



Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Mapping of *Rumex obtusifolius* in nature conservation areas using very high resolution UAV imagery and deep learning

João Valente^{a,*}, Santosh Hiremath^b, Mar Ariza-Sentís^a, Marty Doldersum^c, Lammert Kooistra^c

^a Wageningen University & Research, Information Technology Group, Hollandseweg 1, Wageningen, the Netherlands

^b Aalto University School of Science, Espoo, Finland

^c Wageningen University & Research, Laboratory of Geo-information Science and Remote Sensing, Droevendaalsesteeg 3, Wageningen, the Netherlands

ARTICLE INFO

Keywords:

UAV
Deep learning
Transfer learning
Weed detection
Rumex

ABSTRACT

Rumex obtusifolius (Rumex or broad leaved dock) is one of the most common weeds in grasslands. It spreads quickly, lowers the nutritional value of the grass, and is poisonous for livestock due to its oxalic acid content. Mapping it is important before any control treatment is applied. Current methods for mapping Rumex either involve manual work or the utilization of ground robots, which are not efficient in large fields. This study investigated the feasibility of using aerial images from unmanned aerial vehicles (UAV) and deep learning to map Rumex in grasslands. Seven pre-trained CNN models were tested using transfer learning on UAV images acquired at 10 m, 15 m, and 30 m height. Based on Cross Validation results, MobileNet performed the best in detecting Rumex, with an F1-Score of 78.36% and an AUROC of 93.74%, at 10 m height. At 15 m, the detection performance was relatively lower (F1-score = 72.00%, AUROC = 88.67%), but the results showed that the performance can increase with more data. Experiments also showed that Rumex detection was dependent on the flight height since the algorithm was unable to detect the plants at 30 m height. The code and the datasets used in this work were released in an open access repository to contribute to the advances in grassland management using UAV technology.

1. Introduction

Grasslands are an important agricultural ecosystem in Europe since they represent 35% of the utilized agricultural area (Smit et al., 2008). Most of the obtained production in grasslands is destined for fodder and hence, they are primordial for livestock's nutrition and production. Grasslands have proved to reduce soil erosion (Liu et al., 2020) and to store CO₂ (Yang et al., 2019), which cuts down the carbon losses from agricultural land to the atmosphere and contributes against climate change. Nevertheless, the hectares devoted to grassland in Europe are decreasing in the last years –mainly because of two reasons– (1) the production costs are high and (2) the opportunity cost of sowing another crop is bigger. From all the production costs, pest control represents a large part of the expenditures where chemical products are restricted (organic farming) (Flaten et al., 2020) since weeds are removed manually. Therefore, research should focus on developing methods to detect unwanted weeds at early stages and avoid their infestation through the fields.

Rumex obtusifolius (Rumex or broad leaved dock) is one of the most

common weeds in production grasslands in the Netherlands (Valente et al., 2019). It is a perennial, herbaceous plant of the botanical family of knot weed plants (*Polygonaceae*) and reaches heights of 50 up to 120 cm (Mosyakin, 2005). Rumex is a weed because it competes with grass for water, light, and nutrients (mainly Nitrogen) (Hiremath et al., 2013), reducing crop yield and lowering its edibility since it is poisonous for livestock (Kristálová et al., 2011). Rumex has a high dissemination capacity. A single plant produces over 60,000 seeds per year in their flowering period (Holm et al., 1977). From those, 80% are vital seeds, which will germinate producing new plants. Vital seeds are long-lived and can germinate after burial for 21 years (Toole and Brown, 1946).

Rumex control is mainly done by spraying pesticides. Nevertheless, chemicals are only effective in rosette stadium because they are systemic products, which affect only the leaves and the stem of the plant. As Rumex has a taproot, a lot of energy is stored in the root. Hence, the pesticide affects the leaves and the stem, but not the root, which allows a quick regrowth of the weed (Cavers and Harper, 1964). In extensive used grassland and in conservation areas, a large-scale herbicide usage is banned to preserve biodiversity in grassland. Frequent and early

* Corresponding author.

E-mail address: joao.valente@wur.nl (J. Valente).

<https://doi.org/10.1016/j.jag.2022.102864>

Received 18 October 2021; Received in revised form 23 May 2022; Accepted 8 June 2022

Available online 22 June 2022

1569-8432/© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

mowing is also forbidden to protect soil-breeding bird species and diversity of plants. Both restrictions can result in massive spreading of broad leaved dock and thus complicate the cultivation of the grassland. There are other methods that do not utilize chemicals, for instance crop rotation and biological control (van Evert et al., 2009; Bond and Grundy, 2001). Another alternative is removing the weeds manually or using ground robots (Kounalakis et al., 2019; Lottes et al., 2016; van Evert et al., 2011). However, all the mentioned approaches are inefficient in large grassland fields since they are either expensive, time-consuming and/or labour intensive. Consequently, there is a strong need to find an automatic detection method to control Rumex, among other weeds.

2. Related work

In the last years, there has been significant progress in the fields of object detection and image classification due to the increase in computation power and the availability of large datasets (Ciresan et al., 2012; Russakovsky et al., 2015; Simonyan and Zisserman, 2015). These advances apply also to weed detection, recognition, and management. Already in 2018, (Kamilaris and Prenafeta-Boldú, 2018) published a survey of 40 papers that used deep learning techniques to address agricultural problems, including weed detection, which outperformed the traditional methods implemented until then. The most challenging step for weed detection is to distinguish between weed and crop species (Wang et al., 2019). To overcome the issue, (Brown and Noble, 2005) used both spectral and spatial features to identify weeds in crops. Similarly, (Hamylton et al., 2020) applied a CNN machine learning algorithm that leveraged information from both the high spatial resolution and the spatial context of the raster grids of the UAV images to map island vegetation. A semi-automatic Object-Based Image Analysis procedure developed with Random Forest was implemented by (Gao et al., 2018) to classify crop, weeds, and soil with an accuracy of 94.50%.

Grasslands have a great diversity of vegetation, which is challenging for algorithms to identify targeted weed species from the rest of the grass. Hence, deep learning applications on weed detection in grassland are quite limited. (Gebhardt et al., 2006; Gebhardt and Kühbauch, 2007) carried out two consecutive studies where image segmentation, local homogeneity calculation, and morphological operations were used to segment homogeneous regions, with detection rates from 71% to 95%. Nevertheless, their algorithms were trained with images taken by hand at close range, which is not feasible in large fields. An open source method that can be applied in site-specific weed management was developed by (Lam et al., 2021), with an F1-score of 78.65%. Nevertheless, they only implemented one CNN architecture (VGG16), many elements required manual intervention, and needed the use of commercial software.

Ground robots have already been used to detect weeds in grasslands. However, they have several disadvantages over UAV, mainly the coverage time and the adaptation to the terrain. The robot used in (van Evert et al., 2011) covered one hectare in three hours, whereas a UAV flies one hectare in less than half an hour –depending on the flight height and resolution to be acquired. Secondly, ground robots need to adapt to the irregularities of the terrain (slope, rocks, pits...), which might cause their break or stop their task, but UAV are not affected by ground difficulties. A combination of both technologies is a good solution to speed up the detection time of weeds. A UAV for a faster weed detection was flown in (Binch et al., 2018), which afterwards informed the ground robot about the location of weeds and redirected the robot to the exact location and enabled it to spray weeds from the ground. Their algorithm achieved 83–95% accuracy, with very specific flight parameters and a flight height of 8 feet. Some size distortions were introduced due to a varying flight height, which need to be controlled in future studies.

As a continuation of our previous research (Valente et al., 2019), UAV imagery is used in this research to map Rumex in grasslands. In this study, seven pre-trained convolutional neural network (CNN) models are applied to a high resolution aerial imagery acquired from an

Unmanned Aerial Vehicle (UAV) over wide grasslands to evaluate their feasibility on detecting Rumex plants. To evaluate the performance of the deep learning algorithms, two k-fold cross validation models were implemented for the 10 m and 15 m datasets. To cover the flight height limitation mentioned in our previous paper, three different flight heights (10 m, 15 m, and 30 m) are tested to evaluate the best configuration for Rumex mapping in high-resolution UAV imagery. Moreover, the code and dataset used for this study is available in an open access repository for the scientific community.

3. Materials and methods

The overall methodology of this study is subdivided into two stages: development and mapping (Fig. 1). The development stage consists of two phases, namely preprocessing and training. During preprocessing, an orthomosaic and its ground truth is transformed into a format suitable for training a deep learning model. This transformed data is used to compare different deep learning models using k-fold cross validation. The best model is then trained using all the data. The mapping phase involves applying the trained model to a new orthomosaic to detect and map the Rumex location. The details of both modelling and mapping stages are explained in the remainder of this section.

3.1. Study area

The dataset used in this study was acquired in a grassland field located in Germany, near the city of Kleve. This field belongs to the Salmorth reserve and is used by researchers of the nature conservation center in the Kleve district (Naturschutzzentrum im Kreis Kleve e.V). The exact location of the field is presented in Fig. 2 (Centre of Field 1: 5748143 N, 714617E. Centre of Field 2: 5748301 N, 714692E). All the grassland fields along the Rhine river to the West of the city of Emmerich (close to the Dutch - German border) were infested with *Rumex obtusifolius*.

3.2. Data acquisition and annotation

The aerial imagery was acquired using a Phantom 3 Professional UAV, on April 17th, 2018. Three flight heights were implemented (10 m, 15 m, and 30 m) with spatial resolutions of 5.9 mm, 6.4 mm, and 8.3 mm, respectively. The UAV images and the Ground Control Points (GCP) measured on the field were used to generate very high-resolution orthomosaics with the photogrammetry software Agisoft Metashape (St. Petersburg, Russia, version 1.7.3). The parameters chosen to generate the orthomosaic were the followings: medium accuracy, 400 k key points, and 100 k tie points. The orthomosaics used in this work have been published in a public repository to be used as benchmark for testing future works in this field (Valente and Kooistra, 2021).

Fig. 3 shows the orthomosaic of Field 1 divided into four quadrants (four main cardinal points) and the location of the labelled Rumex plants marked with red bounding boxes. Dividing the orthomosaic into four quadrants was not an indispensable step, but was done to simplify the preprocessing steps (see Section 3.3) and to ease the division of training, validation and test sets.

Experts of the nature conservation centre in the Kleve district located Rumex plants in the orthomosaic and labelled them using Matlab (Massachusetts, USA, version 9.4). They realized that in the 10 m and the 15-m-orthomosaics, they were not able to identify all the broad leaved dock plants, and the amount of labelled ones was approximately 80% of all the Rumex plants present in the field. Regarding the 30-m-orthomosaic, they did not manage to identify any Rumex plant due to the low spatial resolution, which means that no ground truth data was available for Field 2. Table 1 displays the number of labelled Rumex plants in each quadrant of the 10 m and 15-m-orthomosaics. At 10 meters, 481 instances of Rumex were detected by the experts across the four quadrants. Similarly, in the 15-m-orthomosaic, 129 Rumex

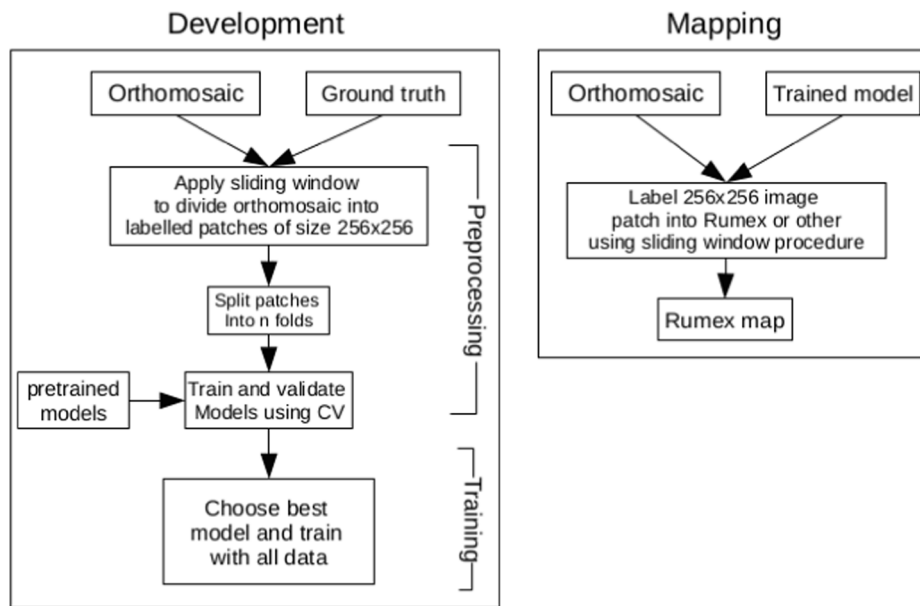


Fig. 1. Overall methodology of the study. In the development stage, the model is trained with the orthomosaic and the corresponding ground truth data. The trained model is used to map Rumex in the mapping stage.

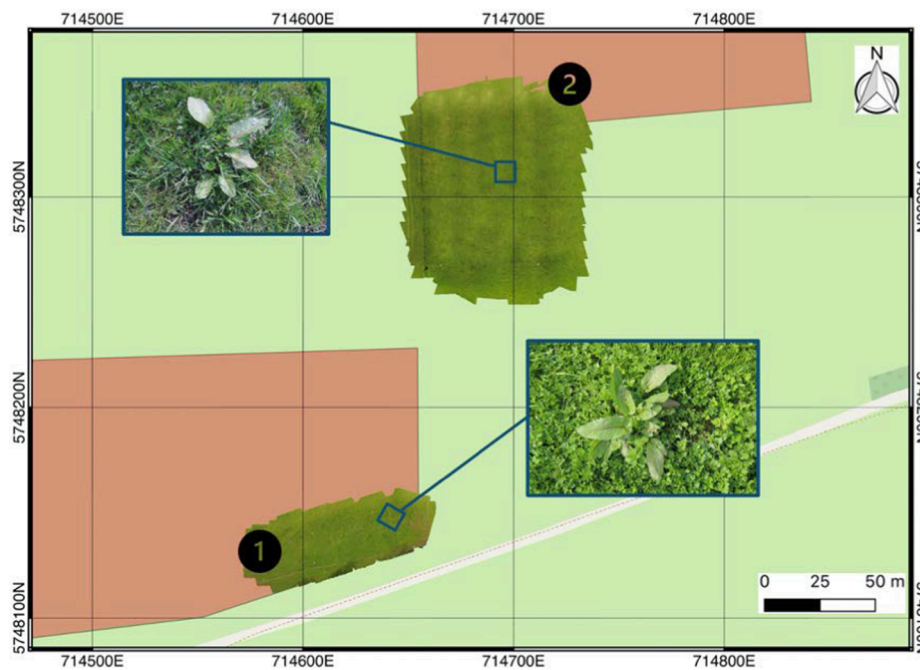


Fig. 2. Location of the fields over which the aerial imagery was acquired. The 10 m and 15-m-data was collected in Field 1 while 30-m-data was acquired in Field 2.

instances were detected by them.

3.3. Preprocessing

The first step during the preprocessing phase was to divide the orthomosaics into non-overlapping image patches of size 256×256 pixels implementing a sliding window operation, as described in (Valente et al., 2019). Each image patch was labelled as *Rumex* (positive case) or *Other* (negative case) based on the ground truth bounding boxes provided by the experts. When the ground truth box overlapped with the image patch, the corresponding patch was labelled as *Rumex*. Otherwise, it was labelled as *Other*.

Fig. 4 shows some examples of image patches with and without

Rumex. The resulting patches were used in a cross validation setting to evaluate the model performance. The details of the CV procedure are explained in Section 3.6.

3.4. Deep learning models

Several off-the-shelf deep learning models were compared for their efficacy in classifying *Rumex* on grasslands with UAV imagery. They were also used to establish a benchmark. The deep learning models considered are listed in Table 2. Large conventional algorithms such as VGG, Resnet and DenseNet were tested, along with smaller models (ShuffleNet, MobileNet, EfficientNet, MNASNet) that were developed for computationally efficiency, meaning that they can be run on devices

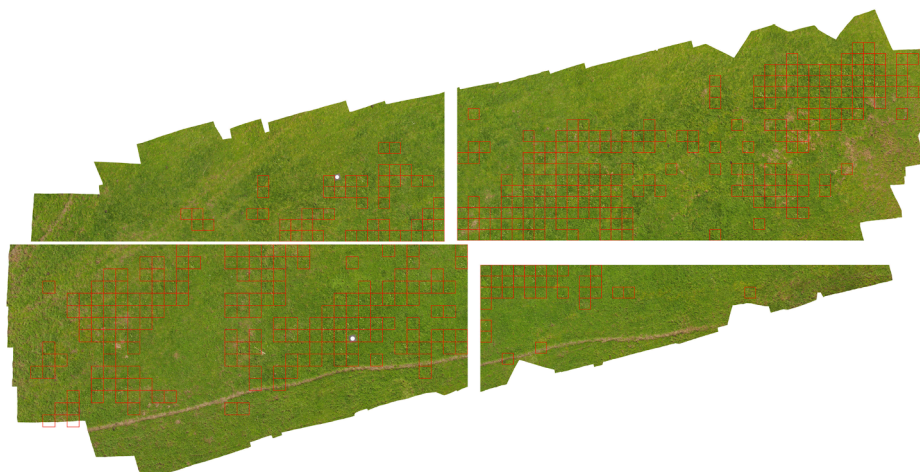


Fig. 3. Orthomosaic of Field 1 divided into four quadrants: NW (top left), NE (top right), SW (bottom left) and SE (bottom right), together with the annotated Rumex plants, marked as red bounding boxes.

Table 1
Number of *Rumex* and *Other* patches in each quadrant of the orthomosaics of 10 m and 15 m.

Quadrant	<i>Rumex</i>	<i>Other</i>
(a) 10 m orthomosaic		
NW	50	291
SW	178	366
NE	219	454
SE	34	177
Total	481	1288
(b) 15 m orthomosaic		
SW	95	245
SE	34	148
Total	129	393

Table 2
Off-the-shelf deep learning models implemented in this work.

Algorithm	Reference
VGG	(Simonyan and Zisserman, 2014)
ResNet	(He et al., 2016)
DenseNet	(Huang et al., 2017)
ShuffleNet V2	(Ma et al., 2018)
MobileNet V2	(Sandler et al., 2018)
EfficientNet	(Tan and Le, 2019)
MNASNet	(Tan et al., 2019)

optimizer and early stopping was used to ensure good generalization (Goodfellow et al., 2016).

All hyperparameters' values were set to default except the learning rate, which was determined separately by testing a range of values as described in (Smith, 2017). The code in this study is available at the following GitHub repository <https://github.com/satnih/rumex>.

3.5. Performance metrics

Several metrics were used to characterize the performance of the classification models, including accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUROC) (Fawcett, 2006). The first four metrics are derived from the confusion matrix of a binary classifier, which consists on a 2×2 matrix that groups the predictions into true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN), as presented in Table 3. The AUROC is a performance metric for discrimination between positive and negative cases. It takes values in the range of [0.5, 1], where a random classifier has a score of around 0.5 and a perfect classifier a 1. It is plotted as false positive rate (FPR) against true positive rate (TPR). While the confusion matrix and the associated metrics are meant to analyze the actual classification decision of a model, AUROC summarises the model performance across all potential classification decisions. All the metrics are defined in Eqs. (1)–(6).

Table 3
Confusion matrix of the Rumex study case.

		Actual class	
		Rumex	Other
Predicted class	Rumex	TP	FP
	Other	FN	TN

with limited hardware capacity.

PyTorch (version 1.7) (Paszke et al., 2017) was chosen to implement and train these models. The library includes these models pretrained on the ImageNet database (Deng et al., 2009), which enables them to classify 1000 different object classes. Since there are only two classes in this study (*Rumex* and *Other*), the final layer of these models were modified to solve a binary classification problem instead of 1000 classes. The model was fine-tuned by means of transfer learning (Yosinski et al., 2014) with the Rumex dataset. The training was carried out using Adam

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

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

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

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

$$FPR = \frac{FP}{FP + TN} \quad (5)$$

$$TPR = \frac{TP}{TP + FN} \quad (6)$$

3.6. Cross Validation

Since there are a limited number of image patches, cross validation (CV) was used to evaluate all the models for both the 10 m and 15 m dataset to ensure there was no bias in the analysis. For the 10 m data set, the patches from the four quadrants were randomly shuffled and split into five folds using stratified sampling. This ensured that each of the five folds had equal number of Rumex and Other patches since there are in total 481 Rumex patches and 1288 Other patches (Table 1). After splitting the data, 5-fold CV was used to evaluate each model. In each iteration, one fold was used as the test set, another was used as the validation set and the remaining three were used as the training set. This procedure was repeated five times, so training was performed on each fold. Each model's performance was determined by averaging the metrics of the five folds. For the 15 m dataset, 2-fold CV was used instead of 5-fold CV due to limited data points (129 Rumex and 393 Other).

4. Results

Table 4 displays the performance of the algorithms trained with the 10 m-orthomosaic. The model's names are suffixed with "-10" to indicate that they are trained with 10 m-data. Due to the high class imbalance, Table 4 also includes the accuracy of a default classifier, that always predicts an image patch to belong to the majority class (in this case it is *Other*). Thus, its accuracy of a default classifier is simply the percentage of *Other* patches in the test set. It can be seen that all models performed better than the default classifier. MobileNet-10 has the highest AUROC (92.68%) and F1-score (77.46%) compared to the other models. However, the other models (except MNASNET) are not too far behind. The minimum AUROC and F1-score obtained (excluding MNASNet) are 90.70% and 73.18%, respectively. However, the standard deviation of the F1-score of Densenet-10 and Efficient-10 are very high (14 and 10, respectively) indicating that their performance was not consistent across different folds. Based on these results, MobileNet-10 was chosen for the rest of the study. Moreover, MobileNet is one of the smallest and most computationally efficient models, with about

Table 4

Cross validation metrics (*mean ± std*) of different models on 10 m data. The Default classifier always predicts the image to belong to *Other* class. All values are in percentage (%).

	F1-score	Precision	Recall	Accuracy	AUROC
MobileNet-10	77.46 ± 4	80.07 ± 11	76.30 ± 5	87.68 ± 4	92.68 ± 1
ShuffleNet-10	75.21 ± 3	82.82 ± 9	70.20 ± 9	87.50 ± 2	91.32 ± 2
ResNet-10	73.33 ± 5	86.76 ± 8	64.44 ± 9	87.39 ± 2	91.11 ± 0
VGG-10	74.50 ± 7	74.54 ± 16	77.78 ± 9	84.90 ± 7	90.70 ± 2
DenseNet-10	73.18 ± 14	75.10 ± 2	80.06 ± 14	80.09 ± 21	92.28 ± 2
EfficientNet-10	75.73 ± 10	93.95 ± 7	72.38 ± 20	88.18 ± 3	90.68 ± 4
MNASNet-10	35.32 ± 20	22.76 ± 13	79.37 ± 44	39.65 ± 20	65.00 ± 30
Default-10	NA	NA	NA	62.40	50.00

300,000 parameters compared to other models.

To better understand the performance of MobileNet-10, the predictions of each fold was analyzed in more detail. For instance, Table 5 shows the confusion matrix of a fold that was the test set in an iteration. Overall, there were 34 misclassified patches that can be divided into three categories based on the type of errors: 1) wrong labels, 2) image artifacts, and 3) small plants. Fig. 5 shows some example patches from these three categories. The first row consists of patches that were mislabelled as *Other*. The model correctly identified them as Rumex, but they were accounted as false positives in the confusion matrix. The second row shows patches with image artifacts because of which the algorithm failed to detect Rumex. Finally, the last row shows patches with very small Rumex which the model could not detect. Overall, MobileNet-10 failed only on difficult cases, indicating its robustness.

4.1. Different Heights

4.1.1. Flight height of 15 meters

MobileNet-10 was also tested on the 15 m-orthomosaic to examine its re-usability. As expected, its performance was poorer on 15 m-data compared to 10 m-data (F1-score = 71.22% vs 77.46%, respectively). A reason for this reduction is that on the 15 m-orthomosaic the Rumex plants were smaller and the algorithm was trained with the size of the 10 m-orthomosaic. To solve this problem, the model was fine-tuned with some images of the 15 m-data. This updated model is referred as MobileNet-15. Despite the limited training data, there was a significant improvement in performance, as shown in Table 6. Particularly, the F1-Score increased from 71.22% to 78.16%, indicating that the model could be adapted to new environments efficiently with a small amount of training data. For future studies, more data can be introduced to the model to check if the model's performance can still increase.

4.1.2. Flight height of 30 meters

MobileNet-10 was also tested on the 30-m-orthomosaic, shown in Fig. 6. The 30-m-orthomosaic did not come with ground truth data because the experts could not identify any Rumex plant in it (mainly due to the spatial resolution). MobileNet-10 also failed to detect Rumex plants in it. These results indicated that the flight height of 30 m and the 8.3 mm spatial resolution were limiting factors for Rumex detection.

5. Discussion

MobileNet-10 was able to correctly classify 93.84% of the image patches of an orthomosaic. Furthermore, it showed promising results when tested on the 15-m-data. Nevertheless, it should be noted that all

Table 5

Confusion matrix from one iteration of 5-fold CV of MobileNet-10.

		Actual	
		Other	Rumex
Predicted	Other	246	12
	Rumex	22	74



Fig. 5. Example patches misclassified by MobileNet-10. The first row shows patches with wrong label (MobileNet-10 correctly identified them as *Other*). The second and the third rows represent patches with image artifacts and small Rumex plants, respectively, in which MobileNet-10 failed to detect Rumex plants.

Table 6

Performance of MobileNet-10 and Mobilenet-15. MobileNet-10 is the Mobilenet model trained with 10 m-data, whereas MobileNet-15 is the updated version of MobileNet-10, which is fine-tuned with 15 m-data. Default-15 is the default classifier that predicts the image to belong always to *Other*.

	F1	Precision	Recall	Accuracy	AUROC
MobileNet-15	78.16 ± 1	85.33 ± 2	72.10 ± 1	90.04 ± 0	93.37 ± 0
MobileNet-10	71.22 ± 0	86.66 ± 0	60.46 ± 1	87.93 ± 0	91.46 ± 0
Default-15	NA	NA	NA	67.00 ± 0	50.00



Fig. 6. No Rumex patches were identified by the experts and neither detected by MobileNet-10 on the 30-m-orthomosaic (no presence of blue squares neither red bounding boxes).

data was acquired on the same day, and under similar light conditions. Therefore, more research is required to assess the robustness of the method, for example, in different weather conditions, and in different latitudes. However, it should be remarked that experiments carried on 15-m-data showed that the same development methodology used to train MobileNet-10 can be used to update the model to new conditions. The reasoning behind is that MobileNet-15 was trained also with images from the 15 m flight height, which were more similar to the test dataset for that case. Therefore, the algorithm learnt some specific patterns of that dataset and consequently Rumex plants were better detected.

In the existing literature on Rumex detection, there are other approaches that achieved similar or higher performance like (van Evert et al., 2011), who managed to achieve 93.00% accuracy and (Ahmed et al., 2014), who achieved an accuracy of 98.50%. Nevertheless, these studies used close range images with spatial resolutions of around 2 mm, either by hand or with a robot, whereas this study uses UAV imagery with 5.9 mm of spatial resolution. While it is difficult to detect Rumex from a larger height, it also makes it possible to cover larger areas, which eases scalability. In another related study (Lam et al., 2021) proposed an automated open-source workflow for mapping Rumex in grasslands. However, the proposed workflow has many elements that require manual intervention, for instance, the resolution parameter of the orthophoto in the WebODM software (ODM, 2020), that needs to be set manually on a case-by-case basis. Further, they claim to provide an open-source methodology while some elements of the workflow require the use of the commercial software Agisoft Metashape. While this software can be replaced by an open-source alternative, like QGIS (QGIS Development Team, 2009), it also requires manual setting of several parameters. Furthermore, the authors do not demonstrate the generalisation ability of the model to new environments, and neither provide a repository to their code and datasets.

It took only 4 s to obtain the results displayed on Table 4 for MobileNet-10 on Field 1. This is yet not enough for a real-time application, but a promising future research direction includes exploring smaller and more efficient deep learning models and methods to map Rumex in a grassland during the UAVs flight.

Finally, while this study focused on mapping *Rumex obtusifolius* in grasslands, the same methodology should be tested to map others weeds or plants in grasslands. Moreover, future studies could focus on the detection and analysis of Rumex plants at an earlier development stage to plan ahead the weed control strategies.

6. Conclusion

This study demonstrated that it is feasible to use deep learning models on UAV imagery to map *Rumex obtusifolius* in grasslands with very limited training data. Seven different deep learning models were explored. Among those, MobileNet was the most suitable for this application, with an F1-score of 78.36%, recall of 79.76%, and an AUROC of 93.74%. Experiments showed that the model trained at a specific flight height does not directly generalize to different heights. Cross-validation made it possible to confirm that using transfer learning on a small amount of additional data can be used to adapt the model to new flying conditions. Moreover, we discovered that 5.9 mm of spatial resolution is the minimum resolution required for Rumex detection using UAV. Finally, the code and the datasets used in this work are published on an open access repository to contribute to the advances in grassland management using UAV technology.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was supported by the SPECTORS project (143081), which is funded by the European cooperation program INTERREG Deutschland-Nederland. It was also partly funded by the AIPSE programme of Academy of Finland through the AI-CropPro project; decision number 315896 (Aalto University, Finland) and 316172 (Luke, Finland). We also would like to acknowledge the computational resources provided by the Aalto Science-IT project and Corinna Roers from Naturschutzzentrum in Kreis Kleve e.V. for labelling the data.

References

- Ahmed, F., Kabir, M.H., Bhuyan, S., Bari, H., Hossain, E., 2014. Automated weed classification with local pattern-based texture descriptors. *Int. Arab J. Inf. Technol.* 11, 87–94.
- Binch, A., Cooke, N., Fox, C., 2018. Rumex and urtica detection in grassland by uav. In: 14th International Conference on Precision Agriculture.
- Bond, W., Grundy, A.C., 2001. Non-chemical weed management in organic farming systems. *Weed Res.* 41, 383–405. <https://doi.org/10.1046/j.1365-3180.2001.00246.x>.
- Brown, R.B., Noble, S.D., 2005. Site-specific weed management: sensing requirements—what do we need to see? *Weed Sci.* 53, 252–258. <https://doi.org/10.1614/WS-04-068R1>.
- Cavers, P.B., Harper, J.L., 1964. *Rumex obtusifolius* L. and *r. crispus* L. *J. Ecol.* 52, 737–766. URL <http://www.jstor.org/stable/2257859>.
- Ciresan, D.C., Meier, U., Schmidhuber, J., 2012. Multi-column deep neural networks for image classification. *CoRR*, abs/1202.2745. arXiv:1202.2745.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. ImageNet: A Large-Scale Hierarchical Image Database. In: *CVPR09*.
- van Evert, F.K., Polder, G., van der Heijden, G.W., Kempenaar, C., Lotz, L.A., 2009. Real-time vision-based detection of rumex obtusifolius in grassland. *Weed Res.* 49, 164–174. <https://doi.org/10.1111/J.1365-3180.2008.00682.X>.
- van Evert, F. K., Samsom, J., Polder, G., Vijn, M., Dooren, H.-J. v., Lamaker, A., van der Heijden, G.W., Kempenaar, C., van der Zalm, T., Lotz, L.A., 2011. A robot to detect and control broad-leaved dock (*rumex obtusifolius* L.) in grassland. *J. Field Robot.*, 28, 264–277. doi:<https://doi.org/10.1002/rob.20377>.
- Fawcett, T., 2006. An introduction to roc analysis. *Pattern Recogn. Lett.* 27, 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>. ROC Analysis in Pattern Recognition.
- Flaten, O., Atsbeha, D.M., Lunnan, T., 2020. Data to estimate costs of producing grass-clover silages. *Data in Brief* 31, 106003. <https://doi.org/10.1016/j.dib.2020.106003>.
- Gao, J., Liao, W., Nuyttens, D., Lootens, P., Vangyeyte, J., Pižurica, A., He, Y., Pieters, J. G., 2018. Fusion of pixel and object-based features for weed mapping using unmanned aerial vehicle imagery. *Int. J. Appl. Earth Obs. Geoinf.* 67, 43–53. <https://doi.org/10.1016/j.jag.2017.12.012>.
- Gebhardt, S., Kühbauch, W., 2007. A new algorithm for automatic rumex obtusifolius detection in digital images using colour and texture features and the influence of image resolution. *Precision Agric.* 8, 1573–1618. <https://doi.org/10.1007/s11119-006-9024-7>.
- Gebhardt, S., Schellbreg, J., Lock, R., Kühbauch, W., 2006. Identification of broad-leaved dock (*rumex obtusifolius* L.) on grassland by means of digital image processing. *Precision Agric.* 7, 1573–1618. <https://doi.org/10.1007/s11119-006-9006-9>.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. <http://www.deeplearningbook.org>.
- Hamylton, S., Morris, R., Carvalho, R., Roder, N., Barlow, P., Mills, K., & Wang, L. (2020). Evaluating techniques for mapping island vegetation from unmanned aerial vehicle (uav) images: Pixel classification, visual interpretation and machine learning approaches. *International Journal of Applied Earth Observation and Geoinformation*, 89, 102085. doi:<https://doi.org/10.1016/j.jag.2020.102085>.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.
- Hiremath, S., Tolpekin, V.A., van der Heijden, G., Stein, A., 2013. Segmentation of rumex obtusifolius using gaussian markov random fields. *Mach. Vis. Appl.* 24, 845–854. <https://doi.org/10.1007/s00138-012-0470-0>.
- Holm, L., Plucknett, D., Pancho, J., Herberger, J., 1977. *Rumex crispus* and *obtusifolius*. the world's worst weeds: distribution and biology. University Press of Hawaii. Honolulu, USA, pp. 401–408.
- Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708.
- Kamilaris, A., Prenafeta-Boldú, F.X., 2018. Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture* 147, 70–90. <https://doi.org/10.1016/j.compag.2018.02.016>.
- Kristálová, V., Hejzman, M., Cervená, K., Pavlu, V., 2011. Effect of nitrogen and phosphorus availability on the emergence, growth and over-wintering of rumex crispus and rumex obtusifolius. *Grass and Forage Science* 66, 361–369.
- Kounalakis, T., Triantafyllidis, G.A., Nalpanitidis, L., 2019. Deep learning-based visual recognition of rumex for robotic precision farming. *Computers and Electronics in Agriculture* 165, 104973.
- Lam, O.H.Y., Dogotari, M., Prüm, M., Vithlani, H.N., Roers, C., Melville, B., Zimmer, F., Becker, R., 2021. An open source workflow for weed mapping in native grassland using unmanned aerial vehicle: using rumex obtusifolius as a case study. *European Journal of Remote Sensing* 54, 71–88. <https://doi.org/10.1080/22797254.2020.1793687>.
- Liu, Y.-F., Liu, Y., Shi, Z.-H., López-Vicente, M., Wu, G.-L., 2020. Effectiveness of re-vegetated forest and grassland on soil erosion control in the semi-arid loess plateau. *CATENA* 195, 104787. <https://doi.org/10.1016/j.catena.2020.104787>.
- Lottes, P., Hoferlin, M., Sander, S., Müter, M., Schulze, P., & Stachniss, L.C. (2016). An effective classification system for separating sugar beets and weeds for precision farming applications. In *IEEE International Conference on Robotics and Automation (ICRA)* (pp. 5157–5163). doi:10.1109/ICRA.2016.7487720.
- Ma, N., Zhang, X., Zheng, H.-T., & Sun, J. (2018). ShuffleNet v2: Practical guidelines for efficient cnn architecture design. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 116–131).
- Mosyakin, S.L. (2005). *Flora of North America* volume 5 of Magnoliophyta: Caryophyllidae. Oxford University Press.
- ODM, O.A. (2020). OpenDroneMap. A command line toolkit to generate maps, point clouds, 3D models and DEMs from drone, balloon or kite images. URL: <https://github.com/OpenDroneMap/ODM>.
- Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., & Lerer, A. (2017). Automatic differentiation in pytorch.
- QGIS Development Team (2009). QGIS Geographic Information System. Open Source Geospatial Foundation. <http://qgis.osgeo.org>.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., & Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. arXiv:1409.0575.
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4510–4520).
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
- Simonyan, K., & Zisserman, A. (2015). Very deep convolutional neural network based image classification using small training sample size. In *3rd IAPR Asian Conference on Pattern Recognition, ACPR 2015, Kuala Lumpur, Malaysia, November 3-6, 2015* (pp. 730–734). IEEE. doi:10.1109/ACPR.2015.7486599.
- Smit, H., Metzger, M., Ewert, F., 2008. Spatial distribution of grassland productivity and land use in europe. *Agric. Syst.* 98, 208–219. <https://doi.org/10.1016/j.agsy.2008.07.004>.
- Smith, L.N. (2017). Cyclical learning rates for training neural networks. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 464–472). IEEE.
- Tan, M., Chen, B., Pang, R., Vasudevan, V., Sandler, M., Howard, A., & Le, Q.V. (2019). Mnasnet: Platform-aware neural architecture search for mobile. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2820–2828).
- Tan, M., & Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning* (pp. 6105–6114). PMLR.
- Toole, E., Brown, E., 1946. Final results of the duvel buried seed experiment. *Journal of Agricultural Research* 72, 201–210.
- Valente, J., Dolderum, M., Roers, C., Kooistra, L., 2019. Detecting rumex obtusifolius weed plants in grasslands from uav rgb imagery using deep learning. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences IV-2/W5*, 179–185. <https://doi.org/10.5194/isprs-annals-IV-2-W5-179-2019>.
- Valente, J., & Kooistra, L. (2021). Dataset on uav high-resolution images from grassland with broad-leaved dock (*rumex obtusifolius*). doi:10.5281/zenodo.5119205.
- Wang, A., Zhang, W., Wei, X., 2019. A review on weed detection using ground-based machine vision and image processing techniques. *Computers and Electronics in Agriculture* 158, 226–240. <https://doi.org/10.1016/j.compag.2019.02.005>.
- Yang, Y., Tilman, D., Furey, G., Lehman, C., 2019. Soil carbon sequestration accelerated by restoration of grassland biodiversity. *Nature Communications* 10. <https://doi.org/10.1038/s41467-019-08636-w>.
- Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? In *Advances in neural information processing systems* (pp. 3320–3328).