Enhancing Agricultural Productivity Monitoring and Forecasting through Deep Transfer Learning

Robin Young
robin@rayoung.org

Rachel Fan

Abstract

Agriculture is increasingly threatened by climate change impacts, especially in lower-income nations where traditional paper survey-based data collection for detailed agricultural reporting can be impractical. Leveraging advancements in remote sensing and machine learning, we propose a Convolutional Neural Network-Long Short-Term Memory model designed to infer farm-level crop yields from one year of high-resolution weather data. We also propose assessing the potential of predicting based on forecasted weather data and implementing transfer learning to predict crop yield in regions outside the US with sparse data collection. If successful, this approach would offer a detailed and adaptive tool for countries with all levels of monitoring resources.

1 Introduction

Agriculture is vital to sustaining the planet's growing population and the global economy. However, the increasing frequency and intensity of extreme weather events, changing precipitation patterns, and rising temperatures all present unprecedented conditions for farmers and pressure agricultural systems and food supply. Climate change challenges are more pronounced in lower-income countries, which often lack resources for active monitoring and risk management of crop yields. Recent advances in technology, specifically remote sensing and machine learning, offer new ways to address these challenges. This paper proposes training a deep learning model to infer farm-level crop yield based on high resolution weather data, which captures the effects of condition variation throughout the year preceding harvest. Furthermore, we discuss the possibility of leveraging US investments in agricultural statistical reporting by training models on USDA data and using transfer learning to predict yield in other countries.

2 Literature Review

The increasing application of machine learning and remote sensing in agricultural monitoring is well documented [1, 2, 3, 4, 5, 6]. Machine learning algorithms have shown promise in mapping soil organic carbon content [1] and identifying crop types [3, 5]. They leverage large datasets, offering potential for detailed and high-resolution agricultural monitoring. In some scenarios, machine learning combined with remote sensing has been able to replace on-site inspections entirely [7]. These technologies have found applications in complex tasks such as predicting crop yields [8, 9, 10, 11] or predicting the impacts of climate change on crops [12, 13, 14, 15]. Neural networks have been used to predict the impact of climate change on crop yield [16, 13]. They can also help in detecting diseased crop trees, thus preventing potential losses [17].

However, existing agricultural monitoring systems do not yet fully leverage these recent developments in remote sensing and machine learning. For instance, the USDA's National Agricultural Statistics Service (NASS) [18] still relies on mailed questionnaire surveys as the main method of data collection, with annual surveys as well as a Census of Agriculture once every five years. The integration of

Preprint. Under review.

machine learning with remote sensing into this process offers a possibility for cheaper and more frequent census-level estimations by inferring agricultural data from other regularly collected data that impact agriculture outcomes.

3 Methodology

We propose three tasks: (1) evaluate the potential of deep learning on observed weather data to infer farm-level crop yield to complement current USDA NASS surveys, (2) assess whether replacing observed weather data with long-range weather forecasts in test data can produce predictions with comparable accuracy, and (3) assess the possibility of transfer learning with the NASS-trained model for inference in other regions.

Weather data for training will come from the PRISM Climate Group [19], and yield data will come from USDA NASS Census of Agriculture and annual surveys [18]. PRISM data will provide daily observations of precipitation, temperature, dew point temperature, vapor pressure, solar radiation, and cloudiness at 4 km resolution starting in 1981 for the conterminous 48 states. PRISM uses information from both remote sensing and ground stations. USDA NASS Census and survey microdata will provide yield data for corn, wheat, and soybean at the farm level at five year intervals for census and yearly for survey, which will be used as the ground truth labels.

Agriculture literature demonstrates weather variation throughout the year has effect on yield [20, 21, 22], and high resolution weather data from PRISM allows weather conditions to be mapped to specific farms. USDA NASS records the address and cropland area for each farm, which can be used together to approximate the area to which to map the weather data. In contrast, typical county level calculations may lose details through aggregation that could be leveraged by a deep learning model to produce more accurate yield estimates by accounting for differences between farms within counties.

Data from these sources will be pre-processed to ensure consistent units and to associate the appropriate weather and yield data for each farm in the conterminous 48 states that produces corn, wheat, or soybean found in NASS records since 1981. Thus, each element of the prepared dataset will include information for one farm over one year for one crop. Then, data from the most recent five years will be reserved for testing, and all earlier years will be used for training.

Given the nature of crop yield predictions, both spatial and temporal patterns can significantly influence the outcome [23]. Thus, we propose a combined spatio-temporal splitting strategy. This method will divide the dataset temporally, ensuring that the training set only consists of data from years preceding the validation/test years. This respects the forward-looking nature of predictions to predict future yields based on past and present data. Within each temporal split, we can further divide the data spatially. Since the USDA ERS defines farm resource regions (FRR) [24] based on similar regional agricultural patterns, it may be beneficial to group data by FRRs to capture effects that are homogeneous within but diverse across regions. Comparing model performance with and without including FRR information for each farm can help evaluate the relevance of FRR for this task.

For the first task, we propose training a Convolutional Neural Network-Long Short-Term Memory model (CNN-LSTM), which is well-suited for scenarios involving sequences of spatial and time-varying data, such as weather observations and forecasts, by combining the spatial feature extraction capabilities of CNNs with the temporal sequence processing strengths of LSTMs [25, 26]. The CNN will be applied to each time step's data to extract spatial features from gridded data, resulting in a sequence of feature vectors. Then, this sequence will be passed through an LSTM network to capture the temporal variation present in the data throughout the year. The LSTM's final output will be connected to a dense layer to predict the annual yield.

The model's performance can be assessed through comparing predicted yield against the reported ground truth yield by calculating the root mean square error (RMSE), mean average error (MAE), and R-squared. These values can be calculated for each crop type and for each farm region (if included) to identify whether the performance differs significantly across these categories. Also, as a simple benchmark, those metrics can also be calculated for multiple linear regression, gradient boosting, or random forest models that uses the same test and training data. Additionally, the importance of using farm-level rather than county-level data can be assessed by training another CNN-LSTM model on county-wide weather averages and yields to predict county-level yield totals and then comparing the results against adding up predicted farm-level yields within each county.

For the second task, we will assess the difference in yield prediction performance between testing the model on empirical weather observations from PRISM and testing on historical long-range weather forecast data from ECMWF [27]. If the differences are reasonably small, then the model can be used to predict yields in a year where the harvest has not occurred by using observed weather up to the point of prediction in conjunction with long-range forecasts to stand in for the rest of that year.

For the third task, the model will be evaluated for the potential of transfer learning. Transfer learning is a machine learning technique where a model developed for one task is repurposed for another related task. Transfer learning can be powerful in compensating for gaps in data. Instead of training a model from scratch, which requires substantial data, transfer learning leverages existing knowledge from the USDA NASS dataset, to enhance predictions in new environments. Thus, by using the model pre-trained on USDA NASS microdata, it may be possible to either infer on agricultural yields in other regions directly or use additional training data to calibrate further where data is available.

Foreseeable challenges include the difficulty of obtaining special permission from USDA for micro-data access. Furthermore, weather forecasts may also prove to be too different from observed weather data for the model to predict well. Additionally, the model may be too specific to the USDA data, leading to overfitting and poor suitability for transfer learning. Differences in agricultural practices, soil types, climatic conditions, and other regional factors between the USDA data region and target regions may limit the generalizability of the model.

4 Pathway to Climate Impact

The proposed method holds significant promise for improving agricultural monitoring practices. Building a successful model as described in the first task above would decrease the barrier of entry for data collection for agricultural monitoring as weather data is already commonly collected and forecasted around the world. Also, providing detailed crop yield estimations at a lower investment than a nationwide paper census would enable more effective risk mitigation and adaptation strategies, strengthening resilience against climate change impacts. If transfer learning is feasible, as listed in the third task, this model would enable lesser-developed countries, which frequently lack the necessary resources for thorough agricultural monitoring [28, 29], to improve monitoring by leveraging the investments made in agricultural statistics reporting in the US.

A released pre-trained model enables predictions that require only publicly available remote sensing data, which reduces dependency on comprehensive physical surveys which may be prohibitive in areas with low infrastructure investments. Access to comprehensive and reliable predicted crop yield data empowers a diverse set of stakeholders, including policymakers, global organizations, and farmers. Policymakers and global organizations can leverage the improved data to design more targeted agricultural policies and interventions. Farmers can utilize the data to make more strategic decisions about planting and harvesting times and other farm management practices. As described in task two above, ideally, a farmer could predict crop yield early in the season by inputting forecasted local weather conditions for the year before scheduled harvest and adjust plans accordingly. Overall, enhanced data inference facilitated by this method will enable stakeholders to make better informed decisions, contributing to more effective climate risk mitigation and adaptation strategies.

References

[1] Ali Sakhaee, Anika Gebauer, Mareike Ließ, and Axel Don. Spatial prediction of organic carbon in German agricultural topsoil using machine learning algorithms. *SOIL*, 8(2):587–604, September 2022. URL: https://soil.copernicus.org/articles/8/587/2022/, doi:10.5194/soil-8-587-2022.

[2] David B. Lobell. The use of satellite data for crop yield gap analysis. *Field Crops Research*, 143:56–64, March 2013. URL: https://linkinghub.elsevier.com/retrieve/pii/S0378429012002754, doi:10.1016/j.fcr.2012.08.008.

[3] José M. Espinosa-Herrera, Antonia Macedo-Cruz, Demetrio S. Fernández-Reynoso, Héctor Flores-Magdaleno, Yolanda M. Fernández-Ordoñez, and Jesús Soria-Ruíz. Monitoring and Identification of Agricultural Crops through Multitemporal Analysis of Optical Images and Machine Learning Algorithms. *Sensors*, 22(16):6106, August 2022. URL: https://www.mdpi.com/1424-8220/22/16/6106, doi:10.3390/s22166106.

[4] Artur Gafurov, Svetlana Mukharamova, Anatoly Saveliev, and Oleg Yermolaev. Advancing Agricultural Crop Recognition: The Application of LSTM Networks and Spatial Generalization in Satellite Data Analysis. *Agriculture*, 13(9):1672, August 2023. URL: https://www.mdpi.com/2077-0472/13/9/1672, doi:10.3390/agriculture13091672.

[5] Meng Zhang, Hui Lin, Guangxing Wang, Hua Sun, and Jing Fu. Mapping Paddy Rice Using a Convolutional Neural Network (CNN) with Landsat 8 Datasets in the Dongting Lake Area, China. *Remote Sensing*, 10(11):1840, November 2018. URL: http://www.mdpi.com/2072-4292/10/11/1840, doi:10.3390/rs10111840.

[6] M. Weiss, F. Jacob, and G. Duveiller. Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment*, 236:111402, January 2020. URL: https://linkinghub.elsevier.com/retrieve/pii/S0034425719304213, doi:10.1016/j.rse.2019.111402.

[7] Francisco Javier López-Andreu, Manuel Erena, Jose Antonio Dominguez-Gómez, and Juan Antonio López-Morales. Sentinel-2 Images and Machine Learning as Tool for Monitoring of the Common Agricultural Policy: Calasparra Rice as a Case Study. *Agronomy*, 11(4):621, March 2021. URL: https://www.mdpi.com/2073-4395/11/4/621, doi:10.3390/agronomy11040621.

[8] Daniel Marc G Dela Torre, Jay Gao, and Cate Macinnis-Ng. Remote sensing-based estimation of rice yields using various models: A critical review. *Geo-spatial Information Science*, 24(4):580–603, October 2021. URL: https://www.tandfonline.com/doi/full/10.1080/10095020.2021.1936656, doi:10.1080/10095020.2021.1936656.

[9] Kersten Clauss, Marco Ottinger, Patrick Leinenkugel, and Claudia Kuenzer. Estimating rice production in the Mekong Delta, Vietnam, utilizing time series of Sentinel-1 SAR data. *International Journal of Applied Earth Observation and Geoinformation*, 73:574–585, December 2018. URL: https://linkinghub.elsevier.com/retrieve/pii/S0303243418303064, doi:10.1016/j.jag.2018.07.022.

[10] Sandeep Kumar Mathivanan and Prabhu Jayagopal. Simulating crop yield estimation and prediction through geospatial data for specific regional analysis. *Earth Science Informatics*, 16(1):1005–1023, March 2023. URL: https://link.springer.com/10.1007/s12145-022-00887-4, doi:10.1007/s12145-022-00887-4.

[11] Yanghui Kang and Mutlu Özdoğan. Field-level crop yield mapping with Landsat using a hierarchical data assimilation approach. *Remote Sensing of Environment*, 228:144–163, July 2019. URL: https://linkinghub.elsevier.com/retrieve/pii/S0034425719301427, doi:10.1016/j.rse.2019.04.005.

[12] P Jones and P Thornton. The potential impacts of climate change on maize production in Africa and Latin America in 2055. *Global Environmental Change*, 13(1):51–59, April 2003. URL: https://linkinghub.elsevier.com/retrieve/pii/S0959378002000900, doi:10.1016/S0959-3780(02)00090-0.

[13] Andrew Crane-Droesch. Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environmental Research Letters*, 13(11):114003, October 2018. URL: https://iopscience.iop.org/article/10.1088/1748-9326/aae159, doi:10.1088/1748-9326/aae159.

[14] Naveen Kumar Arora. Impact of climate change on agriculture production and its sustainable solutions. *Environmental Sustainability*, 2(2):95–96, June 2019. URL: `http://link.springer.com/10.1007/s42398-019-00078-w`, `doi:10.1007/s42398-019-00078-w`.

[15] Saran Aadhar and Vimal Mishra. High-resolution near real-time drought monitoring in South Asia. *Scientific Data*, 4(1):170145, October 2017. URL: `https://www.nature.com/articles/sdata2017145`, `doi:10.1038/sdata.2017.145`.

[16] Michael P. Keane and Timothy Neal. Comparing Deep Neural Network and Econometric Approaches to Predicting the Impact of Climate Change on Agricultural Yield, January 2020. URL: `https://papers.ssrn.com/abstract=3521260`, `doi:10.2139/ssrn.3521260`.

[17] Panu Srestasathiern and Preesan Rakwatin. Oil Palm Tree Detection with High Resolution Multi-Spectral Satellite Imagery. *Remote Sensing*, 6(10):9749–9774, October 2014. URL: `http://www.mdpi.com/2072-4292/6/10/9749`, `doi:10.3390/rs6109749`.

[18] USDA National Agricultural Statistics Service. USDA National Agricultural Statistics Service NASS - Quick Stats, 2017. URL: `https://data.nal.usda.gov/dataset/nass-quick-stats`.

[19] Oregon State University. PRISM Climate Group. URL: `https://prism.oregonstate.edu`.

[20] J. Schuerings, A. Jentsch, V. Hammerl, K. Lenz, H. A. L. Henry, A. V. Malyshev, and J. Kreyling. Increased winter soil temperature variability enhances nitrogen cycling and soil biotic activity in temperate heathland and grassland mesocosms. *Biogeosciences*, 11(23):7051–7060, Dec 2014. `doi:10.5194/bg-11-7051-2014`.

[21] Juergen Kreyling. Winter climate change: a critical factor for temperate vegetation performance. *Ecology*, 91(7):1939–1948, 2010. `doi:10.1890/09-1160.1`.

[22] Hugh A. L. Henry. *Soil Freezing Dynamics in a Changing Climate: Implications for Agriculture*, page 17–27. Springer New York, New York, NY, 2013. `doi:10.1007/978-1-4614-8253-6_2`.

[23] Mariana Oliveira, Luís Torgo, and Vítor Santos Costa. Evaluation Procedures for Forecasting with Spatiotemporal Data. *Mathematics*, 9(6):691, March 2021. URL: `https://www.mdpi.com/2227-7390/9/6/691`, `doi:10.3390/math9060691`.

[24] Ralph E. Heimlich. Farm Resource Regions. Technical report, 2000. URL: `https://ageconsearch.umn.edu/record/33625`, `doi:10.22004/AG.ECON.33625`.

[25] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, and Wang-chun Woo. Convolutional LSTM Network: a machine learning approach for precipitation nowcasting. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'15, pages 802–810, Cambridge, MA, USA, December 2015. MIT Press.

[26] Tara N. Sainath, Oriol Vinyals, Andrew Senior, and Hasim Sak. Convolutional, Long Short-Term Memory, fully connected Deep Neural Networks. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4580–4584, South Brisbane, Queensland, Australia, April 2015. IEEE. URL: `http://ieeexplore.ieee.org/document/7178838/`, `doi:10.1109/ICASSP.2015.7178838`.

[27] European Centre for Medium-Range Weather Forecasts (ECMWF). ECMWF. `https://www.ecmwf.int/`, 2023. Accessed: September 30, 2023.

[28] Clement Atzberger. Advances in Remote Sensing of Agriculture: Context Description, Existing Operational Monitoring Systems and Major Information Needs. *Remote Sensing*, 5(2):949–981, February 2013. URL: `http://www.mdpi.com/2072-4292/5/2/949`, `doi:10.3390/rs5020949`.

[29] United Nations. The state of food and agriculture 2016. Technical report, 2016. URL: `https://www.fao.org/3/i6030e/i6030e.pdf`.