MDPI

Technical Note

Deep-Transfer-Learning Strategies for Crop Yield Prediction Using Climate Records and Satellite Image Time-Series Data

Abhasha Joshi [1], Biswajeet Pradhan [1,*], Subrata Chakraborty [1,2], Renuganth Varatharajoo [3], Shilpa Gite [4,5] and Abdullah Alamri [6]

[1] Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), School of Civil and Environmental Engineering, Faculty of Engineering & IT, University of Technology Sydney, Ultimo, NSW 2007, Australia; abhasha.joshi@student.uts.edu.au (A.J.); subrata.chakraborty@une.edu.au (S.C.)

[2] School of Science and Technology, Faculty of Science, Agriculture, Business and Law, University of New England, Armidale, NSW 2351, Australia

[3] Department of Aerospace Engineering, University Putra Malaysia (UPM), Serdang 43400, Malaysia; renu@upm.edu.my

[4] Department of Artificial Intelligence and Machine Learning, Symbiosis Institute of Technology, Pune 412115, India; shilpa.gite@sitpune.edu.in

[5] Symbiosis Centre for Applied Artificial Intelligence (SCAAI), Symbiosis Institute of Technology, Symbiosis International (Deemed University) (SIU), Pune 412115, India

[6] Department of Geology and Geophysics, College of Science, King Saud University, Riyadh 11451, Saudi Arabia; amsamri@ksu.edu.sa

* Correspondence: biswajeet.pradhan@uts.edu.au or biswajeet24@gmail.com

Abstract: The timely and reliable prediction of crop yields on a larger scale is crucial for ensuring a stable food supply and food security. In the last few years, many studies have demonstrated that deep learning can offer reliable solutions for crop yield prediction. However, a key challenge in applying deep-learning models to crop yield prediction is their reliance on extensive training data, which are often lacking in many parts of the world. To address this challenge, this study introduces TrAdaBoost.R2, along with fine-tuning and domain-adversarial neural network deep-transfer-learning strategies, for predicting the winter wheat yield across diverse climatic zones in the USA. All methods used the bidirectional LSTM (BiLSTM) architecture to leverage its sequential feature extraction capabilities. The proposed transfer-learning approaches outperformed the baseline deep-learning model, with mean absolute error reductions ranging from 9% to 28%, demonstrating the effectiveness of these methods. Furthermore, the results demonstrate that the semi-supervised transfer-learning approach using the two-stage version of TrAdaBoost.R2 and fine-tuning achieved a superior performance compared to the domain-adversarial neural network and standard TrAdaBoost.R2. Additionally, the study offers insights for improving the accuracy and generalizability of crop yield prediction models in diverse agricultural landscapes across different regions.

Keywords: crop yield prediction; deep learning; transfer learning; TrAdaBoost; Bidirectional LSTM; fine-tuning; food security

1. Introduction

Ensuring a stable food supply and food security relies heavily on the timely and dependable prediction of crop yields at a larger scale [1–4]. Recent advancements in computing resources and algorithms have enabled the use of more sophisticated data-driven models, like deep learning, for various prediction problems. A large number of studies have demonstrated that deep learning can offer reliable solutions for crop yield prediction [5–8]. In particular, sequential models such as long short-term memory (LSTM), bidirectional LSTM (BiLSTM), and 1D convolutional neural networks (1DCNNs) have emerged as effective tools for predicting crop yields [9,10].

A key challenge in applying deep-learning models to crop yield prediction is the models' dependence on large training data [6,10]. Insufficient data can lead to overfitting and underfitting [11,12]. In the former condition, models cannot learn from the training data, while, in the latter, models perform well on training data but poorly on unseen test data. This limitation restricts their use in areas with limited historical yield data. Moreover, a model trained on data from one region may not perform well in entirely new locations because of the domain shift [13]. One of the reasons existing deep-learning-based crop yield prediction research has predominantly focused on specific regions of the world is the availability of abundant historical data in those areas [6]. Typically, remote sensing and environmental data are paired with historical crop yield statistics for regional scale crop yield prediction [14]. While remote sensing and environmental data are globally available due to advancements in satellites and sensors, the target data—historical yield statistics—often lack sufficient quantities and regular intervals in many countries.

Transfer learning emerges as a promising technique for overcoming the difficulties of modelling in scenarios where data are scarce. Transfer-learning [15] techniques use information gained in an area with sufficient data to improve the generalisation in an area with limited training data. Transfer learning has proven effective in various tasks, including image classification [12], crop mapping [16,17], vegetation monitoring [18], and water resource management [19]. In crop yield prediction, researchers are exploring the integration of transfer learning with deep learning to improve model generalizability. For instance, Wang, Tran, Desai, Lobell and Ermon [13] successfully applied deep-learning techniques and fine-tuning-based transfer learning to predict soybean yield in Brazil. The study demonstrated the potential of deep learning and transfer learning for crop yield prediction in data-scarce regions. Ma, et al. [20] addressed the generalizability issue of machine-learning models for crop yield prediction by introducing an unsupervised domain adaptation approach. Their unsupervised adaptive domain adversarial neural network, coupled with multiple input variables, demonstrated remarkable performance in both local and transfer settings, indicating its potential to enhance crop yield prediction across diverse regions. Priyatikanto, et al. [21] investigated the generalizability and transferability of maize yield prediction models across the US corn belt by employing three domain adaptation algorithms: the domain adversarial neural network (DANN), Kullback–Leibler importance estimation procedure, and regular transfer neural network (RTNN). Among these algorithms, the DANN exhibited promising results in model generalisation across regions.

While unsupervised domain adaptation methods like the DANN offer promising results in crop yield prediction without requiring labelled target data [20,21], they may not generalise well to unseen target domains that significantly differ from the source domain. Moreover, feature-based methods like the DANN are not applicable to domain adaptation issues where there is a covariate shift [22]; i.e., when the source and target domains share the same labelling functions, potentially impairing learning. Moreover, limited yield data are available in many regions at infrequent intervals or for specific locations. Thus, semi-supervised transfer-learning techniques might be more suitable in such a scenario. Fine-tuning is one of the widely used semi-supervised transfer-learning methods. However, they also not without challenges and are susceptible to negative transfer [23]. Simply applying all source domain data to the target domain for fine-tuning, without proper selection, can lead to negative transfer learning.

TrAdaBoost [24], another semi-supervised transfer-learning approach, combines adaptive-boosting and instance-weighting techniques. Adaptive boosting improves prediction performance by combining multiple weak learners, while instance weighting assigns different weights to samples from the source and target domains [25]. This approach reduces the influence of instances prone to negative transfer and allows the model to focus on more reliable and relevant data. When predicting crop yields across significantly different domains, TrAdaBoost can be a valuable tool to mitigate negative transfer and improve model performance.

This study investigates and compares the effectiveness of various unsupervised and semi-supervised transfer learning (TL) methods for predicting crop yield. The main contributions of the study are as follows:

- This study introduced deep-transfer learning (DTL) strategies that combine the TrAdaBoost algorithm with a BiLSTM model to predict crop yield in data-scarce regions.
- This paper quantitatively evaluates the impacts of four deep-transfer learning (DTL) strategies: fine-tuning (FT), the domain-adversarial neural network (DANN), TrAdaBoost.R2, and a two-stage TrAdaBoost.R2 algorithm on crop yield prediction across different climatic regions. These strategies leverage the sequential feature extraction capabilities of BiLSTM for the task. While previous studies primarily employed multi-layer perceptron networks as feature extractors in their models, our study opts for the BiLSTM model as the base model, given the sequential nature of our input data.

The remainder of this paper is organised as follows: Section 2 details the proposed method for yield prediction, including deep-transfer-learning techniques, the experimental data, and implementation details. Our experimental results are presented in Section 3, and a discussion of the results is presented in Section 4. Finally, Section 5 concludes the paper.

2. Materials and Methods

2.1. Study Area

In this study, winter wheat was selected as the study crop, and the winter-wheat-growing regions in the USA were selected as study areas (Figure 1). Wheat ranks among the top three most commonly consumed staple foods globally [26]. The USA was the second largest wheat exporter in the world in 2021, accounting for 13.1 % of the total wheat exported [27,28]. Moreover, the USA is the fourth largest wheat producer after China, Russia, and India, and around 8.1% of the global wheat was produced here in 2021. Winter wheat varieties, planted in the preceding fall, dominate US wheat production, representing between 70–80% of total wheat production [29]. Prediction of wheat at a regional scale before harvest and mapping the spatial distribution of the wheat area in the USA are important for supply chain management in agribusiness, adapting crop management practices and ensuring their and regional food security. In this study, predictions were made at the county scale.
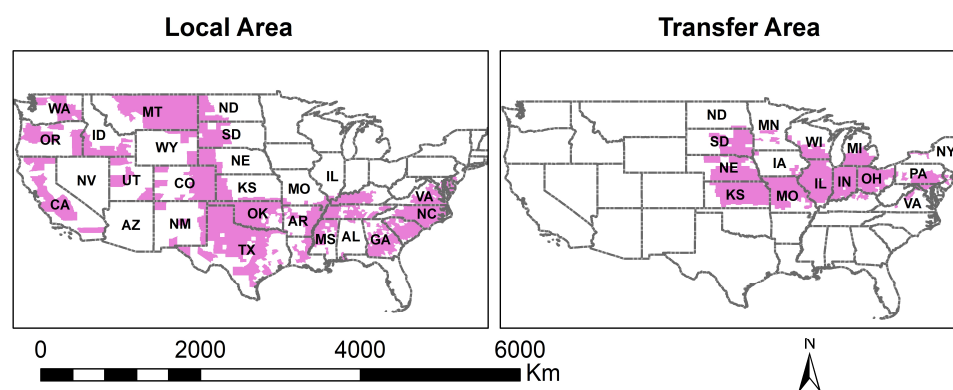
Figure 1. Study area showing local and transfer location. The colored areas on the map denote the study counties.

Transfer experiments were conducted between different climatic regions within the USA. Köppen classification maps of the present day [30] were used to identify the climate classes for each county. The Köppen–Geiger climate classification system categorises climates into six main groups based on monthly temperature and precipitation data. Each group has further subdivisions representing variations within the main class. This classification system is based on the idea that different climate zones support different types of vegetation. For the transfer experiments, we selected counties within the arid and temper-

ate climate classes as the local area (source domain) and counties within the climate class cold and subclass "no dry season, hot summer" climate as the transfer area (target domain). For the counties that fall under more than one zone in the above map, we classified them as the class in which the majority of the area falls.

2.2. Dataset and Pre-Processing

This study utilises remote sensing and weather data as inputs to characterise crop health and growth conditions. Previous studies have shown that time-series remote sensing data and meteorological data are important predictors in regional-scale yield prediction studies [14,31]. Similarly, in transfer learning across different ecological zones for yield prediction, these variables have been found to be applicable [20].

The enhanced vegetation index (EVI) was the remote sensing data used in the study. EVI is a measure of the greenness of vegetation and serves as an indicator of the quantity of healthy vegetation [32]. EVI offers improved sensitivity in high-biomass regions and reduces noise from the canopy background and atmosphere. It has a strong correlation with gross primary production (GPP). The data for the indices were obtained from the MOD13Q1 V6.1 product, a 16-day global 250 m vegetation index and reflectance product of the moderate-resolution imaging spectroradiometer (MODIS). Only pixels with good-quality data (DetailedQA = 0) were used to obtain the time-series data, ignoring data with snow or cloud cover.

EVI is calculated as follows:

$$\text{EVI} = 2.5 \times \frac{(\text{NIR} - \text{Red})}{\text{NIR} + \text{C1} \times \text{Red} - \text{C2} \times \text{Blue} + \text{L}} \tag{1}$$

where NIR, Red, and Blue are reflectance acquired in the near-infrared (841–876 nm), red (620–670 nm), and blue (459–479 nm) portions of the electromagnetic spectrum, respectively. The variable L accounts for soil and canopy background effects, while C1 and C2 are coefficients used to correct atmospheric influences. The standard values used are L = 1, C1 = 6, and C2 = 7.5.

For weather information, we used Terraclimate [33], which is global weather data at monthly intervals prepared by combining the WorldClim dataset with Climatic Research Unit (CRU) Ts4.0 and the Japanese 55-year Reanalysis (JRA55) data. The spatial resolution of the data is 1/24th degree (~4.6 km). The data used in the study are downward surface shortwave radiation, wind speed maximum temperature, and soil moisture. This climatic variable showed a high correlation with crop yield [31]. All the above-mentioned input data are available globally for any part of the world.

The study predicted the end-of-season yield for winter wheat in the study counties. Winter-wheat-growing season in the study is from September–October to May–July in the following year [34]. EVI and weather data from October of the plantation year to June of the harvest year were selected as input data for the predictive models.

We employed Google Earth Engine (GEE) for data collection and pre-processing. The model used monthly EVI and weather data as input. The 16-day EVI data were converted to monthly time-series data using a weighted average scheme, where the weights were based on the degree of temporal overlap. For each input feature, we used the crop map to eliminate irrelevant observations from non-winter-wheat areas. Subsequently, within each administrative unit (county), we extracted all relevant features and aggregated each feature to the administrative division level by calculating the mean value of all extracted pixels within that county.

Target yield data, consisting of county-level winter wheat yields, were obtained from the National Agricultural Statistics Service (NASS) QuickStats database of the United States Department of Agriculture [29]. These data were used to train and test the crop yield prediction model. All yields were reported in units of metric tonnes per cultivated hectare (t/Ha). Winter wheat yields in the transfer area have been generally higher than in the local area during the study period. The Cropland Data Layer (CDL) was used to

delineate the annual cultivation areas of winter wheat within each county. The CDL is an annual georeferenced, crop-specific land cover map dataset produced by the USDA-NASS. The CDL is derived from moderate-resolution satellite imagery combined with extensive agricultural ground truth data [35] to achieve a spatial resolution of 30 m.

2.3. Transfer Learning

The idea of using knowledge from one task to improve learning on another is not new and has existed under different names like inductive transfer [36], multi-task learning [37], and incremental/cumulative learning [38]. However, the rise of deep learning has significantly increased the popularity of transfer learning. Deep neural networks need massive datasets for training, which can be expensive and time-consuming to acquire. Transfer learning helps address this challenge. The goal of transfer learning is to learn knowledge using data from the source domain that can also be applied to the target domain (Figure 2). Transfer-learning approaches can be broadly categorised into four types: instance-based, parameter-based, relation-based, and feature-based [39]. The instance-based transfer-learning approach adjusts the weights of certain data from the source domain and combines them with a few labelled data from the target domain to make predictions in the target domain. Parameter-based transfer learning takes some parameters or prior distributions of hyperparameters from the pre-trained model from the source domain as a starting point. The model's parameters are then fine-tuned on the target data to improve performance on the new task. Feature-based transfer-learning methods aim to discover effective feature representations to minimise domain differences and reduce errors in classification or regression models. Relational-based transfer learning is specifically designed for tasks where data can be represented by relationships between entities. This approach focuses on transferring the logical relationship or rules learned between domains.
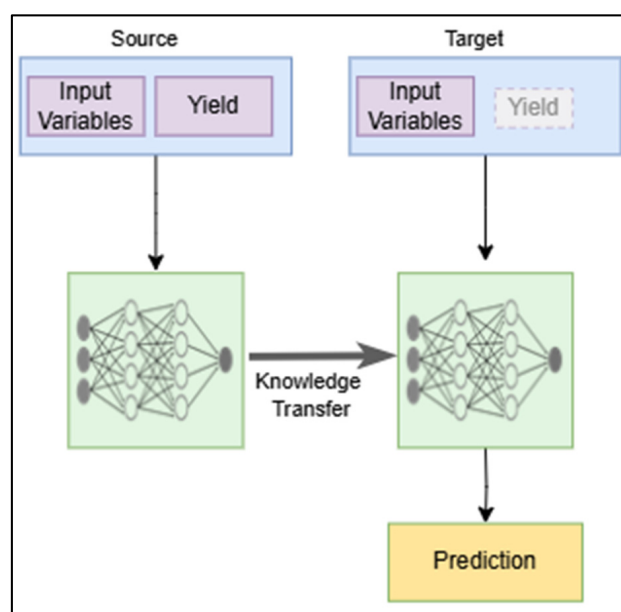
Figure 2. Transfer learning for crop yield prediction. For unsupervised transfer learning, we assume there are no yield data in the target area, and, for semi-supervised scenarios, limited historical yield data are available in the target location.

In this study, we used four transfer-learning approaches: instance-based TrAdaBoost.R2 and two-stage TrAdaBoost.R2, feature-based domain-adversarial neural network, and parameter-based fine-tuning. Across all approaches, BiLSTM was utilised as the base model.

The TrAdaBoost algorithm, proposed by Dai, Yang, Xue and Yu [24], is a transfer-learning algorithm originally developed for the classification field. It assumes that certain source domain data may be effective for learning in the target domain, while others may

not and could even be detrimental. It is based on "reverse boosting". During each boosting iteration, TrAdaBoost strategically adjusts instance weights. When a target instance is misclassified, its weight is increased, encouraging the model to focus on these challenging examples. Conversely, misclassified source instances experience a decrease in weight. This approach helps TrAdaBoost identify and utilise source data points that are most relevant to the target domain while disregarding those that are significantly different. Building upon the principles of AdaBoost.R2 and TrAdaBoost, Pardoe and Stone [40] proposed TrAdaBoost.R2, an instance-based regression transfer algorithm. This algorithm combines the source and target datasets into a single set and handles the reweighting of each training instance independently. TrAdaBoost.R2 can become susceptible to overfitting as the number of boosting iterations increases and decreases in accuracy beyond a certain point. To address this limitation, the authors also introduced the two-stage TrAdaBoost.R2 algorithm. Two-stage TrAdaBoost.R2 assigns weights to the instances in two steps. In the first stage, the algorithm gradually reduces the weights of source data points until reaching a threshold determined by cross-validation. This effectively minimises the influence of potentially irrelevant source data on the model. In stage 2, source instance weights are frozen, while target instance weights are updated according to the standard AdaBoost.R2 procedure. Importantly, only the hypotheses generated in the second stage are retained and used to determine the output of the resulting model. To the best of our knowledge, this is the first study in which these instance-based methods are applied for crop yield prediction.

Another transfer-learning strategy used in this study is fine-tuning [41]. Fine-tuning involves pre-training a model on a data-rich source domain and then refining it with a few labelled samples from a target domain. First, the base neural network is trained on the source domain. Usually, the weights of some of the layers of the trained network are frozen while others are made trainable. One common approach is to freeze the initial few layers responsible for feature extraction of the trained deep-learning model while the predictor part of the model is fine-tuned using data from the target domain. In this study, a transferable BiLSTM model was constructed by keeping the weights of the BiLSTM layers unchanged while fine-tuning the weights of the dense layers of the model.

The final transfer-learning method employed in this study is the DANN [42]. It is an unsupervised technique designed to extract domain-invariant features, meaning features that are relevant to the learning task and remain applicable even when the source and target domains have different data distributions. DANN integrates an adversarial component to align the feature distributions across domains, thereby enhancing the network's generalisation capabilities. DANN typically consists of three main components: the feature extractor, domain classifier, and regressor. The feature extractor is responsible for learning features from the input data. A domain classifier is a network that takes the extracted features and attempts to predict whether the data originated from the source or target domain. This is trained in an adversarial setting. By minimising the domain classifier's ability to distinguish between domains, the model attempts to make the features extracted by the first component indistinguishable between domains. Finally, a regressor utilises the extracted features to perform the main learning task, such as predicting yield.

2.4. BiLSTM Model

The BiLSTM model was selected as the base model in all the transfer-learning approaches. BiLSTM has proven effective in processing time-series remote sensing data for various tasks, including crop detection [43], data imputation [44], and change detection [45]. BiLSTM is a recurrent neural network (RNN) used for processing sequential data [46,47]. It builds upon long short-term memory (LSTM) [48] and is designed to better capture long-term dependencies by addressing the vanishing-gradient problem in RNNs. Unlike standard LSTMs, which process data in a single direction, a BiLSTM model consists of two LSTM components. One LSTM processes the data in the forward direction, while the other processes it in the backward direction. This allows BiLSTM to effectively capture features from sequential data. The model selected for this study consists of two Bidirectional LSTM

(BiLSTM) layers followed by three Dense layers. Additionally, a Dropout layer is inserted between each of the BiLSTM and Dense layers to help mitigate overfitting (Figure 3).
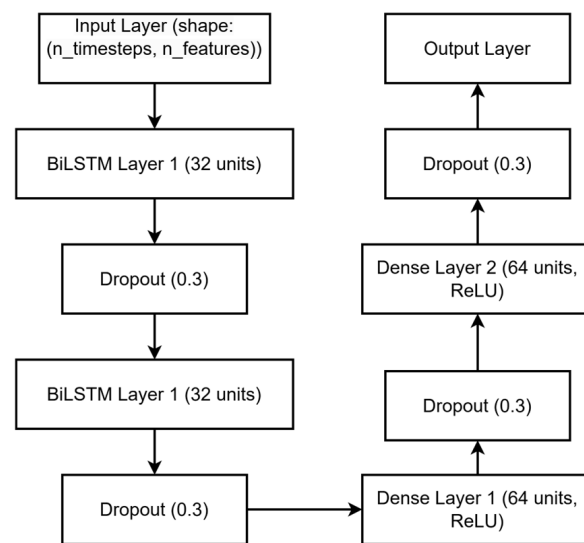
Figure 3. Architecture of the base BiLSTM model.

2.5. Experimental Setup

In this study, the model trained using data from the local area was adapted for prediction on the transfer area using transfer learning. The Koppen classification data divided the wheat-growing counties of the USA into local and transfer areas. Details regarding these areas are provided in the study area section. The years 2019 and 2020 were selected as test years for evaluating the transfer-learning approaches. Data from 2008 to the year preceding the test year were used for model training. The pre-processed dataset consisted of a total of 6121 data points from the local area and 2225 data points from the transfer area. Specifically, for the years 2019 and 2020, the number of data points from the transfer area used to test the model were 104 and 197, respectively.

For the semi-supervised transfer-learning approaches (TrAdaBoost.R2, two-stage TrAdaBoost.R2, and Fine-Tuning), a subset comprising 10% of the available input–target pairs from the transfer area covering the training period was utilised for transfer learning. In contrast, the unsupervised DANN approach utilised all unlabelled input features from the transfer area within the training period. For instance, to predict yields for 2019, the semi-supervised deep-transfer-learning models were trained using input–target data pairs from 2008 to 2018 from the local area, along with 10% of the input–target data pairs from the transfer area for the same period. Meanwhile, the DANN model used input–target data pairs from 2008 to 2018 from the local area and all unlabelled input variables from the transfer area during the same period.

We also compared the results of the transfer-learning approach with those of the base Bi-LSTM model and a Random Forest model trained exclusively on local area data and then directly applied to predict yield in the transfer area. Random Forest [49] is a widely used algorithm that has been found to provide robust performance across a range of tasks, including crop yield prediction [31].

To identify optimal hyperparameters for each model, we employed a grid search technique, and data from 2008 to 2018 were used. The hyperparameter search space and selected hyperparameter for different models are presented in Table 1. The models were implemented within the Python 3.10.6 environment, utilising the TensorFlow framework. The ADAPT [50] library was also used for implementing transfer-learning approaches. Training leveraged a high-performance computing (HPC) server featuring an Intel Xeon Gold 6238R processor clocked at 2.2GHz with 28 cores, 180GB RAM (Six Channel), and a robust NVIDIA Quadro RTX 6000 Passive GPU, boasting 4608 cores, 576 Tensor Cores, and

24GB of memory. The experiment for each model was repeated ten times, and the mean results are presented in the paper.

Table 1. Hyperparameter for different models.

Method	Parameter	Search Space	Selected Parameters
DANN	Lambda	0.0001, 0.001, 0.01, 0.1, 1, 2, 5	0.01
	Learning rate	0.01, 0.001, 0.0001	0.0001
	Epoch	50, 100, 150, 200	150
	Batch size	50, 100	50
TrAdaBoost.R2, two-stage TrAdaBoost.R2	Learning rate	0.01, 0.001	0.01
	Epoch	100, 150	100
	Batch size	50,100	100
Fine-tuning	Epoch (main)	50, 100, 150	100
	Learning rate (main)	0.01, 0.001, 0.0001	0.001
	Batch size (main)	50, 100	100
	Epoch (fine tune)	50, 100, 200, 400, 800	100
	Learning rate (fine tune)	0.01, 0.001, 0.0001	0.001
	Batch size (Fine tune)	20, 50, 100	20
Random Forest	Number of Trees	10, 100, 200, 500	100
	Bootstrap	True, False	True
	Max depth	None, 10, 20	None
	Minimum sample leaf	1, 2, 4	2
	Minimum sample split	2, 5, 10	5

2.6. Performance Evaluation

In this experiment, we utilised the coefficient of determination (R^2) and mean absolute error (MAE) to assess the performance of the model. The R^2 is expressed as a fraction and represents the degree of agreement between the true value and the predicted value, measuring the proportion of variance in the dependent variable explained by the independent variable. MAE is the average absolute difference between the predicted values and the actual value.

$$\text{MAE} = \frac{1}{n}\sum\nolimits_{i=1}^{n}\left|Y_i - \hat{Y}_i\right| \tag{2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2} \tag{3}$$

where Y_i denotes the actual yield values, $\hat{Y}_i$ represents the predicted yield values, and $\overline{Y}$ denotes the mean of the actual yield values.

3. Results

In this study, various transfer-learning techniques were employed with BiLSTM models for winter wheat yield prediction across different climatic zones in the USA. The performance of these methods was evaluated based on MAE and R^2 values over the years 2019 and 2020. The boxplot (Figure 4) illustrates the distribution of winter wheat yields for local and transfer areas over the 13-year period (2008–2020). The transfer area exhibited generally higher winter wheat yields than the local area.
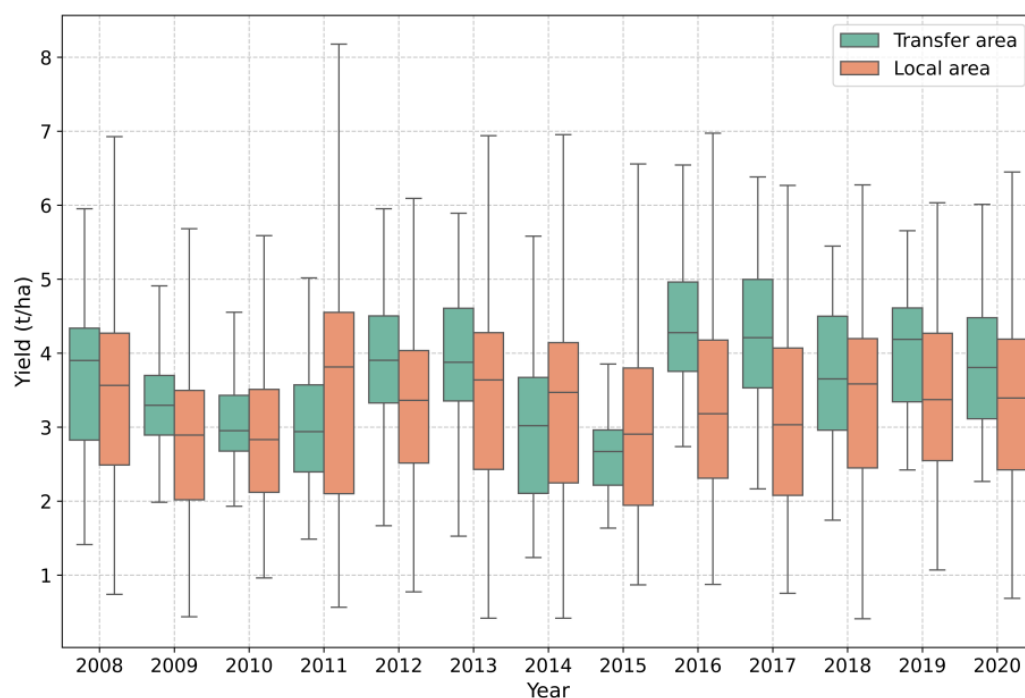
Figure 4. Comparison of crop yields between transfer and local areas from 2008 to 2020.

Table 2 summarises the results obtained from different transfer-learning approaches. The mean of R^2 for the test years for the baseline BiLSTM model without using transfer learning was 0.19, and the MAE was 0.55. The R^2 value implies that the model explains only 19% of the variance in yield in the transfer location, suggesting the model may not be used directly for yield prediction in the transfer location. The Random Forest model also showed poor performance in the transfer location, with a mean MAE of 0.55 and an R^2 of 0.24, indicating a limited predictive capability for yield in that region.

Table 2. R^2 and MAE of different models in transfer location for test years.

|	Transfer Method	2019		2020	
		MAE	R^2	MAE	R^2
	Random Forest	0.56	0.28	0.55	0.20
1	Bi-LSTM Model Without TL	0.53	0.22	0.57	0.16
2	Fine tuning	0.45	0.45	0.41	0.55
3	DANN	0.47	0.41	0.52	0.26
4	TrAdaBoost.R2	0.45	0.41	0.46	0.49
5	Two-stage TrAdaBoost.R2	0.41	0.51	0.42	0.53

The unsatisfactory performance of the models without transfer learning suggests that the relationship learned between the input features and crop yield in the local area is not generalisable to the target domain. The low-dimensional visualisation of the input data of local and transfer locations using t-distributed stochastic embedding (t-SNE) [51] shows distinct clusters for the data for the local and transfer area (Figure 5a), suggesting different distributions of input variables in the local and transfer area. The distribution of yield in the transfer and local area also differs (Figure 5b). The mean, median, and standard deviation across all years and counties within the local area are 3.22 t/ha, 3.26 t/ha, and 1.23 t/ha, respectively, while those for the transfer area are 3.74 t/ha, 3.70 t/ha, and 0.96 t/ha, respectively.
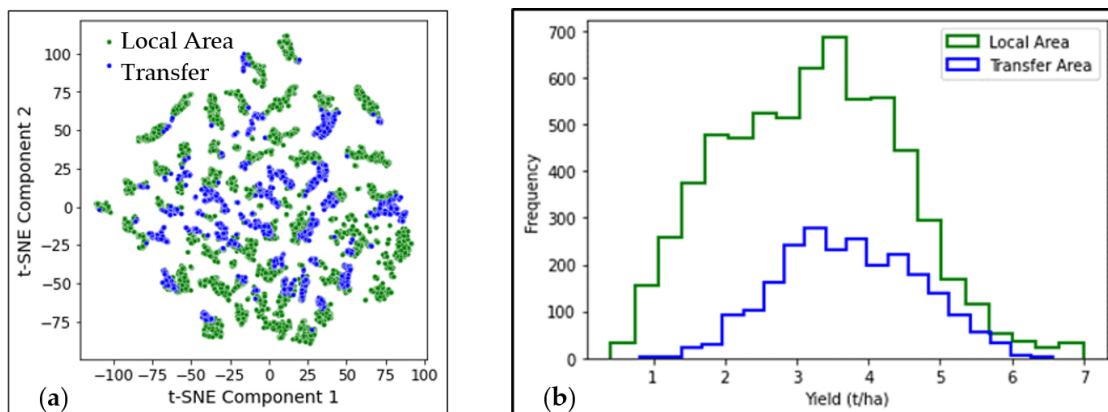
Figure 5. Distribution of data in local and transfer area. (**a**) T-SNE plot of input variables, and (**b**) histograms showing the winter wheat yield distribution of counties in the local area and transfer area from 2008 to 2020.

Compared to the baseline Bi-LSTM model and Random Forest model without transfer learning, all transfer-learning approaches showed improvements in both the MAE and R^2 values (Table 2). The DANN improved the results for both years with a mean MAE of 0.50 and R^2 of 0.34 for test years. All semi-supervised transfer-learning approaches demonstrate notable improvement in performance. In particular, fine-tuning and the two-stage TrAdaBoost.R2 approach achieved the best and had similar performance. Fine-tuning achieved a mean MAE of 0.43 and a mean R^2 of 0.50, and the two-stage TrAdaBoost.R2 achieved a mean MAE of 0.42 and a mean R^2 of 0.52. The standard TrAdaBoost.R2 technique also achieved a comparable performance to both methods, with a mean MAE of 0.46 and a mean R^2 of 0.45. Therefore, for the same base model and hyperparameter setting, the two-stage TrAdaBoost.R2 performs better than TrAdaBoost.R2 for crop yield prediction. Moreover, compared to other models, the two-stage TrAdaBoost.R2 had a consistent performance through the years. However, the computational time for the two-stage TrAdaBoost.R2 is significantly higher than that of other approaches.

Figure 6 presents the spatial distribution of the mean of the absolute error for the winter wheat yield prediction in the years 2019 and 2020. Darker colours indicate larger absolute errors for each model. The spatial distribution of the absolute error shows clusters of highly erroneous counties in the DANN method. The two-stage TrAdaBoost.R2 and fine-tuning methods show a lower absolute error across all the study areas. Similarly, in the scatterplot (Figure 7), both the two-stage TrAdaBoost.R2 and fine-tuning methods exhibited the highest level of agreement between the reported and predicted yield. The scatterplot also reveals that the DANN generally exhibited underprediction. The mean yields for the transfer location in the test years 2019 and 2020 were 4.0 t/ha and 3.8 t/ha, respectively, which is substantially higher than the mean yield in the local area during the study period (3.26 t/ha). The difference in yield distribution in the transfer and local areas is also evident in the boxplot presenting the yield in the transfer and local areas (Figure 4). Since the DANN was not trained on yield data from the transfer location, this likely explains the underprediction. However, to a lesser extent, fine-tuning also showed underprediction, indicating that it was similarly unable to adequately learn the yield distribution from the limited data available from the transfer location. The two-stage TrAdaBoost.R2 exhibited the least occurrence of this issue. The scatterplot of TrAdaBoost.R2 (Figure 7(b1,b2)) shows several points arranged in a straight line parallel to the reported yield axis. This pattern indicates that, for a range of different reported yields, the model predicted similar yields, suggesting an issue of overfitting.
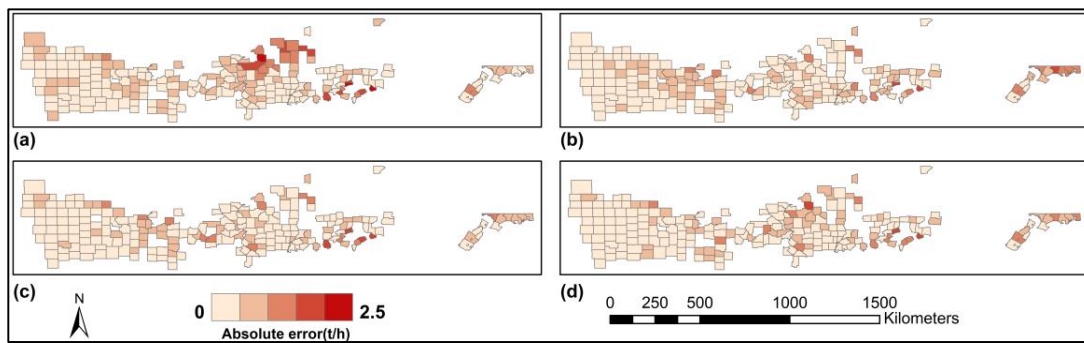
Figure 6. The absolute error maps of (**a**) DANN, (**b**) TrAdaBoost.R2, (**c**) two-stage TrAdaBoost.R2, and (**d**) fine-tuning-based transfer models averaged over the years 2019 and 2020.
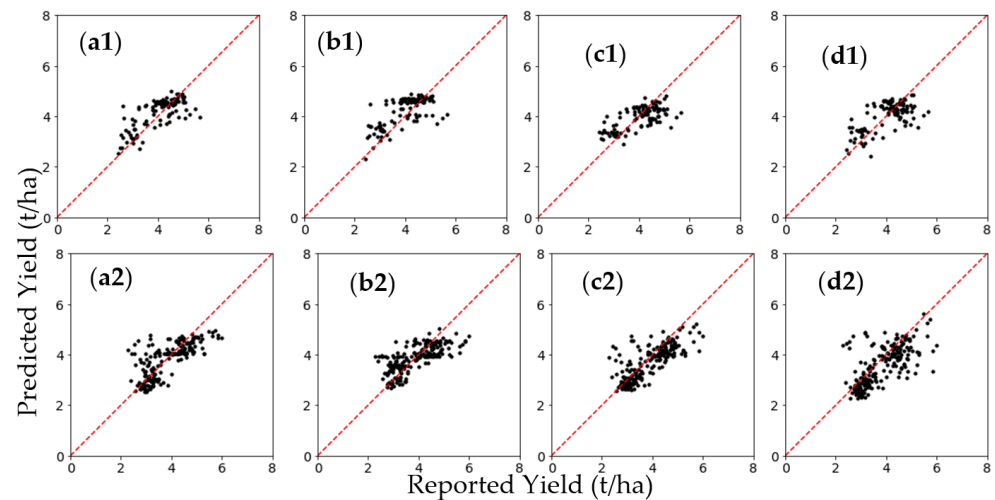
Figure 7. Scatterplot presents the relationship between reported and predicted yields across four transfer learning techniques: (**a**) two-stage TrAdaBoost.R2, (**b**) TrAdaBoost.R2, (**c**) fine tuning, and (**d**) DANN over testing years (1) 2019 and (2) 2020.

To investigate the impact of BiLSTM layers on learning transferable features by the transfer model, we conducted two-stage TrAdaBoost.R2 experiments without incorporating BiLSTM layers. The experiment employed the architecture comprising the final three Dense layers (including the output layer) with Dropout layers between them. We run these experiments 10 times and averaged the results. As depicted in Table 3, transfer models with BiLSTM layers outperformed those without, both in terms of MAE and R2 for the test years. Additionally, we tried using two more Dense layers instead of a BiLSTM layer, but the performance was significantly worse, so we did not include those results.

Table 3. Comparison of performance of two-stage TrAdaBoost.R2 with and without BiLSTM layers.

	Transfer Method	2019		2020	
		MAE	**R^2**	**MAE**	**R^2**
1	Two-stage TrAdaBoost.R2 with BiLSTM component in the base layer	0.41	0.51	0.42	0.53
2	Two-stage TrAdaBoost.R2 without BiLSTM component in the base layer	0.49	0.33	0.55	0.29

4. Discussion

While deep learning and machine learning offer powerful tools for modelling complex, nonlinear relationships between input features and crop yield [5,14], these models are often

limited by their domain-specific nature. As a result, they may not generalise well to different regions with varying data distributions. For instance, the study by Ma, Zhang, Yang and Yang [20] observed a decline in the performance of RF and MLP models when trained on data from a specific region and applied to a different region. A similar observation was found in our study, where models trained without transfer learning (base RF and Bi-LSTM) showed poor performance when applied to the transfer location. We employed transfer learning techniques to address this limitation, which demonstrated significant improvements in model performance with reductions in MAE ranging from 9% to 28%.

The results showed that the fine-tuning and two-stage version of TrAdaboost.R2 exhibited superior performance for crop yield prediction in areas with limited training data. Fine-tuning utilises feature extractors trained on different data, enabling the model to adapt efficiently to new tasks or domains by leveraging existing knowledge [41]. TrAdaboost.R2, which is an instance-based transfer learning method, iteratively assigns weights to the data points based on their contribution to prediction. However, the instance-based Kullback-Leibler Importance Estimation Procedure showed inferior performance for the maise yield prediction in transfer locations, likely due to overfitting [21]. In our study, the scatterplot of TrAdaBoost.R2 (Figure 7(b1,b2)) shows a pattern indicative of the model overly fitting the training data. The two-stage version of TrAdaBoost.R2 likely addresses the overfitting concerns raised by using a staged approach for updating weights. Unsupervised domain adaptation has also proven to be effective for yield prediction in prior studies [52]. However, the DANN method did not perform satisfactorily in our study, particularly for the year 2020. This could be attributed to a greater domain shift in our dataset, as indicated by the lower R^2 values of the base models in our study compared to the R^2 values of the base models in the transfer locations reported in those studies. Also, limited historical yield statistics are available in many regions, so effectively using the available data with a semi-supervised transfer learning method is reasonable.

The robustness of this study lies in its application of an advanced feature extractor, the BiLSTM, unlike previous studies that have mostly relied on the multilayer perceptron. The study demonstrated that the Two-stage Tradaboost model with a Bi-LSTM base layer outperformed the MLP-based version in terms of R^2 and MAE across both test years (Table 3), indicating superior feature learning capabilities of the Bi-LSTM in this context. The transfer learning model with BiLSTM achieved a 16% and 23% reduction in MAE compared to the transfer learning model with MLP for the two test years. Advanced deep learning models, such as LSTM and 1D-CNN, have already demonstrated their effectiveness in yield prediction, outperforming MLP based approaches [6,9,13,53]. The findings of our study suggest that these techniques can provide superior feature representation, in the context of transfer learning for yield prediction as well. Particularly, BiLSTM provides a deeper understanding of context by processing sequences in both directions and then combining these analyses into a single, enhanced representation [44,47]. Additionally, we employed Köppen climate classification data to select distinct areas for source and target locations.

Furthermore, the study indicated that combining fine-tuning with the instance-based two-stage TrAdaBoost.R^2 method can lead to a robust transfer-learning approach for yield prediction. The two-stage TrAdaBoost.R^2 method assigns weights to source data based on its contribution to the prediction, rather than treating all data equally. This approach is particularly valuable when extracting information from multi-source domains with distinct characteristics. Fine tuning updates the model's parameters (weights) using data from the transfer location. Both approaches were found to be effective in our study, and we could potentially combine them to complement each other.

However, it is also important to acknowledge that there are challenges in implementing the approach proposed in this study in certain regions. Firstly, we used MODIS data with a spatial resolution of 250 m for vegetation indices, which is suitable for county-level yield prediction in countries with larger agricultural farms. However, these data may suffer from mixed pixel issues [54] in areas with smaller farms, which is common in many developing countries. One potential solution to this issue is to utilise higher-resolution data

such as Sentinel and Landsat imagery. Another challenge is that this method also requires crop-type data to extract input variables from the area of a particular crop class only. While this study utilises the Cropland Data Layer, such data are not available in many countries. Experimentation with global low-resolution data and static crop-type maps in areas where there is not a significant deviation in farming practices could be a potential solution. It is also important to note that the target region is significantly affected by cloud cover. Consequently, a considerable portion of the data has been removed, leading to data loss. Approximately 35% of the data points from the study area were excluded from this study due to missing EVI values for one or more of the analysed months. Future studies should quantify the impact of noisy remote sensing data and explore alternative strategies like using the interpolation or imputation of missing data techniques to mitigate these data gaps. Additionally, this study integrated remote sensing and weather data for yield prediction. Other variables, such as soil fertility, crop cultivar types, and management practices, may not be fully captured by remote sensing data but these factors could improve the accuracy of yield predictions, particularly when modelling larger areas with distinct agricultural domains. Finally, exploring differences in feature interactions between local and transfer locations would be an interesting area for future research using explainability techniques and other analytical methods. Such studies could help identify which features generalise well across domains and which require domain adaptation techniques to enhance model performance in transfer-learning scenarios.

5. Conclusions

The scarcity of historical crop yield data poses a significant challenge for developing machine-learning models for accurate yield prediction. Moreover, ML models trained on data from one location often fail to predict satisfactorily when applied directly to geographically distinct areas with different environmental conditions. This study proposes different deep-transfer-learning strategies for crop yield prediction, leveraging satellite-derived vegetation indices and meteorological variables in conjunction with a BiLSTM model. The study develops the instance-based two-stage TrAdaBoost.R2 and TrAdaBoost.R2, parameter-based fine-tuning, and feature-based DANN with BiLSTM as the base model and provides a comprehensive quantitative evaluation. The effectiveness of the proposed approach is validated by implementing it in diverse climatic zones within the United States. In conclusion, the proposed deep-transfer-learning strategies show promise in improving the crop yield prediction accuracy in regions with limited historical data. The results also demonstrate that the semi-supervised transfer-learning approach using the two-stage TrAdaBoost.R2 and fine-tuning achieved a superior performance compared to the DANN and vanilla TrAdaBoost.R2. Future studies should evaluate the applicability of this method. This includes incorporating data from different countries and applying it to entirely new geographic regions. Further, a hybrid transfer-learning approach that incorporates parameter updates and instance selection is suggested. As the diversity of source data increases and a more robust methodology is adopted, the transfer-learning approach is expected to become even more powerful in overcoming data scarcity limitations and achieving robust crop yield predictions across various agricultural landscapes.

Data Availability Statement: All the data used are publicly available.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Food and Agriculture Organization. *The Future of Food and Agriculture–Trends and Challenges*; Annual Report; Food and Agriculture Organization: Rome, Italy, 2017.
2. Hoffman, L.A.; Etienne, X.L.; Irwin, S.H.; Colino, E.V.; Toasa, J.I. Forecast Performance of Wasde Price Projections for Us Corn. *Agric. Econ.* **2015**, *46*, 157–171. [CrossRef]
3. Sherrick, B.J.; Lanoue, C.A.; Woodard, J.; Schnitkey, G.D.; Paulson, N.D. Crop Yield Distributions: Fit, Efficiency, and Performance. *Agric. Financ. Rev.* **2014**, *74*, 348–363. [CrossRef]
4. Isengildina-Massa, O.; Irwin, S.H.; Good, D.L.; Gomez, J.K. The Impact of Situation and Outlook Information in Corn and Soybean Futures Markets: Evidence from Wasde Reports. *J. Agric. Appl. Econ.* **2008**, *40*, 89–103. [CrossRef]
5. You, J.; Li, X.; Low, M.; Lobell, D.; Ermon, S. Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017.
6. Joshi, A.; Pradhan, B.; Gite, S.; Chakraborty, S. Remote-Sensing Data and Deep-Learning Techniques in Crop Mapping and Yield Prediction: A Systematic Review. *Remote Sens.* **2023**, *15*, 2014. [CrossRef]
7. Joshi, D.R.; Clay, S.A.; Sharma, P.; Rekabdarkolaee, H.M.; Kharel, T.; Rizzo, D.M.; Thapa, R.; Clay, D.E. Artificial Intelligence and Satellite-Based Remote Sensing Can Be Used to Predict Soybean (Glycine Max) Yield. *Agron. J.* **2023**, *116*, 917–930. [CrossRef]
8. Ma, Y.; Zhang, Z.; Kang, Y.; Özdoğan, M. Corn Yield Prediction and Uncertainty Analysis Based on Remotely Sensed Variables Using a Bayesian Neural Network Approach. *Remote Sens. Environ.* **2021**, *259*, 112408. [CrossRef]
9. Wolanin, A.; Mateo-Garciá, G.; Camps-Valls, G.; Gómez-Chova, L.; Meroni, M.; Duveiller, G.; Liangzhi, Y.; Guanter, L. Estimating and Understanding Crop Yields with Explainable Deep Learning in the Indian Wheat Belt. *Environ. Res. Lett.* **2020**, *15*, 024019. [CrossRef]
10. Jiang, H.; Hu, H.; Zhong, R.; Xu, J.; Xu, J.; Huang, J.; Wang, S.; Ying, Y.; Lin, T. A Deep Learning Approach to Conflating Heterogeneous Geospatial Data for Corn Yield Estimation: A Case Study of the Us Corn Belt at the County Level. *Glob. Change Biol.* **2020**, *26*, 1754–1766. [CrossRef]
11. Torrey, L.; Shavlik, J. Transfer Learning. In *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*; IGI Global: Hershey, PA, USA, 2010; pp. 242–264.
12. Ma, Y.; Chen, S.; Ermon, S.; Lobell, D.B. Transfer Learning in Environmental Remote Sensing. *Remote Sens. Environ.* **2024**, *301*, 113924. [CrossRef]
13. Wang, A.X.; Tran, C.; Desai, N.; Lobell, D.; Ermon, S. Deep Transfer Learning for Crop Yield Prediction with Remote Sensing Data. In Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies, COMPASS 2018, San Jose, CA, USA, 20–22 June 2018.
14. Cai, Y.; Guan, K.; Lobell, D.; Potgieter, A.B.; Wang, S.; Peng, J.; Xu, T.; Asseng, S.; Zhang, Y.; You, L. Integrating Satellite and Climate Data to Predict Wheat Yield in Australia Using Machine Learning Approaches. *Agric. For. Meteorol.* **2019**, *274*, 144–159. [CrossRef]
15. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
16. Xu, J.; Zhu, Y.; Zhong, R.; Lin, Z.; Xu, J.; Jiang, H.; Huang, J.; Li, H.; Lin, T. Deepcropmapping: A Multi-Temporal Deep Learning Approach with Improved Spatial Generalizability for Dynamic Corn and Soybean Mapping. *Remote Sens. Environ.* **2020**, *247*, 111946. [CrossRef]
17. Gadiraju, K.K.; Vatsavai, R.R. Remote Sensing Based Crop Type Classification Via Deep Transfer Learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 4699–4712. [CrossRef]
18. Li, Y.; Liu, H.; Ma, J.; Zhang, L. Estimation of Leaf Area Index for Winter Wheat at Early Stages Based on Convolutional Neural Networks. *Comput. Electron. Agric.* **2021**, *190*, 106480. [CrossRef]
19. Cao, Z.; Ma, R.; Duan, H.; Pahlevan, N.; Melack, J.; Shen, M.; Xue, K. A Machine Learning Approach to Estimate Chlorophyll-a from Landsat-8 Measurements in Inland Lakes. *Remote Sens. Environ.* **2020**, *248*, 111974. [CrossRef]
20. Ma, Y.; Zhang, Z.; Yang, H.L.; Yang, Z. An Adaptive Adversarial Domain Adaptation Approach for Corn Yield Prediction. *Comput. Electron. Agric.* **2021**, *187*, 106314. [CrossRef]
21. Priyatikanto, R.; Lu, Y.; Dash, J.; Sheffield, J. Improving Generalisability and Transferability of Machine-Learning-Based Maize Yield Prediction Model through Domain Adaptation. *Agric. For. Meteorol.* **2023**, *341*, 109652. [CrossRef]
22. Gu, X.; Yu, X.; Sun, J.; Xu, Z. Adversarial Reweighting for Partial Domain Adaptation. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 14860–14872.
23. Qian, F.; Ruan, Y.; Lu, H.; Meng, H.; Xu, T. Enhancing Source Domain Availability through Data and Feature Transfer Learning for Building Power Load Forecasting. In Proceedings of the Building Simulation, Denver, CO, USA, 21–23 May 2024; pp. 1–14.
24. Dai, W.; Yang, Q.; Xue, G.-R.; Yu, Y. Boosting for Transfer Learning. In Proceedings of the 24th International Conference on Machine Learning, Corvallis, OR, USA, 20–24 June 2007; pp. 193–200.
25. Tang, D.; Yang, X.; Wang, X. Improving the Transferability of the Crash Prediction Model Using the Tradaboost. R2 Algorithm. *Accid. Anal. Prev.* **2020**, *141*, 105551. [CrossRef]

26. FAO. Staple Foods: What Do People Eat? Available online: https://www.fao.org/3/u8480e/U8480E07.HTM (accessed on 4 February 2024).
27. FAO. *World Food and Agriculture—Statistical Yearbook 2020*; FAO: Rome, Italy, 2021.
28. USDA. U.S. Wheat Exports in 2021. Available online: https://www.fas.usda.gov/commodities/wheat (accessed on 11 November 2023).
29. USDA. United States Department of Agriculture National Agricultural Statistics Service. Available online: https://quickstats.nass.usda.gov (accessed on 5 December 2023).
30. Beck, H.E.; Zimmermann, N.E.; McVicar, T.R.; Vergopolan, N.; Berg, A.; Wood, E.F. Present and Future Köppen-Geiger Climate Classification Maps at 1-Km Resolution. *Sci. Data* **2018**, *5*, 180214. [CrossRef]
31. Joshi, A.; Pradhan, B.; Chakraborty, S.; Behera, M.D. Winter Wheat Yield Prediction in the Conterminous United States Using Solar-Induced Chlorophyll Fluorescence Data and Xgboost and Random Forest Algorithm. *Ecol. Inform.* **2023**, *77*, 102194. [CrossRef]
32. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the Radiometric and Biophysical Performance of the Modis Vegetation Indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [CrossRef]
33. Abatzoglou, J.T.; Dobrowski, S.Z.; Parks, S.A.; Hegewisch, K.C. Terraclimate, a High-Resolution Global Dataset of Monthly Climate and Climatic Water Balance from 1958–2015. *Sci. Data* **2018**, *5*, 170191. [CrossRef] [PubMed]
34. USDA-NASS. *Crop Production (October 2021)*; USDA-NASS: Washington, DC, USA, 2021.
35. USDA National Agricultural Statistics Service (NASS). Cropland Data Layer. Available online: https://nassgeodata.gmu.edu/CropScape (accessed on 1 November 2022).
36. Ching, J.Y.; Wong, A.K.C.; Chan, K.C.C. Class-Dependent Discretization for Inductive Learning from Continuous and Mixed-Mode Data. *IEEE Trans. Pattern Anal. Mach. Intell.* **1995**, *17*, 641–651. [CrossRef]
37. Yang, Q.; Ling, C.; Chai, X.; Pan, R. Test-Cost Sensitive Classification on Data with Missing Values. *IEEE Trans. Knowl. Data Eng.* **2006**, *18*, 626–638. [CrossRef]
38. Zhu, X.; Wu, X. Class Noise Handling for Effective Cost-Sensitive Learning by Cost-Guided Iterative Classification Filtering. *IEEE Trans. Knowl. Data Eng.* **2006**, *18*, 1435–1440.
39. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2009**, *22*, 1345–1359.
40. Pardoe, D.; Stone, P. Boosting for Regression Transfer. In Proceedings of the 27th International Conference on International Conference on Machine Learning, Haifa, Israel, 21–24 June 2010; pp. 863–870.
41. Oquab, M.; Bottou, L.; Laptev, I.; Sivic, J. Learning and Transferring Mid-Level Image Representations Using Convolutional Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1717–1724.
42. Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; March, M.; Lempitsky, V. Domain-Adversarial Training of Neural Networks. *J. Mach. Learn. Res.* **2016**, *17*, 1–35.
43. Filho, H.C.C.; Júnior, O.A.C.; de Carvalho, O.L.F.; de Bem, P.P.; de Moura, R.S.; de Albuquerque, A.O.; Silva, C.R.; Ferreira, P.H.G.; Guimarães, R.F.; Gomes, R.A.T. Rice Crop Detection Using Lstm, Bi-Lstm, and Machine Learning Models from Sentinel-1 Time Series. *Remote Sens.* **2020**, *12*, 2655. [CrossRef]
44. Chen, B.; Zheng, H.; Wang, L.; Hellwich, O.; Chen, C.; Yang, L.; Liu, T.; Luo, G.; Bao, A.; Chen, X. A Joint Learning Im-Bilstm Model for Incomplete Time-Series Sentinel-2a Data Imputation and Crop Classification. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *108*, 102762. [CrossRef]
45. Li, J.; Hu, M.; Wu, C. Multiscale Change Detection Network Based on Channel Attention and Fully Convolutional Bilstm for Medium-Resolution Remote Sensing Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 9735–9748. [CrossRef]
46. Schuster, M.; Paliwal, K.K. Bidirectional Recurrent Neural Networks. *IEEE Trans. Signal Process.* **1997**, *45*, 2673–2681. [CrossRef]
47. Graves, A.; Schmidhuber, J. Framewise Phoneme Classification with Bidirectional Lstm and Other Neural Network Architectures. *Neural Netw.* **2005**, *18*, 602–610. [CrossRef] [PubMed]
48. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
49. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
50. de Mathelin, A.; Atiq, M.; Richard, G.; de la Concha, A.; Yachouti, M.; Deheeger, F.; Mougeot, M.; Vayatis, N. Adapt: Awesome Domain Adaptation Python Toolbox. *arXiv* **2021**, arXiv:2107.03049.
51. Van der Maaten, L.; Hinton, G. Visualizing Data Using T-Sne. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
52. Ma, Y.; Yang, Z.; Huang, Q.; Zhang, Z. Improving the Transferability of Deep Learning Models for Crop Yield Prediction: A Partial Domain Adaptation Approach. *Remote Sens.* **2023**, *15*, 4562. [CrossRef]
53. Muruganantham, P.; Wibowo, S.; Grandhi, S.; Samrat, N.H.; Islam, N. A Systematic Literature Review on Crop Yield Prediction with Deep Learning and Remote Sensing. *Remote Sens.* **2022**, *14*, 1990. [CrossRef]
54. Keshava, N.; Mustard, J.F. Spectral Unmixing. *IEEE Signal Process. Mag.* **2002**, *19*, 44–57. [CrossRef]