documents/publications/studies/recent-facts-about-photovoltaics-in-germany.pdf (accessed on 8 February 2024).
11. Malof, J.M.; Collins, L.M.; Bradbury, K.; Newell, R.G. A deep convolutional neural network and a random forest classifier for solar photovoltaic array detection in aerial imagery. In Proceedings of the 2016 IEEE International Conference on Renewable Energy Research and Applications (ICRERA), Birmingham, UK, 20–23 November 2016. [CrossRef]
12. Stedin, Kwart van de Zonnepanelen Niet in Beeld. 2018. Available online: https://www.stedin.net/over-stedin/pers-en-media/persberichten/kwart-van-de-zonnepanelen-niet-in-beeld (accessed on 8 February 2024).
13. Breiman, L. Random Forests-Machine Learning. 2001. Available online: https://link.springer.com/article/10.1023/a:1010933404324 (accessed on 8 February 2024).
14. Malof, J.M.; Bradbury, K.; Collins, L.M.; Newell, R.G. Automatic detection of solar photovoltaic arrays in high resolution aerial imagery. *Appl. Energy* **2016**, *183*, 229–240. [CrossRef]
15. Lennert, M. (Université L. de B. (ULB)). A Complete Toolchain for Object-Based Image Analysis with GRASS GIS. No. 163. In FOSS4G Bonn 2016. 2016. Available online: https://av.tib.eu/media/20409 (accessed on 8 February 2024).
16. Grippa, T.; Lennert, M.; Beaumont, B.; Vanhuysse, S.; Stephenne, N.; Wolff, E. An open-source semi-automated processing chain for Urban Object-based classification. *Remote Sens.* **2017**, *9*, 358. [CrossRef]
17. INE. Alicante/Alacant: Población por Municipios y Sexo. 2023. Available online: https://www.ine.es/jaxiT3/Datos.htm?t=2856 (accessed on 8 February 2024).
18. de Crevillent, A. Agenda Urbana Crevillent 2030 | Ayuntamiento de Crevillent. 2023. Available online: https://www.crevillent.es/pagina/agenda-urbana-crevillent-2030/ (accessed on 8 February 2024).
19. IDEV. Ortofoto de 2022 de la Comunitat Valenciana en RGBI y de 25 cm de Resolución. 2023. Available online: https://geocataleg.gva.es/#/search?uuid=spaicv0202_2022CVAL0025&lang=spa (accessed on 8 February 2024).
20. de España, G. Sede Electrónica del Catastro. 2023. Available online: https://www1.sedecatastro.gob.es/ (accessed on 8 February 2024).
21. GRASS GIS. i.pca. 2023. Available online: https://grass.osgeo.org/grass82/manuals/i.pca.html (accessed on 8 February 2024).
22. Li, P.; Zhang, H.; Guo, Z.; Lyu, S.; Chen, J.; Li, W.; Song, X.; Shibasaki, R.; Yan, J. Understanding rooftop PV panel semantic segmentation of satellite and aerial images for better using machine learning. *Adv. Appl. Energy* **2021**, *4*, 100057. [CrossRef]
23. IDEV. Cartografia oficial de la Comunitat Valenciana a Escala 1:5.000 de l'Institut Cartogràfic Valencià. 2023. Available online: https://geocataleg.gva.es/#/results/series_cartograficas (accessed on 8 February 2024).
24. DLR. Orthorectification. 2023. Available online: https://www.dlr.de/eoc/en/desktopdefault.aspx/tabid-6144/10056_read-20918/ (accessed on 8 February 2024).
25. GRASS GIS. r.learn.ml. 2023. Available online: https://grass.osgeo.org/grass82/manuals/addons/r.learn.ml.html (accessed on 8 February 2024).
26. Xia, Z.; Ma, K.; Cheng, S.; Blackburn, T.; Peng, Z.; Zhu, K.; Zhang, W.; Xiao, D.; Knowles, A.J.; Arcucci, R. Accurate identification and measurement of the precipitate area by two-stage deep neural networks in novel chromium-based alloys. *Phys. Chem. Chem. Phys.* **2023**, *25*, 15970–15987. [CrossRef] [PubMed]
27. Cheng, S.; Jin, Y.; Harrison, S.P.; Quilodrán-Casas, C.; Prentice, I.C.; Guo, Y.K.; Arcucci, R. Parameter Flexible Wildfire Prediction Using Machine Learning Techniques: Forward and Inverse Modelling. *Remote Sens.* **2022**, *14*, 3228. [CrossRef]
28. Khurshid, H.; Khan, M. Segmentation and Classification Using Logistic Regression in Remote Sensing Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 224–232. [CrossRef]
29. Lei, T.; Wan, S.; Wu, S.; Wang, H. A New Approach of Ensemble Learning Technique to Resolve the Uncertainties of Paddy Area through Image Classification. 9 November 2020. Available online: https://www.mdpi.com/2072-4292/12/21/3666 (accessed on 8 February 2024).

30. GRASS GIS. r.neighbors. 2023. Available online: https://grass.osgeo.org/grass82/manuals/r.neighbors.html (accessed on 8 February 2024).
31. Ren, S.; Hu, W.; Bradbury, K.; Harrison-Atlas, D.; Valeri, L.M.; Murray, B.; Malof, J.M. Automated Extraction of Energy Systems Information from Remotely Sensed Data: A Review and Analysis. *Appl. Energy* **2022**, *326*, 119876. [CrossRef]
32. Rasouli, Z.; Puig, V. Tilt Angle Optimization of Photovoltaic Panels. 2019. Available online: https://www.iri.upc.edu/publications/show/2260 (accessed on 8 February 2024).
33. So, B.; Nezin, C.; Kaimal, V.; Keene, S.; Collins, L.; Bradbury, K.; Malof, J.M. Estimating the electricity generation capacity of solar photovoltaic arrays using only color aerial imagery. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 1603–1606. [CrossRef]
34. Mapflow. Mapflow AI Models. 2023. Available online: https://docs.mapflow.ai/ (accessed on 8 February 2024).
35. Alsabhan, W.; Dudin, B.; Alotaiby, T. Detecting Buildings and Nonbuildings from Satellite Images Using U-Net. 2022. Available online: https://pubmed.ncbi.nlm.nih.gov/35571708/ (accessed on 8 February 2024).
36. IDEV. Normalized Digital Surface Model (nDSM) of LIDAR of 1 Meter Resolution Covering the Province of Alicante 2016. 2023. Available online: https://catalogo.icv.gva.es/geonetwork/srv/eng/catalog.search#/metadata/spaicv030501_2016PALI0100 (accessed on 8 February 2024).
37. Mayer, K.; Rausch, B.; Arlt, M.L.; Gust, G.; Wang, Z.; Neumann, D.; Rajagopal, R. 3D-PV-Locator: Large-scale detection of rooftop-mounted photovoltaic systems in 3D. *Appl. Energy* **2022**, *310*, 118469. [CrossRef]
38. Hou, X.; Wang, B.; Hu, W.; Yin, L.; Wu, H. SolarNet: A Deep Learning Framework to Map Solar Power Plants in China From Satellite Imagery. arXiv. 10 December 2019. Available online: http://arxiv.org/abs/1912.03685 (accessed on 1 February 2024).
39. Zhuang, L.; Zhang, Z.; Wang, L. The automatic segmentation of residential solar panels based on satellite images: A cross learning driven U-Net method. *Appl. Soft Comput.* **2020**, *92*, 106283. [CrossRef]
40. Wu, A.N.; Biljecki, F. Roofpedia: Automatic mapping of green and solar roofs for an open roofscape registry and evaluation of urban sustainability. *Landsc. Urban Plan.* **2021**, *214*, 104167. [CrossRef]
41. Poon, K.; Kämpf, J.; Tay, S.; Wong, N.; Reindl, T. Parametric Study of urban morphology on building solar energy potential in Singapore context. *Urban Clim.* **2020**, *33*, 100624. [CrossRef]
42. Boccalatte, A.; Thebault, M.; Ménézo, C.; Ramousse, J.; Fossa, M. Evaluating the impact of urban morphology on rooftop solar radiation: A new city-scale approach based on Geneva GIS data. *Energy Build.* **2022**, *260*, 111919. [CrossRef]
43. Carneiro, C.; Morello, E.; Desthieux, G. Assessment of solar irradiance on the urban fabric for the production of renewable energy using LIDAR data and image processing techniques. In *Advances in GIScience: Proceedings of the 12th AGILE Conference*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 83–112. [CrossRef]

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