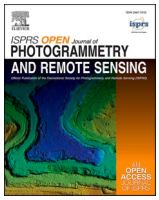
ISPRS Open Journal of Photogrammetry and Remote Sensing 12 (2024) 100064

Contents lists available at ScienceDirect

ISPRS Open Journal of Photogrammetry and Remote Sensing

journal homepage: www.journals.elsevier.com/isprs-open-journal-of-photogrammetry-and-remote-sensing



Improving spatial transferability of deep learning models for small-field crop yield prediction

Stefan Stiller [a,b], Kathrin Grahmann [a], Gohar Ghazaryan [a,c], Masahiro Ryo [a,b,*]

[a] Research Platform "Data Analysis & Simulation", Leibniz Centre for Agricultural Landscape Research (ZALF), Eberswalder Straße 84, 15374, Müncheberg, Germany
[b] Environment and Natural Sciences, Brandenburg University of Technology Cottbus-Senftenberg (BTU-CS), Platz der Deutschen Einheit 1, 03046 Cottbus, Germany
[c] Geography Department, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany

ARTICLE INFO

Keywords:
Smallholder farming
UAV
Crop yield
Convolutional neural network
Spatial cross validation
External validation

ABSTRACT

Predicting crop yield using deep learning (DL) and remote sensing is a promising technique in agriculture. In smallholder agriculture (<2 ha), where 84% of the farms operate globally, it is crucial to build a model that can be useful across several fields (high spatial transferability). However, enhancing spatial model transferability in a small-scale setting faces significant challenges, including spatial autocorrelation, heterogeneity and scale dependence of spatial dynamics, as well as the need to address limited data points. This study aimed to test the hypothesis that spatial cross validation (SCV) is a more suitable model validation practice than random cross validation (RCV) to enhance model transferability for spatial prediction in a small-scale farming setting. We compared the performances of DL models that predict crop yield for several settings including three crop types and two DL architectures based on RCV with and without overlapping samples and SCV. Notably, we conducted model performance tests on external, equally sized fields instead of the field used for training. We used high resolution RGB imagery taken with a drone as input. Our results show that the models using SCV outperformed those using RCV when the models were tested on external fields (on average r = 0.37 for SCV, r = 0.18 for RCV with overlap and r = 0.07 without), even though the models using SCV showed a substantially lower performance for cross validation (CV) than those using RCV (r with SCV and RCV w/o overlap = 0.73 and 0.98/0.73, respectively). The results suggest that RCV leads to over-optimism by overfitting the spatial structure and remembering image-specific information (so called memorization). Our study offers the first empirical evidence in agriculture that SCV is preferable to RCV in small field settings for making DL models more transferable.

1. Introduction

Deep learning (DL) models with computer vision (e.g. proximal and remote sensing) have been widely applied in agriculture (Kamilaris and Prenafeta-Boldú, 2018). Examples of DL applications include land cover and crop type mapping (Kussul et al., 2017), crop yield estimation (Kuwata and Shibasaki, 2015; Nevavuori et al., 2019; Maimaitijiang et al., 2020), drought (Shen et al., 2019) or plant disease spread (Tetila et al., 2020) and monitoring. DL applications can improve agricultural practices across scales ranging from individual organism, field, landscape, to regional and continental scales (Ryo et al., 2022).

Ideally, DL models in agriculture are transferable, meaning that they maintain high predictive accuracy in new settings, such as future years, adjacent fields, and different management practices. The most (84%) of the 540 million farms globally are smallholdings (<2 ha) (Lowder et al., 2016), and crop rotation does not keep the same crop type in the same field. These conditions require to enhance model spatial transferability. Spatial transferability refers to the ability of a trained model to perform well on data beyond the training region (Zhang et al., 2020b). The spatial transferability of DL based prediction models in agriculture was assessed for grassland land-use intensity mapping (Lange et al., 2022), corn and soybean mapping (Xu et al., 2020), or cropland and land cover classification (Zhang et al., 2020a). However, previous studies have been carried out with large field sizes, and little is known about how well a DL model can be spatially transferable in smallholder settings.

For model performance assessment, spatial cross validation (SCV) is a recommended practice over random cross validation (RCV). RCV estimates the ability to make accurate predictions on new, unseen data,

* Corresponding author. Research Platform "Data Analysis & Simulation", Leibniz Centre for Agricultural Landscape Research (ZALF), Eberswalder Straße 84, 15374, Müncheberg, Germany.
E-mail address: Masahiro.Ryo@zalf.de (M. Ryo).

https://doi.org/10.1016/j.ophoto.2024.100064
Received 18 December 2023; Received in revised form 17 April 2024; Accepted 19 April 2024
Available online 23 April 2024

following the *i.i.d.* assumption (Stone, 1974). Spatial data, however, have an underlying structure – spatial autocorrelation – that makes data points have more similar values, the closer they are (Tobler, 1970). Spatial autocorrelation violates the assumption of data independence between training, validation, and test set and hence leads to inflated prediction performance inference (Le Rest et al., 2014; Roberts et al., 2017; Ploton et al., 2020; Kattenborn et al., 2022). SCV can reduce over-optimism for model performance assessment. However, to our best knowledge, no previous study has tested if a DL model trained with SCV performs better than RCV in different agricultural fields where they have different management and soil characteristics.

The proportion of data split for train and test sets is crucial for testing model spatial transferability in small-scale farms. Typically, modeling over large areas can be less sensitive to unequal data splits (e.g., 90:10 or 80:20), while covering sufficient data distributions in both training and testing. However, in smaller areas like smallholder farms, one must carefully balance the proportion of train and test data to ensure the test area captures similar spatial dynamics and scale-dependent patterns as the training data distribution. Different proportions can represent different spatial heterogeneity (Tittonell, 2023) and scale dependent dynamics (Heydari et al., 2023). Scale dependent dynamics in agriculture include for example soil heterogeneity, microclimate (van Wijk, 1965) and biogeochemical properties (Patzold et al., 2008), and management heterogeneity (Shah and Wu, 2019). Furthermore, the challenge of model transferability is amplified by the scarcity of data in smallholder settings, which makes proper model training more challenging (Safonova et al., 2023).

The aim of this study is to examine the effect of model validation techniques on the spatial transferability of DL models for crop yield prediction using multiple crop types in a small-scale farming setting. We applied two convolutional neural network (CNN) architectures for predicting crop yield of three crop types (soybean, maize, and sunflower) using UAV-based RGB images. It is noteworthy that testing multiple crop types and model architectures has rarely been done in previous studies but can enhance the generality of the test. The study site is embedded in a diversified agricultural landscape setting (Grahmann et al., 2024), with high soil heterogeneity, small field sizes and thus, low sample size. We hypothesize that SCV achieves a higher prediction performance than RCV when the models are used on another field site because SCV can alleviate model overfitting.

2. Methods

2.1. Research site and data collection

2.1.1. Study site: landscape experiment patchCROP

The data was collected at the agricultural landscape experiment "patchCROP" (Grahmann et al., 2024), located in the federal state of Brandenburg Germany (70 ha size; lat: 14.141348, long: 52.447421) that has been established in 2020. The experimental site is composed of 30 small field arrangements of each ~0.5 ha (72m × 72m/60m; Figs. 1 and 3). Each field has a unique treatment combination in terms of site-specific crop rotation, soil quality, and management practices including conventional/reduced chemical-synthetic pesticide use and with/without flower strip implementation at the field edge. Thereby, patchCROP aims to study diversified agricultural landscapes in the form

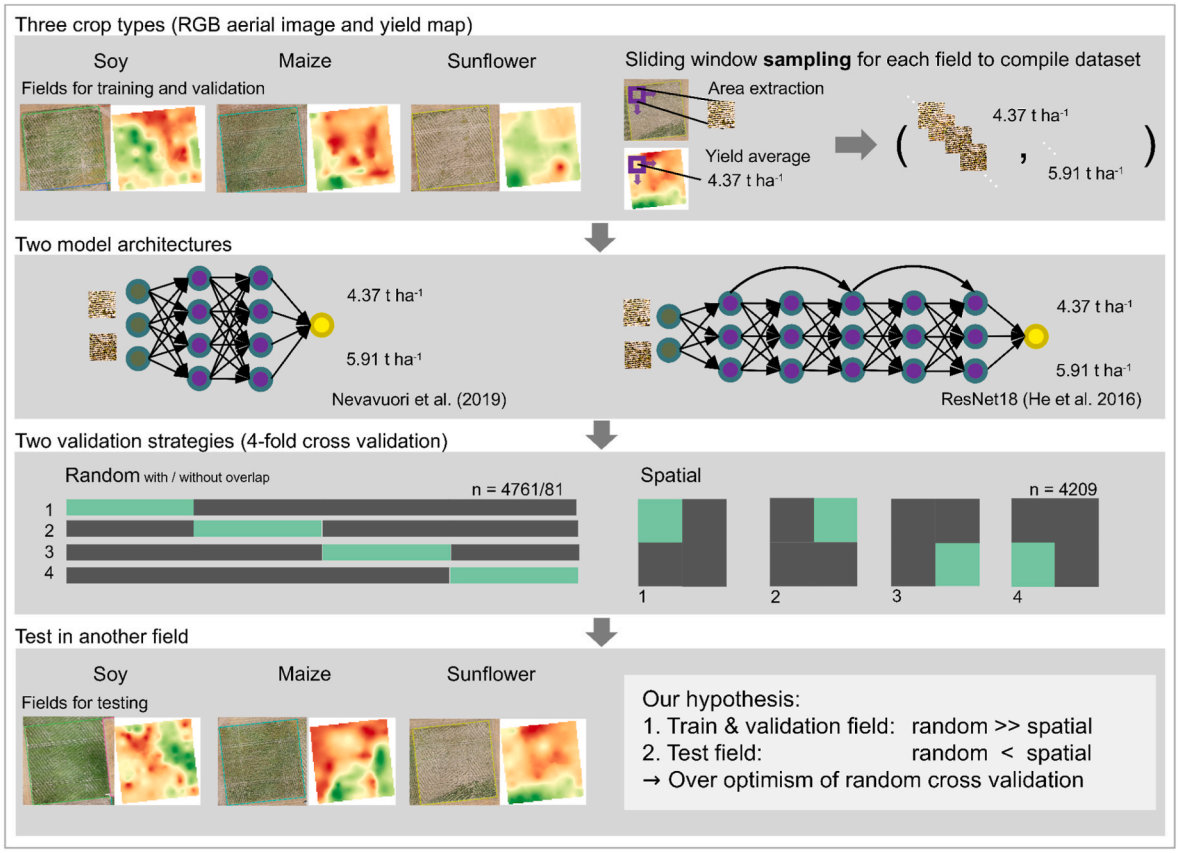


Fig. 1. Conceptual framework for testing different model validation strategies for an honest model assessment of deep learning based crop yield mapping models that utilize remote sensing images to predict crop yield of three crop types: soy, maize and sunflower. We tested different validation strategies (random and spatial cross validation) with model assessment in an unseen area of the same field where the model has been trained (validation) as well as in an external field of that same crop type (test). Two architectures have been tested: ResNet18 and a reimplementation from the literature. All field recordings as well as yield maps have been sampled using a sliding window approach. We hypothesized that random cross validation (RCV) over-optimistically outperforms spatial cross validation (SCV) on the field used for training and validation, but when testing on an external field site SCV is better than RCV.

of new field arrangements, namely patch cropping, and their effects on yield stability, farming system resilience, ecosystem services and biodiversity. This study selected three summer crop types from six field arrangements in the cropping season 2020: soy, maize, and sunflower - for each crop type two fields each. The fields covered different land use intensities, and depict very heterogeneous soil conditions with varying soil texture and topography (Grahmann et al., 2024). Land use intensities comprised: business as usual with conventional pesticide application (soy, maize and sunflower); reduced pesticide application with additional 12m wide flower strips next to the field (soy); and solely reduced pesticide application (maize and sunflower). The selected fields covered high and low yield potential zones (Donat et al., 2022), with both soy fields located in the high yield potential zone, and maize and sunflower fields in the low yield potential zones. Note that we did not make use of all 30 field arrangements because some patches had been harvested already.

2.1.2. UAV RGB image data acquisition

We acquired high-resolution UAV-based RGB images. RGB images were taken with a senseFly-eBee X drone mounted camera (senseFly-S. O.D.A.) on August 6th, 2020, which incorporates an advanced GPS correction system that operates based on real-time kinematic (RTK) and post-processed kinematic (PPK) technologies, allowing accurate georeferencing. The overflight was performed at 84 m height with an average ground sampling distance of 2.22 cm and recorded in the EPSG 25833 coordinate system. Image acquisition dates were synchronized with the phenological growth stages of plants as defined by the BBCH scale (Biologische Bundesanstalt für Land-und Forstwirtschaft, Bundessortenamt und CHemische Industrie; Lancashire et al., 1991). BBCH scale employs a standardized two-digit coding system to categorize phenologically akin plant growth stages. Soy was in late fruit development/early ripening of fruit and seed (BBCH: 79–80), maize in development of fruit (BBCH: 75), and sunflower in ripening of fruit and seed (BBCH: 85). The recording was taken at 10 a.m. for approximately 1 h. Weather conditions were optimal with no strong wind, no rain, or sunny.

2.1.3. Crop yield data acquisition

Soybean was harvested with Claas Lexion 6900 and 10.5m width, maize was harvested with Claas Lexion 770 TT at 6m width; sunflower was harvested with 9m cutting system and Claas Lexion 770 TT. Their crop yields were recorded as geospatially registered point data (Diker et al., 2004; Florin et al., 2009). For each field arrangement between 100 and 300 points were recorded with approx. 1.8m distance in between and a harvester width of 11.5m.

2.2. Data pre-processing

2.2.1. RGB image pre-processing

The landscape orthoimage was computed using Pix4Dmapper 4.5.6 (2020) with automatic keypoint extraction (76300 keypoints per image). Four image tiles have been disabled for orthomosaic computation. The orthomosaic was computed with 5+ overlapping images for each pixel resulting in high pixel fidelity. The mean reprojection error was 0.121 pixels. We confirmed visually that the orthoimage is not affected by major artifacts such as honeycomb artifacts due to pivotable orientation of the camera and low altitude of the sun. The image was visually inspected and no major artifacts were present. We clipped the rectangular shaped fields from the landscape recording (GDAL 3.4.2). Each subset image of the fields has a resolution of 2977x2977 pixels for fields measuring 72 × 72 m, and 2977x2480 pixels for fields measuring 72x60 m.

2.2.2. Crop yield map Cleaning and kriging interpolation

Yield points collected from combine harvesters typically contain a number of defective observations or technical errors that need to be removed to ensure an accurate representation of the ground truth (Arslan and Colvin, 2002). This is especially important in a high-resolution setting at the target scale, when there are only a small number of yield recordings. However, these recordings need to reliably capture the spatial heterogeneity, unlike in low-resolution yield mapping at a coarser spatial scale where a single yield point is less important. The following procedures were taken to reduce the error in yield observations. According to Lyle et al. (2014) yield point errors can be classified among others as related to harvesting dynamics of the combine harvester or to the harvester operator. We removed the erroneous yield points systematically by discarding underestimated yield points at the beginning and end of a harvest path according to (Blackmore, 1999). Lag-time was automatically accounted for by the combine harvester. Operator errors related to speed changes (Arslan and Colvin, 2002) or turns in harvest paths (Lyle et al., 2014) were accounted for by removing the points within the lower ten percentile of grain flow. We also removed the points crossing and intersecting harvest trajectories through visual inspection. We adjusted all remaining recorded yield values to standard moisture content (Mulvaney and Devkota, 2020) (standardized crop moisture content: soy = 13%, maize = 15.5%, and sunflower = 10%). We transformed the coordinate system of the yield point data to match the coordinate system of the UAV-taken image (EPSG 25833). Finally, we spatially interpolated the yield points using ordinary kriging (Cressie, 1988). Hereafter, we refer to the spatially ordinary kriging interpolated yield points as yield map.

2.2.3. Sliding window to compile the datasets

We compiled two datasets for model training and validation by using a sliding window algorithm (Li et al., 2017; Valente et al., 2022) on the RGB image. One dataset contained overlapping samples the other not. The samples that were clipped at each respective position of the sliding window had a resolution of 224x224 pixels (~5.4m × 5.4m) and were shifted with a stride of 30 pixels (~0.73m) or for the other set by 224 pixels (~5.4m). For each of the subset images, an average yield was calculated for the respective sample of the crop yield map. Altogether, we compiled two distinct data sets for each field in which each sample is a tuple that consisted of a remote sensing observation and the average yield within this area. Applying the sliding window amounted to a total of 4761 samples per field for the data set with overlap between samples and 81 samples without overlap.

2.3. Deep learning model architecture

We implemented CNN models for predicting crop yield using the UAV-taken RGB image for each crop type of soy, maize, and sunflower independently.

We tested and compared two architectures, a ResNet18 (He et al., 2016) and a re-implementation of Nevavuori et al. (2019) as baseline model. To perform the task of crop yield prediction using images (i.e. image regression), we exchanged the last fully connected layer of the ResNet18 by one linear layer with rectified linear unit (ReLu) activation (Fukushima, 1975). The implementation by Nevavuori consisted of two fully connected layers with ReLu activation.

All models have been implemented in Python v3.8 language using the DL framework PyTorch v1.13.0. For model training we used a high performance cluster with four Nvidia Tesla V100.

2.4. Model training, validation and Test

2.4.1. Random and spatial cross validation

We tested and compared DL based yield mapping models using 4-fold RCV and SCV and assessed model transferability externally for another field of that same crop type (see *Testing Model Transferability* as well as Figs. 2 and 3). Therefore, we split the data set in four equally sized data partitions and dedicated one alternating partition as a validation set while combining the rest as a train set in successive iterations. For SCV

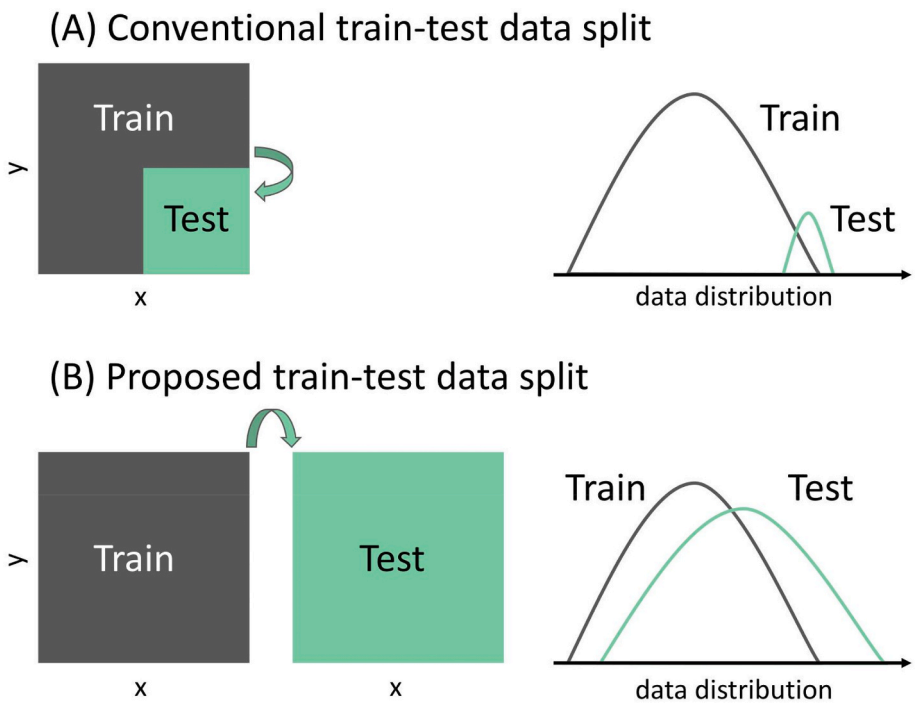


Fig. 2. Proposed scheme for model evaluation. (A) Train-test data split that is typically employed. Within the same single field, the data is split into train and test sets with an uneven proportion, e.g. 80:20. If the test set is randomly sampled across the space, the spatial autocorrelation makes the model overfit. If the test set is spatially blocked, the spatial autocorrelation results in a much narrower data distribution in the test set than that of the train set. Moreover, it does not test model transferability. (B) Proposed train-test data split that prepares the equal proportion and spatial extent of train and test sets can alleviate these issues.

we used the data set with overlap and partitioned the data using spatial blocking, whereas for RCV we used random sampling on the data sets with and without overlap. We removed all samples with any overlap between the training and validation set for SCV to prevent data leakage. Note that we did not remove overlapping samples between training and validation set for RCV as the train set size would become too small for effective model training. Rather we used the data set that contained non-overlapping samples in order to achieve RCV without data leakage.

After removing overlapping samples in the SCV, the combined training set comprised 3279 and the validation set 930 samples, respectively. For RCV with/without overlap, the training set had a size of 3570/60 and the validation set counted 1191/21 samples.

2.4.2. *Augmentations for training samples to facilitate learning*

Image augmentation is a powerful technique in remote sensing (Yu et al., 2017; Chen et al., 2022) and agriculture, for instance for crop-weed classification (Su et al., 2021; Divyanth et al., 2022). We used geometric and color space augmentation, expecting to prevent overfitting (for instance to crops rows) and improve model generalizability via learning more meaningful feature representations rather than data point storage (Scott et al., 2017; Shorten and Khoshgoftaar, 2019). We applied randomly selected geometric augmentations ($p = 0.5$) to each input image used for model training out of a predefined set of augmentations. We used the following geometric transformations in order to prevent overfitting to spatial features like crop rows and traffic lanes: 90° rotation, vertical and horizontal flip, transpose; and brightness to increase robustness for different lighting conditions. Augmentations have been applied using python package 'albumentations' v1.3.0 (Buslaev et al., 2020).

2.4.3. *Model hyper-parameter tuning and model fitting*

Model hyper-parameters regulate a model's learning ability and need to be considered carefully. We used 'flat cross validation' (Wainer and Cawley, 2021), in which the validation set is used for both hyper-parameter tuning and model selection.

We tuned the hyper-parameters 'learning rate', 'weight decay' and 'batch size' on the validation set with Bayesian optimization (Falkner et al., 2018) in 20 trials using the Ray 2.3.0 framework (Liaw et al., 2018). 'Learning rate' was sampled from a log uniform distribution in a range $[10^{-6}, 10^{-1}]$, 'weight decay' uniformly in an interval $[0, 5 * 10^{-3}]$ and for 'batch size' discrete values were selected: 4, 8, 16, 32, 64, 128, 256, and 512.

For model training, we employed Stochastic Gradient Descent with Warm Restarts (SGDR) alongside a cosine annealing learning rate schedule to achieve superior anytime performance and faster convergence compared to other methods, as proposed by Loshchilov and Hutter (2017). We meticulously tuned the SGDR-specific hyper-parameters within the optimization framework previously outlined. These parameters include the number of epochs before each learning rate schedule restart ($T_0 = [1, 10, 50, 100, 200]$) and the multiplication factor for the learning rate increase at each restart ($T_{mult} = [1, 2]$).

Batch sizes affect training speed, gradient stability and regularization, yet options are limited by memory capacity and demand. Thus, we used Automatic Mixed Precision (Micikevicius et al., 2018) to perform some calculations in lower precision arithmetic, while preserving the accuracy of the final result.

All models have been trained from scratch by minimizing the mean squared error between observations and predictions. L2 regularization was applied to the error term using the tuned weight decay hyper-parameter. We trained the models using mini-batch stochastic gradient descent (SGD; Ruder, 2017) with momentum 0.9 for 1000 epochs for data with overlap, and 2000 epochs without overlap to counterbalance reduced sample size.

2.4.4. *Testing model performance in the field of training/validation vs. model transferability*

We assessed model performance i) by using cross validation (CV) on the field of training/validation, and ii) by testing externally on another field as depicted in Figs. 1 and 2. To evaluate model prediction performance in the field of training/validation we reported 'local' and 'global'

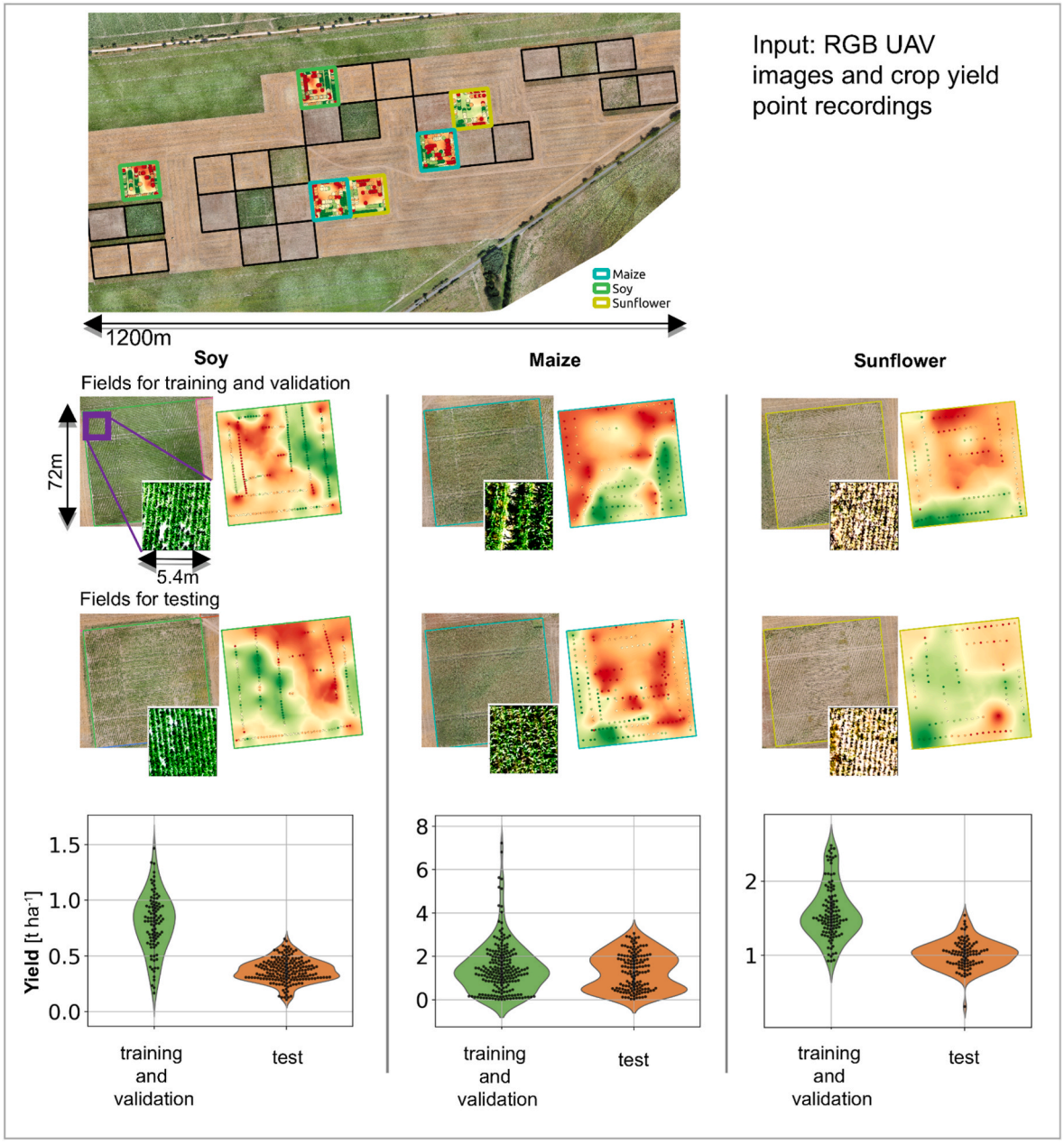


Fig. 3. Dataset from patchCROP landscape experiment in the federal state of Brandenburg, Germany comprising a high-resolution aerial RGB image of small field arrangements and crop yield recordings from which we selected a subset of six fields. The middle row shows the zoomed-in field view of drone imagery with actual training samples (left) and yield point recordings with ordinary kriging interpolated yield maps for the respective crop types: soy, maize and sunflower. Red indicates low yield, green high yield zones. For model training and validation, the same field is used (upper image), whereas model performance is tested on another field of that same crop type (lower image). Crop yield distributions for both fields are shown at the bottom. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

prediction performance using Pearson's correlation coefficient (r) (Meyer et al., 2019) with the models using the distinct type of CV. To assess local prediction performance, we calculated the Pearson's correlation coefficient for predicted vs. observed yields within each fold and then reported the median of these coefficients across all folds (local r). Whereas, the global r was computed from all predictions made across the folds versus the observed values, treating the entire set of predictions as one large dataset. Therefore, we tested both the generalizability of the model across different subsets (local) and the model's overall ability to predict across the entire dataset (global). For external model assessment (i.e. model transferability), the average prediction of the CV fold's model prediction was reported. Combinedly, global performance on the field of training/validation and external assessment allowed a holistic assessment of model predictive capabilities over entire fields. The former was used for assessing overall performance on familiar data, and the latter employed to test the model's transferability and robustness to new, unseen conditions.

3. Results

We trained the models over 1000/2000 epochs based on root mean squared error (RMSE). In Fig. 4 we showcase representative model training and selection of a ResNet18 architecture that predicted maize yield using 4-fold random and spatial CV (for other crop types and architecture, see Figs. S1–S5). RCV fitted the models more tightly to the training data than SCV, with RCV with overlapping samples having the

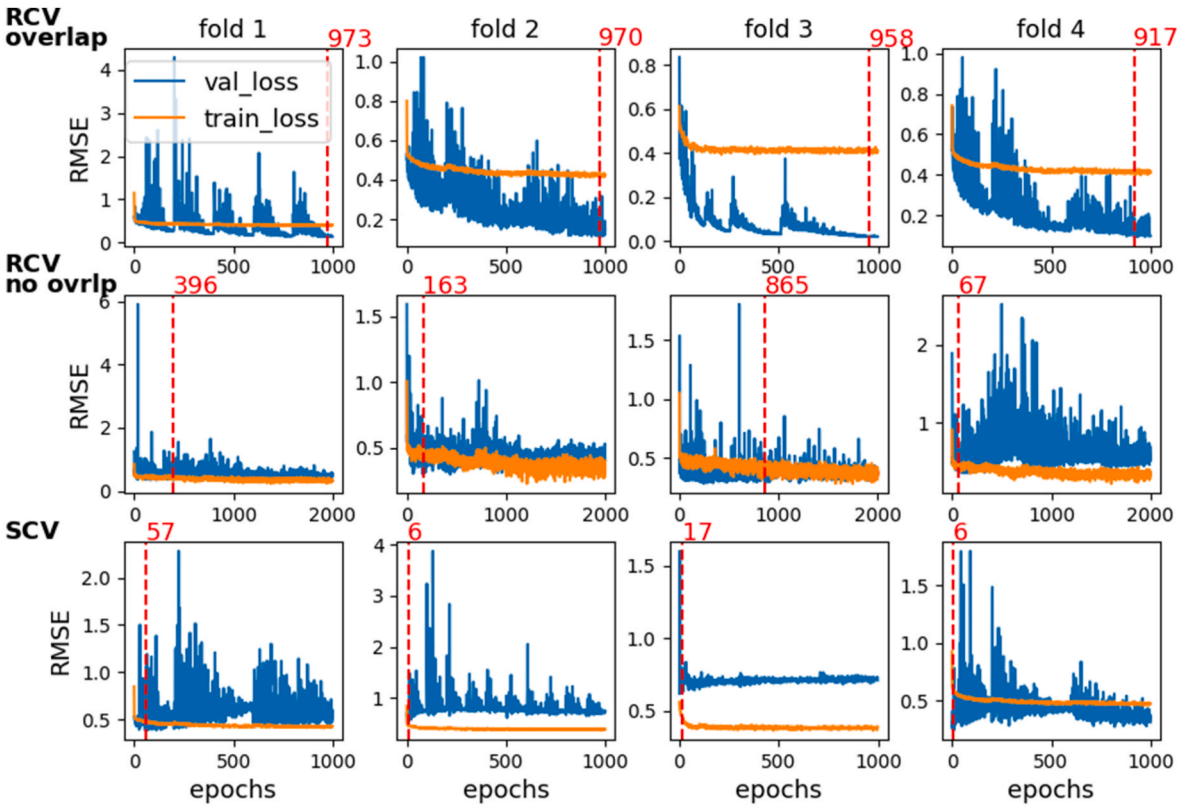


Fig. 4. Training and validation loss for 4-fold random (RCV) with and without overlapping samples and spatial (SCV) cross validation for ResNet18 of maize yield prediction shown as root mean squared error (RMSE). Models that use overlapping samples are trained for 1000 epochs (i.e. training cycles through the entire dataset). Models that use non-overlapping samples are trained for 2000 epochs to counterbalance reduced sample size. Red dashed lines and the associated numbers indicate the epoch of model selection in which the model has the tightest fit to the within-field validation set. In the Supplementary Information the equivalent results for the other crop types and the other model architecture are available. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

tightest fit. The models with the tightest fit to the validation set over all epochs were selected as prediction models (indicated by the red dashed line and number). Epochs of model selection occurred latest for RCV with overlap, earlier for RCV without overlap, and earliest for SCV (917–973 for RCV overlap, 67–865 for RCV non-overlap, 6–57 for SCV).

RCV with overlap resulted in global Pearson's r above 0.98 (i.e. correlation for prediction over all folds) for predicting both the training as well as validation sets (Fig. 5). No remarkable distinction between crop types was observed (Fig. S6). RCV without overlap achieved lower global performance on training (0.68 ± 0.1) and validation set (0.73 ± 0.08) with minor differences between crop types and architectures (Table 1). Conversely, SCV resulted in lower global r for predicted yields on the training (0.58 ± 0.13) and validation set (0.73 ± 0.05) with minor differences between crop types. When we tested prediction on another field (test set), however, we observed a drop in prediction performance. On average both RCV strategies revealed poorer prediction performance than SCV for predicting maize and soy on the test set. For predicting sunflower RCV and SCV performed similarly low (Table 1). The baseline models showed similar behavior on the train and validation sets (Fig. S6).

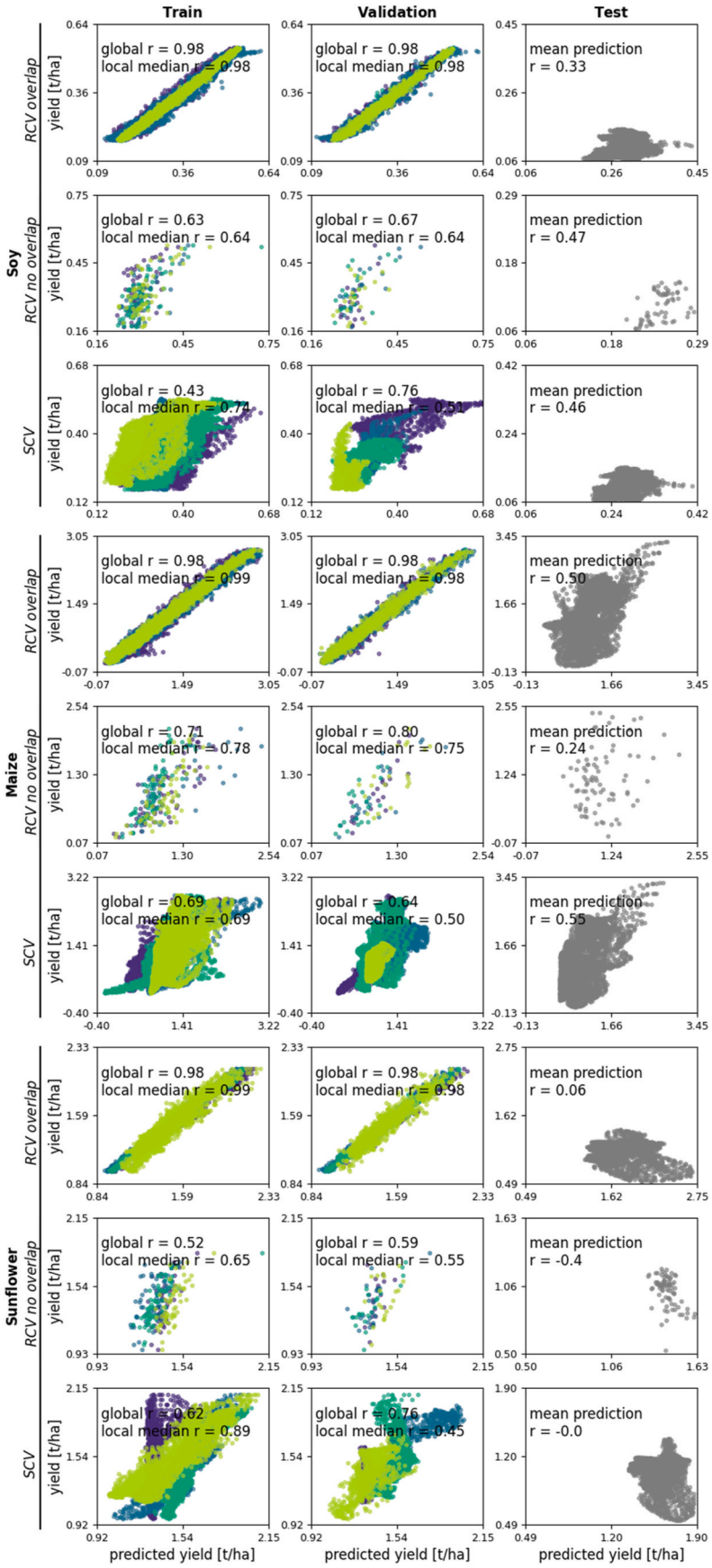
By aggregating all results of 3 crop types and 2 model architectures (Fig. 6), the average model performance supports our hypothesis where SCV achieves higher prediction performance than both RCV strategies when the model is used on another field site (mean r = 0.18 for RCV with overlap, mean r = 0.07 for RCV without overlap, mean r = 0.37 for SCV). Additionally, RCV with overlap made the model performances close to zero standard deviation for the train and validation set, but high deviation for the test set (SD = 0.37). RCV without overlap showed moderate standard deviation on train (SD = 0.11) but high on test (SD = 0.31), while SCV had moderate standard deviation for performances on both train and test set (SD = 0.13 and 0.21, respectively). This may indicate RCV lets the models overfit to the data more than SCV.

4. Discussion

This study aimed to investigate the effect of model validation techniques on the transferability of crop yield prediction models for multiple crop types in a small-scale farming setting that use CNNs and RGB remote sensing images. For that purpose, we i) tested RCV and SCV as model training and validation techniques for DL based crop yield mapping models, and ii) evaluated them by testing externally on another field. We demonstrated that the models using SCV achieved on average a higher performance score and less standard deviation than using RCV on external test sites.

Models using SCV showed an improved prediction for settings beyond the training data than the ones using RCV, which indicates higher *transferability*. As seen in Fig. 3, the data distributions in the external test fields differed from those in the training fields. For instance, in a test field crops could visually differ from crops in the training field and have increased yield due to different soil and management conditions (Herrmann et al., 2020). Hence, devising training and testing data to be equally sized is an important step for model performance tests so that the data sets cover the dynamics at the same spatial scale.

Our findings are especially important for applications, where fields have small spatial extent and are highly heterogeneous (c.f. small data (Safonova et al., 2023),). For example, these settings can be found in diversified agricultural systems with smaller field size that aim to design sustainable cropping systems of the future (Grahmann et al., 2024), as



(caption on next page)

Fig. 5. Observed vs. predicted yield on training, validation and external test set using ResNet18 architecture and color coded 4-fold random (RCV) with and without overlap between samples and spatial (SCV) cross validation for crops soy, maize, and sunflower. For each crop type, the upper row shows the results with RCV with overlap, the middle RCV with non-overlapping samples, and the lower row shows ones with SCV. Global r specifies Pearson's correlation coefficient of predictions and observations across all folds of the respective cross validation, whereas local median r shows the median r of the folds. For external model assessment we report the average Pearson's correlation coefficient across the folds of the respective cross validation models. In the supplementary material section we provide the same figure for the baseline model. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 1
Yield prediction performance on training, validation and external test set using ResNet18 architecture trained and validated using 4-fold random (RCV) with and without sample overlap and spatial (SCV) cross validation for crops soy, maize, and sunflower. Train and validation sets mutually span the same field, whereas the test set is on an external field. Performances on train/validation sets are indicated by global r - the global Pearsson's r over every data point predicted during CV. Test set performance is the average prediction of the CV fold's model prediction.

		ResNet18			Baseline		
		Train	Val	Test	Train	Val	Test
Soy	RCV over.	0.98	0.98	0.33	0.98	0.98	−0.5
	RCV no-ov.	0.63	0.67	0.47	0.85	0.77	−0.1
	SCV	0.43	0.76	0.46	0.42	0.79	0.5
Maize	RCV over.	0.98	0.98	0.5	0.98	0.99	0.47
	RCV no-ov.	0.71	0.8	0.24	0.71	0.8	0.22
	SCV	0.69	0.64	0.55	0.75	0.7	0.47
Sunflower	RCV over.	0.98	0.98	0.06	0.98	0.97	0.2
	RCV no-ov.	0.52	0.59	−0.4	0.68	0.77	0
	SCV	0.62	0.76	0	0.59	0.78	0.22

well as in smallholder farming. Both applications have similar physical structures of small field size and diversified crop production, making our proposed approach applicable for both settings. Typically, DL models require a lot of data to train (cf. big data; Cho et al., 2016; Henighan et al., 2023). Small data with high heterogeneity from model training is a challenge that has rarely been engaged in agriculture and other domains (Safonova et al., 2023). Our results suggest that SCV is a promising training and validation approach for making a DL model more transferable and by that more readily accessible. However, our study's relevance extends beyond the agricultural domain and enhances model generalization, demonstrates testing transferability and showcases how external testing can help to prevent positively biased performance assessments. This can for instance, but not exclusively, be in remote sensing research domains that use DL based prediction models, such as land cover and land use classification (Sefrin et al., 2021), environmental monitoring (Yuan et al., 2020), and change detection over a small spatial extent.

In order for a model to be transferable, a crop prediction model would require to generalize from what it learned from the dynamics and concepts in the training field. Larger test fields can be more heterogeneous. Thus, models are required to be more robust (Stone, 1974) and have higher generalization (Djolonga et al., 2021). Here, we observed that models using SCV outperformed those using RCV of any overlap strategy in an external test field as opposed to the reversed performance order in the training field. Fields for training/validation and testing visually differed quite a lot and had significantly different yield distributions. For soy for instance, the test field looked drier, plants were less green and sparser and the average yield was much smaller as well as its distribution much narrower. That SCV achieved higher prediction performance on the test set than RCV despite the described shift in the data, indicates that the models using SCV learned more meaningful features for better generalization, are more robust and hence more transferable.

Moreover, it suggests that models using RCV overfit the training data more than SCV and have *inflated performance* (i.e. reported performance is biased), which is in line with findings of previous studies (Le Rest

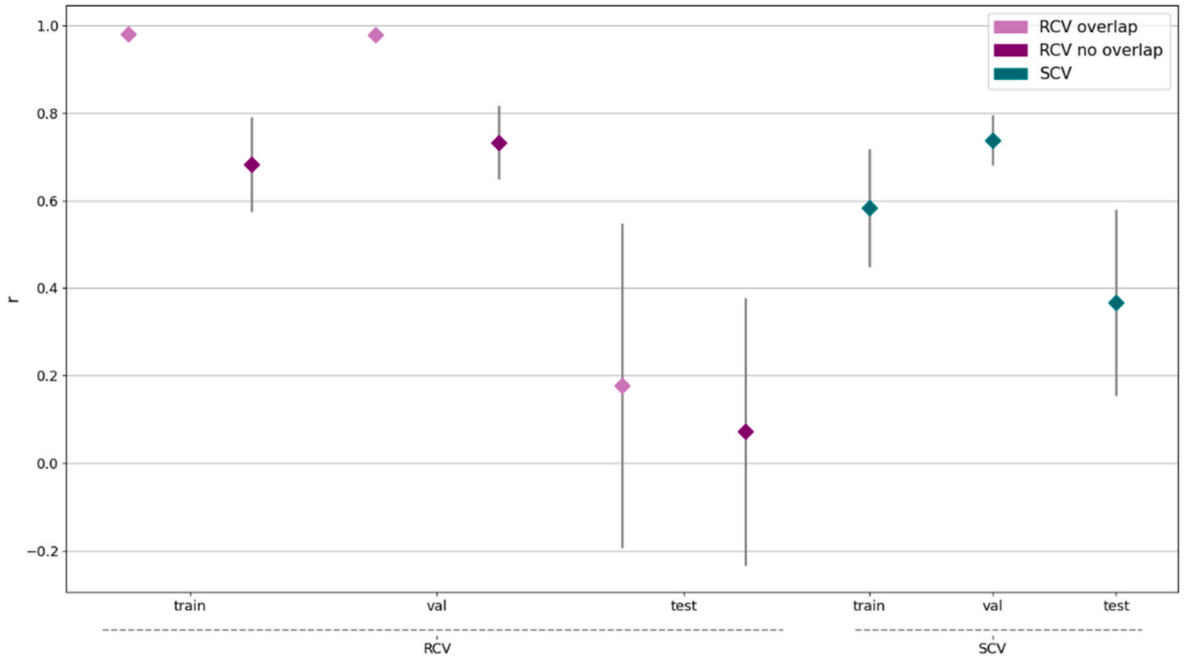


Fig. 6. Aggregated model prediction performance over two deep learning architectures and three crop types (n = 6 for each) for random (RCV) with and without sample overlap and spatial (SCV) cross validation approaches. The rhombus shape represents an aggregated performance score, and the associated bar represents the standard deviation. Aggregated performance metric on training/validation is the average global Pearson's correlation coefficient (r) of predicted vs. observed yield, on test the average r across folds. We can see a significant performance drop from train to test in both RCV cases, while a performance decline from train to test is much smaller for SCV.

et al., 2014; Roberts et al., 2017; Ploton et al., 2020; Kattenborn et al., 2022). Using RCV with remote sensing imagery leads to inflated model performance. The cause of inflated performance can be accounted to: i) overfitting to the spatial structure in the training field, ii) overfitting to training data by storing data points rather than learning generalizable features. Our results suggest that RCV with overlap had inflated performance due overfitting and data points storing due to data leakage, whereas RCV without overlapping samples was inflated solely by overfitting to the spatial structure.

When structural dependencies are present in the data, and the process of splitting training and validation samples does not ensure their independence (cf. spatial autocorrelation; Moran, 1950; Hubert et al., 1981), prediction performance has been shown to be inflated (Le Rest et al., 2014; Roberts et al., 2017; Ploton et al., 2020; Kattenborn et al., 2022). For example, Kattenborn et al. (2022) showed how CNN based segmentation using RCV inflated the prediction performance for tree species classification. They argued that CNNs are especially prone to overfit spatial structures, as it is the exploitation of image patterns that makes them superior predictive tools for images. Here we confirm these findings for crop yield prediction models using high resolution RGB images to have a strong positive bias when trained and validated with RCV as compared to SCV.

Though DL models tend to learn image patterns for predicting rather than *memorizing single pixels* (Krueger et al., 2017), it has been shown that they can overfit small training data by storing data points (e.g. pixels in an image) rather than learning generalized features (Elhage et al., 2022), such as the concept and number of corncobs or ears in a field. Given the fixed size of a small data set consisting of sliding window samples in our study, we argue that the choice of model training and validation technique affects the model to overfit to a present spatial structure, or even memorize points, rather than learning more generalized features. Due to random sampling on images without overlap a model exploits the spatial structure for predicting. With overlap we introduce dependencies between training and validation set (cf. data leakage), such that models trained with RCV tend to memorize and achieve less generalization (i.e., low performance on the test set). With this we can explain the increased prediction performance of models using RCV in the same field as training accompanied by a huge performance drop when tested externally as a combined effect of overfitting and memorization. The model simply remembers what it had previously seen in the training field. On the contrary, sampling by spatial blocking in SCV can encourage the model to learn more generalized features.

Further investigation is needed for understanding what the models learned. For instance, we could imagine that the models might have learned agronomically meaningful features including fruit bodies, plant height, the presence of weeds, and field management paths. Yet, DL models are inherently not interpretable and need to be probed with certain methods and concepts that add explanations for predicting (cf. explainable artificial intelligence (Ryo, 2022). To explain reasons behind predictions, visualization of learned features (Hohman et al., 2020) or adding post hoc explanations (Ribeiro et al., 2016) are relevant steps in order to confirm that learned features follow generalized concepts. Understanding the learning outcome can help validate model generalizability and transferability.

Beyond SCV, other methods and concepts can be applied to further facilitate model transferability. For instance, we used data augmentations to help the model avoid overfitting by learning more robust and generalizable features (Shorten and Khoshgoftaar, 2019). Moreover, the model's performance on the validation set during training can be monitored and the training process can be stopped when the performance starts to degrade in order to tackle overfitting the training data (cf. early stopping; Yao et al., 2007). Further, to facilitate model transferability, models can be pre-trained with self-supervised or transfer-learning strategies, such that the prediction model leverages the knowledge learned from the pre-trained model (Yang et al., 2020). Self-supervised learning efficiently utilizes unlabeled remote sensing imagery, avoiding the uncertainties associated with the interpolated data sets required by traditional supervised learning methods. Additionally U-Net (Ronneberger et al., 2015), with its strength in pixel-to-pixel prediction, presents an interesting approach for leveraging detailed spatial information and circumventing label aggregation. However, label aggregation helps manage uncertainties from harvester-recorded yield points and map interpolation, such that its circumvention introduces a specific challenge for U-Net's detailed spatial analysis. Without claim of completeness, additional methods to facilitate model transferability can be regularization by perturbing the input (dropout) or the distance between ground truth and prediction (such as L1 or L2 Loss) (Kukačka et al., 2017) as well as the use of ensemble methods (Ganaie et al., 2022). It's worth noting, however, that methods like dropout and batch normalization, while beneficial individually, may not harmonize when applied together (Li et al., 2019).

Using CNNs in combination with remote sensing data for prediction often requires a trade-off between spectral, spatial and temporal resolution. The scope of our study is limited to RGB reflectance values with very high spatial resolution (2.22 cm). Yet multispectral data, foremost near-infrared and red edge or the vegetation indices such as normalized difference vegetation index (NDVI) or Normalized difference red edge index (NDRE) have demonstrated effectiveness in closely correlating with plant chlorophyll content, health and phenology, particularly in crops like maize (Herrmann et al., 2020). Hence we expect single RGB pixels to be less important for prediction than multi-spectral pixels, but the image pattern to be spatially more explicit (due to higher resolution) and relevant for prediction. This may affect models to achieve lower prediction performance as for instance to be compared to (Nevavuori et al., 2019). Moreover, the majority of studies use multitemporal data as input (van Klompenburg et al., 2020). Here we showcased an approach based on one image, which is quite unique and also can affect the accuracy. Nevertheless, combinedly they do not affect the main finding of this study but rather highlights the current opportunities and the path to improvement in further studies.

In conclusion, this paper has effectively highlighted the significant role that different training and validation methodologies play in the spatial transferability of crop yield prediction models in a smallholder setting. SCV for model training and validation is a powerful technique for crop yield mapping models in smallholder farming to learn more meaningful features, while alleviating overfitting more effectively than RCV. Testing on an external field unveils a more honest approach to model assessment.

CRediT authorship contribution statement

Stefan Stiller: Data curation, Formal analysis, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. **Kathrin Grahmann:** Data curation, Resources, Writing – review & editing. **Gohar Ghazaryan:** Methodology, Supervision, Validation, Writing – review & editing. **Masahiro Ryo:** Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Writing – original draft, Writing – review & editing.

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.

Acknowledgement

We thank Lars Richter and Mohammad Amin Mohammadibanadaki for data collection and processing. This study was supported by the BMBF funded project 'Multi-modale Datenintegration, domänenspezifische Methoden und KI zur Stärkung der Datenkompetenz in der Agrarforschung (KIKompAG)' (16DKWN089). The

maintenance of the patchCROP experimental infrastructure is supported by the Leibniz Centre for Agricultural Landscape Research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ophoto.2024.100064.

References

Arslan, S., Colvin, T.S., 2002. Grain yield mapping: yield sensing, yield reconstruction, and errors. Precis. Agric. 3, 135–154. https://doi.org/10.1023/A:1013819502827.

Blackmore, S., 1999. Remedial correction of yield map data. Precis. Agric. 1, 53–66. https://doi.org/10.1023/A:1009969601387.

Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, A., Druzhinin, M., Kalinin, A.A., 2020. Albumentations: fast and flexible image augmentations. Information 11, 125. https://doi.org/10.3390/info11020125.

Chen, H., Li, W., Shi, Z., 2022. Adversarial instance augmentation for building change detection in remote sensing images. IEEE Trans. Geosci. Remote Sens. 60, 1–16. https://doi.org/10.1109/TGRS.2021.3066802.

Cho, J., Lee, K., Shin, E., Choy, G., Do, S., 2016. How Much Data Is Needed to Train a Medical Image Deep Learning System to Achieve Necessary High Accuracy? https://doi.org/10.48550/arXiv.1511.06348.

Cressie, N., 1988. Spatial prediction and ordinary kriging. Math. Geol. 20, 405–421. https://doi.org/10.1007/BF00892986.

Diker, K., Heermann, D.F., Brodahl, M.K., 2004. Frequency analysis of yield for delineating yield response zones. Precis. Agric. 5, 435–444. https://doi.org/10.1007/s11119-004-5318-9.

Divyanth, L.G., Guru, D.S., Soni, P., Machavaram, R., Nadimi, M., Paliwal, J., 2022. Image-to-Image translation-based data augmentation for improving crop/weed classification models for precision agriculture applications. Algorithms 15, 401. https://doi.org/10.3390/a15110401.

Djolonga, J., Yung, J., Tschannen, M., Romijnders, R., Beyer, L., Kolesnikov, A., et al., 2021. On robustness and transferability of convolutional neural networks. Available at: http://arxiv.org/abs/2007.08558. (Accessed 10 June 2023).

Donat, M., Geistert, J., Grahmann, K., Bloch, R., Bellingrath-Kimura, S.D., 2022. Patch cropping- a new methodological approach to determine new field arrangements that increase the multifunctionality of agricultural landscapes. Comput. Electron. Agric. 197, 106894 https://doi.org/10.1016/j.compag.2022.106894.

Elhage, N., Hume, T., Olsson, C., Schiefer, N., Henighan, T., Kravec, S., et al., 2022. Toy Models of Superposition. https://doi.org/10.48550/arXiv.2209.10652.

Falkner, S., Klein, A., Hutter, F., 2018. BOHB: Robust and Efficient Hyperparameter Optimization at Scale. https://doi.org/10.48550/arXiv.1807.01774.

Florin, M.J., McBratney, A.B., Whelan, B.M., 2009. Quantification and comparison of wheat yield variation across space and time. Eur. J. Agron. 30, 212–219. https://doi.org/10.1016/j.eja.2008.10.003.

Fukushima, K., 1975. Cognitron: a self-organizing multilayered neural network. Biol. Cybern. 20, 121–136. https://doi.org/10.1007/BF00342633.

Ganaie, M.A., Hu, M., Malik, A.K., Tanveer, M., Suganthan, P.N., 2022. Ensemble deep learning: a review. Eng. Appl. Artif. Intell. 115, 105151 https://doi.org/10.1016/j.engappai.2022.105151.

Grahmann, K., Reckling, M., Hernández-Ochoa, I., Donat, M., Bellingrath-Kimura, S., Ewert, F., 2024. Co-designing a landscape experiment to investigate diversified cropping systems. Agric. Syst. 217, 103950 https://doi.org/10.1016/j.agsy.2024.103950.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.

Henighan, T., Carter, S., Hume, T., Elhage, N., Lasenby, R., Fort, S., et al., 2023. Superposition, memorization, and double descent. Transform. Circuits. Available at: https://www.lesswrong.com/posts/6Ks6p33LQyfFkNtYE/paper-superposition-memorization-and-double-descent. (Accessed 8 June 2023).

Herrmann, I., Bdolach, E., Montekyo, Y., Rachmilevitch, S., Townsend, P.A., Karnieli, A., 2020. Assessment of maize yield and phenology by drone-mounted superspectral camera. Precis. Agric. 21, 51–76. https://doi.org/10.1007/s11119-019-09659-5.

Heydari, L., Bayat, H., Castrignanò, A., 2023. Scale-dependent geostatistical modelling of crop-soil relationships in view of Precision Agriculture. Precis. Agric. 24, 1261–1287. https://doi.org/10.1007/s11119-023-09989-5.

Hohman, F., Park, H., Robinson, C., Polo Chau, D.H., 2020. Summit: scaling deep learning interpretability by visualizing activation and attribution summarizations. IEEE Trans. Vis. Comput. Graph. 26, 1096–1106. https://doi.org/10.1109/TVCG.2019.2934659.

Hubert, L.J., Golledge, R.G., Costanzo, C.M., 1981. Generalized procedures for evaluating spatial autocorrelation. Geogr. Anal. 13, 224–233. https://doi.org/10.1111/j.1538-4632.1981.tb00731.x.

Kamilaris, A., Prenafeta-Boldú, F.X., 2018. Deep learning in agriculture: A survey. Comput. Electron. Agric. 147, 70–90. https://doi.org/10.1016/j.compag.2018.02.016.

Kattenborn, T., Schiefer, F., Frey, J., Feilhauer, H., Mahecha, M.D., Dormann, C.F., 2022. Spatially autocorrelated training and validation samples inflate performance assessment of convolutional neural networks. ISPRS Open J. Photogramm. Remote Sens. 5, 100018 https://doi.org/10.1016/j.ophoto.2022.100018.

Krueger, D., Ballas, N., Jastrzebski, S., Arpit, D., Kanwal, M.S., Maharaj, T., et al., 2017. DEEP NETS DON'T LEARN VIA MEMORIZATION.

Kukačka, J., Golkov, V., Cremers, D., 2017. Regularization for Deep Learning: A Taxonomy. https://doi.org/10.48550/arXiv.1710.10686.

Kussul, N., Lavreniuk, M., Skakun, S., Shelestov, A., 2017. Deep learning classification of land cover and crop types using remote sensing data. IEEE Geosci. REMOTE Sens. Lett. 14, 778–782. https://doi.org/10.1109/LGRS.2017.2681128.

Kuwata, K., Shibasaki, R., 2015. Estimating crop yields with deep learning and remotely sensed data. In: 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 858–861. https://doi.org/10.1109/IGARSS.2015.7325900.

Lancashire, P.D., Bleiholder, H., Boom, T.V.D., Langelüddeke, P., Stauss, R., Weber, E., et al., 1991. A uniform decimal code for growth stages of crops and weeds. Ann. Appl. Biol. 119, 561–601. https://doi.org/10.1111/j.1744-7348.1991.tb04895.x.

Lange, M., Feilhauer, H., Kühn, I., Doktor, D., 2022. Mapping land-use intensity of grasslands in Germany with machine learning and Sentinel-2 time series. Remote Sens. Environ. 277, 112888 https://doi.org/10.1016/j.rse.2022.112888.

Le Rest, K., Pinaud, D., Monestiez, P., Chadoeuf, J., Bretagnolle, V., 2014. Spatial leave-one-out cross-validation for variable selection in the presence of spatial autocorrelation. Glob. Ecol. Biogeogr. 23, 811–820. https://doi.org/10.1111/geb.12161.

Li, W., Fu, H., Yu, L., Cracknell, A., 2017. Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. REMOTE Sens 9. https://doi.org/10.3390/rs9010022.

Li, X., Chen, S., Hu, X., Yang, J., 2019. Understanding the disharmony between dropout and batch normalization by variance shift. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2677–2685. https://doi.org/10.1109/CVPR.2019.00279.

Liaw, R., Liang, E., Nishihara, R., Moritz, P., Gonzalez, J.E., Stoica, I., 2018. Tune: A Research Platform for Distributed Model Selection and Training. https://doi.org/10.48550/arXiv.1807.05118.

Loshchilov, I., Hutter, F., 2017. SGDR: Stochastic Gradient Descent with Warm Restarts. https://doi.org/10.48550/arXiv.1608.03983.

Lowder, S.K., Skoet, J., Raney, T., 2016. The number, size, and distribution of farms, smallholder farms, and family farms worldwide. World Dev. 87, 16–29. https://doi.org/10.1016/j.worlddev.2015.10.041.

Lyle, G., Bryan, B.A., Ostendorf, B., 2014. Post-processing methods to eliminate erroneous grain yield measurements: review and directions for future development. Precis. Agric. 15, 377–402. https://doi.org/10.1007/s11119-013-9336-3.

Maimaitijiang, M., Sagan, V., Sidike, P., Hartling, S., Esposito, F., Fritschi, F.B., 2020. Soybean yield prediction from UAV using multimodal data fusion and deep learning. Remote Sens. Environ. 237 https://doi.org/10.1016/j.rse.2019.111599.

Meyer, H., Reudenbach, C., Wöllauer, S., Nauss, T., 2019. Importance of spatial predictor variable selection in machine learning applications – moving from data reproduction to spatial prediction. Ecol. Model. 411, 108815 https://doi.org/10.1016/j.ecolmodel.2019.108815.

Micikevicius, P., Narang, S., Alben, J., Diamos, G., Elsen, E., Garcia, D., et al., 2018. Mixed Precision Training. https://doi.org/10.48550/arXiv.1710.03740.

Moran, P.A.P., 1950. Notes on continuous stochastic phenomena. Biometrika 37, 17–23. https://doi.org/10.2307/2332142.

Mulvaney, M., Devkota, P., 2020. Adjusting crop yield to a standard moisture content. Environ. Data Inf. Serv. 2020 https://doi.org/10.32473/edis-ag442-2020.

Nevavuori, P., Narra, N., Lipping, T., 2019. Crop yield prediction with deep convolutional neural networks. Comput. Electron. Agric. 163 https://doi.org/10.1016/j.compag.2019.104859.

Patzold, S., Mertens, F.M., Bornemann, L., Koleczek, B., Franke, J., Feilhauer, H., et al., 2008. Soil heterogeneity at the field scale: a challenge for precision crop protection. Precis. Agric. 9, 367–390. https://doi.org/10.1007/s11119-008-9077-x.

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., et al., 2020. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. Nat. Commun. 11, 4540 https://doi.org/10.1038/s41467-020-18321-y.

Ribeiro, M.T., Singh, S., Guestrin, C., 2016. "Why should I trust you?": explaining the predictions of any classifier. ArXiv160204938 Cs Stat. Available at: http://arxiv.org/abs/1602.04938. (Accessed 10 June 2021).

Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., et al., 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40, 913–929. https://doi.org/10.1111/ecog.02881.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. Available at: ArXiv150504597 Cs http://arxiv.org/abs/1505.04597. (Accessed 29 September 2021).

Ruder, S., 2017. An overview of gradient descent optimization algorithms. Available at: http://arxiv.org/abs/1609.04747. (Accessed 13 June 2023).

Ryo, M., 2022. Explainable artificial intelligence and interpretable machine learning for agricultural data analysis. Artif. Intell. Agric 6, 257–265. https://doi.org/10.1016/j.aiia.2022.11.003.

Ryo, M., Schiller, J., Stiller, S., Rivera Palacio, J.C., Mengsuwan, K., Safonova, A., et al., 2022. Deep learning for sustainable agriculture needs ecology and human involvement. J. Sustain. Agric. Environ. 2, 40–44. https://doi.org/10.1002/sae2.12036.

Safonova, A., Ghazaryan, G., Stiller, S., Main-Knorn, M., Nendel, C., Ryo, M., 2023. Ten deep learning techniques to address small data problems with remote sensing. Int. J. Appl. Earth Obs. Geoinformation 125, 103569. https://doi.org/10.1016/j.jag.2023.103569.

Scott, G.J., England, M.R., Starms, W.A., Marcum, R.A., Davis, C.H., 2017. Training deep convolutional neural networks for land–cover classification of high-resolution

imagery. IEEE Geosci. Remote Sens. Lett. 14, 549–553. https://doi.org/10.1109/LGRS.2017.2657778.
Sefrin, O., Riese, F.M., Keller, S., 2021. Deep learning for land cover change detection. Remote Sens 13, 78. https://doi.org/10.3390/rs13010078senseFlyeBeeX n.d.). AgEagle Aer. Syst. Inc. Available at: https://ageagle.com/solutions/drones. (Accessed 6 July 2023) (Accessed July 6, 2023). senseFly-S.O.D.A. (n.d.). AgEagle Aer. Syst. Inc. Available at:
Shah, F., Wu, W., 2019. Soil and crop management strategies to ensure higher crop productivity within sustainable environments. Sustainability 11, 1485. https://doi.org/10.3390/su11051485.
Shen, R., Huang, A., Li, B., Guo, J., 2019. Construction of a drought monitoring model using deep learning based on multi-source remote sensing data. Int. J. Appl. EARTH Obs. GEOINFORMATION 79, 48–57. https://doi.org/10.1016/j.jag.2019.03.006.
Shorten, C., Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. J. Big Data 6, 60. https://doi.org/10.1186/s40537-019-0197-0.
Stone, M., 1974. Cross-validatory choice and assessment of statistical predictions. J. R. Stat. Soc. Ser. B Methodol. 36, 111–133. https://doi.org/10.1111/j.2517-6161.1974.tb00994.x.
Su, D., Kong, H., Qiao, Y., Sukkarieh, S., 2021. Data augmentation for deep learning based semantic segmentation and crop-weed classification in agricultural robotics. Comput. Electron. Agric. 190, 106418 https://doi.org/10.1016/j.compag.2021.106418.
Tetila, E.C., Machado, B.B., Astolfi, G., de Souza Belete, N.A., Amorim, W.P., Roel, A.R., et al., 2020. Detection and classification of soybean pests using deep learning with UAV images. Comput. Electron. Agric. 179 https://doi.org/10.1016/j.compeg.2020.105836.
Tittonell, P., 2023. Spatial heterogeneity in agroecosystems. In: Tittonell, P. (Ed.), A Systems Approach to Agroecology. Springer Nature Switzerland, Cham, pp. 241–280. https://doi.org/10.1007/978-3-031-42939-2_7.
Tobler, W.R., 1970. A computer movie simulating urban growth in the Detroit region. Econ. Geogr. 46, 234–240. https://doi.org/10.2307/143141.
Valente, J., Hiremath, S., Ariza-Sentís, M., Doldersum, M., Kooistra, L., 2022. Mapping of Rumex obtusifolius in nature conservation areas using very high resolution UAV imagery and deep learning. Int. J. Appl. Earth Obs. Geoinformation 112, 102864. https://doi.org/10.1016/j.jag.2022.102864.
van Klompenburg, T., Kassahun, A., Catal, C., 2020. Crop yield prediction using machine learning: a systematic literature review. Comput. Electron. Agric. 177, 105709 https://doi.org/10.1016/j.compag.2020.105709.
van Wijk, W.R., 1965. Soil microclimate, its creation, observation and modification. In: Waggoner, P.E., Gates, D.M., Webb, E.K., van Wijk, W.R., Businger, J.A., Crawford, T.V., et al. (Eds.), Agricultural Meteorology. American Meteorological Society, Boston, MA, pp. 59–73. https://doi.org/10.1007/978-1-940033-58-7_3.
Wainer, J., Cawley, G., 2021. Nested cross-validation when selecting classifiers is overzealous for most practical applications. Expert Syst. Appl. 182, 115222 https://doi.org/10.1016/j.eswa.2021.115222.
Xu, J., Zhu, Y., Zhong, R., Lin, Z., Xu, J., Jiang, H., et al., 2020. DeepCropMapping: a multi-temporal deep learning approach with improved spatial generalizability for dynamic corn and soybean mapping. Remote Sens. Environ. 247 https://doi.org/10.1016/j.rse.2020.111946.
Yang, X., He, X., Liang, Y., Yang, Y., Zhang, S., Xie, P., 2020. Transfer learning or self-supervised learning? A Tale of Two Pretraining Paradigms. https://doi.org/10.48550/arXiv.2007.04234.
Yao, Y., Rosasco, L., Caponnetto, A., 2007. On early stopping in gradient descent learning. Constr. Approx. 26, 289–315. https://doi.org/10.1007/s00365-006-0663-2.
Yu, X., Wu, X., Luo, C., Ren, P., 2017. Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework. GIScience Remote Sens. 54, 741–758. https://doi.org/10.1080/15481603.2017.1323377.
Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., et al., 2020. Deep learning in environmental remote sensing: achievements and challenges. Remote Sens. Environ. 241 https://doi.org/10.1016/j.rse.2020.111716.
Zhang, D., Pan, Y., Zhang, J., Hu, T., Zhao, J., Li, N., et al., 2020a. A generalized approach based on convolutional neural networks for large area cropland mapping at very high resolution. Remote Sens. Environ. 247 https://doi.org/10.1016/j.rse.2020.111912.
Zhang, W., Liljedahl, A.K., Kanevskiy, M., Epstein, H.E., Jones, B.M., Jorgenson, M.T., et al., 2020b. Transferability of the deep learning mask R-CNN model for automated mapping of ice-wedge polygons in high-resolution satellite and UAV images. Remote Sens 12, 1085. https://doi.org/10.3390/rs12071085.