

Chapter 16

Data-Driven Decision Support in Breeding Programs

S Harishkumar^{*1}, T. Sterlin² and Thirumalraj³

¹*B.Sc. (Hons) Agriculture, Vels Institute of Science and Technology Advanced Studies (VISTAS)*

²*B.Sc. (Hons) Agriculture, Vels Institute of Science and Technology Advanced Studies (VISTAS)*

³*Assistant Professor, Department of Genetics and Plant Breeding, Vels Institute of Science and Technology Advanced Studies (VISTAS)*

***Corresponding Author Email:** harishsenthil2401005@gmail.com

Abstract

Data-driven decision support systems have fundamentally transformed the landscape of modern breeding programs, providing a framework for integrating vast and complex datasets into actionable insights that accelerate crop improvement. The convergence of genomics, phenomics, environmental monitoring, and computational analytics has enabled breeders to make informed decisions with unprecedented precision and efficiency, moving beyond traditional trial-and-error approaches. High-throughput phenotyping platforms, combined with UAVs, satellite imagery, and sensor networks, allow continuous monitoring of plant growth, stress responses, and trait expression across multiple environments, producing high-resolution datasets that feed into predictive models. Genomic selection and machine learning techniques have enhanced the ability to predict breeding values, optimize parent selection, and identify superior lines early in the breeding cycle, reducing both time and resource requirements. Decision support tools, including open-source and proprietary platforms, provide user-friendly interfaces that facilitate real-time analysis, visualization, and scenario simulation, empowering breeders to design optimized crossing schemes, advance or cull lines strategically, and allocate resources efficiently. The integration of emerging technologies such as artificial intelligence, digital twins, and blockchain ensures secure, traceable, and transparent management of breeding data while enabling autonomous and adaptive breeding strategies. Climate-resilient breeding is increasingly guided by predictive analytics, allowing programs to anticipate genotype-by-environment interactions and develop varieties capable of withstanding abiotic stresses such as drought, heat, and salinity.

Keywords: *Data-driven breeding, Genomic selection, High-throughput phenotyping, Predictive analytics*

Introduction to Data-Driven Decision Support in Breeding Programs

Definition and scope of data-driven decision support

Data-driven decision support in breeding programs refers to the systematic use of quantitative data, computational tools, and predictive models to guide key breeding decisions such as selection, crossing, and advancement of lines. It integrates diverse datasets including genomic markers, phenotypic traits, and environmental variables to enhance decision accuracy. The core concept involves converting large volumes of raw data into actionable insights using statistical and machine learning approaches. The scope spans the entire breeding pipeline, from germplasm characterization to multi-location trials. With high-throughput sequencing technologies generating massive datasets, structured decision support systems have become essential for efficient data handling and interpretation. The main objective is to increase genetic gain per unit time by improving selection accuracy and reducing breeding cycle duration. Research evidence suggests that genomic-enabled decision systems can significantly outperform traditional selection approaches, particularly for complex traits controlled by multiple genes. These systems also enable better handling of genotype-by-environment interactions, improving the stability of selected cultivars across diverse conditions.

Evolution from conventional to data-intensive breeding approaches

Breeding methodologies have undergone a major transformation from phenotype-based selection to data-intensive approaches driven by genomic and computational advances (James *et.al.*, 2016). Earlier breeding programs relied heavily on visual selection and repeated field evaluations, often requiring many years to develop a new cultivar. The introduction of molecular markers marked a critical transition, enabling indirect selection for traits governed by major genes. Advances in sequencing technologies drastically reduced the cost of genotyping, making genome-wide analysis feasible. This led to the adoption of genomic selection, where thousands of markers are used simultaneously to predict breeding values. Technological developments such as high-throughput phenotyping, remote sensing, and automated data collection systems have further expanded data availability. These innovations allow breeders to capture detailed spatial and temporal variation in crop performance. As a result, breeding cycles have been shortened, and selection decisions have become more precise. Modern breeding programs now operate as dynamic systems that continuously integrate new data to refine predictions and improve outcomes.

Importance in modern crop improvement under climate and resource constraints

Crop improvement efforts today must address challenges such as climate variability, declining natural resources, and increasing global food demand. Data-driven decision support systems provide a powerful framework for tackling these issues by enabling precise and efficient selection strategies. By combining genomic, phenotypic, and environmental data, breeders can identify genotypes with improved

tolerance to stresses such as drought, heat, and salinity. Predictive modeling allows simulation of crop performance under varying environmental scenarios, supporting the development of resilient cultivars. Resource limitations such as water scarcity and nutrient constraints require crops with enhanced efficiency traits, which can be effectively identified using integrated data analysis. Studies have shown that data-assisted breeding approaches can accelerate yield improvement rates compared to conventional methods. These systems also optimize experimental design and resource allocation, reducing costs while improving accuracy. The ability to make informed, data-backed decisions contributes significantly to sustainable agricultural intensification and long-term food security.

Precision genetics and its integration with decision support systems

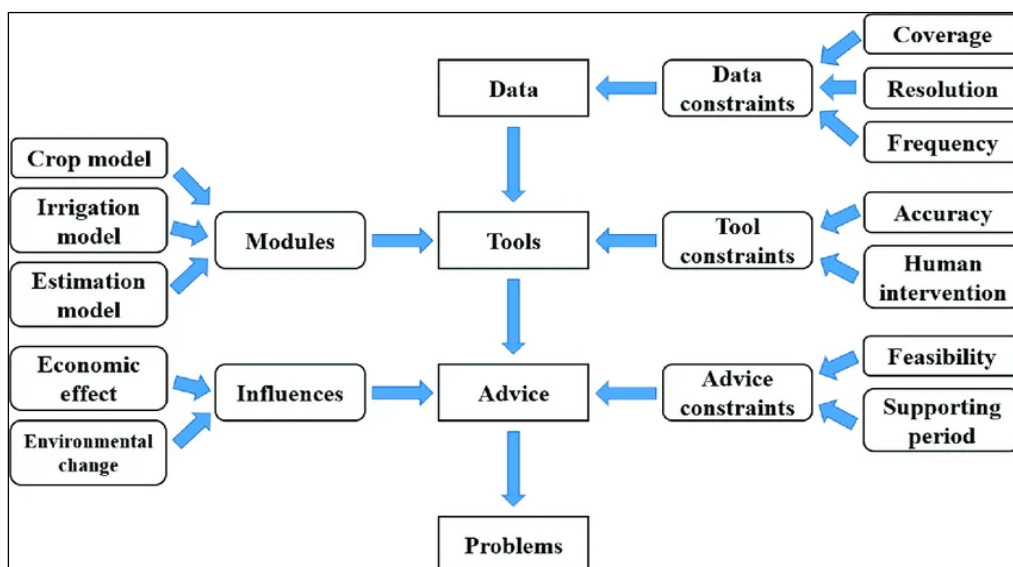
Precision genetics focuses on the targeted analysis and utilization of genetic variation to improve crop performance with high accuracy. It combines genomic tools, phenotyping technologies, and computational analytics to dissect complex traits and identify beneficial alleles. The integration of precision genetics with decision support systems creates a powerful framework for informed breeding decisions. Genomic prediction models use large-scale marker data to estimate breeding values, enabling early and accurate selection of superior genotypes. Techniques such as genome-wide association studies help identify genetic regions linked to important traits, while gene editing technologies provide opportunities for precise genetic modifications. Decision support platforms incorporate these tools to analyze complex datasets and generate recommendations tailored to specific breeding objectives. This integration allows breeders to consider multiple factors simultaneously, including genetic potential, environmental adaptability, and management conditions. The combined use of precision genetics and data-driven decision systems enhances breeding efficiency, reduces uncertainty, and accelerates the development of improved crop varieties suited to diverse and changing environments.

Foundations of Decision Support Systems (DSS) in Plant Breeding

Key components of DSS: data, models, and interfaces

Decision Support Systems in plant breeding are structured around three fundamental components: data, analytical models, and user interfaces. Data forms the backbone and includes genotypic, phenotypic, environmental, and management datasets collected across breeding cycles. Modern breeding programs generate millions of data points per season through high-throughput phenotyping platforms and next-generation sequencing technologies. These datasets require robust storage and preprocessing pipelines to ensure quality and usability. Models represent the analytical core, transforming raw data into predictions and recommendations. Statistical approaches such as mixed linear models and genomic prediction algorithms are widely used to estimate breeding values and selection indices. The increasing use of machine learning models has improved prediction accuracy for

complex traits by capturing non-linear relationships. Interfaces serve as the interaction layer between breeders and the system, providing visualization tools, dashboards, and decision outputs in an interpretable format. Effective interfaces enhance usability by simplifying complex analyses into actionable insights. The synergy among these components determines the efficiency of a DSS, enabling breeders to make timely and informed decisions while managing large-scale datasets with precision.



Types of DSS used in agriculture and breeding

Various types of Decision Support Systems are employed in agriculture and plant breeding, each designed to address specific decision-making needs. Model-driven DSS rely on mathematical and statistical models to simulate biological processes and predict outcomes such as yield performance or trait inheritance. Data-driven DSS focus on extracting patterns and relationships from large datasets using data mining and machine learning techniques. Knowledge-driven DSS incorporate expert knowledge, rules, and heuristics to guide breeding decisions, often used in trait selection and disease management. Communication-driven DSS facilitate collaboration among breeders, researchers, and stakeholders by integrating shared data platforms and decision tools. Document-driven DSS manage and retrieve structured and unstructured information, supporting literature-based decision-making. In breeding programs, hybrid systems that combine multiple DSS types are increasingly common, offering greater flexibility and analytical power. For example, integrating crop simulation models with genomic prediction tools allows breeders to evaluate genotype performance across diverse environments. The adoption of cloud-based DSS platforms has further expanded accessibility, enabling real-time data analysis and decision-making across geographically distributed breeding programs.

Role of artificial intelligence and machine learning

Artificial intelligence and machine learning have become central to modern Decision Support Systems in plant breeding by enabling the analysis of complex, high-dimensional datasets. Machine learning algorithms such as random forests, support vector machines, and neural networks are widely used to predict phenotypic traits from genomic and environmental data. These approaches can handle non-linear interactions and large feature sets more effectively than traditional statistical models. Deep learning techniques, particularly convolutional neural networks, are increasingly applied in image-based phenotyping to assess traits such as canopy structure, disease severity, and biomass. AI-driven models can process data from drones, satellites, and field sensors, generating detailed insights into crop performance. Studies indicate that machine learning models can improve prediction accuracy for complex traits by 10–20% compared to conventional methods under certain conditions. AI also supports optimization problems in breeding, such as selecting optimal parent combinations and designing efficient crossing schemes. The integration of AI into DSS enables automated data analysis, reducing manual effort and accelerating decision-making processes. As computational power continues to grow, AI is expected to play an even greater role in enhancing the precision and scalability of breeding programs.

Integration with breeding workflows and pipelines

The integration of Decision Support Systems into breeding workflows transforms traditional breeding pipelines into data-driven, iterative processes (Zhang *et.al.*, 2025). Modern breeding programs consist of multiple stages, including germplasm evaluation, crossing, selection, and multi-environment testing. DSS tools are embedded at each stage to support decision-making through predictive analytics and data visualization. For instance, genomic prediction models are used during early selection stages to identify promising lines before extensive field testing, reducing time and resource requirements. During field trials, DSS platforms integrate phenotypic and environmental data to assess genotype performance and stability. These systems also support resource allocation by optimizing trial design, location selection, and replication strategies. Integration is facilitated through digital breeding platforms that connect databases, analytical tools, and user interfaces into a unified system. This allows seamless data flow across different stages of the breeding cycle. The result is a continuous feedback loop where new data refine existing models, improving prediction accuracy over time. Such integration enhances efficiency, shortens breeding cycles, and increases the likelihood of developing superior crop varieties adapted to diverse environmental conditions.

Types and Sources of Data in Modern Breeding Programs*Genotypic data: SNPs, sequencing, and marker platforms*

Genotypic data forms a foundational component of modern breeding programs by providing detailed information on the genetic makeup of plant materials. Single

nucleotide polymorphisms (SNPs) are the most widely used markers due to their abundance and distribution across the genome, with high-density SNP arrays capable of generating tens of thousands to millions of markers per sample. Advances in next-generation sequencing technologies such as genotyping-by-sequencing (GBS) and whole-genome sequencing have significantly reduced costs, making large-scale genotyping feasible. Marker platforms such as Illumina and Affymetrix arrays enable high-throughput and reproducible data generation. These datasets are used to construct linkage maps, perform genome-wide association studies, and implement genomic selection models. The availability of reference genomes for major crops has enhanced the accuracy of variant detection and annotation. Genotypic data allows breeders to estimate genetic relationships, identify quantitative trait loci, and track favorable alleles across generations. Integration with analytical models enables prediction of breeding values with high precision, particularly for complex traits controlled by multiple genes. The scale and resolution of genotypic data continue to expand, driving the shift toward predictive and precision-based breeding strategies.

Phenotypic data: field trials, high-throughput phenotyping, and sensor data

Phenotypic data captures observable traits such as yield, plant height, disease resistance, and physiological characteristics, serving as a critical link between genotype and performance. Traditional field trials remain essential, involving replicated experiments across multiple locations and seasons to assess genotype performance under varying conditions. These trials generate large datasets that reflect genotype-by-environment interactions. High-throughput phenotyping technologies have transformed data collection by enabling rapid, non-destructive measurement of traits using imaging systems, drones, and ground-based sensors. Platforms equipped with RGB, hyperspectral, and thermal cameras can capture detailed information on plant growth, canopy structure, and stress responses. Sensor-based data collection allows continuous monitoring of plant traits, increasing temporal resolution and reducing human error. For example, UAV-based phenotyping can assess thousands of plots within hours, compared to manual methods that require several days. The integration of phenotypic data with genomic information enhances the accuracy of predictive models. Despite these advances, challenges remain in data standardization and processing due to the volume and complexity of phenotypic datasets. Effective utilization of phenotypic data is essential for improving selection decisions and achieving higher genetic gains.

Environmental and climatic data: weather, soil, and geospatial layers

Environmental and climatic data play a crucial role in understanding the performance of genotypes across diverse conditions. Weather variables such as temperature, rainfall, humidity, and solar radiation directly influence crop growth and development. Soil properties including texture, nutrient content, pH, and water-holding capacity also affect plant performance. Geospatial data layers, derived from

satellite imagery and geographic information systems, provide spatial context for field trials and breeding environments. These datasets enable detailed characterization of target environments and help identify patterns of genotype adaptation. Integration of environmental data with phenotypic and genotypic information supports the analysis of genotype-by-environment interactions, which are critical for selecting stable and adaptable varieties. Crop simulation models often use environmental inputs to predict performance under different scenarios, including future climate conditions. The increasing availability of real-time weather data and remote sensing technologies has improved the precision of environmental characterization. Accurate environmental data allows breeders to design trials more effectively and select genotypes suited to specific agro-ecological zones, enhancing the relevance and impact of breeding programs.

Management and agronomic practice data

Management and agronomic practice data provide essential context for interpreting breeding trial results and optimizing crop performance. These data include information on planting dates, irrigation schedules, fertilizer application, pest and disease control measures, and tillage practices. Variations in management practices can significantly influence phenotypic expression, making it important to record and integrate these factors into analysis. Digital tools and farm management systems now enable precise documentation of agronomic practices, improving data accuracy and accessibility. For example, sensor-based irrigation systems can generate detailed records of water application, while precision agriculture technologies track input usage at fine spatial scales. Incorporating management data into decision support systems allows breeders to evaluate genotype performance under specific management regimes. This is particularly important for developing varieties tailored to different production systems, such as low-input or intensive farming. Analytical models that include management variables can better explain variation in crop performance and improve prediction accuracy. Effective use of management data supports the development of integrated breeding strategies that consider both genetic potential and agronomic conditions.

Socio-economic and market-driven datasets

Socio-economic and market-driven datasets are increasingly important in aligning breeding objectives with end-user needs and market demands. These data include information on consumer preferences, market prices, value chain requirements, and adoption patterns of crop varieties.

Data Management and Integration

Data collection protocols and standardization

Effective data management in breeding programs begins with well-defined data collection protocols and standardization procedures that ensure consistency, accuracy, and reproducibility across experiments (Dipta *et.al.*, 2023). Standard

operating procedures guide the collection of genotypic, phenotypic, and environmental data, minimizing variability introduced by human error or differing methodologies. Standardization involves the use of controlled vocabularies, ontologies, and measurement units, which enable uniform interpretation of traits across locations and time periods. Ontologies such as the Crop Ontology and Plant Trait Ontology provide structured frameworks for trait description, facilitating data integration and comparison. Digital tools such as field data collection apps and barcode-based sample tracking systems enhance precision and reduce transcription errors. High-throughput phenotyping platforms generate large volumes of standardized data by using calibrated sensors and automated workflows. Consistent protocols are especially critical in multi-location trials, where variations in data collection methods can obscure true genetic differences. Studies show that standardized phenotyping can reduce experimental error by up to 20%, improving the reliability of downstream analyses. Clear documentation of protocols also supports reproducibility and data sharing, making datasets more valuable for long-term breeding objectives.

Data storage systems: databases, cloud platforms, and repositories

Modern breeding programs rely on advanced data storage systems to handle the vast amounts of information generated from genomics, phenomics, and environmental monitoring. Relational databases such as MySQL and PostgreSQL are commonly used to store structured breeding data, enabling efficient querying and management. Specialized breeding databases like the Breeding Management System and Integrated Breeding Platform provide tailored solutions for storing genotypic and phenotypic datasets. Cloud computing platforms have become increasingly important due to their scalability, allowing storage and processing of terabytes of data without the need for extensive local infrastructure. Services such as distributed cloud storage support collaborative access and real-time data updates across geographically dispersed teams. Data repositories serve as centralized archives for long-term storage and sharing of datasets, often supporting version control and metadata documentation. Secure storage systems also incorporate backup and recovery mechanisms to prevent data loss. The use of cloud-based platforms has been shown to improve data accessibility and computational efficiency, enabling faster analysis and decision-making. As data volumes continue to grow, robust storage solutions remain essential for maintaining data integrity and supporting advanced analytical workflows.

Interoperability and data sharing frameworks

Interoperability is a critical aspect of data integration, enabling different systems, tools, and datasets to work together seamlessly within breeding programs. It relies on standardized data formats, metadata schemas, and communication protocols that allow data exchange across platforms. Frameworks such as the Breeding API facilitate interoperability by providing standardized interfaces for accessing and

sharing breeding data. Data sharing frameworks promote collaboration among research institutions, breeding organizations, and stakeholders by enabling secure and efficient data exchange. These frameworks often incorporate access controls and data governance policies to protect sensitive information while supporting collaboration. Interoperability also supports the integration of diverse datasets, including genomic, phenotypic, and environmental data, into unified analytical pipelines. This integration enhances the ability to perform comprehensive analyses and improve prediction accuracy. Collaborative platforms enable breeders to access shared datasets, reducing duplication of efforts and accelerating innovation. The adoption of interoperable systems has been associated with increased efficiency in breeding programs, as it allows seamless data flow and reduces the time required for data preparation and analysis.

Data quality control and curation

Data quality control and curation are essential processes that ensure the reliability and usability of datasets in breeding programs. Quality control involves detecting and correcting errors, inconsistencies, and missing values in datasets. Techniques such as outlier detection, validation checks, and consistency analysis are commonly used to identify data anomalies. For genotypic data, quality control steps include filtering markers based on missing data rates, minor allele frequency, and genotyping errors. Phenotypic data require careful validation to account for environmental variation and measurement errors. Data curation involves organizing, annotating, and maintaining datasets to ensure long-term usability. Metadata documentation plays a key role by providing context about data collection methods, experimental design, and environmental conditions. Automated data pipelines and validation tools help streamline quality control processes, reducing manual effort and improving accuracy. High-quality datasets are crucial for reliable model predictions and decision-making. Poor data quality can lead to biased results and reduced genetic gain. Studies indicate that rigorous data curation can significantly improve prediction accuracy in genomic selection models. Maintaining high data quality standards is therefore fundamental to the success of data-driven breeding programs.

FAIR data principles (Findable, Accessible, Interoperable, Reusable)

The FAIR data principles provide a comprehensive framework for managing and sharing data in a way that maximizes its value and usability. Findable data are supported by clear metadata, unique identifiers, and searchable repositories, allowing users to easily locate relevant datasets. Accessible data are stored in systems that enable authorized users to retrieve information using standardized protocols. Interoperable data are formatted using standardized vocabularies and ontologies, allowing integration with other datasets and analytical tools. Reusable data are well-documented, with clear usage licenses and detailed metadata that enable future applications beyond the original study. Adoption of FAIR principles

enhances transparency, reproducibility, and collaboration in breeding programs. Many international research initiatives and funding agencies now require compliance with FAIR guidelines to promote open science. Implementing these principles involves the use of standardized data formats, persistent identifiers such as DOIs, and comprehensive metadata documentation. FAIR-compliant data systems support advanced analytics by enabling seamless data integration and reuse. The application of these principles has been shown to improve research efficiency and reduce duplication of efforts, contributing to more effective and collaborative crop improvement strategies.

Analytical Tools and Modeling Approaches

Statistical models in breeding (BLUP, mixed models)

Statistical models form the backbone of analytical frameworks in plant breeding, enabling accurate estimation of genetic effects and selection decisions. Best Linear Unbiased Prediction (BLUP) and mixed linear models are widely used due to their ability to separate genetic variance from environmental noise. These models incorporate both fixed effects, such as treatment conditions, and random effects, such as genetic variation among lines. BLUP has become a standard approach for estimating breeding values, especially in multi-location trials, as it accounts for unbalanced data and complex experimental designs. The use of restricted maximum likelihood methods improves parameter estimation, ensuring unbiased variance components. Mixed models also facilitate the analysis of genotype-by-environment interactions, which are critical for identifying stable and adaptable genotypes. Studies have shown that BLUP-based selection can improve prediction accuracy by 20–30% compared to simple phenotypic selection. These models are computationally efficient and scalable, making them suitable for large breeding datasets. Their integration into decision support systems enables breeders to rank genotypes more reliably and optimize selection intensity, ultimately enhancing genetic gain.

Genomic selection and prediction models

Genomic selection has revolutionized breeding by enabling the prediction of genetic merit using genome-wide marker data. Unlike marker-assisted selection, which focuses on a limited number of loci, genomic selection uses thousands of markers simultaneously to capture the effects of both major and minor genes. Prediction models such as genomic BLUP, ridge regression, Bayesian methods, and kernel-based approaches are commonly employed. These models estimate genomic estimated breeding values, allowing early selection of superior genotypes without extensive phenotypic evaluation. This approach significantly reduces breeding cycle duration, in some cases by up to 50%, accelerating the development of improved varieties. Prediction accuracy depends on factors such as training population size, marker density, and trait heritability. For complex traits like yield, genomic selection has demonstrated substantial gains in efficiency compared to conventional

methods. Integration with high-throughput phenotyping and environmental data further enhances model performance. As sequencing costs continue to decline, genomic selection is becoming increasingly accessible, supporting large-scale implementation across diverse crops and breeding programs.

Machine learning and deep learning applications

Machine learning and deep learning techniques are transforming data analysis in breeding by enabling the handling of large, complex datasets with non-linear relationships (Crossa *et.al.*, 2025). Algorithms such as random forests, gradient boosting, support vector machines, and neural networks are widely used for trait prediction and classification tasks. These methods can model interactions among genomic, phenotypic, and environmental variables more effectively than traditional approaches. Deep learning models, particularly convolutional neural networks, are applied in image-based phenotyping to assess traits such as disease severity, canopy structure, and biomass accumulation. These models can process high-resolution images from drones and field sensors, generating detailed and accurate trait measurements. Machine learning approaches have been shown to improve prediction accuracy for complex traits by up to 15–25% in certain scenarios. They also support feature selection, identifying the most informative variables for prediction. Integration of these techniques into decision support systems enables automated analysis and real-time insights. As computational resources expand, machine learning is expected to play an increasingly central role in enhancing breeding efficiency and precision.

Crop growth and simulation models

Crop growth models simulate the physiological processes of plant development, providing insights into how genotypes respond to environmental and management conditions. Models such as DSSAT, APSIM, and WOFOST use inputs including weather data, soil properties, and crop parameters to predict growth stages, biomass accumulation, and yield. These models are valuable tools for evaluating genotype performance under different environmental scenarios, including future climate projections. By integrating genetic parameters, crop models can support the selection of traits associated with stress tolerance and resource use efficiency. Simulation studies allow breeders to test hypotheses and optimize breeding strategies without the need for extensive field trials. These models can also guide trial design by identifying environments that maximize the expression of target traits. The combination of crop models with genomic prediction frameworks enhances the ability to predict genotype performance across diverse conditions. Advances in computing have improved the scalability and accuracy of these models, making them an integral part of modern decision support systems in breeding programs.

Multi-environment trial (MET) analysis

Multi-environment trial analysis is essential for evaluating genotype performance

across diverse locations and conditions. MET data capture variability arising from environmental factors, enabling the assessment of genotype stability and adaptability. Statistical methods such as analysis of variance, genotype plus genotype-by-environment interaction models, and additive main effects and multiplicative interaction models are commonly used. These approaches help identify genotypes that perform consistently across environments or those specifically adapted to particular conditions. Stability parameters and biplot analyses provide visual representations of genotype performance, aiding interpretation and decision-making. MET analysis also supports the identification of mega-environments, which are regions with similar environmental characteristics. This classification helps breeders target specific environments more effectively. The use of advanced models that incorporate environmental covariates improves the understanding of genotype-by-environment interactions. Accurate MET analysis enhances selection decisions by balancing yield performance and stability, ensuring the development of varieties that perform well under variable conditions.

Decision trees and optimization algorithms

Decision trees and optimization algorithms play a crucial role in guiding complex decision-making processes in breeding programs. Decision trees provide a structured and interpretable framework for selecting among alternative breeding strategies based on data-driven criteria. These models split data into subsets based on key variables, enabling clear identification of decision rules. Optimization algorithms, including linear programming, genetic algorithms, and simulated annealing, are used to solve complex problems such as parent selection, crossing design, and resource allocation. These methods aim to maximize genetic gain while minimizing costs and time. For example, optimization techniques can identify the best combination of parental lines to achieve desired trait improvements. Decision trees are particularly useful for integrating multiple data sources, including genomic and environmental data, into a single decision framework. The combination of decision trees and optimization methods enhances the efficiency and effectiveness of breeding programs. Their integration into decision support systems enables automated and data-driven decision-making, supporting the development of improved crop varieties with greater precision and speed.

Decision-Making in Breeding Programs

Selection strategies: phenotypic vs genomic selection

Selection strategies in breeding programs have evolved from purely phenotypic approaches to integrated genomic-based frameworks. Phenotypic selection relies on observable traits measured in field conditions, often across multiple seasons and environments to ensure reliability. This method is effective for traits with high heritability but becomes less efficient for complex traits influenced by multiple genes and environmental interactions. Genomic selection uses genome-wide marker information to predict breeding values, enabling early selection without extensive

phenotyping. Studies indicate that genomic selection can increase selection accuracy by up to 30% for low-heritability traits and reduce breeding cycle time significantly. The choice between phenotypic and genomic strategies depends on factors such as trait architecture, resource availability, and breeding objectives. Combining both approaches through integrated selection frameworks often yields the best results, leveraging the strengths of each method. Decision support systems assist in evaluating these strategies by simulating outcomes and optimizing selection intensity, leading to improved genetic gain and efficiency.

Parent selection and crossing decisions

Parent selection is a critical step that determines the genetic potential of future breeding populations. The goal is to choose parental lines that maximize the probability of producing superior progeny by combining desirable alleles. Traditional approaches rely on phenotypic performance and pedigree information, while modern methods incorporate genomic data to estimate genetic diversity and complementarity. Genomic relationship matrices and marker-based diversity analyses enable precise identification of suitable parents. Crossing decisions are guided by predictions of progeny performance using simulation models and genomic estimated breeding values. Optimization algorithms can evaluate thousands of possible crosses to identify combinations with the highest expected genetic gain. Maintaining genetic diversity while improving target traits is a key consideration in parent selection. Studies show that strategic crossing based on genomic information can enhance selection response by 10–20% compared to random mating. Decision support tools provide breeders with ranked crossing plans and predicted outcomes, improving the efficiency and effectiveness of breeding strategies.

Advancement and culling decisions in breeding pipelines

Breeding pipelines involve multiple stages where lines are evaluated, advanced, or discarded based on performance criteria. Advancement decisions determine which genotypes progress to the next stage, while culling removes inferior lines to conserve resources. These decisions are increasingly supported by data-driven frameworks that integrate phenotypic, genotypic, and environmental information. Statistical models such as BLUP and genomic prediction are used to estimate breeding values and rank candidates. Early-generation selection using genomic data allows rapid elimination of low-potential lines, reducing the number of entries in later stages. This approach can decrease evaluation costs by up to 40% while maintaining or improving selection accuracy. Decision thresholds are often defined based on performance metrics, stability, and specific breeding goals. Automated decision support systems can process large datasets and provide recommendations in real time, enhancing consistency and reducing subjective bias. Efficient advancement and culling strategies are essential for maintaining a streamlined pipeline and achieving faster genetic progress.

Resource allocation and trial design optimization

Efficient use of resources is essential for maximizing the impact of breeding programs (Sanchez *et.al.*, 2023). Resource allocation involves decisions on how to distribute budget, labor, and experimental capacity across different stages of the breeding pipeline. Trial design optimization focuses on structuring experiments to obtain reliable data while minimizing costs. Statistical designs such as randomized complete block designs, alpha-lattice designs, and augmented designs are commonly used to improve experimental precision. Decision support systems use simulation and optimization algorithms to identify optimal trial configurations, including the number of replications, locations, and genotypes. These tools can evaluate trade-offs between cost and accuracy, helping breeders allocate resources effectively. For example, optimizing the number of test locations can improve the detection of genotype-by-environment interactions without unnecessary expenditure. Data-driven approaches have been shown to increase experimental efficiency by up to 25%. Integration of environmental and genomic data further enhances trial design by targeting representative testing conditions. Effective resource allocation and trial design contribute to more reliable results and improved decision-making.

Risk analysis and uncertainty management

Risk and uncertainty are inherent in breeding programs due to environmental variability, genetic complexity, and market dynamics. Decision-making frameworks incorporate risk analysis to evaluate the probability of different outcomes and their potential impacts. Statistical models and simulation tools are used to assess uncertainty in trait expression, genotype performance, and environmental conditions. Sensitivity analysis helps identify key factors that influence outcomes, guiding more robust decision-making. Probabilistic models enable breeders to quantify risks associated with different strategies, such as selecting specific parents or advancing certain lines. Incorporating environmental variability into predictive models improves the reliability of performance estimates. Decision support systems provide scenario analysis capabilities, allowing breeders to explore multiple strategies and select those with the highest expected benefits and acceptable risk levels. Effective uncertainty management enhances resilience in breeding programs, ensuring consistent progress even under variable conditions. This approach supports the development of stable and adaptable crop varieties capable of performing under diverse and changing environments.

High-Throughput Phenotyping and Digital Agriculture*Advances in sensor technologies and imaging platforms*

High-throughput phenotyping has transformed trait measurement by enabling rapid, precise, and non-destructive data collection across large breeding populations. Modern sensor technologies include RGB cameras, multispectral and hyperspectral sensors, thermal imaging devices, and LiDAR systems, each capturing distinct

aspects of plant physiology and structure. Hyperspectral sensors can record reflectance across hundreds of wavelengths, allowing detection of subtle biochemical changes linked to stress, nutrient status, or disease. Thermal cameras measure canopy temperature, which is strongly associated with transpiration efficiency and drought tolerance. LiDAR systems generate three-dimensional plant architecture data, supporting biomass estimation and growth analysis. Imaging platforms integrated with automated systems can phenotype thousands of plots per day, a scale unattainable through manual methods. Sensor calibration and standardization are critical to ensure data accuracy across time and environments. Studies indicate that high-throughput phenotyping can reduce labor requirements by over 60% while increasing data resolution. These technologies provide continuous and objective measurements, improving the reliability of trait assessment and enabling breeders to capture dynamic plant responses throughout the growth cycle.

UAVs, satellites and ground-based systems

Digital agriculture platforms rely on multiple data acquisition systems, including unmanned aerial vehicles, satellites, and ground-based sensors, to monitor crop performance at different spatial and temporal scales. UAVs equipped with advanced sensors can capture high-resolution imagery at centimeter-level detail, making them ideal for plot-level phenotyping. They enable frequent monitoring, allowing detection of early stress symptoms and growth patterns. Satellite systems provide broader coverage, capturing field and regional-level data with revisit intervals ranging from daily to weekly. Though spatial resolution is lower than UAVs, satellites offer consistent long-term datasets useful for environmental characterization and trend analysis. Ground-based systems, such as phenotyping carts and stationary sensors, provide highly accurate measurements and continuous data streams. Integration of these platforms allows multi-scale analysis, combining detailed local observations with broader environmental context. Data collected from these systems can include vegetation indices, canopy height, biomass estimates, and soil moisture levels. The combined use of aerial, satellite, and ground technologies enhances monitoring efficiency and supports comprehensive evaluation of breeding trials across diverse environments.

Integration of phenomics with genomics

The integration of phenomics and genomics represents a major advancement in understanding the relationship between genotype and phenotype. Phenomics provides high-resolution trait data, while genomics offers detailed information on genetic variation. Combining these datasets enables the identification of genetic regions associated with complex traits and improves prediction accuracy in breeding programs. Genome-wide association studies and genomic prediction models benefit from precise phenotypic measurements, increasing their ability to detect meaningful genetic signals. High-throughput phenotyping generates large datasets that capture temporal changes in plant growth, which can be linked with

genomic markers to study trait dynamics. This integration supports the dissection of quantitative traits influenced by multiple genes and environmental factors. Advanced analytical frameworks use combined datasets to predict performance across environments, enhancing selection decisions. Research indicates that integrating phenomic and genomic data can improve predictive ability by up to 20% compared to using either data type alone. This approach enables a more comprehensive understanding of plant biology and accelerates the development of improved crop varieties with desirable traits.

Data processing and interpretation challenges

The rapid expansion of high-throughput phenotyping and digital agriculture technologies has led to significant challenges in data processing and interpretation (Yang *et.al.*, 2020). Large datasets generated from sensors, imaging platforms, and remote sensing systems require substantial computational resources for storage, processing, and analysis. Data preprocessing steps such as noise reduction, image correction, and feature extraction are critical for ensuring accuracy. Variability in data quality due to environmental conditions, sensor calibration, and measurement errors can complicate analysis. Extracting meaningful traits from raw data often involves complex algorithms and machine learning techniques, which require specialized expertise. Standardization of data formats and analysis protocols remains a challenge, limiting interoperability across platforms. Interpreting high-dimensional data and translating it into actionable insights for breeding decisions is another key issue. Visualization tools and decision support systems play an important role in simplifying complex outputs. Studies highlight that a significant proportion of phenotypic data remains underutilized due to processing limitations. Addressing these challenges requires improved computational infrastructure, standardized workflows, and interdisciplinary collaboration, ensuring that the full potential of digital phenotyping technologies is realized in breeding programs.

Table: High-Throughput Phenotyping and Digital Agriculture

Category	Tool/Technology	Description	Application in Agriculture
Imaging Technologies	RGB Imaging	Captures visible plant traits	Monitors growth, biomass, and morphology
	Hyperspectral Imaging	Captures wide spectral information	Detects stress, nutrient status, and disease early
	Thermal Imaging	Measures canopy temperature	Assesses plant water status and

			drought stress
Remote Sensing Platforms	UAVs (Drones)	Aerial data collection systems	Field-scale monitoring and precision farming
	Satellites	Large-scale earth observation	Crop mapping and yield prediction
Ground-Based Platforms	Phenotyping Robots	Automated field or greenhouse systems	Collect high-resolution plant trait data
	Sensor Networks (IoT)	Distributed environmental sensors	Real-time monitoring of soil, weather, and crops
Data Acquisition Tools	LiDAR	3D scanning of plant structure	Measures plant height, canopy architecture
	GPS & GIS	Spatial mapping technologies	Tracks field variability and site-specific management
Data Processing & Analytics	Big Data Platforms	Cloud-based storage and processing	Handles large phenotypic datasets
	Machine Learning & AI	Predictive modeling and pattern recognition	Enhances decision-making and yield prediction
Decision Support Systems	Farm Management Software	Integrates data for farm decisions	Optimizes irrigation, fertilization, and pest control
	Digital Twins	Virtual crop models	Simulates crop growth and management scenarios
Applications	Precision Agriculture	Site-specific crop management	Increases efficiency and reduces input

			costs
	Stress Detection	Early identification of abiotic/biotic stress	Improves crop health management
	Yield Prediction	Forecasting production using data models	Supports planning and supply chain decisions
Emerging Trends	Integration with Genomics	Linking phenotypic and genetic data	Accelerates breeding programs
	Automation & Robotics	Autonomous farming systems	Reduces labor and increases efficiency

Integration of Genomics and Big Data Analytics

Role of omics technologies (genomics, transcriptomics, proteomics)

Omics technologies provide a comprehensive framework for understanding biological systems at multiple levels, supporting advanced breeding strategies. Genomics focuses on DNA sequence variation and enables identification of markers, genes, and quantitative trait loci associated with agronomic traits. Transcriptomics examines gene expression patterns, revealing how genes respond to developmental cues and environmental stresses. Proteomics extends this understanding by analyzing protein abundance, structure, and interactions, offering insights into functional mechanisms underlying phenotypic traits. Together, these technologies enable a systems-level understanding of plant biology. High-throughput sequencing platforms can generate billions of reads per experiment, producing detailed datasets that capture molecular complexity. Integration of omics data allows identification of regulatory networks and pathways influencing traits such as yield, stress tolerance, and nutrient efficiency. This multi-layered approach enhances the ability to dissect complex traits controlled by numerous genes and interactions. Omics-driven breeding strategies improve the precision of selection by linking molecular information with observable characteristics, accelerating the development of improved crop varieties.

Big data infrastructure and computational tools

The integration of genomics with breeding programs requires robust big data infrastructure capable of handling massive and complex datasets. Modern sequencing technologies can generate terabytes of data in a single experiment, necessitating scalable storage and processing solutions. High-performance computing systems and cloud-based platforms provide the computational capacity

required for large-scale data analysis. Distributed computing frameworks such as Hadoop and Spark enable parallel processing, reducing analysis time for complex workflows. Specialized bioinformatics tools and pipelines are used for sequence alignment, variant calling, and functional annotation. Databases and data management systems store structured and unstructured data, ensuring accessibility and organization. Visualization tools and dashboards help interpret complex datasets, making them accessible to breeders and researchers. Efficient computational infrastructure supports real-time analysis and integration of diverse data types. Investments in hardware, software, and technical expertise are essential for maintaining effective data systems. The availability of advanced computational tools has significantly enhanced the ability to analyze and utilize genomic data in breeding programs.

Predictive analytics and breeding value estimation

Predictive analytics plays a central role in modern breeding by enabling the estimation of breeding values using large-scale datasets. Statistical and machine learning models analyze relationships between genotype, phenotype, and environment to predict the performance of breeding lines. Genomic estimated breeding values are calculated using genome-wide marker information, allowing early selection of superior genotypes. These predictions reduce the need for extensive field evaluations, saving time and resources. Prediction accuracy depends on factors such as dataset size, trait heritability, and model selection. Advanced algorithms can incorporate non-linear relationships and interactions, improving prediction performance for complex traits. Integration of environmental and management data enhances the robustness of predictions by accounting for variability in growing conditions. Studies have demonstrated that predictive analytics can significantly increase selection efficiency and genetic gain. Decision support systems use these predictions to rank candidates and guide breeding decisions. The application of predictive analytics enables a shift from reactive to proactive breeding strategies, improving overall program efficiency.

Data fusion techniques for multi-layered datasets

Data fusion techniques enable the integration of diverse datasets, including genomic, phenotypic, environmental, and management information, into unified analytical frameworks. These techniques are essential for capturing the complexity of biological systems and improving prediction accuracy. Methods such as multi-view learning, Bayesian integration, and network-based approaches are commonly used to combine data from different sources. Data fusion allows identification of relationships that may not be apparent when analyzing datasets independently. For example, linking genomic markers with environmental variables can reveal genotype-specific responses to stress conditions. Integration of multi-layered data enhances the ability to model genotype-by-environment interactions and predict performance across diverse conditions. Advanced computational methods can

handle high-dimensional data, extracting meaningful patterns and reducing noise. Effective data fusion requires standardized data formats, robust preprocessing, and appropriate modeling techniques. Studies indicate that combining multiple data sources can improve predictive accuracy by up to 25% compared to single-source models. This approach supports more informed decision-making and contributes to the development of resilient and high-performing crop varieties.

Decision Support Tools and Platforms

Existing software and platforms (e.g., Breeding Management System, Field Book)

A wide range of software platforms support data-driven decision-making in breeding programs by integrating data collection, management, and analysis into unified environments. The Breeding Management System (BMS), developed under the Integrated Breeding Platform, is widely used for managing genotypic and phenotypic datasets, experimental designs, and selection workflows. Field Book is a mobile-based application designed for efficient field data collection, enabling real-time recording of phenotypic observations with reduced transcription errors. Other platforms such as Agrobase, Flapjack, and TASSEL provide functionalities for data visualization, genetic analysis, and marker-trait association studies. These tools support high-throughput data handling and enable breeders to process large datasets with improved accuracy. Many platforms incorporate statistical models and genomic prediction modules, allowing direct integration of analytics into breeding workflows. Adoption of such tools has been shown to improve data management efficiency and reduce manual workload significantly. Continuous updates and community-driven development ensure that these platforms remain relevant and capable of handling emerging data types and analytical requirements in modern breeding programs.

Custom-built DSS tools for specific crops

Custom-built decision support systems are developed to address the unique requirements of specific crops, breeding objectives, and production environments. These tools are tailored to incorporate crop-specific traits, growth models, and genetic architectures, providing more precise and relevant recommendations. For example, breeding programs for cereals, legumes, and horticultural crops often require specialized modules for trait evaluation, disease resistance screening, and quality assessment. Custom DSS tools can integrate local environmental data, management practices, and breeding goals, enhancing their applicability. These systems often include simulation models that predict crop performance under different scenarios, supporting strategic decision-making. Development of custom tools involves collaboration between breeders, data scientists, and software engineers to ensure that the system meets practical needs. While development costs can be high, the benefits include improved accuracy, flexibility, and alignment with program objectives. Custom solutions are particularly valuable in complex breeding

programs that require integration of diverse datasets and specialized analytical workflows.

Open-source vs proprietary solutions

Decision support platforms in breeding programs can be broadly categorized into open-source and proprietary solutions, each offering distinct advantages and limitations (Rathore *et.al.*, 2018). Open-source tools provide free access to source code, enabling customization and community-driven improvements. Examples include TASSEL and Flapjack, which are widely used for genetic analysis and visualization. These platforms promote transparency and collaboration, allowing users to adapt tools to specific requirements. Proprietary solutions, such as Agrobase and certain commercial breeding software, offer dedicated support, user-friendly interfaces, and integrated features, often with subscription or licensing costs. These systems are typically more polished and easier to deploy but may limit flexibility due to restricted access to underlying code. The choice between open-source and proprietary solutions depends on factors such as budget, technical expertise, and program scale. Many breeding programs adopt hybrid approaches, combining open-source tools for analysis with proprietary platforms for data management. This combination allows optimization of cost, functionality, and usability, supporting efficient decision-making processes.

User interface design and usability considerations

User interface design plays a critical role in the effectiveness of decision support tools by determining how easily users can interact with complex data and analytical outputs. Well-designed interfaces present information in a clear and intuitive manner, using dashboards, visualizations, and interactive elements to simplify data interpretation. Features such as customizable views, real-time updates, and graphical representations of results enhance user experience and decision-making efficiency. Usability considerations include ease of navigation, responsiveness, and compatibility across devices such as desktops, tablets, and smartphones. Training requirements and user support systems also influence adoption rates, as complex tools may be underutilized if they are difficult to operate. Incorporating user feedback during the design process helps ensure that tools meet the practical needs of breeders. Studies indicate that improved interface design can significantly increase the adoption and effective use of digital tools in agriculture. Integration of visualization techniques such as heat maps, scatter plots, and geospatial mapping further enhances the interpretability of data. Effective user interface design ensures that advanced analytical capabilities are accessible to users with varying levels of technical expertise.

Challenges and Limitations

Data heterogeneity and integration issues

Data heterogeneity is a major challenge in modern breeding programs due to the

diversity of data types, formats, and sources. Breeding datasets include genomic sequences, SNP arrays, high-throughput phenotyping images, environmental measurements, and management records, each with varying scales, units, and quality. Integrating these heterogeneous datasets into unified analytical frameworks is complex, requiring standardization, normalization, and robust data mapping protocols. Differences in data collection methods across locations and seasons can introduce inconsistencies that compromise analysis and predictive accuracy. Multi-environment trials, in particular, generate data with variable quality due to environmental fluctuations and management practices. Without proper harmonization, combining genotypic, phenotypic, and environmental datasets can produce biased results and limit the utility of decision support systems. Ontologies and controlled vocabularies, such as the Crop Ontology, help reduce heterogeneity by providing consistent trait definitions and metadata standards. Despite these tools, the challenge of integrating high-dimensional and multi-modal datasets remains significant. Addressing this issue requires coordinated efforts in data standardization, automated preprocessing pipelines, and scalable integration frameworks to ensure that heterogeneous data can be analyzed reliably for breeding decisions.

Computational and infrastructure constraints

The volume and complexity of data generated by modern breeding programs place heavy demands on computational infrastructure. High-throughput sequencing, multi-sensor phenotyping, and environmental monitoring generate terabytes of data per season, necessitating robust storage, processing, and backup solutions. Traditional local servers often lack the capacity to manage these datasets efficiently, making high-performance computing clusters or cloud-based systems essential. Analytical workflows for genomic prediction, machine learning, and crop simulation require substantial processing power and memory. Limited infrastructure can lead to delays in data analysis, restricting the speed and accuracy of breeding decisions. Data transfer, security, and accessibility are additional concerns when large datasets are distributed across locations. Computational bottlenecks can hinder real-time decision support, reducing the efficiency of selection and advancement cycles. Developing scalable, flexible, and secure infrastructure is critical to support modern data-driven breeding programs. Investment in cloud computing, distributed processing frameworks, and database optimization can help overcome these constraints, enabling the effective use of large and complex datasets in breeding pipelines.

Skill gaps and capacity building needs

Effective implementation of data-driven breeding requires a workforce with expertise in genomics, bioinformatics, data science, and breeding principles. Many breeding programs face skill gaps, as traditional breeding training often does not include advanced computational and analytical methods. Managing high-throughput

phenotyping data, developing predictive models, and integrating multi-layered datasets require specialized knowledge in statistical genetics, machine learning, and software tools. Capacity building is necessary to train breeders, data managers, and technical staff in these areas, ensuring they can use decision support systems effectively. Collaborative training programs, workshops, and interdisciplinary teams have proven successful in enhancing skills and promoting adoption of modern analytical tools. Without adequate capacity, the potential of data-driven breeding cannot be fully realized, and misinterpretation of complex datasets may lead to suboptimal selection decisions. Building institutional support, developing user-friendly tools, and fostering continuous learning are essential to close skill gaps and enhance the efficiency and impact of breeding programs.

Ethical, legal and data privacy concerns

Data-driven breeding raises important ethical, legal, and privacy issues, particularly when sharing genomic and phenotypic datasets (Zhang *et.al.*, 2025). Intellectual property rights, data ownership, and access permissions must be clearly defined to prevent misuse and ensure equitable benefits. Sensitive data, such as proprietary germplasm information or farm-level environmental and management data, require secure storage and controlled access. Legal frameworks governing data sharing, licensing, and cross-border collaboration vary widely, complicating multi-institutional projects. Ethical considerations also arise in participatory breeding and data collection from farmers, requiring informed consent and transparency about how data will be used. Maintaining confidentiality while promoting open science requires careful balance. Failure to address these issues can hinder collaboration, reduce data availability, and slow progress in breeding programs. Adoption of clear governance policies, secure data platforms, and standardized licensing agreements are critical to mitigate risks, ensure compliance, and foster trust among stakeholders.

Cost and scalability issues

Implementing data-driven breeding approaches often requires substantial financial investment, creating challenges for scalability and sustainability. High-throughput genotyping, phenotyping platforms, and computational infrastructure involve significant upfront costs. Operating expenses include consumables, maintenance, software licenses, and personnel salaries. Small and medium-scale breeding programs may face difficulty in adopting these technologies due to budget constraints. Scaling data-driven approaches across multiple locations and crops requires harmonized protocols, trained staff, and interoperable systems, which increase complexity and costs. Despite the efficiency gains in selection and cycle reduction, the initial financial and logistical burden can be prohibitive for widespread implementation. Cost-effective strategies, such as shared infrastructure, open-source software, and cloud-based computational resources, can partially address these constraints. Prioritizing investment in high-impact components and

integrating modular tools allows programs to scale progressively while maintaining efficiency and predictive accuracy. Strategic planning is essential to ensure that the benefits of data-driven breeding outweigh costs and that programs remain sustainable over time (Fadda *et.al.*, 2020).

Conclusion

Data-driven decision support systems have revolutionized modern breeding by integrating genomics, phenomics, environmental data, and computational analytics to enable precise and efficient selection of superior crop varieties. High-throughput phenotyping, UAVs, and sensor technologies generate large-scale datasets that, when combined with genomic prediction and machine learning models, accelerate breeding cycles and improve accuracy for complex traits. Decision support platforms facilitate resource optimization, trial design, and real-time analysis, while emerging technologies such as digital twins, artificial intelligence, and blockchain enhance automation, traceability, and predictive capability. Climate-resilient breeding is increasingly guided by predictive analytics, enabling development of varieties tolerant to abiotic stresses. Participatory approaches supported by digital tools ensure relevance to end-user needs. Despite challenges in data integration, infrastructure, and capacity, these innovations collectively drive sustainable crop improvement and global food security.

References

1. Crossa, J., Montesinos-Lopez, O. A., Costa-Neto, G., Vitale, P., Martini, J. W., Runcie, D., ... & Ortiz, R. (2025). Machine learning algorithms translate big data into predictive breeding accuracy. *Trends in Plant Science*, 30(2), 167-184.
2. Dipta, B., Sood, S., Devi, R., Bhardwaj, V., Mangal, V., Thakur, A. K., ... & Singh, A. K. (2023). Digitalization of potato breeding program: Improving data collection and management. *Heliyon*, 9(1).
3. Fadda, C., Mengistu, D. K., Kidane, Y. G., Dell'Acqua, M., Pè, M. E., & Van Etten, J. (2020). Integrating conventional and participatory crop improvement for smallholder agriculture using the seeds for needs approach: A review. *Frontiers in Plant Science*, 11, 559515.
4. James, R. A., Campbell, I. M., Chen, E. S., Boone, P. M., Rao, M. A., Bainbridge, M. N., ... & Shaw, C. A. (2016). A visual and curatorial approach to clinical variant prioritization and disease gene discovery in genome-wide diagnostics. *Genome medicine*, 8(1), 13.
5. Rathore, A., Singh, V. K., Pandey, S. K., Rao, C. S., Thakur, V., Pandey, M. K., ... & Das, R. R. (2018). Current status and future prospects of next-generation data management and analytical decision support tools for enhancing genetic gains in crops. *Plant genetics and molecular biology*, 277-292.

6. Sanchez, D., Sadoun, S. B., Mary-Huard, T., Allier, A., Moreau, L., & Charcosset, A. (2023). Improving the use of plant genetic resources to sustain breeding programs' efficiency. *Proceedings of the National Academy of Sciences*, 120(14), e2205780119.
7. Yang, W., Feng, H., Zhang, X., Zhang, J., Doonan, J. H., Batchelor, W. D., ... & Yan, J. (2020). Crop phenomics and high-throughput phenotyping: past decades, current challenges, and future perspectives. *Molecular plant*, 13(2), 187-214.
8. Zhang, Y., Huang, G., Zhao, Y., Lu, X., Wang, Y., Wang, C., ... & Zhao, C. (2025). Revolutionizing crop breeding: Next-generation artificial intelligence and big data-driven intelligent design. *Engineering*, 44, 245-255.
9. Zhang, Y., Huang, G., Zhao, Y., Lu, X., Wang, Y., Wang, C., ... & Zhao, C. (2025). Revolutionizing crop breeding: Next-generation artificial intelligence and big data-driven intelligent design. *Engineering*, 44, 245-255.