Deep Transfer Learning For Crop Yield Prediction

Jiamian He*
Senjuti Bala*
s3790525@vuw.leidenuniv.nl
s4043499@vuw.leidenuniv.nl
LIACS, Leiden University
Leiden, Netherlands

Figure 1: Terra MODIS surface reflectance band 1-4-3 data from the MOD09A1 product over the western United States, 2018

Abstract

In 2015, the United Nations (UN) introduced the Sustainable Development Goals (SDGs) to address global challenges by 2030. SDG 2, "Zero Hunger," emphasizes the need to end hunger, achieve food security, and promote sustainable agriculture. Crop yield prediction plays a vital role in achieving these goals, offering insights into agricultural productivity. While previous studies focused on transfer learning for soybean yield prediction in geographically proximate regions like Argentina and Brazil, there is limited exploration of its broader applicability. This paper extends these efforts, evaluating transfer learning for diverse crops (e.g., corn) and countries (e.g., the United States), while incorporating additional environmental features to enhance model performance. Experiments conducted on SustainBench datasets aim to refine accuracy and reduce RMSE, addressing current gaps in crop yield prediction.

Experiments show that migration learning can quickly apply the Argentina soybean prediction model to US soybean prediction, reducing training time and cost, with accuracy close to that of the benchmark model. Transfer learning is also conducted from Argentina soybean to US corn data, while the R^2 data suggests the model fits quite well, there is still space to reduce the RMSE.In addition, although simply adding precipitation as additional environmental data does not improve model performance, different ways of adding environmental data deserve further experimentation. This experiment highlights the potential of migration learning to address the capabilities and gaps in current crop yield prediction and provide scalable solutions for sustainable agriculture.

Keywords

Deep Learning, Transfer Learning, Crop Yield Prediction, Satellite Image, SustainBench, SDG2

ACM Reference Format:
Jiamian He and Senjuti Bala. 2024. Deep Transfer Learning For Crop Yield Prediction. In . ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

*Both authors contributed equally to this research.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

In 2015, the United Nations proposed 17 SDGs to drive progress toward sustainability by 2030. Among them, SDG 2, "Zero Hunger," highlights the importance of sustainable agriculture and food security. Accurate crop yield prediction is essential for developing policies that can double the productivity of smallholder farms. However, traditional survey-based methods are costly and have sparse temporal and spatial coverage. Satellite imagery provides an affordable, scalable alternative for monitoring agricultural fields, offering rich, high-dimensional data.

Deep learning (DL) methods, particularly those leveraging transfer learning (TL), are well-suited for analyzing such data. Unlike traditional machine learning methods, DL enables automatic feature extraction and shows promise in adapting pre-trained models to related tasks through fine-tuning. The selected paper[6], "Deep Transfer Learning for Crop Yield Prediction with Remote Sensing Data," demonstrated TL's effectiveness in soybean yield prediction for Argentina and Brazil. However, its applicability across larger regions and for different crops remains unvalidated. Additionally, environmental variables such as soil moisture and precipitation, which are critical to yield prediction, were not considered.

This project aims to address these gaps by:

- **Evaluating transfer learning performance:** Extending soybean-based models to the United States and corn yield predictions.
- **Exploring environmental features:** Investigating how integrating additional variables enhances prediction accuracy.
- **Expanding scope:** Validating the methodology for diverse crops and geographies.

Thus, in this project, we mainly focus on evalutating the transfer learning ability, fine-tuning for tranfer learning, and explore on other features inputting to increase the model performance. There are 3 research questions we would like to experiment in this project:

- What if we apply the LSTM model to soybean yield predictions in America? How much extra training time and what fine-tuning skills would be required for this transfer learning process?
- If we extend the model to predict corn or wheat, would it still perform effectively? How about the time and tuning skills?
- The selected paper trained the CNN and LSTM models only using the histogram images. Will adding more environmental features improve performance? How to modify the model will fit this task?

To answer these questions, our approach is based on building an LSTM model trained on the SustainBench dataset and fine-tuning it for cross-region and cross-crop predictions, providing insights into SDG 2's actionable implementation.

2 Related work

Recent advancements in machine learning (ML) have enabled scalable solutions for agricultural monitoring, particularly using satellite imagery. SustainBench, introduced by Yeh et al. (2021), is a benchmark dataset for SDG-related tasks, including crop yield prediction. It simplifies high-dimensional satellite data into histograms, providing efficient inputs for ML models [7].

The selected paper, "Deep Transfer Learning for Crop Yield Prediction with Remote Sensing Data," demonstrated the effectiveness of transfer learning (TL) using LSTM models. It achieved robust results in soybean yield prediction for Brazil by transferring a model pre-trained on Argentina data. However, its scope was limited to soybeans, and no additional environmental features were considered, leaving opportunities for broader validation [6] [7]. Recent research has emphasized the benefits of multi-modal approaches in agricultural yield estimation. For instance, Lu et al. (2024) showed that incorporating environmental data such as precipitation and soil moisture significantly enhanced prediction accuracy [3]. Similarly, Wang et al. (2018) established that fine-tuning pre-trained models could yield effective results in cross-region yield prediction tasks [6]. These studies provide a foundation for exploring transfer learning across diverse crops and regions.

Our work builds upon these advancements by expanding the use of TL for new crops, such as corn, and integrating environmental features. These enhancements address the limitations in existing approaches, improving accuracy and generalizability in real-world agricultural applications.

3 Methods

Following the original paper, we train an LSTM model[2] on Argentina soybean data as the foundation model and use the transfer learning technique to quickly train new models for other crop types and countries. Once having the foundation model, we design 3 experiments to evaluate the model transfer learning performance, and also explore adding more environmental features to improve the prediction accuracy:

- Experiment on US soybean data to evaluate the transfer learning performance.
- Experiment on Argentina corn data to evaluate the transfer learning performance.
- Experiment with adding more environmental features to improve prediction accuracy.

All experiments are compared with the baselines of the original paper by the unified metrics. The possible weaknesses and strengths of these approaches will be evaluated with the experiment results.

3.1 Foundation Model: LSTM

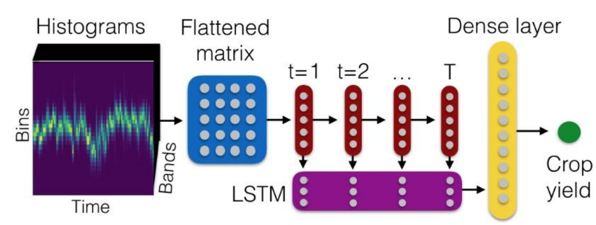

Figure 2: LSTM model for Deep Transfer Learning for Crop Yield Prediction

Long short-term memory (LSTM) is a type of recurrent neural network (RNN). With input, output and forget gates, it can efficiently learn from the sequential data. Shown in the figure2, the histograms are split into 32 bins ordered by time T (t1,t2...t32). Each T means each location sample measurement (every 8 days), in a total of 32 times per year. After the sequential learning in LSTM, the dense layer fully connects the hidden results and finally outputs the prediction.

3.2 Transfer Learning

Using pre-trained weights, we removed the final dense layer from the pre-trained model and replaced it with an untrained dense layer of the same dimensions. According to the different experiment settings, we also will change the Y true label for example from soybean to crop prediction.

3.3 Metrics

The formula for R^2 (coefficient of determination) and RMSE (Root Mean Squared Error) are:

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

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

Here, y_i represents the actual value, $\hat{y}_i$ is the predicted value, $\bar{y}$ is the mean of the actual values, and n is the total number of samples. The R^2 value ranges from [-1, 1], the closer to 1 the better model prediction is. Similarly, the RMSE is used to measure the overall mean error between predicted and actual values, with smaller values the better performance.

For evaluating the transfer learning ability, we also consider the training speed, such as the number of epochs the model starts to converge.

3.4 Novelty

Compared to previous works, which focused solely on transfer learning for a single crop (soybean), our study extends this approach by evaluating the transfer learning capabilities of the soybean-based model on other main crops like corn. Additionally, we explore whether adding more potentially relevant environmental features can improve the model's performance.

4 Data

The dataset we use is SustainBench, which is a set of sustainability benchmarks aligned with the UN SDGs, including No Poverty, Zero Hunger, Good Health, Quality Education, Clean Water, Climate Action, and Life on Land.

For the crop yield prediction problem, we use the data[4] under the SustainBench dataset–SDG2 Zero Hunger catalog. This dataset provides valuable preprocessed and dimensionality-reduced data for analyzing soybean yields from the original Satellite imageries (MOD09A1, MYD11A2 and MCD12Q1). It includes yield records for 857 counties in the United States, 135 counties in Argentina, and 32 counties in Brazil from years 2005 to 2016. In total, the dataset contains 9,049 data points for the United States, 1,615 for Argentina, and 384 for Brazil.

The input data consists of 9 histograms for each county during the harvest season (imageries were taken every 8 days for a total of 32 times), derived from MODIS satellite imagery. The input format is 32x32x9 pixels, which means that for each of 7 surface reflectance and 2 surface temperature bands, the MODIS pixel values were binned into 32 ranges and 32 timesteps per harvest season. The outputs are soybean yields in metric tonnes per harvested hectare over the counties.

5 Experimental setup

We setup the whole experiment in 4 steps, including first building a basic LSTM model and fine-tuning it based on Argentina soybean data, then exploring the transfer learning ability on US soybean and both corn prediction. Additionally, adding more features should also be considered to optimize the model.

For every experiment, we split the dataset into 60-20-20 as train-validate-test set. Every experiment runs once with 20 epochs. For prediction accuracy, we use RMSE and R^2 as the metrics and compare the overall baseline results in the original paper. For the transfer learning ability evaluation, we use the number of epochs to converge as the standard.

5.1 Baseline model and fine-tuning

First, we set the hyperparameters including batch 32, LSTM units 200, optimizer to 'adam', loss metrics to RMSE and R^2, learning rate to 0.005 to have a baseline model and result, with Argentina's soybean data. Then we use the grid search technique for fine-tuning, LSTM units [200,300], learning rate [0.1, 0.01, 0.005], dropout rate [0.5, 0.75] to have the optimized hyperparameters combination with the same data samples.

5.2 Transfer learning on US soybean

Next, we use the pre-trained LSTM hidden layer weights and replace the dense layer with a new untrained dense layer. Evaluate transfer learning capacity using accuracy metrics and epoch times in US soybean data samples.

5.3 Transfer learning on US corn

Because the original SustainBench dataset does not have the prediction values on corn, so we need to change the prediction (Y label) accordingly by ourselves. Then we use the same method and input as the second sub-experiment, and have the new corn prediction results. Transfer Learning for US Soybean and US Corn For both US soybean and US corn, we utilized pre-trained weights from the Argentina soybean model. The output dense layer was replaced with an untrained layer suitable for the respective prediction tasks. While the same pipeline was used for both experiments, US corn predictions required adjusting the labels to reflect corn yield data. The hyperparameters for the corn experiment are summarized in the table below:

Parameter	Value
Batch size	32
Dropout rate	0.75
Learning rate	0.01
LSTM units	200
Loss function	Mean Squared Error (MSE)
Optimizer	Adam

The AgML-CY-Bench dataset [5], specifically designed for crop yield prediction, provided data for the corn experiments. This dataset complements the SustainBench data by offering detailed crop-specific and environmental features.

5.4 Adding new feature - precipitation

Other environmental data such as elevation and slope, precipitation levels, and other values can be considered as important features and merged into the input for training. Using the same Argentina soybean data, the new model result is compared with the fine-tuned result in the first sub-experiment. Here we try to add the rainfall record.

Region1	Region2	Year	Soybean_label	Rainfall
chaco	9 de julio	2007	1.7	668.9
santa fe	constitucion	2014	3.2	502.3
entre rios	gualeguay	2009	1.43943	516.1
santa fe	san martin	2015	4.02035	502.3
santa fe	san cristobal	2011	2.47568	502.3

Table 1: Example of adding precipitation values in Argentina Soybean dataset

As in table 1, every data sample has "Region1" as province, and "Region2" as county in "Region1". According to the rain record [1] from the official website of Argentina, every "Region1" has its average precipitation value by months in last 30 years. Because from November to February is the main period in which rainfall can affect Argentina's soybean yields production, so we only sum up these 4 months precipitation value as the rainfall value.

After adding the rainfall values for every data samples, the rainfall values are transferred to (32,32,1) shape and concatenated with the (32,32,9) histograms, so the final input data shape should be (32, 32, 10). Thus, we also need to modify the original LSTM model with the new config B, W, C as 32,32,10.

6 Results

In general, we reproduce the baseline result similar to the paper[6] and optimize our hyper-parameter combination by fine-tuning. Based on this pre-trained model, we train the new dense layer and output layer with the LSTM weight of the pre-trained model by transfer learning method. The US soybean result shows the good tansfer capacity. However, adding the rainfall feature does not achieve the desired enhancement, which may be related to the way of adding precipitation value. All our experiment codes and data can be accessed in here: https://github.com/holeiden/UC_final_project_Group2/tree/main

6.1 Baseline model and fine-tuning result

The baseline LSTM model training is shown in the figure3 below. The baseline model learns very quickly within 3 epochs, both in the train and the eval split dataset. However, after around 15 epochs, the model finally converges and the RMSE becomes lowest.

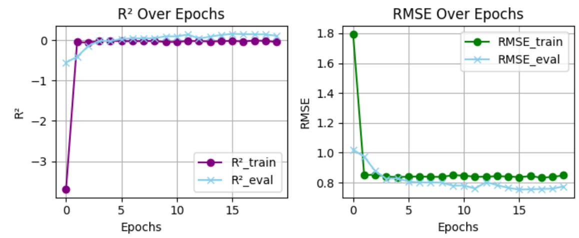

Figure 3: Training the LSTM baseline model

Since the original paper only provides separate test results for each year from 2012-2016, but our test data does not split by years, here we first average the 5-year results from 2012-2016 of the original paper, and take the average results of R^2 0.31 and RMSE 0.62 as the reference benchmark as in table2. From the results of the table, the baseline model result RMSE 0.78 is a bit higher than the paper result 0.62, and the R^2 of -0.18 indicates that the baseline model predicts poorly as the model is biased, and it does not capture the features in the data very well. So further modification and training should be done before using the model in next sub-experiment.

Model	Test Country	Year	R^2	RMSE
Paper	Argentina	2012	0.49	0.54
Paper	Argentina	2013	0.49	0.6
Paper	Argentina	2014	-0.13	0.73
Paper	Argentina	2015	0.57	0.54
Paper	Argentina	2016	0.12	0.7
Paper	Argentina	2012-2016	0.31	0.62
Baseline	Argentina	2005-2016	-0.18	0.85
Fine-tuning	Argentina	2005-2016	/	0.78

Table 2: Comparison between paper's baseline and our models

Model	Dropout	LR	Units	R^2	RMSE
baseline	0.75	0.005	200	-0.1801	0.8451
fine-tuning	0.5	0.1	200	/	0.7789
fine-tuning	0.5	0.1	300	/	0.7828
fine-tuning	0.5	0.01	200	/	0.7793
fine-tuning	0.5	0.01	300	/	0.7925
fine-tuning	0.5	0.005	200	/	0.7842
fine-tuning	0.5	0.005	300	/	0.7926
fine-tuning	0.75	0.1	200	/	0.8072
fine-tuning	0.75	0.1	300	/	0.8012
fine-tuning	0.75	0.01	200	/	0.7786
fine-tuning	0.75	0.01	300	/	0.7794
fine-tuning	0.75	0.005	200	/	0.8214
fine-tuning	0.75	0.005	300	/	0.7883

Table 3: Build baseline model and fine-tuning with Argentina soybean dataset

To have a better model, we use grid-search fine-tuning skill to have a better pre-trained model. All hyper-parameter combination results are shown in table3. The RMSE are quite similar but the dropout rate, learning rate, LSTM units (0.75, 0.01, 200) has the lowest RMSE, so we take this combination to have pre-trained model weights, and the later on sub-experiment will use these hyper-parameters and weights for transfer learning.

6.2 Transfer learning on US soybean result

Shown in the figure4, with the pre-trained lstm model and new us soybean dataset, the transferred model learns very fast, after 1 epoch it starts to converge. Although the R^2 is around 0, which means the model's prediction is not so good as the paper, but it is better than the baseline model. Also, the RMSE is 0.69, a bit lower than the fine-tuned model and the baseline model. As a result, both the R^2, RMSE loss metrics and the transfer learning speed (epoch times) are lower, it shows the model has a good transfer capacity, it might be scalable for other countries.

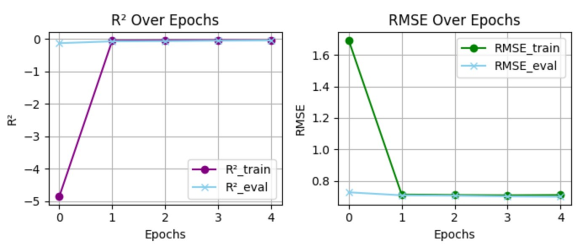

Figure 4: Transfer learning with US soybean dataset

The comparison4 of training epochs shows that the baseline model requires the most training epochs, followed by fine-tuning, while transfer learning requires significantly fewer epochs. This shows the efficiency of transfer learning.

Model	Test Country	Year	R^2	RMSE	Training Epochs
Paper	Argentina	2012-2016	0.31	0.62	/
Baseline	Argentina	2005-2016	-0.18	0.85	15
Fine-tuning	Argentina	2005-2016	/	0.78	5
Transfer	USA	2005-2016	-0.02	0.69	1

Table 4: Transfer learning result in USA soybean

6.3 Transfer Learning on US Corn Result

This experiment evaluates the transfer learning ability of the pre-trained LSTM model for predicting corn yields in the United States. Using the AgML-CY-Bench dataset [5], we adapted the model to predict corn yields by replacing the dense output layer with a new untrained layer. The experiment maintained consistency with the soybean experiments, employing the same preprocessing pipeline and evaluation metrics.

The model demonstrated rapid convergence within the 10 epochs. The final evaluation metrics for the test set are summarized in Table 5. The R^2 score indicates moderate predictive performance, while the RMSE suggests room for improvement.

The training and validation loss over epochs are shown in Figure 5. The R^2 and RMSE trends over epochs are presented in Figure 6. Additionally, the predicted vs. actual corn yield plot (Figure 7) illustrates the model's predictions against ground truth.

Metric	Value
R^2 (Test Set)	0.3096
RMSE (Test Set)	2.3378

Table 5: Performance metrics for transfer learning on US corn dataset

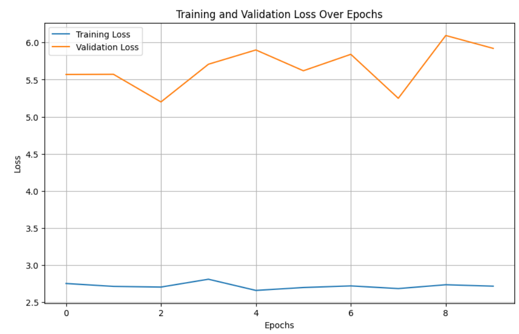

Figure 5: Training and Validation Loss Over Epochs for US Corn Dataset

This graph tells us that the consistent decline in training loss reflects on the model's capacity to fit the data. Also indicates potential over-fitting towards the later epochs.

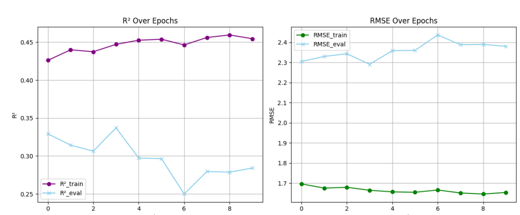

Figure 6: R^2 and RMSE Trends Over Epochs for US Corn Dataset

The R^2 scores for both training and evaluation shows a sign of improvement across epochs; on the other hand the RMSE trends show consistent reduction. This highlights the model's capacity to minimuze prediction errors over time.

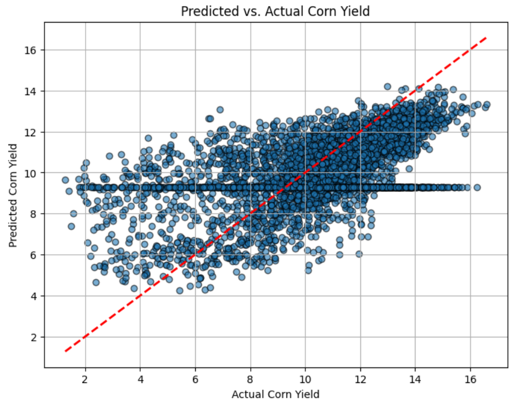

Figure 7: Predicted vs. Actual Corn Yield for US Corn Dataset.

The scatter plot shows the comparison between predicted corn yields to their actual values for the US dataset. The data points to the diagonal red line indicates a moderate agreement between predictions however it still needs a lot of work for better precision.

6.4 Adding precipitation feature on Argentina soybean result

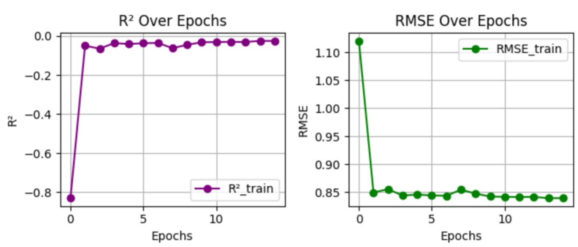

Figure 8: Training a new model after adding new precipitation feature

As shown in figure8, although the model also converges quickly, from the training with the train set, both R^2 and RMSE results are worse than the results from fine-tuned only without adding rainfall values. The additional feature does not work well because of some reasons:

- **Spatial problem**: Precipitation only added by the province as "Region 1" (a total of 9 provinces), not by county level as "Region 2", resulting in many different county data using the same rainfall value. It might cause the loss of local information and variation, leading to the over-smoothing problem and biasing the model training.
- **Temporal problem**: The precipitation values have no difference within different years or months, only using the same sum of November - February average rainfall values across 30 years. This will lead to the similar problem as the spatial one, lacking data variation will reduce the model performance.
- **Concatenate problem**: The same precipitation values are repeated from (1,,) to (32,32,1) in order to align with the histograms for concatenation. However, this repetition might lead to the information redundancy, adding lots of noise into the model without extra useful information.
- **Feature engineering problem**: Other environmental factors, such as elevation and slope, have not yet been experimented with. Identifying which features contribute most to prediction performance still needs to be done to address through further feature engineering, before adding the new feature into the model.

7 Conclusion

In this study, we replicate the LSTM model from the selected paper using the SustainBench dataset, which is processed from satellite images. By applying fine-tuning technique, we improve the model's performance similar to the benchmark. Then, we explore the transfer learning technique by adapting the Argentina soybean prediction model to US soybean. Next, transfer learning is applied from Argentina soybean to US corn data. Additionally, we try incorporating new environmental variables, such as precipitation, aiming to increase prediction accuracy.

Our experiment demonstrates that transfer learning effectively adapts the Argentina soybean prediction model to US soybean prediction, significantly reducing training time and cost while achieving accuracy similar to the benchmark model. When applying TL from Argentina soybean to US corn, while R2 indicates a good fit to the model, RMSE still leaves room for improvement. Although adding precipitation as an additional environmental variable does not improve model performance, other methods deserve further investigation. These findings show that transfer learning can address limitations in current crop yield prediction models and offer scalable, efficient solutions for sustainable agriculture.

8 Acknowledgments

We would like to express our gratitude to our advisors (the Lecturer and the Teaching Assistants) and peers for their guidance and support throughout this project. Their insights and feedback were invaluable in refining our model and analysis.

References

[1] Argentina official protal 2024. . https://www.smn.gob.ar/estadisticas
[2] Codes of selected paper 2018. . https://github.com/sustainlab-group/Deep-Transfer-Learning-Crop-Yield-Prediction
[3] Jian Lu et al. 2024. GOA-optimized deep learning for soybean yield estimation using multi-source remote sensing data. *Scientific Reports* (2024).
[4] SustainBench SDG2 Crop Yield Dataset 2018. . https://sustainlab-group.github.io/sustainbench/docs/datasets/sdg2/crop_yield.html
[5] AgML Team. 2024. AgML-CY-Bench: Crop Yield Benchmark. https://github.com/WUR-AI/AgML-CY-Bench
[6] Anna X. Wang et al. 2018. Deep Transfer Learning for Crop Yield Prediction with Remote Sensing Data. *Association for Computing Machinery* (2018).
[7] Christopher Yeh, Chenlin Meng, Sherrie Wang, et al. 2021. SUSTAINBENCH: Benchmarks for Monitoring the Sustainable Development Goals with Machine Learning. *arXiv preprint arXiv:2111.04724* (2021).