A Multiview UAV Imagery-Based Method for Assessing Spruce Tree Health at the Individual Tree Level

Alva Anttonen, Axel Päivänsalo, Emma Turkulainen, Raquel Alves de Oliveira, Kirsi Karila, Niko Koivumäki, Roope Näsi, Eija Honkavaara

Finnish Geospatial Research Institute (FGI), 02150 Espoo, Finland – alva.anttonen@nls.fi, axel.paivansalo@gmail.com, emma.turkulainen@nls.fi, raquel.alvesdeoliveira@nls.fi, kirsi.karila@nls.fi, niko.koivumaki@nls.fi, roope.nasi@nls.fi, eija.honkavaara@nls.fi

Keywords: Bark beetle, Tree health classification, Deep learning, Remote sensing, UAV, Multiview images.

Abstract

Assessing the health of individual spruce trees in forests is critical for early detection of bark beetle infestations and effective forest management. This study presents a novel methodology that leverages multi-view uncrewed aerial vehicle (UAV) imagery to improve tree health classification at the individual tree level. High-resolution images were collected over a spruce-dominated forest and processed to extract multiple perspectives of each tree crown. Compared to the typical case of using one orthophoto per tree, our process yielded on average 31 images per tree crown. Deep learning models, including VGG16 and a simple CNN, were trained to classify trees as healthy, infested, dead, or non-spruce. Results demonstrate that incorporating multi-view images increased classification accuracy, particularly for the challenging infested and non-spruce categories, compared to traditional orthophoto-based approaches. The best-performing model achieved an overall accuracy of 0.94 and a macro F1-score of 0.85, with notable improvements in detecting infested trees.

1. Introduction

In uncrewed aerial vehicle (UAV) photogrammetry, image datasets are captured with high forward and side overlaps to enable robust image orientation and support accurate 3D modeling and surface reconstruction. Usually, when analysing UAV remote sensing datasets, processed products like point clouds or orthophotos are utilised. However, the raw high-resolution UAV images represent a substantial volume of data that is often not utilised in analysis tasks. Using the raw images might be beneficial, especially for deep learning (DL) methods that benefit from large amounts of data.

In the context of forest monitoring, high-resolution UAV image datasets enable detailed and precise analysis of forests at the individual tree level. Using orthorectified images provides only a single nadir perspective of each tree crown, limiting the view to one image per tree. On the other hand, we could use images of the tree crown from each image it is visible in, giving typically tens of perspectives per tree from different directions. The objective of this study was to develop a novel methodology to utilize multi-view images of individual trees from aerial images, and to study how utilising the multiview images helps in a tree health classification machine learning (ML) task.

This work focuses on detecting individual spruces in a spruce dominated forest, with an emphasis of identifying trees infected by the European spruce bark-beetle (*I. typographus*). Detected spruces are classified as healthy, infested, or dead, while non-spruce trees make up a fourth class. Bark beetles are a major forest pest and Europe has seen a surge of bark-beetle caused damage in the last decades (Patacca et al., 2023). The visible signs of a bark beetle infestation are sawdust-like powder at the base of the tree, entrance holes at the lower part of the trunk, bark shedding at the higher parts of the trunk, and changes in crown colour (Bárta et al., 2022). The earliest signs, sawdust and entrance holes, are small and close to the ground, and thus not visible in aerial photos. Analysis of orthophotos can only detect coloration symptoms that affect the top of the tree crown. Perspective photos have the advantage that they can show colouration changes at the lower parts of the tree crown.

While this study uses multi-view perspective images to detect bark beetle symptoms, the approach might be valuable in other tasks as well. For example, the shape of the tree crown from an oblique view may be relevant for tree species classification or the assessment of tree quality, such as identifying anomalies due to nutrient deficiencies or snow damages.

2. Materials and Methods

2.1 Data gathering

The study area was located in the Helsinki Central Park and it was of size of 300 m x 220 m. The dataset was captured using the DJI Zenmuse P1 RGB camera onboard DJI Matrice 300RTK quadcopter drone on 27.9.2023. The images were taken at an altitude 97 metres above ground from the nadir angle. Spruce trees in the area were on average 30 m tall, so the flying altitude was 67 metres above the forest canopy. The study area was divided into training, validation and test areas, shown in Figure 1. The area included a total of 1609 trees from classes healthy (903), infested (25), dead (192) and non-spruce (489). Table 1 additionally shows how the trees were distributed among the training, validation and test splits.

2.2 Photogrammetric processing

Photogrammetric processing was used to generate a 3D model of the study area based on UAV images. This processing produced a point cloud representing the 3D geometry of the area, which was further converted into a digital surface model (DSM), which shows the canopy height of the forest, and a digital ortho map (DOM). The processing also yielded the precise

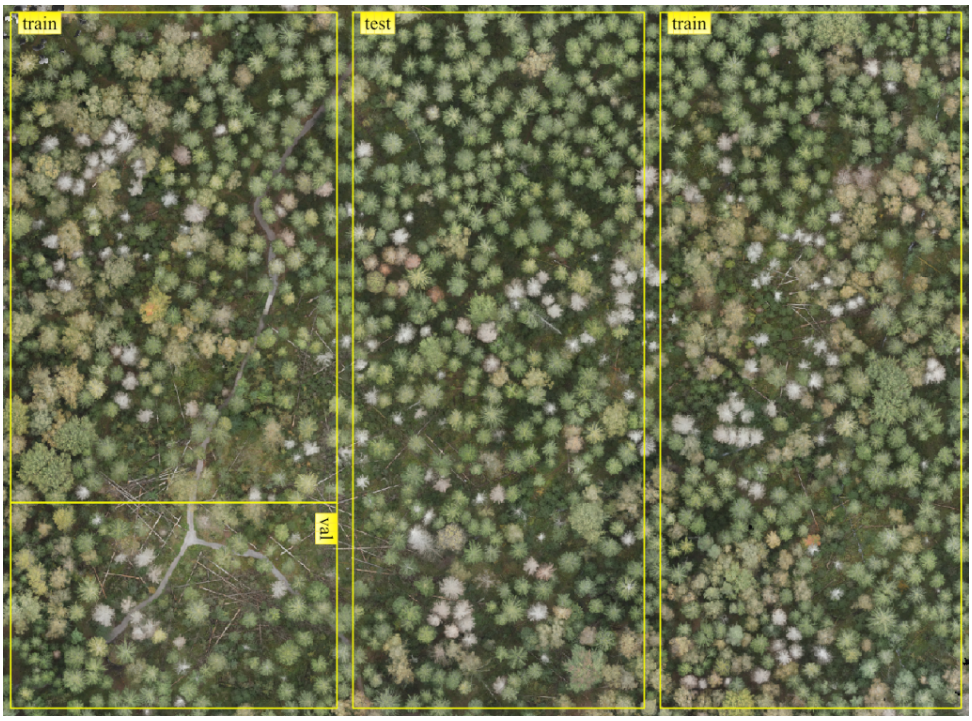

Figure 1. Orthoimage of the study area divided to train, validation and test areas. The areas were split in such a way that the number of infected trees would be sufficiently high in both the train and test areas.

Split	Healthy	Infested	Dead	Non-Spruce	Total
Train	495	11	107	389	1002
Val.	72	3	16	37	128
Test	336	11	69	63	479
Total	903	25	192	489	1609

Table 1. Summary of dataset splits.

camera positions and orientations of the images, as well as the camera distortion parameters. Datasets were processed using Agisoft Metashape Professional photogrammetric software.

2.3 Tree detection

In this study, we used two methods to locate the tree bases. The training and test datasets were created using tree locations from the reference tree database with individual tree information generated from the national 5 points/m LiDAR dataset (Metsäkanta, Hyyppä et al., 2024). The tree crown diameter was estimated based on the trunk diameter estimate of the data and trees were assumed to be circular. The reference tree locations in the test set were partially manually cleaned.

For a fully automatic tree detection method, we use bounding boxes obtained by the detection ML model YOLOv11(Redmon et al., 2016). The bounding boxes were converted to ellipses with the same width and height as the bounding box, because an ellipse represents the tree shape more naturally and appears more consistent from multiple viewing angles than a rigid north-east-aligned rectangle. The ML detection method uses only the DOM as input. The YOLO11m model, pretrained on the COCO dataset, was fine-tuned for the one-class tree detection task on our test set for 200 epochs with default parameter values.

The ML tree detection method is presented here in order to show that a fully automatic pipeline from UAV images to multi-view detection is feasible. However, the subsequent segmentation and classification tasks utilise only the reference tree locations. This is to maximize the size and quality of datasets used for classification, which is the focus of this study. Even though the ML tree detection results are satisfactory (see section 3.1), it misses some trees, and missing even a few trees in the crucial infected class makes evaluation uncertain, since there are only 25 of them in the dataset.

2.4 Individual tree multi-view image extraction

To preprocess UAV data for single-tree multi-view analysis, it is necessary to identify each tree in each image. To achieve this, the 3D geometries of the trees were approximated and then projected to the original images. The projected geometries provided image segments for each individual tree in every image it appears in. An example of the results of the projection and segmentation are shown in Figure 2.

To approximate the 3D shapes of trees we utilized a digital terrain model (DTM), a digital surface model (DSM), and a DOM of the area. The tree bases were detected as polygons in the DOM. The z-coordinate of the base was determined from the DTM and the height of the top was determined from the DSM. The DSM and DOM were obtained by photogrammetric processing, while the DTM is based on the national lidar survey. In our approximation trees were assumed to have the geometry of an upright prism with an elliptical base, although any other convex shape could have been used as the base.

Projection of a point from a 3D scene to the 2D image plane of a camera is possible given the extrinsic and intrinsic parameters

Figure 2. An example of the projetion results for one image. Original image (top), segmentation based on projected cylinders (middle), and cut-out images of individual trees (bottom).

of the camera (Förstner and Wrobel, 2016, chap. 12.1). The camera parameters were obtained during the photogrammetric processing, and were thus available in our analysis. The projection of a scene point $\mathbf{x}_x = (x_s, y_s, z_s)$ to the corresponding point in the camera coordinate system $\mathbf{x}_c = (x_c, y_c, z_c)$ uses the collinearity equation (Förstner and Wrobel, 2016, chap. 12.1.3.1).

Given the projection equations for a single point, projecting the upright prisms representing trees is simply a matter of projecting enough points of the prism for a sufficient resolution. In practice, we approximated the base and top ellipses with 16 equally spaced points each and projected those. The convex hull of the 32 projected points gives the image segment of the prism.

The tree image segments will overlap as some of the trees occlude others behind them. Which tree is in front can be determined based on the horizontal distance from the ground nadir point of the camera to the tree. This works in the special case of the simple upright geometry of trees. For more complex geometries the overlaps might need to be resolved by a more sophisticated technique like ray casting. Segments with area less than 25% of the mean of all segment areas were deemed too small to contain useful information and filtered out.

2.5 Image classification

The dataset of single-tree images was used to train supervised DL models, which classify trees to four classes (healthy, infested, dead, non-spruce). The VGG16 (Simonyan and Zisserman, 2015) and a simple 2D convolutional neural network (CNN2D as described in Turkulainen et al., 2023) were used in this study. The CNN was chosen as a simple model with few parameters, but still suitable to the task. The VGG16 on the other hand is a pretrained model with more layers and parameters. Specifically, the model used was VGG16-BN with pretraining on the ImageNet dataset. A visual transformer model (Dosovitskiy et al., 2021) was also considered, but dropped due to poor performance in preliminary runs.

The classification models were trained for 500 epochs, and the weights that maximize accuracy on the validation set were used. The models required hyperparameter tuning to avoid overfitting. A hyperparameter optimisation based on random sampling was used to select optimal hyperparameters for each model. The training data was augmented with random rotations and flips for all models. The models and the optimised hyperparameters are summarised in Table 2. Each model was trained 5 times with different seeds, so that average performance could be evaluated.

Since each tree appears in multiple images, the predictions for one tree are aggregated to get the final prediction. Multiple aggregation methods were tested, including choosing the most commonly predicted class (mode), weighing the predictions by confidence and taking the class with largest total weight (weighted mode), and taking the prediction with highest confidence.

2.6 Evaluation

To test whether using multi-view data gives better results than using orthophotos, a DL model was trained also with the orthophoto dataset. The results of the orthophoto model were then compared to the multi-view model results.

The models were evaluated on data from a held-out test portion of the study area. Classification accuracy was evaluated with overall accuracy (OA), precision, recall and F_1-score. Out of these OA takes into account all classes, while precision, recall and F_1-score are class specific. The averaged F_1-score across all four classes was also calculated (macro F_1).

Overall accuracy was calculated as the fraction of correct predictions out of all predictions. The precision and recall metrics focus on different types of errors. Precision depends on false positives, while recall depends on false negatives. Usually, high recall can be achieved by sacrificing on precision and vice versa, and the challenge is to do well on both at the same time. The F_1-score is the harmonic mean of precision and recall, and is used as a compromise metric between the two. Given the number of true positives (tp), false positives (fp), and false negatives (fn), precision, recall and F_1-score are calculated as follows:

$$\text{precision} = \frac{tp}{tp + fp}\,, \tag{1}$$

$$\text{recall} = \frac{tp}{tp + fn}\,, \tag{2}$$

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}\,. \tag{3}$$

All of the aforementioned metrics range from 0 to 1, where higher is better. When reporting results we report percentage points ranging from 0 to 100 for conciseness and readability.

View	Model	Parameters	Batch size	Learning rate	Weight decay
Orthophoto	CNN	$\sim$ 200 000	20	0.000002	0.002
Orthophoto	VGG16	$\sim$ 134 million	20	0.0000006	0.02
Multi-view	CNN	$\sim$ 200 000	48	0.000003	0.003
Multi-view	VGG16	$\sim$ 134 million	44	0.0000001	0.003

Table 2. Classifier models listed with their number of trainable parameters and the optimised hyperparameters that were used in training.

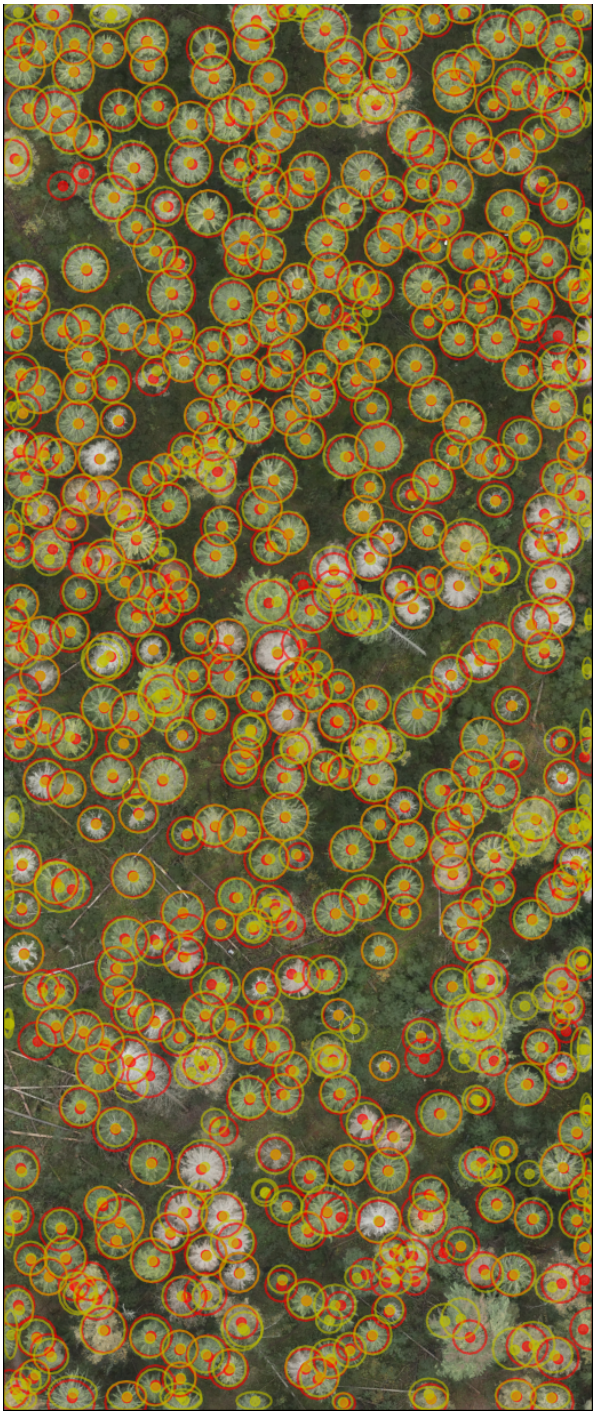

Figure 3. Detected trees (yellow) and reference trees (red) in the test area.

3. Results and discussion

3.1 Tree detection results

The YOLOv11 based detection correctly located 460 trees in the test area (true positives), missed 19 trees (false negatives), and detected 111 extra trees (false positives). Based on this, the detection can be said to have precision 80.6, recall 96.0 and an F_1-score of 87.6. As can be seen in Figure 3, most of the false positives were due to duplicate detections of the same tree. This happened especially often to the non-spruce trees that had larger crowns compared to spruces. The successful detections are very similar to the reference detection, which means they would produce similar projections in the multi-view image extraction process.

3.2 Multi-view image extraction results

Using the reference tree locations, each tree had on average 31 multi-view projections. The distribution of projection counts among the trees is visualised in Figure 4. This means utilising multi-view data increased the amount of available data by an order of magnitude. One tree in the test area had zero projections, removing it from the dataset. However, this tree was inspected and found to be a mistake in the reference tree locations – there were two overlapping locations for the same tree. Apart from the one outlier, all trees had at least 15 views. The quality of the projections was varying. Some of the cut-out images barely showed the tree crown, only showed the trunk of the tree, or in the worst case, didn't show the correct tree at all.

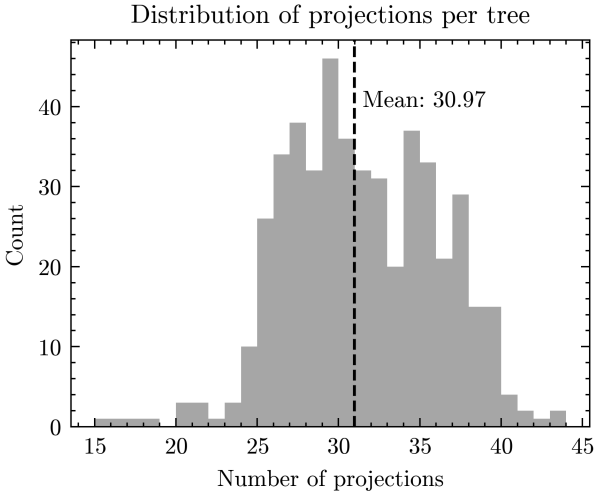

Figure 4. Histogram of the number of multi-view projections per tree.

View	Model	OA	Macro F_1	Class	Precision	Recall	F_1
Orthophoto	CNN	91.7±0.2	76.8±0.6	Healthy	94.9±0.5	97.2±0.5	96.1±0.1
				Infested	48.9±5.8	41.8±5.0	44.6±2.1
				Dead	88.2±2.5	**92.8**±0.0	90.4±1.3
				Non-Spruce	83.4±2.1	69.8±1.6	76.0±1.4
	VGG16	91.0±1.1	79.5±2.5	Healthy	94.4±0.8	96.0±1.0	95.2±0.7
				Infested	52.1±10.6	65.5±4.1	57.5±6.3
				Dead	92.1±3.0	90.4±1.3	91.2±1.0
				Non-Spruce	79.6±4.0	69.8±6.6	74.2±4.2
Multi-view	CNN	93.3±0.3	84.4±1.3	Healthy	95.5±0.1	97.1±0.5	96.3±0.2
				Infested	53.4±6.2	**80.0**±4.1	64.0±5.6
				Dead	91.6±1.3	88.4±0.0	90.0±0.6
				Non-Spruce	**95.1**±1.6	80.6±0.7	87.3±0.9
	VGG16	**94.0**±0.4	**85.1**±1.7	Healthy	**96.2**±0.2	**97.6**±0.4	**96.9**±0.2
				Infested	**56.9**±4.2	74.5±10.0	**64.4**±5.8
				Dead	**93.4**±1.5	89.9±1.8	**91.6**±0.7
				Non-Spruce	92.6±1.4	**82.9**±1.7	**87.4**±1.4

Table 3. Classification accuracy metrics. Each model was trained five times with different seeds. The mean and standard deviation of the five runs is reported for each metric. Higher is better for all shown metrics. The highest mean score in each metric is bolded.

3.3 Classification results

The accuracy metrics of trained classifiers are presented in Table 3. The displayed results only measure classification accuracy, omitting error caused by the object detection step. Although many ways of aggregating the multi-view predictions were tested, the different aggregation methods yielded very similar results. Thus, all results displayed are using the simplest aggregation method (mode).

The table shows for each metric the mean and standard deviation of five runs with different seeds. Because of the limited test set the standard deviations of the rare infected class are sometimes quite high (up to 11 percentage points). Despite the variance between runs, a clear ordering can be seen in the macro F_1-scores. The VGG16 model is better than the CNN model and the multi-view data is better than the orthophoto data. Furthermore, the improvement from incorporating multi-view data was larger than the improvement from using a better model. All models have a high overall accuracy (higher than 90 percentage points), but a significantly lower macro F_1-score. This is mainly due to the difficult but small infected class, which doesn't have much effect on overall accuracy, but has a large effect on the macro F_1-score. The orthophoto-based CNN performs very poorly on the infected class, but improves a lot when multi-view data is used – the F_1-score jumps from 44.6 to 64.0 percentage points. Overall, the best classification results were achieved using the multi-view VGG16, yielding F_1-scores of 96.9 for healthy, 64.4 for infested, 91.6 for dead, and 87.4 for non-spruce trees.

3.4 Conclusions and further research

The multiview method provided promising results, significantly improving the health classification of spruce trees in a study area impacted by the bark beetle. The improvements were the greatest for the challenging classes infested and non-spruce trees. Our further research will focus on additional testing of the method in different analysis tasks and environments.

Acknowledgements

This research was funded by the European Union within project "Network for novel remote sensing technologies in forest disturbance ecology" (decision no. 101078970) and by the Academy of Finland within project "Learning techniques for autonomous drone based hyperspectral analysis of forest vegetation" (decision no. 357380). This study has been performed with affiliation to the Academy of Finland Flagship Forest–Human–Machine Interplay—Building Resilience, Redefining Value Networks and Enabling Meaningful Experiences (UNITE) (decision no. 357908).

References

Bárta, V., Hanuš, J., Dobrovolný, L., Homolová, L., 2022. Comparison of field survey and remote sensing techniques for detection of bark beetle-infested trees. Forest Ecology and Management 506, 119984. https://doi.org/10.1016/j.foreco.2021.119984

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. Presented at the International Conference on Learning Representations (ICLR 2021).

Förstner, W., Wrobel, B.P., 2016. Photogrammetric Computer Vision, Geometry and Computing. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-11550-4

Hyyppä, M., Turppa, T., Hyyti, H., Yu, X., Handolin, H., Kukko, A., Hyyppä, J., Virtanen, J.-P., 2024. Concepts Towards Nation-Wide Individual Tree Data and Virtual Forests. ISPRS International Journal of Geo-Information 13, 424. https://doi.org/10.3390/ijgi13120424

Patacca, M., Lindner, M., Lucas-Borja, M.E., Cordonnier, T., Fidej, G., Gardiner, B., Hauf, Y., Jasinevičius, G., Labonne,

S., Linkevičius, E., Mahnken, M., Milanovic, S., Nabuurs, G.-J., Nagel, T.A., Nikinmaa, L., Panyatov, M., Bercak, R., Seidl, R., Ostrogović Sever, M.Z., Socha, J., Thom, D., Vuletic, D., Zudin, S., Schelhaas, M.-J., 2023. Significant increase in natural disturbance impacts on European forests since 1950. Global Change Biology 29, 1359–1376. https://doi.org/10.1111/gcb.16531

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You Only Look Once: Unified, Real-Time Object Detection, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Las Vegas, NV, USA, pp. 779–788. https://doi.org/10.1109/CVPR.2016.91

Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition. 3rd International Conference on Learning Representations (ICLR 2015).

Turkulainen, E., Honkavaara, E., Näsi, R., Oliveira, R.A., Hakala, T., Junttila, S., Karila, K., Koivumäki, N., Pelto-Arvo, M., Tuviala, J., Östersund, M., Pölönen, I., Lyytikäinen-Saarenmaa, P., 2023. Comparison of Deep Neural Networks in the Classification of Bark Beetle-Induced Spruce Damage Using UAS Images. Remote Sensing 15, 4928. https://doi.org/10.3390/rs15204928