

# Artificial Intelligence (AI) For Crop Yield Forecasting<sup>1</sup>

Clyde Fraise, Yiannis Ampatzidis, Sandra Guzmán, Wonsuk Lee, Christopher Martinez, Sanjay Shukla, Aditya Singh, and Ziwen Yu<sup>2</sup>

## Introduction

This publication aims to introduce readers to recent crop yield forecasting approaches based on artificial intelligence (AI) and to provide examples of how AI can potentially improve yield forecasting at the field and regional levels. It is intended primarily for Extension agents, crop consultants, and members of the public interested in applications of AI to agriculture.

The ability to promptly and reliably forecast crop yield is an important aspect of regional and global food security. Forecasting year-to-year variations in the yields of major crops at the regional and national levels can strengthen the ability of societies to respond to food production shocks and food price spikes triggered by extreme events. When wheat prices spiked in 2010–2011 due to droughts in Russia, Ukraine, China, and Argentina, increased bread prices combined with high unemployment caused the governments in several Arab countries to fall in what is known as the “Arab Spring.” Although an improved and more timely wheat crop forecast would probably not have ensured a different outcome, it could certainly help governments improve their planning and mitigate the potential consequences of drought. Within-season crop forecasting is also important for farmers to make more informed crop management and financial decisions (Hansen et al. 2004; Hansen and Indeje 2004; Newlands et al. 2014).

Many approaches have been used to forecast yield at the regional and field levels (Figure 1). Field surveys, mathematical models that simulate crop development and yield, statistical models, remote sensing, and combinations of surveys, models, and remote sensing have been used to forecast yield at the field and regional scales (Cai et al. 2019; Lizumi et al. 2018; Guan et al. 2017; Newlands et al. 2014; de Wit and van Diepen 2007; Jagtap and Jones 2002). However, crop yield forecasting is challenging due to the many factors involved, such as crop and variety, soil type, management practices, pests and diseases, and climate and weather patterns during the season. A crop’s response to these factors and their interactions is highly nonlinear and frequently difficult to understand. More recently, approaches based on AI algorithms have been gaining popularity, and encouraging results on the use of

*machine learning* for crop yield forecasting have been reported.

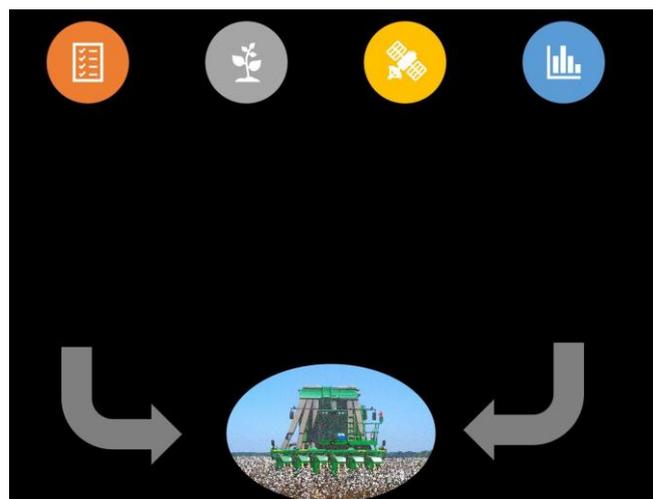


Figure 1. Approaches used for crop forecasting include field surveys, mathematical models that simulate crop development and yield, statistical models, remote sensing and, more recently, machine learning. Combinations of two or more approaches have also been used for crop yield forecasting.

Credit: UF/IFAS

## What is machine learning?

Machine learning (ML) is an application of AI that gives computers the ability to learn without being explicitly programmed. It is comprised of various mathematical algorithms that make learning possible. A machine learning algorithm is a process that is used to fit a mathematical model to a dataset through training or learning. The learned model is subsequently used against an independent dataset to determine how well it performs against the unused data in a process called testing (Witten et al. 2016). Mathematical models based on machine learning generally improve when more training data are made available. We will use the diagram in Figure 2 to describe a hypothetical example of a project utilizing ML to forecast crop yield based on climatic patterns and remote sensing-based vegetation indices.

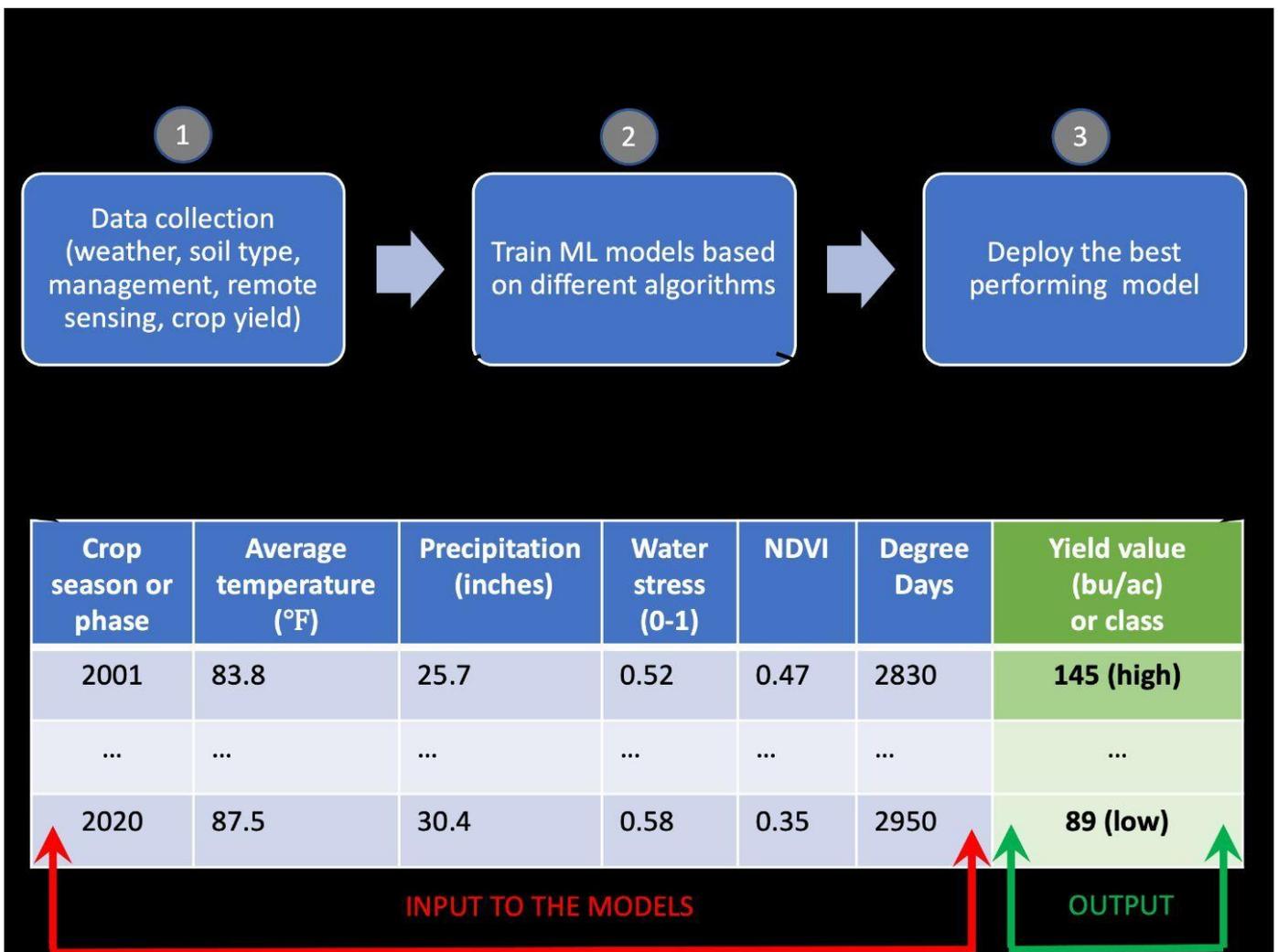


Figure 2. Machine learning projects based on supervised learning are generally composed of three phases or activities: (1) Data collection; (2) Training of the ML models; (3) Deployment of the best performing model. Credit: Clyde Fraisse, UF/IFAS

The first step of a ML project is *data collection*. In the case of yield forecasting based on climatic variables and vegetation indices, we would need to have climate records and regional or field imagery for the locations and crop whose yield we want to forecast. In this case, we would use a “supervised-learning” process as we have a good idea of what climate variables are important for crop yield. As shown in Figure 2, the climate and remote sensing indicators used as input to the model include precipitation amount, air temperature, levels of water stress, accumulation of growing degree-days, and NDVI (Normalized Difference Vegetation Index). Vegetation indices such as the NDVI are calculated based on imagery collected by sensors that can be mounted in different platforms (e.g., satellite, airborne, drones). NDVI values estimated during the cropping season help in the yield forecast process because they provide information about plant health, canopy coverage, and/or water status of canopies.

In order to “train” the model, we also need to have crop yield observations during several years (as many as

possible) and hopefully under different climatic conditions to provide examples of “good years” and “bad years” in the dataset. With the input variables (climatic variables and vegetation indices) and the observed output (crop yield observations), we can now start the process of *training the model* using ML algorithms. There are several types of ML algorithms; normally, more than one mathematical model is produced based on the different algorithms used. To *test* the models and determine which one is the best performing model, part of the dataset is left unused and subsequently used to determine how well the learned model performs against the unused data. Once the best performing model is selected, it can be *deployed* to provide operational crop forecast outputs on a regular basis.

Model outputs can be a *numeric* output (e.g., bushels/acre) or a *category* such as low, average, or high yield. As in any situation in which mathematical models are used to extract information from data, the results will be as good as the data used to generate them.

## Examples of Machine Learning Applications for Crop Yield Forecasting

Several recent studies have reported encouraging results obtained by using machine learning algorithms to predict crop yield. The variables that were most used to predict yield include air temperature, precipitation, and soil type. The table below summarizes a few recent examples of ML used to predict yields for different crops.

## Conclusions

The importance of crop yield forecasting for food security is expected to increase as the global population grows and extreme weather events such as droughts, heat waves, and large storms become more frequent or more intense due to climate change. Several approaches have been used to predict crop yield such as field surveys, crop models, remote sensing, traditional statistical models, and more recently, machine learning algorithms. Machine learning approaches by themselves or combined with other approaches such as crop modeling and remote sensing have shown promising results and can improve the accuracy of crop yield prediction at the field and regional levels. This publication aimed to introduce the topic of machine learning applied to crop yield forecasting and to describe a few examples of applications. The application of artificial intelligence in agriculture is very promising and expected to increase in the next few years. We can expect improved crop yield forecasting based on artificial intelligence approaches as more data become available and the technology evolves.

## References

- Cai, Y., K. Guan, D. Lobell, A. B. Potgieter, S. Wang, J. Peng, T. Xu, S. Asseng, Y. Zhang, L. You, and B. Peng. 2019. "Integrating Satellite and Climate Data to Predict Wheat Yield in Australia Using Machine Learning Approaches." *Agric. and Forest Met.* 274:144–159. <https://doi.org/10.1016/j.agrformet.2019.03.010>
- Chen, Y., W. S. Lee, H. Gan, N. Peres, C. Fraisse, Y. Zhang, and Y. He. 2019. "Strawberry Yield Prediction Based on a Deep Neural Network Using High-Resolution Aerial Orthoimages." *Remote Sensing* 11:1584. doi:10.3390/rs11131584.
- Costa, L., J. McBreen, Y. Ampatzidis, J. Guo, M. Reisi Gahrooei, and A. Babar. 2021. "Using UAV-Based Hyperspectral Imaging and Functional Regression to Assist in Predicting Grain Yield and Related Traits in Wheat under Heat-Related Stress Environments for the Purpose of Stable Yielding Genotypes." *Precision Agriculture*. <https://doi.org/10.1007/s11119-021-09864-1>
- de Wit, A. J. W., and C. A. van Diepen. 2007. "Crop Model Data Assimilation with the Ensemble Kalman Filter for Improving Regional Crop Yield Forecasts." *Agric. and Forest Met.* 146(1-2): 38–56. <https://doi.org/10.1016/j.agrformet.2007.05.004>
- Guan, K., J. Wu, J. S. Kimball, M. C. Anderson, S. Frolking, B. Li, C. R. Hain, and D. B. Lobell. 2017. "The Shared and Unique Values of Optical, Fluorescence, Thermal and Microwave Satellite Data for Estimating Large-Scale Crop Yields." *Remote Sens. Environ.* 199:333–349. <https://doi.org/10.1016/j.rse.2017.06.043>
- Hansen, J. W., and M. Indeje. 2004. "Linking Dynamic Seasonal Climate Forecasts with Crop Simulation for Maize Yield Prediction in Semi-Arid Kenya." *Agricultural and Forest Meteorology* 125:143–157. <https://doi.org/10.1016/j.agrformet.2004.02.006>
- Hansen, J. W., A. Potgieter, and M. Tippet. 2004. "Using a General Circulation Model to Forecast Regional Wheat Yields in Northeast Australia." *Agric. For. Meteorol.* 127:77–92. <http://dx.doi.org/10.1016/j.agrformet.2004.07.005>
- Jagtap, S., and J. W. Jones. 2002. "Adaptation and Evaluation of the CROPGRO-Soybean Model to Predict Regional Yield and Production." *Agric., Ecosystems & Env.* 93(1-3): 73–85. [https://doi.org/10.1016/S0167-8809\(01\)00358-9](https://doi.org/10.1016/S0167-8809(01)00358-9)
- Khaki, S. and Wang, L. 2019. "Crop yield prediction using deep neural networks." *Frontiers in Plant Science*. doi: 10.3389/fpls.2019.00621.
- Lee, W. S., F. Wu, A. Abd-Elrahman, N. Peres, and S. Agehara. 2020. "Strawberry Yield Prediction Models Based on Imagery Information." *FSREF Research Report 2019–20*.
- Lizumi, T., Y. Shin, W. Kima, M. Kim, and J. Choi. 2018. "Global Crop Yield Forecasting Using Seasonal Climate Information from a Multi-Model Ensemble." *Climate Services* 11:13–23.
- Newlands, N. K., D. S. Zamar, L. A. Kouadio, Y. Zhang, A. Chipanshi, A. Potgieter, S. Toure, and H. S. J. Hill. 2014. "An Integrated, Probabilistic Model for Improved Seasonal Forecasting of Agricultural Crop Yield under Environmental Uncertainty." *Front. Environ. Sci.* 2:17. <https://doi.org/10.3389/fenvs.2014.00017>
- Shahhosseini, M., H. Guiping, I. Huber, and S. V. Archontoulis. 2021. "Coupling Machine Learning

and Crop Modeling Improves Crop Yield Prediction in the US Corn Belt.” *Scientific Reports* 11:1606. doi:[10.1038/s41598-020-80820-1](https://doi.org/10.1038/s41598-020-80820-1).

van Klompenburg, T., Kassahun, A., and Catal, C. 2020. “Crop yield prediction using machine learning: A systematic review.” *Computers and Electronics in Agriculture* 177. doi: [10.1016/j.compag.2020.105709](https://doi.org/10.1016/j.compag.2020.105709)

Vijayakumar, V., L. Costa, and Y. Ampatzidis. 2021. “Prediction of Citrus Yield with AI Using Ground-Based Fruit Detection and UAV Imagery.” *ASABE Paper No. 2100493*. doi:[10.13031/aim.202100493](https://doi.org/10.13031/aim.202100493).

Witten, I. H., E. Frank, M. A. Hall, and C. J. Pal. 2016. *Data Mining: Practical Machine Learning Tools and Techniques*. 4<sup>th</sup> Edition. San Francisco, CA: Morgan Kaufmann.

## Tables

Crop (Region of the Study)	Objective	Input Variables	ML Algorithms Evaluated	Reference
Generic	Systematic literature review of algorithms used for crop yield forecasting	Most used features are temperature, rainfall, and soil type.	Most applied algorithm is Artificial Neural Networks	van Klompenburg et al. (2020)
Corn (US Corn Belt: Indiana, Illinois, Iowa)	Countywide corn yield levels	Outputs from crop model (APSIM) including biomass, and simulated yield combined with weather, management, and soil type	LASSO, linear regression, random forest, gradient boosting regressions	Shahhosseini et al. (2021)
Corn (US, several locations where the performance of corn hybrids were tested)	Understand the relationship between yield and genotype and environmental factors	Genes and environmental components	Deep neural network, Lasso, shallow neural networks, regression tree	Khaki and Wang (2019)
Wheat (Australia)	Statistical division yield levels	Climate and satellite-derived vegetation indices	Support vector machine, random forest, and neural network	Cai et al. (2019)
Wheat (Florida)	Field plots of breeding programs	Hyperspectral imaging from UAVs	Functional regression	Costa et al. (2021)
Citrus (Florida)	Yield at the tree level	Combination of UAV multispectral imaging and ground-collected color imagery	Gradient boosting and random forest regressions, linear regression, and partial least squares regression	Vijayakumar et al. (2021)
Strawberry (Florida)	Yield at the plot level	UAV imaging and ground-collected color imagery	Region-based convolutional neural networks	Chen et al. (2019); Lee et al. (2020)

<sup>1</sup> This document is AE571, one of a series of the Department of Agricultural and Biological Engineering, UF/IFAS Extension. Original publication date April 2022. Revised October 2025. Visit the EDIS website at <https://edis.ifas.ufl.edu> for the currently supported version of this publication.

<sup>2</sup> Clyde Fraisse, professor, agrometeorology, Department of Agricultural and Biological Engineering; Yiannis Ampatzidis, professor, precision agriculture/automation, smart machines, UF/IFAS Southwest Florida Research and Education Center, Immokalee, FL; Sandra M. Guzmán, assistant professor, water resources, Department of Agricultural and Biological Engineering, UF/IFAS Indian River Research and Education Center; Won Suk Lee, professor, machine vision and precision agriculture, UF/IFAS Department of Agricultural and Biological Engineering, Gainesville, FL; Christopher J. Martinez, former associate professor, UF/IFAS Department of Agricultural and Biological Engineering, Gainesville, FL; Sanjay Shukla, professor, water quality, Department of Agricultural and Biological Engineering, UF/IFAS Southwest Florida Research and Education Center, Immokalee, FL; Aditya Singh, assistant professor, remote sensing and spectroscopy, UF/IFAS Department of Agricultural and Biological Engineering, Gainesville, FL; Ziwen Yu, assistant professor, big data analytics, UF/IFAS Department of Agricultural and Biological Engineering, Gainesville, FL; UF/IFAS Extension, Gainesville, FL 32611.

The Institute of Food and Agricultural Sciences (IFAS) is an Equal Opportunity Institution authorized to provide research, educational information and other services only to individuals and institutions that function with non-discrimination with respect to race, creed, color, religion, age, disability, sex, sexual orientation, marital status, national origin, political opinions or affiliations. For more information on obtaining other UF/IFAS Extension publications, contact your county's UF/IFAS Extension office. U.S. Department of Agriculture, UF/IFAS Extension Service, University of Florida, IFAS, Florida A & M University Cooperative Extension Program, and Boards of County Commissioners Cooperating. Andra Johnson, dean for UF/IFAS Extension.