International Journal of Environmental & Agriculture Research (IJOEAR)
ISSN: [2454-1850]
[Vol-12, Issue-2, February- 2026]
Website: www.ijoear.com
Journal DOI: 10.25125/agriculture-journal

Applications of Artificial Intelligence and Machine Learning in Plant Breeding

V. Sandeep Varma[1*]; P. Neelima[2]

Director, Achievers Agricet Academy, Bapatla-522 101
Director, Evergreen School of Agriculture, Bapatla. -522 101
*Corresponding Author

Received:- 15 February 2026/ Revised:- 23 February 2026/ Accepted:- 01 March 2026/ Published: 08-03-2026

Abstract— *Plant breeding plays a vital role in meeting the needs of ever-increasing global food demands, climate change and sustainable agricultural practices. Artificial Intelligence and Machine Learning algorithms are used in plant breeding for several activities, including genotype-phenotype prediction, genomic selection, trait discovery, and the optimization of breeding methods. These methods help to determine the location of genetic markers that are related to certain traits based on the analysis of big data sets containing genomic and phenotypic information, which in turn allows the breeders to choose the plants with the desired traits effectively. The use of AI technologies can enhance the breeding process through the use of simulation of breeding results, hence cutting down on the time and resources needed for the conventional trial and error methods. Concerns on data quality, model interpretability and ethical issues need to be addressed so that the application of AI in breeding is reliable and devoid of ethical concerns. Also, the lack of advanced computing infrastructure and skilled personnel is a challenge to many breeders especially in developing countries. The prospects of artificial intelligence (AI) and machine learning (ML) in plant breeding exhibit considerable promise. The continuous advancements in computational biology, genomics, and data analytics will substantially enhance the capabilities of artificial intelligence-driven breeding systems. The integration of artificial intelligence (AI) and machine learning (ML) into plant breeding methodologies has the potential to revolutionize crop improvement efforts, therefore laying the foundation for sustainable agriculture and food security in the context of a changing climate.*

Keywords— Artificial Intelligence, Plant Breeding, Machine Learning.

I. INTRODUCTION

1.1 Significance and Context of Plant Breeding:

Plant breeding is basic to agricultural development and has a long history of human endeavor with the aid of science consistently over centuries (Acquaah, 2012; Tanksley & McCouch, 1997). It began when farmers in the early years of agriculture picked and collected seeds from plants that possessed desirable characteristics, which implied the beginning of crop domestication. For centuries, farmers and developers gradually improved and crossbred crop plants to meet their needs, laying the groundwork for the current plant breeding techniques. It began with the discovery of genetics in the 19th century, with Mendel's work, and entered the modern era in the twentieth century in connection with molecular biology, genomics, and biotechnologies.

1.2 The Advancement of Artificial Intelligence (AI) and Machine Learning (ML) in Plant Breeding:

AI and ML, during the early forms of its implementation, were mostly operating as marginal aids in the typical breeding methodology used to improve the data managing and handling system. Nevertheless, with the emergence and development of high-throughput sequencing technology and continuous accumulation of genetic resources, AI and ML have gradually changed the modern breeding mode in the present breeding mode.

The early applications of AI and ML in plant breeding was focused on utilizing statistic models and algorithm to solve the genotype-phenotype prediction challenges (Lopez et al., 2018). Genomic prediction and genome-wide association studies (GWAS) have helped in flagging several marker traits which has given breeders knowledge on the genetic architecture of crops (Crossa et al., 2017). Because AI and ML technologies are evolving rapidly and rapidly, breeders were forced to search for better algorithms to predict genotypes and phenotypes or discover traits, such as deep learning algorithms (Zhang et al., 2019).

The most significant application of Deep Learning Models, especially Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs), was seen in handling large-scale genomics and phenotypic data, establishing complex genotype-phenotype correlations, and identification of novel associated genetic variants (Kadam & Yin, 2020). Previously, Artificial intelligence algorithms were designed and implemented in decision support systems which helped the breeders in formulating the breeding strategies, choosing the parent organisms and to manage the populations so as to enable the breeders to formulate breeding schemes favorable to attain their goals and constraints put in place by the available resources (Jarquín et al., 2014).

Robotic technology with remote sensing and phenomics platforms improved the application of AI and ML in plant breeding by allowing high-throughput phenotyping and non-destructive trait measurements (Golzarian et al., 2011). Robotic phenotyping systems using AI solution methods enable rapid and accurate assessment of plant and crop characteristics such as biomass yield, disease resistance, and stress tolerance features. Cumulatively, it allowed breeders to have adequate knowledge on selection differences (Mulla et al., 2020).

II. GENOTYPE-PHENOTYPE PREDICTION

2.1 Conventional Techniques Versus Artificial Intelligence and Machine Learning Strategies:

The evolution of genotype-phenotype prediction methods in plant breeding encapsulates a transformative journey from traditional approaches to cutting-edge Artificial Intelligence (AI) and Machine Learning (ML) techniques. Traditional methods relied on phenotype observation, pedigree analysis, and controlled crosses to predict the performance of crop varieties. These approaches were time-consuming, labor-intensive, and limited by the complexity of genetic interactions. They involved field trials, where plant phenotypes were manually observed and recorded over multiple growing seasons, depending on breeder intuition and experience for decision-making.

TABLE 1
DIFFERENCES BETWEEN TRADITIONAL METHODS AND AI AND ML METHODS

Traditional Methods	AI and ML Approaches
Relied on phenotype observation, pedigree analysis, and controlled crosses to predict the performance of crop varieties	Uses the power of computational algorithms to analyze large-scale genomic and phenotypic datasets, enabling breeders to make more accurate and efficient predictions
Time-consuming, labor-intensive, and limited by the complexity of genetic interactions	Less time required
Involves field trials, where plant phenotypes were manually observed and recorded over multiple growing seasons	Involves a variety of algorithms, including neural networks, support vector machines, and random forests, to identify patterns and associations between genotypic and phenotypic information.

In contrast, AI and ML approaches use the power of computational algorithms to analyze large-scale genomic and phenotypic datasets, enabling breeders to make more accurate and efficient predictions. These methods require less time and involve a variety of algorithms, including neural networks, support vector machines, and random forests, to identify patterns and associations between genotypic and phenotypic information. AI and ML technologies provide objective, data-driven insights into genotype-phenotype connections. AI models expedite the breeding cycle by assessing the impact of specific genetic markers on phenotypic variation, hence prioritizing candidate genes for subsequent validation and functional research (Gonzalez et al., 2012).

2.2 Processing of Genomic Data and Feature Extraction:

During the pre-processing phase of genomic data management, numerous essential activities are implemented to guarantee the quality and reliability of the data for further analysis. The process starts with data cleansing, entailing the discovery and elimination of noise, mistakes, and inconsistencies within the dataset (Figure 1). Common methodologies encompass the elimination of low-quality reads, the removal of duplicate sequences, and the rectification of sequencing mistakes (Chen et al., 2018). After the data is cleansed and standardized, feature extraction is crucial for obtaining pertinent information from the genomic dataset. Feature extraction is converting raw genomic data into a collection of useful features that include biological traits and patterns pertinent to the study inquiry. Common feature extraction approaches encompass dimensionality reduction techniques, including principal component analysis (PCA), and feature selection algorithms, such as mutual information-based methods and LASSO regression (Guyon and Elisseeff, 2003; Ringner, 2008).

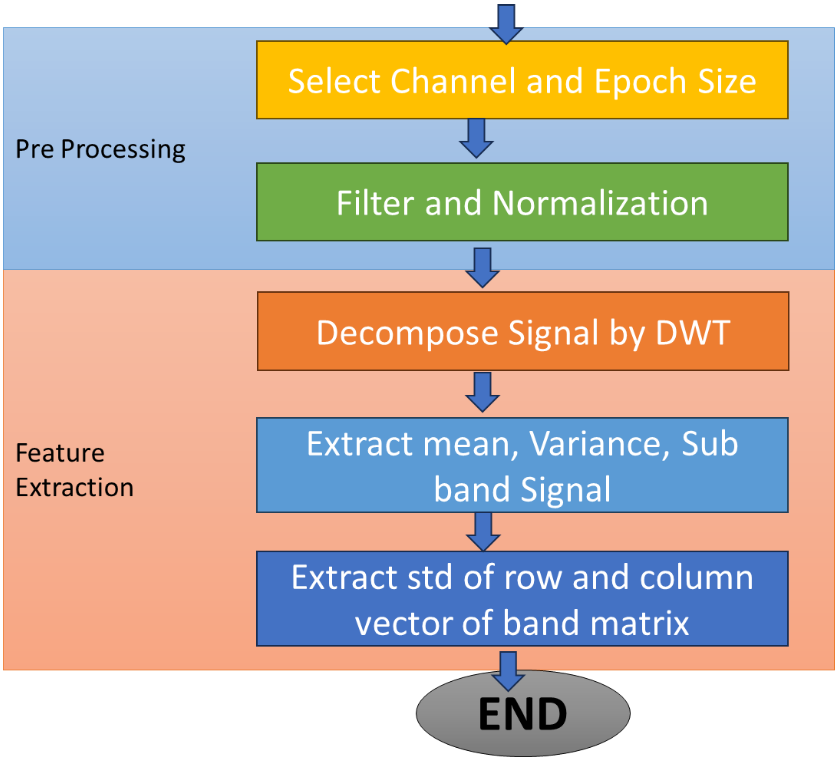

FIGURE 1: Flowchart of Pre-processing and Feature Extraction
Source: Adapted from Chum et al. (2011)

Techniques for feature selection and dimensionality reduction are essential in genomic research, since they tackle the issues associated with high-dimensional genomic information. Genomic data typically comprises thousands to millions of variables (e.g., gene expression levels, single nucleotide polymorphisms). Feature selection techniques assist in identifying the most pertinent subset of features for specific analyses, thereby diminishing computational complexity and enhancing model interpretability.

2.3 Forecasting Models and Algorithms:

A multitude of prediction models and algorithms have been utilized in genotype-phenotype prediction to elucidate the intricate connections between genetic variation and observable features in plants. Conventional regression models, including linear regression and logistic regression, function as essential instruments for forecasting phenotypic results based on genotypic information. These models offer a direct method to measure the linear correlation between genotype and phenotype, rendering them to extensively utilize in association studies and quantitative trait locus (QTL) mapping (Gianola et al., 2009).

Categories of machine learning algorithms include:

1. **Random forests** is a widely utilized ensemble learning technique that leverages the capabilities of decision trees to predict phenotypic qualities by synthesizing predictions from several trees trained on distinct data subsets (Breiman, 2001) (Figure 2).

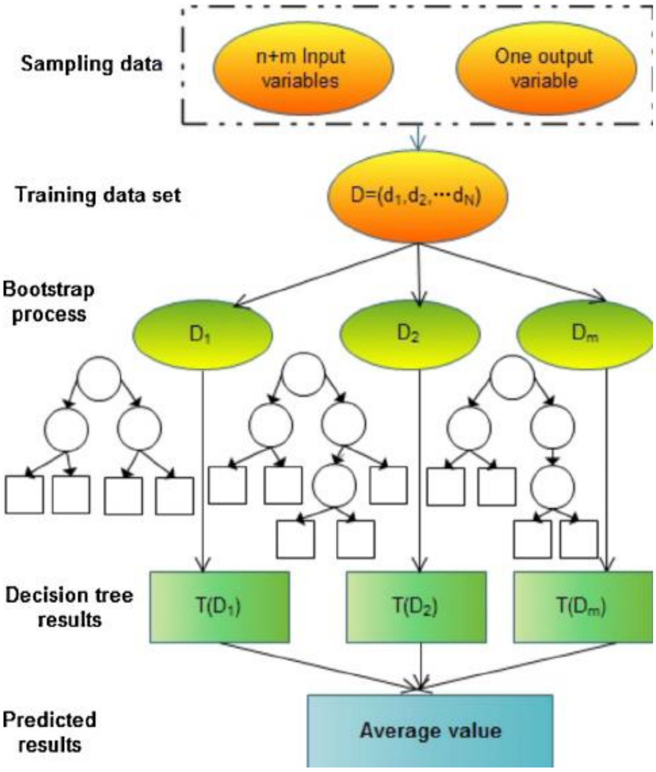

FIGURE 2: Flowchart of Random Forests

Source: Adapted from Liu et al. (2014)

2. **Neural networks**, specifically deep learning architectures, have emerged as potent instruments for elucidating complex genotype-phenotype interactions and extracting latent characteristics from genomic data.
3. **Convolutional neural networks (CNNs) and recurrent neural networks (RNNs)** have been modified for genomic sequence data, allowing researchers to forecast phenotypic outcomes from DNA sequences, gene expression profiles, and epigenetic modifications (Alipanahi et al., 2015; Angermueller et al., 2016).

III. GENOMIC SELECTION

3.1 Fundamental Principles and Concepts:

Genomic selection (GS) is an advanced breeding methodology that utilizes genomic data to forecast the breeding values of individuals for particular characteristics, hence enhancing the efficiency and precision of selection decisions in plant breeding initiatives. The predictive models employed in GS incorporate data from hundreds to millions of genetic markers and their corresponding impacts on target traits, facilitating the computation of Genomic Estimated Breeding Values (GEBVs) for each person within a breeding population (Meuwissen et al., 2001). These GEBVs function as indicators of an individual's performance for the trait of interest, enabling the discovery and prioritizing of exceptional prospects for future breeding cycles.

Diverse statistical and machine learning algorithms, including ridge regression, Bayesian methods, and other machine learning techniques, are utilized to create predictive models that consider the polygenic characteristics of traits and effectively manage high-dimensional genomic data (Habier et al., 2007; Crossa et al., 2017). Genomic selection models, by analyzing the complete genome, offer an extensive assessment of an individual's genetic potential, mitigating the effects of environmental variability and enhancing the predictability of breeding results (Goddard and Hayes, 2007).

3.2 Genomic Prediction Models Utilizing AI and ML:

A variety of advanced Artificial Intelligence (AI) and Machine Learning (ML) methods have been created to precisely forecast phenotypic outcomes based on genetic data in the field of genomic prediction.

- **Genomic Best Linear Unbiased Prediction (GBLUP)** is a statistical method that determines breeding values by applying a linear mixed model on genomic data. GBLUP employs genomic links among people, represented by marker-based kinship matrices, to estimate genetic effects throughout the genome, while considering population structure and genetic relatedness (VanRaden, 2008). Notwithstanding its simplicity, GBLUP has exhibited strong performance across several crops and features, functioning as a benchmark approach for genomic prediction.

- **The Bayesian technique** is a distinct category of algorithms utilized in genomic prediction, providing adaptability and the capacity to integrate previous information and uncertainty into the modeling framework. Bayesian regression methods, including Bayesian LASSO and Bayesian Ridge Regression, facilitate the estimation of marker effects and coefficient shrinkage, adeptly representing the polygenic structure of characteristics while mitigating overfitting (Habier et al., 2007). Moreover, Bayesian approaches enable the incorporation of supplementary information sources, like pedigree and environmental data, into the predictive model, hence improving accuracy and dependability (Perez and Misztal, 2011).

3.3 Convergence of Phenotypic and Genomic Information:

The integration of phenotypic and genomic data offers a potential approach to enhance prediction accuracy and breeding value assessment in plant breeding operations (Figure 3). Joint modelling frameworks, like Genomic Best Linear Unbiased Prediction (GBLUP) with phenotypic covariates, use phenotypic measures as fixed effects in the prediction model, facilitating the concurrent assessment of genetic and environmental influences on target traits (Heslot et al., 2015). This method improves predictive accuracy by including supplementary sources of variation and minimizing residual error, especially for characteristics affected by environmental influences or non-genetic interactions.

Jiang and Reif (2015) proposed multi-trait prediction models that use genomic and phenotypic data to estimate breeding values for multiple variables. These models use genetic correlations among characteristics to enhance prediction accuracy and capture pleiotropic effects in shared genetic architecture. Breeders can use these models to optimize agronomic performance and trait combinations in breeding populations. AI technology, including phenotypic variety of crops, SNP polymorphisms, QTL analysis, GWAS analysis, genomic selection, and genome sequencing, is used to forecast crop phenotypes and develop innovative breeding techniques.

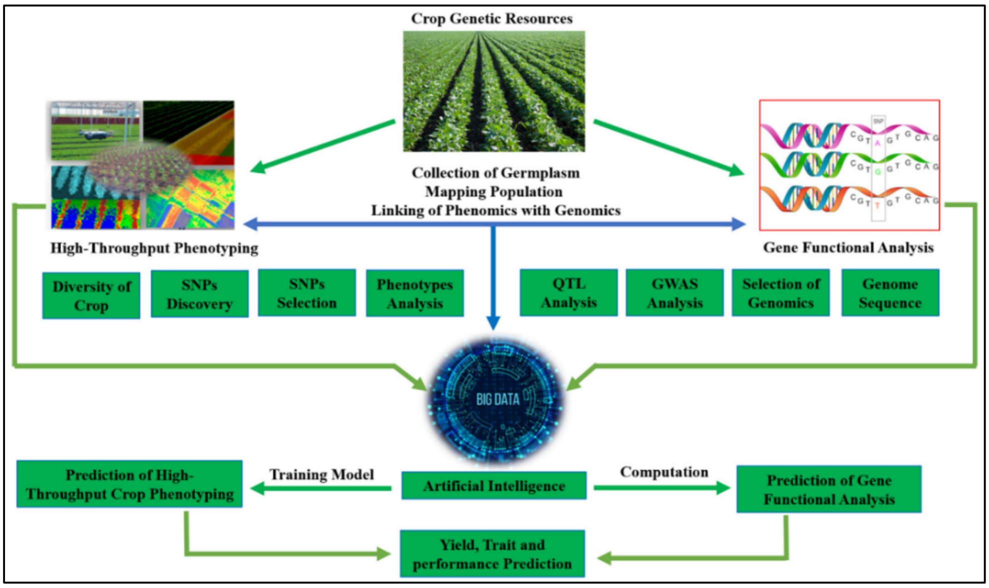

FIGURE 3. Artificial Intelligence functions as a powerful tool for forecasting high-throughput crop phenotyping and performing gene functional analysis in modern crop breeding
Source: Khan et al. (2022)

3.4 Strategies for Implementation and Case Studies:

Collard and Mackill (2008) highlight the importance of marker-assisted selection (MAS) in breeding operations to improve prediction accuracy, genetic gain, and operational efficiency. MAS uses marker data to identify individuals with advantageous alleles for desired characteristics, accelerating the breeding process. By integrating MAS with genomic selection, breeders can use both marker-based and genome-wide data for predicting breeding values, enhancing the precision and efficiency of selection.

Case studies illustrate the effective implementation of genomic selection (GS) in enhancing crop quality:

- **Maize breeding:** A study by Rincent et al. (2014) demonstrated the successful use of genomic selection (GS) in a large-scale maize breeding program, leading to a substantial enhancement in grain production and resilience to biotic and abiotic challenges. By amalgamating genetic information with phenotypic data, breeders successfully predicted breeding values for several characteristics, resulting in the creation of top maize hybrids exhibiting enhanced performance across broad environmental circumstances.
- **Wheat breeding:** A research by Rutkoski et al. (2012) demonstrated the effectiveness of genomic selection in forecasting breeding values for grain yield and resistance to Fusarium head blight in wheat breeding populations. Utilizing genetic information, breeders attained significant enhancements in grain production and disease resistance relative to traditional breeding techniques, underscoring the capacity of genomic selection to expedite trait advancement in wheat.
- **Rice breeding:** A joint initiative among researchers and breeders successfully employed genomic selection (GS) to create blast-resistant rice varieties with enhanced yield and grain quality characteristics (Spindel et al., 2015). By integrating genomic data into selection processes, breeders accelerated the creation of blast-resistant rice varieties, reducing production losses from this destructive disease and safeguarding food security in rice-dependent areas.
- **Horticultural Breeding:** Kumar et al. (2012) employed genomic selection (GS) to forecast breeding values for fruit firmness and acidity in apple breeding populations, resulting in the creation of novel apple cultivars with enhanced post-harvest storage attributes and market attractiveness. Likewise, GS has been utilized in strawberry breeding to improve disease resistance and fruit quality characteristics, allowing breeders to create high-yielding, disease-resistant cultivars with exceptional flavor and shelf life (Bassil et al., 2015).

IV. TRAIT DISCOVERY AND CHARACTERIZATION

4.1 Identification of Genetic Markers:

Artificial Intelligence (AI) and Machine Learning (ML) methodologies are essential for detecting genetic markers linked to advantageous features in plant breeding. A primary use of AI and ML in this area is Genome-Wide Association Studies (GWAS), which employ statistical learning algorithms to identify significant correlations between genetic markers and phenotypic variation in various breeding populations (Zhang et al., 2019). Machine learning methods, including logistic regression, random forests, and gradient boosting machines, are frequently employed in genome-wide association studies (GWAS) to discern marker-trait relationships, while considering population structure and mitigating false positives. Moreover, deep learning architectures, like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), provide robust methodologies for assessing genomic sequences and detecting regulatory elements and functional variations linked to certain characteristics (Zou et al., 2020). When combined with functional validation, these methodologies demonstrate significant potential for clinical translation, offering a robust evidence-based framework to guide post-GWAS research (Nicholls et al., 2020).

4.2 Integration of Multi-Omics:

The amalgamation of multi-omics data—genomics, transcriptomics, metabolomics, and epigenomics—provides a holistic methodology for trait characterization in plant breeding, enhancing the comprehension of the molecular mechanisms governing complex traits and promoting more accurate trait prediction and selection, such as abiotic stress tolerance (Figure 4). Genomic data offer insights into DNA sequence variations and genetic structure, enabling breeders to pinpoint candidate genes and genetic markers linked to desired characteristics using genome-wide association studies (GWAS) and linkage mapping (Hirsch et al., 2014). Transcriptomic data elucidate gene expression patterns and regulatory networks, revealing transcriptional alterations that contribute to trait variation and provide insights into gene regulatory mechanisms and pathways (Shi et al., 2020). Metabolomic data provide an overview of metabolic profiles and biochemical pathways, elucidating the metabolic signatures linked to trait phenotypes and identifying possible biomarkers and metabolic pathways that contribute to trait variance (Saito and Matsuda, 2010). Furthermore, epigenomic data elucidate DNA methylation, histone changes, and chromatin accessibility, hence enhancing understanding of the epigenetic control of gene expression and phenotypic diversity (Law and Jacobsen, 2010). Through the integration of multi-omics data layers, breeders can develop extensive trait networks and pinpoint essential molecular determinants and regulatory hubs that influence trait variation, facilitating enhanced trait prediction, marker-assisted selection, and focused breeding strategies in plant breeding initiatives.

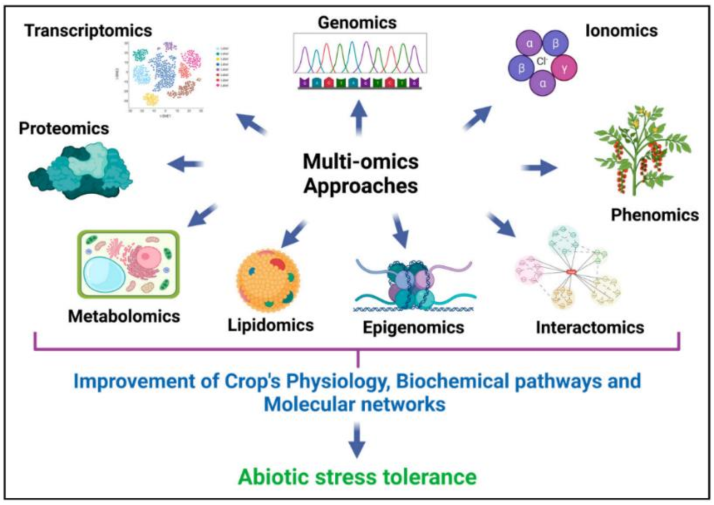

FIGURE 4: Integrative multi-omics approaches to improve abiotic stress resistance in plants
Source: Roychowdhury et al. (2023)

4.3 Artificial Intelligence and Machine Learning Strategies for Trait Prioritization:

AI-driven methodologies provide novel solutions for ranking traits according to their significance for crop enhancement and agronomic relevance, utilizing sophisticated data analytics and predictive modeling techniques to discern traits with the highest potential for genetic advancement and agricultural influence. This entails employing machine learning algorithms (MLAs), including random forests, support vector machines, and deep learning architectures, to analyze extensive phenotypic and genotypic datasets and prioritize traits according to their heritability, economic significance, and breeding goals (Crossa et al., 2017). Integrating AI-driven methodologies into trait prioritization processes enables breeders to improve the efficacy and efficiency of crop enhancement initiatives, expediting the creation of high-yielding, stress-resistant, and nutritious crop varieties designed to address the changing demands of agricultural systems and food markets.

Heslot et al. (2015) has highlighted the increasing use of machine learning algorithms in breeding programs for trait ranking and selection. These algorithms prioritize characteristics based on their agronomic significance and potential impact on crop enhancement. Supervised learning algorithms like random forests, support vector machines, and gradient boosting machines are used to develop prediction models. These models can identify essential features that significantly influence desired phenotypic results, aiding breeders in selecting viable candidates for further assessment and selection. Mohsen et al. (2021) further evaluated the efficacy of three machine learning algorithms for soybean seed yield forecasting (Figure 5).

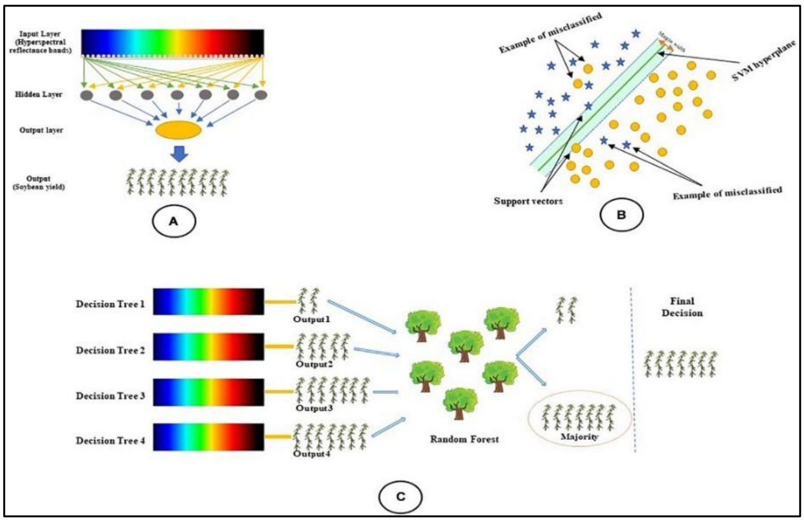

FIGURE 5: A schematic depiction of the machine learning techniques employed in this work for classifying soybean yield utilizing reflectance bands: (A) Multilayer perceptron, (B) Support vector machine, and (C) Random forest
Source: Yoosefzadeh et al. (2021)

V. OPTIMIZATION OF BREEDING SCHEMES

5.1 Models for Simulation and Optimization:

Simulation models are essential for creating and improving breeding schemes, enabling breeders to investigate various tactics, assess their effectiveness, and determine optimal decision routes to efficiently meet breeding objectives. Simulation models replicate breeding populations and cycles throughout numerous generations, encompassing genetic processes such as recombination, selection, and drift to forecast genetic gain and response to selection under various scenarios (Voss-Fels et al., 2019). These models frequently integrate genetic factors, concepts of population genetics, and breeding objectives to simulate breeding programs and evaluate their effects on trait enhancement and genetic diversity (Cooper et al., 2014).

Optimization models employ mathematical programming and computational techniques to determine optimum breeding methods and decision routes that optimize genetic gain while adhering to breeding limitations and objectives (Crossa et al., 2017). Optimization models articulate breeding challenges as mathematical optimization problems, wherein decision variables denote breeding actions (e.g., selection intensity, mating designs) and objective functions evaluate breeding objectives (e.g., maximizing genetic gain, minimizing breeding costs) (Whishart et al., 2019). These models employ optimization methods, including linear programming, mixed-integer programming, and genetic algorithms, to identify the optimal combination of breeding decisions that provide the intended breeding outcomes (Ducrocq et al., 2015). Optimization models furnish breeders with quantitative instruments to enhance breeding tactics, distribute resources efficiently, and adjust breeding programs to evolving aims and restrictions.

Artificial Intelligence and Machine Learning tools empower breeders to enhance breeding strategies and decision-making processes via optimization and simulation models. Optimization models represent breeding issues as mathematical optimization challenges, whereas simulation models replicate breeding populations and cycles to assess various tactics and situations (Cooper et al., 2014).

5.2 Decision Support Systems:

AI-driven decision support systems (DSS) have become essential tools for breeders, providing advanced solutions for genotype selection, parental line optimization, and population management in plant breeding initiatives. These systems employ AI and ML algorithms to scrutinize extensive genomic and phenotypic information, deliver actionable insights, and inform breeding decisions to optimize genetic gain and breeding efficiency.

Decision support systems provide many benefits in optimizing breeding operations and enhancing genetic improvement in plant breeding projects. A primary benefit is the capacity to easily integrate and evaluate extensive genetic and phenotypic data, allowing breeders to make informed decisions and successfully prioritize breeding goals (Hickey et al., 2017). DSS improve cooperation and communication among breeders, researchers, and stakeholders by offering open and accessible platforms for data exchange, analysis, and decision assistance (Rutkoski et al., 2019). By promoting cooperation and knowledge sharing, DSS enables breeders to use collective expertise and insights, hence improving the efficacy and influence of breeding efforts.

5.3 Expediting Breeding Cycles:

AI and ML techniques, including speed breeding, marker-assisted selection (MAS), and genomic prediction, are essential strategies for accelerating breeding cycles in contemporary plant breeding programs, allowing breeders to hasten the creation of enhanced crop varieties to satisfy increasing global food demands.

- **Speed breeding** is a technique that alters environmental variables to reduce generation durations and expedite breeding cycles, gaining recognition for its capacity to swiftly enhance breeding populations and enable high-throughput phenotyping (Watson et al., 2018). Further illustrated the effective application of speed breeding to expedite the advancement of wheat varieties exhibiting enhanced characteristics, including yield, disease resistance, and resilience to environmental challenges (Figure 6). The researchers expedited the development of elite wheat varieties by manipulating environmental conditions to shorten generation times, thereby advancing breeding populations, enabling high-throughput phenotyping, and accelerating selection decisions.

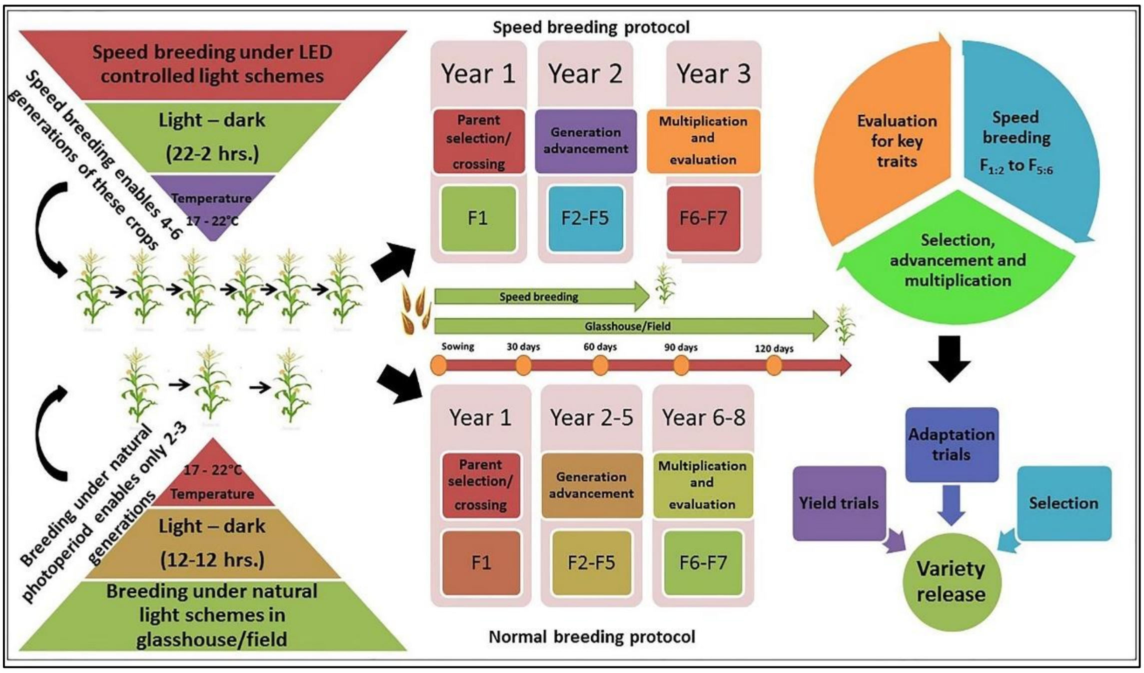

FIGURE 6: A summary of the speed breeding technique and its implications for expediting breeding cycles to enhance growth and production relative to traditional breeding methods under standard photoperiod conditions

Source: Rai (2022)

- **Marker-assisted selection (MAS)** has revolutionized plant breeding by allowing breeders to select for specific traits using molecular markers associated with target genes or quantitative trait loci (QTLs) (Collard and Mackill, 2008). Artificial Intelligence and Machine Learning methodologies augment Marker-Assisted Selection by refining marker-trait association research, streamlining marker selection, and forecasting phenotypic performance utilizing genomic data (Crossa et al., 2017). Integrating MAS with genomic prediction models enables breeders to prioritize marker-assisted crossings, expedite trait introgression, and improve selection precision, thereby reducing breeding cycles and speeding genetic gain.

- **Genomic prediction** employs genome-wide genetic markers to forecast breeding values and identify superior individuals for breeding, eliminating the necessity for comprehensive phenotyping (Heffner et al., 2009). Artificial Intelligence and Machine Learning algorithms augment genomic prediction by elucidating intricate genotype-phenotype correlations, enhancing predictive accuracy, and expediting breeding advancements (Hickey et al., 2017). Utilizing modern machine learning models, including deep learning architectures and Bayesian approaches, genomic prediction facilitates breeders in expediting selection decisions, minimizing generation intervals, and hastening variety creation in plant breeding projects.

- **Case Study:** Jarquin et al. (2014) demonstrated that by merging genomic prediction models with multi-environment trials and field data, breeders could precisely forecast breeding values for yield and other agronomic characteristics in soybean breeding populations. The implementation of genomic selection has allowed breeders to expedite genetic improvement, shorten breeding cycles, and create soybean varieties with enhanced production potential and adaptability to various conditions.

5.4 Artificial Intelligence-Enhanced Breeding Techniques:

Several scientists contributed to the advancement of AI and ML-driven breeding strategies in plant breeding. Genetic design, a sophisticated approach that combines genomic prediction models with experimental designs, has been used to enhance mating schemes and selection decisions. This approach allows breeders to efficiently manage resources, create targeted crosses, and enhance genetic variety while selecting for specific characteristics. Genetic prediction-based mating designs, as proposed by Crossa et al. (2017), use genetic information to improve mating decisions and breeding strategies. This approach aids in selecting ideal parental combinations and creating enhanced hybrids. Ducrocq et al. (2015) also highlighted the importance of optimal breeding methods, which use optimization algorithms, simulation models, and decision support systems to assess breeding situations and determine optimal breeding options. These methods can minimize generation intervals, increase selection intensity, and expedite variety creation, improving the productivity and resilience of agricultural breeding programs. Hickey et al. (2017) have highlighted the expedited genetic advancement in crop breeding operations, resulting in the identification of superior germplasm, acceleration of trait introgression, and reduced generation intervals.

VI. BEYOND CONVENTIONAL PLANT BREEDING

6.1 Robotics and Automation in Phenotyping:

Robotics and automation are essential for enhancing high-throughput phenotyping (HTP) and data acquisition in plant breeding and agricultural research. They facilitate researchers in the effective collection of extensive phenotypic data from plant populations, hence expediting breeding advancements and improving the precision of trait assessments.

- **High-Throughput Phenotyping:** A notable use of robots and automation in HTP is in plant imaging and sensing. Automated imaging platforms with high-resolution cameras, sensors, and robotic systems can accurately record precise phenotypic data, including plant growth dynamics, canopy architecture, leaf morphology, and physiological features, with high accuracy and throughput (Paulus et al., 2014). These systems provide non-destructive monitoring of plant characteristics during the development cycle, aiding in the characterization of genotype-phenotype interactions and the discovery of essential qualities for crop enhancement.
- **Data gathering:** Robotics and automation technologies are essential in optimizing data gathering processes and minimizing manual labor in both field and laboratory environments. Autonomous vehicles, drones, and robotic systems outfitted with sensors and actuators may traverse field areas, gather samples, and conduct measurements with less human involvement (Haghighattalab et al., 2016). These tools provide swift and systematic data collection over extensive field trials, improving the efficiency and scalability of phenotyping initiatives.
- **Controlled environment phenotyping (CEP):** Robotics and automation facilitate the establishment of controlled environment phenotyping (CEP) facilities, wherein environmental parameters may be meticulously regulated to replicate various growth circumstances and stress factors (Fahlgren et al., 2015). Automated growth chambers, greenhouses, and phenotyping platforms, outfitted with environmental sensors and robotic systems, allow researchers to perform reproducible experiments under controlled conditions, thereby facilitating the investigation of genotype-environment interactions and the characterization of plant responses to abiotic and biotic stresses.

6.2 Predictive Breeding for Climate Change Adaptation:

Artificial Intelligence and Machine Learning offer significant promise in predictive breeding for climate change adaptation by facilitating the creation of stress-tolerant and climate-resilient crop varieties. Given the rising frequency and severity of climate-related phenomena, including droughts, floods, heatwaves, and insect infestations, there is an urgent necessity to cultivate crops capable of enduring these adversities and flourishing in evolving climatic circumstances.

AI and ML significantly contribute to predictive breeding by analysing extensive genomic and phenotypic information to uncover genetic markers and characteristics linked to stress tolerance and climate resistance. Machine learning algorithms may utilize multi-omics data, encompassing genotypic, transcriptomic, and metabolomic information, to elucidate the genetic foundations of complex characteristics like drought tolerance, heat tolerance, and disease resistance (Singh et al., 2016).

Artificial intelligence and machine learning enable the execution of genomic selection and marker-assisted breeding techniques for climate change adaption. Utilizing genomic prediction models and marker-trait relationships, breeders may discern and choose individuals possessing advantageous alleles for stress tolerance characteristics, hence enhancing the efficiency and precision of selection decisions (Crossa et al., 2017). These methods allow breeders to swiftly incorporate desirable features into superior germplasm and create climate-resilient crop varieties suited to certain agro-ecological areas and climatic circumstances.

6.3 AI-Enhanced Gene Editing and Genetic Engineering:

The amalgamation of AI and ML methodologies with gene editing technologies, such as CRISPR-Cas9, presents substantial potential for precise trait alteration in plants, delivering unparalleled accuracy, efficiency, and rapidity in agricultural enhancement initiatives. Artificial Intelligence and Machine Learning algorithms can improve the design and optimization of CRISPR-Cas9 editing techniques, allowing researchers to forecast and prioritize target locations with high specificity and minimum off-target effects (Zhou et al., 2018). Machine learning models developed using extensive genomic datasets may discern sequence characteristics, chromatin accessibility patterns, and epigenetic markers linked to effective CRISPR-Cas9 editing, therefore facilitating the selection of appropriate target locations and enhancing editing efficiency (Kuang et al., 2019).

Moreover, AI-driven methodologies enhance the analysis and interpretation of CRISPR-Cas9 editing results, allowing researchers to evaluate the functional implications of genetic alterations and anticipate the phenotypic effects of certain gene modifications (Ding et al., 2019). Machine learning algorithms may examine multi-omics data, including transcriptomic, proteomic, and metabolomic profiles, to clarify the molecular pathways and regulatory networks that govern trait expression and to suggest potential genes for targeted change (Kandavelou et al., 2019). Integrating AI-driven analytics with CRISPR-Cas9 editing enables researchers to expedite the identification and confirmation of gene targets for trait enhancement and to elucidate the intricate genotype-phenotype interactions in plants.

VII. OBSTACLES AND CONSTRAINTS

7.1 Data Quality and Challenges:

A major difficulty is the accessibility of extensive and high-quality datasets necessary for building strong and dependable prediction models (Deb et al., 2018). To tackle these problems, researchers must use stringent data collecting processes, quality control techniques, and data integration strategies to improve the integrity and dependability of agricultural statistics. Standardizing data formats, metadata standards, and data sharing protocols can enhance data interoperability and enable the smooth integration of heterogeneous information from many sources (Araus and Cairns, 2014).

7.2 Data Standardization and its Challenges:

Data standardization is crucial for improving the dependability of AI-driven breeding models by fostering data interoperability, transparency, and reproducibility among various research environments and stakeholders. Standardized data formats, metadata standards, and ontologies enhance the integration and interchange of diverse information from many sources, allowing researchers to aggregate, harmonize, and analyse data more efficiently (Barchi et al., 2019). Standardized formats for genomic data, including Variant Call Format (VCF) and Genomic Data Format (GDF), allow researchers to systematically store and exchange genetic variation data, thereby enhancing subsequent analyses such as genotype imputation, marker-trait association, and genomic prediction (Crossa et al., 2017).

7.3 Data Dissemination and Obstacles:

Data sharing programs foster open research principles, cooperation, and information exchange throughout the scientific community (Poldrack et al., 2017). Making research data publicly accessible via repositories, databases, and data portals enables researchers to enhance the impact and reutilization of their datasets, promote multidisciplinary partnerships, and expedite scientific discovery (Hirschhorn et al., 2005). Data sharing improves the openness and reproducibility of AI-driven studies, facilitating independent validation and verification of study findings by other academics and stakeholders (Whetzel et al., 2019).

7.4 Interpretability and Transparency of Artificial Intelligence Models:

Interpreting intricate AI models in plant breeding is a considerable difficulty, especially in comprehending the fundamental principles influencing predictions. Although AI-driven models, such as deep learning neural networks, can demonstrate remarkable predictive accuracy, they are commonly perceived as "black boxes," complicating the interpretation of their reasoning behind certain findings or predictions (Rosenberg et al., 2019). This deficiency in interpretability might impede researchers' comprehension of the molecular mechanisms governing trait expression and genotype-phenotype connections in plant breeding.

Researchers are devising techniques for model interpretation, sensitivity analysis, and visualization to improve transparency and explainability in AI-driven plant breeding. Methods such feature significance analysis, saliency maps, and decision trees allow breeders to discern the most significant factors and comprehend their contributions to model predictions (Lundberg and Lee, 2017).

7.5 Computational Infrastructure and Resource Availability:

Tallis et al. (2018) and Gupta et al. (2020) emphasize the importance of computational infrastructure, including hardware, software, and technical support, for data-intensive analyses and modelling. However, insufficient financing and investment in scientific research infrastructure in developing nations hinder access to advanced computing resources and tools. Insufficient internet connectivity and digital infrastructure in rural regions further impede access to online materials, collaborative platforms, and cloud-based computing services. Deb et al. (2018) emphasize the need for coordinated actions from governments, funding bodies, and international organizations to enhance scientific capacity, upgrade digital infrastructure, and promote collaboration.

7.6 Ethical Considerations and Socio-economic Consequences:

Gong et al. (2019) and Feldman et al. (2019) highlight the ethical challenges in AI-driven plant breeding, highlighting the need for data privacy and informed consent to protect sensitive information. They emphasize the importance of equitable access to technology, including seed patents and licensing contracts, to stimulate innovation and tackle social justice issues. They also emphasize the need for fair allocation of benefits, recognizing stakeholder contributions, and advocating for free access to genetic resources to ensure the proper and fair application of AI in plant breeding.

7.7 Mitigating Bias and Inequity in AI Implementations:

Bias in AI models can occur when the training data lacks diversity or fails to reflect the populations or characteristics under examination, resulting in erroneous or inequitable predictions and results (Tallis et al., 2018). In plant breeding, biased training data can provide distorted representations of genetic diversity, resulting in inconsistencies in trait prediction and selection among various populations or geographic areas (Van Evert et al., 2020). A primary problem in mitigating bias in AI-driven breeding applications is guaranteeing the representativeness and inclusiveness of training data. Historical biases, sample mistakes, and systematic under representation of specific groups or characteristics can result in flawed models that sustain inequality and intensify current inequities (Gupta et al., 2020).

VIII. PROSPECTS AND CONCLUSIONS

8.1 Emerging Trends and Innovations:

AI and ML are revolutionizing plant breeding, enhancing genetic gain, refining breeding tactics, and tackling agricultural issues. Abadi et al. (2016) highlight the use of federated learning, meta-learning, and deep reinforcement learning in plant breeding. Vanschoren et al. (2020) emphasize the importance of leveraging prior breeding experiments for model training and optimization. Miotto et al. (2018) highlight the potential of deep reinforcement learning in enhancing breeding techniques and resource allocation. These advancements are crucial for global food security and sustainability.

8.2 Implications for Policy and Societal Consequences:

Rosenberg et al. (2019) and Tallis et al. (2018) emphasize the importance of regulatory frameworks in promoting responsible innovation and fair access to AI-driven plant breeding. They emphasize the need for explicit rules to protect farmers' rights, biodiversity, and traditional knowledge, while also facilitating technology transfer among research institutions and producers.

8.3 Inter-disciplinary Cooperation and Prospective Paths:

Multidisciplinary collaboration is crucial for enhancing AI-driven plant breeding, leveraging the expertise of plant breeders, data scientists, biotechnologists, and agronomists. Lee et al. (2019) highlight the importance of integrating genetic, genomic, and phenotypic data to create customized solutions for agronomic challenges. Kamilaris et al. (2018) highlight the role of agronomists in improving breeding techniques, evaluating prediction models, and applying research findings. Singh et al. (2016) highlight the potential of remote sensing, image sensors, and Internet of Things (IoT) devices in real-time monitoring of crop health and yield potential.

IX. CONCLUSIONS

The revolutionary capacity of AI and ML in transforming plant breeding and tackling global agriculture concerns is significant. These technologies provide unparalleled opportunity to expedite breeding cycles, improve forecast accuracy, and create crop varieties that are hardy, productive, and suited to evolving environmental circumstances. Artificial intelligence and machine learning algorithms allow breeders to examine extensive genomic and phenotypic data with greater efficiency and precision than conventional approaches. AI-driven methodologies enhance genotype-phenotype prediction, trait discovery, and breeding value calculation by detecting intricate patterns, genetic connections, and predictive models with increased accuracy and efficiency (Miotto et al., 2018). Furthermore, AI and ML methodologies enable breeders to enhance breeding methods, experimental frameworks, and resource distribution, resulting in more efficient and economical crop enhancement initiatives (Bhatia et al., 2020).

These technologies also provide potential solutions for urgent agricultural issues, including climate change, food hunger, and resource depletion. AI-driven breeding methodologies can create crop varieties with enhanced stress tolerance, disease resistance, and nutritional quality, hence providing food security and sustainability amid shifting climatic circumstances and increasing pest pressures (Tallis et al., 2018). Moreover, AI and ML methodologies empower breeders to maximize resource utilization, mitigate environmental effects, and improve agricultural output, hence facilitating the efficient and sustainable management of land, water, and inputs (Van Evert et al., 2020).

Moreover, AI and ML has the capacity to democratize access to breeding tools and resources, enabling farmers, especially in poor nations, to engage in crop enhancement initiatives and reap the advantages of technical advancements. AI-driven breeding technologies can enhance capacity building, technology transfer, and inclusive innovation by promoting open science principles, collaborative platforms, and knowledge-sharing initiatives, thereby advancing agricultural development and economic growth in rural communities (Chen et al., 2018).

In summary, AI and ML has the revolutionary capacity to reinvent plant breeding and tackle global agricultural concerns by increasing breeding efficiency, enhancing crop performance, and fostering sustainability. Utilizing these tools, researchers, breeders, and governments may expedite innovation, enhance resilience, and guarantee food security for future generations.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest

REFERENCES

[1] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., ... Zheng, X. (2016). TensorFlow: Large-scale machine learning on heterogeneous distributed systems. *arXiv*. https://arxiv.org/abs/1603.04467

[2] Acquaah, G. (2012). *Principles of plant genetics and breeding* (2nd ed.). John Wiley & Sons.

[3] Alipanahi, B., Delong, A., Weirauch, M. T., & Frey, B. J. (2015). Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nature Biotechnology, 33*(8), 831–838.

[4] Angermueller, C., Pärnamaa, T., Parts, L., & Stegle, O. (2016). Deep learning for computational biology. *Molecular Systems Biology, 12*(7), Article 878.

[5] Araus, J. L., & Cairns, J. E. (2014). Field high-throughput phenotyping: The new crop breeding frontier. *Trends in Plant Science, 19*(1), 52–61.

[6] Barchi, L., Pietrella, M., Venturini, L., Minio, A., Toppino, L., Acquadro, A., & Lanteri, S. (2019). A genome-wide association analysis of the anthocyanin and brix content in wild and cultivated strawberry highlights the diversification of their metabolic pathways. *BMC Plant Biology, 19*(1), 1–17.

[7] Bassil, N. V., Davis, T. M., Zhang, H., Ficklin, S., Mittmann, M., Webster, T., & Main, D. (2015). Development and preliminary evaluation of a 90 K Axiom® SNP array for the allo-octoploid cultivated strawberry *Fragaria × ananassa*. *BMC Genomics, 16*(1), 1–14.

[8] Bhatia, A., Hasan, M. M., Palchowdhury, S., Paul, S., & Asif, M. (2020). A review on machine learning approaches for breeding trait prediction and recommendation in crop plants. *Computers and Electronics in Agriculture, 178*, Article 105743.

[9] Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

[10] Chen, Y., Wang, Z., Li, Y., Truong, E., & Banerjee, A. (2018). Meta-transfer learning for few-shot learning. *Proceedings of the 35th International Conference on Machine Learning, 80*, 4343–4352.

[11] Chum, P., Park, S., Ko, K., & Sim, K. (2011). Optimal EEG feature extraction using DWT for classification of imagination of hands movement. *Journal of Korean Institute of Intelligent Systems, 21*(6), 786–791.

[12] Collard, B. C., & Mackill, D. J. (2008). Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1491), 557–572.

[13] Cooper, M., Gho, C., Leafgren, R., Tang, T., & Messina, C. (2014). Breeding drought-tolerant maize hybrids for the US corn-belt: Discovery to product. *Journal of Experimental Botany, 65*(21), 6191–6204.

[14] Crossa, J., Pérez, P., Cuevas, J., López, M. O., Jarquín, D., de los Campos, G., & Burgueño, J. (2017). Genomic selection in plant breeding: Methods, models, and perspectives. *Trends in Plant Science, 22*(11), 961–975.

[15] Deb, D., Garg, T., Mahajan, A., & Sharma, D. (2018). Quality attributes and quality assessment techniques of open data: A comprehensive review. *Data Technologies and Applications, 52*(1), 54–79.

[16] Ding, W., Li, Y., & Gao, Y. (2019). Using machine learning models to select CRISPR-Cas9 target sites based on chromatin accessibility. *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, 96–103.

[17] Ducrocq, V., Besbes, B., & Casellas, J. (2015). Genetic algorithms for optimization in livestock management and breeding. *Genetics Selection Evolution, 47*(1), Article 54.

[18] Fahlgren, N., Gehan, M. A., & Baxter, I. (2015). Lights, camera, action: High-throughput plant phenotyping is ready for a close-up. *Current Opinion in Plant Biology, 24*, 93–99.

[19] Feldman, M. W., Khoury, C. K., & van de Wouw, M. (2019). Agricultural biotechnology and intellectual property: Seeds of change. *Nature Plants, 5*(11), 1142–1146.

[20] Gianola, D., Okut, H., Weigel, K. A., Rosa, G. J., & Bacheller, L. R. (2009). Predicting complex quantitative traits with Bayesian neural networks: A case study with Jersey cows and wheat. *BMC Genetics, 10*(1), Article 37.

[21] Goddard, M. E., & Hayes, B. J. (2007). Genomic selection. *Journal of Animal Breeding and Genetics, 124*(6), 323–330.

[22] Golzarian, M. R., Frick, R. A., Rajendran, K., Berger, B., Roy, S., Tester, M., & Lun, D. S. (2011). Accurate inference of shoot biomass from high-throughput images of cereal plants. *Plant Methods, 7*(1), 1–11.

[23] Gong, L., Sun, X., Sui, L., & Zhao, H. (2019). Ethical considerations in AI-driven breeding. In O. B. Smith (Ed.), *Ethics in agriculture – An African perspective* (pp. 103–121). Springer.

[24] González-Camacho, J. M., de los Campos, G., Pérez, P., Gianola, D., Cairns, J. E., Mahuku, G., & Crossa, J. (2012). Genome-enabled prediction of genetic values using radial basis function neural networks. *Theoretical and Applied Genetics, 125*(4), 759–771.

[25] Gupta, R., Bhatt, S., & Patel, K. (2020). Digital divide in India: A systematic literature review. *Telematics and Informatics, 56*, Article 101456.

[26] Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research, 3*, 1157–1182.

[27] Habier, D., Fernando, R. L., Kizilkaya, K., & Garrick, D. J. (2007). Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics, 8*(1), 1–15.

[28] Haghighattalab, A., González Pérez, L., Mondal, S., Singh, D., Schinstock, D., Rutkoski, J., & Poland, J. (2016). Application of unmanned aerial systems for high throughput phenotyping of large wheat breeding nurseries. *Plant Methods, 12*(1), Article 35.

[29] Heffner, E. L., Jannink, J. L., & Sorrells, M. E. (2009). Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *The Plant Genome, 2*(2), 191–197.

[30] Heslot, N., Yang, H. P., Sorrells, M. E., & Jannink, J. L. (2015). Genomic selection in plant breeding: A comparison of models. *Crop Science, 55*(1), 36–45.

[31] Hickey, L. T., Hafeez, A. N., Robinson, H., Jackson, S. A., Leal-Bertioli, S. C. M., Tester, M., & Dieters, M. J. (2017). Breeding crops to feed 10 billion. *Nature Biotechnology, 35*(10), 927–937.

[32] Hirsch, C. N., Foerster, J. M., Johnson, J. M., Sekhon, R. S., Muttoni, G., Vaillancourt, B., & de Leon, N. (2014). Insights into the maize pan-genome and pan-transcriptome. *The Plant Cell, 26*(1), 121–135.

[33] Hirschhorn, J. N., & Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics, 6*(2), 95–108.

[34] Jarquín, D., Crossa, J., Lacaze, X., Du Cheyron, P., Daucourt, J., Lorgeou, J., & Gay, L. (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theoretical and Applied Genetics, 127*(3), 595–607.

[35] Jiang, Y., & Reif, J. C. (2015). Modeling epistasis in genomic selection. *Genetics, 201*(2), 759–768.

[36] Kadam, S., & Yin, X. (2020). Applications of artificial intelligence in plant breeding. *Trends in Plant Science, 25*(7), 672–680.

[37] Kamilaris, A., Prenafeta-Boldú, F. X., & Fountas, S. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture, 147*, 70–90.

[38] Kandavelou, K., Manivannan, K., Balasubramanian, S., & Lakshmanan, G. (2019). Artificial intelligence in plant genome editing. In P. K. Gupta (Ed.), *Artificial intelligence applications in genetic improvement* (pp. 83–97). Elsevier.

[39] Khan, M. H. U., Wang, S., Wang, J., Ahmar, S., Saeed, S., Khan, S. U., Xu, X., Chen, H., Bhat, J. A., & Feng, X. (2022). Applications of artificial intelligence in climate-resilient smart-crop breeding. *International Journal of Molecular Sciences, 23*(19), Article 11156.

[40] Rai, K. K. (2022). Integrating speed breeding with artificial intelligence for developing climate-smart crops. *Molecular Biology Reports, 49*(12), 11385–11402.

[41] Kuang, Z., Ping, Y., Hao, Y., Fang, Z., Li, J., & Yin, H. (2019). Highly efficient RNA-guided base editing in rabbit. *Nature Communications, 10*(1), 1–10.

[42] Kumar, S., Garrick, D. J., Bink, M. C., Whitworth, C., & Chagné, D. (2012). Genomic selection for fruit quality traits in apple (*Malus × domestica* Borkh.). *PLoS ONE, 7*(5), Article e36674.

[43] Law, J. A., & Jacobsen, S. E. (2010). Establishing, maintaining, and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics, 11*(3), 204–220.

[44] Lee, K., Zhang, Y., & Kleinstiver, B. P. (2019). Activities and specificities of CRISPR/Cas9 and Cas12a nucleases for targeted mutagenesis in maize. *Plant Biotechnology Journal, 17*(2), 362–372.

[45] Liu, M., Liu, X., Li, J., Ding, C., & Jiang, J. (2014). Evaluating total inorganic nitrogen in coastal waters through fusion of multi-temporal RADARSAT-2 and optical imagery using random forest algorithm. *International Journal of Applied Earth Observation and Geoinformation, 33*, 192–202.

[46] López-Cruz, M. A., Crossa, J., de los Campos, G., Alvarado, G., Mondal, S., & Rutkoski, J. (2018). A review of computational tools for prediction and improvement of crop traits. *Frontiers in Plant Science, 10*, Article 422.

[47] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*, 4765–4774.

[48] Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics, 157*(4), 1819–1829.

[49] Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: Review, opportunities and challenges. *Briefings in Bioinformatics, 19*(6), 1236–1246.

[50] Yoosefzadeh-Najafabadi, M., Earl, H., Tulpan, D., Sulik, J., & Eskandari, M. (2021). Application of machine learning algorithms in plant breeding: Predicting yield from hyperspectral reflectance in soybean. *Frontiers in Plant Science, 11*, Article 2169.

[51] Mulla, D. J., Joshi, N., Chang, A., Abel, K., & Venugopal, S. (2020). Automation and robotics in plant breeding: Current status and future prospects. *Plant Phenomics, 2020,* Article 8849382.

[52] Nicholls, H. L., John, C. R., Watson, D. S., Munroe, P. B., Barnes, M. R., & Cabrera, C. P. (2020). Reaching the end-game for GWAS: Machine learning approaches for the prioritization of complex disease loci. *Frontiers in Genetics, 11*, Article 350.

[53] Paulus, S., Dupuis, J., Mahlein, A. K., Kuhlmann, H., & Kersting, K. (2014). Automatic early plant disease detection using machine learning-based image analysis. *Plant Pathology, 63*(6), 1302–1312.

[54] Pérez-Enciso, M., & Misztal, I. (2011). Qxpak5: Old mixed model solutions for new genomics problems. *BMC Bioinformatics, 12*(1), Article 202.

[55] Poldrack, R. A., Gorgolewski, K. J., & Varoquaux, G. (2017). Computational and informatic advances for reproducible data analysis in neuroimaging. *Annual Review of Biomedical Engineering, 19*, 259–287.

[56] Rincent, R., Charpentier, J. P., Faivre-Rampant, P., Paux, E., Le Gouis, J., Bastien, C., & Moreau, L. (2014). Phenomic selection is a low-cost and high-throughput method based on indirect predictions: Proof of concept on wheat and poplar. *G3: Genes, Genomes, Genetics, 4*(8), 1603–1610.

[57] Ringnér, M. (2008). What is principal component analysis? *Nature Biotechnology, 26*(3), 303–304.

[58] Rosenberg, N. A., Sella, G., & Blum, M. G. (2019). Interpreting polygenic scores, polygenic adaptation, and human phenotypic differences. In M. A. Goldman (Ed.), *Evolutionary medicine* (pp. 115–125). Oxford University Press.

[59] Roychowdhury, R., Das, S. P., Gupta, A., Parihar, P., Chandrasekhar, K., Sarker, U., Kumar, A., Ramrao, D. P., & Sudhakar, C. (2023). Multi-omics pipeline and omics-integration approach to decipher plant's abiotic stress tolerance responses. *Genes, 14*(6), Article 1281.

[60] Rutkoski, J., Benson, J., Jia, Y., Brown-Guedira, G., Jannink, J. L., & Sorrells, M. (2012). Evaluation of genomic prediction methods for Fusarium head blight resistance in wheat. *The Plant Genome, 5*(2), 51–61.

[61] Rutkoski, J., Poland, J., Mondal, S., Autrique, E., Pérez, L. G., & Crossa, J. (2019). Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *G3: Genes, Genomes, Genetics, 9*(7), 2495–2508.

[62] Saito, K., & Matsuda, F. (2010). Metabolomics for functional genomics, systems biology, and biotechnology. *Annual Review of Plant Biology, 61*, 463–489.

[63] Shi, T., Shi, L., Fang, H., Weng, Z., & Schadt, E. E. (2020). Investigating and suppressing batch effects in single-cell RNA-Seq data. *Genome Biology, 21*(1), 1–19.

[64] Singh, A., Ganapathysubramanian, B., Singh, A. K., & Sarkar, S. (2016). Machine learning for high-throughput stress phenotyping in plants. *Trends in Plant Science, 21*(2), 110–124.

[65] Spindel, J., Begum, H., Akdemir, D., Collard, B., Redoña, E., Jannink, J. L., & McCouch, S. R. (2015). Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement. *Heredity, 116*(4), 395–408.

[66] Tallis, H., Kreis, K., O'Hare, M., O'Connell, D., Hawkins, E., Folarin, A., & Rahman, M. (2018). *The role of digital agriculture in food security*. The Nature Conservancy.

[67] Tanksley, S. D., & McCouch, S. R. (1997). Seed banks and molecular maps: Unlocking genetic potential from the wild. *Science, 277*(5329), 1063–1066.

[68] van Evert, F. K., Franke, A. C., de Rooij-van der Goes, P. C., van Laar, H. H., & Schans, D. A. (2020). Applications of AI and machine learning in agricultural systems: A review. *Engineering Applications of Artificial Intelligence, 91*, Article 103529.

[69] VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science, 91*(11), 4414–4423.

[70] Vanschoren, J., van Rijn, J. N., Bischl, B., & Torgo, L. (2020). OpenML: Networked science in machine learning. *ACM Transactions on Intelligent Systems and Technology, 11*(3), 1–36.

[71] Voss-Fels, K. P., Cooper, M., Hayes, B. J., & Wu, H. (2019). Genomic selection in plant breeding programs: A comparison of simulation models and field experiments in wheat. *Frontiers in Plant Science, 10,* Article 179.

[72] Watson, A., Ghosh, S., Williams, M. J., Cuddy, W. S., Simmonds, J., Rey, M. D., & Hickey, L. T. (2018). Speed breeding is a powerful tool to accelerate crop research and breeding. *Nature Plants, 4*(1), 23–29.

[73] Whetzel, P. L., Noy, N. F., Shah, N. H., Alexander, P. R., Nyulas, C., Tudorache, T., & Musen, M. A. (2019). BioPortal: Ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Research, 37*(Suppl. 1), W170–W173.

[74] Whishart, J., Arief, V. N., & Smith, A. B. (2019). Operations research in agriculture: A review. *Computers and Electronics in Agriculture, 157*, 8–27.

[75] Yoosefzadeh-Najafabadi, M., Earl, H., Tulpan, D., Sulik, J., & Eskandari, M. (2021). Application of machine learning algorithms in plant breeding: Predicting yield from hyperspectral reflectance in soybean. *Frontiers in Plant Science, 11*, Article 2169.

[76] Zhang, X., Pérez-Rodríguez, P., Burgueño, J., Olsen, M., Buckler, E., Atlin, G., & Crossa, J. (2019). Rapid cycling genomic selection in a multiparental tropical maize population. *G3: Genes, Genomes, Genetics, 9*(7), 2299–2312.

[77] Zhou, H., Cheng, N., & Yu, X. (2018). Utilizing machine learning approaches for precise and fast CRISPR editing. *Trends in Biotechnology, 36*(10), 1017–1027.

[78] Zou, J., Huss, M., Abid, A., Mohammadi, P., Torkamani, A., & Telenti, A. (2020). A primer on deep learning in genomics. *Nature Genetics*, *52*(1), 12–18.