

ARTICLE

Typology-Based Verification of Green-Roof Potential Sites in Lisbon through WorldView-2, Sentinel-2, and Building-Objects Validation

Antônio T. Reis^{1,*} and José Pedro¹

¹ Federal University of Rio Grande do Sul (UFRGS), Brazil

* Correspondence: tarcísio@orion.ufrgs.br

Abstract

Green-roof inventories for municipalities should be able to discriminate between confirmed roofs greening and vegetation intersecting building footprints. It is true especially for Lisbon with historical red clay roofs, flat individual roofs, trees on the borders of streets, and communal podium-type green spaces in a dense urban fabric. This paper investigates whether Lisbon's candidate inventory reflects the condition which should be prioritized for field validation when the accuracy of image-class classification is high but the object assignment differs by morphology. The dataset consists of an eight-band WorldView-2 image obtained on 18 February 2022, twelve observations made monthly with Sentinel-2 in 2023, 54,800 municipality building footprints, 4,015 grouped-building detections, 371 final candidates, and 196 verified green roofs. The investigation follows five-stage procedure that is built into the Lisbon record: class-specific reliability of the image classification, successive elimination, beta-binomial posterior credibility, seasonal NDVI margins, and inspection burden distribution. The classification audit of the WorldView-2 image is quite successful as evidenced by the overall accuracy value of 0.950, Cohen's κ of 0.914, vegetation F1 of 0.960, and unsuitable-roof F1 of 0.947. These numbers confirm the greenness and materials-based screening but not the roof attribution. Validation of the objects is very typologically specific: 50 out of 176 individual-roof candidates pass the test, versus 146 out of 195 shared space candidates. The verification odds for the shared space path are 7.51 higher than for the individual-building path, and the reliability-adjusted posterior intervals stay well-separated. The seasonal NDVI discriminates the two types too, since top-level roofs span 0.25-0.31 margins around the 0.30 gate, while shared space roofs are located above the gate at 0.32-0.49. In summary, Lisbon's candidate map includes two separate streams of verification: individual roofs need specific error detection while common space roofs present the high yielding approach to completion of the city's inventory.

Keywords: Lisbon; green roofs; WorldView-2; Sentinel-2; NDVI; validation probability; roof morphology

1. Introduction

Green roofs can now be considered a part of the working infrastructure of dense cities rather than architectural constructions. While they have the potential to serve as contributors to runoff delay, thermal moderation, roof-surface protection, habitat provision, and amenity of the locality, they depend on construction parameters including the

presence of drainage layer, substrate depth, vegetation assemblage, irrigation regimen, maintenance status, and local climate exposure [1–4]. The problems of Mediterranean and Atlantic-Mediterranean cities make the identification of green roofs particularly challenging due to stress in dry season, intense sun, heterogeneous building ages, and a large proportion of historical tie roofing when visible greenness becomes an incomplete indication of a constructed green roof [5, 6]. For municipal management, the problem is not only to estimate the possibility of the provision of the ecosystem service by green roofs but also to distinguish reliable candidates in the inventory. Thus, incentive policies and plans of green infrastructure should have spatial evidence sufficient to support inspections, registration, prioritization of maintenance, and investments in the future [7, 8].

The use of remote sensing technology allows for making the roof greening visible at the scales which are inaccessible for municipal teams. High-resolution optical images can detect the vegetation on the roofs, while building footprint layers provide the information about the boundaries of the objects to proceed from pixels to candidates of the roof greening. NDVI is the useful tool in this regard because it shows the difference between the red absorption and the near-infrared reflectance in active vegetation [9–12]. However, a green roof inventory is not a map of vegetation. In dense urban environment, the NDVI-positive signal can belong to the roof garden, tree crown crossing the footprint boundary, ground-level courtyard, vegetated podium over underground construction, or mixed pixel caused by the mismatch between building geometry and image resolution. This explains the need to separate classification accuracy and the credibility of objects. The literature on the assessment of classification accuracy proves the necessity of separate consideration of class-specific and allocation-specific aspects hidden under the single overall value in the case when mapped classes have different planning function [13, 14].

Modern studies of roofscapes expand the range of methods to be used for mapping green roofs. The study conducted by Roofpedia proves the capability of automatic mapping of green and solar roofs in multiple cities as well as the importance of public roofscape registry for the sustainability assessments [15]. The case study of London roofscape uses the aerial image segmentation for the detection of the green roofs on the building level to prove the promise of automated inventory and the problem arising from the rare positive cases and complicated urban texture [16]. The other studies combine information about the roof geometry, land cover data, and suitability criteria to detect areas suitable for the roof greening [17, 18]. All these studies prove the feasibility of automated generation of green roof candidates but also highlight another question to be answered: what type of the candidates should be inspected first and what type mostly helps to understand false positives.

Object detection is one of the tools available for detecting roof conditions which cannot be estimated with the help of building footprints. Faster R-CNN introduced the approach based on the region-proposal learning for the object detection, and residual networks provided the improvement of feature learning in deeper models [19, 20]. Such techniques found application in the remote sensing of buildings, greenhouses, and other spatial objects where the shape and context become as important as spectral responses [21–24]. Green roof applications of Faster R-CNN and other detectors work well where there are repeated building arrangements or podium-like surfaces allowing to recognize spatial patterns [25]. Nevertheless, the detector confidence does not mean the validation. Thus, it is possible to say that a detector recognizes the appropriate geometry of residential block, NDVI confirms the greenness, and the field inspection is still necessary to check that the surface belongs to a roofed structure and should be included in the municipal green roof registry.

Lisbon becomes the illustrative example in the case when there are two roof conditions in the same city and the same evidence sequence [26]. In the case of the individual buildings, it is possible to start with 54,800 municipal footprints, filter out unsuitable roof materials, apply the 0.30 NDVI gate, reduce edge contamination with the help of 3.6 m inward buffer, and have 176 final candidates. In the case of the grouped buildings, it is necessary to use 4,015 Faster R-CNN detections of the repetitive residential block shapes, apply the filtering according to the detector confidence and NDVI and have 195 candidates. The validation results in 50 individual green roofs and 146 shared space green roofs. Therefore, there are two similar in terms of candidate numbers but very different in terms of validation yield. This situation leads to the research question which is more practical than the general mapping question: when the Lisbon inventory is analyzed with the help of classification audit, attrition sequence, posterior validation probability, seasonal NDVI envelopes, and inspection burden, which roof condition becomes the verification priority and how should the field effort be allocated?

This study evaluates the Lisbon mapping record with the help of the fixed evidence sequence and keeps the detector, NDVI gate, buffer distance, and candidate generation rules. The main question is to identify whether the inventory provides high-yield candidate stream and targeted error diagnosis stream. Image-class reliability measures the ability to recognize the vegetation and screen out roof materials. Attrition sequence detects the places where the loss of objects occurs. The beta-binomial posterior credibility evaluates the uncertainty of validation results rather than using only the point rates. Seasonal NDVI margins allow to understand whether the 0.30 gate works equally well for top-level roofs and shared space roofs. The inspection-burden allocation will help to evaluate which route contributes more confirmed roofs and which contributes more non-confirmed candidates. Thus, the final interpretation of the Lisbon mapping record provides the city with a practical verification order without any changes in the mapping procedure.

The roofscape context represented in Figure 1 highlights the empirical reason to keep two verification streams. Historical red clay roofs, individual flat roofs, and grouped residential blocks lead to different geometric relationship between the vegetation and buildings. Thus, the vegetation signal over the flat roof, courtyard edge, and shared podium surface should not have the same verification meaning if it passes the same NDVI threshold.



Figure 1. Lisbon roofscape context.

The illustration highlights the empirical justification for maintaining dual verification pathways because individual high-level roof types need an accurate footprint-to-vegetation association, whereas common green areas need to be aware of their surrounding built form. The rest of the article follows this dichotomy through the classification audit, candidate thinning, posterior verification, seasonal margins of greenness, and inspection burden interpretation.

2. Materials and Lisbon Verification Record

2.1. Case-study setting and remote-sensing record

The case study is Lisbon, Portugal, a concentrated Atlantic-Mediterranean capital with a municipal area of 100.05 km², a population density of 5,438 people per km², and an assorted building stock characterized by multi-family residential buildings. The green roof mapping record relates to a city in which historical districts comprise numerous sloping roofs with red clay roofing materials, but new areas include condominium-type building complexes with private green roof gardens located above underground garages. It is critical in the sense that it affects the interpretation of vegetation located in close proximity to buildings. Individual-building roof gardens can be assessed using building footprints and NDVI-positive pixels while shared green spaces require another type of spatial indicator because their underlying structure cannot be recorded in building footprints. Furthermore, Lisbon

experiences summer heat stress and urban heat island effect, making roof greening relevant to heat mitigation, outdoor comfort, and public health considerations [27, 28].

The Lisbon dataset includes three major sources of spatial information [26]. The first is a 1.6 m-resolution multispectral eight-band WorldView-2 satellite image acquired on 18 February 2022 to map building and condominium green roofs across the city. The second source is a time series of 12 Sentinel-2 images taken in each month of 2023, which are used to monitor the greenness dynamics in the Gulbenkian case area throughout the year. Finally, there is a municipal building footprint layer with 54,800 buildings. The WorldView-2 image is used for a classification of water bodies, built environment, vegetation, and unsuitable roofs; the building footprint layer identifies the universe of individual roofs; and the Sentinel-2 time series provides the NDVI envelope. The NDVI threshold equals 0.30 and is calibrated based on 219 green-roof small polygons mapped in 2022 in municipal spaces such as underground parks, public buildings, and infrastructure. Such green roofs cover approximately 0.3% of the Lisbon city area and constitute the mean greenness value used as the candidate gate.

The classification check involved 100 validation samples. Confusion matrix is provided in Table 1. The diagonal includes 95 correct classification cases, which means that the overall classification accuracy is 0.950. Marginal structure of the table is equally important because not all classes play the same role in the green roof identification process. Vegetation class is the positive spectral evidence used for the acceptance of candidate areas. Unsuitable roofs, which are mostly red tile or roof material exclusions, provide the grounds for the exclusion of candidate buildings. The built environment class is the constructed context in which the vegetation signal should be assigned.

Table 1. WorldView-2 audit counts.

Reference class	Water body	Built environment	Vegetation	Unsuitable roofs	Row total
Water body	6	2	0	0	8
Built environment	0	56	0	0	56
Vegetation	0	2	24	0	26
Unsuitable roofs	0	1	0	9	10
Column total	6	61	24	9	100

The images of class context are analyzed with the class-specific criteria in the Results and discussion section, where the audit categories are tied to verification burden rather than shown as interchangeable labels.

2.2. Candidate-generation and validation record

Starting from the individual-building approach, 54,800 municipality-building polygons have been used initially. Building with the red-tiled roof area exceeding the median value, 64.71%, were omitted, thus reducing the building universe to 27,401 polygons. The next step included clipping of the NDVI layer to the polygons and converting this layer into points with the values equal to or larger than 0.30. Polygons without such points were excluded, thus leaving 8,656 polygons. As the WorldView-2 NDVI composition had a resolution of around 1.8 m and allowed capturing the tree canopy close to building edges, the buffer zone of 3.6 m was created for each polygon and the common areas of the buffered layer and the 8,656-polygon layer were extracted, resulting in 3,423 polygons. The final choice was to discriminate large buildings ($\geq 2,000 \text{ m}^2$) and apply either green-roof area or green-roof ratio criteria for extracting 176 individual-roof candidates. 50 individual-roof green roofs were validated.

The second method starts from a different assumption: the shared-space green roofs are typically linked with repeating building configurations, surrounded by podium or ground-level green surfaces. Therefore, the ordinary building polygons do not directly represent the underground or semi-underground structures. Falcão et al. built a Faster R-CNN object detector in ArcGIS Pro, which allows identifying two main building shapes A and B, with their variations A1, A2, B1, and B2 [26]. The dataset for training consisted of 150 samples. The augmentation included a 6° rotation and resulted in 116,824 features, including 82,324 class-A features and 34,500 class-B features. The detector included ResNet50 backbone and learning rate of 5.2481×10^{-5} . Average precision became 0.92 and 0.93 for classes A and B respectively. The detector produced 4,015 bounding geometries; geometries

below the confidence threshold of 0.87 were discarded. The remaining geometries were intersected with NDVI points and detections below the greenness gate of 0.30 were removed, leaving 195 shared-space candidates. 146 shared-space green roofs were validated.

These details are crucial, since they imply that the second route does not consist of the direct importing of an undifferentiated neural network output into the green-roof inventory. This approach implies using the hypothesis about the urban form: the repeated condominium geometries reveal the green surfaces, which cannot be found if the city is represented solely in terms of separate building footprints. In this case, the 0.87 detector confidence gate and 0.30 NDVI gate function as complementary filters: detector confidence provides morphological evidence of the candidate belonging to a repeated building configuration, while NDVI provides greenness evidence. Finally, validation shows whether the combination of morphological and greenness evidence corresponds to an actual shared green roof. Therefore, the analysis presented in the manuscript separates these evidence types in order to make the results interpretable as a verification problem rather than as a single classification label.

Therefore, the validation records include 371 candidate polygons: 176 individual-roof candidates and 195 shared-space candidates. 196 green roofs were validated, 25.5% of them being individual-roof cases, and 74.5% being shared-space cases. Lisbon parish distribution is consistent with the construction chronology: Parque das Nações, which experienced the major expansion in the 1990s, includes 64 shared-space green roofs, while there are no mapped green roofs in Beato and Penha de França parishes. In parishes, the correlation between the share of buildings constructed after the 1990s and the number of mapped green roofs is 0.85. Avenidas Novas underwent the exhaustive on-site validation, which revealed seven green roofs (two individual-roof and five shared-space), with no additional green roofs found.

The parameters provided in Table 2 provide the scope of the investigation. The analysis uses the Lisbon satellite dates, footprints counts, training counts, detector counts, candidate counts, validation counts, and NDVI seasonal range presented in the table. No other imagery, training data, candidate threshold, and validation sample are used. The restriction ensures that the interpretation is limited to 54,800 footprints for individual buildings and 4,015 detections for groups of buildings routes.

Table 2. Lisbon inputs.

Input group	Principal values	Analytical role
WorldView-2 classification	18 February 2022; eight bands; 100 validation samples; four classes; accuracy 0.950; $\kappa = 0.914$	Quantifies vegetation recognition and roof-material exclusion reliability.
Individual-building sequence	54,800 footprints; 27,401 after material exclusion; 8,656 after NDVI retention; 3,423 after inward buffering; 176 candidates; 50 validated	Quantifies contraction from municipal footprints to validated top-level roofs.
Grouped-building sequence	150 training samples; 116,824 augmented features; 4,015 detections; 195 candidates; 146 validated	Quantifies validation yield for shared green spaces associated with recurring condominium forms.
Seasonal NDVI record	Twelve monthly Sentinel-2 observations for 2023; top-level range 0.25–0.31; ground-level range 0.32–0.49	Compares the 0.30 greenness gate for top-level and ground-level roof conditions.

3. Verification Computations for Lisbon

3.1. Calculations of NDVI and class reliability

Morphospectral approach starts with NDVI calculation, which is expressed as

$$\text{NDVI} = \frac{\text{NIR} - R}{\text{NIR} + R}, \quad (1)$$

where R is reflectance of the red band, and NIR is reflectance of the near-infrared band. The Lisbon dataset uses 0.30 as the greenness threshold. In this research the threshold is regarded as a gate for candidates selection, not as proof of green-roof presence. The distinction is required as the verification of roofs' status depends on the

relationship between the vegetation signal and the urban object it belongs to. The same threshold can be used for highly credible roof, a less credible tree-contaminated roof edge, or even lowly credible green area above a garage.

The class reliability is calculated based on Table 1. For a class c , precision, recall, and F1 are

$$P_c = \frac{n_{cc}}{n_{\cdot c}}, \quad R_c = \frac{n_{cc}}{n_c}, \quad F_c = \frac{2P_c R_c}{P_c + R_c}, \quad (2)$$

where n_{cc} is the diagonal count, $n_{\cdot c}$ is the column total, and n_c is the row total. Overall accuracy is

$$A_o = \frac{\sum_c n_{cc}}{N}, \quad (3)$$

and Cohen's agreement coefficient is

$$\kappa = \frac{A_o - A_e}{1 - A_e}, \quad A_e = \frac{\sum_c n_{c\cdot} n_{\cdot c}}{N^2}. \quad (4)$$

The vegetation F1 value is retained as the greenness reliability term because vegetation is the positive class supporting candidate inclusion. The unsuitable-roof F1 value is retained for interpretation because material exclusion is a major operation in the individual-building route.

For each roof condition g , the point validation support is

$$Q_g = \frac{V_g}{C_g}, \quad (5)$$

where V_g is the number of validated green roofs and C_g is the number of final candidates. Verification odds are

$$O_g = \frac{Q_g}{1 - Q_g} = \frac{V_g}{C_g - V_g}, \quad (6)$$

and the relative odds of shared-space verification against individual-roof verification are

$$\Lambda_{ST} = \frac{O_S}{O_T}. \quad (7)$$

The reliability-adjusted morphospectral support for roof condition g is

$$M_g = Q_g F_{\text{veg}} \kappa. \quad (8)$$

This product is deliberately conservative. It is not a pixel-level probability. It combines the object validation rate with image-class reliability, preventing object-level success from being read independently of the classification audit.

3.2. Candidate contraction by roof condition

Sequential attrition is used to measure where each candidate-generation route contracts. For a route with stages $i = 0, \dots, k$, the retention ratio between two adjacent stages is

$$r_i = \frac{N_i}{N_{i-1}}, \quad (9)$$

and the attrition elasticity is

$$E_i = -\ln r_i. \quad (10)$$

This suggests that very few observations are lost in the corresponding stage. A large value shows that there is strong contraction. The logarithmic transformation makes sense since it ensures comparability of the individual-building and grouped-building paths despite the different initial sizes of the populations. It is also capable of determining if the uncertainty is solved right away by applying wide exclusions or if it is limited to the final small candidate pool.

3.3. Posterior validation probability

Validation proportions are not assumed to be known constants. Posterior probability of validity for each type of roof is estimated using the beta-binomial distribution. Assuming a half-unit beta prior, the posterior distribution of validation probability θ_g is

$$\theta_g \mid V_g, C_g \sim \text{Beta} \left(V_g + \frac{1}{2}, C_g - V_g + \frac{1}{2} \right). \quad (11)$$

The posterior mean equals

$$\bar{\theta}_g = \frac{V_g + \frac{1}{2}}{C_g + 1}, \quad (12)$$

and 95% credible intervals are obtained using beta quantiles 0.025 and 0.975. The same reliability adjustment applied to M_g confidence intervals is used here as well. Binomial interval estimation is provided because of the moderate number of candidates and the difference between 50/176 and 146/195 that cannot be evaluated based on point estimates alone [29, 30].

3.4. Seasonal NDVI margins and inspection burden

Seasonal NDVI margins measure the stability of the 0.30 threshold value. Margins for roof condition g are

$$U_g = N_{\max,g} - \tau, \quad L_g = N_{\min,g} - \tau, \quad (13)$$

where $\tau = 0.30$ is the greenness gate. The threshold-instability share is

$$H_g = \begin{cases} 0, & N_{\min,g} \geq \tau, \\ \tau - N_{\min,g}N_{\max,g} - N_{\min,g}, & N_{\min,g} < \tau < N_{\max,g}, \\ 1, & N_{\max,g} \leq \tau. \end{cases} \quad (14)$$

This measure shows how many percent of the seasonal envelope are located under the gate. The inspection burden calculation comes from the candidate shares, validated shares, and non-validated shares. A roof condition that gives rise to many validated objects, but very few non-validated candidates, may be considered a rich confirmation stream. A roof condition that gives rise to few validated objects, but many non-validated candidates can be considered a targeted confirmation stream.

4. Results and Discussion

4.1. Reliability of image-class classification and object assignment

The classification audit shows that a robust image base exists for the Lisbon inventory. The diagonal sum in Table 1 is 95 out of 100, thus providing an accuracy of 0.950. The expected agreement using the marginal sums of the rows and columns is 0.4178, providing κ of 0.914. Those figures clearly demonstrate the reliability of the WorldView-2 classification as an image classification task; however, they cannot prove on their own that the object assigned to NDVI-positive class is a confirmed green roof. The problem of green roofs mapping is a dual one: the image has to identify vegetation and the roof material, and after this the object assignment process has to determine whether the vegetation belongs to the given roof condition.

Class-specific accuracies from Table 3 clarify how it is possible to have high image accuracies alongside an imbalance in the validation results. Precisions of vegetation are 1.000 and F1 of vegetation is 0.960, so the green component in and of itself is not the weak component in the Lisbon classification. Precision of unsuitable roof is also 1.000, thus confirming that the exclusion of unsuitable materials is justified by image components. It is unclear, however, whether each green vegetation area corresponds to a correct roof object. This issue is in line with the broader discussions on accuracy assessment in remote sensing, where aggregate accuracy values are useful, but not sufficient [31].

Table 3. Classification reliability.

Class	Precision	Recall	F1	Interpretation for verification burden
Water body	1.000	0.750	0.857	Contextual class with limited direct effect on roof decisions.
Built environment	0.918	1.000	0.957	Establishes the constructed surface background against which vegetation is assigned.
Vegetation	1.000	0.923	0.960	Supplies the positive greenness evidence used by the NDVI gate.
Unsuitable roofs	1.000	0.900	0.947	Supports removal of red-tile and other material candidates unlikely to be green roofs.

The four classes in Figure 2 represent the use of the audit categories. Water and built environment characterize the urban landscape, vegetation serves as the inclusion signal, and roofs provide the exclusion signal. A process that treats these classes as a decision chain is more helpful than one that views them as a thematic map. For Lisbon, the main interpretive risk is not the absence of vegetation on the image. It is the incorrect attachment of vegetation to the built object.

This result shapes the direction of the analysis. The individual-building approach relies upon the presence of NDVI-positive points inside the building footprints, hence, continues to rely on the same set of errors — tree crown, footprint edge, misaligned roof boundary, and vegetated areas near building perimeter. The grouped-building approach is different in that it uses residential block geometry to find shared green surfaces which may not be represented by the standard roof polygon geometry. As the two approaches solve different problems, the validation results should be compared separately from the start.



Figure 2. WorldView-2 class contexts.

4.2. Candidate attrition and route-specific information loss

Individual-building approach is highly selective. After the roof material exclusion applied to a total of 54,800 municipal footprints, 27,401 buildings are left in the sample. Further reduction to 8,656 buildings comes after

NDVI gate is applied, then there are 3,423 buildings left after the inward buffer, and finally, 176 buildings remain due to the candidate rules. After validation, there are only 50 valid candidates left. Thus, the whole chain of the selection steps transforms the municipal footprint layer to a small list for further verification.

From Table 4, one can observe that there is the most serious attrition at the step of final candidate selection when only 5.142% of the buffered NDVI-positive buildings remain. In other words, the attrition elasticity of this step equals 2.968, which is considerably higher than the material exclusion and buffering steps. Hence, the individual-building route does not fail earlier on, but carries the object ambiguity through some filters and concentrates it into a small number of candidates for direct validation.

Table 4. Individual-route attrition.

Stage	Count	Share of starting footprints	Share from previous stage	$-\ln r_i$
Municipal building footprints	54,800	100.000%	—	—
After roof-material exclusion	27,401	50.002%	50.002%	0.693
After NDVI-point gate	8,656	15.796%	31.590%	1.153
After inward-buffer gate	3,423	6.246%	39.544%	0.928
Final individual-roof candidates	176	0.321%	5.142%	2.968
Validated individual green roofs	50	0.091%	28.409%	1.258

From Table 4, one can observe that there is the most serious attrition at the step of final candidate selection when only 5.142% of the buffered NDVI-positive buildings remain. In other words, the attrition elasticity of this step equals 2.968, which is considerably higher than the material exclusion and buffering steps. Hence, the individual-building route does not fail earlier on, but carries the object ambiguity through some filters and concentrates it into a small number of candidates for direct validation.

As can be seen from the contraction pattern shown in Figure 3, the significance of the numbers can be made clear by their inspection. However, big losses prior to the validation step cannot be taken as the sign that the last step produces a credible result. Although the final candidate set is small, out of 176 candidates 126 are not validated. The possible reasons include residual edge vegetation, trees overhanging the building boundary, heterogeneity of the roof surface, displacement of footprints, and small green spots that satisfy the spectral gate but do not act as a green roof. Thus, in municipal studies, the considered route may serve as an error diagnosis flow that shows where the footprints–vegetation assignment requires verification.

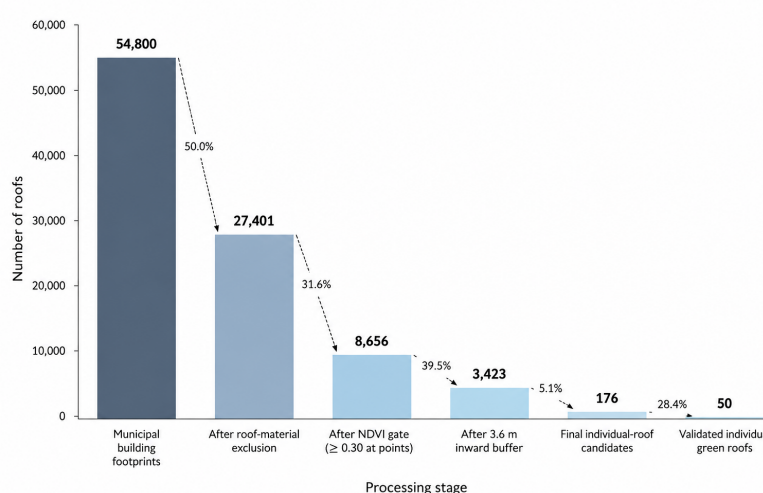


Figure 3. Individual-roof contraction.

The grouped-building approach has a different pattern of the candidate pool reduction. First of all, there were 4,015 grouped-building objects detected, and then 195 remained after applying the detector confidence and NDVI filters.

Out of them, 146 were successfully validated as green roofs. As a result, the first contraction is very high, but the attrition at the validation step is low.

From Table 5, one can note that the grouped route values show a more intense signal prior to validation. The 4.857% retention ratio of the detections to candidates ratio is close to the individual roof shrinkage ratio, but the validation step keeps 74.872% of the candidates from the shared space. In other words, the grouped route leaves its uncertainty ratio mainly prior to candidate extraction, whereas the individual route maintains it mostly during validation. Such an aspect is important to consider when planning inspections.

Table 5. Grouped-route attrition.

Stage	Count	Share of starting geometries	Share from previous stage	$-\ln r_i$
Grouped-building detections	4,015	100.000%	—	—
Shared-space candidates	195	4.857%	4.857%	3.025
Validated shared-space green roofs	146	3.636%	74.872%	0.289

The sequence of visuals provided in Figure 4 demonstrates why morphology plays such a crucial role. Shared green roofs can usually be found on areas that cannot be covered with the built structure of a building using one ordinary footprint. Therefore, a detector that recognizes typical configurations of residential blocks will bring us information that is not available for individual-building route. The NDVI gate adds greenness evidence after that, and validation process makes sure that spatial and spectral evidence are aligned with the actual presence of roofed green surfaces. This interpretation pattern fits perfectly into object-based remote sensing approach, where analysis unit and its context play an important role in interpretation [32, 33].

The two attrition patterns provide the answer to our research question for the first time. The route that has less false candidates does not start from less objects, but rather from the combination of morphology and greenness recognition prior to validation. Lisbon's shared space candidates should get priority in terms of inventory completion, while individual roof candidates need more attention to checking trees overlap, alignment of the footprints and roof edge contamination.



Figure 4. Grouped-building recognition.

4.3. Posterior credibility and validation asymmetry

Validation counts make the typological distinction clear. Individual roofs generate 176 candidates and 50 validations with validation rate of 0.284. Shared space roofs generate 195 candidates and 146 validations with validation rate of 0.749. Relative verification-odds ratio is calculated as the ratio between validation probabilities: 50/126 for individual roofs and 146/49 for shared spaces. Relative verification-odds ratio equals 7.51, and it is not a marginal result.

This example demonstrates how two equal numbers of candidates could end up having drastically different inspection results. It is clear from the bars shown in Figure 5 that the individual roof bar is made up mostly of non-validated

items, whereas the shared space bar consists mainly of validated items. Clearly, this means that one must avoid an error of assigning visits in proportion to the number of candidates.

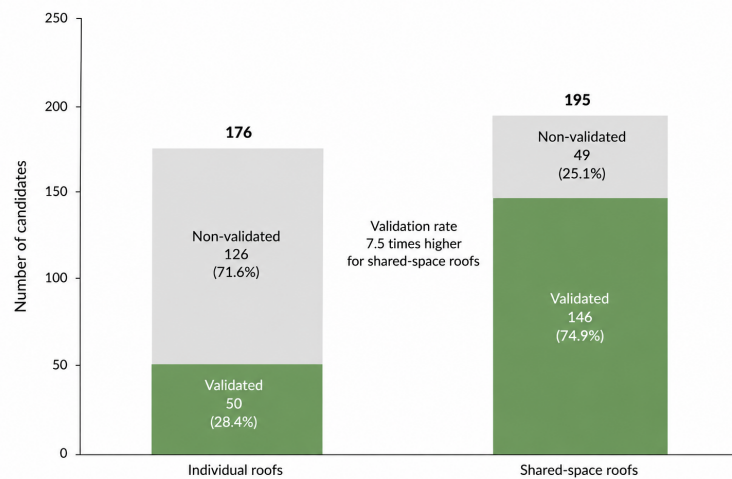


Figure 5. Validation by roof condition.

Table 6. Posterior validation probability.

Candidate group	Candidates	Validated	Posterior mean	95% credible interval	Reliability-adjusted interval	Support M_g
Individual roofs	176	50	0.285	0.221–0.354	0.194–0.310	0.249
Shared-space roofs	195	146	0.747	0.684–0.806	0.601–0.707	0.657
Combined candidates	371	196	0.528	0.477–0.579	0.419–0.508	0.464

Thus, the posterior intervals in Figure 6 provide a concise uncertainty statement about inspection planning. One can expect that the field team will receive a large number of non-confirmations in case of individual roof candidates even with the image reliability accounted for. In contrast, one can expect confirmations from the shared-space candidates at a considerably higher rate. The statistical statement reinforces the spatial statement: the shared-space candidates represent a high-yield inventory completion stream, while individual roofs represent a necessary but low-yield confirmation stream.

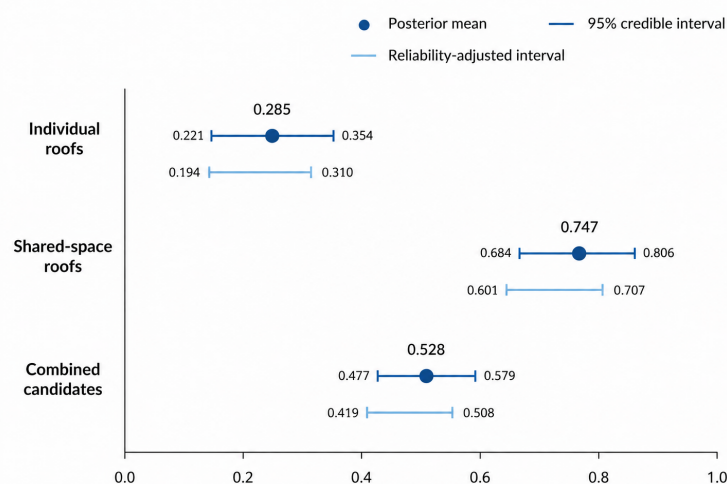


Figure 6. Reliability-corrected credibility.

The credibility separation presented above also provides protection from overstatement of the image-classification values. The classification audit performed on the WorldView-2 images is quite good. But the posterior calculation

performs the validation at the object level where municipal decision-making occurs. It is important for remote sensing inventories because an image-class label can be assigned to the wrong planning object even though the label itself is correct.

4.4. NDVI Seasonality and Threshold Stability

The NDVI of 0.30 reacts differently when applied to both roof types. Top level roofs span between 0.25 and 0.31 within the annual envelope of Sentinel-2 and pass through the gate but exceed it by a little bit at the highest point. The ground level roofs of the shared space span between 0.32 and 0.49 and remain above the gate throughout the year. The same threshold therefore provides weaker seasonal support for top-level roofs and stronger seasonal support for shared-space roofs.

The margins presented in Table 7 explain the rationale for regarding 0.30 as a candidate-forming rule rather than a universal proof rule. A top-level roof approaching the threshold may fall below it as a consequence of drying out, changing substrate moisture, or changing imaging conditions. A shared green surface with a margin of 0.02 at the bottom and 0.19 at the top is relatively safe inside the greenness range. This pattern is compatible with findings that green-roof signals depend on substrate depth, irrigation, and plant condition in Mediterranean climates [34–36].

Table 7. NDVI threshold margins.

Roof condition	Min.	Max.	Lower	Upper	Interpretation
Top-level roofs	0.25	0.31	-0.05	0.01	Crosses the 0.30 gate; threshold support is seasonally fragile.
Shared-space roofs	0.32	0.49	0.02	0.19	Remains above the 0.30 gate; threshold support is seasonally stable.

The Figure 7 provides visual support for the aforementioned posterior difference. Shared-space candidates are not only more likely to become valid but also have a greenness level remaining higher than the threshold. Candidates from the top-level space approach the threshold very closely, and hence one date may determine their fragile candidate status. The threshold is not invalid. Its evidence differs under varying roof conditions.

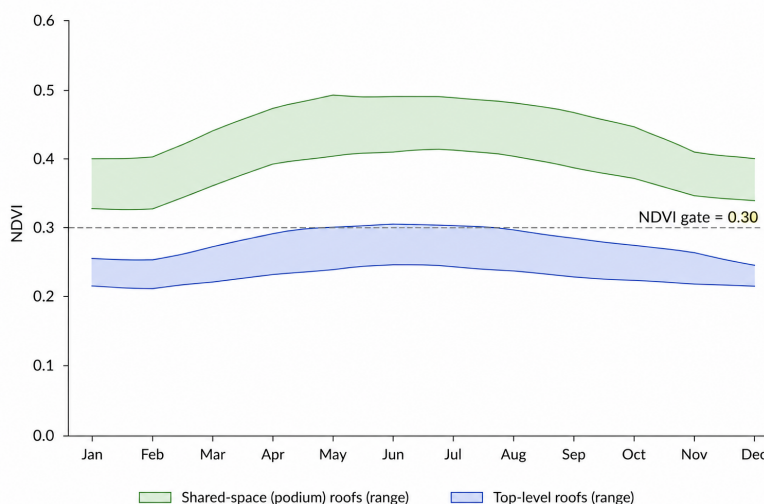


Figure 7. Seasonal NDVI envelope.

It follows that the seasonal behaviour informs the timing of field work. While individual roof validation requires attention to borderline surfaces, dry season vegetation stress, and possible difference between green patches and roof systems, shared-space inspection is focused on structure classification, owner identification, accessibility, and enrichment of attributes due to a lower greenness margin. Thus, the annual Sentinel-2 dataset brings practical value without changing the candidate-generating rules.

4.5. Workload of inspection and city’s interpretation

The final list includes 371 candidates and 196 validations. Individual roofs account for 47.4% of the number of candidates and only 25.5% of validated green roofs, whereas shared space roofs account for 52.6% of the candidates and 74.5% of the validated green roofs. On the other hand, non-validated candidates are distributed in an inverted way: individual roofs account for 72.0% of the number of non-validation, while shared space roofs – for 28.0%. This distribution is the practical core of the paper because it translates validation probability into expected field workload.

The inspection rates shown in Table 8 provide a clear management implication. A visit to a shared-space candidate will yield, on average, 75 confirmations out of every 100 inspections, while a visit to an individual-roof candidate will yield only 29 confirmations out of every 100 inspections. The individual approach is also necessary, but its primary role is not the rapid expansion of the inventory. Its primary role is the identification of the areas where the automated roof attribution process is susceptible to errors.

Table 8. Inspection burden.

Roof condition	Candidates	Validated	Non-validated	Candidate share	Validated share	Non-validated share
Individual roofs	176	50	126	47.4%	25.5%	72.0%
Shared-space roofs	195	146	49	52.6%	74.5%	28.0%
Total	371	196	175	100.0%	100.0%	100.0%

Figure 8 provides a visualisation of the same burden in terms of a countable field. The validated shared-space category is represented by a large block, and the non-validated individual-roof category is also represented by a large block. This pair of blocks gives more useful information than the overall 52.8% of the validation rate. One number would have concealed the fact that one type of candidates is responsible for the building of the confirmed register, and the other type accounts for the false-candidate workload.

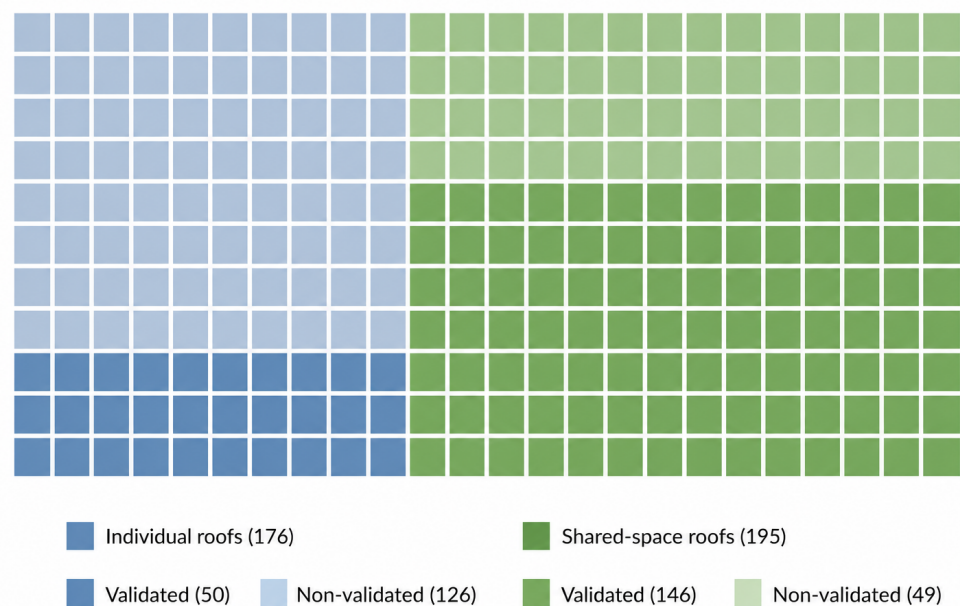


Figure 8. Inspection-burden mosaic.

Policy considerations come from this imbalance. The shared-space candidates should be considered a high-yield inventory-completion category, fit for confirmation of area, ownership context, roofed-structure status, access, and maintenance condition. Individual candidates should be treated as a targeted-confirmation category, with the forms

of the inspection recording edge trees, roof slope, roof material, footprint alignment, and whether the vegetated surface represents a true roof installation. The approach is consistent with the green-roof policy literature that focuses on barriers, maintenance requirements, and information needs [37–39].

4.6. Verification classes for planning use

The findings suggest two major categories of verification and one minor diagnostic category. The high-yield category includes shared-space candidates with high validation probability, credible adjusted posterior, and stable seasonal NDVI. The targeted-confirmation category includes individual-roof candidates with lower validation probability and higher threshold sensitivity. The diagnostic category consists primarily of non-validated individual-roof candidates showing the reasons why edge vegetation, footprint geometry, and material screening still produce false candidates.

Table 9 includes those classes that address the planning aspect of the research question. The idea is not to eliminate the low-yield stream. Individual candidates are essential since they help identify weaknesses of the footprint-based strategy. Shared-space candidates are essential since they hold the majority of the confirmed inventory. A credible city register should take into account both streams, give them different inspection forms, and not treat them as equivalent candidate polygons.

Table 9. Verification classes.

Class	Evidence basis	Planning use
High-yield shared-space review	146/195 validated; posterior mean 0.747; adjusted support 0.657; seasonal NDVI above 0.30	Complete the municipal register and enrich attributes for confirmed shared green surfaces.
Targeted individual-roof confirmation	50/176 validated; posterior mean 0.285; adjusted support 0.249; seasonal NDVI crosses 0.30	Inspect roof attribution carefully and separate real top-level installations from edge or material artefacts.
False-candidate diagnosis	Individual roofs form 72.0% of non-validated candidates	Record why candidates fail so that later mapping cycles can improve footprint, buffer, and material rules.

Verification priority	Primary roof type	Key evidence	Reliability support	Action focus
High-yield shared-space review		146 of 195 validated - Posterior mean 0.747 - NDVI consistently above 0.30	Support 0.657	Prioritize shared-space candidates for field verification
Targeted individual-roof confirmation		50 of 176 validated - Posterior mean 0.285 - NDVI near 0.30 threshold	Support 0.249	Apply targeted checks for individual roofs
False-positive diagnosis & sampling		126 of 176 not validated 72% of non-validated - NDVI and edge effects likely	Support 0.464	Investigate false positives and refine filters

Overall: 371 candidates, 196 validated (52.8%); shared-space support = 2.64× individual support.

Figure 9. Priorities of verification.

The priority register in Figure 9 states the ultimate operational interpretation. The review of shared-space candidates is the first path to extending the confirmed register. The confirmation of individual-roof candidates is the path for solving difficult object-attribution issues. The diagnosis of false candidates is the path for making future inspections

and object-attribution protocols more reliable. It is especially true for cities that have both older and newer buildings. In such cities, red-tile roofs, yards, trees, and podiums can create vegetation signals around buildings.

A candidate-based municipal database would strengthen the interpretation during future cycles. Each polygon should preserve its material exclusion ratio, number of NDVI points, buffered NDVI points, roof area, greening ratio, detection confidence, min/max NDVI, validation status, parish, and construction period context. With the help of those features, the same validation process could be stratified by building age, neighbourhood, roof area, confidence score, and threshold distance. The aggregate Lisbon values already show the primary asymmetry; a more detailed candidate table would uncover the mechanism behind it.

5. Conclusions

The paper asked which Lisbon green-roof candidate stream was more valuable for verification when the accuracy of the image-class classification was high but roof-object assignment depended on morphology. The answer is that shared-space candidates were more valuable for inventory extension, whereas individual-roof candidates were essential for confirmation and error diagnosis. This conclusion follows from the Lisbon numbers and not from any other procedure or dataset. The audit of the WorldView-2 classification is excellent, with the accuracy of 0.950, $\kappa = 0.914$, vegetation F1 of 0.960, and unsuitable roof F1 of 0.947. The weakness, therefore, was not the recognition of vegetation as a class. The weakness was the assignment of vegetation to the right roof condition.

The individual-building strategy reduces 54,800 municipal footprints to 176 candidates and 50 confirmed top-level green roofs. Its validation rate is 0.284, and the reliability-adjusted posterior interval is 0.194–0.310. The strategy is very selective, but it keeps many false candidates since it is still hard to separate footprint edges, roof materials, tree crowns, and small vegetated patches even after the filtering. The grouped-building strategy reduces 4,015 detections to 195 candidates and 146 confirmed shared-space roofs. Its validation rate is 0.749, and the reliability-adjusted posterior interval is 0.601–0.707. Those two intervals do not overlap, and the shared-space strategy has 7.51 times higher verification odds than the individual-roof strategy.

Seasonal NDVI shows that the two roof conditions did not provide the same threshold support. The range of the top-level roofs is from 0.25 to 0.31 and it crosses the 0.30 gate, so the candidate status of top-level roofs is seasonally unstable. The range of the shared-space roofs is from 0.32 to 0.49 and it does not cross the gate, so the greenness evidence of the shared-space roofs is more stable. The inspection burden also points in the same direction: individual roofs make up 72.0% of non-validated candidates, whereas shared-space roofs make up 74.5% of validated green roofs. A pooled candidate rate would conceal this operational asymmetry.

The practical implication for Lisbon is, thus, specific. Shared-space candidates should be inspected first if the aim is the expansion and enrichment of the confirmed municipal green-roof register. Individual-roof candidates should be inspected with a different aim: the diagnosis of false positives, the confirmation of borderline top-level installations, and the improvement of the roof-edge vegetation, red-tile material, and footprint mismatch handling in future mapping cycles.

References

- [1] Berardi, U., GhaffarianHoseini, A., & GhaffarianHoseini, A. (2014). State-of-the-art analysis of the environmental benefits of green roofs. *Applied energy*, 115, 411–428.
- [2] Shafique, M., Kim, R., & Rafiq, M. (2018). Green roof benefits, opportunities and challenges—A review. *Renewable and Sustainable Energy Reviews*, 90, 757–773.
- [3] Manso, M., Teotónio, I., Silva, C. M., & Cruz, C. O. (2021). Green roof and green wall benefits and costs: A review of the quantitative evidence. *Renewable and Sustainable Energy Reviews*, 135, 110111.
- [4] Vijayaraghavan, K. (2016). Green roofs: A critical review on the role of components, benefits, limitations and trends. *Renewable and sustainable energy reviews*, 57, 740–752.

- [5] Fioretti, R., Palla, A., Lanza, L. G., & Principi, P. (2010). Green roof energy and water related performance in the Mediterranean climate. *Building and environment*, 45(8), 1890-1904.
- [6] Nardini, A., Andri, S., & Crasso, M. (2012). Influence of substrate depth and vegetation type on temperature and water runoff mitigation by extensive green roofs: shrubs versus herbaceous plants. *Urban Ecosystems*, 15(3), 697-708.
- [7] Joshi, M. Y., & Teller, J. (2021). Urban integration of green roofs: Current challenges and perspectives. *Sustainability*, 13(22), 12378.
- [8] Dong, J., Guo, F., Lin, M., Zhang, H., & Zhu, P. (2022). Optimization of green infrastructure networks based on potential green roof integration in a high-density urban area—A case study of Beijing, China. *Science of The Total Environment*, 834, 155307.
- [9] Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote sensing of environment*, 25(3), 295-309.
- [10] Addabbo, P., Focareta, M., Marcuccio, S., Votto, C., & Ullo, S. L. (2016). Contribution of Sentinel-2 data for applications in vegetation monitoring. *Acta Imeko*, 5(2), 44-54.
- [11] Dobrinić, D., Gašparović, M., & Medak, D. (2021). Sentinel-1 and 2 time-series for vegetation mapping using random forest classification: A case study of Northern Croatia. *Remote Sensing*, 13(12), 2321.
- [12] Eisfelder, C., Asam, S., Hirner, A., Reiners, P., Holzwarth, S., Bachmann, M., ... & Kuenzer, C. (2023). Seasonal vegetation trends for Europe over 30 years from a novel normalised difference vegetation index (NDVI) time-series—The TIMELINE NDVI product. *Remote Sensing*, 15(14), 3616.
- [13] Pontius Jr, R. G., & Millones, M. (2011). Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International journal of remote sensing*, 32(15), 4407-4429.
- [14] Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., & Wulder, M. A. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote sensing of Environment*, 148, 42-57.
- [15] Wu, A. N., & Biljecki, F. (2021). Roofpedia: Automatic mapping of green and solar roofs for an open roofscape registry and evaluation of urban sustainability. *Landscape and Urban Planning*, 214, 104167.
- [16] Simpson, C. H., Brousse, O., Mohajeri, N., Davies, M., & Heaviside, C. (2023). An open-source automatic survey of green roofs in London using segmentation of aerial imagery. *Earth System Science Data*, 15(4), 1521-1541.
- [17] Santos, T., Tenedório, J. A., & Gonçalves, J. A. (2016). Quantifying the city's green area potential gain using remote sensing data. *Sustainability*, 8(12), 1247.
- [18] Gandini, A., Navarro, D., & Feliu, E. (2023, May). Identification and mapping of areas and buildings with high roof greening potential. In *2023 Joint Urban Remote Sensing Event (JURSE)* (pp. 1-4). IEEE.
- [19] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28.
- [20] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- [21] Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS journal of photogrammetry and remote sensing*, 152, 166-177.
- [22] Bai, T., Pang, Y., Wang, J., Han, K., Luo, J., Wang, H., ... & Zhang, H. (2020). An optimized faster R-CNN method based on DRNet and RoI align for building detection in remote sensing images. *Remote Sensing*, 12(5), 762.

- [23] Li, M., Zhang, Z., Lei, L., Wang, X., & Guo, X. (2020). Agricultural greenhouses detection in high-resolution satellite images based on convolutional neural networks: Comparison of faster R-CNN, YOLO v3 and SSD. *Sensors*, 20(17), 4938.
- [24] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of big data*, 6(1), 1-48.
- [25] Marcos, L. (2024). Detection of Green Roofs through Satellite Imagery Using Faster R-CNN (Doctoral dissertation, Toronto Metropolitan University).
- [26] Falcão, A. P., Pernes, J., Miranda, V., & Silva, C. M. (2025). Mapping Green Roofs on Buildings Using Vegetation Indices and Deep Learning Techniques. *Remote Sensing*, 17(21), 3657.
- [27] Santamouris, M. (2014). Cooling the cities—a review of reflective and green roof mitigation technologies to fight heat island and improve comfort in urban environments. *Solar energy*, 103, 682-703.
- [28] Rodrigues, M., Santana, P., & Rocha, A. (2021). Modelling of temperature-attributable mortality among the elderly in Lisbon metropolitan area, Portugal: a contribution to local strategy for effective prevention plans. *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, 98(4), 516.
- [29] Agresti, A., & Coull, B. A. (1998). Approximate is better than “exact” for interval estimation of binomial proportions. *The American Statistician*, 52(2), 119-126.
- [30] Brown, L. D., Cai, T. T., & DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical science*, 16(2), 101-133.
- [31] Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote sensing of environment*, 80(1), 185-201.
- [32] Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS journal of photogrammetry and remote sensing*, 65(1), 2-16.
- [33] Myint, S. W., Gober, P., Brazel, A., Grossman-Clarke, S., & Weng, Q. (2011). Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote sensing of environment*, 115(5), 1145-1161.
- [34] Silva, J., Paço, T. A., Sousa, V., & Silva, C. M. (2021). Hydrological performance of green roofs in mediterranean climates: A review and evaluation of patterns. *Water*, 13(18), 2600.
- [35] Cascone, S. (2019). Green roof design: State of the art on technology and materials. *Sustainability*, 11(11), 3020.
- [36] Jamei, E., Chau, H. W., Seyedmahmoudian, M., Mekhilef, S., & Hafez, F. S. (2023). Green roof and energy—role of climate and design elements in hot and temperate climates. *Heliyon*, 9(5).
- [37] Liberalesso, T., Cruz, C. O., Silva, C. M., & Manso, M. (2020). Green infrastructure and public policies: An international review of green roofs and green walls incentives. *Land use policy*, 96, 104693.
- [38] Zhang, G., & He, B. J. (2021). Towards green roof implementation: Drivers, motivations, barriers and recommendations. *Urban forestry & urban greening*, 58, 126992.
- [39] Nguyen, C. N., Muttill, N., Tariq, M. A. U. R., & Ng, A. W. (2021). Quantifying the benefits and ecosystem services provided by green roofs—A review. *Water*, 14(1), 68.