Received: 1 September 2024 | Accepted: 25 February 2025

DOI: 10.1111/2041-210X.70024

Methods in Ecology and Evolution

PRACTICAL TOOLS

Towards enhancing field-based vegetation monitoring: A deep learning approach for species coverage estimation from ground-level imagery

Pauline Müller | Stefano Puliti | Johannes Breidenbach

Norwegian Institute of Bioeconomy Research (NIBIO), Division of Forest and Forest Resources, Ås, Norway

Correspondence
Stefano Puliti
Email: stefano.puliti@nibio.no

Funding information
Norges Forskningsråd, Grant/Award Number: 309671; HORIZON EUROPE Climate, Energy and Mobility, Grant/Award Number: 101056907

Handling Editor: Abebe Ali

Abstract

1. Field-based vegetation mapping is important for environmental assessments. Often, the area covered by a species is estimated visually within a reference frame. However, such assessments are prone to observer bias and a large variability.
2. We developed a deep learning pipeline relying on YOLOv8 models to segment species and estimate the percentage cover (%) of *Vaccinium myrtillus* (blueberry) and *Vaccinium vitis-idaea* (lingonberry), two key understory species in boreal forests. We used 138 nadir and downward-looking images of the forest floor captured in correspondence with 50 × 50 cm vegetation sub-plots assessed within National Forest Inventory (NFI) plots. First, we trained a bounding-box frame detection model to crop the image to the same area assessed in the field. Second, we trained an instance segmentation model to classify species. Third, we flattened the class values into a semantic raster and estimated the species-specific cover by pixel counting.
3. We evaluated our method against an independent test set of 156 images and found a root mean squared error (RMSE) of 8.82% for blueberry and 3.49% for lingonberry and no substantial systematic errors. An additional comparison with ocular estimation by various field workers for the same plots showed that the model estimates were within the range of estimates by field workers 8 out of 9 times for blueberry and 7 out of 9 times for lingonberry.
4. The developed method shows promise in reducing observer bias and variability in vegetation surveys, thereby improving their consistency while significantly reducing the time needed for species-specific coverage estimation. This is particularly beneficial for repeated measurements and monitoring vegetation cover dynamics. However, as the method relies on RGB data, it is limited to estimating the percentage of visible species that are not obscured by others. Expanding the method to include a broader range of cover classes (e.g. grasses, rocks, logs) or species could automate the capture of crucial information

from widely available ground-based images, thereby enhancing our ability to characterize a broader range of ecosystems.

KEYWORDS
biodiversity monitoring, botanical surveys, CNNs, deep learning, in-situ observations, National Forest Inventory, observer bias, YOLOv8

1 | INTRODUCTION

Field inventories are the basis for spatial planning, resource management and nature conservation (Haga et al., 2021; Ullerud et al., 2018). Accurate vegetation monitoring becomes more important with increasing pressure on natural resources and rising stress factors such as climate change (Fuchs et al., 2015). The quality of the data obtained during fieldwork is crucial. Poor-quality data, due to surveyor bias and inaccurate estimates of coverage, can lead to inaccurate ecological field assessments (Cherrill, 2016) and hinder the effective use of this data for monitoring vegetation changes through time.

Vegetation monitoring usually records two important pieces of information: the presence and abundance of species (Vittoz & Guisan, 2007). Cover fractions between different species are a common measure of abundance and are often estimated through ocular estimation within a defined frame (Bergstedt et al., 2009) or through line intercept sampling (e.g. US National Resource Inventory) (e.g. US National Resource Inventory Nusser & Goebel, 1997). Compared to other methods such as biomass harvesting or counting individuals, such cover estimates are less time-consuming (Bergstedt et al., 2009) and thus more scalable. However, the subjective nature of ocular estimation makes them prone to surveyor bias and potentially inconsistent through time (Bergstedt et al., 2009; Hahn & Scheuring, 2003). To investigate the bias from ocular coverage estimation, Bergstedt et al. (2009) compared coverage estimations of species from different fieldworkers for the same plots in the context of a national forest vegetation survey in Sweden. Nearly 20% of the variance in the data was explained by observer identity, indicating that the observer bias makes up a major part of the variability of cover estimates (Bergstedt et al., 2009).

One potential way to automate vegetation monitoring while reducing the surveyor bias and variability is to use imagery captured using off-the-shelf smartphone cameras (Morrison, 2015). Such imagery is increasingly collected by botanists, national monitoring programs and citizen science initiatives. When combined with advancements in deep learning, it offers an efficient and objective method to study vegetation communities (Kattenborn et al., 2021).

Images have been used to detect and identify species (Wäldchen & Mäder, 2018) and established datasets (e.g. iNaturalist; Van Horn et al., 2018), frameworks and models (e.g. Pl@ntNet-300K; Garcin et al., 2021) are now available. However, these efforts are largely confined to image classification tasks, which limit their application for fine-grained quantitative estimates of species-specific vegetation cover. Traditionally, the study of vegetation communities has relied on remotely sensed data, with deep learning showing great potential to provide more accurate and robust methods (Kattenborn et al., 2019, 2020). Research into close-range nadir vegetation survey images, such as those from smartphones, is largely unexplored, highlighting the need to adapt deep learning methods for this emerging and cost-effective data source.

This study aimed to develop an automated approach to estimate the vegetation cover (%) of two key target species: blueberry (*Vaccinium myrtillus*) and lingonberry (*Vaccinium vitis-idea*) within a reference frame using ground-level images taken by smartphones. By using our data and code, the provided method can efficiently be extended to other species and ecosystems. We first trained a YOLOv8 object detection model to crop the images captured in the field to the inside of a reference frame. A second model was trained to segment the target species, where the pixels per species were counted to estimate the percentage coverage. Species-specific cover estimates are common measures in vegetation surveys (Higgins et al., 1996), including in the Norwegian NFI (Breidenbach et al., 2020), and currently often rely on fieldworkers' ocular estimates, which vary considerably. Such an automated method offers a potential solution to reduce observer bias and improve the data accuracy and consistency of multi-temporal measurements.

2 | METHODS

2.1 | Data

This study was conducted using images and species cover estimates from the Norwegian NFI, thus spanning broad geographical ranges (i.e. >10° latitude) from coastal to mountain forests (>1000 m altitudinal gradient), and the range of blueberry and lingonberry cover distributions (see Figure 1a–d) representative of the variation of vegetation communities in Norwegian boreal forests (see Appendix S1 for examples).

Within the NFI, the percent coverage of blueberry and lingonberry is estimated for systematically distributed square (50 × 50 cm) vegetation plots, located 5 m from the plot centre in each of the cardinal directions (Breidenbach et al., 2020). During a pilot in the 2023 field season (May–October), seven experienced field workers collected nadir images of such vegetation plots. The images were captured using seven different camera models with varying resolutions (see Figure 1e,f) and under diverse light conditions (see

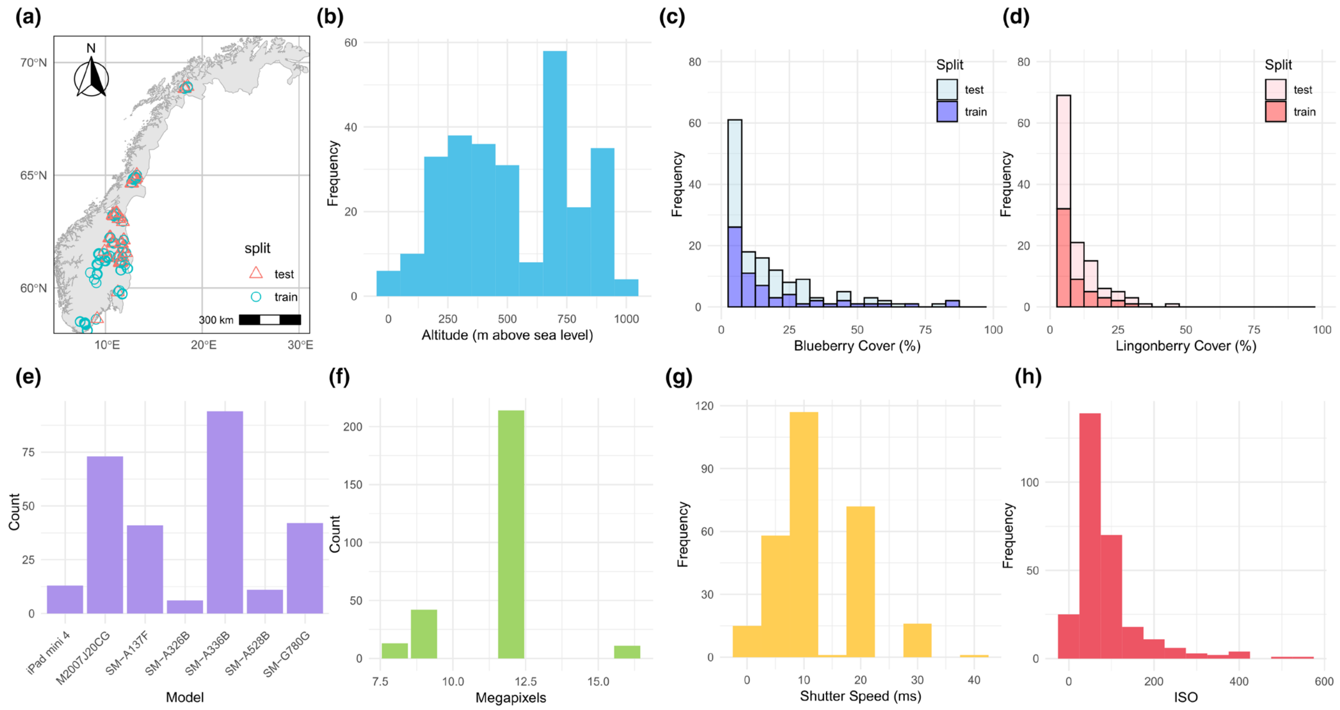

FIGURE 1 Summary information on the available data for this study, including (a, b) geographical, altitudinal distribution; (c, d) berry cover distributions; (e, f) camera specifications; and (g, h) image capture settings.

Figure 1g,h). In the remaining text, *italic* font indicates datasets, scripts or functions that can be found in our GitHub repository (see Data Availability Statement section). The individual datasets are listed below:

- *Field_data_NFI*: consists of the whole 294 ground-level images with blueberry and lingonberry coverage estimates. These are further divided into:
  - *Frame_data*: 108 images labelled with a bounding box inside the reference frame. These were randomly split into training (*frame_data_train*; 80%) and validation (*frame_data_val*; 20%).
  - *Species_segmentation_data*: 138 images (including the above Frame_data) that have been labelled for instance segmentation of blueberry or lingonberry. The species-specific instance labelling took approximately 50 h and was done using a semi-automated segmentation procedure based on SegmentAnything (Kirillov et al., 2023) and consequent manual editing. A randomly selected 80% of the images (*Species_data_train*) were used for training the model and the remaining 20% (*Species_data_val*) for validation within the model training.
  - *Species_cover_data_test*: 156 images (independent from the above datasets) with species-specific coverage estimates. These were the key data to evaluate the accuracy of the species-specific cover estimates from the proposed method against an independent and varied dataset (i.e. NFI plots scattered throughout Norway).
- *Field_course_data*: This data set consists of nine ground-level images with coincident estimations of blueberry and lingonberry coverage by 20 fieldworkers (multiple surveyors) from the annual NFI training course for calibration in 2023. These data are exclusively for testing to evaluate whether the estimates from our proposed automated method were falling within the human variation.

2.2 | CNN models

In this study, we utilized Ultralytics' YOLOv8 version 8.2.48 (Python 3.10.12), a mainstream framework thanks to its ease of use, performance, lean architecture and versatility in downstream tasks (Jocher, 2024). In this study, we relied on the object detection (i.e. bounding boxes) and instance segmentation (i.e. instance masks) tasks. All models were trained using data augmentation techniques, including transformations in the geometric (flipping, scaling, cropping, translation) and colour spaces (hue, saturation and brightness), mosaicking (combining multiple 4 images into a single mosaic image), erasing and cropping part of the image. All analysis was performed on a Google Colab instance with NVIDIA Tesla K80 with 12GB of VRAM.

2.2.1 | Frame object detection model

A YOLOv8 object detection model was trained using the *Frame_data_train* dataset, and the best model was identified using the *Frame_data_val* to mask field images to the frame's inner dimensions (see Figure 2a). Cropping to the inner frame ensured scale consistency across images and species cover estimates (Figure 2). For this

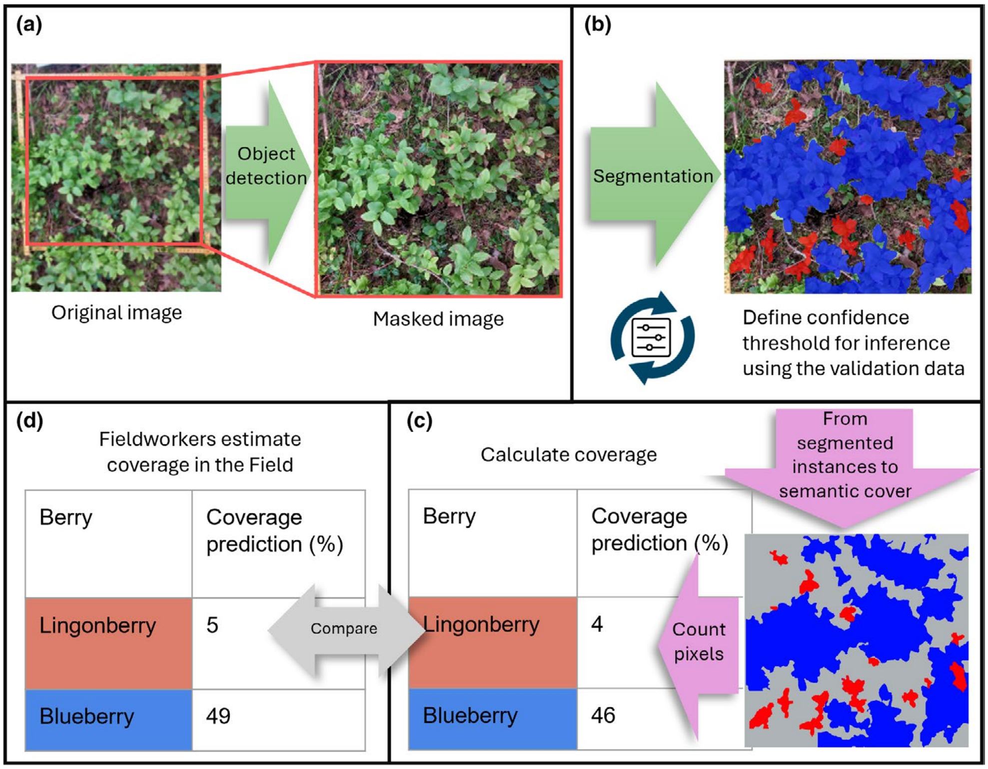

FIGURE 2 Workflow overview: (a) masking images, (b) applying the instance segmentation model, (c) converting instances to semantic rasters and estimating coverage, (d) evaluating the model against field-assessed species cover.

simple task (single-class and clear object), we opted for the leanest architecture (YOLOv8 nano; 2.6M parameters) with an image size of 640×640 pixels. The model weights were initialized using Ultralytics' COCO-pretrained model and trained for 50 epochs, with a batch size of 4. Given that each image contains only one vegetation quadrant, during inference, output was limited to the object with the largest confidence score.

2.2.2 | Species instance segmentation model

The instance segmentation model was trained using *Species_data_train*, and the best model was identified using *Species_data_val* to segment the target species in cropped images (Figure 2b). In this study, instances represent contiguous groups of individuals rather than single specimens. For this model, we opted for a YOLOv8 small architecture (*yolov8s-seg.pt*) model at high image resolution (1024 pixels), thus allowing us to capture fine species-specific structural traits (e.g. branch architecture and leaf traits).

For inference, the confidence threshold for predictions was optimized based on species coverage (%) rather than default instance segmentation metrics (i.e. F1 score-confidence curve). To achieve this, estimated species proportions from labelled images from *Species_data_val* (see next section) were used to iteratively evaluate the impact of different confidence thresholds on the root mean squared error (RMSE). The confidence threshold yielding the lowest RMSE was used in the following predictions.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{1}$$

where y_i is the reference species-specific coverage computed from the manual labelling, $\hat{y}_i$ is the estimated species coverage from the model and n is the number of images.

2.3 | Percentage coverage estimation

In the third step, the predicted instance masks (in YOLO text format) were converted to semantic masks (raster; .tif) using the *predict_cover* function (Figure 2c). Within the *predict_cover* function, each pixel is classified as either blueberry, lingonberry or background, and the cover of each species is calculated as the proportion of the pixels assigned to a species relative to the total number of pixels. To prevent overlapping instances, the function uses the confidence scores of individual instances, ensuring the mask includes only the highest-confidence instances in overlapping areas. The *predict_cover* function also provides parameters to control the number of species (number_of_classes), the image size (img_size_frame and img_size_species) and the confidence threshold (conf_treshold).

Methods in Ecology and Evolution

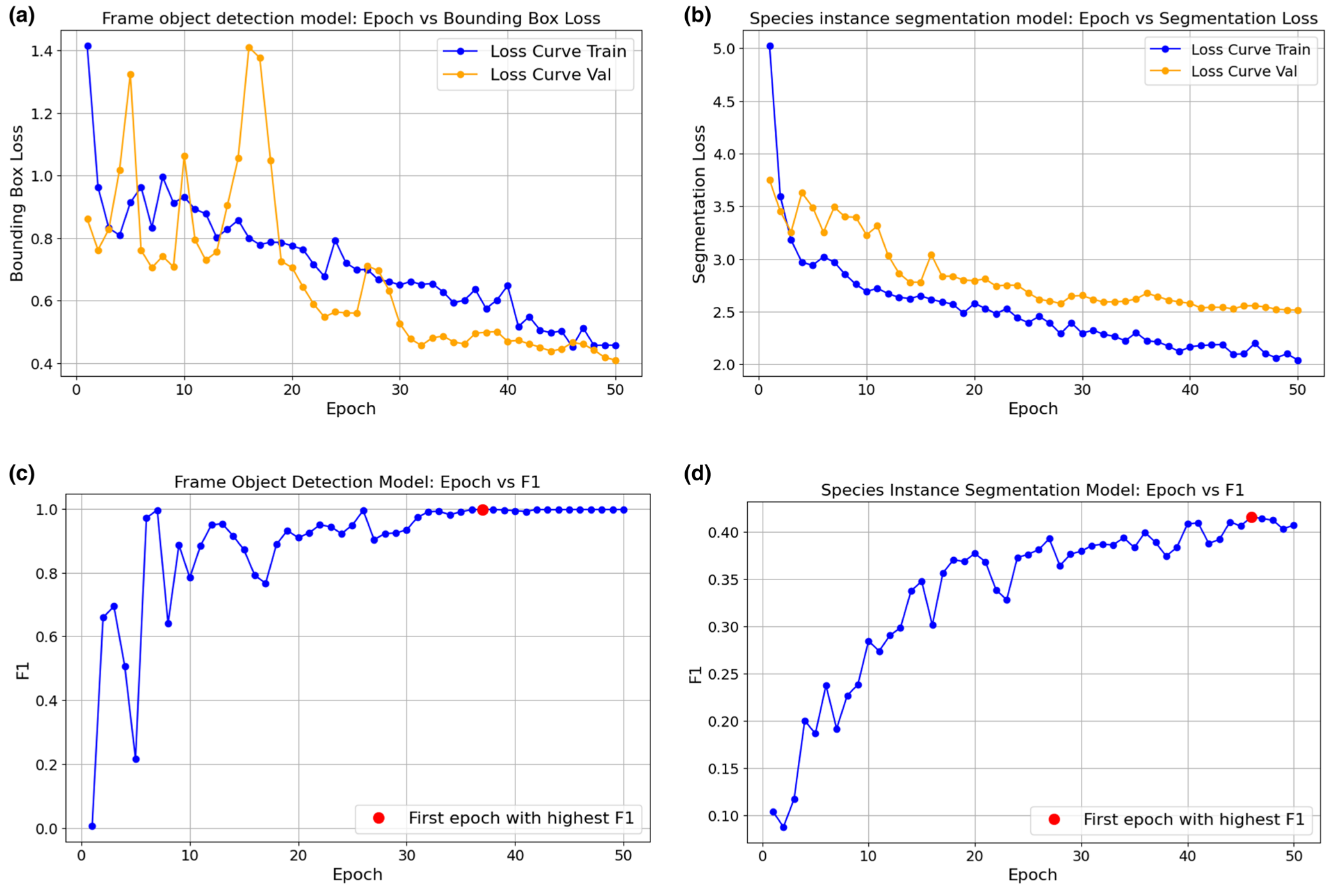

FIGURE 3 Epoch versus loss curve and epoch versus $F1$ curve for the frame object detection model (a, c) and species instance segmentation model (b, d).

2.4 | Evaluation

The developed method was evaluated based on the following metrics:

- Machine-learning metrics were computed using the validation datasets (*frame_data_val*: 22 plots and *Species_data_val*: 28 plots) and were intended to evaluate the performance of the object detection and segmentation models under a machine-learning perspective, and thus useful only for completeness and for future comparison with alternative models but with no direct link to the target species cover variable. These included analysing the true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). A TP was defined as a bounding box or mask with an intersection over union (IoU) ≥0.6 (default value in YOLOv8) and an FP if IoU <0.6 (Sportelli et al., 2023). Based on the TP, TN, FP and FN, we computed:

$$\text{Precision}(P) = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall}(R) = \frac{TP}{TP + FN} \quad (3)$$

To summarize P and R, we computed the $F1$ score ($F1$), which corresponds to the harmonic mean of precision and recall and is computed according to

$$F1 = \frac{2 \times P \times R}{P + R} \quad (4)$$

Domain-specific metrics were computed using the *Species_cover_data_test* (156 plots) and were intended to assess the predictive errors of the developed deep learning pipeline in the downstream application of species-specific coverage (%) estimation. Being directly linked to biologically meaningful outputs such metrics provide a clearer understanding of the impact of the developed model in "real-world" applications. The presence of both random and systematic errors was assessed by using the RMSE (see Equation 1) and MD metrics. Where y_i is the reference field estimated cover, $\hat{y}_i$ is the estimated species coverage from the model and $n = 156$.

$$\text{MD} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right) \quad (5)$$

In addition, to gain deeper insights into the objectivity of our method compared to traditional assessments, particularly in relation to surveyor variability, we applied our models to the dataset *Field_course_data*.

Here, we compared our species cover estimates with the range of the field estimates from 20 surveyors.

3 | RESULTS

3.1 | CNN models

The model training for the frame object detection did not show signs of overfitting after 50 epochs (Figure 3a) and the best model was found at the 37th epoch (Figure 3c). Based on the dataset *Frame_data_val*, the *F*1 was maximized (1.00) at a confidence threshold of 0.831. Such a high *F*1 is likely due to the combination of performing a very simple task and having a limited validation dataset (22 images). However, the step of cropping the image to the inside of the frame is key to ensuring a consistent scale for estimating the percentage coverage of different images from different plots.

On the other hand, the training for the species instance segmentation model showed initial signs of overfitting at approximately 30–40 epochs and the best model was found at epoch 46 (Figure 3b,d). The highest *F*1 (0.41) was found for a confidence threshold of 0.179. Nevertheless, at such confidence level, noticeable under-detection of blueberries was observed. Since the *F*1 is heavily influenced by the number of instances—an aspect irrelevant to our method, which flattens instances into species-wise semantic masks—we further fine-tuned the confidence threshold (Figure 4). This was done by minimizing the RMSE of cover estimates between the manually labelled field validation data *Species_data_val* and our method's predictions across confidence thresholds of 0.08 and 0.18.

While the RMSEs for blueberry fluctuate more than for lingonberry, but both species show minimum RMSEs between confidence thresholds of 0.10 and 0.13 (Figure 4), with a minimum average between the two species at 0.10, which was therefore selected for further predictions.

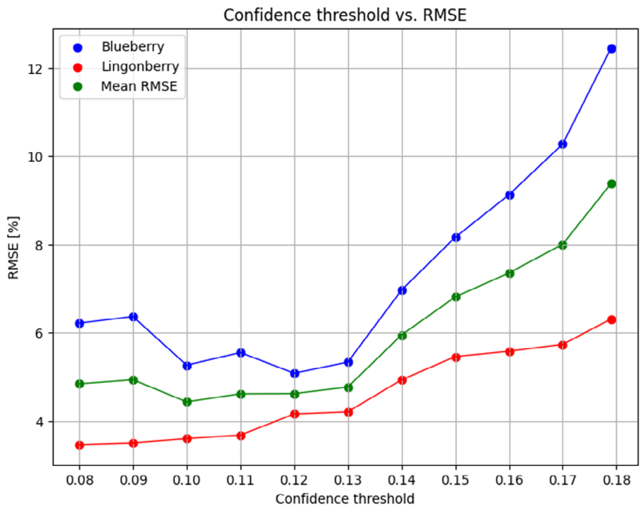

FIGURE 4 RMSE values concerning calculated coverages from labelled data in comparison to predictions with different confidence thresholds (28 validation images from *Species_data_val*).

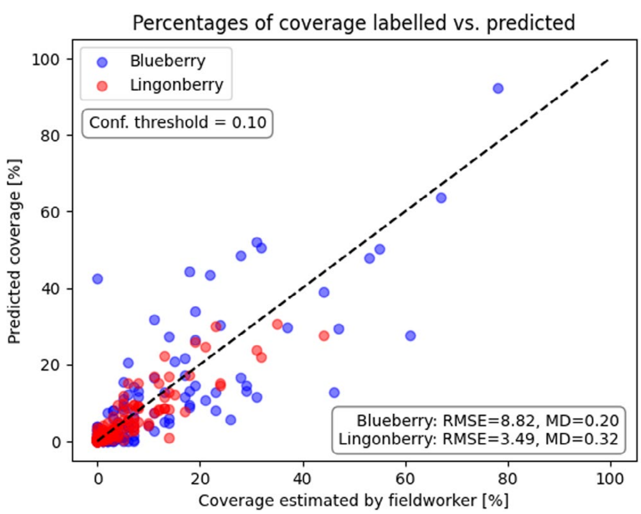

FIGURE 5 Coverage estimated in fieldwork versus prediction (156 test images from *Species_cover_data_test*).

3.2 | Percentage cover estimation

The domain-level evaluation of the predictions of our deep learning pipeline against the field-estimated values (single-surveyor) in the test dataset *Species_cover_data_test* (Figure 5) showed the main following characteristics:

- Our method could accurately predict the percentage cover for both species (RMSE 10%) and across the range of vegetation coverage 0%–80% without any noticeable signs of saturation. Further, we found a slight tendency to underestimate the percentage cover but this was always less than 1% of the mean species-specific cover.
- The percentage cover predictions for lingonberry were more accurate (RMSE = 3.49%) compared to those for blueberry (RMSE = 8.82%). This can be explained by the compound effect of a narrower range in the reference lingonberry cover (i.e. 0%–44%) compared to blueberry and the more distinct morphological features of lingonberry. Lingonberry's smaller, glossier leaves with revolute margins, low creeping growth, reddish stems, stable phenology and vibrant red berries create a higher contrast in images compared to blueberry. In contrast, blueberry plants have larger, smoother leaves, a taller erect form and thicker stems, making lingonberry easier to detect and segment (Kattenborn et al., 2020).

The evaluation of our method against the nine vegetation plots with coincident measurements by 20 surveyors (Figure 6) showed that, for most plots, our predictions were within or close to the range of the estimates of the fieldworkers and often close to the mean of the fieldworkers. This suggests our method offers a more

BRITISH ECOLOGICAL SOCIETY Methods in Ecology and Evolution

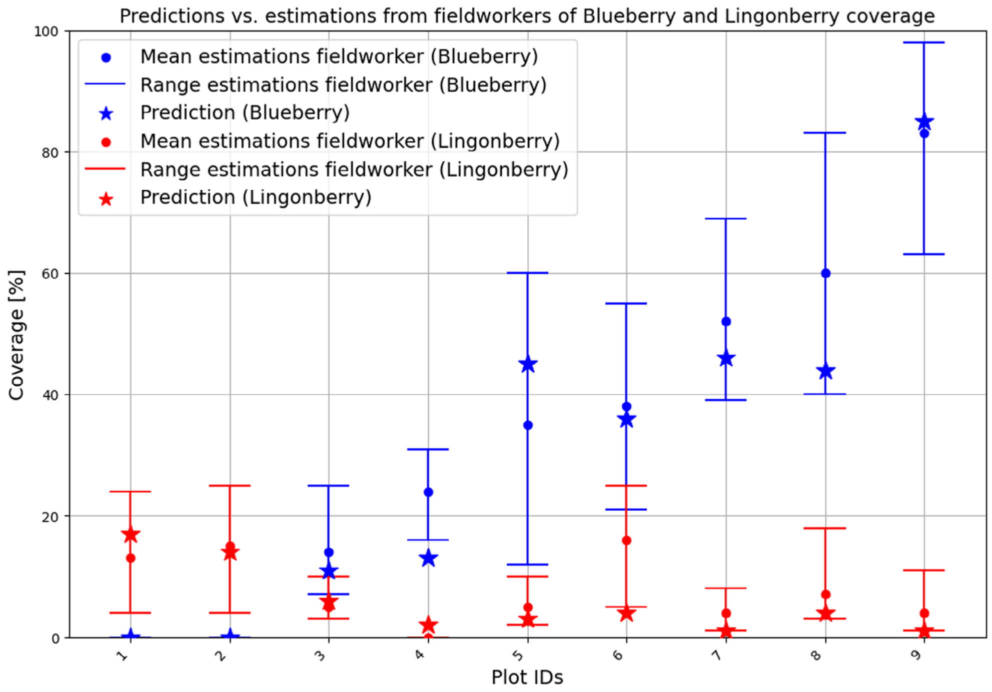

FIGURE 6 Blueberry and lingonberry estimations from the field course by 20 fieldworkers versus the predictions of the model (*Field_course_data*).

objective estimate that is as good as the field interpretation from several surveyors.

Interestingly, for plots with >40% blueberry coverage, lingonberry cover predictions were consistently lower than field estimates, likely due to the occlusion effect of the taller blueberry layer (e.g. plot 6 in Figures 6 and 7c). In contrast, images with relatively low species cover (e.g. plots IDs 1 and 2 in Figure 6) with a clear distinction amongst the target species and contrast against the background were more accurately estimated.

4 | DISCUSSION

Our study demonstrates how the developed deep learning pipeline can accurately and objectively estimate species percentage cover, effectively addressing key challenges such as surveyor variation and potential bias from the subjective ocular estimation in field vegetation surveys (Bergstedt et al., 2009). This data-driven approach ensures consistency in time-series vegetation surveys, which is crucial for vegetation monitoring programs like National Forest Inventories (NFIs) or grassland monitoring systems. Beyond boosting objectivity, the proposed method offers the advantages of speed (e.g. using a simple camera snapshot) and accessibility, relying on widely available and inexpensive smartphone technology. Using lightweight CNN models, our method can be deployed on edge devices like smartphones, enabling offline use and supporting citizen science projects. By streamlining species cover estimation, the proposed method can either reduce survey costs or increase the number of observations collected, thereby enhancing the accuracy and scale of large-area assessments. In a survey protocol, our method could give an additional estimate that a fieldworker can use to support or improve their visual estimate. Extending this method to include a broader range of species or cover classes (e.g. lichens, mosses, mineral soil, deadwood) would further broaden its applicability across diverse ecosystems (e.g. forests, grasslands, peatlands) and applications. Incorporating additional species and cover types into monitoring frameworks would enable a more comprehensive evaluation of vegetation communities, soil characteristics, habitat quality and ecosystem function. This, in turn, would provide researchers and practitioners with better data, ultimately strengthening conservation efforts and ecosystem management strategies.

Despite the promising results, this study represents an initial exploration of using deep learning for species cover estimation, based on a limited number of species and labelled data points. This study showed that the model worked better for species with limited seasonal variation (i.e. lingonberry is evergreen) and with traits that clearly separate them from the background (Kattenborn et al., 2020). The recognition of blueberry yielded larger RMSE and MD values in the evaluation, as the appearance of the plants differs greatly in the images. MacEachern et al. (2023) encountered a similar problem when they trained YOLO models to recognize the ripening stages of wild blueberries and estimate the yield. In their study, they analysed different stem samples and found that the number of berries, the position of the berries on the stem and the number of leaves on a stem vary greatly from genotype to genotype.

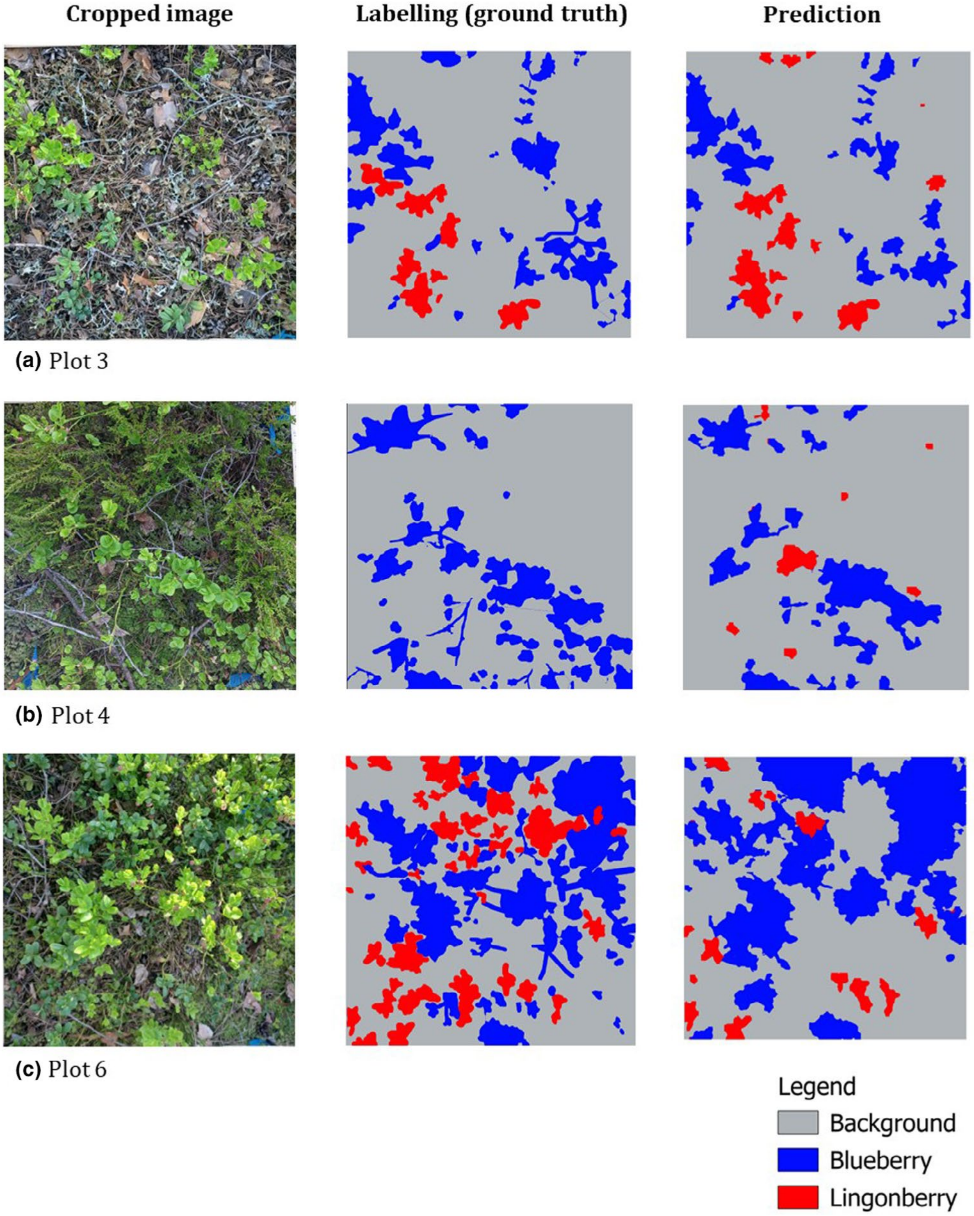

FIGURE 7 Cropped, manually labelled and predicted images from plots 3, 4 and 6 (*Field_ course_data*).

Future work should expand species coverage and refine models with more labelled data to improve scope, accuracy and transferability. Species instance segmentation in diverse, occluded and crowded environments remains challenging, and adding more species may reduce overall performance due to increased complexity. A potential way to overcome such an issue is by developing AI pipelines with vision model families focused on individual species or cover classes. This approach could boost overall performance while ensuring versatility in including new species and broader vegetation classes. It also enables the creation of species-specific datasets (i.e. where only one species is labelled), allowing flexible data adjustments to address species-specific performance and bias issues. Finally, the proposed method is intrinsically limited to the cover estimation of species in the upper vegetation layers, which can result in underestimating the cover of species in low-ground vegetation layers within more structurally complex environments. This study represents a starting point towards AI-assisted in-situ vegetation surveys, offering a scalable and cost-effective tool that extends beyond NFIs to broader ecological applications, including biodiversity monitoring, conservation planning and citizen science initiatives.

AUTHOR CONTRIBUTIONS

Pauline Müller, Stefano Puliti and Johannes Breidenbach conceived the study. Stefano Puliti and Pauline Müller designed the methodology; Pauline Müller analysed the data; Johannes Breidenbach acquired funding. Pauline Müller wrote the first draft. All authors contributed to the manuscript and gave final approval for publication. All researchers are located in the region where the study is carried out. This study is based on an issue arising from fieldwork that combines practical work with technical tools. There is a close dialogue with the people responsible for the fieldwork to apply this method in the field.

ACKNOWLEDGEMENTS

This study received funding from the EU under GA101056907 (PathFinder). This work is part of the Center for Research-based Innovation SmartForest: Bringing Industry 4.0 to the Norwegian forest sector (NFR SFI project no. 309671, smartforest.no).

CONFLICT OF INTEREST STATEMENT

There are no conflicts of interest.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.70024.

DATA AVAILABILITY STATEMENT

Data and code are available via https://zenodo.org/records/13361905 (Müller et al. 2025). Updates to the code base and models will be accessible via the following Github repository: https://github.com/Paulineemu/VegCover.

ORCID

Pauline Müller https://orcid.org/0009-0001-6084-8279
Stefano Puliti https://orcid.org/0000-0003-4624-8987

REFERENCES

Bergstedt, J., Westerberg, L., & Milberg, P. (2009). In the eye of the beholder: Bias and stochastic variation in cover estimates. *Plant Ecology*, *204*, 271–283.

Breidenbach, J., Granhus, A., Hylen, G., Eriksen, R., & Astrup, R. (2020). A century of National Forest Inventory in Norway—Informing past, present, and future decisions. *Forest Ecosystems*, *7*, 46.

Cherrill, A. (2016). Inter-observer variation in habitat survey data: Investigating the consequences for professional practice. *Journal of Environmental Planning and Management*, *59*, 1813–1832.

Fuchs, R., Herold, M., Verburg, P. H., Clevers, J. G. P. W., & Eberle, J. (2015). Gross changes in reconstructions of historic land cover/use for Europe between 1900 and 2010. *Global Change Biology*, *21*, 299–313.

Garcin, C., Joly, A., Bonnet, P., Affouard, A., Lombardo, J.-C., Chouet, M., Servajean, M., Lorieul, T., & Salmon, J. (2021). Pl@ntNet-300K: A plant image dataset with high label ambiguity and a long-tailed distribution. *NeurIPS 2021-35th Conference on Neural Information Processing Systems*. https://github.com/plantnet/PlantNet-300K/

Haga, H. E. E. S., Nilsen, A. B., Ullerud, H. A., & Bryn, A. (2021). Quantification of accuracy in field-based land cover maps: A new method to separate different components. *Applied Vegetation Science*, *24*(2), e12578. https://doi.org/10.1111/avsc.12578

Hahn, I., & Scheuring, I. (2003). The effect of measurement scales on estimating vegetation cover: A computer-assisted experiment. *Community Ecology*, *4*, 29–33.

Higgins, K. F., Jenkins, K. J., Clambey, G. K., Uresk, D. W., Naugle, D. E., Norland, J. E., & Barker, W. T. (1996). Vegetation sampling and measurement. *Research and Management Techniques for Wildlife and Habitats*, *5*, 567–591.

Jocher, G. (2024). Ultralytics yolov8 docs. https://docs.ultralytics.com/de

Kattenborn, T., Eichel, J., & Fassnacht, F. E. (2019). Convolutional neural networks enable efficient, accurate and fine-grained segmentation of plant species and communities from high-resolution UAV imagery. *Scientific Reports*, *9*, 17656.

Kattenborn, T., Eichel, J., Wiser, S., Burrows, L., Fassnacht, F. E., & Schmidtlein, S. (2020). Convolutional neural networks accurately predict cover fractions of plant species and communities in unmanned aerial vehicle imagery. *Remote Sensing in Ecology and Conservation*, *6*, 472–486.

Kattenborn, T., Leitloff, J., Schiefer, F., & Hinz, S. (2021). Review on convolutional neural networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *173*, 24–49.

Kirillov, A., et al. (2023). Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. Computer Vision Foundation (pp. 4015–4026).

MacEachern, C. B., Esau, T. J., Schumann, A. W., Hennessy, P. J., & Zaman, Q. U. (2023). Detection of fruit maturity stage and yield estimation in wild blueberry using deep learning convolutional neural networks. *Smart Agricultural Technology*, *3*, 100099.

Morrison, L. W. (2015). Observer error in vegetation surveys: A review. *Journal of Plant Ecology*, *9*, 367–379.

Nusser, S. M., & Goebel, J. J. (1997). The National Resources Inventory: A long-term multi-resource monitoring programme. *Environmental and Ecological Statistics*, *4*, 181–204.

Sportelli, M., Apolo-Apolo, O. E., Fontanelli, M., Frasconi, C., Raffaelli, M., Peruzzi, A., & Perez-Ruiz, M. (2023). Evaluation of YOLO object detectors for weed detection in different turfgrass scenarios. *Applied Sciences*, *13*, 8502.

Ullerud, H. A., Bryn, A., Halvorsen, R., & Hemsing, L. Ø. (2018). Consistency in land-cover mapping: Influence of field workers, spatial scale and classification system. *Applied Vegetation Science*, *21*, 278–288.

Van Horn, G., Aodha, O. M., Song, Y., Cui, Y., Sun, C., Shepard, A., Adam, H., Perona, P., & Belongie, S. (2018). The inaturalist species classification and detection dataset. Computer Vision Foundation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 8769–8778).

Vittoz, P., & Guisan, A. (2007). How reliable is the monitoring of permanent vegetation plots? A test with multiple observers. *Journal of Vegetation Science*, *18*, 413–422.

Wäldchen, J., & Mäder, P. (2018). Machine learning for image based species identification. *Methods in Ecology and Evolution*, *9*, 2216–2225.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1. Figure A1. Example test images showing the variation in the *Species_cover_data_test* data in terms of vegetation communities including herbs, grasses, ferns, lichens, mosses and presence of other berries (*Empetrum nigrum* as *Vaccinium uliginosum*) and various other cover types such as woody debris, litter, trees, shrubs (*Juniperus communis*) water, rocks.

How to cite this article: Müller, P., Puliti, S., & Breidenbach, J. (2025). Towards enhancing field-based vegetation monitoring: A deep learning approach for species coverage estimation from ground-level imagery. *Methods in Ecology and Evolution*, *16*, 949–957. https://doi.org/10.1111/2041-210X.70024