Harnessing AI to Transform Agriculture and Inform Agricultural Research

Debra P. C. Peters
USDA Agricultural Research Service

Adam Rivers
USDA Agricultural Research Service

Jerry L. Hatfield
USDA Agricultural Research Service

Danielle G. Lemay
USDA Agricultural Research Service

Simon Liu
USDA Agricultural Research Service

Bruno Basso
Michigan State University

***Abstract*—We provide an overview of the Special Issue on current advances, challenges, and opportunities for AI technologies in agriculture. We illustrate the potential of AI using four major components of the food system: production, distribution, consumption, and uncertainty. We recognize that the transformation of agriculture will require new tools to more precisely manage fields to increase production while minimizing the environmental risk to water and air quality. Combining AI with other technologies will be needed to provide effective production management strategies for a given combination of soil, climate, pest complexes, and vegetation. New methods will be needed to determine production limitations, and effective management options. The agricultural enterprise is prime for the use of AI and other technologies if they can be adapted for the unique characteristics of agroecosystems, including variability and directional changes in climate and other global change drivers as well as novel management and policy decisions, and economic market volatility.**

Digital Object Identifier 10.1109/MITP.2020.2986124
Date of current version 21 May 2020.

U.S. AGRICULTURE IS challenged to produce, store, and distribute nutritious, safe food from local to global scales at higher efficiencies with less environmental impacts under variable climate and changes in other environmental drivers.[1] Data availability for this food system is exploding, yet the agricultural research enterprise is often viewed as data rich and information poor. The research enterprise includes U.S. government agencies (e.g., U.S. Department of Agriculture—Agricultural Research Service [USDA ARS]), Land Grant Universities, private companies, and nonprofits, who often collaborate to deliver information to facilitate decision making by individuals or collectives who are directly influencing one or more components of the food system (stakeholders) or are in-charge of policies dealing with these components (decision-makers). There is a critical need to integrate these large amounts of data from many sources, types, and scales of resolution (time, space) into information that users can understand and use for management and decision making. AI technologies have the potential to transform agriculture and help meet these challenges. The use of AI technologies in agriculture is still in the early stages. In this overview paper, we illustrate the potential of AI technologies using examples for each of the four components of the food system (see Figure 1).

PRODUCTION

Estimating how much of what type of food will be produced from which part of the country or world PRIOR to its actual production is one of the most challenging tasks of U.S. agriculture. Variability in climate, weather, and extreme events can negatively impact crop yields and result in hard to predict losses and yields (see Figure 1). Yet this information governs the rest of the food production system, from the amount of surface and ground water, soil, and nutrients needed to grow that food to the biotic properties of the plants, livestock, and microbes that determine the nutritious quality of the food available to consumers (see Figure 1). At the farm-level, the

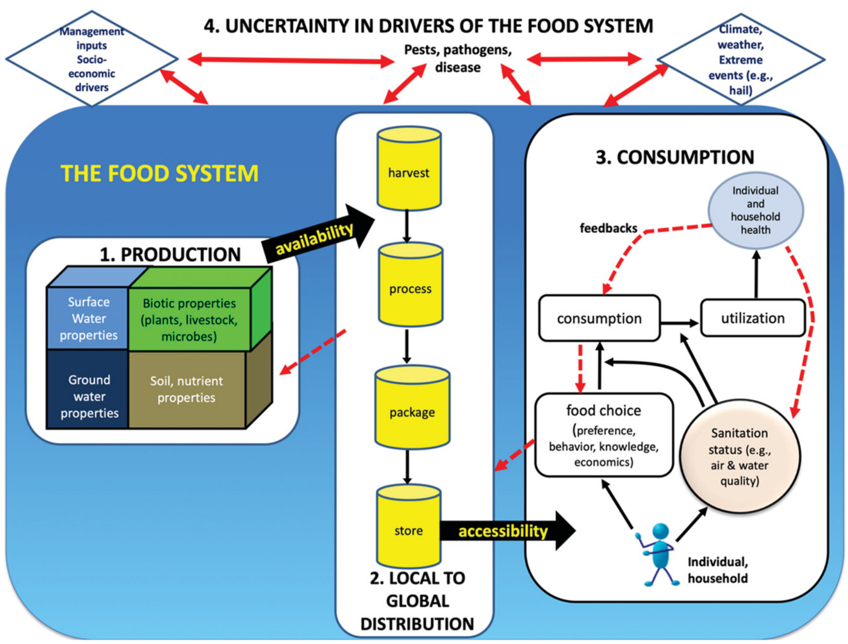

Figure 1. Four components of the food system. AI has an increasing role in all four. 1) On farm/ranch production is related to properties of plants, livestock, microbes, and nutrients, soils, and water, both surface and ground, that lead to available food that is; 2) distributed from local to global scales after being harvested, processes, packaged, and stored. 3) The individual or household make food choices related to variables, including preferences, behavior, knowledge, and economics, leading to food that is consumed, some of which is utilized, that depends on the sanitary conditions of the food and home, leading to the health of the individual and household. 4) Uncertainty in the drivers include human activities (socio-economic factors, management decisions), biotic factors (pests, pathogens, disease), and atmospheric conditions (climate, weather, extreme events such as hail, drought, hurricanes). Feedbacks (red arrows) are possible for many of the processes.

use of AI for crop nutrient management has provided precise and efficient approaches under these variable conditions,[2] and has been instrumental in capturing downstream effects of runoff and losses of nutrients.

The increasing availability of comparable data on climate, soils, and biotic variables representing critical processes within agroecosystems distributed at many sites across large geographic extents is allowing AI to be successfully applied to a variety of large-scale problems, often with nonintuitive results.[4] For example, an integrated approach based on AI that extrapolated results from within-field variability to the entire corn-belt demonstrated that reactive nitrogen (*N*) losses occurred from constantly underperforming areas within fields, and not from high performing areas.[3]

New insights can be gained by the integration and analysis of large datasets using AI and other analytical tools at large spatial extents over long time periods that were not possible previously. Recent analyses using ML for habitat modeling of animal disease at the continental-scale showed that the disease has two phases with environmental variables suitable for the habitat for different vectors in incursion years compared with expansion years.[5] These hypothesized vectors of importance are now being field-tested. In addition, an early warning system being developed based on environmental variables in months preceding onset of disease is providing actionable information to practitioners and livestock owners.[5]

DISTRIBUTION

The amount of food produced is important to the human capital and infrastructure required to harvest, process, package, and store food safely, and the distribution network of needed to transport and redistribute it (see Figure 1). Food access represents the complete cycle from field to table. One of the critical components of food access is ensuring the quality of food is acceptable to the consumer. Development of automated grading and sorting provides a better quality product, and can increase the speed and accuracy of evaluation.[6]

Another challenge for AI is to identify pathogens and predict consequences. For example, Salmonellosis sickens an estimated 1.35 million in the U.S. annually, and the economic impact has been placed at $3.4 billion annually. To track outbreaks, public health agencies need to type strains of Salmonella and predict their resistance to antibiotics. To control Salmonella in food production facilities, food safety and animal production regulators will culture, type, and monitor resistance to antibiotics and sanitizers. Traditionally, typing has been done by culturing bacteria then typing them genetically (today, whole-genome sequencing). These strains can also be screened for antibiotic resistance, or less commonly for sanitizer susceptibility. The increased popularity of clinical culture-independent diagnostic tests (CIDT) means that public health labs receive fewer clinical isolates for typing and antibiotic resistance profiling.[7] This is a major threat to the system for mitigating Salmonella outbreaks. The untargeted sequencing of all DNA in a sample (metagenomics) is a promising tool to type Salmonella from CIDTs but current methods of whole genome multilocus sequence typing methods require all genes to be present and complete. However, metagenomics data are often fragmented and contain missing regions.

We are developing an ML method to use gap-filled metagenomic data to predict the Salmonella type, sanitizer resistance, and antimicrobial resistance. This method creates reduced representations of the genome fingerprint using disentangled variational autoencoders. The method provides the best starting representation to train models for typing and phenotype prediction. Nearly a quarter of a million Salmonella genomes are now available to train these models. These data and smaller datasets containing antimicrobial and sanitizer resistance profiles provide a starting point to train supervised ML models, such as regularized feed-forward neural networks that predict type and resistance from incomplete or missing data, with high accuracy.

CONSUMPTION

Assuming the accessibility of high quality food, food choices are then made at the individual and household level based on numerous additional factors (see Figure 1). The food that is consumed and utilized is determined by the health of the individual and sanitation status of the food and household, such as air and water quality, that ultimately influence the health of the individual with feedbacks to consumption, sanitation status, and food choice. AI can be used to address the various challenges associated with food consumption as part of a comprehensive personalized nutrition approach.

A major challenge for AI is to track what individuals are eating and to predict the impact on their health. The USDA maintains food composition databases which are used by automated applications, such as ASA24,[8] to estimate nutrient intake. However, the time required and errors due to consumers' inability to accurately recall past dietary choices suggests that AI-augmented systems could potentially reduce burden and increase accuracy of dietary assessment. After quantifying diet, the next step is to predict the health impact of that diet; increasing evidence suggests that this needs to be personalized. For example, ML models demonstrated that person-specific characteristics are stronger predictors of glucose response ($R = 0.7$) to a challenge meal than the carbohydrate content of the meal ($R = 0.4$).[9] AI technologies and extensive, high-quality labeled data sets are needed to help consumers make food choices that maximize health and well-being while minimizing personal and societal costs.

UNCERTAINTY

Probably the greatest challenge to using AI technologies is uncertainty in that drivers of local agroecosystems, and agroecosystem dynamics themselves, may not reflect past drivers; thus training data used in AI technologies need to constantly evolve to reflect these changing conditions. The problem is complicated further for agroecosystems because these drivers can exhibit intra- and interannual variability as well as directional changes through time, and the drivers can interact with each other and with the food system that itself has feedbacks (see red-dashed arrows, Figure 1).

The agricultural enterprise is prime for the use of AI and other technologies if they can be adapted for the unique needs of agroecosystems that account for these changing drivers and their complex interactions. For example, these systems are particularly susceptible to variability in weather because annual plants in row crops lack storage organs to offset weather variability, and perennial crops and range plants are long-term investments with long lifespans and large storage organs that are unable to adapt to directional changes in climate. The agricultural enterprise is data rich given that some data have been collected since the 1860s. However, these data have not been fully utilized or integrated to connect local, farm, or ranch-scale data with regional to global knowledge networks. For example, many commercial products often use single season data for on-farm predictions rather than many years of data synthesized in a way for farmers to understand. Frameworks that integrate ML technologies with transdisciplinary expert knowledge and harmonized datasets of diverse variables across large, spatially heterogenous areas through time will be needed to address complex agroecosystem problems.

APPLICATIONS

The future of farming to meet the needs of the ever increasing human population requires that agriculture become more efficient in production. This increase will need to be conducted with less land area for production and increased climate variation.[1] To accomplish the task of increasing both the quality and quantity of agricultural production to meet demand will require a new research paradigm and more importantly, how the research is extended into the production, processing, storage, and distribution system.

Transformation of agriculture will require new tools to more precisely manage fields to increase production while minimizing the environmental risk to water and air quality. Combining tools to provide effective production management strategies for a given combination of soil, climate, pest complexes, and vegetation will require new methods of determining production limitations and using the predictive tools from AI to estimate the most effective management options (see Figure 2). One example in this issue illustrates how management decisions could be made in each yield stability zone and then linked with real-time weather data to run crop models and determine the effectiveness of management decisions.[10]

CONCLUSIONS

In this issue, we provide examples of the power of AI to address agricultural problems for food production at multiple scales from the local farm or ranch[2] to regional and continental

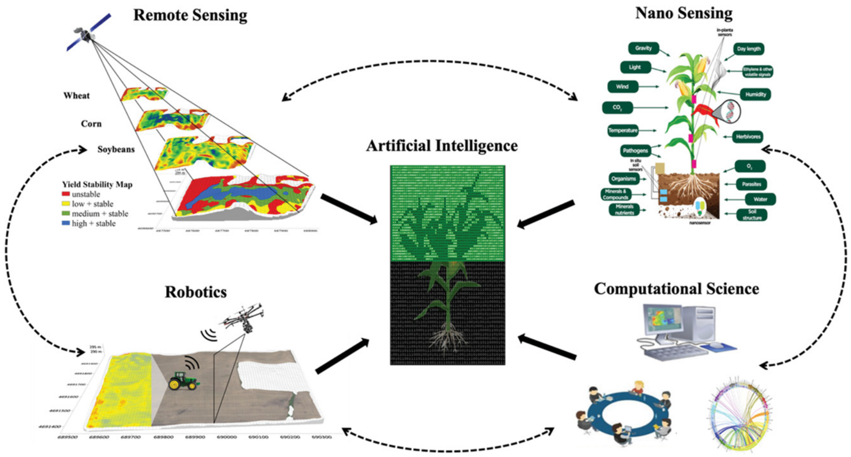

Figure 2. Future AI farming scenario. Imagery can provide georeferenced growth and yield by crop type for fields where nanosensors are collecting detailed measurements of plant growth as input to crop numerical models synthesized in AI predictive models to estimate the most effective management options applied as precision robotics by the farmer.

scales,[4] to determine accessibility and quality of food[6] leading to dietary choices and human health. The greatest challenge to applying AI technologies to agriculture lies with uncertainty in the food system where new advances will be needed to accommodate changes in environmental drivers, management decisions, and economic markets that will require evolving training data.

Although AI provides great promise for agriculture, these technologies will be insufficient for the problems of the future. AI needs to be integrated with approaches that allow for changes in training data and parameters with time. We encourage an integration of AI technologies with numerical models and human-guided ML as future research tools. In this issue, we show how recommender systems with ML use a broad range of techniques to learn from past experiences to better inform current choices.[11] We provide examples where an integration of nonnumeric forms of data, such as text, images, and sounds, can provide greater understanding of agroecosystems.[12] AI approaches that focus on extreme events, such as drought, are particularly relevant[13] and will likely need to be extended in the future as these weather events become more extreme than occurred historically.

ACKNOWLEDGMENTS

This work was supported by USDA-ARS CRIS Projects 6235-11210-007, 2032-51530-026-00D, and 5030-11610-005-00D.

REFERENCES

1. J. L. Hatfield and C. L. Walthall, “Meeting global food needs: Realizing the potential via genetics x environment x management interactions,” *Agronomy J.*, vol. 107, no. 4, pp. 1215–1226, 2015.
2. K. A. Sudduth, M. J. Woodward-Greene, B. W. Penning, M. A. Locke, A. Rivers, and K. S. Veum, “AI down on the farm,” *IT Professional*, vol. 22, no. 3, May/Jun. 2020.
3. B. Basso, G. Shuai, J. Zhang, and G. P. Robertson, “Yield stability analysis reveals sources of large-scale nitrogen loss from the US Midwest,” *Scientific Rep.*, vol. 9, 2019, Art. no. 5774.
4. B. T. Bestelmeyer *et al.*, “Scaling up agricultural research with AI,” *IT Professional*, vol. 22, no. 3, May/Jun. 2020.
5. D. P. C. Peters *et al.*, “Developing multi-scale early warning strategies for vector-borne disease using big data-model integration and machine learning,” *Ecosphere*. (in press)
6. B. W. Penning, W. M. Snelling, and M. J. Woodward-Greene, “Machine learning in the assessment of meat quality,” to be published.
7. G. Langley *et al.*, “Effect of culture-independent diagnostic tests on future emerging infections program surveillance,” *Emerg. Infectious Disease*, vol. 21, no. 9, pp. 1582–1588, 2015.
8. U.S. Department of Agriculture, Agricultural Research Service. Food Data Central, 2019. [Online]. Available: fdc.nal.usda.gov. https://epi.grants.cancer.gov/asa24
9. D. Zeevi *et al.*, “Personalized nutrition by prediction of glycemic responses,” *Cell*, vol. 163, no. 5, pp. 1079–1094, 2015.

10. J. L. Hatfield, M. Cryder, and B. Basso, "Remote sensing: advancing the science and the applications to transform agriculture," *IT Professional*, vol. 22, no. 3, May/Jun. 2020.
11. D. P. C. Peters, H. M. Savoy, G. A. Ramirez, and H. H. Huang, "AI with ML recommender systems for agricultural research," *IT Professional*, vol. 22, no. 3, May/Jun. 2020.
12. C. S. Parr, D. G. Lemay, C. L. Owen, M. J. Woodward-Greene, and J Sun, "Multi-modal AI to improve agriculture," to be published.
13. P. Ramos-Giraldo, C. Reberg-Horton, A. M. Locke, S. Mirsky, and E. Lobaton, "Drought stress detection using low-cost computer vision system and machine learning techniques," *IT Professional*, vol. 22, no. 3, May/Jun. 2020.

Debra P. C. Peters is currently the Acting Chief Science Information Officer with the USDA Agricultural Research Service, Beltsville, MD, USA, and a Research Scientist with the USDA-ARS Jornada Experimental Range, Las Cruces, NM, USA. Contact her at deb.peters@usda.gov.

Jerry Hatfield is currently a Laboratory Director and Supervisory Plant Physiologist with the National Laboratory for Agriculture and the Environment, Ames, IA, USA. Contact him at jerry.hatfield@usda.gov.

Simon Liu is currently an Associate Administrator with the USDA-ARS, Beltsville, MD, USA. Contact him at simon.liu@usda.gov.

Adam Rivers is currently a Computational Biologist with the USDA ARS, Agricultural Microbiomes Group, Gainsville, FL, USA. Contact him at adam.rivers@usda.gov.

Danielle Lemay is currently a Research Scientist with USDA ARS Western Human Nutrition Research Center, Davis, CA, USA. Contact her at danielle.lemay@usda.gov.

Bruno Basso is currently a Professor with Michigan State University, East Lansing, MI, USA. Contact him at basso@msu.edu.