Multi-source Unsupervised Domain Adaptation on Corn Yield Prediction

Yuchi Ma[1], Zhou Zhang[1*]

Department of Biological Systems Engineering, University of Wisconsin-Madison[1]

* Author to whom correspondence should be addressed.

Email: ma286@wisc.edu; zzhang347@wisc.edu

Abstract

Recently, supervised machine learning methods based on remote sensing observations have achieved satisfactory results in crop yield prediction. However, supervised learning models tend to have poor transferability. Due to domain shifts between observations in different regions, models trained with data from one spatial region (i.e., source domain) often lose their validity when directly applied to another region (i.e., target domain). To address this issue, we proposed a Bayesian Multi-source Maximum Predictor Discrepancy (BMMPD) neural network which is an unsupervised domain adaptation (UDA) approach to improve the model's transferability for corn yield prediction at the county level. We proposed to maximize the discrepancy between two yield predictors' outputs to detect unlabeled target samples that are far from the support of the source domain. A feature extractor then learned to align source and target domains by minimizing the predictor discrepancy. Moreover, we applied Bayesian learning to prevent overfitting. A case study was conducted in the U.S. corn belt to evaluate the proposed BMMPD model. Time-series vegetation indices and weather observations were collected and aggregated to the county level and used as the input predictors. Experiment results demonstrated that the proposed BMMPD has effectively reduced domain shifts and outperformed several state-of-art domain adaptation methods.

Introduction

Accurate crop yield prediction is central to the prevention of famine, stability of commodity markets, and sustainable development of agriculture (Feng et al., 2020; Johnson 2014; Sun et al., 2019). Leading crop yield prediction techniques, such as survey-based methods and biophysical simulation models, rely heavily on locally sensed data, which are expensive to collect for economic and manpower reasons. Recently, with the advent of satellite missions and artificial intelligence techniques, supervised machine learning models have been developed to correlate crop yield with remote sensing observations. For example, Johnson (2014) estimated the U.S. county-level yield for corn and soybean using tree-based models with sequential satellite vegetation indices (VIs) and weather observations as input variables. Khaki and Wang (2019) incorporated weather and soil observations with a fully connected deep neural network (DNN) to forecast U.S. corn yield. Ma et al. (2021a) developed a Bayesian neural network (BNN) based on multi-source remote sensing observations, which can predict not only corn yield but also the uncertainty associated with the prediction.

Due to the phenomenon known as "domain shift" (Kouw and Loog, 2019) caused by spatial heterogeneity of meteorological conditions, soil properties, and farming practice, machine learning models established between reference (reported) yields and remote sensing measurements within a specific region often lose their validity when directly applied to new regions. To address this challenge, domain adaptation, a new machine learning technique that transfers knowledge learned from a local region with rich ground reference data to the target region with limited or no ground truth data, has become a viable solution. To perform domain adaptation for DNNs, a widely used supervised strategy is to first pre-train a model on a source domain with abundant ground reference data, and then apply it in a target domain (e.g., a new region), with the parameters of the pre-trained model are fine-tuned using labeled samples from the target domain. Despite several successful cases (Russello, 2018; Wang et al., 2018), a certain number of labeled data (e.g., yield records) from the target domain is still needed to perform the fine-tuning. Because collecting yield data can be financially expensive, labor-intensive, and time-consuming, many agricultural production areas may lack reliable ground reference yield data for either directly training or fine-tuning supervised transfer learning models.

To address this issue, unsupervised domain adaptation (UDA) has been proposed to improve model transferability without requiring any labeled data in the target domain. Widely used existing UDA methods include discrepancy-

based methods and adversarial-based methods. Discrepancy-based methods try to align features from source and target domains by minimizing the distance between feature distributions (Long et al., 2015; Luo et al., 2017). Adversarial-based methods address the domain shift by learning good representations that are discriminative for the main learning task and indiscriminative between source and target domains (Ganin et al., 2016). While significant progress has been made on domain adaptation for image classification and segmentation, UDA studies on remote sensing are rare, especially for regression tasks such as yield prediction. There are three major bottlenecks in applying UDA methods to yield prediction with remote sensing observations. First, most current UDA methods are used for single-source domain adaptation which assumes labeled data are acquired from a single region (domain). In practice, labeled training samples may be collected from multiple source domains with different feature distributions (Zhao et al., 2020). Since domain shifts exist not only between source and target but also among different source domains, single-source domain adaptation methods could have a poor performance when samples from different sources interfere with each other (Riemer et al., 2019). Second, current UDA methods mostly align distributions of source and target without considering specific tasks. For remote sensing measurements with significant domain shifts, input data are likely to be projected to ambiguous feature spaces that lack meaningful information. Moreover, since the training set for county-level yield prediction is relatively small, overfitting could happen during model training.

To address those issues, inspired by maximum classifier discrepancy (Saito et al., 2018) and Bayesian learning (Ma et al., 2021a), we proposed a Bayesian Multi-source Maximum Predictor Discrepancy (BMMPD) neural network UDA approach. Specifically, labeled samples collected from different regions were grouped into multiple sources for multi-source domain adaptation. Then, by using maximum predictor discrepancy, the BMMPD model aimed to align the distributions of source and target domains by considering task-specific regression models. In addition, Bayesian learning was applied to prevent overfitting. Experiments in the U.S. corn belt showed that the proposed method outperformed state-of-art deep learning models with improved model transferability.

Related Work

Existing UDA methods mainly focus on the single-source scenario, i.e., labeled data samples are assumed to be from one source domain. Single-source UDA algorithms commonly employ a conjugated architecture with two objectives (Zhao et al., 2020). One objective is to learn a task model based on the labeled source samples by corresponding task losses, such as mean square error loss (MSE) for regression (Feng et al., 2021) and cross-entropy loss for classification (Wang et al., 2021). The other objective is to reduce the domain shift and align the source and target domains. One of the most representative single UDA methods is domain adversarial neural networks (DANN) (Ganin et al., 2017), which employs an adversarial objective with a domain discriminator to extract domain-invariant features from source and target domains. Ma et al. (2021b), for the first time in the precision agriculture community, employed the DANN model for county-level yield prediction and demonstrated its effectiveness.

Recently, there has been growing interest in multi-source UDA. For example, Peng et al. (2019) proposed a multi-source UDA model named Moment Matching for Multi-Source Domain Adaptation (M^3SDA) for image classification. M^3SDA reduces source-target divergence and inter-source divergence by minimizing the moment-related distance between each domain. Xu et al. (2018) proposed a deep cocktail network that uses multi-way adversarial learning to minimize the discrepancy between the target and source domains. Tasar et al. (2020) proposed a StandardGAN which standardizes multiple source domains and target domains for satellite image segmentation. However, to our best knowledge, no multi-source UDA studies have been conducted for yield prediction which is a regression task that differs from classification applications.

Methodology

In the scenario of yield prediction, predictors $\mathbf{x}$ are remote sensing observations, and response variable y is crop yield. Domain shifts exist between the source domain $(\mathbf{x}_s, y_s) \sim \mathcal{D}_s$ and the target domain $(\mathbf{x}_t, y_t) \sim \mathcal{D}_t$. When there are limited or even no yield records in the target domain, supervised learning models can only be trained with labeled source samples and would have poor performance in the target domain when they are directly applied without domain adaptation. Representative UDA methods, such as DANN and M^3SDA, try to align source and target domains by adapting the marginal distributions $p(\mathbf{x}_s)$ and $p(\mathbf{x}_t)$ under the assumption that $p(y_s|\mathbf{x_s})$ and $p(y_t|\mathbf{x_t})$ are the same in both domains. This assumption could be invalid in the scenario of yield prediction since the crop yield response to a given remote sensing observations $\mathbf{x}$ which may be different from region to region due to environmental variations (Deines et al., 2021).

To better align $p(y_s|\mathbf{x_s})$ and $p(y_t|\mathbf{x_t})$, we proposed a new deep multi-source UDA approach for yield prediction, named BMMPD. The innovation of the model is threefold. First, the setting of multiple source domains was adopted to prevent negative interference between samples from different sources. Second, Bayesian learning was incorporated to prevent overfitting. Third, a pair of predictors was trained for each source and used to measure discrepancy on unlabeled target samples.

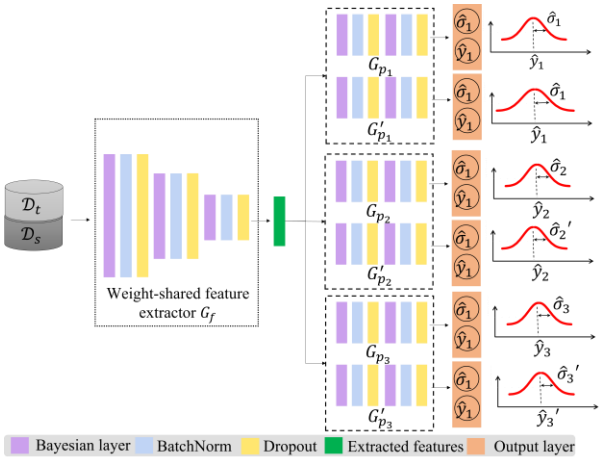

Figure 1. The architecture of BMMPD given three sources.

Specifically, given unlabeled target domain $\mathcal{D}_t$ and N labeled source domains $\mathcal{D}_s = \{\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_N\}$, the BMMPD model has a weight-shared feature extractor G_f and N pairs of source-specific yield predictors $\{G_{p_i}, G'_{p_i}\}_{i=1}^{N}$ (Figure 1). The feature extractor and yield predictors were designed as Bayesian Neural Networks and each predictor has two endpoints to estimate the yield distribution in the form of a normal distribution (Ma et al., 2021a).

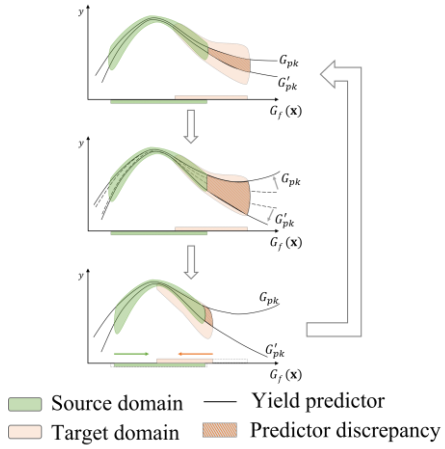

Figure 2. Example of aligning target samples to the kth source domain by the BMMPD model.

BMMPD is trained recursively by three steps (Figure 2). In step 1, labeled data from each source is forwarded through the feature extractor and then to the domain-specific predictors for yield distribution estimation. The yield prediction loss in each source is calculated using Eq. (2)-(3) and the model is updated to minimize the total yield prediction loss in all sources (Eq. (1)):

$$\min_{G_f, \{G_{p_i}, G'_{p_i}\}_{i=1}^{N}} L_y(\mathcal{D}_s) = \sum_{i=1}^{N} L_y(\mathcal{D}_i) + L'_y(\mathcal{D}_i) \quad (1)$$

$$L_y(\mathcal{D}_i) = \mathbb{E}_{(\mathbf{x_s}, y_s)\sim\mathcal{D}_i}[-\log p(y_s|G_{p_i}(G_f(\mathbf{x_s})))] \quad (2)$$

$$L'_y(\mathcal{D}_i) = \mathbb{E}_{(\mathbf{x_s}, y_s)\sim\mathcal{D}_i}[-\log p(y_s|G'_{p_i}(G_f(\mathbf{x_s})))] \quad (3)$$

In step 2, the feature extractor is fixed while yield predictors are kept trainable. Both labeled source data $\mathcal{D}_s$ and unlabeled target data $\mathcal{D}_t$ are used for model training. Like step 1, labeled source data from each source domain are used to calculate the prediction loss $L_y(\mathcal{D}_s)$. Meanwhile, the unlabeled target data are fed into the feature extractor and then forwarded to all predictors. The predictor discrepancy $L_d(\mathcal{D}_t)$ is calculated as the *KL*-divergence between the estimated target yield distributions by each pair of predictors (Eq. (5)-(6)). Since each pair of predictors $\{G_{p_i}, G'_{p_i}\}$ is initialized differently and trained with identical source samples, they would agree well on their predictions for source samples but have random agreement on samples outside the support of the source. Therefore, given unlabeled target data $\mathbf{x_t}\sim\mathcal{D}_t$, the prediction discrepancy between G_{p_i} and G'_{p_i} indicates the similarity between $p(y_s|\mathbf{x_s})$ and $p(y_t|\mathbf{x_t})$. During backpropagation, G_{p_i} and G'_{p_i} are updated to minimize the prediction loss while maximizing the discrepancy loss (Eq. (4)). By doing this, G_{p_i} and G'_{p_i} are trained to better discriminate target samples with large discrepancy loss while keeping low prediction loss on the source domain:

$$\min_{\{G_{p_i}, G'_{p_i}\}_{i=1}^{N}} L(\mathcal{D}_s, \mathcal{D}_t) = L_y(\mathcal{D}_s) - L_d(\mathcal{D}_t) \quad (4)$$

$$L_d(\mathcal{D}_t) = \sum_{i=1}^{N} L_{d_i}(\mathcal{D}_t) \quad (5)$$

$$L_{d_i}(\mathcal{D}_t) = \mathbb{E}_{\mathbf{x_t}\sim\mathcal{D}_t}[KL(G_f(G_{p_i}(\mathbf{x_t})), G_f(G'_{p_i}(\mathbf{x_t})))] \quad (6)$$

where $KL(p, q)$ denotes the *KL*-divergence between two distributions p and q.

In step 3, the feature extractor is kept trainable while the N pairs of yield predictors are fixed. Only unlabeled target data $\mathbf{x_t}$ is input into the network and used to calculate the predictor discrepancy $L_d(\mathcal{D}_t)$. During backpropagation, the feature extractor is updated towards minimizing the predictor discrepancy $L_d(\mathcal{D}_t)$. By doing this, the feature extractor is trained to extract domain-invariant features and align source and target distributions in a task-specific way. Finally, in the testing phase, the final prediction is an ensemble result of outputs from each predictor.

Experiments and Results

Experimental Setup

Counties in the study area were grouped into four domains according to eco-regions partitioned by National Ecological Observatory Network (NEON), a continental-scale research platform for understanding ecosystems (Kampe, 2010). NEON partitions the U.S. into eco-regions with different ecosystem performances. Counties in the U.S. corn belt are in seven NEON eco-regions. Since some eco-regions consist of very few counties from the study area, we merged small

eco-regions and finally resulted in four eco-domains (Figure 3). In experiments, each eco-domain was alternatively treated as the target domain and the other three were treated as sources. Reported yield records for each county were collected from USDA National Agricultural Statistics Service (NASS) database (USDA, 2020).

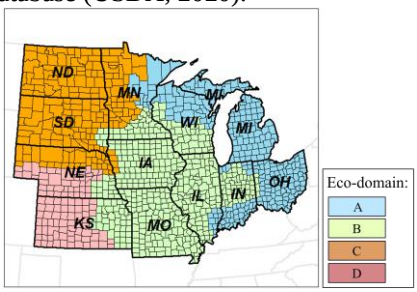

Figure 3. Four eco-domains in the study area.

Multi-source remote sensing data from 2006 to 2019 were collected and used as input predictors. Specifically, to measure the biomass, canopy chlorophyll content, and vegetation moisture content, three complementary VIs have been extracted from the daily MODIS MCD43A4 product (Schaaf and Wang, 2015), including Enhanced Vegetation Index (EVI), Green Chlorophyll Index (GCI), and Normalized Difference Water Index (NDWI). In addition, daytime and nighttime land surface temperature were collected from the MODIS MOD11A2 product (Park et al., 2005). Daily total precipitation, maximum air temperature, and mean air temperature were extracted from the PRISM dataset (Daly et al., 2008). These variables were first aggregated spatially to the county level. After that, sequential predictors were aggregated to a 16-day interval from April to October to cover the growing season of corn. Finally, time-series predictors and yield records were paired for model training. BMMPD was compared to three approaches, including DNN, DANN, and M^{3}SDA. DNN was trained with all labeled source samples and applied directly to the target domain without domain adaptation. When training DANN, three source eco-domains were grouped into one source domain since DANN is a single-source UDA method. M^{3}SDA was trained following the multi-source setting. We used all preceding years since 2006 for model training and tested models in 2016–2019. Each model was evaluated based on root mean square error (RMSE) and mean absolute relative error (MARE).

Experiment Results

The evaluation results in each target domain were averaged over four testing years with the best performing one highlighted in bold (Table 1). BMMPD was observed to have a better spatial transferability than DNN, DANN, and M^{3}SDA in eco-domain A, C, and D.

Table 1. Model comparison results of RMSE (t/ha) and MARE (%) in each target eco-domains.

Target	DNN		DANN		M^{3}SDA		BMMPD	
	RMSE	MARE	RMSE	MARE	RMSE	MARE	RMSE	MARE
A	1.15	9.18	1.04	7.84	1.25	10.24	**0.96**	**7.60**
B	1.35	9.76	**1.30**	**9.11**	2.19	15.43	1.54	10.75
C	1.56	12.31	1.56	14.53	1.77	15.60	**1.19**	**9.95**
D	1.36	12.14	1.32	11.22	2.12	17.56	**1.13**	**9.87**

We further presented the average absolute error maps for each model. Although BMMPD performed well in most cases, it performed worse than DANN in ecoregion B (Figure 4), which agrees with the evaluation results in Table 1. The degraded performance of BMMPD in ecoregion B is due to the training sets from source domains are not large enough. When conducting multi-source domain adaptation, each domain-specific yield predictor of BMMPD was underfitted due to the small training set from each source domain. It demonstrated that sufficient labeled data from source domains are required to guarantee the success of the multi-source UDA.

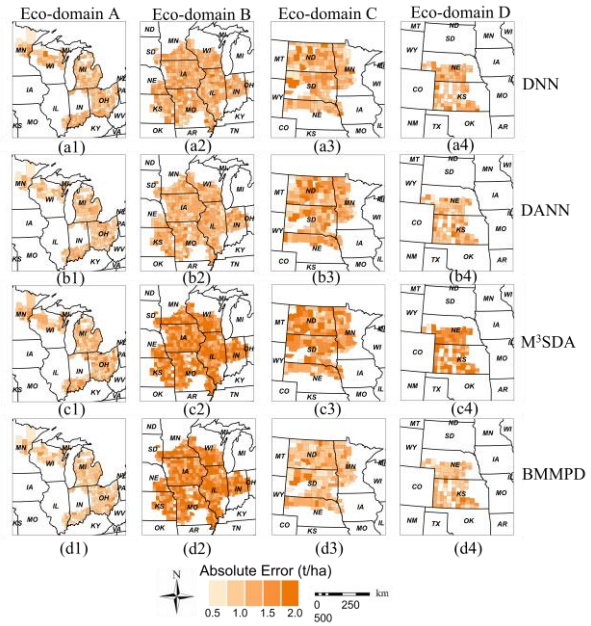

Figure 4. The average absolute error maps for model (a) DNN, (b) DANN, (c) M^{3}SDA, (d) BMMPD in eco-domain (1) A, (2) B, (3) C, (4) D.

Conclusion

In this study, we proposed a UDA method named BMMPD for corn yield prediction. It aims to accurately predict corn yield in the target domain without using any labeled samples from that region. By using maximum predictor discrepancy, BMMPD could align the distributions of source and target domains by considering task-specific regression models. The multiple domain setting was adapted to address negative interference between different sources. Bayesian learning was applied to prevent overfitting. More datasets will be used to further validate the proposed method in the future.

References

Astola, H., Seitsonen, L., Halme, E., Molinier, M., Lönnqvist, A., 2021. Deep neural networks with transfer learning for forest variable estimation using sentinel-2 imagery in boreal forest. Remote Sens. 13. https://doi.org/10.3390/rs13122392

Daly, C., Halbleib, M., Smith, J.I., Gibson, W.P., Doggett, M.K., Taylor, G.H., Curtis, J., Pasteris, P.P., 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. Int. J. Climatol. a J. R. Meteorol. Soc. 28, 2031–2064.

Deines, J.M., Patel, R., Liang, S.Z., Dado, W., Lobell, D.B., 2021. A million kernels of truth: Insights into scalable satellite maize yield mapping and yield gap analysis from an extensive ground dataset in the US Corn Belt. Remote Sens. Environ. 253, 112174. https://doi.org/10.1016/j.rse.2020.112174

Feng, L., Zhang, Z., Ma, Y., Du, Q., Williams, P., Drewry, J., 2020. Alfalfa Yield Prediction Using UAV-Based Hyperspectral Imagery and Ensemble Learning. Remote Sens. 12, 2028. https://doi.org/10.3390/rs12122028

Feng, L., Zhang, Z., Ma, Y., Sun, Y., Du, Q., Williams, P., Drewry, J., Luck, B., 2021. Multitask Learning of Alfalfa Nutritive Value From UAV-Based Hyperspectral Images. IEEE Geosci. Remote Sens. Lett.

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V., 2017. Domain-adversarial training of neural networks. Adv. Comput. Vis. Pattern Recognit. 17, 189–209. https://doi.org/10.1007/978-3-319-58347-1_10

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V., 2016. Domain-adversarial training of neural networks. J. Mach. Learn. Res. 17, 2030–2096.

Hao, P., Di, L., Zhang, C., Guo, L., 2020. Transfer Learning for Crop classification with Cropland Data Layer data (CDL) as training samples. Sci. Total Environ. 733, 138869. https://doi.org/10.1016/j.scitotenv.2020.138869

Johnson, D.M., 2014. An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. Remote Sens. Environ. 141, 116–128. https://doi.org/10.1016/j.rse.2013.10.027

Kampe, T.U., 2010. NEON: the first continental-scale ecological observatory with airborne remote sensing of vegetation canopy biochemistry and structure. J. Appl. Remote Sens. 4, 043510. https://doi.org/10.1117/1.3361375

Khaki, S., Wang, L., 2019. Crop yield prediction using deep neural networks. Front. Plant Sci. 10, 1–10. https://doi.org/10.3389/fpls.2019.00621

Kouw, W. M., & Loog, M. (2019). A review of domain adaptation without target labels. IEEE transactions on pattern analysis and machine intelligence, 43(3), 766-785.

Long, M., Cao, Y., Wang, J., Jordan, M.I., 2015. Learning transferable features with deep adaptation networks. 32nd Int. Conf. Mach. Learn. ICML 2015 1, 97–105.

Luo, Z., Zou, Y., Hoffman, J., Li, F.F., 2017. Label efficient learning of transferable representations across domains and tasks. Adv. Neural Inf. Process. Syst. 2017-Decem, 165–177.

Ma, Y., Zhang, Z., Kang, Y., Özdoğan, M., 2021a. Corn yield prediction and uncertainty analysis based on remotely sensed variables using a Bayesian neural network approach. Remote Sens. Environ. 259, 112408.

Ma, Y., Zhang, Z., Lexie, H., Yang, Z., 2021b. An adaptive adversarial domain adaptation approach for corn yield prediction. Comput. Electron. Agric. 187, 106314. https://doi.org/10.1016/j.compag.2021.106314

Park, S., Feddema, J.J., Egbert, S.L., 2005. MODIS land surface temperature composite data and their relationships with climatic water budget factors in the central Great Plains. Int. J. Remote Sens. 26, 1127–1144.

Peng, X., Bai, Q., Xia, X., Huang, Z., Saenko, K., Wang, B., 2019. Moment matching for multi-source domain adaptation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1406–1415.

Riemer, M., Cases, I., Ajemian, R., Liu, M., Rish, I., Tu, Y., Tesauro, G., 2019. Learning to learn without forgetting by maximizing transfer and minimizing interference. 7th Int. Conf. Learn. Represent. ICLR 2019 1–31.

Russello, H., 2018. Convolutional Neural Networks for Crop Yield Prediction using Satellite Images. IBM Cent. Adv. Stud.

Saito, K., Watanabe, K., Ushiku, Y., Harada, T., 2018. Maximum classifier discrepancy for unsupervised domain adaptation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3723–3732.

Schaaf, C., Wang, Z., 2015. MCD43A4 MODIS/Terra+ Aqua BRDF/Albedo Nadir BRDF Adjusted RefDaily L3 Global-500m V006. NASA EOSDIS L. Process. DAAC.

Sun, C., Feng, L., Zhang, Z., Ma, Y., Crosby, T., Naber, M., Wang, Y., 2019. Prediction of End-Of-Season Tuber Yield and Tuber Set in Potatoes Using In-Season UAV-Based Hyperspectral Imagery and Machine Learning.

Tasar, O., Tarabalka, Y., Giros, A., Alliez, P., Clerc, S.,

2020. StandardGAN: Multi-source domain adaptation for semantic segmentation of very high resolution satellite images by data standardization. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work. 2020-June, 747–756. https://doi.org/10.1109/CVPRW50498.2020.00104

USDA, 2020. United States Department of Agriculture National Agricultural Statistics Service [WWW Document]. URL https://quickstats.nass.usda.gov/ (accessed 12.27.20).

Wang, A.X., Tran, C., Desai, N., Lobell, D., Ermon, S., 2018. Deep transfer learning for crop yield prediction with remote sensing data. Proc. 1st ACM SIGCAS Conf. Comput. Sustain. Soc. COMPASS 2018. https://doi.org/10.1145/3209811.3212707

Wang, Y., Zhang, Z., Feng, L., Ma, Y., Du, Q., 2021. A new attention-based CNN approach for crop mapping using time series Sentinel-2 images. Comput. Electron. Agric. 184, 106090. https://doi.org/10.1016/j.compag.2021.106090

Xu, R., Chen, Z., Zuo, W., Yan, J., Lin, L., 2018. Deep Cocktail Network: Multi-source Unsupervised Domain Adaptation with Category Shift. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 3964–3973. https://doi.org/10.1109/CVPR.2018.00417

Zhao, S., Li, B., Reed, C., Xu, P., Keutzer, K., 2020. Multi-source domain adaptation in the deep learning era: A systematic survey. arXiv.

Zhao, S., Wang, G., Zhang, S., Gu, Y., Li, Y., Song, Z., Xu, P., Hu, R., Chai, H., Keutzer, K., 2019. Multi-source distilling domain adaptation. arXiv. https://doi.org/10.1609/aaai.v34i07.6997