Artificial Intelligence (AI) For Crop Yield Forecasting

Clyde Fraisse, Yiannis Ampatzidis, Sandra Guzmán, Wonsuk Lee, Christopher Martinez, Sanjay Shukla, Aditya Singh, and Ziwen Yu

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.
Artificial Intelligence (AI) For Crop Yield Forecasting

**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.

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.

---

**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. Credits: UF/IFAS

**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. Credits: Clyde Fraisse, UF/IFAS
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


<table>
<thead>
<tr>
<th>Crop (Region of the Study)</th>
<th>Objective</th>
<th>Input Variables</th>
<th>ML Algorithms Evaluated</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corn (US Corn Belt: Indiana, Illinois, Iowa)</td>
<td>Countywide corn yield levels</td>
<td>Outputs from crop model (APSIM) including biomass, and simulated yield combined with weather, management, and soil type</td>
<td>LASSO, linear regression, random forest, gradient boosting regressions</td>
<td>Shahhosseini et al. (2021)</td>
</tr>
<tr>
<td>Wheat (Australia)</td>
<td>Statistical division yield levels</td>
<td>Climate and satellite-derived vegetation indices</td>
<td>Support vector machine, random forest, and neural network</td>
<td>Cai et al. (2019)</td>
</tr>
<tr>
<td>Wheat (Florida)</td>
<td>Field plots of breeding programs</td>
<td>Hyperspectral imaging from UAVs</td>
<td>Functional regression</td>
<td>Costa et al. (2021)</td>
</tr>
<tr>
<td>Citrus (Florida)</td>
<td>Yield at the tree level</td>
<td>Combination of UAV multispectral imaging and ground-collected color imagery</td>
<td>Gradient boosting and random forest regressions, linear regression, and partial least squares regression</td>
<td>Vijayakumar et al. (2021)</td>
</tr>
<tr>
<td>Strawberry (Florida)</td>
<td>Yield at the plot level</td>
<td>UAV imaging and ground-collected color imagery</td>
<td>Region-based convolutional neural networks</td>
<td>Chen et al. (2019); Lee et al. (2020)</td>
</tr>
</tbody>
</table>