Evaluating a modified StarDist Implementation for Individual Tree Detection and Crown Delineation in heterogeneous Landscapes

Nora Tilly [1], Andreas Tilly [2], Georg Bareth [1]

[1] GIS & Remote Sensing Group, Institute of Geography, University of Cologne, Germany – nora.tilly@uni-koeln.de; g.bareth@uni-koeln.de
[2] Independent Researcher, Cologne, Germany

Keywords: Dehesa, Oak trees, Aerial orthophotos, Deep Learning, Segmentation, ITDCD

Abstract

Individual tree detection and crown delineation (ITDCD) in dehesa landscapes is complicated by geometric distortions from steep terrain, varying tree densities, and the partly multi-crown 'broccoli-like' structure of holm and cork oaks. This study evaluates the usability of a modified StarDist deep learning model, which has recently shown effectiveness for ITDCD in Canadian forests. Moreover, this study develops a workflow transforming the original StarDist, designed for microscopy images, into an ITDCD solution, taking the georeferencing of geospatial data into account. The tile-wise organized ground truth dataset is created with the pretrained *Tree Segmentation* model available in the *ArcGIS Living Atlas*, combined with manual revision. Several augmentation methods are applied, resulting in 960 images, which are split into 85 % for training and 15 % for validation. Following the approach of the Canadian forest study, the StarDist implementation is modified by introducing a constraint to the probability loss function. Rather than computing loss across all pixels, the modified loss function considers only pixels explicitly annotated as objects, while background pixels are excluded. An additional dataset of 1,200 trees serves as ground truth for testing the prediction across the entire study area. Using an Intersection over Union of 0.5, this test demonstrates good performance (Accuracy: 87.50 %; F1-score: 0.85). The accuracy varies with tree density: in areas with sparse tree cover, nearly all tree crowns are detected; in moderately dense areas, a number of tree crowns are missed; whereas in very dense tree layers, the frequency of missed detections increases.

1. Introduction

Oak trees are fundamental components of the agro-silvopastoral systems in the south-western Iberian Peninsula, known as dehesas in Spain and montados in Portugal. Holm oaks (*Quercus ilex*) and cork oaks (*Quercus suber*) dominate the sparse tree layer, which integrates extensive grazing and arable farming (Plieninger and Wilbrand, 2014). Beyond providing shade for crops, pastures, and herbs, these trees yield key socio-economic benefits, e.g., acorns serve as fodder for Iberian pigs, while cork oaks supply the cork industry. Thus, the trees are essential here, as they have positive effects on entire ecosystems (Moreno Marcos et al., 2007). However, these prime examples of a sustainable balance between land use and conservation are threatened due to mismanagement, disease, and climate change (Camilo-Alves et al., 2013; López-Sánchez et al., 2017).

While spectral remote sensing is frequently used for assessing tree vitality and oak decline on a regional scale (Godinho et al., 2018; Navarro et al., 2019), individual tree segmentation in remote sensing imagery remains challenging. As noted by Tilly et al. (2020) for the study area investigated here, the strong geometric distortions due to the steep terrain represent a key obstacle. Moreover, the highly variable tree densities, and the characteristic 'broccoli-like' structure of oaks, with some trees having multiple crowns, further complicate the task. Identifying and delineating individual trees is, however, essential to extract spectral information from single crowns for tree-level analysis.

Reviews by Zhao et al. (2023) and Zheng et al. (2025), which provide comprehensive overviews of current approaches and algorithms, reveal growing research interest on individual tree detection and crown delineation (ITDCD) over the past decade. As in many remote sensing applications, the rise of deep learning-based methods has been particularly influential. While Zheng et al. (2025) reported only two deep learning-based ITDCD studies published in 2017, this number increased to approximately 30 publications annually in 2022 and 2023. Recently, Tong and Zhang (2025) demonstrated that a modified StarDist deep learning model is well-suited for ITDCD in Canadian forests. This study investigates whether it is also useful for delineating crowns of oak trees in a heterogeneous dehesa landscape. Moreover, this study develops a workflow for model training and testing, which takes the georeferencing information of geospatial data into account.

2. Methods

2.1 Study Area

The *Dehesa San Francisco*, managed by the Fundación Monte Mediterráneo (www.fundacionmontemediterraneo.com) served as study site (Figure 1). Located in the *Sierra Morena*, one of southern Spain's principal mountain systems, the dehesa lies close to the village of Santa Olalla del Cala, approximately 60 km north of Seville in the autonomous region of Andalusia. The territory of the dehesa covers about 500 ha with an altitude of 350 to 500 m above sea level. The bedrock is mostly schist with a shallow soil layer (≈ 25 cm), mainly Leptosols, rarely interrupted by Cambisols (Sapp et al., 2019).

A continental Mediterranean climate prevails in Andalusia, characterized by hot and dry summers, mild winters, and rainfall concentrated from autumn to spring. The geomorphological and climatic conditions create a heterogeneous landscape and vegetation structure. While sunny, higher south-facing slopes show a sparse tree layer with about 10 trees/ha or complete lack of vegetation, the density on lower south-facing slopes and ridges increases to 30 to 45 trees/ha. In contrast, the shadier north-facing slopes host the highest densities of 70 to 85 trees/ha.

https://doi.org/10.5194/isprs-archives-XLIX-B3-2026-1139-2026 |

2.2 Remote Sensing Data & Ground truth

The freely available digital RGB orthophotos (CNIG 2022a) from the National Aerial Orthophotography Plan were used. A composite raster dataset was generated by adding the NIR band of the false-color image (CNIG 2022b) as a fourth layer, since near-infrared spectral information is known to be well suited for the remote sensing of vegetation. All four bands, with a spatial resolution of 0.25 m, were captured with one multispectral camera but were provided in two products by the Spanish authority. The datasets were provided as 8-bit unsigned raster data sets, corresponding to a radiometric resolution of 256 discrete grey values.

Studies on ITDCD commonly use manually annotated tree crowns as ground truth. A different strategy was applied in this study, because manual annotating objects is both time-consuming and error-prone. The pretrained *Tree Segmentation* model based on the *DeepForest* Python package available in the *ArcGIS Living Atlas* (Esri, 2025) was initially applied to generate a vector dataset of tree crowns in ArcGIS Pro 3.5. In four areas (tiles for training and validation in Figure 1), the results of the *Tree Segmentation* model were revised, and missing tree crowns were manually annotated. This process yielded a complete ground truth dataset comprising 948 trees organized into 64 tiles. Reflecting the principle that ground truth quality outweighs quantity (Roscher et al., 2024) this complete dataset was particularly valuable for validation purposes.

Additionally, 1,200 tree crowns from the initial vector dataset, located outside these 64 tiles, were reviewed, revised if necessary, and used as ground truth for testing the model prediction with unseen data. These tree crowns were selected across the entire dehesa to evaluate the model performance under varying site conditions with different tree densities.

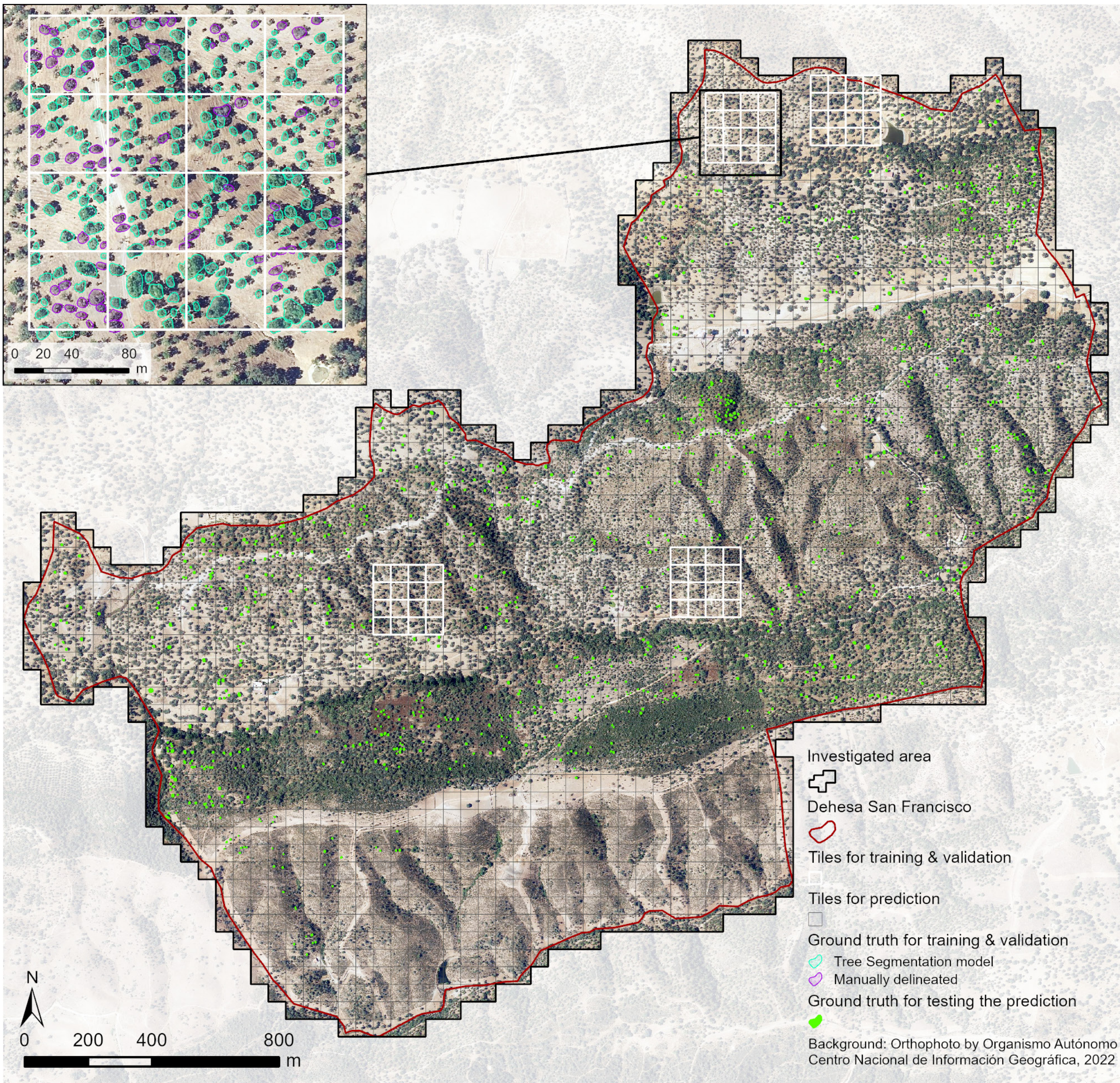

Figure 1. Ground truth used for training, validation, and testing the modified StarDist deep learning model at the *Dehesa San Francisco*, Andalusia, Spain.

2.3 StarDist

The StarDist deep learning model was introduced by Schmidt et al. (2018) for cell and nuclei segmentation in microscopy images. It represents objects as star-convex polygons, where a straight line from an interior point (kernel point) to any boundary point lies entirely within the shape. The training pipeline consists of three steps: prediction, filtering, and evaluation.

2.3.1 Object Prediction: StarDist employs a U-Net architecture to process input images of 256 × 256 pixels, producing two predictions for every pixel: a probability indicating whether that pixel serves as a kernel point, and a set of radial distances describing the object's boundary relative to that kernel point. In the encoding phase, the input image is downsampled to a resolution of 32 × 32 pixels through repeated blocks of 3 × 3 convolutional layers followed by max-pooling operations. Convolutions act as learnable local filters, extracting increasingly abstract features: early layers capture low-level details like edges and textures, while deeper stages integrate these into broader contextual understanding, such as potential object structures. Max-pooling downsamples by retaining the strongest activation in each local window, enabling each feature to consider larger regions of the original image. The decoding phase then reconstructs high-resolution representations by upsampling these compressed features back to 256 × 256 and refining them with additional convolutions for local consistency. Skip connections play a crucial role in the U-Net architecture, as high-resolution feature maps from encoder levels are directly concatenated to the upsampled decoder features at matching scales. This fusion preserves fine spatial details lost during downsampling while retaining the rich semantic context from deeper layers, ensuring that the final maps support pixel-accurate predictions. The reconstructed 256 × 256 feature maps undergo another 3 × 3 convolutional layer with 128 channels. This step reduces feature contention, ensuring shared features effectively serve both kernel point detection and boundary prediction, thereby providing a solid base for the two output heads. The first head generates the object probability map via a 1 × 1 convolution followed by sigmoid activation, producing a single-channel output where each pixel's value (0 to 1) quantifies its likelihood as kernel point. The second head creates the radial distance map through another 1 × 1 convolution with linear activation, outputting one channel per predefined ray direction evenly spaced around 360 degrees. Connecting these points in angular order reconstructs the star-convex polygon.

2.3.2 Filtering: The U-Net prediction results in a probability value and a number of radial distances (32 with the default settings) for each pixel of the input image. As the number of potential objects obviously exceeds the number of actually existing objects, two filtering stages are applied. First, assuming that object probabilities are highest for pixels close to kernel points, all pixels with a prediction probability below 0.50 are discarded. Second, non-maximum suppression (NMS) eliminates overlapping polygons by identifying groups of candidates representing the same object based on their Intersection over Union (IoU), retaining only the candidate with the highest prediction probability score per group. During the training, the probability and NMS thresholds are optimized based on the training data.

2.3.3 Evaluation: The filtered final set of objects is then evaluated with a set of standard metrics for different IoU values. These metrics include accuracy, precision, recall, and F1-Score, as will be shown below.

2.4 Implementation and Modification of StarDist

The Python implementation for training and testing the StarDist model is available as a well-documented Jupyter notebook via GitHub (Schmidt, 2021). Several modifications were necessary for the delineation of tree crowns and geospatial data in general.

2.4.1 Generating Training Data: As StarDist was originally developed for microscopy images, the recommended procedures for generating training data were suboptimal for georeferenced remote sensing data. This study therefore developed an alternative approach in ArcGIS Pro. As stated above, objects are represented as star-convex polygons in the StarDist framework. A script-based validation was therefore implemented to verify the star-convexity of ground truth tree crown polygons. Polygons which were not star-convex were identified, listed, and manually corrected. After star-convexity was achieved for all polygons, the image data sets for the StarDist implementation were created.

A mask image corresponding to each original image is required for training the StarDist model. An example of an input image and the corresponding mask image is shown in Figure 2. Incomplete tree crowns located at image edges were ignored during the training process. The 64 tiles for training and validation were designed with an object-to-tile size ratio similar to the data in the original repository. Each tile had a size of 55 × 55 m. Several augmentation methods were applied to each tile to increase the training sample diversity and to enhance the model robustness and generalization to unseen data.

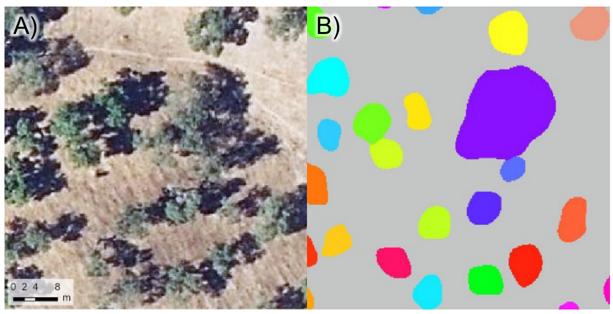

Figure 2: Example input image (A) and corresponding mask image (B) used for training. Incomplete tree crowns located at image edges are ignored during the training process.

2.4.2 Modification for ITDCD: The original StarDist algorithm computes a probability loss across all image pixels, encompassing background and annotated objects. However, this approach presents a challenge when working with incomplete annotations, which are prevalent in ITDCD applications. When similar-looking unannotated objects exist in the background, the standard loss function misclassifies them as background, thereby confusing the model and reducing its confidence in detecting actual annotated objects. Additionally, in the case of Dehesa imagery, background pixels are far more diverse compared to the rather homogeneous backgrounds in microscopy images, suggesting that training the model to recognize background patterns may be less effective. Following the approach of Tong and Zhang (2025), the implementation was modified by introducing a constraint to the probability loss function. Rather than computing loss across all pixels, the modified loss function considers only those pixels annotated as objects, while background pixels are excluded. Due to the commonly sparse annotations, this model is called $\text{StarDist}_{\text{Sparse}}$ in the following.

In contrast to the commonly sparse ground truth, a complete ground truth dataset was created for all 64 tiles in this study. Hence, the impact of neglecting background training in the StarDist$_{\text{Sparse}}$ implementation could be tested. This model with the standard loss function is referred to as StarDist$_{\text{Original}}$ in this study.

2.4.3 Training the Modified Model: The StarDist training workflow was adapted to incorporate StarDist$_{\text{Sparse}}$ with its modified probability loss function and StarDist$_{\text{Original}}$ with the original loss function. The 960 images (original and augmented) were evenly split, with 85 % assigned to training and 15 % to validation. Most training parameters from the original implementation were kept at their default values for this initial study. However, the number of rays per object was increased from 32 to 64 to allow a more detailed tree crown representation. Standard metrics (accuracy, precision, recall, F1-score) were calculated for the fully annotated validation dataset across multiple IoU thresholds, as shown in formulas (1) to (4).

$$\text{Accuracy (\%)} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} * 100 \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

$$\text{F1-Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

where TP = true positive
TN = true negative
FP = false positive
FN = false negative

2.4.4 Testing the Prediction: The original workflow for testing generates only two outputs per image, which are likely to be sufficient for spatially independent microscopy images: a mask image, similar to that shown in Figure 2, and ray visualization of all star-convex polygons in an image, as shown in Figure 3. However, these image-wise results are suboptimal for geospatial data, as the georeferencing information is lost and no spatial connection between tiles exists. Hence, this workflow was adapted to preserve georeferencing and enable subsequent merging of all individual objects into a complete, georeferenced polygon layer.

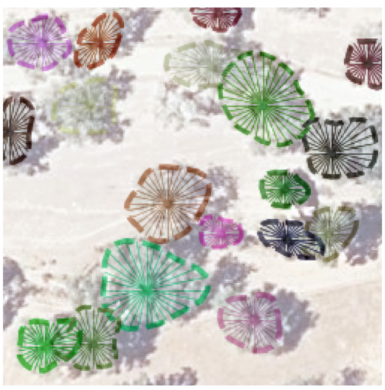

Figure 3: Default output of the standard StarDist prediction workflow showing ray visualization of star-convex polygons.

The 4-band composite raster dataset covering the entire *Dehesa San Francisco* study area was used to evaluate the model performance on unseen data across the heterogeneous dehesa landscape. The dataset was divided into 55 × 55 m tiles and exported as GeoTIFF images. The grid is shown in Figure 1. The GeoTIFF format preserves georeferencing information and embedded geospatial metadata, including map projection, coordinate reference system, and reference ellipsoid.

Four custom utility functions were implemented to transfer georeferencing information and metadata from raster tiles to prediction outputs and to export star-convex polygons as a georeferenced vector dataset. In particular, these functions performed:

1. Extraction of GeoTIFF metadata and tags
2. Transformation of prediction output coordinates from image space to real-world coordinates
3. Polygon reconstruction from ray endpoint vertices
4. Export of georeferenced vector data in GeoJSON format (selected due to superior processing speed compared to alternatives like Shapefile or GeoPackage)

StarDist$_{\text{Sparse}}$ and StarDist$_{\text{Original}}$ were used to predict tree crowns across all 1,849 images covering the entire study area. After applying the utility functions to generate georeferenced vector datasets, the main outputs were two GeoJSON files: the first contained all predicted tree crown kernel points as points, and the second contained the delineated tree crown boundaries as polygon objects. The prediction probability from the object probability map was stored for each crown in both datasets. The GeoJSON files were imported into a file geodatabase in ArcGIS Pro for validation against ground truth and result visualization. Figure 4 shows an example of a predicted tree crown with kernel point, delineated boundary, and reconstruction rays (here reduced from 64 to 32 rays for clearer visualization).

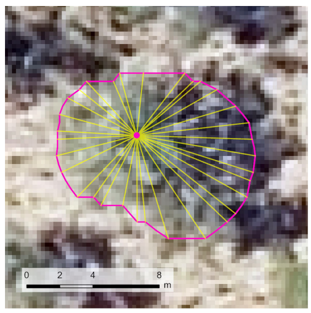

Figure 4: Visualization of predicted tree crown with kernel point, delineated boundary, and reconstruction rays.

The statistical evaluation for testing the prediction was adapted following Tong and Zhang (2025) to mitigate the effects of the incomplete ground truth coverage across the entire dehesa. Accuracy, precision, recall, and F1-Score were script-based calculated to evaluate the prediction performance. However, in contrast to the standard formulas mentioned above, these metrics were adapted to consider only matched crown pairs.

3. Results

The StarDist deep learning model, which had already demonstrated very good performance for ITDCD in a Canadian forest (Tong and Zhang, 2025), was tested for its suitability in delineating holm and cork oak trees within the heterogeneous dehesa landscapes. These trees present a challenging morphology due to their characteristic 'broccoli-like' structure with some trees having multiple crowns. StarDist was originally developed by Schmidt et al. (2018) for cell and nuclei detection in microscopy images. Thus, this study developed an adapted workflow, including training data generation, model definition, and application to unseen data, for geospatial data and individual tree detection and crown delineation (ITDCD) in particular. In contrast to the original workflow, where mask images are generated using software for annotating microscopy images, this study developed an alternative approach in ArcGIS Pro. All raster datasets used for training and model application were stored in GeoTIFF format to preserve georeferencing information. Predicted objects were exported as georeferenced vector data in GeoJSON format, inheriting the georeferencing from the input GeoTIFFs.

3.1 Performance during Training

During the training workflow, probability and NMS thresholds were optimized and standard metrics were calculated for different IoU thresholds. The optimized probability and NMS thresholds for $\text{StarDist}_{\text{Sparse}}$ were 0.49 and 0.40, respectively. With an IoU value of 0.50, the validation dataset (with complete ground truth) yielded strong performance (Accuracy: 77.95%; Precision: 0.92; Recall: 0.84; F1-score: 0.88). Results across multiple IoU thresholds are presented in Table 1 and Figure 5. As evident from the metric trends, performance remains relatively stable at lower IoU values but drops sharply at higher thresholds.

The optimized probability and NMS thresholds for $\text{StarDist}_{\text{Original}}$ were 0.50 and 0.30, respectively. Results across multiple IoU thresholds are presented in Table 2 and Figure 6, showing a similar performance of $\text{StarDist}_{\text{Original}}$ compared to $\text{StarDist}_{\text{Sparse}}$.

IoU	Accuracy	Precision	Recall	F1-Score
0.3	83.33 %	0.95	0.87	0.91
0.5	77.95 %	0.92	0.84	0.88
0.7	63.79 %	0.82	0.74	0.78
0.9	38.96 %	0.59	0.54	0.56

Table 1. Metrics for different Intersection over Union (IoU) values with $\text{StarDist}_{\text{Sparse}}$.

IoU	Accuracy	Precision	Recall	F1-Score
0.3	82.41 %	0.96	0.86	0.90
0.5	76.98 %	0.92	0.83	0.87
0.7	63.48 %	0.82	0.74	0.78
0.9	34.85 %	0.55	0.49	0.52

Table 2. Metrics for different Intersection over Union (IoU) values with $\text{StarDist}_{\text{Original}}$.

3.2 Performance during Testing across the Study Area

$\text{StarDist}_{\text{Sparse}}$ and $\text{StarDist}_{\text{Original}}$, using their respectively optimized probability and NMS thresholds, were applied to delineate individual tree crowns across the entire study area. The herein developed workflow used the georeferencing information from the input GeoTIFF raster images to transform the predicted objects, more precisely, the coordinates of the predicted kernel points and ray endpoints, from image space into real-world coordinates. The polygon vector data objects reconstructed from the ray endpoint vertices were imported into ArcGIS Pro and validated against an independent ground truth dataset comprising 1,200 trees. Figure 7 presents the tree crowns delineated by $\text{StarDist}_{\text{Sparse}}$ for the entire study area along with three zoomed-in sections, illustrating the spatial distribution of correctly delineated and missed tree crowns. In areas with sparse tree cover (Figure 7 B), nearly all tree crowns were identified. In moderately dense regions (Figure 7 A), few crowns were missed, whereas in very dense canopy layers (Figure 7 C), the number of missed detections increased markedly.

The model performance was evaluated following the approach of Tong and Zhang (2025) for incomplete ground truth scenarios. With an IoU threshold of 0.5, $\text{StarDist}_{\text{Sparse}}$ achieved an accuracy of 87.50 %, a precision of 0.80, and a recall of 0.91, resulting in an F1-score of 0.85. Figure 8 presents a histogram of the IoU values between the ground truth and the predicted tree crowns, calculated for the correctly delineated objects. Out of the successfully delineated tree crowns (IoU > 0.5), 64 % achieved IoU values exceeding 0.7. These results indicate a good reliability of the delineation approach, particularly in sparse and moderate canopy densities. $\text{StarDist}_{\text{Original}}$ achieved similar results, with an accuracy of 89.60 %, a precision of 0.81, and a recall of 0.90, resulting in an F1-score of 0.85.

Figure 9 provides a visual comparison between tree crowns, predicted with both models and the corresponding ground truth outlines. Figure 9 A) and B) illustrate exemplary differences in the background composition. It can be seen that the delineation performs very well in areas with a sparse tree cover and a homogeneous background, while larger discrepancies occur in areas of moderate vegetation density. In particular, textured background, like shown in Figure 9 B) may contribute to less accurate edge detection, potentially leading to the more pronounced differences between both of the predicted and the ground truth tree crowns. The comparison between $\text{StarDist}_{\text{Sparse}}$ and $\text{StarDist}_{\text{Original}}$ suggests that background recognition during

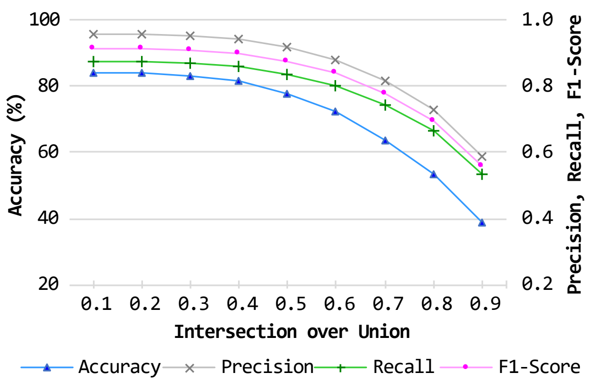

Figure 5.Graphs for validation metrics for different Intersection over Union values with $\text{StarDist}_{\text{Sparse}}$.

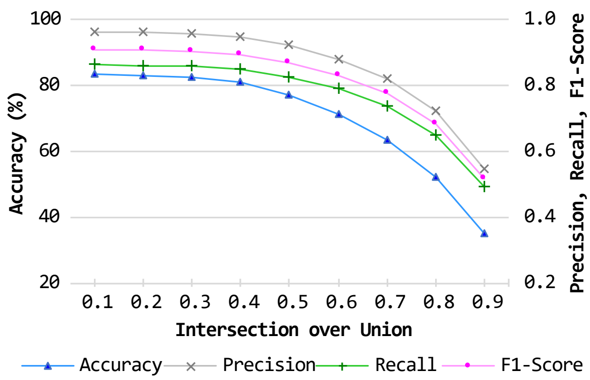

Figure 6. Graphs for validation metrics for different Intersection over Union values with $\text{StarDist}_{\text{Original}}$.

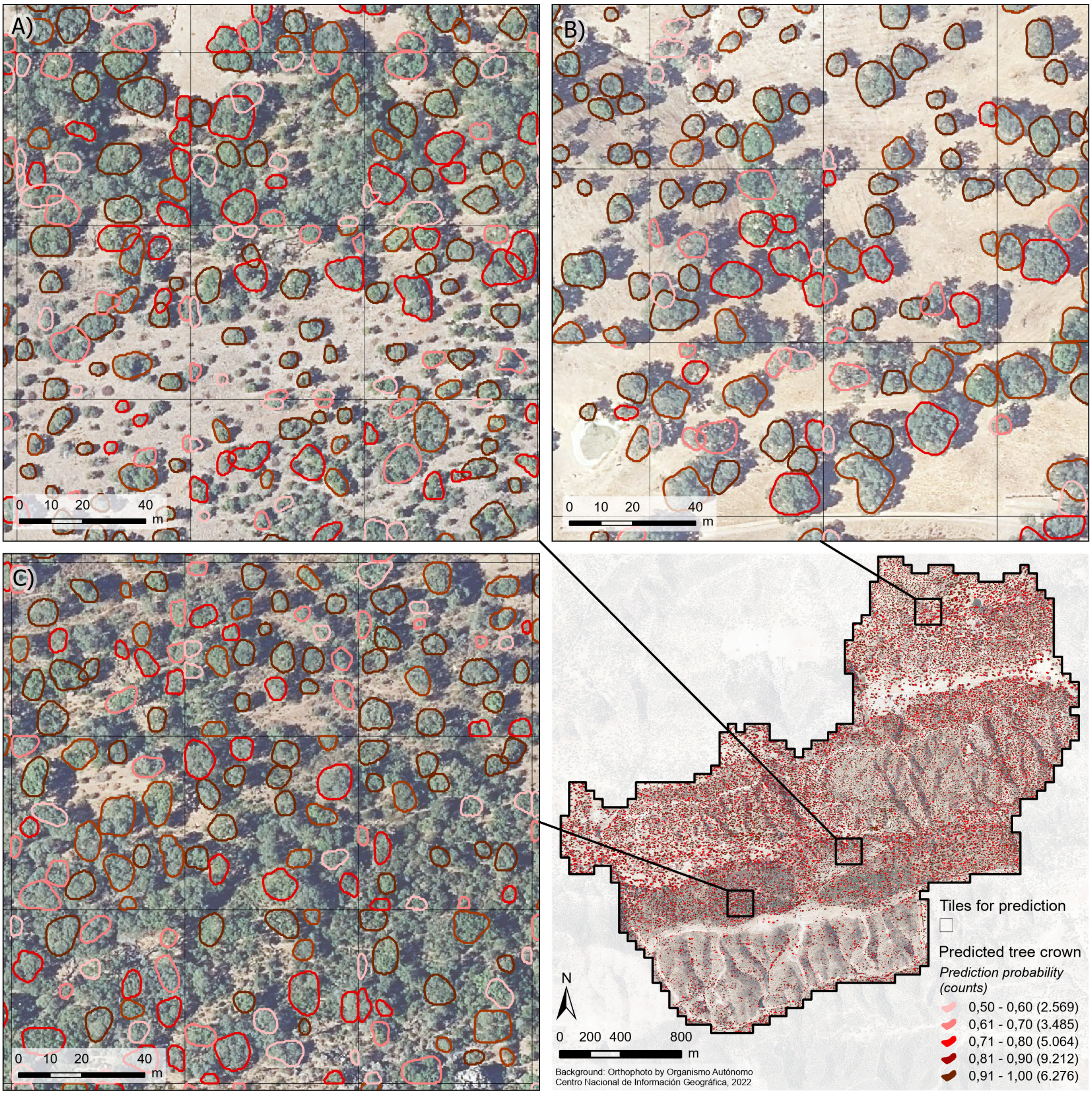

Figure 7. Tree crowns predicted with StarDist$_{Sparse}$ for the entire *Dehesa San Francisco*, Andalusia, Spain and three zoomed-in sections A), B), C).

training has little to no impact on the tree crown delineation accuracy, even though noisy, low-contrast backgrounds might still challenge edge detection.

Overall, the results indicate a quite robust performance of StarDist$_{Sparse}$ for ITDCD across the study area. Tree crowns were delineated reliably, while the accuracy varied with canopy density and background structure. The statistical metrics and visual comparisons illustrate consistent delineation quality in sparsely vegetated areas and more pronounced deviations in complex canopy conditions.

4. Discussion

This study presents the first evaluation of a modified StarDist implementation for individual tree detection and crown delineation (ITDCD) in structurally heterogeneous landscapes.

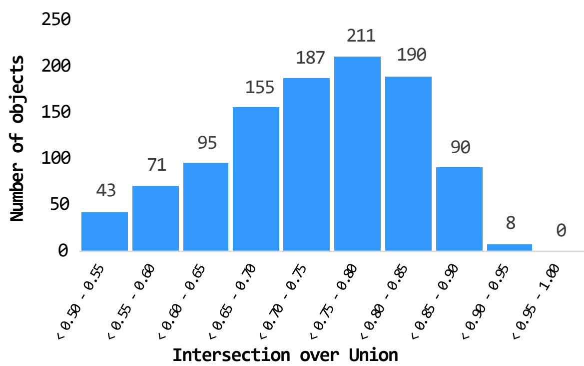

Figure 8. Histogram of the Intersection over Union values between matched ground truth and tree crowns predicted with StarDist$_{Sparse}$ across the entire study area (n = 1,050).

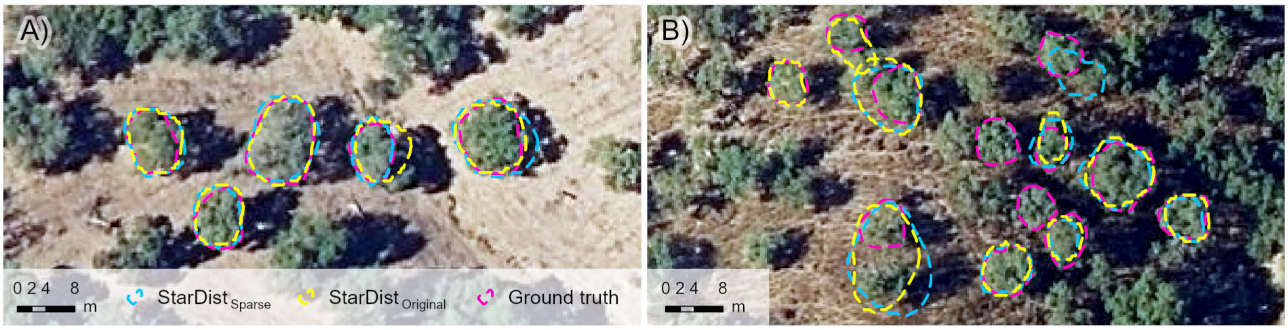

Figure 9. Comparison of tree crown predicted with $\text{StarDist}_{\text{Sparse}}$ and $\text{StarDist}_{\text{Original}}$ with ground truth. Layers of predicted tree crowns were here reduced to those overlapping with ground truth polygons.

Originally developed by Schmidt et al. (2018) for cell and nuclei segmentation in microscopy images, the StarDist algorithm demonstrates essential features which are beneficial for ITDCD. As shown by Tong and Zhang (2025) in a Canadian forest study, star-convex polygons effectively capture rounded objects such as tree crowns, similar to cells, as can also be seen in Figure 4. However, unlike the comparatively homogeneous Canadian forest, the *Dehesa San Francisco* study area on the southern Iberian Peninsula is characterized by a steep terrain and continental Mediterranean climate conditions. These factors cause both geometric image distortions and strong variations in tree density (10–85 trees/ha) (Figure 1). Additionally, the sparse tree cover generates extensive shadow effects, a common challenge in remote sensing imagery.

As evident from the zoomed-in sections of Figure 7 and from Figure 9, StarDist effectively differentiated tree crowns from shadows across most areas. This robust performance is likely due to the U-Net architecture, which fuse high-resolution feature maps from early encoder stages with context-rich representations from later decoder stages to preserve fine-grained boundary details and enable reliable edge detection under challenging shadow conditions. Studies such as Hu and Jung (2021) and Freudenberg et al. (2022) confirm that U-Net is well-suited for individual tree crown delineation.

A further challenge probably arose from the variability in background composition across the study area. Unlike the homogeneous backgrounds of microscopy images, the imagery here shows highly diverse and textured backgrounds (Figure 9 A and B). The comparison between $\text{StarDist}_{\text{Sparse}}$ and $\text{StarDist}_{\text{Original}}$ suggests that explicit background pixel training during model training can be neglected for tree crown delineation. Given that complete annotations are rarely available for ITDCD approaches, excluding background pixels via a modified loss function is advantageous. However, low-contrast backgrounds may still challenge edge detection and future studies should investigate the impact of background heterogeneity on delineation accuracy and explore preprocessing strategies.

Figure 7 shows that tree crowns located at image edges cannot be fully delineated. For this first evaluation, the study area was divided into non-overlapping tiles and it was ensured that the 1,200 ground truth trees, used to evaluate the prediction with unseen data, were fully contained in the centers of tiles. Future studies should incorporate established strategies, like using overlapping tiles and retaining only objects fully contained in central regions (Cira et al., 2024; Marmanis et al., 2018).

Overall, this initial evaluation indicates the suitability of the modified StarDist implementation for delineating cork and holm oaks, characterized by their distinctive 'broccoli-like' structure with some trees having multiple crowns. A prerequisite for applying StarDist was that the tree crown polygons were star-convex, and fewer than ten crowns required manual correction. However, the extent to which star-convex polygons are suitable for cork and holm oak crowns requires further evaluation. Furthermore, limitations become apparent in areas with dense tree layers, where crown separation remains challenging. Higher-resolution imagery data and additional information could improve discrimination in these regions. For example, height information derived from airborne laser scanning is well-known to enhance delineation accuracy (Hamraz et al., 2016; Hu et al., 2014). Data with sufficient point density, however, were not available for the present study area, but further investigations are planned.

5. Conclusion and Outlook

This study adapted the StarDist deep learning model for individual tree detection and crown delineation (ITDCD) in the heterogeneous dehesa landscape of the southern Iberian Peninsula. Originally designed for microscopy image segmentation, the workflow for training and testing was modified for geospatial data characteristics and a modified model, which mitigates challenges of incomplete ground truth annotations through a modified probability loss function was evaluated against the original model.

The adapted implementation, $\text{StarDist}_{\text{Sparse}}$, achieved a good overall performance, with an accuracy of 87.50 % and an F1-score of 0.85 when evaluated against an independent ground truth dataset of 1,200 trees and an IoU threshold of 0.5. $\text{StarDist}_{\text{Original}}$ achieved similar results, with an accuracy of 89.60 % and an F1-score of 0.85. Most crowns in areas of sparse to moderate canopy density were well delineated, while performance decreased in dense vegetation, primarily due to adjacent and overlapping crowns.

These findings suggest that StarDist’s star-convex polygon representation and U-Net as backbone could be suitable for delineating rounded yet irregular crown shapes typical of holm and cork oaks. This approach provides a promising starting point for extracting spectral information from individual tree crowns in satellite data, which can then be used for vitality assessment and oak decline monitoring on tree-level. The developed workflow of integrating ArcGIS Pro for data preparation, GeoTIFF-derived georeferencing, and GeoJSON vector export, provides a reproducible approach for geospatial application of StarDist in remote sensing analysis.

Future work should focus on optimizing the training data composition across varying tree-layer densities, implementing overlapping tile strategies to minimize boundary effects, and refining feature extraction for complex background conditions. Additionally, integrating higher-resolution imagery or laser scanning-derived height data could improve crown separation in dense stands.

References

Camilo-Alves, C. de S. e P., Esteves da Clara, M.I., Ribeiro, N.M.C.A., 2013. Decline of Mediterranean oak trees and its association with Phytophthora cinnamomi: A review. *Eur. J. For. Res.* 132, 411–432. doi.org/10.1007/s10342-013-0688-z

Cira, C.I., Manso-Callejo, M.Á., Alcarria, R., Iturrioz, T., Arranz-Justel, J.J., 2024. Insights into the Effects of Tile Size and Tile Overlap Levels on Semantic Segmentation Models Trained for Road Surface Area Extraction from Aerial Orthophotography. *Remote Sensing* 16. doi.org/10.3390/rs16162954

CNIG (Centro Nacional de Información Geográfica), 2022a. Ortofotos PNOA Máxima Actualidad. https://centrodedescargas.cnig.es/CentroDescargas/ortofoto-pnoa-maxima-actualidad

CNIG (Centro Nacional de Información Geográfica), 2022b. Ortofotos PNOA Falso Color Infrarrojo. https://centrodedescargas.cnig.es/CentroDescargas/ortofoto-pnoa-falso-color-infrarrojo

Esri, 2025. Tree Segmentation. https://www.arcgis.com/home/item.html?id=6d910b29ff38406986da0abf1ce50836

Freudenberg, M., Magdon, P., Nölke, N., 2022. Individual tree crown delineation in high-resolution remote sensing images based on U-Net. *Neural Comput. Appl.* 34, 22197–22207. doi.org/10.1007/s00521-022-07640-4

Godinho, S., Guiomar, N., Gil, A., 2018. Estimating tree canopy cover percentage in a mediterranean silvopastoral systems using Sentinel-2A imagery and the stochastic gradient boosting algorithm. *Int. J. Remote Sens.* 39, 4640–4662. doi.org/10.1080/01431161.2017.1399480

Hamraz, H., Contreras, M.A., Zhang, J., 2016. A robust approach for tree segmentation in deciduous forests using small-footprint airborne LiDAR data. *Int. J. of Applied Earth Observation and Geoinformation* 52, 532–541. doi.org/10.1016/j.jag.2016.07.006

Hu, B., Jung, W., 2021. Individual tree crown delineation from high spatial resolution imagery using U-Net. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLIII-B3, 61–66. doi.org/10.5194/isprs-archives-XLIII-B3-2021-61-2021

Hu, B., Li, J., Jing, L., Judah, A., 2014. Improving the efficiency and accuracy of individual tree crowndelineation from high-density LiDAR data. *Int. J. of Applied Earth Observation and Geoinformation* 26, 145–155. doi.org/10.1016/j.jag.2013.06.003

López-Sánchez, A., Bareth, G., Bolten, A., Linstädter, A., 2017. Effects of declining oak vitality on ecosystem functions: Lessons from a Spanish oak woodland. *Geophysical Research Abstracts* 19, 17578.

Marmanis, D., Schindler, K., Wegner, J.D., Galliani, S., Datcu, M., Stilla, U., 2018. Classification with an edge: Improving semantic image segmentation with boundary detection. *ISPRS Journal of Photogrammetry and Remote Sensing* 135, 158–172. doi.org/10.1016/j.isprsjprs.2017.11.009

Moreno Marcos, G., Obrador, J.J., García, E., Cubera, E., Montero, M.J., Pulido, F., Dupraz, C., 2007. Driving competitive and facilitative interactions in oak dehesas through management practices. *Agroforestry Systems* 70, 25–40. doi.org/10.1007/s10457-007-9036-y

Navarro, A., Catalao, J., Calvao, J., 2019. Assessing the use of Sentinel-2 time series data for monitoring Cork Oak decline in Portugal. *Remote Sensing* 11. doi.org/10.3390/rs11212515

Plieninger, T., Wilbrand, C., 2014. Die Dehesas in Spanien, in: Konold, W., Böcker, R., Hampicke, U. (Eds.), Handbuch Naturschutz Und Landschaftspflege. Landsberg, pp. 1–13. doi.org/10.1002/9783527678471.hbnl2001012

Roscher, R., Russwurm, M., Gevaert, C., Kampffmeyer, M., Dos Santos, J.A., Vakalopoulou, M., Hänsch, R., Hansen, S., Nogueira, K., Prexl, J., Tuia, D., 2024. Better, Not Just More: Data-centric machine learning for Earth observation. *IEEE Geosci. Remote Sens. Mag.* 12, 335–355. doi.org/10.1109/MGRS.2024.3470986

Sapp, M., Rose, L.E., Tyborski, N., Linstädter, A., Sánchez, A.L., Mansfeldt, T., Waldhoff, G., Bareth, G., Bonkowski, M., 2019. Site-specific distribution of oak rhizosphere-associated oomycetes revealed by cytochrome c oxidase subunit II metabarcoding. *Ecol. Evol.* 1–15. doi.org/10.1002/ece3.5577

Schmidt, U., 2021. StarDist - Object Detection with Star-convex Shapes. https://github.com/stardist/stardist

Schmidt, U., Weigert, M., Broaddus, C., Myers, G., 2018. Cell Detection with Star-convex Polygons, in: Medical Image Computing and Computer Assisted Intervention– MICCAI 2018: 21st International Conference. Granada, Spain, pp. 265–273.

Tilly, N., Reddig, F., Lussem, U., Bareth, G., 2020. First Investigation of Mediterranean Oak Tree Vitality with High-Resolution WorldView-3 Satellite Data: Comparing ten Vegetation Indices and three Machine Learning Classifiers. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* XLIII-B3, 1069–1076. doi.org/10.5194/isprs-archives-XLIII-B3-2020-1069-2020

Tong, F., Zhang, Y., 2025. Individual tree crown delineation in high resolution aerial RGB imagery using StarDist-based model. *Remote Sens. Environ.* 319. doi.org/10.1016/j.rse.2025.114618

Zhao, H., Morgenroth, J., Pearse, G., Schindler, J., 2023. A Systematic Review of Individual Tree Crown Detection and Delineation with Convolutional Neural Networks (CNN). *Current Forestry Reports* 9, 149–170. doi.org/10.1007/s40725-023-00184-3

Zheng, J., Yuan, S., Li, W., Fu, H., Yu, L., Huang, J., 2025. A Review of Individual Tree Crown Detection and Delineation From Optical Remote Sensing Images: Current progress and future. *IEEE Geosci. Remote Sens. Mag.* 13, 209–236. doi.org/10.1109/MGRS.2024.3479871