CHAPTER FIVE

Use of artificial intelligence in soybean breeding and production

Asheesh K. Singh[a,*], Sarah E. Jones[a], Liza Van der Laan[a], Timilehin T. Ayanlade[b], Joscif Raigne[a], Nasla Saleem[b], Shambhavi Joshi[b], Muhammad Arbab Arshad[c], Hossein ZareMehrjerdi[d], Ashlyn Rairdin[a], Juan Di Salvo[a], Dinakaran Elango[a], Leonardo De Azevedo Peixoto[a], Talukder Z. Jubery[b], Adarsh Krishnamurthy[b], Arti Singh[a], Soumik Sarkar[b], and Baskar Ganapathysubramanian[b]

[a]Department of Agronomy, Iowa State University, Ames, IA, United States
[b]Department of Mechanical Engineering, Iowa State University, Ames, IA, United States
[c]Department of Computer Science, Iowa State University, Ames, IA, United States
[d]Department of Computer Engineering, Iowa State University, Ames, IA, United States
*Corresponding author. e-mail address: singhak@iastate.edu

Contents

Advances in Agronomy, Volume 190
ISSN 0065-2113, https://doi.org/10.1016/bs.agron.2024.11.003

Abstract

Artificial intelligence (AI) in soybean research has revolutionized various crop improvement and production aspects. This review provides predominant areas that have seen the use of AI. AI applications in phenomics have enabled collecting and analyzing high-dimensional data in soybean plants, from below- to above-ground traits, predicting phenotypes, and identifying complex patterns. In genomics, AI has improved genomic selection accuracy and identified genomic regions associated with traits of interest, such as resistance to biotic and abiotic stresses. AI has also been extensively used in detecting and managing biotic and abiotic plant stresses using RGB, multispectral, and thermal imagery from ground-based and aerial platforms. Additionally, AI has shown significant potential in yield prediction, incorporating factors such as vegetation indices, weather data, and soil properties. This review explains the concept of cyber-agricultural systems (CAS) that integrates AI, advanced sensing, computational modeling, and scalable cyberinfrastructure to optimize soybean production, enhance resource management, reduce environmental impact, and improve farm efficiency. We explain the use of CAS in crop improvement as well. We provide an exhaustive listing of challenges and future direction in the integration of AI in soybean production and crop improvement, including multi-modal and layered sensing, data availability and quality, computational modeling, AI models and tools, Cyberinfrastructure, Explainability and interpretability of AI models, AI-related impacts on privacy, ethics, and policy, Impact on Smallholder Farmers, Digital Twin, Large Soybean Datasets for community usage, and Immersive environments.

1. Introduction

Soybeans are an essential source of both protein and oil, playing a critical role in human and animal nutrition. Soybean offers a high-quality protein source that helps address global nutritional and feed needs, particularly in developing regions (Hartman et al., 2011). In animal diets, soybean meal significantly boosts livestock productivity by providing essential amino acids, thereby supporting the global meat supply chain

(Graham and Vance, 2003). The versatility of soybean extends beyond nutrition, as they are also pivotal in various industries, including sustainable aviation fuel and biodiesel production, further underscoring their economic importance (Hartman et al., 2011). Moreover, soybean contributes to sustainable agricultural practices through biological nitrogen fixation. This process enhances soil fertility by converting atmospheric nitrogen into a form usable by plants, thereby reducing the reliance on synthetic nitrogen fertilizers (Graham and Vance, 2003).

Soybean production has faced numerous challenges as acreage has expanded. The effects of biotic and abiotic pressure individually, and in combination, can negatively impact soybean yield, resulting in billions of dollars in U.S. crop insurance payments, economic losses for farmers, and increased prices for consumers (Dice and Rodziewicz, 2020). These stresses can be exacerbated by extreme weather events, which are becoming more prevalent due to climate change (Raymond et al., 2020). Across crop species, the main abiotic stress factors associated with yield losses are drought, heat, cold, and soil salinity (Oerke, 2006). Water deficit, or drought, is a major abiotic factor that affects the yield of crop species around the world, and is considered as one of the main constraints on yield potential in the highly productive US Corn Belt region (Yang and Wang, 2023). Soybean yield losses due to drought stem from reduced growth and development rate, and have been found to cause a yield reduction of 40 % (Specht et al., 1999). A recent study on the magnitude, frequency, duration, and timing of droughts showed that more than two thirds of global soybean acreage is at high risk of severe droughts causing reduced yield in soybean (Santini et al., 2022). Rising global temperatures, that often accompany drought stress, are predicted to increase the prevalence of heat stress in many soybean production regions (Teixeira et al., 2013). While heat stress has generally not been considered a significant constraint on soybean production worldwide, end-of-the-century climate projections show it to be an increasingly important factor affecting the yields of numerous crop species, including soybean (Bezner Kerr et al., 2022), which is projected to have yield losses of up to 22 % due to heat stress (Yang and Wang, 2023). Additionally, cold temperatures can negatively impact soybean. Depending on the cultivation area, cold stress can affect the germination and seedling establishment, particularly if planting dates are earlier, and early frosts can affect seed development at the end of the season (Ohnishi et al., 2010).

In addition to abiotic stresses, soybean is also exposed to yield limiting biotic stresses such as insect pests, diseases, and weeds. In 2022, the most severe biotic stress in both the Northern and Southern United States as well as Ontario, Canada was soybean cyst nematode (*Heterodera glycines*) causing over 90 million bushels in yield loss (Allen et al., 2023). In the Northern United States and Ontario, Canada, the fungal disease sudden death syndrome (*Fusarium virguliforme*) ranks second causing over 19 million bushels yield loss, while in the southern region of the US, root knot nematode (*Meloidogyne spp.*) takes second place causing over 13 million bushels in yield loss (Allen et al., 2023). Poorly controlled weeds, especially in late-season, can cause yield losses up to 48 % (Landau et al., 2022). Insects are constantly on the move and new insect pests can emerge to cause damage such as the soybean gall midge (*Resseliella maxima* Gagné) first reported in 2019 in the Midwest United States (Gagné et al., 2019) and can cause yield losses from 17–31 % in soybean (McMechan et al., 2021). Biotic stresses pose a unique challenge in that they are also constantly evolving to meet and overcome genetic and management strategies developed for biotic stress mitigation. Unfortunately, the United States leads the world with 132 species of herbicide resistant weeds, with a growing number of species developing resistance to multiple herbicide modes of action (Heap, 2024). To make matters worse, soybean cyst nematode has developed the ability to reproduce on a highly utilized source of resistance from PI 88788 (McCarville et al., 2017) that was predominantly used in breeding due to its effectiveness and incorporation in high-yielding genetic materials. These unique challenges highlight the need for both diverse genetic and management strategies to safeguard the genetic and chemical strategies for as long as possible. As the problems exacerbate and new issues emerge in crop production, we must investigate state-of-the-art technology and novel methods to breed higher-yielding, more resilient crops.

Soybean breeders have adapted and developed new technology to face the various challenges in today's production of soybeans. Among these tools is artificial intelligence (AI) where computer models are trained to process information in a manner similar to the human brain. AI and deep learning (DL) have emerged as transformative tools in modern agriculture, reshaping traditional practices and fostering innovation (Pathan et al., 2020). The integration of AI in soybean improvement and production represents new approaches to meet the global protein and oil needs. AI technologies can help improve plant breeding efficiency and success, optimize crop production and management, and help control diseases and pests, leading to increased productivity and sustainability. These technological advancements are crucial for meeting the rising global demand for soybeans, ensuring food security, and maintaining economic stability Negus et al. (2024). The soybean industry can achieve more efficient and resilient production systems by leveraging AI, ultimately contributing to a more secure and sustainable agricultural future.

This paradigm shift is propelled by the exponential growth of data and the advent of high-throughput imaging technologies, which generate vast amounts of valuable information regarding plant health and environmental conditions (Araus and Cairns, 2014). The emergence of advanced technologies such as drones, ground robots, and sensors, has brought high-throughput phenotyping and phenomics to the forefront, transforming the measurement of multiple plant traits across various growth stages and facilitating rapid, precise, and accurate data collection (Feng et al., 2021; Guo et al., 2021). DL plays a pivotal role in harnessing this deluge of data for the identification, classification, quantification, and prediction (ICQP) of plant stress phenomena (Singh et al., 2016). By leveraging its capacity for sophisticated data analysis and pattern recognition, DL enables the extraction of valuable insights from complex datasets Feuer et al. (2024). This empowers farmers and researchers to make informed decisions regarding crop management practices, resource allocation, and stress mitigation strategies. AI advancement, in conjunction with data generation and sensing technology, is solving complex problems.

This review article provides a comprehensive overview of the literature on AI in soybean breeding and production, along with basics of machine learning, deep learning, and artificial intelligence. Several AI application areas are covered, including phenomics, genomics, and cyber-agricultural systems. We present challenges that need to be overcome for a full realization of AI potential and present future directions in plant breeding and crop production.

2. Machine learning, deep learning, and artificial intelligence

Machine Learning (ML), Deep Learning (DL), and Artificial Intelligence (AI) are key components of modern data-driven technology, each with unique roles that are closely connected (Fig. 1). AI is the broadest term and refers to creating systems that can perform tasks that

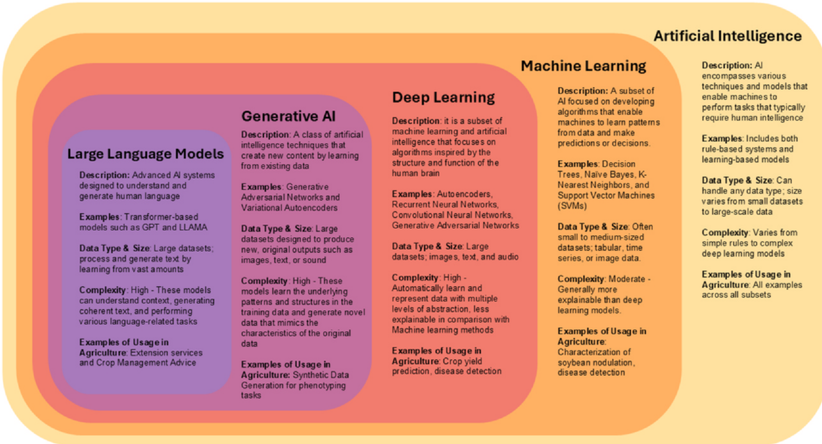

Fig. 1 Artificial intelligence, Machine learning, Deep learning, Generative AI and Large language models: Description of the terms, Data type and size, Complexity and Examples of usage.

usually require human intelligence, such as making decisions, solving problems, and learning from experience. For example, an AI system in agriculture might analyze weather data and plant health to recommend the best time to harvest crops.

ML is a part of AI that focuses on creating algorithms that allow computers to learn from data and make predictions. Think of it as teaching a computer to recognize patterns and make decisions based on those patterns. ML can be divided into three main types: (a) supervised learning, (b) unsupervised learning, and (c) reinforcement learning. In supervised learning, the computer is trained on a labeled dataset, meaning each piece of data has a correct answer provided. For example, labeled images of healthy and diseased plants can be used to teach the computer to identify diseases. Unsupervised learning involves the computer looking for patterns in data without any labels, similar to sorting a mixed bag of seeds into groups without knowing what each seed type is. Reinforcement learning involves computer learning by trial and error, receiving rewards for correct actions, much like training a dog with treats for good behavior.

Numerous ML techniques have demonstrated their efficacy in soybean phenotyping. Common regression methods used include linear regression, logistic regression, stepwise regression, ridge regression, partial least squares regression, elastic net regression, piecewise regression, tree regression, and Gaussian process regression. Classification methods frequently employed are Naive Bayes, decision trees, random forests, K-nearest neighbor, linear discriminant analysis, quadratic discriminant analysis, support vector machines (SVM), and extreme learning machines. These methods have been applied to various phenotyping tasks, providing valuable insights into soybean traits and growth patterns. These are covered in more detail in (Singh et al., 2016; Gill et al., 2022b).

DL is a more advanced subset of ML. It uses complex structures called neural networks with many layers (hence "deep") to process and learn from large amounts of data. DL is especially powerful for tasks involving images, audio, and text. In plant breeding and phenotyping, DL can be used to analyze images of plants to detect diseases, measure growth, and predict yields. This advanced capability allows breeders, researchers, and farmers to gain valuable insights from complex datasets, leading to more precise and efficient agricultural practices.

DL algorithms have shown significant promise in extracting valuable insights from plant phenotype data. Advances in automation, computation, and sensor technology have facilitated the collection of high-resolution phenotype data across extensive geographical areas with high temporal resolution. This influx of data has enabled the successful application of DL algorithms to a wide range of plant phenotyping tasks. Deep learning methods, particularly Convolutional Neural Networks (CNN), have excelled in challenging tasks such as plant disease classification (Singh et al., 2018). Additionally, DL methods have achieved state-of-the-art performance in complex image-based phenotyping problems, such as root and shoot feature identification and localization (Jubery et al., 2021). Other commonly used deep learning models in soybean phenotyping include Multilayer Perceptrons (MLP), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) (Shook et al., 2021). Other research areas include generative deep learning, super-resolution, dehazing, and spectral reconstruction, which aim to enhance sensor-based phenotype information (Shoeiby et al., 2019). Reinforcement learning (RL) is also a developing area, particularly useful for optimizing phenotyping strategies and improving decision-making processes in soybean research as prototyped in (Hitti et al., 2024). A comprehensive discussion on DL for plant phenotyping can be found in (Singh et al., 2018).

Currently, transformer-based models, such as vision transformers (Dosovitskiy et al., 2020) have emerged as powerful tools in deep learning due to their ability to capture long-range dependencies and contextual information. The performance of these models has led to their adoption in phenotyping tasks for improved phenotyping performance (Bi et al., 2023). Transformer-based models are particularly effective in handling large-scale phenotyping data, offering improved performance and scalability for soybean research tasks (Yang et al., 2022).

In soybean research, the collection of vast amounts of data has outpaced the available expertise for labeling, resulting in a substantial amount of unlabeled data. This imbalance poses a significant challenge for training effective deep learning models. However, several techniques can help address this issue, some of which are explained below:

- Active Learning: This approach involves selectively querying the most informative data points for labeling, thereby maximizing the efficiency of the labeling process. By focusing on the most uncertain or diverse samples, active learning can significantly reduce the amount of labeled data required for training while maintaining model performance This approach has been successfully implemented in soybean leaf stress classification (Nagasubramanian et al., 2021).
- Transfer Learning: Transfer learning leverages pre-trained models on related tasks or domains to improve performance on the target task. In soybean research, models pre-trained on large, annotated datasets from similar crops or agricultural tasks can be fine-tuned with a smaller amount of soybean-specific data, thereby overcoming the labeling bottleneck (Yang et al., 2021).
- Self-Supervised Learning: This technique uses the data itself to generate supervisory signals, enabling the model to learn useful representations from unlabeled data. For example, by learning augmented views of the data from various angles and parts, the model can learn robust features without requiring extensive labeled datasets (Chiranjeevi et al., 2023).

Computer Vision (CV), a field of AI that enables machines to interpret visual information, is increasingly applied in soybean breeding and production for high-throughput phenotyping. Using image processing techniques and machine learning algorithms, including deep learning models like Convolutional Neural Networks, CV automates the extraction of plant traits from images. These pipelines can measure plant height, leaf area, and canopy cover from aerial images; detect diseases based on leaf symptoms; and assess pod and seed characteristics. This approach allows breeders to evaluate large plant populations more efficiently than manual methods, potentially capturing subtle variations that human observers might miss. Recent developments in imaging technologies, such as multispectral and hyperspectral cameras, 3D imaging including LiDAR (Light Detection and Ranging), and photogrammetry, have expanded CV applications in soybean phenotyping. For example, hyperspectral imaging has been used to assess drought stress responses and nitrogen status in soybeans Li et al. (2020). These 3D techniques provide more accurate measurements of plant architecture and biomass Paulus (2019). As imaging technologies progress, CV may provide new insights into soybean phenotypes, contributing to crop improvement efforts.

3. Phenotyping and AI

Automated plant phenotyping pipelines are essential for measuring plant traits efficiently and accurately. Sensors, when integrated across multiple platforms, capture diverse data types that are critical for these phenotyping pipelines. They offer powerful tools for gathering data at various scales and resolutions. These pipelines rely heavily on robust image processing and AI algorithms to extract meaningful insights from the data (see Fig. 1). In addition, they offer promising approaches for faster and more efficient analytics, enabling researchers to uncover patterns and features from the large volumes of data generated by high-throughput phenotyping platforms (Ghosal et al., 2018). Researchers have developed and applied various image processing algorithms to measure different plant traits. Techniques such as segmentation, classification, feature extraction, skeletonization, graph-based algorithms, and morphological operations have been widely used for this purpose (Hamuda et al., 2016; Arnal Barbedo, 2013; Kumar and Raghavendra, 2019). These tools are essential for extracting meaningful information from images, enabling accurate and high-throughput phenotyping of soybean plants.

High-throughput phenotyping (HTP) is an advanced method that allows for the rapid measurement and analysis of plant traits using automated imaging and data processing techniques. Combining various platforms and sensors, HTP systems can capture large volumes of phenotypic data in a short amount of time, significantly enhancing the efficiency of phenotyping beyond traditional manual and visual methods that can be time-consuming and labor-intensive. Furthermore, large breeding programs and farm fields benefit from the use of technology that can span a field faster than a human alone to allow for increased population sizes and full coverage scouting in a farm field. HTP can be applied at multiple levels of crop growth and development from microscopic (Akintayo et al., 2018), to single leaf (Yu et al., 2024), root (Carley et al., 2023), plant canopy (Naik et al., 2017), test plots (Parmley et al., 2019a), to large scale field evaluation (Song et al., 2017) assisted by specially designed HTP systems. These advancements in HTP provide valuable insights into plant growth patterns and improve the precision of phenotypic evaluations in soybean breeding programs (Singh et al., 2021b).

3.1 Sensors for data collection

Data is the foundation for developing AI applications in soybean breeding and production. Large and high quality data sets are important for development of useful techniques and models. The sensors commonly used in high throughput phenotyping can be organized by their data output such as digital or numeric, by their radiation source (active or passive), and by the range of the electromagnetic (EM) spectrum utilized (Singh et al., 2021b). Below we discuss sensors, data types, and the various platforms utilized to facilitate data collection.

RGB sensors, much like the human eye, are able to capture color differences in the visual range (400–700 nm), specifically red, green, and blue (RGB) channels. RGB sensors are commonly used for agricultural purposes due to their accessibility, affordability, high resolution, and versatility (Singh et al., 2021a). The ability to detect and differentiate colors is a fundamental strength of these cameras, which makes them particularly useful in a variety of applications in soybean phenotyping, such as nutrient deficiency screening (Dobbels and Lorenz, 2019; Naik et al., 2017), disease classification (Ghosal et al., 2018), disease quantification (Rairdin et al., 2022), and even yield and agronomic trait prediction (Yuan et al., 2019). RGB sensors excel in morphological phenotyping of various plant parts such as roots and nodules, extracting traits that would be insurmountable to collect manually (Falk et al., 2020a,b; Carley et al., 2023). In associated agricultural tasks, RGB sensors have been used for insect detection (Chiranjeevi et al., 2023), weed density assessment dos Santos Ferreira et al. (2017), weed identification Zou et al. (2023), and weed management through precision agriculture (Staff, 2022).

Multispectral sensors have a wider spectral range than RGB sensors and typically utilize about 3–10 wavebands (Singh et al., 2021a). Wavebands are generally wider, containing several wavelengths, and non-continuous. In addition to RGB bands, multispectral cameras often contain red-edge and/or near-infrared (NIR) bands which can capture non-visual clues for plant health assessment and early detection (Jones et al., 2024). Unlike RGB sensors, multispectral sensors enable calculation of many more vegetation indices that augment data, including commonly used vegetation indices such as NDVI index (Rouse et al., 1973). Multispectral sensors excel in situations where information beyond the visual spectrum is beneficial and are often used in soybean stress screening (Zhou et al., 2021, 2020), and have been used in soybean yield prediction as well (Herrero-Huerta et al., 2020).

Hyperspectral sensors are found in two forms including radiometers with purely spectral dimensions in digital number output, as well as imaging sensors with spatial and spectral dimensions. Hyperspectral sensors cover a larger range of the EM spectrum at a much higher and continuous density compared to multispectral sensors. The range and density of hyperspectral sensors enable calculation of a wider range of vegetation indices targeting very specific single wavelength bands and utilizing wavelengths from the ultra-violet, to infrared, and short-wave infrared regions. In soybean, several studies utilizing hyperspectral imaging have been performed in the lab or greenhouse for early and accurate disease detection and severity estimation (Nagasubramanian et al., 2018, 2019). Studies utilizing hyperspectral radiometers have been successful in the field for yield prediction and predictive breeding (Parmley et al., 2019b,a).

Thermal sensors capture infrared radiation at a far range of the EM spectrum compared to RGB, multispectral, and hyperspectral sensors. In soybean, research found that thermal data can help identify candidate genes for drought tolerance in soybean (Bazzer and Purcell, 2020), is closely related to drought induced canopy wilting (Bai and Purcell, 2018) as well as soybean crop water status (Crusiol et al., 2020). Thermal sensors commonly used in plant phenotyping come in several forms including non-contact imagers and radiometers which capture thermal infrared radiation emitted, and direct contact thermocouples. Non-contact thermal imagers have an advantage in phenotyping speed as they can be mounted on drones, however, these methods do not come without challenges. Accuracy of thermal imaging can vary widely and is sensitive to environmental parameters such as air temperature, flight direction, solar angle, humidity, and cloud cover among other factors (Perich et al., 2020) and should be evaluated closely for accurate phenotyping.

3D sensor (three-dimensional) is a device that attains information about the physical/natural world in three dimensions (X, Y, Z), in addition to the information gathered by 2D image data. The primary usage of 3D sensors is to gather depth, which tells the distance between the object and the sensor for every point in the image. The sensors can also measure the size, volume, shape, contour, surface texture, and light reflectivity, making them useful for applications in diverse industries. These sensors allow access to the plant architecture, enabling tracking of the physical development, and define different parameters of plant organs and canopies (Paulus, 2019; Young et al., 2023). 3D sensors have the potential to provide both qualitative and quantitative measures of plants, capturing detailed information such as the number and shapes of leaves, size, surface area, and the architecture of the plant (Paulus, 2019; Salter et al., 2021). They can accurately assess branch and leaf angles, contributing significantly to our understanding of plant morphology and behavior under various environmental conditions (Liu et al., 2019a; Zhou, 2022; Young et al., 2024). Importantly, all this data is gathered non-destructively, allowing continuous monitoring and analysis without harming the plant. A range of 3D sensors offer varied outputs, crucial for detailed spatial analysis and precise modeling across different applications.

- Point Cloud: Point cloud data is produced by specific 3D sensors that capture the spatial data as points. Point cloud data have coordinates and color information for each point of the environment detected by sensors. This data type is useful for applications requiring high-precision modeling and mapping.
  - (a) LiDaR: Light detecting and ranging system (LiDAR) technology emits pulses of light by laser beams and then calculates how long it takes for each light pulse to return, hence calculating the distance to objects. This results in dense point clouds that accurately represent 3D representations of the environment. The LiDAR system can collect data by airborne laser scanner, terrestrial, or mobile laser scanner, serving diverse applications. LiDAR systems can be utilized for measuring crop features, detecting objects, evaluating biomass, and planning agricultural activities (Rivera et al., 2023).
  - (b) Structured Light Scanners: Similar to LiDAR, Structured Light Scanners detect distances using light. These scanners emit a known pattern of light onto an object and calculate the deformation of this pattern when it reflects (Georgopoulos et al., 2010). This process creates a dense 3D point cloud of the object. Additionally, in detailed agricultural and environmental studies, these sensors provide depth and RGB data (RGB-D), facilitating the classification and detection of various objects, such as leaves, branches, flowers, and fruits (Harandi et al., 2023).

- (c) Photogrammetry: Photogrammetry uses overlapping 2D images captured from different perspectives to reconstruct detailed 3D models with texture mappings. This technology creates lifelike and textured 3D representations, ideal for precise modeling of objects and environments. It can be utilized in any situation where the object can be photographically recorded, reconstructing them into multiple digital formats like coordinates, point clouds, and meshes (Luhmann et al., 2023). Significantly, photogrammetry is proficient at generating detailed digital twins of agricultural fields, facilitating precise soil roughness measurement, and enhancing crop management strategies (Gilliot et al., 2017).

- Depth Map: Depth map data encodes the distance of objects from the camera lens, where each pixel represents the distance from the camera. This data typically employs a grayscale or color scale to visually indicate varying distances, clearly depicting the object's depth. This format is an effective way to visually assess objects' spatial arrangement and depth in an environment (Chen et al., 2023).
  - (a) Stereo Vision Cameras: Stereo Vision Cameras measure 3D shapes and typically consist of two or more image sensors positioned at slightly different angles. These cameras capture images simultaneously from multiple perspectives, allowing them to collect in-depth information based on the disparity between the images (Rosell-Polo et al., 2015). This technique is highly effective in applications such as autonomous navigation for robots in various environments and generating detailed 3D terrain maps (Rovira-Más et al., 2008).
  - (b) Time-of-Flight Cameras: Time-of-flight (ToF) cameras measure full-range distances in real time. They first light the scene with modulated infrared light and then measure the phase shift between the reference signal and the reflected light. The system allows for precise depth mapping, which is essential in environments requiring rapid and precise distance measurements (Lindner et al., 2010). These low-cost ToF sensors can be directly utilized to derive crop height models, eliminating the need for prior terrain measurements (Hämmerle and Höfle, 2016).

All of the above-mentioned sensors capture information about plants from different point of view, providing in-depth phenotyping possibilities.

The integration of ML, DL, and AI with 3D sensor data enhances plant phenotyping. ML algorithms, such as Random Forest (RF), can be utilized to analyze complex 3D datasets for yield prediction, leaf area index, and biomass (Randelović et al., 2023). DL excels in extracting characteristics from 3D imaging by using CNNs, improving detection and classification of plant structures. For example, (Zhao et al., 2022) demonstrated that DL-based 3D reconstruction from single RGB photos could effectively estimate phenotypic variables. Their technique involves applying deep learning to construct 3D models of plants from simple 2D photos, allowing for phenotypic measurements of plant height, trunk diameter, canopy size and analysis of plant growth status. AI leverages ML and DL results from numerous sensors to optimize crop management and breeding strategies, allowing for real-time monitoring and decision-making. Recent studies show that incorporating AI with sensors and robots can assess plant features, measure physiological parameters, detect diseases, and predict crop yields and performances (Qiao et al., 2022).

3.2 Platforms for carrying sensors

Numerous platform options exist to carry suites of sensors for use in phenotyping soybean, as well as other crop species. The broad categories for such platforms include proximal and aerial platforms. Proximal platforms enable proximal phenotyping of plants, which is the phenotyping of crop plants via ground-based, non-destructive approaches for in situ measurements. Field carts, or proximal sensing carts (PSCs), are a low-cost platform that requires manual pushing or pulling to move through the field, although more advanced versions are motorized to reduce the labor of moving such carts through the field (Alison et al., 2018). PSCs are often lightweight for minimal soil disturbance and can be customized to suit the height and row spacing as required for the field. Additionally, sensors can easily be added, removed, or re-positioned for different requirements, allowing for a flexible platform, such as collecting canopy data in soybean (Parmley et al., 2019a). Another proximal platform is field rovers, which are semi-autonomous and require less human labor for operation. Rovers can be built narrow to fit within rows of soybean and other crops or have a high clearance to go over crops. Today, multiple commercial options of rovers have been developed for use in agriculture research (Farm-ng, 2024; EarthSense, 2021). Rovers equipped with LiDAR and RGB cameras and has been used in soybean to estimate yield by tracking and counting soybean pod numbers (McGuire et al., 2021; Riera et al., 2021).

Uncrewed aerial vehicles (UAVs), commonly called drones, have increased in popularity for plant phenotyping due to their high throughput ability.

UAVs can collect data in a shorter amount of time compared to the ground-based methods, have less of a risk of damaging the plants, and no risk of soil compaction due to the traffic of the platform in the field. The speed of UAVs allows for a higher temporal and spatial resolution of data (Xie and Yang, 2020), which provides for phenotyping with greater accuracy. Different classes of UAVs exist, with four broad categories being single-rotors, multi-rotors, fixed wings, and vertical takeoff and landing (VTOL) (Guo et al., 2021). These different categories of UAVs differ in their payload capacity, maximum flight time, and ease of operation, which are all factors to consider when deploying UAVs in soybean research and production. An additional platform for aerial sensing is the use of satellites, which can carry sensors for panchromatic, multispectral, and hyperspectral imagery. Satellites have been used in soybean for on-farm yield forecasting (Schwalbert et al., 2020), as well as for predicting soybean seed composition (Hernandez et al., 2023). While satellites do not require a trained operator like UAVs do, several limitations need to be considered. These limitations include the resolution of the cameras, weather conditions, and the revisit frequency for imaging. As of 2023, the average farm size in the United States is 464 acres, making crop stress scouting a time-consuming challenge by traditional on-the-ground methods (USDA-NASS, 2024). Drones and satellite platforms offer rapid methods for large-scale imaging of crop fields, enable measurement of wavelengths beyond human vision, and facilitate high-throughput phenotyping in major crop breeding programs (Herr et al., 2023).

4. Applications of AI in soybean improvement and production

4.1 Phenomics

The soybean phenotype results from the interaction of a plant's genotype, environment, and management practices (Furbank and Tester, 2011). Plant phenomics involves large-scale collection of high-dimensional data across an organism (Houle et al., 2010). Advanced phenotyping platforms and sensors have created a deluge of data requiring sophisticated analysis methods (Singh et al., 2016). Characterizing soybean phenotypes has direct advantages in breeding by enhancing the understanding of genotype-phenotype interactions. This facilitates rapid selection in breeding programs by identifying genes of interest (Rairdin et al., 2022) and is key to understanding the genetic basis of complex traits (Houle et al., 2010).

Deep learning models, particularly CNNs and RNNs, have revolutionized phenotype prediction in crops. CNNs excel at analyzing spatial data from genomic sequences, while RNNs capture temporal changes over plant developmental stages (Gao et al., 2023; Ray et al., 2023). These models excel in processing and analyzing high-dimensional phenotypic data, such as images capturing plant morphology and developmental stages (Liu et al., 2019b). By leveraging large datasets, CNNs can identify complex patterns and subtle variations in phenotypic traits, such as leaf shape, plant height, and biomass, which are crucial for assessing plant health and vigor. In addition to CNNs, advanced AI techniques have been applied to enhance the precision and accuracy of phenomic predictions (Xu et al., 2022). The integration of diverse datasets provides comprehensive insights into plant responses and adaptations, supporting the selection of superior genotypes for breeding. Additional challenges, such as the integration of multiple modalities of sensing has been met by new models such as the RGB and Infrared Feature Fusion Segmentation Network (RIFSeg-Net) that utilizes a Res-Net backbone (Yu et al., 2024). The model combines images from multiple modalities creating a single mask for segmentation enabled by the Segment Anything Model (SAM) (Kirillov et al., 2023) that accurately extracts individual leaves from canopies.

AI enabled phenomics is transforming soybean breeding, enabling collection of previously difficult-to-measure or labor intensive traits (Fig. 2). A mobile, low-cost root phenotyping system using computer vision and machine learning was developed for high-throughput analysis of root system architecture traits (Falk et al., 2020a). Novel methods pair AI with phenotyping to increase automation and feasible population sizes. Examples include using hyperspectral wavebands and 3D deep CNNs to measure internal stem disease symptoms (Nagasubramanian et al., 2019), and employing RetinaNet and UNet architectures for root nodulation trait collection (Jubery et al., 2021; Carley et al., 2023). These technologies introduce new traits, such as nodule size and location on the root, that can be measured via high-throughput systems. At field scale, supervised machine learning techniques like LASSO enable early yield prediction and selection using UAV data, allowing screening of large populations comparable to breeders' selection (Zhou et al., 2022). By increasing the speed and automation of trait collection, these methods allow breeders and scientists to investigate topics not previously attempted. As satellite technology advances, the application of these models to larger-scale platforms becomes feasible.

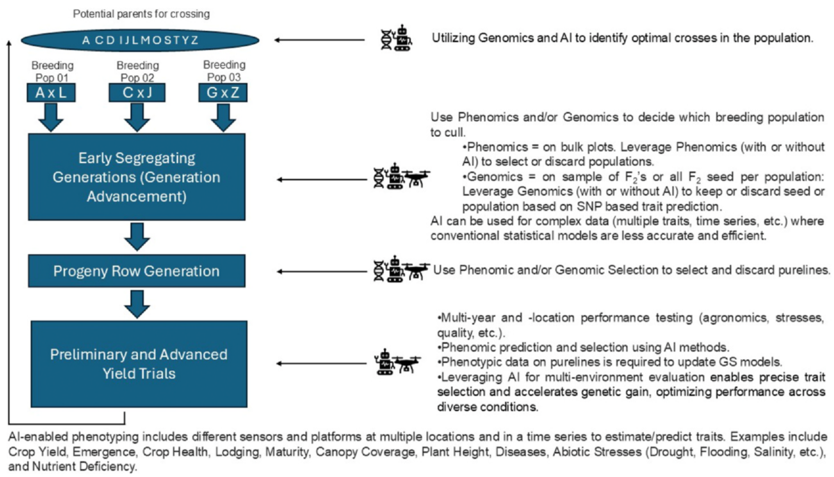

Fig. 2 Overview of soybean breeding strategy using genomics, phenomics, and artificial intelligence (AI) for multi-environment evaluation leading to the identification of a new variety. Breeding populations (e.g., Pop 01, Pop 02, Pop 03) undergo multiple generations of selection. Genomics (including Marker-assisted selection and Genomic Prediction) and Phenomics with AI are useful to optimize the cross-selection, trait estimate and prediction, and culling and selection. Marker-assisted selection and genomic selection models can be applied to enrich allele and gene frequencies over generations. AI enhances the accuracy of complex trait prediction, particularly in multi-environment trials, accelerating genetic gains and improving performance across diverse locations.

From a breeding perspective, having a fully characterized soybean phenotype is key to developing ideotypes – ideal plant types that optimize desired traits for specific environments for prescriptive breeding. The use of DL in soybean breeding is transforming the field, enabling breeders to collect previously difficult to measure traits such as root shape, length, number, mass, and angle made possible through a mobile, low-cost root phenotyping system using computer vision and ML for high-throughput analysis of root system architecture (RSA) traits (Falk et al., 2020a). Furthermore, development of shape profiles could assist breeders in understanding characteristics ideal for certain environments. This capability facilitates location-dependent prescriptive breeding, and enhances the selection of superior lines via additional traits.

The development of reliable, high-throughput methods is crucial for screening large populations for important traits. Integrating AI in phenomics offers a robust framework for image-to-trait pipelines, combining genetic data with environmental insights to enhance trait predictions and cultivate soybean varieties tailored to global agricultural demands. The use of deep learning and big data analytics not only improves the efficiency of phenomic predictions but also enhances the understanding of genotype-environment interactions, leading to more resilient and productive soybean varieties. Overall, the application of AI in phenomic prediction provides a powerful tool for advancing soybean breeding and production. The ability to analyze and interpret complex phenotypic data enables breeders to make more informed decisions, accelerating the development of high-performing soybean cultivars for target environments.

4.2 Genomics

As climate change is expected to decrease overall crop productivity and soybean demand is projected to increase (Ray et al., 2013), addressing these challenges requires developing genomics-based approaches for crop improvement, leveraging the large quantities of genomic data produced in recent years (Abberton et al., 2016). Many valuable agronomic traits are quantitative, controlled by numerous small-effect loci, complicating traditional phenotypic selection methods (Merrick et al., 2022). The release of the soybean reference genome (Schmutz et al., 2010) and development of standardized marker arrays (Song et al., 2013) have accelerated marker-assisted selection (MAS), enabling early-generation selection and reducing plot testing expenses. Genomic selection, which estimates all gene effects simultaneously (Meuwissen et al., 2001), produces genomic estimated breeding values (GEBVs) for parent selection and line advancement (Fig. 2). This approach has the potential to reduce breeding cycle length (Ma et al., 2018) and accelerate genetic gains (Voss-Fels et al., 2019), supported by methods such as speed breeding (Watson et al., 2018).

Non-linear prediction algorithms, including ML and DL, have shown promise in improving genomic selection accuracy (Crossa et al., 2017; Cuevas et al., 2016; Pérez-Rodríguez et al., 2012). These methods offer advantages in handling complex data and potentially addressing issues like epistasis effects and genomic imprinting (Varona et al., 2018). Deep learning, particularly CNNs and Deep Neural Networks (DNNs), has demonstrated the ability to identify complex multidimensional patterns in large datasets (Zou et al., 2019). CNNs have shown success in selecting high-value phenotypes from genomic data (Ma et al., 2018).

Innovative approaches like the hyperspectral wide association study (HypWAS) integrate phenomic and genomic data to identify key hyperspectral reflectance bands linked to soybean yield, offering indirect selection criteria for breeding programs (Yoosefzadeh-Najafabadi et al., 2021). DL techniques are also transforming phenotyping for disease resistance, using DL frameworks for image-based phenotyping to provide more insightful results than traditional visual methods, identifying significant SNP markers linked to sudden death syndrome (SDS) resistance (Rairdin et al., 2022). ML-based genome-wide association studies (GWAS) have unveiled novel genomic regions associated with resistance to Southern root-knot nematode (SRKN), identifying minor effect SNPs missed by traditional methods (Vieira and Chen, 2021). The G2PDeep web server exemplifies the potential of DL frameworks in genomic prediction, offering a user-friendly platform for creating, training, and deploying models for quantitative phenotype prediction (Zeng et al., 2021).

Recent studies have demonstrated the superiority of ML models like XGBoost and random forest over DL models for genotype-to-phenotype predictions using genome-wide molecular markers, significantly enhancing prediction accuracy and reducing marker inputs by up to 90 % (Gill et al., 2022a). Conversely, DL-based models such as SoyDNGP have shown remarkable precision in predicting complex traits, providing an accessible web server for trait estimation, thus enhancing breeding programs (Gao et al., 2023). In addition, CNNs have been explored for genomic selection to predict quantitative traits from single nucleotide polymorphisms (SNPs) without the need for genotype imputation, outperforming traditional statistical methods (Liu et al., 2019b). Comparative studies of various genomic prediction (GP) methods have shown that traditional models like the genomic best linear unbiased predictor (GBLUP) often outperform DL models, particularly when accounting for genotype × environmental interaction (G × E) effects are high (Ray et al., 2023). While DL algorithms can capture nonlinear patterns and integrate diverse data sources, potentially improving prediction accuracy for large breeding datasets, their superiority over conventional models in prediction power is not definitive (Montesinos-López et al., 2021). DL applications in genomic selection need high-quality, large training datasets for effective use. More importantly, the nature of molecular marker data type is not complex, and further research is needed that integrates multi-omics datastreams to compare traditional genomic prediction and DL methods.

Exploring genomic prediction models for traits with varying heritabilities helps refine the accuracy of breeding selections, optimizing the predictive performance of these models (Kaler et al., 2022). By examining different marker sets and training population sizes, researchers can improve the efficiency of selection processes in breeding programs. The development of SoyDNGP, a DL model demonstrating high predictive accuracy for complex traits across different crops, illustrates the potential of these technologies in developing customized cultivars (Gao et al., 2023). This model performs with minimal parameter tuning, highlighting significant advancements in trait prediction.

Deep learning and AI are revolutionizing soybean breeding programs by offering powerful tools for analyzing vast amounts of genomic data. The integration of AI and pHENOMICS in soybean breeding offers promising avenues for improving selection accuracy and accelerating genetic gains. These technologies accelerate the identification of superior lines, enhancing the efficiency of breeding programs through early selection of lines that are likely to perform well under specific conditions (Gao et al., 2023; Ray et al., 2023; Liu et al., 2019b). This rapid identification process is not only about speed but also precision, allowing for the early selection of lines that are predicted to yield well under specific conditions, thereby enhancing the development of tailored soybean varieties with desired characteristics.

4.3 Plant stresses

Soybean, like many crops, face stress that is in part due to a varied climate, which is expected to worsen due to climate change (Bezner Kerr et al., 2022). Abiotic and biotic stresses prevent soybean from reaching maximum yield potential and pose complex challenges for plant breeders. The economic impact of soybean diseases is substantial, with estimated average losses of $41.66 per acre in Iowa and $44.83 per acre across the United States and Ontario, Canada between 2014 and 2019 (Bradley et al., 2021). On average from 1980 to 2020, droughts caused over 7 billion dollars in damage each year to the agricultural sector, ranking third in billion dollar environmental events to impact the United States (NOAA National Centers for Environmental Information (NCEI), 2024). Management strategies exist to mitigate some of the stresses, such as irrigating fields with insufficient rainfall and applying fertilizers when soils contain insufficient levels of necessary nutrients for crop growth and development. Furthermore, pest management can include fungicide, herbicides, and insecticides applications. In 2012, herbicides were applied to 98 % of soybean acres, insecticides applied to 18 % of acres, and fungicides applied to 11 % of acres (USDA-NASS, 2013). In 2018 herbicide use expanded to 99 % of soybean acres, insecticides applied to 16 % of acres and fungicides applied to 15 % of acres (USDA-NASS, 2019). Nitrogen, Phosphorus, and Potassium application have increased across soybean acres from

2012 to 2018 by 2%, 5%, and 6% respectively (USDA-NASS, 2019). However, the mitigation practices are often costly and are becoming increasingly environmentally unsustainable (Liu et al., 2017; Good and Beatty, 2011).

The key methods for application of AI for addressing plant stress is identification, classification, quantification, and prediction (Singh et al., 2016). Soybean stress identification and quantification present significant challenges due to the difficulty in distinguishing between various sources of stress. Traditional methods rely on individuals trained in symptom and pest identification. Visual severity field ratings can also be susceptible to intra- and inter-rater variability (Akintayo et al., 2018; Singh et al., 2021a). AI and DL have emerged as popular research areas for stress identification, addressing the limitations of traditional methods. These technologies are being applied across various scales, from small-scale platforms using ground-based images to medium-scale platforms such as UAVs. Researchers have used DL to develop a model capable of identifying nine different abiotic and biotic stresses and classifying their severity levels using soybean leaf images (Ghosal et al., 2018). Further use of the dataset and model created from that study were used and led to advancements in data augmentation to improve classification accuracy (Saleem et al., 2024).

4.3.1 Abiotic stresses

Traditionally, direct selection for yield stability under multiple locations in stressed environments has been used to develop crop cultivars with stress tolerance (Singh et al., 2021c), although this process is labor and resource intensive. Indirect selection is another method used in which morphological or physiological characteristics that contribute to stress resistance are selected, generally in specialized nurseries (Singh et al., 2021c). This approach requires in-depth knowledge on how a species responds to different stressors, and what characteristics will be beneficial to the plant under stress.

Advances have been made in phenotyping for rating soybean stress responses to flooding (Zhou et al., 2021), drought (Peirone et al., 2018; Zhou et al., 2020), and iron deficiency chlorosis (Naik et al., 2017; Dobbels and Lorenz, 2019), which can be used for identifying candidate genes to be used in breeding for abiotic stress tolerance. These advances in phenotyping have largely been made possible due to advances in phenomics, ML, and DL models. For instance, Naik et al. (2017) used ML to successfully identify and classify the severity of iron deficiency chlorosis (IDC) stress symptoms using cell phone images. In another study, the levels of dicamba treatment were able to be estimated from ground based hyperspectral wavelengths using a random forest model, although the model only worked in situations were the soybean crop was still recoverable (Zhang et al., 2019a). On medium-scale platforms, UAVs have proven effective for various applications in abiotic stress research. Dobbels and Lorenz (2019) demonstrated the capability of AI to identify and classify IDC from aerial platforms using neural networks and random forests. Multispectral and thermal cameras mounted on UAVs were used to determine flood injury scores in soybeans via a feedforward neural network (FNN) model (Zhou et al., 2021). Researchers applied RGB, multispectral, and thermal cameras for assessing leaf wilting and drought responses by using a SVM model (Zhou et al., 2020). Dicamba injury ratings of tolerant and susceptible soybean varieties were successful using aerial RGB images wtih artificial neural network (ANN) and random forest models (Vieira et al., 2022). For additional abiotic stresses such as heat stress, leaf temperature measurements can provide valuable insights (Jagadish et al., 2021). UAV-mounted thermal sensors can collect data on large trials at crucial time points for heat stress, from early vegetative to reproductive stages. Deep learning models excel at processing large datasets to optimize irrigation practices (Umutoni and Samadi, 2024). These models utilize vast amounts of data, including real-time weather conditions and soil moisture levels, to accurately predict crop water needs. Weather data, including parameters such as temperature, humidity, solar radiation, and wind speed, is crucial for determining irrigation requirements. Precise measurements of these parameters, obtained through Internet of Things (IoT) based weather stations and various sensors, significantly influence water loss rates (Abdelmoamen Ahmed et al., 2021). Researchers have demonstrated that such IoT-based weather monitoring systems can analyze the crop environment in real-time, utilizing sensors for these weather variables along with soil moisture content (Pramanik et al., 2022). Real-time estimation of reference evapotranspiration (ETO) using these weather variables provides insights into water loss from plants and soil (Bounajra et al., 2024). Integrating weather forecasts into AI-driven irrigation systems enables proactive adjustments to irrigation schedules. Wireless Sensor Networks (WSNs) facilitate precision monitoring across large cropping areas, allowing real-time analysis and immediate adjustments based on predefined thresholds (Glória et al., 2021; Jamal et al., 2023). In soybean cultivation, AI/DL-driven irrigation systems optimize water usage, enhance yield, and reduce environmental impact. These systems analyze real-time data to assess water needs, significantly reducing waste through runoff and deep percolation (Sarkar et al., 2024). Machine learning algorithms, such as CNNs, coupled with IoT systems, enable targeted irrigation based on crop stress levels (Tace et al., 2022), preserving soil integrity and preventing issues like nutrient leaching and salinization leading to healthier soil ecosystem and more resilient crop growth (Goap et al., 2018). This technology offers cost savings through reduced water and energy consumption, contributed to broader environmental goals, such as lower carbon footprints and improved water resource management, supporting farm viability and aligning with global sustainability targets (Blessy and Kumar, 2021). Machine learning and DL models have advanced abiotic stress phenotyping in soybeans, enabling large-scale data collection on experimental lines. This benefits geneticists by facilitating the identification of stress-tolerance genes through large panel screenings. Soybean breeders can leverage these advancements and candidate genes to develop cultivars with enhanced abiotic stress tolerance. Additionally, these models show potential for farmer applications, potentially enabling near real-time detection and scouting of stress symptoms in fields, thus improving crop management strategies.

4.3.2 Biotic stresses

Biotic stresses encompass a wide range of organisms such as bacteria, virus, fungi, weeds, insects, and nematodes. Unfortunately, biotic stresses have a clear strength in their ability to overcome genetic sources of tolerance and management strategies, as well as their unpredictability due to the interplay with weather, the environment, and host susceptibility. Accurate identification can be complicated by confounding visual symptoms. For example, Sudden Death Syndrome (*Fusarium virguliforme*), Brown stem rot (*Cadophora gregata*), and Southern stem canker (*Diaporthe phaseolorum var. merdionalis*) produce very similar interveinal chlorosis and necrosis symptoms and often require examination of plant stem and roots to differentiate signs and symptoms for correct diagnostics (Hartman et al., 2015). The interplay of various biological organisms can also complicate management strategies. Some diseases can be spread by insects, such as the soybean dwarf virus spread by aphids or soybean vein necrosis virus spread by thrips (order Thysanoptera). Therefore, early identification and control of insect pests can prevent the potential secondary spread of viral infection (Hartman et al., 2015).

Deep learning models, have demonstrated remarkable success in leveraging high-resolution images collected from various sensing platforms, pre-processing to enhance features relevant to the desired task, such as classification and prediction. The training process involves feeding annotated datasets through multiple layers, enabling the model to learn critical feature representations in identifying defects and diseases in soybean leaves with high accuracy (LeCun et al., 2015). The deep learning approach has proven particularly effective in identifying common soybean diseases (Nagasubramanian et al., 2020). The intricacies of training DL models for soybean disease classification encompass several key aspects: optimizing hyperparameters to fine-tune model performance, employing data augmentation techniques to improve model robustness and generalization, and leveraging transfer learning to enhance performance when annotated data is limited. These methodologies, as highlighted in Ferentinos (2018), have demonstrated the efficacy of DL models in plant disease detection and diagnosis. Similarly, Rairdin et al. (2022) trained a DL model to classify and quantify sudden death syndrome in soybeans using ground-based canopy images. Nagasubramanian et al. (2019) developed a 3D CNN for classifying charcoal rot in soybean using hyperspectral imagery, showcasing the possibilities for advanced imaging technologies in field applications. Another technique, rare object detection, via a deep convolutional selective autoencoder, enabling automated counting of soybean cyst nematode eggs, a process necessary for rating levels of resistance in soybean SCN resistance screening (Akintayo et al., 2018).

DL models are now increasingly equipped to process extensive datasets of plant and pest images, effectively recognizing various disease patterns and symptoms Ghosal et al. (2018). Recent advancements have introduced more sophisticated DL architectures and training techniques that enhance the model's ability to generalize from training data to real-world conditions, significantly improving detection accuracy even under variable field conditions (Ahmad et al., 2023). High throughput platforms such as UAVs, ground robots, and insect traps equipped with advanced imaging sensors enhance these capabilities, enabling rapid identification of pest and disease types over large areas and detailed assessment of infestation severity. Significant enhancements in crop disease detection capabilities have been facilitated by the integration of these systems with DL techniques (Wiesner-Hanks et al., 2019; Bouguettaya et al., 2021).

Innovations in classification systems specifically developed for soybean diseases leverage DL to analyze images and accurately differentiate disease types (Yu et al., 2022). These systems facilitate rapid responses to disease outbreaks, potentially reducing the spread and severity of infections. Recent models incorporate techniques such as transfer learning and semi-supervised learning, allowing for effective training with limited annotated datasets, a common challenge in agricultural settings (Fang et al., 2020;

Tetila et al., 2020; Bouguettaya et al., 2021). Applying self-supervised learning methods has improved the classification of agriculturally important insects with minimal annotations, enhancing model performance under conditions of low annotation availability (Karmakar et al., 2023). Out-of-distribution detection algorithms ensure effective pest detection and classification even under varied field conditions, maintaining high accuracy and reliability across different scenarios (Saadati et al., 2024).

4.3.3 Insect, weed, and disease ICQP and management

Deep learning models have expanded rapidly in plant stress phenotyping due to their ability to handle highly dimensional data, recognize important data features, and contribute to identification, classification, quantification, and prediction of plant stress including insects, weeds and diseases (Singh et al., 2016, 2018). These models attempt to mimic the learning process of the human brain by utilizing a multi-layer neural network framework to learn more abstract, discriminative features of the data (Singh et al., 2018).

Pest control in agriculture faces significant challenges, including the ineffectiveness of manual field scouting, difficulties in disease identification, and the increasing prevalence of herbicide-resistant weeds. Between 1990 and 2015, an average of five new herbicide-resistant weed cases emerged annually (Kniss, 2018). To address these issues, robotic technology and AI offer promising solutions. Robots designed to be lightweight and navigate between variable row sizes can minimize crop disruption and soil compaction. These robots can be equipped with sensors and compact AI models for pest identification, such as the InsectNet model for insects (Chiranjeevi et al., 2023). Autonomous robots with computer vision capabilities can accurately detect and map weeds in real-time (Bawden et al., 2017), while DL-based weed detection systems for UAVs enable large-scale, high-resolution weed mapping (Sa et al., 2017). The See & Spray system is a notable example of AI-powered weed identification and targeted herbicide application (Chostner, 2017).

The implementation of AI-driven pest classification enables more informed decision-making regarding pesticide application, potentially helping to combat herbicide resistance development. In soybean, economic benefits of precision spraying technology in the field can save from 43.9% to 90.6% herbicide application resulting in average savings of $38.78/hectare (Houser et al., 2024). In soybean production, AI is transforming disease prediction and pesticide optimization. AI models can accurately forecast potential disease outbreaks by leveraging environmental factors, historical crop performance, and current crop health indicators, enabling early detection and management of soybean foliar diseases (Kashyap et al., 2022; Nayar et al., 2023). The integration of AI with Integrated Crop and Pest Management (ICPM) strategies offers comprehensive insights into crop health, pest levels, and environmental conditions, supporting informed decision-making and promoting sustainable farming practices (Miranowski, 1980; Greene et al., 1985). Real-time data collection and analysis technologies, such as UAV and IoT devices, are instrumental in implementing precision agriculture, accurately detecting affected areas and enabling precise pesticide application (Singh et al., 2021a; Balaji et al., 2023). These advancements in AI and robotics offer promising solutions for sustainable and efficient pest management in soybean cultivation and agriculture as a whole (Oberti et al., 2016).

4.3.4 Early detection of stresses

Early detection is one of the key areas of advancement in plant stress detection. Early detection is critical as it allows for timely intervention, potentially preventing widespread disease outbreaks and minimizing yield losses. This proactive approach helps prevent pest resistance and uncontrolled epidemics. Given the importance of swift and accurate stress identification in mitigation efforts, deep learning and AI-based solutions are gaining prominence due to their versatility and accuracy.

A primary method of early detection involves the utilization of wavelengths beyond the visible spectrum, including infrared and hyperspectral imaging. These wavelengths can identify physiological stress in plants that is not yet visible to the human eye, enabling the detection of diseases before symptoms appear (Lowe et al., 2017; Moghadam et al., 2017; Golhani et al., 2018; Khaled et al., 2018; Seshaiah et al., 2024). Machine learning also plays a crucial role in selecting the most informative spectral bands from highly correlated data, which is common in hyperspectral imaging. This selection process improves the efficiency and accuracy of disease detection models by reducing data dimensionality and focusing on the most relevant features. In addition to plant sensors, soil sensors represent a vital technology to assess early disease development. These sensors gather real-time data on soil conditions, which can be crucial for predicting potential disease outbreaks. A study has shown that combining soil sensor data with ML techniques allows for the efficient diagnosis of various fungal diseases, achieving prediction accuracy greater than 98% (Kumar et al., 2020). By continuously monitoring real-time agricultural information, soil sensors can help mitigate disease risks, leading to improved crop management and sustainable farming practices.

Combining HTP platforms with DL/AI methods opens new opportunities for dynamic and precise monitoring of crop diseases and pests over extensive areas. This integration allows for continuous monitoring and real-time data delivery, essential for tracking disease progression and evaluating the effectiveness of treatment strategies.

4.4 Seed yield prediction

An important and relevant application of ML in crop production is the in-season prediction of yield. Accurate estimates of crop yields help breeders make timely decisions for selecting and advancing experimental lines. Beyond plant breeders, predictions of local and regional yields have several benefits. These benefits include better planning of the use of the harvested crop, price discovery for futures contracts, price regulation, and providing farmers a baseline yield for planning input costs to increase profitability (Johnson, 2014). The United States Department of Agriculture (USDA) predicts crop yields across the United States as part of their service to agriculture (USDA-NASS, 2023). The National Agricultural Statistics Service (NASS), which is the statistical group of the USDA, conducts annual surveys throughout the growing season by contacting farmers for on-farm yield estimates, as well as sampling sections of growing fields for indicators of crop development (Johnson, 2014; USDA-NASS, 2023). However, such estimates for yield prediction are limited and time-consuming to collect, which lends the feasibility of remote sensing paired with AI methods to predict an estimated end-of-season yield. Numerous prediction factors have been considered and researched, including vegetation indices (VIs), weather and climate data, and soil properties. Additionally, the scales for prediction and platforms used have varied, with applications spanning from field level to county level and utilizing different ground and aerial-based platforms for sensing.

Along with the different prediction factors and platforms, different ML and DL models have been explored, and this exploration continues as new models are released. Through their complex architectures involving layers of neural networks, DL models can integrate various types of data, including remote sensing imagery, soil properties, and historical yield data. This integration enables a holistic approach to crop modeling that traditional methods, which often handle fewer data types and require extensive pre-processing, cannot achieve (Toledo and Crawford, 2023). In particular, LSTM networks have demonstrated remarkable capabilities in predicting soybean yield and other agronomic traits, potentially surpassing traditional ML methods. LSTM networks, a type of RNN, are particularly suited for sequential data, making them ideal for time-series predictions in agricultural forecasting. These networks can model seasonality and other temporal dynamics influencing crop development, providing more accurate yield predictions and agronomic trait analysis (Shook et al., 2021).

Vegetation indices are a popular predictor for in-season soybean yield prediction. A study by Maimaitijiang et al. (2020) collected VIs from a UAV platform. This data was combined with other UAV derived variables related to canopy structure and texture to predict the yields on three soybean genotypes in large field plots. Five different ML models were compared, with a DNN being found to have the greatest accuracy, with a R^2 of 0.72 and a RMSE of 15.9% (Maimaitijiang et al., 2020). A separate study looked at eight different soil properties as variables to predict yield on a field-scale in Canada, and compared four different models for these predictions. Random forest was found to be the most successful, with a R^2 of 0.94 for soybean, and the importance of the different soil variables in prediction was investigated (Burdett and Wellen, 2022). More commonly, yield prediction models utilize both VIs and weather data to develop more robust and accurate models. One such baseline model to predict crop yield, including soybean yield, used satellite-based Normalized Difference Vegetation Index (NDVI), surface temperature, and precipitation as input into multiple regression models to predict the county level yields (Johnson, 2014). A similar study used VIs, weather data, and maturity group in a polynomial and ridge regression model to predict the yield of specific field sites (Mourtzinis et al., 2014). Interestingly, these two studies differed in the importance of precipitation for predicting yield.

In a similar study for predicting county-level yields in Illinois and Iowa, a SVM was found to have the most accurate results and found that the variable importance rankings changed throughout the growing season (Ju et al., 2021). In the Brazilian Cerrado region, researchers used NASA-POWER weather data to compare random forest, ANNs, and SVM for yield forecasts, and found the random forest model to have the highest performance (Barbosa Dos Santos et al., 2022). This study also looked at the importance of climatic variables at different soybean phenological stages, and found that the magnitude and order of importance changes throughout the season. A comprehensive study used weather parameters, soil characteristics, and crop management to predict yields (Ansarifar et al., 2021). While several models were compared in this study, their newly proposed interaction regression model had the best accuracy and performance in predicting the yields of both soybean and corn. Additionally, they observed the additive and interaction effects of predictor variables and the temporal variations of these effects (Ansarifar et al., 2021). Ground-based collected VIs were used to train a random forest model to predict and rank the yield of numerous soybean genotypes in small plots for breeding application (Parmley et al., 2019a). In another study, multi-spectral images from the MODIS satellite were used to predict soybean yields on the county level throughout parts of the United States (You et al., 2017). In this study, they compared baseline methods of ridge regression, decision trees, and DNNs to previous CNNs and LSTM approaches and reported that DL models outperform the popular baseline methods (You et al., 2017).

In one study, researchers used historical soybean yield data from breeding trials to train an LSTM model to predict yields, and determine which weather parameters were most relevant for such predictions (Shook et al., 2021). This study also utilized pedigree-related measures, and the combination with weather parameters resulted in the LSTM model having a significantly higher prediction accuracy than SVR and LASSO. An additional benefit of the LSTM model in this study was the temporal attention mechanisms, which offer insights into critical periods during the growing season that most affect crop yield (Shook et al., 2021). Researchers in southern Brazil used MODIS satellite imagery-derived VIs, along with temperature and precipitation data, to predict municipal soybean yields (Schwalbert et al., 2020). Comparing ordinary least squares regression, random forest, and LSTM models, they found that LSTM outperformed the others at most time points. The inclusion of weather parameters improved prediction accuracy, reducing MAE, RMSE, and MSE, underscoring the importance of weather data in yield forecasting.

Deep learning models, particularly LSTM, often outperform traditional algorithms in agricultural applications, demonstrating superior accuracy and generalization across diverse environments, although it is not always superior when there is a small number of features (Kang et al., 2020). However, for certain agricultural traits, traditional ML models like XGBoost and random forest can still excel (Gill et al., 2022a). This indicates that model selection should be tailored to the specific trait and dataset under consideration.

4.5 Cyber-agricultural systems

The development of advanced sensors, platforms, AI algorithms, and tools—as discussed in the previous sections— has driven a transformation in agriculture where these technologies are integrated under the framework of cyber-physical systems (CPS). This integration enables the creation of interconnected systems that can monitor, analyze, and optimize agricultural processes in real time, leading to more efficient and sustainable practices. Cyber-physical systems (CPS) are engineered systems resulting from the continuous integration of computation and physical components. They involve a close interaction between sensors, computing devices, control and actuation systems, and networking infrastructure. In CPS, the physical space serves as the source of information, and the cyberspace uses this information to make decisions, which are then implemented back into the physical space.

Building upon this foundation, a new paradigm of Cyber-Agricultural Systems (CAS), which instantiates CPS specifically for agriculture was introduced in Sarkar et al. (2024). CAS represents a transformative approach that integrates advanced sensing, computational modeling, AI, and smart actuators to revolutionize agricultural practices (Fig. 3). The core of CAS is its integration of various technological pillars—sensing, modeling, actuation, and Internet of Things (IoT) to create a more interconnected and intelligent

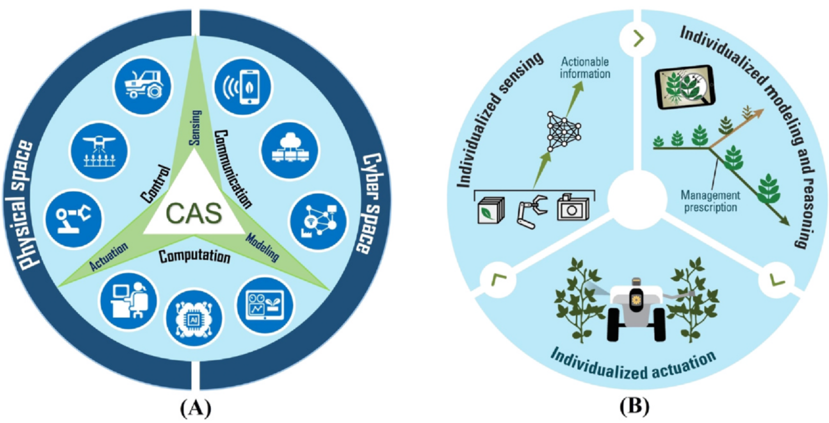

Fig. 3 (A) Cyber-physical systems (CPS) are engineered systems with deep integration between the physical and cyberspace. The three technical modules of CPS—sensing, modeling, and actuation—leverage the three functional pillars: communication, computation, and control. Cyber-agricultural systems (CAS) are built on the CPS concept. (B) The future vision of CAS—an individualized plant management paradigm that senses and models up to individual plants and organs providing unprecedented insights for decision making in breeding and production. *Figure and caption reprinted with permission from Elsevier. Original article: Sarkar et al. (2024). Cyber-agricultural systems for crop breeding and sustainable production. Trends in Plant Science 29(2): 130–149.*

agricultural environment (Sarkar et al., 2024; Sharma et al., 2020; Dumitrache et al., 2017). This integration leads to ultra-precision agriculture, enabling individualized phenotyping and actuation at finer scales.

4.5.1 Key components of CAS

The following sections outline the key components of CAS, emphasizing their roles and impact on modern agricultural practices.

Advanced Sensing: Advanced sensing technologies play a pivotal role in CAS systems by providing real-time data collection essential for informed decision-making. IoT devices, including soil moisture sensors, weather stations, and plant health monitors, are deployed across fields to continuously gather data on critical parameters such as soil moisture, temperature, humidity, and precipitation, (see Fig. 4A). This granular data enables precise monitoring and management of crop conditions, leading to optimized water usage, timely interventions for pest and disease control, and overall improved crop health (Shaikh et al., 2022). The deployment of such sensors ensures that farmers can make data-driven decisions, enhancing productivity and sustainability in soybean cultivation.

Artificial Intelligence and Machine Learning: AI and ML are at the heart of CAS, offering powerful tools for analyzing vast amounts of agricultural data. AI and ML models can identify patterns and correlations that are not easily discernible to humans. For instance, they can analyze crop information, weather forecasts, and soil conditions to predict yield (Torsoni et al., 2023). The integration of AI and ML in soybean production facilitates precision agriculture, enhancing both efficiency and output, (see Fig. 4B).

Robotics: Robotics technology revolutionizes traditional agricultural practices by automating labor-intensive tasks such as planting, harvesting, and crop monitoring. Autonomous robots equipped with AI capabilities can navigate fields, plant seeds at precise depths and intervals, and harvest crops with minimal human intervention (Mahmud et al., 2020). These robots not only increase operational efficiency but also reduce the reliance on manual labor, which is often scarce and expensive. In soybean production, robotic systems can ensure timely planting and harvesting, thereby aligning agricultural activities with optimal growing conditions and reducing crop losses, (see Fig. 4C).

Wireless Communication: Wireless communication is a critical component of CAS, enabling seamless data exchange between various devices and platforms. Technologies such as 4G/5G, Wi-Fi, and low-range (LoRa) high-bandwidth wireless connectivity ensure that data collected by

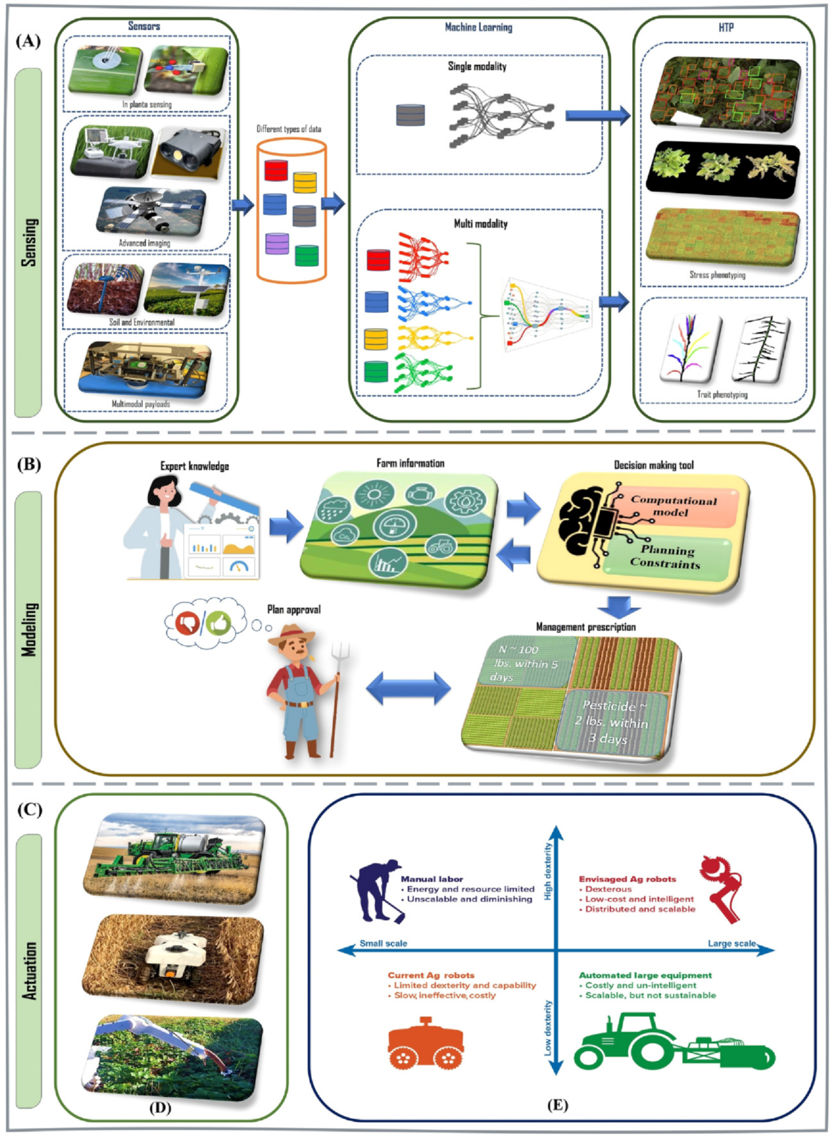

Fig. 4 (A) CAS sensing: advanced sensing technology of different modalities, leveraging heterogeneous platforms; recent advances in information processing methods, enabled by computer vision and machine learning, lead to high-throughput phenotyping (HTP) of important plant traits and stresses. (B) CAS modeling and reasoning: computational modeling at a plant to field to regional scale involving domain knowledge and data; computational models are then used to make optimal reasoning, planning, and control for agricultural decisions. (C) CAS actuation and in-field intelligence. (D) Advanced actuation such as precision spraying, autonomous scouting

IoT sensors and transmitted by robotic systems is relayed to central databases in real time. This connectivity allows for the integration of diverse data sources, facilitating comprehensive analysis and decision-making. Efficient data communication is essential for the coordinated functioning of all components within a CAS (Elijah et al., 2018; Parween et al., 2021).

Scalable Computing Infrastructure: The vast amounts of data generated by advanced sensing technologies, AI models, and robotic systems necessitate a scalable computing infrastructure capable of handling large datasets and complex computations (Mekala and Viswanathan, 2017). Cloud computing platforms and high-performance computing (HPC) systems provide the necessary computational power and storage capacity to process and analyze agricultural data effectively. These infrastructures support the local computing systems (edge devices) that may be mounted on sensors and robots for the deployment of AI models and the real-time processing of data streams, enabling swift and accurate decision-making. This scalability is crucial for adapting to the growing data demands of modern agriculture.

Security: As CAS becomes more integrated and data-driven, ensuring the security of these systems is paramount. Cyber-security measures are essential to protect sensitive agricultural data from unauthorized access and cyber threats. This includes implementing encryption protocols, secure data transmission methods, and robust authentication mechanisms. In soybean production, secure CAS can prevent data breaches that could compromise farm operations and intellectual property (Alahmadi et al., 2022). Ensuring the security of these systems is vital for maintaining trust and reliability in digital agriculture.

CAS is poised to revolutionize the agricultural sector by offering new avenues for enhancing efficiency, productivity, sustainability, and resilience. In the context of soybean cultivation, CAS can significantly improve crop management through precision farming, by providing precise, real-time data and automated responses to various agricultural challenges. One possible application of precision farming in the context of CAS

robots, and dexterous robotic arms for plant manipulation. (E) Technical challenges in CAS is to create highly dexterous robots that are scalable to large fields at low cost. *Figure and caption reprinted with permission from Elsevier. Original article: Sarkar et al. 2024. Cyber-agricultural systems for crop breeding and sustainable production. Trends in Plant Science 29(2): 130–149.*

could be pesticide application. Precision pesticide application starts with pest scouting using sensors and cameras on autonomous devices or hand-held tools to collect high-resolution data. Through efficient networks for communication, AI-driven decision tools analyze this data, and robots or drones apply chemicals only in target areas, reducing soil compaction, minimizing environmental impact, and potentially increasing yield (Shaheb et al., 2021; Frene et al., 2024). This targeted approach minimizes chemical use, reduces environmental impact, and ensures effective pest control.

In addition, CAS enables smart irrigation and water management, offering significant benefits for sustainable crop production. These advanced systems optimize water usage through precise scheduling, leading to cost savings and enhanced environmental sustainability by reducing water waste and chemical inputs (Choudhary et al., 2019). Integration of real-time weather data allows for dynamic irrigation adjustments, maintaining optimal soil moisture levels, particularly with changing climate (Campoverde and Palmieri, 2022; Nobles et al., 2022; Sacala et al., 2017). These applications leverage CAS technologies to address agricultural challenges, ensuring sustainable management.

4.5.2 Challenges and considerations in CAS

Implementing CAS presents several challenges and considerations that must be addressed to ensure their effectiveness and sustainability. One major challenge is the high initial cost of deploying advanced technologies such as IoT devices, AI models, and robotics, which can be prohibitive for small-scale farmers (Yang et al., 2023). Additionally, the integration of diverse technologies requires robust and scalable computing infrastructure, which can be difficult to maintain and upgrade. Data privacy and security are also critical concerns, as the increasing digitization of agricultural operations makes them vulnerable to cyber threats and data breaches. Ensuring interoperability among different devices and platforms is another significant consideration, as it is essential for seamless data exchange and system functionality. Furthermore, there is a need for continuous training and support for farmers to effectively utilize these advanced systems, which can be a barrier to widespread adoption. Addressing these challenges requires a collaborative approach involving technology providers, policymakers, and the agricultural community to develop cost-effective, secure, and user-friendly solutions (Yang et al., 2023).

5. Challenges and future directions for the use of AI in soybean breeding and production

5.1 Multi-modal and layered sensing

The integration of multi-modal and layered sensing technologies is a promising approach for enhancing the accuracy and robustness of data acquisition in precision agriculture. Multi-modal sensing involves the combination of multiple sensing modalities, such as RGB, multispectral, hyperspectral, thermal, and LiDAR, to capture complementary information about crop traits and field conditions (Karmakar et al., 2023). Ground-based sensors, such as proximal sensing carts and stationary sensor networks, provide high-resolution data on individual plants or small plots, while UAVs and satellites offer a broader spatial coverage and the ability to monitor large agricultural areas (Bruckstein et al., 2009). The fusion of data from multiple modalities and layers presents both opportunities and challenges. Integrating different modalities presents challenges due to the distinct statistical properties, formats, and processing requirements of each modality, complicating their unification into a single model. To overcome these challenges, techniques such as cross-modal alignment, hierarchical fusion strategies, feature concatenation, and attention mechanisms can be employed to manage this complexity and ensure robust cross-modal interactions (Xu et al., 2023). Additionally, advanced data fusion methods, such as deep learning-based approaches, can effectively integrate and analyze heterogeneous data streams (Lu et al., 2024). Moreover, the development of standardized data formats and protocols is crucial for ensuring interoperability and facilitating data sharing among researchers and stakeholders. The integration of sensing data with crop growth models and decision support systems will enable more accurate yield predictions and informed management decisions. The integration of GPS coordinates with multi-modal sensing data enhances spatial accuracy and facilitates precise georeferencing of crop traits and field conditions (Weiss et al., 2020). This spatial context is crucial for implementing site-specific management practices and for tracking temporal changes across different field locations. Recent advancements in multi-modal data integration for crop phenotyping have shown promising results. Yu et al. (2024) demonstrated the effectiveness of a novel approach combining RGB and infrared imaging for soybean canopy analysis, achieving high accuracy in segmenting soybean canopies from field images. Such innovative multi-modal approaches not only improve the accuracy of crop trait estimation but also provide a foundation for developing more comprehensive understanding of crop growth dynamics in field conditions (Zhang et al., 2019b).

5.2 Data availability and quality

The development and application of AI models in agriculture heavily depend on the availability and quality of datasets. Various online platforms, such as Mesonet, iNaturalist, and Kaggle, offer diverse datasets including weather station data and animal and insect image collections, creating numerous opportunities for AI applications. However, despite their accessibility, these datasets present several challenges that need to be addressed for effective AI model training and implementation (Sarkar et al., 2024).

Data structure is a critical consideration when evaluating datasets for AI model training. The variation in storage structures across different datasets poses a significant challenge when combining multiple sources, potentially hindering the development of comprehensive models that could benefit from diverse data inputs. To facilitate the accumulation of large datasets for model training, it is essential to develop common data storage protocols. Implementing folder structures that enable easy and fast labeling can greatly aid in preparing large image datasets for training models, which is particularly important for computer vision applications in agriculture, such as pest and disease identification.

The infrastructure for data storage and download must be robust to support the compilation of extensive datasets, such as the 13 million images used in training an insect identification model (Chiranjeevi et al., 2023). This highlights the need for scalable and efficient data management systems in agricultural AI research. Moreover, the development of standardized data formats and metadata schemas specific to agricultural data could significantly enhance interoperability and facilitate the integration of diverse datasets from multiple sources.

The quality of data used in training AI models, particularly foundational models, is crucial. High-quality datasets should capture natural variability to build more robust models that can generalize well to real-world agricultural scenarios. This includes ensuring diversity in environmental conditions, crop varieties, and stress factors represented in the datasets. Additionally, the accuracy of data labels is paramount. For instance, the insect identification model utilized data from iNaturalist, a citizen science project where domain experts verify data labels (Chiranjeevi et al., 2023). This approach of expert validation can be crucial in ensuring the reliability of training data, especially in domains where specialized knowledge is required.

Ensuring the correctness of data provided during model training can significantly enhance model accuracy, while messy or inaccurate data can lead to confusion and lower accuracy models. In the agricultural context, this could involve rigorous validation processes for field data, including cross-verification of sensor readings, standardization of measurement techniques, and careful documentation of data collection methodologies. The challenge of data quality is particularly acute in agriculture due to the variability of environmental conditions and the potential for human error in field observations.

To prepare for future applications of AI in agriculture, it is necessary to address these data-related challenges comprehensively (WorldFAIR Project, 2024; Wilkinson et al., 2016). Developing infrastructure and protocols for foundational models would allow for the creation of a few AI models that can be fine-tuned on smaller datasets. This approach could significantly advance the field by providing a solid foundation for various agricultural AI applications, from crop yield prediction to automated pest management systems. Furthermore, the agricultural sector could benefit from the development of centralized, curated data repositories specifically designed for AI applications (Swetnam et al., 2024; Hugging Face, 2024). These repositories could serve as benchmarks for model development and evaluation, ensuring that researchers and practitioners have access to high-quality, standardized datasets (Yang et al., 2024a; Arshad et al., 2024b). Collaborative efforts between academic institutions, industry partners, and government agencies could be instrumental in establishing such resources.

As the agricultural sector continues to embrace AI technologies, addressing these data-related challenges will be crucial in realizing the full potential of AI in improving agricultural practices and outcomes. This includes not only improving data collection and storage methods but also developing robust data validation techniques, creating standardized benchmarks, and fostering a culture of data sharing and collaboration within the agricultural research community.

5.3 Computational modeling

Computational modeling in agriculture employs numerical methods to predict plant growth, biomass, and yield by analyzing the interactions between crops and their environments. Crop modeling, a critical component of this field, simulates or predicts plant growth, development, and yield under various environmental conditions, helping to assess climate change impacts on agriculture (Phuoc et al., 2023). By integrating data on weather, soil properties, and crop genetics, crop models evaluate how factors such as planting density, irrigation, and fertilization affect crop performance. Several biophysical process-based models, such as ORYZA, APSIM, DSSAT, and MLCan, are widely used in crop modeling for decision-making purposes (Bouman and Van Laar, 2006; Keating et al., 2003; Jones et al., 2003; Drewry et al., 2010). Despite their utility, these models face significant challenges, including incomplete mechanistic knowledge, difficulty in measuring latent variables, and brittleness due to mismatches in the scales of input parameters. To address these limitations, hybrid approaches combining data-driven methods and process-based models are being developed.

Early data-driven efforts utilized single data modalities for crop yield prediction, disease identification, and irrigation optimization (Balakrishnan and Muthukumarasamy, 2016; Ramesh and Vardhan, 2015; Ahmad et al., 2010; Mohanty et al., 2016; Karandish and Šimůnek, 2016). The advent of IoT devices has enabled the collection of multi-modal data, enhancing decision-making processes beyond single-mode ML. Studies have shown that integrating publicly available weather and soil data can effectively predict county-level corn yield in the US Midwest (Jiang et al., 2020). Moreover, DL models have been developed to combine genotype and environmental variables for crop yield prediction, with explainable DL models providing insights into significant predictors (Shook et al., 2021; Khaki et al., 2020; Barbosa et al., 2020; Gangopadhyay et al., 2020; Akhavizadegan et al., 2021). However, purely data-driven models often fail to provide accurate outcomes beyond their training data, but integrating biophysical knowledge can mitigate this issue and reduce the need for extensive data. Recent advancements have seen the integration of high-throughput imaging and sensing data with biophysical knowledge to create flexible, hybrid AI models, such as knowledge-guided ML models for rice growth simulation (Han et al., 2023). Similarly, coupling of ML and crop modeling was shown to improve crop yield prediction in the US Corn Belt (Shahhosseini et al., 2021). While the best approaches for integrating knowledge and ML are still being refined, the field is progressing and holds great promise for future advancements in crop modeling.

5.4 AI models and tools

The integration of AI in agriculture has led to the development of various off-the-shelf tools that agricultural professionals and non-technical individuals can easily explore and utilize. These tools leverage satellite imagery, weather data, and machine learning algorithms to provide valuable insights for precision farming and crop management.

One such tool is OneSoil, a digital agriculture platform that aids in remote crop monitoring, yield increase, and optimization of seed and fertilizer costs (OneSoil, 2024). This user-friendly application demonstrates the potential of AI in making complex agricultural data accessible to a wide range of users. Another significant development in this field is Agromonitoring, which offers satellite and weather data for precision farming (Agromonitoring, 2024). This platform processes large amounts of satellite and climate data to provide vegetation indices, weather forecasts, and analytical reports. Its dashboard feature allows users to monitor field states throughout the year, integrating satellite imagery and weather data with advanced machine learning technologies. For those requiring more advanced spatial analysis capabilities, QGIS offers a comprehensive, open-source geographical information system (QGIS.org, 2024). While not exclusively an AI tool, QGIS supports various data formats and provides a framework for integrating AI-driven analyses. Its ability to handle raster, vector, mesh, and point cloud data makes it a versatile tool for agricultural applications, particularly when combined with AI models for crop monitoring and land use analysis. Google Earth Engine represents a significant leap in the accessibility of large-scale geospatial analysis (Google, 2024). This platform combines a vast catalog of satellite imagery and geospatial datasets with powerful analysis capabilities. While it requires some technical expertise to use effectively, Earth Engine's ability to detect changes, map trends, and quantify differences on the Earth's surface makes it an invaluable tool for researchers and developers working on agricultural applications. For more specialized crop intelligence, Taranis offers a platform focused on providing leaf-level insights for crop advisors and growers (Taranis, 2024). Using high-resolution drone imagery and AI-powered analysis, Taranis can detect and analyze various crop threats, including weed severity, disease, insect damage, and nutrient deficiencies. This tool demonstrates the power of combining AI with targeted data collection methods to provide actionable insights for agricultural decision-making.

These tools represent a spectrum of AI applications in agriculture, from user-friendly mobile apps to powerful analytical platforms. They showcase the potential of AI to democratize access to complex agricultural data and insights, enabling both experts and non-technical users to make more informed decisions about crop management and resource allocation. In addition to specialized agricultural tools, several general-purpose AI models with user-friendly interfaces can be applied to agricultural tasks. The Segment Anything Model (SAM), developed by Meta AI, offers powerful image segmentation capabilities that can be used for crop analysis and field mapping (Kirillov et al., 2023). For object detection, models like You Only Look Once (YOLO) can be adapted to identify various agricultural elements such as crop types, pests, or equipment in fields (Redmon et al., 2016). In the realm of image classification, ResNet variants have shown promise in detecting diseases in crops, demonstrating the potential for early identification of plant health issues (He et al., 2016). While these models may require some adaptation for specific agricultural use cases, they represent accessible entry points for users to experiment with AI applications in agriculture without extensive technical expertise.

5.5 Cyberinfrastructure

The implementation of Cyber-Agricultural systems (CAS) relies on a robust cyberinfrastructure to support essential functions such as efficient data transfer, real-time decision support, management information delivery, and storage of heterogeneous data formats from various sensors and platforms. This infrastructure facilitates querying data based on research needs, organizing trained models, providing visualization, and storing scripts and models for future use. In an analogy to infrastructure such as roads, bridges, rail lines, power grids, and telephony networks that underlie an industrial economy, cyberinfrastructure refers to the collective of advanced computing systems, data, and information management, and high-performance networks that power 21st-century science and engineering research and education. Advanced cyberinfrastructure comprises not only hardware systems but also the software that links all the components and makes the system useful and usable, as well as the human expertise that operates and helps researchers utilize the resources. Cyberinfrastructure encompasses various technological solutions tailored to support the specific needs of Cyber-Ag. Examples include data transfer solutions that facilitate efficient data transfer from fog-edge-cloud devices, decision support systems (DSS) providing real-time or scheduled decision support, data management solutions handling diverse data formats from different sensors and platforms, visualization tools offering visualization capabilities to make data insights accessible, and storage solutions organizing and storing trained models, scripts, and data for future use. While cyberinfrastructure is already proving productive in industrial settings, agriculture presents unique challenges requiring different approaches.

Agriculture's dynamic and variable environments necessitate flexible and adaptive data management solutions. Key challenges include rural connectivity, as many agricultural operations are in rural areas with limited high-speed internet access. Solutions like LoRaWAN and edge computing are needed to manage data locally. Additionally, distributed data processing is crucial for handling large volumes of data from multiple sources across edge, fog, and cloud layers to ensure timely decision-making. Agricultural devices often operate with limited power and computational resources, necessitating efficient data compression, lightweight machine-learning models, and energy-aware computing frameworks. The infrastructure must also withstand extreme weather, dust, and temperature variations. Effective communication between edge devices can enhance decision support through distributed learning, networking, and weight sharing.

Addressing these challenges involves a multi-faceted approach. Shared data storage systems facilitate efficient and secure transfer of large datasets, while high-memory computing provides access to high-memory computers and virtual machines for data analysis. Metadata labeling enables descriptive metadata for efficient data retrieval, and identity management systems ensure secure data sharing. Computational efficiency is crucial, involving learning from compressed sensor data sets and converting them into actionable insights, typically through scheduled computations performed in central or distributed units connected to the cloud.

Initiatives like the AI Institute for Resilient Agriculture (AIIRA) (Ganapathysubramanian et al., 2024) and the ICICLE AI Institute (Intelligent CyberInfrastructure with Computational Learning in the Environment) (Panda et al., 2024) are leading efforts to address these challenges. AIIRA focuses on integrating advanced technologies to enhance agricultural productivity and sustainability, while ICICLE aims to democratize AI by developing intelligent cyberinfrastructure spanning the edge-cloud-HPC computing continuum. Cybershuttle (Marru et al., 2023) is another initiative to support an end-to-end computational science research continuum, enabling seamless movement from local laptops to preprocessing, simulation, visualization, and analysis stages. This infrastructure supports scaling computational resources, captures metadata, and facilitates iterative processes.

Cyberinfrastructure is key in enabling information flow across different disciplines and platforms. For example, in soybean plant physiology, empirical response curves for stomatal conductance with environmental conditions can be used by crop modelers to simulate crop yield, with data and models shared as web resources through platforms like CyVerse (Swetnam et al., 2024) and the Open Ag Data Alliance (Ault et al., 2022). Ongoing research and development, improvements in rural connectivity, and engagement with the farming community are essential for the continued advancement and adoption of cyberinfrastructure in agriculture. Collaborative projects like INFEWS (Innovations at the Nexus of Food, Energy, and Water Systems) and PCHES (Program on Coupled Human and Earth Systems) are also making strides by developing container-based modeling infrastructures to understand and address the impacts of agricultural production on sustainable water use.

5.6 Explainability and interpretability of AI models

Explainable deep learning aims to address the "black box" nature of many AI models by providing interpretable tools that clarify why a model makes specific decisions or behaves in a certain way. This approach is crucial in agricultural applications, particularly in plant phenotyping, where understanding the model's decision-making process is essential for scientific validity and practical implementation. Techniques such as saliency maps highlight the most important pixels in imagery data, revealing spatial regions crucial for classification (Simonyan et al., 2013). This methodology has been applied in soybean stress phenotyping, where a 3D-CNN model using hyperspectral imagery simultaneously learned spectral and spatial disease signatures correlated to charcoal rot symptom severity (Nagasubramanian et al., 2019).

Explainable AI (XAI) techniques enhance the trustworthiness of image-based phenotypic information used in food production systems (Mostafa et al., 2023; Harfouche et al., 2023). Ghosal et al. (2018) demonstrated how techniques such as Grad-CAM (Selvaraju et al., 2017) can improve transparency in stress phenotyping by highlighting specific regions that are critical for the model in determining different types of stress. Additionally, the technique discussed in (Chefer et al., 2021) can be applied to explore the transparency of vision transformer models (Dosovitskiy et al., 2020). Furthermore, explainable DL output can derive stress severity scores that show high agreement with expert ratings (Ghosal et al., 2018), providing an unsupervised method for quantitative stress measurement. Zhou et al. (2024) enhanced the explainability of Support Vector Regression models used in predicting soybean branching by employing SHapley Additive exPlanations (SHAP) (Lundberg and Lee, 2017). This technique helps identify key genes influencing branching, allowing breeders to optimize their programs more effectively. Other XAI techniques, such as rule extraction and feature importance analysis, provide valuable insights into AI models' decision-making processes in agriculture (Samek et al., 2021).

Rule extraction methods, for instance, can generate human-interpretable rules describing the conditions for particular predictions, enabling agronomists to understand and validate the model's reasoning (Guidotti et al., 2018). By providing interpretable and trustworthy insights, XAI enables plant scientists and breeders to make more informed decisions. The stress severity ratings derived from explainable DL could be input as phenotypic data into genomic studies such as GWAS and Quantitative Trait Loci mapping, or automatically incorporated into genomic selection methodologies in breeding programs. This integration of explainable AI in agricultural research and practice represents a significant advancement in leveraging AI technologies for improved crop management and breeding strategies.

5.7 AI related impacts on privacy, ethics, and policy

The implementation of AI in crop production and breeding raises significant ethical concerns, particularly regarding data security, privacy, and policy issues. Large amounts of data are regularly collected on farms, providing valuable information for developing robust AI models. However, farmers often show reluctance in sharing this data due to lack of transparency (Wiseman et al., 2019). This hesitation stems from concerns about data ownership, potential misuse, and the fear of losing competitive advantages. To address these concerns, transparency regarding the intended use and outcomes of the data is essential, along with assurances that farmers will benefit from their data contributions. AI technologies in agriculture are governed by diverse regulations across different countries and regions, reflecting the global nature of both agriculture and AI development. The European Union has taken a proactive approach with its Artificial Intelligence Act, which categorizes AI applications based on risk levels and mandates transparency and human oversight for high-risk applications in agriculture (European Commission, 2024). This comprehensive framework aims to foster innovation while ensuring that AI systems in agriculture do not compromise safety or ethical standards.

In the United States, the USDA has developed a data strategy regarding transparency, security, and accountability of data usage, including applications in AI (United States Department of Agriculture, 2023). This data plan also features objectives to develop training programs to increase availability of education about ethical use of data and AI. The United States government continues to encourage regulation regarding AI with an Executive Order that has eight guiding principles that focus on ensuring AI is used safely, securely, fairly, and with the protection of Americans in mind (Executive Office of the President, 2023). These guidelines reflect a growing awareness of the potential risks associated with AI, including data breaches, algorithmic bias, and the concentration of market power. Some states have enacted specific legislation to protect against the misuse of AI and are moving to create local groups to monitor the use of AI within their respective states (State of Utah, 2024; State of Colorado, 2024). Other regions, including Australia and Canada, are focusing on confidentiality, safety, responsibility, transparency and protecting users' data rights (Australian Government, 2024; Government of Canada, 2021). These efforts highlight the global recognition of the need for standardized approaches to data management and AI deployment in agriculture, while also acknowledging the unique challenges faced by different agricultural systems worldwide.

Fairness, accountability, transparency, and equitable access are crucial principles in the ethical development of AI in agriculture. These principles aim to prevent discrimination, establish clear responsibilities for AI deployment, make AI decisions understandable to stakeholders, and ensure that all farmers can benefit from AI advancements, regardless of their size or technological expertise. The implementation of these principles requires careful consideration of how AI systems are designed, deployed, and monitored in agricultural settings. Data security and privacy are paramount concerns in the agricultural use of AI. The vast amount of data collected from farms, including sensitive information about crop yields, soil conditions, and farm management practices, necessitates robust security measures to protect against unauthorized access or breaches. Farmers need assurances that their data will be used ethically and that they retain control over it. Addressing these challenges requires comprehensive policy support, including financial subsidies, infrastructure development, and robust training programs to equip farmers with the necessary skills. Clear regulations on data privacy and ownership are also essential to protect farmers' interests. With appropriate policy interventions, smallholder farmers can leverage AI to improve their productivity and contribute to sustainable agricultural development. Ensuring the privacy of this data is crucial not only for individual farmers but also for maintaining fair competition in the agricultural sector. Policies regarding data ownership and privacy are being developed to address these concerns, with a focus on giving farmers control over their data and ensuring that they understand how it will be used.

The societal impacts of AI in agriculture extend beyond the farm, potentially affecting labor markets and changing the structure of agricultural communities. As AI technologies automate certain tasks, there are concerns about job displacement and the changing skill requirements for agricultural workers. Policy makers must consider these broader implications when developing regulations for AI in agriculture, ensuring that the benefits of these technologies are balanced against potential negative impacts on rural communities. Governments and regulatory bodies should create frameworks that encourage responsible AI adoption while balancing innovation and protection. This balance is critical to ensure that the agricultural sector can benefit from AI advancements without compromising ethical standards or exacerbating existing inequalities. Collaboration among industry stakeholders, researchers, and policymakers is essential to shape effective policies that address the complex ethical and practical challenges posed by AI in agriculture.

To foster trust and adoption of AI technologies in agriculture, ongoing communication with stakeholders, careful consideration of societal impacts, and the development of supportive policies are necessary. This includes educating farmers about the potential benefits and risks of AI technologies, involving them in the development of AI solutions, and ensuring that they have a voice in policy-making processes. Researchers must engage with policymakers to ensure that research can continue while protecting farmers' data, striking a balance between scientific advancement and ethical considerations.

By addressing these ethical, policy, data security, and privacy issues, the agricultural sector can leverage AI to enhance productivity, sustainability, and resilience in the face of growing global challenges. The holistic approach to AI implementation in agriculture aims to harness its potential while mitigating risks and ensuring widespread distribution of benefits. As AI technologies continue to evolve, ongoing evaluation and adjustment of policies and practices will be necessary to maintain ethical standards and maximize the positive impact of AI in agriculture and breeding.

5.8 Impact on smallholder farmers

Smallholder farmers' adoption of AI technologies is hindered by high initial costs, inadequate digital infrastructure, and a lack of technical knowledge. Many farmers may find the investment in AI hardware, software, and training prohibitive, while poor internet connectivity in rural areas further complicates the deployment and maintenance of these technologies (Felz et al., 2022). Moreover, the need for technical training poses a barrier, as many farmers may lack the skills required to use AI systems effectively.

Data privacy concerns also arise, as farmers need assurances that their data will be used ethically and that they retain control over it. Addressing these challenges requires comprehensive policy support, including financial subsidies, infrastructure development, and robust training programs to equip farmers with the necessary skills. Clear regulations on data privacy and ownership are also essential to protect farmers' interests. With appropriate policy interventions, smallholder farmers can leverage AI to improve their productivity and contribute to sustainable agricultural development.

5.9 Digital twin

Digital Twins (DTs) have emerged as a transformative technology in agriculture, unifying sensing, modeling, control, and actuation aspects of Cyber-Agricultural Systems (CAS) under a single framework. This integration positions DTs as a potential game-changer for CAS, offering unprecedented opportunities for precision agriculture and sustainable farming practices. While widely used in engineering systems (Madni et al., 2019; Schleich et al., 2017; Torzoni et al., 2024) and supply chain management (Ivanov and Dolgui, 2021; Ivanov, 2024), DTs are now rapidly being adopted in agriculture, spanning applications from fundamental research (Pylianidis et al., 2021) to breeding (Moghadam et al., 2020), precision agriculture (Angin et al., 2020; Alves et al., 2019; Goldenits et al., 2024), and policy-making (Delgado et al., 2019). DT in agriculture is a data and software framework that serves as a digital replica of the agricultural physical system (Jones et al., 2020). These digital replicas mirror the behavior of their real-world counterparts throughout their life cycle, from seed to harvest. These advanced models simulate the physiological state, growth, and development of plants or fields by incorporating diverse components such as historical data, crop models, AI models for phenotyping and ICQ assessment, decision-making algorithms, and field maps with 3D crop models for robotic navigation. Crop models, integral to DTs, are essential for understanding plant physiology, growth, development, and management. Unlike traditional agricultural simulators, DTs must continuously or periodically update their digital state using real-time measurements from their physical counterparts, including phenotyping, physiological measurements, and environmental data like soil, weather, and management practices. Moreover, DTs provide a structured approach to reconcile known dynamics, encoded in crop models, with unknown dynamics derived from real-world measurements. This integration is critical for biological systems where comprehensive first-principle models are not available, unlike engineered systems with fully describable behaviors. DTs have been utilized in various agricultural contexts, covering species such as row crops, orchards, viticulture, gardens, and horticulture. They can be designed and implemented at different scales, from individual organs to entire fields and greenhouses (Chaux et al., 2021; Howard et al., 2020; Kamburjan et al., 2024; Reyes Yanes et al., 2022). Recent research has applied DTs at these varied scales. Applications of DTs extend from monitoring and real-time diagnostics to optimizing yield, profitability, breeding decisions, and autonomous field operations (Laryukhin et al., 2019; Skobelev et al., 2020; Defraeye et al., 2021). Technologies like augmented reality (AR) and virtual reality (VR) can further augment DT applications, providing immersive visualization and interaction with digital agricultural systems and enhancing user experience and operational efficiency. The future of DT research lies in developing 'Intelligent Digital Twins' (IDTs), capable of self-learning and making autonomous decisions for farm management (Laryukhin et al., 2019). Machine learning techniques are being intensively studied to imbue DTs with intelligence, enabling them to adapt to dynamic environmental conditions and optimize farm operations in real-time. Preliminary studies on IDTs, primarily utilizing generative models, show promise for self-learning capabilities with varying levels of data integration (Tsialiamanis et al., 2021). The combination of ML, extensive sensing, and autonomous systems presents significant opportunities to advance agriculture through DTs.

5.10 Large soybean datasets for community usage

The plant breeding and production community leverages diverse datasets—including large images, sensor data, and multi-omics information—to advance AI-driven research and crop improvement in soybeans and other field crops. The USDA-ARS Ag Data Commons (USDA National Agricultural Library, 2024) provides a comprehensive repository of agricultural datasets, including environmental conditions, soil properties, and plant health monitoring, which are essential for studying and improving crop performance. The PlantVillage dataset (PlantVillage, 2024) includes over 50,000 images across various crop species, offering valuable resources for disease detection and health monitoring in soybeans and other crops. Similarly, the CropDeep Agricultural Dataset (Jiang, 2023) focuses on images of tomato diseases collected via IoT and mobile cameras, which provide methodologies applicable to other crops, including soybeans. The

TERRA REF initiative (TERRA-REF, 2024) offers extensive high-resolution sensor data from sorghum breeding trials, which can be utilized to study similar traits in soybean cultivation. Additionally, the Quantitative Plant project (Quantitative Plant, 2024) hosts a range of datasets, such as root system and shoot images, pivotal for phenotyping and growth analysis in crops like wheat and soybean. Resources like GrowStuff (Growstuff Team, 2024) and Open Plant Pathology (Del Ponte and Sparks, 2024) contribute further by providing open-access data and tools for crop record-keeping and disease research, enhancing our understanding and management of field crops.

SoyBase (Grant et al., 2010), the Soybean Genetics and Genomics Database, provides access to diverse genomic and phenomic datasets for soybean research. It is a comprehensive resource for soybean geneticists and breeders, offering data on genetic maps, markers, QTL (Quantitative Trait Loci) information, and sequences. SoyBase also includes various tools for visualizing and analyzing genomic data, making it an invaluable resource for identifying genes associated with important agronomic traits. An Illumina Infinium BeadChip containing over 50,000 SNPs from soybean has been developed (Song et al., 2013, 2015).

The Soybean Knowledge Base (SoyKB) (Joshi et al., 2014), SoyMD (Yang et al., 2024b), and SoyOmics (Liu et al., 2023) integrate multi-omics datasets essential for soybean research. SoyKB combines various omics data with molecular breeding information. SoyMD offers transcriptomic, proteomic, and metabolomic datasets, aiding in understanding gene expression and protein modifications. SoyOmics provides high-throughput sequencing, quantitative proteomics, and metabolite profiles, supporting soybean cultivar improvement and stress response studies. These comprehensive datasets support the development of precision agriculture technologies and improve crop productivity and sustainability. These open-source datasets are ideally positioned to assist the soybean research community develop and deploy ML- and AI-based solutions.

5.11 Immersive environments

AI and ML have revolutionized 3D modeling of real-world environments. These technologies enable quick conversion of 2D images or video into detailed 3D models, with ML algorithms improving quality and filling gaps (Arshad et al., 2024a). This advancement has significantly facilitated the creation of immersive environments across various industries, making the process faster and more accessible.

An immersive environment is a digital environment designed to fully engage and envelop the user's senses, creating a sense of presence and interaction with the environment. The environment stimulates the physical world by engaging one or more senses like sight, sound, touch, and possibly smell. This type of environment is becoming increasingly relevant in fields like education, healthcare, entertainment, marketing, manufacturing, and many more diverse research fields (Suh and Prophet, 2018). Different technologies are deployed to give a sense of presence in these environments, such as virtual reality (VR) and augmented reality (AR). Virtual Reality technology uses input devices such as head-mounted displays (HMDs) and controllers to immerse the user in a computer-generated, three-dimensional environment. VR takes place in the artificial/virtual environment, where users can generally manipulate real-life objects with the help of input devices. Augmented Reality focuses on the intersection of the real and virtual worlds, where digital information is overlaid onto the real-world environment. Unlike VR, AR does not aim to fully immerse the user in an artificial world, but instead enhances the real-world experience, providing users with additional information and control over their surroundings (Ardiny and Khanmirza, 2018).

In education, AR and VR make learning engaging across math, physics, astronomy, and biology (Ardiny and Khanmirza, 2018). A recent study has shown that VR can help students enhance their learning and observation with the visualization of complex problems, especially in subjects where visual understanding is important (Campos et al., 2022). In the tourism industry, VR technology enables tourists to explore destinations virtually and meticulously plan their visits ahead of time. Additionally, many hotels are leveraging VR to offer virtual tours of their rooms, enhancing their marketing efforts and allowing potential guests to experience accommodations before booking (Pestek and Sarvan, 2020). In agronomy, AR/VR can be crucial in applications involving sensing, reasoning, and future remote robotic manipulation. These technologies can enable researchers and (eventually) farmers to perform various experiments and operations with considerably less effort (Hurst et al., 2021). It has multiple applications that can help both small- and large-scale farmers. These technologies can provide critical training to alleviate labor shortages and improve worker skills, lowering the risk of fatalities and injuries among inexperienced workers. The deployment of VR in training has shown many advantages, such as cost-effective, safe learning environments, enabling trainees to practice various scenarios repeatedly, and ensuring proficiency in skills while minimizing exposure to real-world dangers (Xie et al., 2021). VR can also help with virtual tours that allow researchers to explore remote agricultural locations from their homes, which is helpful in times like the COVID-19 pandemic. With the use of AR prototypes that teach insect identification, this technology can also teach farmers about disease outbreaks and pest control. For example, researchers have developed an AR system that helps farmers directly distinguish between beneficial and harmful pests in their fields with mobile phones (Nigam et al., 2011). Furthermore, AR can also help farmers/researchers retain specific information on water, soil, and fertilizer requirements, thereby lowering costs and improving crop management (Isafiade and Mabiletsa, 2020).

Immersion technology in agriculture has an inspiring future ahead of it and will wholly transform farming methods. With the ongoing development of immersive technology, farmers and researchers can conduct precise and efficient crop management by observing data in real time. This includes keeping track of plant health, soil conditions, and watering requirements to maximize resource utilization and yield. Moreover, combining AR/VR, AI, and IoT will allow for smarter farming approaches, making agriculture more sustainable and resilient. These advances will address labor shortages and improve overall farm management, significantly contributing to long-term food production (de Oliveira and Corrêa, 2020). Even though immersive technology has advanced considerably in recent years, its widespread application still has challenges and drawbacks. One of the major challenges is the cost of implementing AR/VR systems; HMDs are expensive and require substantial investment. Furthermore, these technologies necessitate high-performance hardware and frequent software upgrades, which might be unaffordable for small-scale farmers and organizations. Technical limitations such as limited battery life, consistent internet connectivity, and consumer discomfort during extended use also hinder adoption. Additionally, integrating AR/VR into established agricultural practices requires training and technological skills, which may not be readily accessible in all locations.

5.12 Soybean variety development

Within breeding programs, AI has been applied in many ways in soybean including pod counting (Riera et al., 2021), disease classification (Nagasubramanian et al., 2018), root trait extraction (Carley et al., 2023; Falk et al., 2020b), abiotic stress classification (Dobbels and Lorenz, 2019; Zhou et al., 2020, 2021) and many more. For a breeder developing cultivars, these models can provide insightful information previously unavailable or difficult to obtain due to phenotyping/measurement challenges; however, several challenges remain.

Deep learning models can often be complex and difficult to interpret due to the lack of understanding in how they make predictions (McGovern et al., 2019). More basic ML models, such as decision trees, are inherently interpretable, while other models such as neural networks use feature attribution methods to make up for the lack of inherent interpretability (Paudel et al., 2023). These feature attribution methods, along with automatic feature learning capacity of DL, make DL models more interpretable to users, including plant breeders, to utilize in research and cultivar development. Model interpretability can provide useful insights which plant breeders can use for decision making. Interpretable DL models allow for an understanding of associations between the features and the outcomes (Azodi et al., 2020), which can be interpreted as a goal for guiding hypotheses. An example of an interpretable model is (Nagasubramanian et al., 2020), in which several DL methods with interpretability were used to detect and classify eight different soybean stresses. The interpretability of the model allowed for identification of infected regions on leaves, which could be used to generate hypotheses for the response mechanisms to the stresses, as well as allowing for biological interpretations. (Newman and Furbank, 2021) argues that interpretability of ML models should not be reliant on only the ranking of variable importance, and that for greatest utility in biology, models should be made understandable even at the cost of predictive accuracy. By understanding and being able to interpret the model, scientists can seek to understand the systems, rather than being limited to only predicting the system. In understanding why a prediction was made, a breeder can better select cultivars for specific production systems, and can work towards ideotype development for different environments. Essentially, a variety development program that makes hundreds of decision in the breeding pipeline can benefit from optimization and interpretability for processes and the overall system.

Fully understanding soybean is key to developing novel and custom varieties equipped for resiliency and high yield under stress and uniquely placed in their ideal environment. Predictive breeding combines environmental and management considerations for appropriate placement of novel lines (Parmley et al., 2019b). Development of ideotypes relies on ability to fully characterize the soybean genetically and phenotypically, both above and below ground. AI based techniques such as 3D canopy fingerprinting which allow for querying of related soybean canopy structure can help breeders understand how canopy architecture is related to preferential agronomic traits (Young et al., 2023). While other high throughput plant phenotyping techniques implementing AI for root ideotype characterization can illuminate parts of the plant not commonly phenotyped (Falk et al., 2020b). Furthermore, AI's integration into genetic studies of soybeans enables the development of customized crop varieties. By analyzing vast datasets on genetic markers, AI algorithms can predict plant traits that optimize yield, disease resistance, and adaptability to specific climatic conditions. Additional design strategies include genome editing techniques which can have a significant impact in the development of new soybean varieties. These technologies can aid breeders in the development of non-transgenic cultivars with traits that otherwise would be unfeasible to develop. Gene editing techniques provide human control over genetic information (Kumar and Jain, 2015), accelerating the improvement of crop traits. Combining structural, physiological, and genetic understanding of soybean under varying conditions is fundamental to digital twin development. Through AI in concert with environmental and genetic data, soybean growth and development can be simulated to predict desirable traits, allowing breeders to design customized soybean varieties and determine ideal placement. Furthermore, AI will benefit in the next generation of gene targets adn transformation as it will integrate insights from multi-omics.

Advances in genomics and phenomics are resulting in increased integration of these tools into modern plant breeding, which is allowing breeders to further increase yield and genetic gain. Additional opportunities include the application of omics technologies, which can be difficult due to the generation of large datasets which are often heterogeneous and complex to analyze, resulting in a big data problem (Harfouche et al., 2019). AI is able to assist with the problem of large datasets and streamline analyses, and is being successfully applied to various breeding objectives that were previously outlined. One growing area in breeding that AI has potential to affect is in developing novel varieties with unique combination of traits meeting the needs of prescriptive cultivars (Parmley et al., 2019b). Another avenue is the integration of soil-crop-weather parameters to develop adaptive ideotypes through digital twins. These technologies paired with AI have the potential for breeders to tailor-made soybean varieties that meet specific agronomic and climate needs. This enhances crop performance under varied environmental conditions and aligns with sustainable agricultural practices by reducing dependency on chemical inputs (Poonia et al., 2022; Wang et al., 2023).

6. Concluding remarks

The integration of AI and DL methodologies in soybean improvement and production presents a paradigm shift in addressing the escalating challenges of productivity and sustainability. These advanced computational approaches demonstrate superior capacity in processing and analyzing multidimensional, high-throughput resolution, integrating AI, developing decision support tools and informed decision making. On-board computing on machines represents a quantum leap in CAS, enabling high-fidelity detection of biotic and abiotic stressors and informing data-driven management strategies. The computational prowess of AI in assimilating and interpreting diverse datasets, encompassing genomic, meteorological, and historical agronomic information, facilitates evidence-based decision-making and resource optimization. This capability has profound implications for varietal selection, targeted pesticide application, and site-specific management protocols, potentially yielding both economic and environmental dividends in soybean production systems, and meet future production goals in the face of climate variability. However, the implementation of AI methodologies in agriculture is not without challenges. These include substantial computational resource requirements, the necessity for expansive, high-quality training datasets, and the inherent opacity of many deep learning algorithms, which complicates model interpretability and validation. Furthermore, issues pertaining to data security, privacy, and equitable access to AI technologies necessitate careful consideration to ensure ethical and unbiased implementation across diverse agricultural contexts. Future research trajectories should focus on addressing these challenges to fully harness the potential of AI in agriculture. Priority areas include enhancing model interpretability through explainable AI techniques, improving data quality and accessibility through standardized protocols, and developing scalable solutions adaptable to diverse agricultural systems. Advanced research avenues may explore multi-trait prediction models and the integration of dynamic environmental variables to enhance the robustness and applicability of AI systems in soybean breeding and production. While AI and deep learning approaches offer significant advantages in deciphering complex agricultural data and optimizing decision-making processes, their successful implementation necessitates a multidisciplinary approach addressing both technical and ethical considerations and partnerships with farmers and communication with policymakers. As these technologies continue to evolve, they hold the potential to significantly enhance the sustainability, efficiency, and resilience of soybean production systems, contributing to global food security in the face of increasing environmental volatility and resource constraints.

Acknowledgments

AI Institute for Resilient Agriculture (USDA-NIFA #2021-647021-35329), COALESCE: COntext Aware LEarning for Sustainable CybEr-Agricultural Systems (CPS Frontier #1954556), Smart Integrated Farm Network for Rural Agricultural Communities (SIRAC) (NSF S & CC #1952045), Iowa Soybean Association, USDA CRIS project IOW04714, R.F. Baker Center for Plant Breeding, G.F. Sprague Chair in Agronomy, and Plant Sciences Institute.

CRediT authorship contribution statement

A.K. Singh: Conceptualization, Writing and Editing. **S.E. Jones**: Writing and Editing. **L.V. Laan**: Writing and Editing. **T.T. Ayanlade**: Writing and Editing. **J. Raigne**: Writing and Editing. **N. Saleem**: Writing and Editing. **S. Joshi**: Writing and Editing. **M.A. Arshad**: Writing and Editing. **H. ZareMehrjerdi**: Writing and Editing. **A. Rairdin**: Writing and Editing. **J.D. Salvo**: Writing and Editing. **D. Elango**: Writing and Editing. **L.A. Peixoto**: Writing and Editing. **T.Z. Jubery**: Writing and Editing. **A. Krishnamurthy**: Writing and Editing. **A. Singh**: Writing and Editing. **S. Sarkar**: Writing and Editing. **B. Ganapathysubramanian**: Writing and Editing.

Conflicts of interest

No conflict of interest.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/bs.agron.2024.11.003.

References

Abberton, M., Batley, J., Bentley, A., Bryant, J., Cai, H., Cockram, J., Costa de Oliveira, A., Cseke, L.J., Dempewolf, H., De Pace, C., Edwards, D., Gepts, P., Greenland, A., Hall, A., Henry, R., Hori, K., Howe, G.T., Hughes, S., Humphreys, M., Lightfoot, D., Marshall, A., Mayes, S., Nguyen, H.T., Ogbonnaya, F.C., Ortiz, R., Paterson, A.H., Tuberosa, R., Valliyodan, B., Varshney, R.K., Yano, M., 2016. Global agricultural intensification during Climate Change: a role for genomics. Plant Biotechnology Journal 14, 1095–1098. https://doi.org/10.1111/pbi.12467.

Abdelmoamen Ahmed, A., AlOmari, S., Awal, R., Fares, A., Chouikha, M., 2021. A distributed system for supporting smart irrigation using internet of things technology. Engineering Reports 3, e12352. https://doi.org/10.1002/eng2.12352.

Agromonitoring, 2024. Agromonitoring: Satellite and weather data for precision farming in fast and easy ways. https://agromonitoring.com. Accessed: 2024-08-05.

Ahmad, A., Gamal, A.E., Saraswat, D., 2023. Toward generalization of deep learning-based plant disease identification under controlled and field conditions. Institute of Electrical and Electronics Engineers Access 11, 9042–9057. https://doi.org/10.1109/ACCESS.2023.3240100.

Ahmad, S., Kalra, A., Stephen, H., 2010. Estimating soil moisture using remote sensing data: a machine learning approach. Advances in Water Resources 33, 69–80. https://doi.org/10.1016/j.advwatres.2009.10.008.

Akhavizadegan, F., Ansarifar, J., Wang, L., Huber, I., Archontoulis, S.V., 2021. A time-dependent parameter estimation framework for crop modeling. Scientific Reports 11, 11437. https://doi.org/10.1038/s41598-021-90835-x.

Akintayo, A., Tylka, G.L., Singh, A.K., Ganapathysubramanian, B., Singh, A., Sarkar, S., 2018. A deep learning framework to discern and count microscopic nematode eggs. Scientific Reports 8, 9145–9211. https://doi.org/10.1038/s41598-018-27272-w.

Alahmadi, A.N., Rehman, S.U., Alhazmi, H.S., Glynn, D.G., Shoaib, H., Solé, P., 2022. Cyber-security threats and side-channel attacks for digital agriculture. Sensors 22, 3520. https://doi.org/10.3390/s22093520.

Alison, T., Alex, C.M., Henry, S., Bryce, T., Charles, M., Talon, M.D.J., 2018. Professor: a motorized field-based phenotyping cart. HardwareX 4, e00025. https://doi.org/10.1016/j.ohx.2018.e00025.

Allen, T., Mueller, D., Sisson, A., 2023. Soybean disease loss estimates from the United States and Ontario, Canada—2022. Crop Protection Network 1018. https://doi.org/10.31274/cpn-20230421-1. https://cropprotectionnetwork.org/publications/soybean-disease-loss-estimates-from-the-united-states-and-ontario-canada-2022.

Alves, R.G., Souza, G., Maia, R.F., Tran, A.L.H., Kamienski, C., Soininen, J.P., Aquino, P.T., Lima, F., 2019. A digital twin for smart farming. In: 2019 IEEE Global Humanitarian Technology Conference (GHTC), IEEE, pp. 1–4. https://doi.org/10.1016/j.agsy.2020.103046.

Angin, P., Anisi, M.H., Göksel, F., Gürsoy, C., Büyükgülcü, A., 2020. AgriLoRa: a digital twin framework for smart agriculture. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications 11, 77–96. https://doi.org/10.22667/JOWUA.2020.12.31.077.

Ansarifar, J., Wang, L., Archontoulis, S.V., 2021. An interaction regression model for crop yield prediction. Scientific Reports 11, 17754. https://doi.org/10.1038/s41598-021-97221-7.

Araus, J.L., Cairns, J.E., 2014. Field high-throughput phenotyping: the new crop breeding frontier. Trends in Plant Science 19, 52–61. https://doi.org/10.1016/j.tplants.2013.09.008.

Ardiny, H., Khanmirza, E., 2018. The role of AR and VR technologies in education developments: opportunities and challenges. In: 2018 6th RSI International Conference on Robotics and Mechatronics (ICRoM), IEEE, pp. 482–487. https://doi.org/10.1109/ICRoM.2018.8657615.

Arnal Barbedo, J.G., 2013. Digital image processing techniques for detecting, quantifying and classifying plant diseases. SpringerPlus 2, 660. https://doi.org/10.1186/2193-1801-2-660.

Arshad, M.A., Jubery, T., Afful, J., Jignasu, A., Balu, A., Ganapathysubramanian, B., Sarkar, S., Krishnamurthy, A., 2024a. Evaluating neural radiance fields (NeRFs) for 3D plant geometry reconstruction in field conditions. Plant Phenomics. https://doi.org/10.34133/plantphenomics.0235.

Arshad, M.A., Jubery, T.Z., Roy, T., Nassiri, R., Singh, A.K., Singh, A., Hegde, C., Ganapathysubramanian, B., Balu, A., Krishnamurthy, A., Sarkar, S., 2024b. Ageval: A benchmark for zero-shot and few-shot plant stress phenotyping with multimodal llms. arXiv. https://doi.org/10.48550/arXiv.2407.19617.

Ault, A., Palacios, S., Evans, J., 2022. Agriculture data sharing: conceptual tools in the technical toolbox and implementation in the open ag data alliance framework. Agronomy Journal 114, 2681–2692. https://doi.org/10.1002/agj2.21007.

Australian Government, 2024. Policy for the responsible use of AI in government.

Azodi, C.B., Tang, J., Shiu, S.H., 2020. Opening the black box: Interpretable machine learning for geneticists. Trends in Genetics 36, 442–455. https://doi.org/10.1016/j.tig.2020.03.005.
Bai, H., Purcell, L.C., 2018. Aerial canopy temperature differences between fast- and slow-wilting soya bean genotypes. Journal of Agronomy and Crop Science (1986) 204, 243–251. https://doi.org/10.1111/jac.12259.
Balaji, P., Devasena, D., Kayalvizhi, S., Aarthikannan, N., 2023. IoT-empowered precision agricultural multi-rotor drones: a revolutionary approach for sustainable farming. In: 2023 International Conference on Recent Advances in Science and Engineering Technology (ICRASET), pp. 1–5. https://doi.org/10.1109/ICRASET59632.2023.10419911.
Balakrishnan, N., Muthukumarasamy, G., 2016. Crop production-ensemble machine learning model for prediction. International Journal of Computer Science and Software Engineering 5, 148.
Barbosa, A., Trevisan, R., Hovakimyan, N., Martin, N.F., 2020. Modeling yield response to crop management using convolutional neural networks. Computers and Electronics in Agriculture 170, 105197. https://doi.org/10.1016/j.compag.2019.105197.
Barbosa Dos Santos, V., Moreno Ferreira Dos Santos, A., da Silva Cabral de Moraes, J.R., de Oliveira Vieira, I.C., de Souza Rolim, G., 2022. Machine learning algorithms for soybean yield forecasting in the brazilian cerrado. Journal of the Science of Food and Agriculture 102, 3665–3672. https://doi.org/10.1002/jsfa.11713.
Bawden, O., Kulk, J., Russell, R., McCool, C., English, A., Dayoub, F., Lehnert, C., Perez, T., 2017. Robot for weed species plant-specific management. Journal of Field Robotics 34, 1179–1199. https://doi.org/10.1002/rob.21727.
Bazzer, S.K., Purcell, L.C., 2020. Identification of quantitative trait loci associated with canopy temperature in soybean. Scientific Reports 10, 17604. https://doi.org/10.1038/s41598-020-74614-8.
Bezner Kerr, R., Hasegawa, T., Lasco, R., Bhatt, I., Deryng, D., Farrell, A., Gurney-Smith, H., Ju, H., Lluch-Cota, S., Meza, F., Nelson, G., Neufeldt, H., Thornton, P., 2022. Food, fibre and other ecosystem products. In: Pörtner, H.O., Roberts, D.C., Tignor, M., Poloczanska, E.S., Mintenbeck, K., Alegría, A., Craig, M., Langsdorf, S., Löschke, S., Möller, V., Okem, A., Rama, B. (Eds.), Climate Change 2022: Impacts, adaptation and vulnerability. Contribution of working group II to the sixth assessment report of the intergovernmental panel on climate change. Cambridge University Press, pp. 713–906. https://doi.org/10.1017/9781009325844.007.
Bi, L., Wally, O., Hu, G., Tenuta, A.U., Kandel, Y.R., Mueller, D.S., 2023. A transformer-based approach for early prediction of soybean yield using time-series images. Frontiers in Plant Science 14, 1173036. https://doi.org/10.3389/fpls.2023.1173036.
Blessy, J.A., Kumar, A., 2021. Smart irrigation system techniques using artificial intelligence and IoT. In: 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Institute of Electrical and Electronics Engineers, Tirunelveli, India, pp. 1355–1359. https://doi.org/10.1109/ICICV50876.2021.9388444.
Bouguettaya, A., Zarzour, H., Kechida, A., Mohammed Taberkit, A., 2021. Recent advances on UAV and deep learning for early crop diseases identification: A short review. In: 2021 International Conference on Information Technology (ICIT), Institute of Electrical and Electronics Engineers, Amman, Jordan. pp. 334–339. https://doi.org/10.1109/ICIT52682.2021.9491661.
Bouman, B., Van Laar, H., 2006. Description and evaluation of the rice growth model ORYZA2000 under nitrogen-limited conditions. Agricultural Systems 87, 249–273. https://doi.org/10.1016/j.agsy.2004.09.011.
Bounajra, A., ElGuemmat, K., Mansouri, K., Akef, F., 2024. Towards efficient irrigation management at field scale using new technologies: a systematic literature review. Agricultural Water Management 295, 108758. https://doi.org/10.1016/j.agwat.2024.108758.

Bradley, C.A., Allen, T.W., Sisson, A.J., Bergstrom, G.C., Bissonnette, K.M., Bond, J., Byamukama, E., Chilvers, M.I., Collins, A.A., Damicone, J.P., Dorrance, A.E., Dufault, N.S., Esker, P.D., Faske, T.R., Fiorellino, N.M., Giesler, L.J., Hartman, G.L., Hollier, C.A., Isakeit, T., Jackson-Ziems, T.A., Jardine, D.J., Kelly, H.M., Kemerait, R.C., Kleczewski, N.M., Koehler, A.M., Kratochvil, R.J., Kurle, J.E., Malvick, D.K., Markell, S.G., Mathew, F.M., Mehl, H.L., Mehl, K.M., Mueller, D.S., Mueller, J.D., Nelson, B.D., Overstreet, C., Padgett, G.B., Price, P.P., Sikora, E.J., Small, I., Smith, D.L., Spurlock, T.N., Tande, C.A., Telenko, D.E.P., Tenuta, A.U., Thiessen, L.D., Warner, F., Wiebold, W.J., Wise, K.A., 2021. Soybean yield loss estimates due to diseases in the United States and Ontario, Canada, from 2015 to 2019. Plant Health Progress 22, 483–495. https://doi.org/10.1094/PHP-01-21-0013-RS.

Bruckstein, A.M., Donoho, D.L., Elad, M., 2009. From sparse solutions of systems of equations to sparse modeling of signals and images. SIAM Review 51, 34–81. https://doi.org/10.1137/060657704.

Burdett, H., Wellen, C., 2022. Statistical and machine learning methods for crop yield prediction in the context of precision agriculture. Precision Agriculture 23, 1553–1574. https://doi.org/10.1007/s11119-022-09897-0.

Campos, E., Hidrogo, I., Zavala, G., 2022. Impact of virtual reality use on the teaching and learning of vectors. Frontiers in Education 7. https://doi.org/10.3389/feduc.2022.965640.

Campoverde, L.M.S., Palmieri, N., 2022. A reinforcement learning approach for smart irrigation systems. Autonomous air and ground sensing systems for agricultural optimization and phenotyping VII. SPIE, pp. 40–48. https://doi.org/10.1117/12.2623059.

Carley, C.N., Zubrod, M.J., Dutta, S., Singh, A.K., 2023. Using machine learning enabled phenotyping to characterize nodulation in three early vegetative stages in soybean. Crop Science 63, 204–226. https://doi.org/10.1002/csc2.20861.

Chaux, J.D., Sanchez-Londono, D., Barbieri, G., 2021. A digital twin architecture to optimize productivity within controlled environment agriculture. Applied Sciences 11, 8875. https://doi.org/10.3390/app11198875.

Chefer, H., Gur, S., Wolf, L., 2021. Generic attention-model explainability for interpreting bi-modal and encoder-decoder transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 397–406. https://doi.org/10.1109/ICCV48922.2021.00045.

Chen, Z., Wang, J., Jin, J., 2023. Fully automated proximal hyperspectral imaging system for high-resolution and high-quality in vivo soybean phenotyping. Precision Agriculture 24, 2395–2415. https://doi.org/10.1007/s11119-023-10045-5.

Chiranjeevi, S., Sadaati, M., Deng, Z.K., Koushik, J., Jubery, T.Z., Mueller, D., Neal, M. E., Merchant, N., Singh, A., Singh, A.K., Sarkar, S., Singh, A., Ganapathysubramanian, B., 2023. Deep learning powered real-time identification of insects using citizen science data. arXiv. https://doi.org/10.48550/arXiv.2306.02507.

Chostner, B., 2017. See & spray: the next generation of weed control. Resource Magazine 24, 4–5.

Choudhary, S., Gaurav, V., Department of Computer Science, Sir M. Visvesvaraya Institute of Technology, Bengaluru, India,Singh, A., Department of Computer Science, Sir M. Visvesvaraya Institute of Technology, Bengaluru, India,Agarwal, S., 2019. Autonomous Crop Irrigation System using Artificial Intelligence. International Journal of Engineering and Advanced Technology 8, 46–51. https://doi.org/10.35940/ijeat.E1010.0585S19.

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., De Los Campos, G., Burgueño, J., González-Camacho, J.M., Pérez-Elizalde, S., Beyene, Y., Dreisigacker, S., Singh, R., Zhang, X., Gowda, M., Roorkiwal, M., Rutkoski, J., Varshney, R.K., 2017. Genomic selection in plant breeding: methods, models, and perspectives. Trends in Plant Science 22, 961–975. https://doi.org/10.1016/j.tplants.2017.08.011.

Crusiol, L.G.T., Nanni, M.R., Furlanetto, R.H., Sibaldelli, R.N.R., Cezar, E., Mertz-Henning, L.M., Nepomuceno, A.L., Neumaier, N., Farias, J.R.B., 2020. UAV-based thermal imaging in the assessment of water status of soybean plants. International Journal of Remote Sensing 41, 3243–3265. https://doi.org/10.1080/01431161.2019.1673914.

Cuevas, J., Crossa, J., Soberanis, V., Pérez-Elizalde, S., Pérez-Rodríguez, P., Campos, G.d.l., Montesinos-López, O., Burgueño, J., 2016. Genomic prediction of genotype × environment interaction kernel regression models. The Plant Genome 9, 1–20. https://doi.org/10.3835/plantgenome2016.03.0024.

de Oliveira, M. E., Corrêa, C.G., 2020. Virtual reality and augmented reality applications in agriculture: a literature review. In: 2020 22nd symposium on virtual and augmented reality (svr), IEEE, Porto de Galinhas, Brazil, pp. 1–9. https://doi.org/10.1109/SVR51698.2020.00017.

Defraeye, T., Shrivastava, C., Berry, T., Verboven, P., Onwude, D., Schudel, S., Bühlmann, A., Cronje, P., Rossi, R.M., 2021. Digital twins are coming: Will we need them in supply chains of fresh horticultural produce? Trends in Food Science & Technology 109, 245–258. https://doi.org/10.1016/j.tifs.2021.01.025.

Del Ponte, E., Sparks, A., 2024. Open Plant Pathology. https://openplantpathology.org/. Accessed: 2024-08-30.

Delgado, J.A., Short Jr, N.M., Roberts, D.P., Vandenberg, B., 2019. Big data analysis for sustainable agriculture on a geospatial cloud framework. Frontiers in Sustainable Food Systems 3, 54. https://doi.org/10.3389/fsufs.2019.00054.

Dice, J., Rodziewicz, D., 2020. Drought risk to the agriculture sector. Economic Review 105, 61–86. https://www.kansascityfed.org/Economic.

Dobbels, A.A., Lorenz, A.J., 2019. Soybean Iron deficiency chlorosis high throughput phenotyping using an unmanned aircraft system. Plant Methods 15, 97. https://doi.org/10.1186/s13007-019-0478-9.

dos Santos Ferreira, A., Matte Freitas, D., Gonçalvesda Silva, G., Pistori, H., Theophilo Folhes, M., 2017. Weed detection in soybean crops using ConvNets. Computers and Electronics in Agriculture 143, 314–324. https://doi.org/10.1016/j.compag.2017.10.027.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv. https://doi.org/10.48550/arXiv.2010.11929.

Drewry, D., Kumar, P., Long, S., Bernacchi, C., Liang, X.Z., Sivapalan, M., 2010. Ecohydrological responses of dense canopies to environmental variability: 1. Interplay between vertical structure and photosynthetic pathway. Journal of Geophysical Research: Biogeosciences 115. https://doi.org/10.1029/2010JG001340.

Dumitrache, I., Caramihai, S.I., Sacala, I.S., Moisescu, M.A., 2017. A Cyber Physical Systems Approach for Agricultural Enterprise and Sustainable Agriculture. In: 2017 21st International Conference on Control Systems and Computer Science (CSCS), Institute of Electrical and Electronics Engineers, Bucharest, Romania, pp. 477–484. https://doi.org/10.1109/CSCS.2017.74.

EarthSense, 2021. The Terra Family of AI-powered Robots. https://www.earthsense.co/. Accessed: 2024-08-21.

Elijah, O., Rahman, T.A., Orikumhi, I., Leow, C.Y., Hindia, M.N., 2018. An overview of internet of things (IoT) and data analytics in agriculture: benefits and challenges. Institute of Electrical and Electronics Engineers Internet of Things Journal 5, 3758–3773. https://doi.org/10.1109/JIOT.2018.2844296.

European Commission, 2024. Regulation of the European Parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act), com (2021) 0206. http://data.europa.eu/eli/reg/2024/1689/oj. Accessed: 2024-08-21.

Executive Office of the President, 2023. Executive order no. 14110, 3c.f.r. 75191. https://www.govinfo.gov/content/pkg/FR-2023–11-01/pdf/2023–24283.pdf. Accessed: 2024-08-21.
Falk, K.G., Jubery, T.Z., Mirnezami, S.V., Parmley, K.A., Sarkar, S., Singh, A., Ganapathysubramanian, B., Singh, A.K., 2020a. Computer vision and machine learning enabled soybean root phenotyping pipeline. Plant Methods 16, 5. https://doi.org/10.1186/s13007-019-0550-5.
Falk, K.G., Jubery, T.Z., O'Rourke, J.A., Singh, A., Sarkar, S., Ganapathysubramanian, B., Singh, A.K., 2020b. Soybean root system architecture trait study through genotypic, phenotypic, and shape-based clusters. Plant Phenomics 2020, 1925495. https://doi.org/10.34133/2020/1925495.
Fang, W., Yue, L., Dandan, C., 2020. Classification system study of soybean leaf disease based on deep learning. In: 2020 International Conference on Internet of Things and Intelligent Applications (ITIA), Institute of Electrical and Electronics Engineers, Zhenjiang, China, pp. 1–5. https://doi.org/10.1109/ITIA50152.2020.9312252.
Farm-ng, 2024. Introducing the amiga. https://farm-ng.com/. Accessed: 2024-08-21.
Felz, D., Keifer Peretti, K., Austin, A., 2022. Privacy, cyber & data strategy advisory: AI regulation in the U.S.: What's coming, and what companies need to do in 2023. https://www.alston.com/en/insights/publications/2022/12/ai-regulation-in-the-us.
Feng, L., Chen, S., Zhang, C., Zhang, Y., He, Y., 2021. A comprehensive review on recent applications of unmanned aerial vehicle remote sensing with various sensors for high-throughput plant phenotyping. Computers and Electronics in Agriculture 182, 106033. https://doi.org/10.1016/j.compag.2021.106033.
Ferentinos, K.P., 2018. Deep learning models for plant disease detection and diagnosis. Computers and Electronics in Agriculture 145, 311–318. https://doi.org/10.1016/j.compag.2018.01.009.
Feuer, B., Joshi, A., Cho, M., Chiranjeevi, S., Deng, Z.K., Balu, A., Singh, A.K., Sarkar, S., Merchant, N., Singh, A., Ganapathysubramanian, B., Hegde, C., 2024. Zero-shot insect detection via weak language supervision. Plant Phenome Journal 7. https://doi.org/10.1002/ppj2.20107.
Frene, J.P., Pandey, B.K., Castrillo, G., 2024. Under pressure: elucidating soil compaction and its effect on soil functions. Plant and Soil. https://doi.org/10.1007/s11104-024-06573-2.
Furbank, R.T., Tester, M., 2011. Phenomics – technologies to relieve the phenotyping bottleneck. Trends in Plant Science 16, 635–644. https://doi.org/10.1016/j.tplants.2011.09.005.
Gagné, R.J., Yukawa, J., Elsayed, A.K., McMechan, A.J., 2019. A new pest species of *Resseliella* (Diptera: Cecidomyiidae) on soybean (Fabaceae) in North America, with a description of the genus. Proceedings of the Entomological Society of Washington 121, 168–177. https://doi.org/10.4289/0013-8797.121.2.168.
Ganapathysubramanian, B., Bell, J.M.P., Kantor, G., Merchant, N., Sarkar, S., Schnable, P.S., Segovia, M., Singh, A., Singh, A.K., 2024. Aiira: AI institute for resilient agriculture. AI Magazine 45, 94–98. https://doi.org/10.1002/aaai.12151.
Gangopadhyay, T., Shook, J., Singh, A.K., Sarkar, S., 2020. Interpreting the impact of weather on crop yield using attention. In: NeurIPS Workshop on AI for Earth Sciences, pp. 1–6.
Gao, P., Zhao, H., Luo, Z., Lin, Y., Feng, W., Li, Y., Kong, F., Li, X., Fang, C., Wang, X., 2023. SoyDNGP: a web-accessible deep learning framework for genomic prediction in soybean breeding. bbad349. Briefings in Bioinformatics 24. https://doi.org/10.1093/bib/bbad349.
Georgopoulos, A., Ioannidis, C., Valanis, A., 2010. Assessing the performance of a structured light scanner. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 38, 251–255.

Ghosal, S., Blystone, D., Singh, A.K., Ganapathysubramanian, B., Singh, A., Sarkar, S., 2018. An explainable deep machine vision framework for plant stress phenotyping. Proceedings of the National Academy of Sciences 115, 4613–4618. https://doi.org/10.1073/pnas.1716999115.

Gill, M., Anderson, R., Hu, H., Bennamoun, M., Petereit, J., Valliyodan, B., Nguyen, H.T., Batley, J., Bayer, P.E., Edwards, D., 2022a. Machine learning models outperform deep learning models, provide interpretation and facilitate feature selection for soybean trait prediction. BMC Plant Biology 22, 180. https://doi.org/10.1186/s12870-022-03559-z.

Gill, T., Gill, S.K., Saini, D.K., Chopra, Y., de Koff, J.P., Sandhu, K.S., 2022b. A comprehensive review of high throughput phenotyping and machine learning for plant stress phenotyping. Phenomics 2, 156–183. https://doi.org/10.1007/s43657-022-00048-z.

Gilliot, J., Vaudour, E., Michelin, J., 2017. Soil surface roughness measurement: a new fully automatic photogrammetric approach applied to agricultural bare fields. Computers and Electronics in Agriculture 134, 63–78. https://doi.org/10.1016/j.compag.2017.01.010.

Glória, A., Cardoso, J., Sebastião, P., 2021. Sustainable irrigation system for farming supported by machine learning and real-time sensor data. Sensors 21, 3079. https://doi.org/10.3390/s21093079.

Goap, A., Sharma, D., Shukla, A., RamaKrishna, C., 2018. An IoT based smart irrigation management system using machine learning and open source technologies. Computers and Electronics in Agriculture 155, 41–49. https://doi.org/10.1016/j.compag.2018.09.040.

Goldenits, G., Mallinger, K., Raubitzek, S., Neubauer, T., 2024. Current applications and potential future directions of reinforcement learning-based digital twins in agriculture. Smart Agricultural Technology 8, 100512. https://doi.org/10.1016/j.atech.2024.100512.

Golhani, K., Balasundram, S.K., Vadamalai, G., Pradhan, B., 2018. A review of neural networks in plant disease detection using hyperspectral data. Information Processing in Agriculture 5, 354–371. https://doi.org/10.1016/j.inpa.2018.05.002.

Good, A.G., Beatty, P.H., 2011. Fertilizing nature: a tragedy of excess in the commons. PLoS Biology 9, e1001124. https://doi.org/10.1371/journal.pbio.1001124.

Google, 2024. Google Earth Engine. https://earthengine.google.com/. Accessed: 2024-08-04.

Government of Canada, 2021. Bill c-27: An act to enact the consumer privacy protection act, the personal information and data protection tribunal act, and the artificial intelligence and data act. https://www.parl.ca/DocumentViewer/en/44–1/bill/C-27/first-reading. Accessed: 2024-08-21.

Graham, P.H., Vance, C.P., 2003. Legumes: Importance and constraints to greater use. Plant Physiology 131, 872–877. https://doi.org/10.1104/017004.

Grant, D., Nelson, R.T., Cannon, S.B., Shoemaker, R.C., 2010. SoyBase, the USDA-ARS soybean genetics and genomics database. Nucleic Acids Research 38, D843–D846. https://doi.org/10.1093/nar/gkp798.

Greene, C.R., Kramer, R.A., Norton, G.W., Rajotte, E.G., Robert, M., 1985. An economic analysis of soybean integrated pest management. American Journal of Agricultural Economics 67, 567–572. https://doi.org/10.2307/1241077.

Growstuff Team, 2024. Growstuff – An open gardening platform. https://github.com/Growstuff/growstuff. Accessed: 2024-08-30.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., Pedreschi, D., 2018. A survey of methods for explaining black box models. ACM Computing Surveys (CSUR) 51, 1–42.

Guo, W., Carroll, M.E., Singh, A., Swetnam, T.L., Merchant, N., Sarkar, S., Singh, A.K., Ganapathysubramanian, B., 2021. UAS-based plant phenotyping for research and breeding applications. Plant Phenomics 2021, 9840192. https://doi.org/10.34133/2021/9840192.

Hämmerle, M., Höfle, B., 2016. Direct derivation of maize plant and crop height from low-cost time-of-flight camera measurements. Plant Methods 12, 1–13. https://doi.org/10.1186/s13007-016-0150-6.
Hamuda, E., Glavin, M., Jones, E., 2016. A survey of image processing techniques for plant extraction and segmentation in the field. Computers and Electronics in Agriculture 125, 184–199. https://doi.org/10.1016/j.compag.2016.04.024.
Han, J., Shi, L., Pylianidis, C., Yang, Q., Athanasiadis, I.N., 2023. DeepOryza: A knowledge guided machine learning model for rice growth simulation. In: 2nd AAAI Workshop on AI for Agriculture and Food Systems, pp. 1–6.
Harandi, N., Vandenberghe, B., Vankerschaver, J., Depuydt, S., Van Messem, A., 2023. How to make sense of 3D representations for plant phenotyping: a compendium of processing and analysis techniques. Plant Methods 19, 60. https://doi.org/10.1186/s13007-023-01031-z.
Harfouche, A.L., Jacobson, D.A., Kainer, D., Romero, J.C., Harfouche, A.H., Mugnozza, G.S., Moshelion, M., Tuskan, G.A., Keurentjes, J.J., Altman, A., 2019. Accelerating climate resilient plant breeding by applying next-generation artificial intelligence. Trends in Biotechnology 37, 1217–1235. https://doi.org/10.1016/j.tibtech.2019.05.007.
Harfouche, A.L., Nakhle, F., Harfouche, A.H., Sardella, O.G., Dart, E., Jacobson, D., 2023. A primer on artificial intelligence in plant digital phenomics: Embarking on the data to insights journey. Trends in Plant Science 28, 154–184. https://doi.org/10.1016/j.tplants.2022.08.021.
Hartman, G.L., Rupe, J.C., Sikora, E.J., Domier, L.L., Davis, J.A., Steffey, K.L., 2015. Compendium of soybean diseases and pests. Disease Compendium Series, 5th ed. APS Press, The American Phytopathological Society, St. Paul, Minnesota, U.S.A.
Hartman, G.L., West, E.D., Herman, T.K., 2011. Crops that feed the World 2. Soybean—worldwide production, use, and constraints caused by pathogens and pests. Food Security 3, 5–17. https://doi.org/10.1007/s12571-010-0108-x.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.
Heap, I., 2024. The international herbicide-resistant weed database. www.weedscience.org.
Hernandez, C.M., Correndo, A., Kyveryga, P., Prestholt, A., Ciampitti, I.A., 2023. On-farm soybean seed protein and oil prediction using satellite data. Computer and Electronics in Agriculture 212, 108096. https://doi.org/10.1016/j.compag.2023.108096.
Herr, A.W., Adak, A., Carroll, M.E., Elango, D., Kar, S., Li, C., Jones, S.E., Carter, A.H., Murray, S.C., Paterson, A., Sankaran, S., Singh, A., Singh, A.K., 2023. Unoccupied aerial systems imagery for phenotyping in cotton, maize, soybean, and wheat breeding. Crop Science 63, 1722–1749. https://doi.org/10.1002/csc2.21028.
Herrero-Huerta, M., Rodriguez-Gonzalvez, P., Rainey, K.M., 2020. Yield prediction by machine learning from UAS-based mulit-sensor data fusion in soybean. Plant Methods 16, 78. https://doi.org/10.1186/s13007-020-00620-6.
Hitti, Y., Buzatu, I., Del Verme, M., Lefsrud, M., Golemo, F., Durand, A., 2024. Growspace: a reinforcement learning environment for plant architecture. Computers and Electronics in Agriculture 217, 108613. https://doi.org/10.1016/j.compag.2024.108613.
Houle, D., Govindaraju, D.R., Omholt, S., 2010. Phenomics: The next challenge. Nature Reviews. Genetics 11, 855–866. https://doi.org/10.1038/nrg2897.
Houser, D., Darr, M., Huffman, R., 2024. Precision spraying technology. Integrated Crop Management News.Iowa State University Extension and Outreach. https://www.extension.iastate.edu/icm/precision-spraying-technology.

Howard, D.A., Ma, Z., Aaslyng, J.M., Jørgensen, B.N., 2020. Data architecture for digital twin of commercial greenhouse production. In: 2020 RIVF international conference on computing and communication technologies (RIVF), IEEE, pp. 1–7. https://doi.org/10.1109/RIVF48685.2020.9140726.

Hugging Face, 2024. Hugging Face: The AI community building the future.https://huggingface.co/. Accessed: 2024-08-03.

Hurst, W., Mendoza, F.R., Tekinerdogan, B., 2021. Augmented reality in precision farming: Concepts and applications. Smart Cities 4, 1454–1468. https://doi.org/10.3390/smartcities4040077.

Isafiade, O., Mabiletsa, O., 2020.Immersive technologies for development: An analysis of agriculture. In: 2020 ITU Kaleidoscope: Industry-Driven Digital Transformation (ITU K), Institute of Electrical and Electronics Engineers, Ha Noi, Vietnam, pp. 1–8. https://doi.org/10.23919/ITUK50268.2020.9303205.

Ivanov, D., 2024. Conceptualisation of a 7-element digital twin framework in supply chain and operations management. International Journal of Production Research 62, 2220–2232. https://doi.org/10.1080/00207543.2023.2217291.

Ivanov, D., Dolgui, A., 2021. A digital supply chain twin for managing the disruption risks and resilience in the era of industry 4.0. Production Planning & Control 32, 775–788. https://doi.org/10.1080/09537287.2020.1768450.

Jagadish, S.V.K., Way, D.A., Sharkey, T.D., 2021. Plant heat stress: concepts directing future research. Plant Cell Environment 44, 1992–2005. https://doi.org/10.1111/pce.14050.

Jamal, A., Cai, X., Qiao, X., Garcia, L., Wang, J., Amori, A., Yang, H., 2023. Real-time irrigation scheduling based on weather forecasts, field observations, and human-machine interactions. Water Resources Research 59, e2023WR035810. https://doi.org/10.1029/2023WR035810.

Jiang, L.E.A., 2023. CropDeep: an open agricultural dataset focusing on tomato diseases based on IoT and mobile cameras. Sensors 23, 2134. https://doi.org/10.3390/s23042134.

Jiang, Z., Liu, C., Ganapathysubramanian, B., Hayes, D.J., Sarkar, S., 2020. Predicting county-scale maize yields with publicly available data. Scientific Reports 10, 14957. https://doi.org/10.1038/s41598-020-71898-8.

Johnson, D.M., 2014. An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. Remote Sensing of Environment 141, 116–128. https://doi.org/10.1016/j.rse.2013.10.027.

Jones, D., Snider, C., Nassehi, A., Yon, J., Hicks, B., 2020. Characterising the digital twin: a systematic literature review. CIRP Journal of Manufacturing Science and Technology 29, 36–52. https://doi.org/10.1016/j.cirpj.2020.02.002.

Jones, J., Hoogenboom, G., Porter, C., Boote, K., Batchelor, W., Hunt, L., Wilkens, P., Singh, U., Gijsman, A., Ritchie, J., 2003. The DSSAT cropping system model. European Journal of Agronomy 18, 235–265. https://doi.org/10.1016/S1161-0301(02)00107-7.

Jones, S.E., Ayanlade, T., Fallen, B., Jubery, T.Z., Singh, A., Ganapathysubramanian, B., Sarkar, S., Singh, A., 2024. Multi-sensor and multi-temporal high-throughput phenotyping for monitoring and early detection of water-limiting stress in soybean. arXiv. https://doi.org/10.48550/arxiv.2402.18751.

Joshi, T., Fitzpatrick, M.R., Chen, S., Liu, Y., Zhang, H., Endacott, R.Z., Gaudiello, E.C., Stacey, G., Nguyen, H.T., Xu, D., 2014. Soybean knowledge base (SoyKB): a web resource for integration of soybean translational genomics and molecular breeding. Nucleic Acids Research 42, D1245–D1252. https://doi.org/10.1093/nar/gkt905.

Ju, S., Lim, H., Ma, J.W., Kim, S., Lee, K., Zhao, S., Heo, J., 2021. Optimal county-level crop yield prediction using MODIS-based variables and weather data: a comparative study on machine learning models. Agricultural and Forest Meterology 307, 108530. https://doi.org/10.1016/j.agrformet.2021.108530.

Jubery, T.Z., Carley, C.N., Singh, A., Sarkar, S., Ganapathysubramanian, B., Singh, A.K., 2021. Using machine learning to develop a fully automated soybean nodule acquisition pipeline (SNAP). Plant Phenomics 2021, 9834746. https://doi.org/10.34133/2021/9834746.

Kaler, A.S., Purcell, L.C., Beissinger, T., Gillman, J.D., 2022. Genomic prediction models for traits differing in heritability for soybean, rice, and maize. BMC Plant Biology 22, 87. https://doi.org/10.1186/s12870-022-03479-y.

Kamburjan, E., Sieve, R., Prabhu Baramashetru, C., Amato, M., Barmina, G., Occhipinti, E., Broch Johnsen, E., 2024. GreenhouseDT: An exemplar for digital twins. In: Proceedings of the 19th International Symposium on Software Engineering for Adaptive and Self-Managing Systems, pp. 175–181. https://doi.org/10.1145/3643915.3644108.

Kang, Y., Ozdogan, M., Zhu, X., Ye, Z., Hain, C., Anderson, M., 2020. Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the US Midwest. Environmental Research Letters 15, 064005. https://doi.org/10.1088/1748-9326/ab7df9.

Karandish, F., Šimůnek, J., 2016. A comparison of numerical and machine-learning modeling of soil water content with limited input data. Journal of Hydrology 543, 892–909. https://doi.org/10.1016/j.jhydrol.2016.11.007.

Karmakar, P., Teng, S.W., Murshed, M., Pang, S., Li, Y., Lin, H., 2023. Crop monitoring by multimodal remote sensing: a review. Remote Sensing Applications: Society and Environment 33, 101093. https://doi.org/10.1016/j.rsase.2023.101093.

Kashyap, Y., Shrivastava, S.S., Sharma, R., 2022. An improved soybean foliar disease detection system using deep learning. In: 2022 IEEE Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), pp. 1–6. https://doi.org/10.1109/IATMSI56455.2022.10119330.

Keating, B., Carberry, P., Hammer, G., Probert, M., Robertson, M., Holzworth, D., Huth, N., Hargreaves, J., Meinke, H., Hochman, Z., McLean, G., Verburg, K., Snow, V., Dimes, J., Silburn, M., Wang, E., Brown, S., Bristow, K., Asseng, S., Chapman, S., McCown, R., Freebairn, D., Smith, C., 2003. An overview of APSIM, a model designed for farming systems simulation. European Journal of Agronomy 18, 267–288. https://doi.org/10.1016/S1161-0301(02)00108-9.

Khaki, S., Wang, L., Archontoulis, S.V., 2020. A CNN-RNN framework for crop yield prediction. Frontiers in Plant Science 10, 1750. https://doi.org/10.3389/fpls.2019.01750.

Khaled, A.Y., Abd Aziz, S., Bejo, S.K., Nawi, N.M., Seman, I.A., Onwude, D.I., 2018. Early detection of diseases in plant tissue using spectroscopy-applications and limitations. Applied Spectroscopy Reviews 53, 36–64. https://doi.org/10.1080/05704928.2017.1352510.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollár, P., Girshick, R., 2023. Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4015–4026. https://doi.org/10.48550/arXiv.2304.02643.

Kniss, A.R., 2018. Genetically engineered herbicide-resistant crops and herbicide-resistant weed evolution in the United States. Weed Science 66, 260–273. https://doi.org/10.1017/wsc.2017.70.

Kumar, M., Kumar, A., Palaparthy, V.S., 2020. Soil sensors-based prediction system for plant diseases using exploratory data analysis and machine learning. Institute of Electrical and Electronics Engineers Sensors Journal 21, 17455–17468. https://doi.org/10.1109/JSEN.2020.3046295.

Kumar, S.S., Raghavendra, B., 2019. Diseases detection of various plant leaf using image processing techniques: a review. In: 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), Institute of Electrical and Electronics Engineers, pp. 313–316. https://doi.org/10.1109/ICACCS.2019.8728325.

Kumar, V., Jain, M., 2015. The CRISPR-Cas system for plant genome editing: advances and opportunities. Journal of Experimental Botany 66, 47–57. https://doi.org/10.1093/jxb/eru429.

Landau, C.A., Hager, A.G., Williams, M.M., 2022. Deteriorating weed control and variable weather portends greater soybean yield losses in the future. Science of the Total Environment 830, 154764. https://doi.org/10.1016/j.scitotenv.2022.154764.

Laryukhin, V., Skobelev, P., Lakhin, O., Grachev, S., Yalovenko, V., Yalovenko, O., 2019. The multi-agent approach for developing a cyber-physical system for managing precise farms with digital twins of plants. Cybernetics and Physics 8, 257–261. https://doi.org/10.1109/SMARTCOMP50058.2020.00062.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.

Li, Z., Guo, R., Li, M., Chen, Y., Li, G., 2020. A review of computer vision technologies for plant phenotyping. Computers and Electronics in Agriculture 176, 105672. https://doi.org/10.1016/j.compag.2020.105672.

Lindner, M., Schiller, I., Kolb, A., Koch, R., 2010. Time-of-flight sensor calibration for accurate range sensing. Computer Vision and Image Understanding 114, 1318–1328. https://doi.org/10.1016/j.cviu.2009.11.002.

Liu, J., Hertel, T.W., Lammers, R.B., Prusevich, A., Baldos, U.L.C., Grogan, D.S., Frolking, S., 2017. Achieving sustainable irrigation water withdrawals: global impacts on food security and land use. Environmental Research Letters 12, 104009. https://doi.org/10.1088/1748-9326/aa88db.

Liu, S., Martre, P., Buis, S., Abichou, M., Andrieu, B., Baret, F., 2019a. Estimation of plant and canopy architectural traits using the digital plant phenotyping platform. Plant Physiology 181, 881–890. https://doi.org/10.1104/19.00554.

Liu, Y., Wang, D., He, F., Wang, J., Joshi, T., Xu, D., 2019b. Phenotype prediction and genome-wide association study using deep convolutional neural network of soybean. Frontiers in Genetics 10, 486384. https://doi.org/10.3389/fgene.2019.01091.

Liu, Y., Zhang, Y., Liu, X., Shen, Y., Tian, D., Yang, X., Liu, S., Ni, L., Zhang, Z., Song, S., Tan, Z., 2023. SoyOmics: Aa deeply integrated database on soybean multi-omics. Molecular Plant 16, 794–797. https://doi.org/10.1016/j.molp.2023.03.011.

Lowe, A., Harrison, N., French, A.P., 2017. Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress. Plant Methods 13, 80. https://doi.org/10.1186/s13007-017-0233-z.

Lu, J., Fu, H., Tang, X., Liu, Z., Huang, J., Zou, W., Chen, H., Sun, Y., Ning, X., Li, J., 2024. GOA-optimized deep learning for soybean yield estimation using multi-source remote sensing data. Scientific Reports 14, 7097. https://doi.org/10.1038/s41598-024-57278-6.

Luhmann, T., Robson, S., Kyle, S., Boehm, J., 2023. Close-range photogrammetry and 3D imaging. De Gruyter textbook, second ed. Walter de Gruyter GmbH & Co KG.

Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions. arXiv. https://doi.org/10.48550/arxiv.1705.07874.

Ma, W., Qiu, Z., Song, J., Li, J., Cheng, Q., Zhai, J., Ma, C., 2018. A deep convolutional neural network approach for predicting phenotypes from genotypes. Planta 248, 1307–1318. https://doi.org/10.1007/s00425-018-2976-9.

Madni, A.M., Madni, C.C., Lucero, S.D., 2019. Leveraging digital twin technology in model-based systems engineering. Systems 7, 7. https://doi.org/10.3390/systems7010007.

Mahmud, M.S.A., Abidin, M.S.Z., Emmanuel, A.A., Hasan, H.S., 2020. Robotics and automation in agriculture: present and future applications. Applications of Modelling and Simulation 4, 130–140.

Maimaitijiang, M., Sagan, V., Sidike, P., Hartling, S., Esposito, F., Fritschi, F.B., 2020. Soybean yield prediction from UAV using multimodal data fusion and deep learning. Remote Sensing of Environment 237, 111599. https://doi.org/10.1016/j.rse.2019.111599.

Marru, S., Pierce, M., Plale, B., Pamidighantam, S., Wannipurage, D., Christie, M., Ranawaka, I., Abeysinghe, E., Quick, R., Tajkhorshid, E., Koric, S., Basney, J., Spivak, M., Isralewitz, B., Bernardi, R., Gomes, D., Krishnan, G., Bazhenov, M., Smallen, S., Majumdar, A., Arkhipov, A., Dai, K., Liu, X.P., Yoshimoto, K., 2023. Cybershuttle: an end-to-end cyberinfrastructure continuum to accelerate discovery in science and engineering. In: Practice and Experience in Advanced Research Computing, pp. 26–34. https://doi.org/10.1145/3569951.359360.

McCarville, M.T., Marett, C.C., Mullaney, M.P., Gebhart, G.D., Tylka, G.L., 2017. Increase in soybean cyst nematode virulence and reproduction on resistant soybean varieties in Iowa from 2001 to 2015 and the effects on soybean yields. Plant Health Progress 18, 146–155. https://doi.org/10.1094/PHP-RS-16-0062.

McGovern, A., Lagerquist, R., Gagne, D.J., Eli Jergensen, G., Elmore, K.L., Homeyer, C.R., Smith, T., 2019. Making the black box more transparent: Understanding the physical implications of machine learning. Bulleting of the American Meteorological Society 100, 2175–2199. https://doi.org/10.1175/BAMS-D-18-0195.1.

McGuire, M., Soman, C., Diers, B., Chowdhary, G., 2021. High throughput soybean pod-counting with in-field robotic data collection and machine-vision based data analysis. arXiv. arXiv:2105.10568.

McMechan, A.J., Hodgson, E.W., Varenhorst, A.J., Hunt, T., Wright, R., Potter, B., 2021. Soybean Gall Midge (Diptera: Cecidomyiidae), a new species causing injury to soybean in the United States. Journal of Integrated Pest Management 12, 1. https://doi.org/10.1093/jipm/pmab001.

Mekala, M.S., Viswanathan, P., 2017. A survey: Smart agriculture IoT with cloud computing. In: 2017 international conference on microelectronic devices, circuits and systems (ICMDCS), Institute of Electrical and Electronics Engineers, pp. 1–7. https://doi.org/10.1109/ACCESS.2019.2949703.

Merrick, L.F., Herr, A.W., Sandhu, K.S., Lozada, D.N., Carter, A.H., 2022. Optimizing plant breeding programs for genomic selection. Agronomy (Basel) 12, 714. https://doi.org/10.3390/agronomy12030714.

Meuwissen, T.H.E., Hayes, B.J., Goddard, M.E., 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157, 1819–1829. https://doi.org/10.1093/genetics/157.4.1819.

Miranowski, J.A., 1980. Estimating the relationship between pest management and energy prices, and the implications for environmental damage. American Journal of Agricultural Economics 62, 995–1000. https://doi.org/10.2307/1240300.

Moghadam, P., Lowe, T., Edwards, E.J., 2020. Digital twin for the future of orchard production systems. In: Proceedings, MDPI, p. 92. https://doi.org/10.3390/proceedings2019036092.

Moghadam, P., Ward, D., Goan, E., Jayawardena, S., Sikka, P., Hernandez, E., 2017. Plant disease detection using hyperspectral imaging. In: 2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Institute of Electrical and Electronics Engineers, pp. 1–8. https://doi.org/10.1109/DICTA.2017.8227476.

Mohanty, S.P., Hughes, D.P., Salathé, M., 2016. Using deep learning for image-based plant disease detection. Frontiers in Plant Science 7, 1419. https://doi.org/10.3389/fpls.2016.01419.

Montesinos-López, O.A., Montesinos-López, A., Pérez-Rodríguez, P., Barrón-López, J.A., Martini, J.W., Fajardo-Flores, S.B., Gaytan-Lugo, L.S., Santana-Mancilla, P.C., Crossa, J., 2021. A review of deep learning applications for genomic selection. BMC Genomics 22, 1–23. https://doi.org/10.1186/s12864-020-07319-x.

Mostafa, S., Mondal, D., Panjvani, K., Kochian, L., Stavness, I., 2023. Explainable deep learning in plant phenotyping. Frontiers in Artificial Intelligence 6, 1203546. https://doi.org/10.3389/frai.2023.1203546.

Mourtzinis, S., Rowntree, S.C., Suhre, J.J., Weidenbenner, N.H., Wilson, E.W., Davis, V.M., Naeve, S.L., Casteel, S.N., Diers, B.W., Esker, P.D., Specht, J.E., Conley, S.P., 2014. The use of reflectance data for in-season soybean yield prediction. Agronomy Journal 106, 1159–1168. https://doi.org/10.2134/agronj13.0577.

Nagasubramanian, K., Jones, S., Sarkar, S., Singh, A.K., Singh, A., Ganapathysubramanian, B., 2018. Hyperspectral band selection using genetic algorithm and support vector machines for early identification of charcoal rot disease in soybean stems. Plant Methods 14, 1–13. https://doi.org/10.1186/s13007-018-0349-9.

Nagasubramanian, K., Jones, S., Singh, A.K., Sarkar, S., Singh, A., Ganapathysubramanian, B., 2019. Plant disease identification using explainable 3D deep learning on hyperspectral images. Plant Methods 15, 98. https://doi.org/10.1186/s13007-019-0479-8.

Nagasubramanian, K., Jubery, T., Fotouhi Ardakani, F., Mirnezami, S.V., Singh, A.K., Singh, A., Sarkar, S., Ganapathysubramanian, B., 2021. How useful is active learning for image-based plant phenotyping? The Plant Phenome Journal 4, e20020. https://doi.org/10.1002/ppj2.20020.

Nagasubramanian, K., Singh, A.K., Singh, A., Sarkar, S., Ganapathysubramanian, B., 2020. Usefulness of interpretability methods to explain deep learning based plant stress phenotyping. arXiv. https://doi.org/10.48550/arXiv.2007.05729.

Naik, H.S., Zhang, J., Lofquist, A., Assefa, T., Sarkar, S., Ackerman, D., Singh, A., Singh, A.K., Ganapathysubramanian, B., 2017. A real-time phenotyping framework using machine learning for plant stress severity rating in soybean. Plant Methods 13, 23. https://doi.org/10.1186/s13007-017-0173-7.

Nayar, P., Chhibber, S., Dubey, A.K., 2023. Detection of plant disease and pests using coherent deep learning algorithms. In: 2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), pp. 8–12. https://doi.org/10.1109/CISES58720.2023.10183522.

Negus, K.L., Li, X., Welch, S.M., Yu, J., 2024. The role of artificial intelligence in crop improvement. Advances in Agronomy. Elsevier, pp. 1–66. https://doi.org/10.1016/bs.agron.2023.11.001. volume 184.

Newman, S.J., Furbank, R.T., 2021. Explainable machine learning models of major crop traits from satellite-monitored continent-wide field trial data. Nature Plants 7, 1354–1363. https://doi.org/10.1038/s41477-021-01001-0.

Nigam, A., Kabra, P., Doke, P., 2011. Augmented reality in agriculture. In: 2011 IEEE 7th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), IEEE, pp. 445–448. https://doi.org/10.1109/WiMOB.2011.6085361.

NOAA National Centers for Environmental Information (NCEI), 2024. U.S. billion-dollar weather and climate disasters.https://www.ncei.noaa.gov/access/billions/, https://doi.org/10.25921/stkw-7w73. Accessed: 2024-07-25.

Nobles, C., Burrell, D.N., Waller, T., Cusak, A., 2022. Food sustainability, cyber-biosecurity, emerging technologies, and cybersecurity risks in the agriculture and food industries. International Journal of Environmental Sustainability and Green Technologies (IJESGT) 13, 1–17. https://doi.org/10.4018/IJESGT.309744.

Oberti, R., Marchi, M., Tirelli, P., Calcante, A., Iriti, M., Tona, E., Hočevar, M., Baur, J., Pfaff, J., Schütz, C., Ulbrich, H., 2016. Selective spraying of grapevines for disease control using a modular agricultural robot. Biosystems Engineering 146, 203–215. https://doi.org/10.1016/j.biosystemseng.2015.12.004.

Oerke, E.C., 2006. Crop losses to pests. The Journal of Agricultural Science 144, 31–43. https://doi.org/10.1017/S0021859605005708.

Ohnishi, S., Miyoshi, T., Shirai, S., 2010. Low temperature stress at different flower developmental stages affects pollen development, pollination, and pod set in soybean. Environmental and Experimental Botany 69, 56–62. https://doi.org/10.1016/j.envexpbot.2010.02.007.

OneSoil, 2024. Free app for precision farming. https://onesoil.ai/en/. Accessed: 2024-08-04.
Panda, D.K., Chaudhary, V., Fosler-Lussier, E., Machiraju, R., Majumdar, A., Plale, B., Ramnath, R., Sadayappan, P., Savardekar, N., Tomko, K., 2024. Creating intelligent cyberinfrastructure for democratizing AI. AI Magazine 45, 22–28. https://doi.org/10.1002/aaai.12166.
Parmley, K., Nagasubramanian, K., Sarkar, S., Ganapathysubramanian, B., Singh, A.K., 2019a. Development of optimized phenomic predictors for efficient plant breeding decisions using phenomic-assisted selection in soybean. Plant Phenomics 2019. https://doi.org/10.34133/2019/5809404.
Parmley, K.A., Higgins, R.H., Ganapathysubramanian, B., Sarkar, S., Singh, A.K., 2019b. Machine learning approach for prescriptive plant breeding. Scientific Reports 9, 17132. https://doi.org/10.1038/s41598-019-53451-4.
Parween, S., Hameed, R.S., Sinha, K., 2021. IoT and its real-time application in agriculture. Handbook of Research on Knowledge and Organization Systems in Library and Information Science. IGI Global, pp. 103–123. https://doi.org/10.4018/978-1-7998-7258-0.ch006.
Pathan, M., Patel, N., Yagnik, H., Shah, M., 2020. Artificial cognition for applications in smart agriculture: a comprehensive review. Artificial Intelligence in Agriculture 4, 81–95. https://doi.org/10.1016/j.aiia.2020.06.001.
Paudel, D., de Wit, A., Boogaard, H., Marcos, D., Osinga, S., Athanasiadis, I.N., 2023. Interpretability of deep learning models for crop yield forecasting. Computers and Electronics in Agriculture 206, 107663. https://doi.org/10.1016/j.compag.2023.107663.
Paulus, S., 2019. Measuring crops in 3D: using geometry for plant phenotyping. Plant Methods 15, 103. https://doi.org/10.1186/s13007-019-0490-0.
Peirone, L.S., Pereyra Irujo, G.A., Bolton, A., Erreguerena, I., Aguirrezábal, L.A.N., 2018. Assessing the efficiency of phenotyping early traits in a greenhouse automated platform for predicting drought tolerance of soybean in the field. Frontiers in Plant Science 9, 587. https://doi.org/10.3389/fpls.2018.00587.
Pérez-Rodríguez, P., Gianola, D., González-Camacho, J.M., Crossa, J., Manès, Y., Dreisigacker, S., 2012. Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. G3: Genes Genomes Genetics 2, 1595–1605. https://doi.org/10.1534/g3.112.003665.
Perich, G., Hund, A., Anderegg, J., Roth, L., Boer, M.P., Walter, A., Liebisch, F., Aasen, H., 2020. Assessment of multi-image unmanned aerial vehicle based high-throughput field phenotyping of canopy temperature. Frontiers in Plant Science 11, 150. https://doi.org/10.3389/fpls.2020.00150.
Pestek, A., Sarvan, M., 2020. Virtual reality and modern tourism. Journal of Tourism Futures 7, 245–250. https://doi.org/10.1108/JTF-01-2020-0004.
Phuoc, L.H., Suliansyah, I., Arlius, F., Chaniago, I., Xuan, N.T.T., Van Quang, P., 2023. Literature review crop modeling and introduction a simple crop model. Journal of Applied Agricultural Science and Technology 7, 197–216. https://doi.org/10.55043/jaast.v7i3.123.
PlantVillage, 2024. PlantVillage. https://plantvillage.psu.edu. Accessed: 2024-08-30.
Poonia, R.C., Singh, V., Nayak, S.R., 2022. Deep learning for sustainable agriculture(Eds.). Cognitive Data Science in Sustainable Computing. Academic Press, Elsevier, San Diego.
Pramanik, M., Khanna, M., Singh, M., Singh, D., Sudhishri, S., Bhatia, A., Ranjan, R., 2022. Automation of soil moisture sensor-based basin irrigation system. Smart Agricultural Technology 2, 100032. https://doi.org/10.1016/j.atech.2021.100032.
Pylianidis, C., Osinga, S., Athanasiadis, I.N., 2021. Introducing digital twins to agriculture. Computers and Electronics in Agriculture 184, 105942. https://doi.org/10.1016/j.compag.2020.105942.

QGIS.org, 2024. QGIS Geographic Information System. Open Source Geospatial Foundation Project. https://www.qgis.org/ Accessed: 2024-08-04.
Qiao, Y., Valente, J., Su, D., Zhang, Z., He, D., 2022. AI, sensors and robotics in plant phenotyping and precision agriculture. Frontiers in Plant Science 13, 1064219. https://doi.org/10.3389/fpls.2022.1064219.
Quantitative Plant, 2024. Quantitative Plant: a website presenting image analysis software tools and models for plants. http://www.quantitative-plant.org/. Accessed: 2024-08-30.
Rairdin, A., Fotouhi, F., Zhang, J., Mueller, D.S., Ganapathysubramanian, B., Singh, A.K., Dutta, S., Sarkar, S., Singh, A., 2022. Deep learning-based phenotyping for genome wide association studies of sudden death syndrome in soybean. Frontiers in Plant Science 13, 966244. https://doi.org/10.3389/fpls.2022.966244.
Ramesh, D., Vardhan, B.V., 2015. Analysis of crop yield prediction using data mining techniques. International Journal of Research in Engineering and Technology 4, 47–473. https://doi.org/10.15623/ijret.2015.0401071.
Randelović, P., Dordević, V., Miladinović, J., Prodanović, S., Ćeran, M., Vollmann, J., 2023. High-throughput phenotyping for non-destructive estimation of soybean fresh biomass using a machine learning model and temporal UAV data. Plant Methods 19, 89. https://doi.org/10.1186/s13007-023-01054-6.
Ray, D.K., Mueller, N.D., West, P.C., Foley, J.A., 2013. Yield trends are insufficient to double global crop production by 2050. PloS One 8, e66428. https://doi.org/10.1371/journal.pone.0066428.
Ray, S., Jarquin, D., Howard, R., 2023. Comparing artificial-intelligence techniques with state-of-the-art parametric prediction models for predicting soybean traits. The Plant Genome 16, e20263. https://doi.org/10.1002/tpg2.20263.
Raymond, C., Horton, R.M., Zscheischler, J., Martius, O., Agha Kouchak, A., Balch, J., Bowen, S.G., Camargo, S.J., Hess, J., Kornhuber, K., Oppenheimer, M., Ruane, A.C., Wahl, T., White, K., 2020. Understanding and managing connected extreme events. Nature Climate Change 10, 611–621. https://doi.org/10.1038/s41558-020-0790-4.
Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779–788. https://doi.org/10.1109/CVPR.2016.91.
Reyes Yanes, A., Abbasi, R., Martinez, P., Ahmad, R., 2022. Digital twinning of hydroponic grow beds in intelligent aquaponic systems. Sensors 22, 7393. https://doi.org/10.3390/s22197393.
Riera, L.G., Carroll, M.E., Zhang, Z., Shook, J.M., Ghosal, S., Gao, T., Singh, A., Bhattacharya, S., Ganapathysubramanian, B., Singh, A.K., Sarkar, S., 2021. Deep multiview image fusion for soybean yield estimation in breeding applications. Plant Phenomics 2021. https://doi.org/10.34133/2021/9846470.
Rivera, G., Porras, R., Florencia, R., Sánchez-Solís, J.P., 2023. LiDAR applications in precision agriculture for cultivating crops: a review of recent advances. Computers and Electronics in Agriculture 207, 107737. https://doi.org/10.1016/j.compag.2023.107737.
Rosell-Polo, J., Auat Cheein, F., Gregorio, E., Andújar, D., Puigdomènech, L., Masip, J., Escolà, A., 2015. Chapter three – Advances in structured light sensors applications in precision agriculture and livestock farming. In: Sparks, D. (Ed.), Advances in agronomy. Academic Press, pp. 71–112. https://doi.org/10.1016/bs.agron.2015.05.002. volume 133.
Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1973. Monitoring vegetation systems in the great plains with ERTSNASA Special Publications, pp. 309–317. https://api.semanticscholar.org/CorpusID:133358670.
Rovira-Más, F., Zhang, Q., Reid, J.F., 2008. Stereo vision three-dimensional terrain maps for precision agriculture. Computers and Electronics in Agriculture 60, 133–143. https://doi.org/10.1016/j.compag.2007.07.007.

Sa, I., Chen, Z., Popović, M., Khanna, R., Liebisch, F., Nieto, J., Siegwart, R., 2017. WeedNet: Dense semantic weed classification using multispectral images and MAV for smart farming. IEEE Robotics and Automation Letters 3, 588–595. https://doi.org/10.1109/LRA.2017.2774979.

Saadati, M., Balu, A., Chiranjeevi, S., Jubery, T.Z., Singh, A.K., Sarkar, S., Singh, A., Ganapathysubramanian, B., 2024. Out-of-distribution detection algorithms for robust insect classification. Plant Phenomics 6, 0170. https://doi.org/10.34133/plantphenomics.0170.

Sacala, I.S., Dumitrache, I., Moisescu, M.A., Stanescu, A.M., Caramihai, S.I., 2017. Agricultural enterprise architecture based on cyber physical systems paradigm. In: 2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC), Institute of Electrical and Electronics Engineers, Funchal, pp. 1306–1311. https://doi.org/10.1109/ICE.2017.8280031.

Saleem, N., Balu, A., Jubery, T.Z., Singh, A., Singh, A.K., Sarkar, S., Ganapathysubramanian, B., 2024. Class-specific data augmentation for plant stress classification. The Plant Phenome Journal 7, e20112. https://doi.org/10.1002/ppj2.20112.

Salter, W.T., Shrestha, A., Barbour, M.M., 2021. Open source 3D phenotyping of chickpea plant architecture across plant development. Plant Methods 17, 1–16. https://doi.org/10.1186/s13007-021-00795-6.

Samek, W., Montavon, G., Lapuschkin, S., Anders, C.J., Müller, K.R., 2021. Explaining deep neural networks and beyond: A review of methods and applications. Proceedings of the Institute of Electrical and Electronics Engineers 109, 247–278. https://doi.org/10.1109/JPROC.2021.3060483.

Santini, M., Noce, S., Antonelli, M., Caporaso, L., 2022. Complex drought patterns robustly explain global yield loss for major crops. Scientific Reports 12, 5792. https://doi.org/10.1038/s41598-022-09611-0.

Sarkar, S., Ganapathysubramanian, B., Singh, A., Fotouhi, F., Kar, S., Nagasubramanian, K., Chowdhary, G., Das, S.K., Kantor, G., Krishnamurthy, A., Merchant, N., Singh, A.K., 2024. Cyber-agricultural systems for crop breeding and sustainable production. Trends in Plant Science 29, 130–149. https://doi.org/10.1016/j.tplants.2023.08.001.

Schleich, B., Anwer, N., Mathieu, L., Wartzack, S., 2017. Shaping the digital twin for design and production engineering. CIRP Annals 66, 141–144. https://doi.org/10.1016/j.cirp.2017.04.040.

Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, Q., Thelen, J.J., Cheng, J., Xu, D., Hellsten, U., May, G.D., Yu, Y., Sakurai, T., Umezawa, T., Bhattacharyya, M.K., Sandhu, D., Valliyodan, B., Lindquist, E., Peto, M., Grant, D., Shu, S., Goodstein, D., Barry, K., Futrell-Griggs, M., Abernathy, B., Du, J., Tian, Z., Zhu, L., Gill, N., Joshi, T., Libault, M., Sethuraman, A., Zhang, X.C., Shinozaki, K., Nguyen, H.T., Wing, R.A., Cregan, P., Specht, J., Grimwood, J., Rokhsar, D., Stacey, G., Shoemaker, R.C., Jackson, S.A., 2010. Genome sequence of the palaeopolyploid soybean. Nature 463, 178–183. https://doi.org/10.1038/nature08670.

Schwalbert, R.A., Amado, T., Corassa, G., Pott, L.P., Prasad, P.V.V., Ciampitti, I.A., 2020. Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil. Agricultural and Forest Meterology 284, 107886. https://doi.org/10.1016/j.agrformet.2019.107886.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision, pp. 618–626. https://doi.org/10.1109/ICCV.2017.74.

Seshaiah, M., Shravya, D., Tarun, K., Vandana, R., Varshitha, R., 2024. Recognition of plant disease using deep learning approaches. In: 2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS), pp. 1–6. https://doi.org/10.1109/ICKECS61492.2024.10616775.

Shaheb, M.R., Venkatesh, R., Shearer, S.A., 2021. A review on the effect of soil compaction and its management for sustainable crop production. Journal of Biosystems Engineering 46, 417–439. https://doi.org/10.1007/s42853-021-00117-7.
Shahhosseini, M., Hu, G., Huber, I., Archontoulis, S.V., 2021. Coupling machine learning and crop modeling improves crop yield prediction in the us corn belt. Scientific Reports 11, 1606. https://doi.org/10.1038/s41598-020-80820-1.
Shaikh, F.K., Karim, S., Zeadally, S., Nebhen, J., 2022. Recent trends in internet-of-things-enabled sensor technologies for smart agriculture. Institute of Electrical and Electronics Engineers Internet of Things Journal 9, 23583–23598. https://doi.org/10.1109/JIOT.2022.3210154.
Sharma, R., Parhi, S., Shishodia, A., 2020. Industry 4.0 applications in agriculture: cyber-physical agricultural systems (CPASs). In: Advances in Mechanical Engineering: Select Proceedings of ICAME 2020. Springer, pp. 807–813. https://doi.org/10.1007/978–981-15–3639-7_97.
Shoeiby, M., Robles-Kelly, A., Timofte, R., Zhou, R., Lahoud, F., Süsstrunk, S., Xiong, Z., Shi, Z., Chen, C., Liu, D., Zha, Z.J., Wu, F., Wei, K., Zhang, T., Wang, L., Fu, Y., Nagasubramanian, K., Singh, A.K., Singh, A., Sarkar, S., Ganapathysubramanian, B., 2019. PIRM2018 challenge on spectral image super-resolution: methods and results. In: Leal-Taixé, L., Roth, S. (Eds.), Computer Vision – ECCV 2018 Workshops. Springer International Publishing, Cham, pp. 356–371. https://doi.org/10.1007/978-3-030-11021-5_22.
Shook, J., Gangopadhyay, T., Wu, L., Ganapathysubramanian, B., Sarkar, S., Singh, A.K., 2021. Crop yield prediction integrating genotype and weather variables using deep learning. PLoS One 16, e0252402. https://doi.org/10.1371/journal.pone.0252402.
Simonyan, K., Vedaldi, A., Zisserman, A., 2013. Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv. https://doi.org/10.48550/arxiv.1312.6034.
Singh, A., Ganapathysubramanian, B., Singh, A.K., Sarkar, S., 2016. Machine learning for high-throughput stress phenotyping in plants. Trends in Plant Science 21, 110–124. https://doi.org/10.1016/j.tplants.2015.10.015.
Singh, A., Jones, S., Ganapathysubramanian, B., Sarkar, S., Mueller, D., Sandhu, K., Nagasubramanian, K., 2021a. Challenges and opportunities in machine-augmented plant stress phenotyping. Trends in Plant Science 26, 53–69. https://doi.org/10.1016/j.tplants.2020.07.010.
Singh, A.K., Ganapathysubramanian, B., Sarkar, S., Singh, A., 2018. Deep learning for plant stress phenotyping: Trends and future perspectives. Trends in Plant Science 23, 883–898. https://doi.org/10.1016/j.tplants.2018.07.004.
Singh, A.K., Singh, A., Sarkar, S., Ganapathysubramanian, B., Schapaugh, W., Miguez, F.E., Carley, C.N., Carroll, M.E., Chiozza, M.V., Chiteri, K.O., Falk, K.G., Jones, S.E., Jubery, T.Z., Mirnezami, S.V., Nagasubramanian, K., Parmley, K.A., Rairdin, A.M., Shook, J.M., Van der Laan, L., Young, T.J., Zhang, J., 2021b. High-throughput phenotyping in soybean. In: Zhou, J., Nguyen, H.T. (Eds.), High-throughput crop phenotyping. Springer International Publishing, pp. 129–163. https://doi.org/10.1007/978-3-030-73734-4_7.
Singh, D.P., Singh, A.K., Singh, A., 2021c. Plant Breeding and Cultivar Development. Elsevier Science, Cambridge, Massachusetts.
Skobelev, P., Mayorov, I., Simonova, E., Goryanin, O., Zhilyaev, A., Tabachinskiy, A., Yalovenko, V., 2020. Development of models and methods for creating a digital twin of plant within the cyber-physical system for precision farming management. Journal of Physics: Conference Series. IOP Publishing, 012022. https://doi.org/10.1088/1742-6596/1703/1/012022.
Song, Q., Hyten, D.L., Jia, G., Quigley, C.V., Fickus, E.W., Nelson, R.L., Cregan, P.B., 2013. Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. PLoS One 8, e54985. https://doi.org/10.1371/journal.pone.0054985.

Song, Q., Hyten, D.L., Jia, G., Quigley, C.V., Fickus, E.W., Nelson, R.L., Cregan, P.B., 2015. Fingerprinting soybean germplasm and its utility in genomic research. G3: Genes, Genomes, Genetics 5, 1999–2006. https://doi.org/10.1534/g3.115.019000.

Song, X.P., Potapov, P.V., Krylov, A., King, L., Di Bella, C.M., Hudson, A., Khan, A., Adusei, B., Stehman, S.V., Hansen, M.C., 2017. National-scale soybean mapping and area estimation in the United States using medium resolution satellite imagery and field survey. Remote Sensing of Environment 190, 383–395. https://doi.org/10.1016/j.rse.2017.01.008.

Specht, J.E., Hume, D.J., Kumudini, S.V., 1999. Soybean yield potential – A genetic and physiological perspective. Crop Science 39, 1560–1570. https://doi.org/10.2135/cropsci1999.3961560x.

Staff, A., 2022. See & spray TM select by John Deere. Resource Magazine 29, 7–8.

State of Colorado, 2024. Concerning the creation of the artificial intelligence impact task force, colorado revised statutes § 2-3-1701. https://leg.colorado.gov/sites/default/files/2024a_1468_signed.pdf. Accessed: 2024-08-21.

State of Utah, 2024. Artificial intelligence amendments, utah code annotated § 13-–2-12. https://le.utah.gov/~2024/bills/static/SB0149.html. Accessed: 2024-08-21.

Suh, A., Prophet, J., 2018. The state of immersive technology research: a literature analysis. Computers in Human Behavior 86, 77–90. https://doi.org/10.1016/j.chb.2018.04.019.

Swetnam, T.L., Antin, P.B., Bartelme, R., Bucksch, A., Camhy, D., Chism, G., Choi, I., Cooksey, A.M., Cosi, M., Cowen, C., Culshaw-Maurer, M., Davey, R., Davey, S., Devisetty, U., Edgin, T., Edmonds, A., Fedorov, D., Frady, J., Fonner, J., Gillan, J.K., Hossain, I., Joyce, B., Lang, K., Lee, T., Littin, S., McEwen, I., Merchant, N., Micklos, D., Nelson, A., Ramsey, A., Roberts, S., Sarando, P., Skidmore, E., Song, J., Sprinkle, M.M., Srinivasan, S., Stanzione, D., Strootman, J.D., Stryeck, S., Tuteja, R., Vaughn, M., Wali, M., Wall, M., Walls, R., Wang, L., Wickizer, T., Williams, J., Wregglesworth, J., Lyons, E., 2024. CyVerse: cyberinfrastructure for open science. PLOS Computational Biology 20, e1011270. https://doi.org/10.1371/journal.pcbi.1011270.

Tace, Y., Tabaa, M., Elfilali, S., Leghris, C., Bensag, H., Renault, E., 2022. Smart irrigation system based on IoT and machine learning. Energy Reports 8, 1025–1036. https://doi.org/10.1016/j.egyr.2022.07.088.

Taranis, 2024. Leading crop management software. https://www.taranis.com/. Accessed: 2024-08-04.

Teixeira, E.I., Fischer, G., van Velthuizen, H., Walter, C., Ewert, F., 2013. Global hot-spots of heat stress on agricultural crops due to climate change. Agriculture and Forest Meteorology 170, 206–215. https://doi.org/10.1016/j.agrformet.2011.09.002.

TERRA-REF, 2024. TERRA phenotyping reference platform. https://terraref.org. Accessed: 2024-08-30.

Tetila, E.C., Machado, B.B., Astolfi, G., de Souza Belete, N.A., Amorim, W.P., Roel, A.R., Pistori, H., 2020. Detection and classification of soybean pests using deep learning with UAV images. Computers and Electronics in Agriculture 179, 105836. https://doi.org/10.1016/j.compag.2020.105836.

Toledo, C.A., Crawford, M., 2023. Deep learning models using multi-modal remote sensing for prediction of maize yield in plant breeding experiments. In: IGARSS 2023–2023 IEEE International Geoscience and Remote Sensing Symposium, Institute of Electrical and Electronics Engineers, pp. 487–490. https://doi.org/10.1109/IGARSS52108.2023.10281741.

Torsoni, G.B., de Oliveira Aparecido, L.E., dos Santos, G.M., Chiquitto, A.G., da Silva Cabral Moraes, J.R., de Souza Rolim, G., 2023. Soybean yield prediction by machine learning and climate. Theoretical and Applied Climatology 151, 1709–1725. https://doi.org/10.1007/s00704-022-04341-9.

Torzoni, M., Tezzele, M., Mariani, S., Manzoni, A., Willcox, K.E., 2024. A digital twin framework for civil engineering structures. Computer Methods in Applied Mechanics and Engineering 418, 116584. https://doi.org/10.1016/j.cma.2023.116584.
Tsialiamanis, G., Wagg, D.J., Dervilis, N., Worden, K., 2021. On generative models as the basis for digital twins. Data-Centric Engineering 2, e11. https://doi.org/10.3390/s22197393.
Umutoni, L., Samadi, V., 2024. Application of machine learning approaches in supporting irrigation decision making: a review. Agricultural Water Management 294, 108710. https://doi.org/10.1016/j.agwat.2024.108710.
United States Department of Agriculture, 2023. Fiscal year 2024–2026 data strategy. https://www.usda.gov/sites/default/files/documents/fy-2024–2026-usda-data-strategy.pdf. Accessed 2024-09-19.
USDA-NASS, 2013. 2012 Agricultural chemical use survey soybeans. https://www.nass.usda.gov/Surveys/Guide_to_NASS_Surveys/Chemical_Use/2012_Soybeans_Highlights/ChemUseHighlights-Soybeans-2012.pdf. Accessed: 2024-08-21.
USDA-NASS, 2019. 2018 Agricultural chemical use survey soybeans. https://www.nass.usda.gov/Surveys/Guide_to_NASS_Surveys/Chemical_Use/2018_Peanuts_Soybeans_Corn/ChemUseHighlights_Soybeans_2018.pdf. Accessed: 2024-08-21.
USDA-NASS, 2023. The yield forecasting program of NASS. https://www.nass.usda.gov/Publications/Methodology_and_Data_Quality/Advanced_Topics/Yield. Accessed: 2024-08-21.
USDA-NASS, 2024. Farms and land in farms 2023 summary. https://downloads.usda.library.cornell.edu/usda-esmis/files/5712m6524/b2775h03z/ns065w04d/fnlo0224.pdf. Accessed: 2024–09-02.
USDA National Agricultural Library, 2024. Ag Data Commons. https://www.nal.usda.gov/services/agdatacommons. Accessed: 2024-08-30.
Varona, L., Legarra, A., Toro, M.A., Vitezica, Z.G., 2018. Non-additive effects in genomic selection. Frontiers in Genetics 9, 341678. https://doi.org/10.3389/fgene.2018.00078.
Vieira, C.C., Chen, P., 2021. The numbers game of soybean breeding in the United States. Crop Breeding and Applied Biotechnology 21, 1–10. https://doi.org/10.1590/1984-70332021v21Sa23.
Vieira, C.C., Sarkar, S., Tian, F., Zhou, J., Jarquin, D., Nguyen, H.T., Zhou, J., Chen, P., 2022. Differentiate soybean response to off-target dicamba damage based on UAV imagery and machine learning. Remote Sensing 14, 1618. https://doi.org/10.3390/rs14071618.
Voss-Fels, K.P., Cooper, M., Hayes, B.J., 2019. Accelerating crop genetic gains with genomic selection. Theoretical and Applied Genetics 132, 669–686. https://doi.org/10.1007/s00122-018-3270-8.
Wang, H., Guohui, S., Zizhong, S., Xiangdong, H., 2023. Effects of climate and price on soybean production: empirical analysis based on panel data of 116 prefecture-level Chinese cities. PLoS One 18, e0273887. https://doi.org/10.1371/journal.pone.0273887.
Watson, A., Ghosh, S., Williams, M.J., Cuddy, W.S., Simmonds, J., Rey, M.D., Asyraf Md Hatta, M., Hinchliffe, A., Steed, A., Reynolds, D., Adamski, N.M., Breakspear, A., Korolev, A., Rayner, T., Dixon, L.E., Riaz, A., Martin, W., Ryan, M., Edwards, D., Batley, J., Raman, H., Carter, J., Rogers, C., Domoney, C., Moore, G., Harwood, W., Nicholson, P., Dieters, M.J., DeLacy, I.H., Zhou, J., Uauy, C., Boden, S.A., Park, R.F., Wulff, B.B.H., Hickey, L.T., 2018. Speed breeding is a powerful tool to accelerate crop research and breeding. Nature Plants 4, 23–29. https://doi.org/10.1038/s41477-017-0083-8.
Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: a meta-review. Remote Sensing of Environment 236, 111402. https://doi.org/10.1016/j.rse.2019.111402.

Wiesner-Hanks, T., Wu, H., Stewart, E., DeChant, C., Kaczmar, N., Lipson, H., Gore, M.A., Nelson, R.J., 2019. Millimeter-level plant disease detection from aerial photographs via deep learning and crowdsourced data. Frontiers in Plant Science 10. https://doi.org/10.3389/fpls.2019.01550.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR guiding principles for scientific data management and stewardship. Scientific Data 3, 1–9. https://doi.org/10.1038/sdata.2016.18.

Wiseman, L., Sanderson, J., Zhang, A., Jakku, E., 2019. Farmers and their data: an examination of farmers' reluctance to share their data through the lens of the laws impacting smart farming. Netherlands Journal of Agricultural Science-Wageningen Journal of Life Sciences 90, 100301. https://doi.org/10.1016/j.njas.2019.04.007.

WorldFAIR Project, 2024. The WorldFAIR Project. https://worldfair-project.eu/. Accessed: 2024-08-03.

Xie, B., Liu, H., Alghofaili, R., Zhang, Y., Jiang, Y., Lobo, F.D., Li, C., Li, W., Huang, H., Akdere, M., Mousas, C., Yu, L.F., 2021. A review on virtual reality skill training applications. Frontiers in Virtual Reality 2, 645153. https://doi.org/10.3389/frvir.2021.645153.

Xie, C., Yang, C., 2020. A review on plant high-throughput phenotyping traits using UAV-based sensors. Computer and Electronics in Agriculture 178, 105731. https://doi.org/10.1016/j.compag.2020.105731.

Xu, P., Zhu, X., Clifton, D.A., 2023. Multimodal learning with transformers: a survey. Institute of Electrical and Electronics Engineers Transactions on Pattern Analysis and Machine Intelligence 45, 12113–12132. https://doi.org/10.1109/TPAMI.2023.3275156.

Xu, Y., Zhang, X., Li, H., Zheng, H., Zhang, J., Olsen, M.S., Varshney, R.K., Prasanna, B.M., Qian, Q., 2022. Smart breeding driven by big data, artificial intelligenceand integrated genomic-enviromic prediction. Molecular Plant 15, 1664–1695. https://doi.org/10.1016/j.molp.2022.09.001.

Yang, C.H., Feuer, B., Jubery, Z., Deng, Z.K., Nakkab, A., Hasan, M.Z., Chiranjeevi, S., Marshall, K., Baishnab, N., Singh, A.K., Singh, A., Sarkar, S., Merchant, N., Hedge, C., Ganapathysubramanian, B., 2024a. Arboretum: A large multimodal dataset enabling AI for biodiversity. arXiv. https://doi.org/10.48550/arXiv.2406.17720.

Yang, M., Wang, G., 2023. Heat stress to jeopardize crop production in the US corn belt based on downscaled CMIP5 projections. Agricultural Systems 211, 103746. https://doi.org/10.1016/j.agsy.2023.103746.

Yang, S., Zheng, L., Chen, X., Zabawa, L., Zhang, M., Wang, M., 2022. Transfer learning from synthetic in-vitro soybean pods dataset for in-situ segmentation of on-branch soybean pods. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1666–1675. https://doi.org/10.1109/CVPRW56347.2022.00173.

Yang, S., Zheng, L., He, P., Wu, T., Sun, S., Wang, M., 2021. High-throughput soybean seeds phenotyping with convolutional neural networks and transfer learning. Plant Methods 17, 50. https://doi.org/10.1186/s13007-021-00749-y.

Yang, Z., Luo, C., Pei, X., Wang, S., Huang, Y., Li, J., Liu, B., Kong, F., Yang, Q.Y., Fang, C., 2024b. SoyMD: a platform combining multi-omics data with various tools for soybean research and breeding. Nucleic Acids Research 52, D1639–D1650. https://doi.org/10.1093/nar/gkad786.

Yoosefzadeh-Najafabadi, M., Torabi, S., Tulpan, D., Rajcan, I., Eskandari, M., 2021. Genome-wide association studies of soybean yield-related hyperspectral reflectance bands using machine learning-mediated data integration methods. Frontiers in Plant Science 12, 777028. https://doi.org/10.3389/fpls.2021.777028.

You, J., Li, X., Low, M., Lobell, D., Ermon, S., 2017. Deep gaussian process for crop yield prediction based on remote sensing data. In: Proceedings of the AAAI conference on artificial intelligence, pp. 4559–4565.

Young, T.J., Chiranjeevi, S., Elango, D., Sarkar, S., Singh, A.K., Singh, A., Ganapathysubramanian, B., Jubery, T.Z., 2024. Soybean canopy stress classification using 3D point cloud data. Agronomy 14, 1181. https://doi.org/10.3390/agronomy14061181.

Young, T.J., Jubery, T.Z., Carley, C.N., Carroll, M., Sarkar, S., Singh, A.K., Singh, A., Ganapathysubramanian, B., 2023. "Canopy fingerprints" for characterizing three-dimensional point cloud data of soybean canopies. Frontiers in Plant Science 14, 1141153. https://doi.org/10.3389/fpls.2023.1141153.

Yu, H., Weng, L., Wu, S., He, J., Yuan, Y., Wang, J., Xu, X., Feng, X., 2024. Time-series field phenotyping of soybean growth analysis by combining multimodal deep learning and dynamic modeling. Plant Phenomics 6, 0158. https://doi.org/10.34133/plantphenomics.0158.

Yu, M., Ma, X., Guan, H., Liu, M., Zhang, T., 2022. A recognition method of soybean leaf diseases based on an improved deep learning model. Frontiers in Plant Science 13, 878834. https://doi.org/10.3389/fpls.2022.878834.

Yuan, W., Wijewardane, N.K., Jenkins, S., Bai, G., Ge, Y., Graef, G.L., 2019. Early prediction of soybean traits through color and texture features of canopy RGB imagery. Scientific Reports 9, 14089. https://doi.org/10.1038/s41598-019-50480-x.

Zeng, S., Mao, Z., Ren, Y., Wang, D., Xu, D., Joshi, T., 2021. G2PDeep: a web-based deep-learning framework for quantitative phenotype prediction and discovery of genomic markers. Nucleic Acids Research 49, W228–W236. https://doi.org/10.1093/nar/gkab407.

Zhang, J., Huang, Y., Reddy, K.N., Wang, B., 2019a. Assessing crop damage from dicamba on non-dicamba-tolerant soybean by hyperspectral imaging through machine learning. Pest Management Science 75, 3260–3272. https://doi.org/10.1002/ps.5448.

Zhang, J., Virk, S., Porter, W., Kenworthy, K., Sullivan, D., Schwartz, B., 2019b. Applications of unmanned aerial vehicle based imagery in turfgrass field trials. Frontiers in Plant Science 10, 279. https://doi.org/10.3389/fpls.2019.00279.

Zhao, G., Cai, W., Wang, Z., Wu, H., Peng, Y., Cheng, L., 2022. Phenotypic parameters estimation of plants using deep learning-based 3-D reconstruction from single RGB image. IEEE Geoscience and Remote Sensing Letters 19, 1–5. https://doi.org/10.1109/LGRS.2022.3198850.

Zhou, J., 2022. High-throughput plant phenotyping. In: Zhang, Q. (Ed.), Encyclopedia of Smart Agriculture Technologies. Springer International Publishing, Cham, pp. 1–10. https://doi.org/10.1007/978-3-030-89123-7_153-1.

Zhou, J., Beche, E., Vieira, C.C., Yungbluth, D., Zhou, J., Scaboo, A., Chen, P., 2022. Improve soybean variety selection accuracy using UAV-based high-throughput phenotyping technology. Frontiers in Plant Science 12, 768742. https://doi.org/10.3389/fpls.2021.768742.

Zhou, J., Mou, H., Zhou, J., Ali, M.L., Ye, H., Chen, P., Nguyen, H.T., 2021. Qualification of soybean responses to flooding stress ssing UAV-based imagery and deep learning. Plant Phenomics 2021, 9892570. https://doi.org/10.34133/2021/9892570.

Zhou, J., Zhou, J., Ye, H., Ali, M.L., Nguyen, H.T., Chen, P., 2020. Classification of soybean leaf wilting due to drought stress using UAV-based imagery. Computers and Electronics in Agriculture 175, 105576. https://doi.org/10.1016/j.compag.2020.105576.

Zhou, W., Yan, Z., Zhang, L., 2024. A comparative study of 11 non-linear regression models highlighting autoencoder, DBN, and SVR, enhanced by SHAP importance analysis in soybean branching prediction. Scientific Reports 14, 5905. https://doi.org/10.1038/s41598-024-55243-x.

Zou, J., Huss, M., Abid, A., Mohammadi, P., Torkamani, A., Telenti, A., 2019. A primer on deep learning in genomics. Nature Genetics 51, 12–18. https://doi.org/10.1038/s41588-018-0295-5.

Zou, K., Wang, H., Yuan, T., Zhang, C., 2023. Multi-species weed density assessment based on semantic segmentation neural network. Precision Agriculture 24, 458–481. https://doi.org/10.1007/s11119-022-09953-9.