



International Journal of Research in Agronomy

E-ISSN: 2618-0618
P-ISSN: 2618-060X
© Agronomy
NAAS Rating (2026): 5.20
www.agronomyjournals.com
2026; SP-9(6): 194-200
Received: 02-04-2026
Accepted: 05-05-2026

Aditya Kumar
Student, Department of Genetics
and Plant Breeding, Lovely
Professional University, Phagwara,
Punjab, India

Tanu Sandhu
Student, Department of Genetics
and Plant Breeding, Lovely
Professional University, Phagwara,
Punjab, India

Dilbag Singh
Student, Department of Genetics
and Plant Breeding, Lovely
Professional University, Phagwara,
Punjab, India

Dr. Rahul Singh
Assistant Professor, Department of
Genetics and Plant Breeding,
Lovely Professional University,
Phagwara, Punjab, India

Harnessing artificial intelligence in genomic selection: A new era in plant breeding

Aditya Kumar, Tanu Sandhu, Dilbag Singh and Rahul Singh

DOI: <https://www.doi.org/10.33545/2618060X.2026.v9.i6Sc.5764>

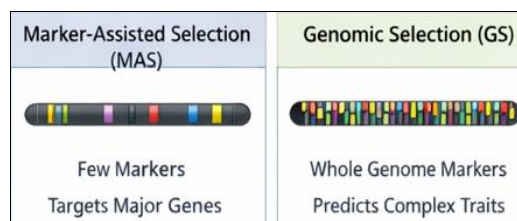
Abstract

Artificial intelligence (AI) and genomic selection (GS) is transforming modern breeding of plants whereby the intricate characteristics can be precisely forecasted using abundant genomic and phenomic and environmental data sets. This review provides an overview of the current advance in deep learning, machine learning and high-throughput phenotyping (HTP) of genomic prediction on crops of key crops. The AI models, CNNs, LSTM and Transformers, are better models of the non-linear interactions between genes and gene and environment interactions compared to the classical statistical models. The integration with UAV-based phenotyping, ground sensors and multi-omics data will further go a step further to enhance the accuracy of the prediction and minimise the breeding cycles. Besides, the novel solutions like multi-task learning, federated learning, and explainable AI improve the model performance, its scalability, and comprehensibility. No matter whether the problem of data quality, computational needs, and model transparency is present, AI-based genomic selection is a prospective model to increase the genetic gain and contribute to climate-resilient crop production. Such innovations are a significant advancement towards Breeding 5.0 and sustainable global food security which involve data-led approaches.

Keywords: Artificial intelligence (AI), genomic selection (GS), High-throughput phenotyping (HTP), Deep learning (DL), machine learning (ML), UAV phenotyping, multi-omics integration, genotype x environment (GxE), genomic prediction

1. Introduction

Due to climatic change, increased population, new pests, and degradation of natural resources, the farming system in the world is changing. Plant breeding needs to become more rapid to sustain the food supply; it has to produce food that grows extensively, and can endure harsh weather conditions and be healthy. The ancient breeding systems such as sorting plants by their appearance or by a limited set of genetic indicators (Marker-assisted selection (MAS)) have been useful, but continue to struggle with the impacts of many genes and the interaction of the genes with the environment. The genomic selection (GS) method was introduced in 2001. It relies on markers data on the entire genome to make an approximation of the goodness of the plant they are expected to be. When compared to MAS, which considers a small number of markers, GS can more effectively and faster predict the best plants prior to the breeding process and reduce each cycle in less time.



Source: ^[14, 6] Heffner *et al.*, 2009; Crossa *et al.*, 2017

Fig 1: Difference between Marker-Assisted Selection and Genomic Selection approaches in Plant Breeding

Corresponding Author:
Aditya Kumar
Student, Department of Genetics
and Plant Breeding, Lovely
Professional University, Phagwara,
Punjab, India

The effectiveness of GS is determined by the models under which the genetic data is related to the traits. The traditional approaches like genomic best linear unbiased prediction (GBLUP) or Bayesian models presume that relationships are simple and additive and do not model those effects like dominance, interactions between genes and other non-linear effects. This is why more and more artificial intelligence, in particular, deep learning (DL) is being utilized. DL is able to acquire complicated patterns. An analysis of 14 datasets of crops by International Maize and Wheat Improvement Center (CIMMYT) in 2025 revealed that tuned DL models tended to be superior to GBLUP, particularly where the trait being assessed was of complex inheritance and when the training data was small. Predictions using single-nucleotide polymorphisms (SNPs) were significantly improved in rice using structural variants (SVs), particularly binary or non-linear traits. The same was observed in maize and wheat in which DL had the ability to capture non-additive effects and compound interaction between genes and environment.

New DL designs are challenging prediction quality. In 2025, Li and others developed a hybrid architecture combining the convolutional neural networks (CNNs), residual networks (ResNets), and long short-term memory (LSTM) layers. This combination is able to learn spatial patterns, temporal changes and sequences better than single DL models on several crops. As the data on most of the environments and all of the layers of omics are available, such tools as DeepCCR provide rice breeders with the interface of a user-friendly DL. Such advances are useful in the direction of Breeding 5.0 an intervention that combines AI, big data, and genome editing.

Meanwhile, scientists are incorporating enviromics, which are data about the environment (soil type, temperature, rainfall and satellite indices) to enhance predictions of the traits. Developed DL models are becoming popular with small or structured data sets. As an example, in 2025, the Learnable Group Transform (LGT) by Dong and colleagues increased the resource-constrained performance of resource-limited settings (increasing rice yield by 15% and height prediction accuracy by 15 per cent). LSTM frame works based on autoencoders have also performed well with barley despite paucity of training data.

Even breeding choices are being optimized with the help of AI. Reinforcement learning (RL) allows the system to learn optimal actions in the long-term. In 2024, Younis and colleagues demonstrated that an RL strategy outperformed conventional selection in maize, increased genetic gain over the long term. However, such approaches require computing resources, high quality and clean datasets and are not always easy to understand. Explainable AI, combining high-throughput phenotyping with environmental data and open-source software are working to resolve such barriers, taking us one step closer to a data-driven, AI-assisted breeding age.

The review closely observes the integration of AI, in particular, DL, into GS in plant breeding. It reviews the current advances of AI predictive applications, their application in large crops, the importance of high-throughput phenotyping and multi-omics, and the challenges facing the field. The idea is to provide plant breeders, data scientists and researchers with a clear image of how AI may be used to increase genetic gain and aid the next generation of breeding practices, as part of the Breeding 5.0 vision.

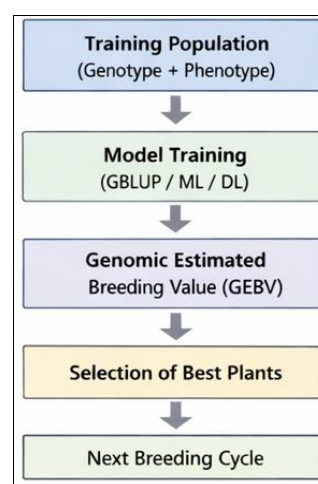
2. Genomic Selection: Concepts and Tools

2.1 Overview of Genomic Selection Methodology

Genomic Selection (GS) is a relevant novel breeding method of

plants. It assists in searching and selecting the most suitable plants based on the data of markers of the entire genome, rather than simply gazing at the appearance of the plants. Meuwissen and others had originally proposed GS in 2001. It involves several markers that are located throughout the genome to identify the effect of all significant genes, including those with small effects that are frequently missed using other approaches. GS operates by making a prediction model based on a training group of plants that has both predictive and performance data. The models then use only marker data to make breeding predictions of other plants. This allows breeders to make decisions earlier in the breeding cycle, which save both time and money and enhance the genetics.

GS also examines the entire genome and adds the effect of numerous markers in comparison to Marker Assisted Selection (MAS), which makes use of the markers near the main genes. This is highly suitable to complicated attributes that require a significant number of genes, as well as those influenced by the environmental conditions.



Source: Meuwissen *et al.*, 2001;
Crossa *et al.*, 2017

Fig 2: General workflow of genomic selection in plant breeding

2.2 Important Genomic Selection Terms

2.2.1 Genomic Estimated Breeding Value (GEBV)

GEBVs are predictors of genetic value of a plant based on marker data. They are indispensable in GS since they allow breeders to rank and choose the plants in case of missing performance data^[13] (Hayes *et al.*, 2009).

2.2.2 Training Population (TP)

It is an assortment of plants with both marker and performance data. The representativeness, diversity and size of the TP have a great impact on the accuracy of the model. Preferably, the TP must resemble the breeding group with regard to genetics and traits (Spindel and McCouch, 2016).

2.2.3 Prediction Accuracy

This is the level of correspondence of GEBVs with true breeding values or actual performance. It demonstrates the model efficacy and relies on such variables as trait heritability, density of markers, population structure, TP size, and the method of model^[6] (Crossa *et al.*, 2017).

2.3 Genotyping Data and Technologies

The technology of genotyping has been improved, which has made GS easier and cheaper. In the initial years, GS applied the

high-density SNP arrays such as Illumina Infinium or Affymetrix Axiom which provided good reliable data but were costly to use in large projects. Genotyping-by-sequencing (GBS) and low-coverage whole-genome sequencing (LC-WGS) are cheaper alternatives that are currently trendy. These approaches reach the genome in a broader way at reduced cost through the sequencing of particular regions or entire genome at reduced coverage^[10, 35] (Elshire *et al.*, 2011; Scheben *et al.*, 2017). They may contain missing data or errors and require imputation or statistical correction in order to maintain accuracy. Specific sequencing, exome capture and transcript-based genotyping are aimed at significant genomic regions. The inclusion of other forms of data like transcriptomics, metabolomics and epigenomics provide a more comprehensive view of traits, particularly those of a complex nature that is affected by both gene regulation and metabolism.

Data collection and management are very important. Digital tools, drones, sensor phenotyping are capable of collecting extensive amounts of accurate performance data as time and space vary. Together with more comprehensive genotypic data, these technologies generate more powerful GS models and a more in-depth insight into growth by climate relationships.

3. Genomic Prediction by use of Machine Learning and AI
In current genomics prediction, machine learning and AI are now vital. They are able to capture complex, non-linear as well as high-dimensional relationships that simple statistical models fail to capture (Montesinos Lopez *et al.*, 2022). Such methods are particularly applicable to plant breeding, in which relationships among genes, environment, and various types of data introduce complexity that can be addressed with the help of flexible, data-driven approaches.

3.1 Genomic prediction core algorithms

Table 1: Algorithms used for genomic predictions

Algorithm	Key Features	Use in Genomic Prediction
Random Forest (RF)	Ensemble of decision trees; robust to overfitting	Effective in noisy, high-dimensional data and G×E studies ^[8] (Cuevas-Casado <i>et al.</i> , 2023)
Support Vector Machine (SVM)	Kernel-based; handles non-linear patterns	Competitive with DL in structured, small datasets (Su <i>et al.</i> , 2025)
Artificial Neural Networks (ANN)	Fully connected nodes; adaptive learning	Can capture additive and non-additive effects but needs careful tuning (Okeke <i>et al.</i> , 2021)
Deep Learning	CNN, RNN, LSTM, ResNet, Transformers	Strong in high-throughput data and sequence modelling ^{[11][12][31]} (Fang <i>et al.</i> , 2025; Ramos-Onsins <i>et al.</i> , 2024)

3.2 Applications in Major Crops

Maize: On large maize datasets, DL models like multilayer perceptrons (MLPs) and CNNs have been applied. These models outperform older GS methods particularly on traits such as the grain yield, flowering time, and the height of the plant. Their results become even better with the inclusion of environmental data (Montesinos Lopez *et al.*, 2022).

0.76, which is 30 per cent higher than traditional GS^[1] (Cao *et al.*, 2025).

Wheat

In numerous wheat trials, it was found that DL models worked. DL out-performed GBLUP in most of the cases, particularly when leaving one of the environments unsolved. The example is that CNN-LSTM models that utilized drone (UAV) images and map data forecasted winter wheat grain yield with an R² of

Rice

By the use of SNP data in conjunction with structural variant (SV) data, Ramos Onsins *et al.* (2024) enhanced their prediction accuracy of rice traits using dL pipelines. Such mixed inputs resulted into improved predictions of most traits. There was also an analysis by Fang *et al.* (2025) which demonstrated that a transformer based DL model that utilized tokenized k-mers of genome sequences performed better than CNNs in the Rice3K dataset.

3.3 Performance Metrics and Model Comparison

Table 2: Common performance metrics used for evaluating genomic prediction models

Metric	Description	Importance
R ² (Coefficient of determination)	Proportion of phenotypic variance explained	Higher = better model fit
RMSE (Root Mean Square Error)	Measures prediction error	Lower =higher accuracy
Pearson’s r	Linear correlation coefficient	Reflects strength of prediction
MAE(Mean Absolute Error)	Average absolute prediction error	More robust to outliers

4. High-throughput phenotyping integration

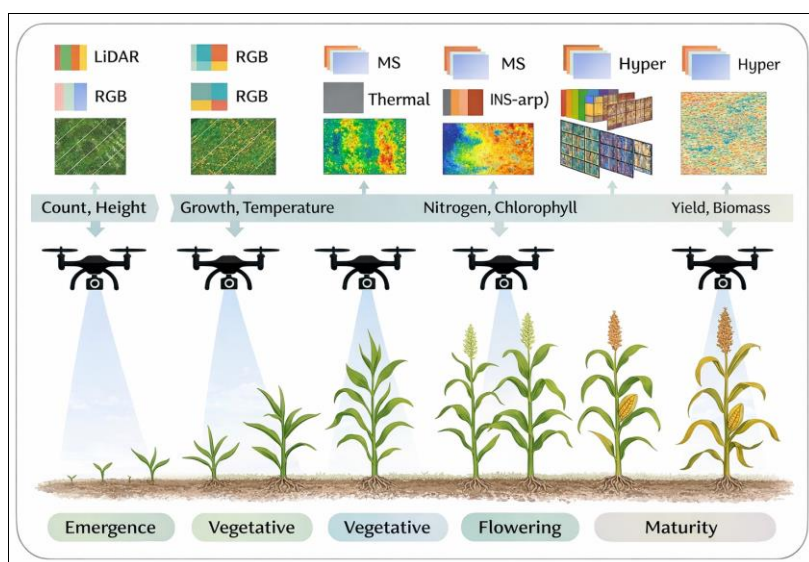
High-throughput phenotyping (HTP) tools are becoming more commonly utilized in modern breeding of plants. These instruments can be used to measure traits in large numbers, fast, and without being invasive. With the help of advanced imaging, drones, ground robots, and multi sensor platforms, breeders will be able to gather accurate and timely data in terms of phenotypes in large environments and over time. These technologies provide the additional knowledge about the characteristics that vary over time and allow a more precise evaluation of the performance of genotypes in the real world.

4.1 Drones, Sensors and Imaging Tools Role

RGB, multispectral, hyperspectral, thermal, and LiDAR drones (UAVs) have become established as instruments of collecting detailed canopy and field characteristics. The measurements taken with these tools are vegetation indices (e.g., NDVI), biomass, structural traits, and stress responses. As an illustration, drone imaging has been applied in wheat breeding to forecast the grain yield, weight as well as protein content at different growth stages with high accuracy of prediction^[18] (Kaushal *et al.*, 2024). Autonomous robots are also common and ground-based phenotyping systems. High-resolution phenotypic information

(plant height, stem diameter, and leaf area index) is collected on thousands of plots using robots like Terra Sentia. These systems supplement aerial platforms by resolving features missed in the canopy, and offer fine-scale resolution that is usually needed when modeling gene-by-environment (GxE) interactions. Especially useful is time-series phenotyping, in which data are

collected at various growth stages. It enables the evaluation of the dynamic trait expression and growth patterns. Examples of drone imaging over time building phenomic models that predict complex traits, even in the absence of genetic data, are in maize, rice, and forest species including pine.



Source: ^[18] Kaushal *et al.*, 2024

Fig 3: UAV-based RGB, multispectral, hyperspectral, thermal, and LiDAR imaging for high-throughput field phenotyping

4.2 Linking Phenomics with AI and Genomics

Due to the broad range of its applications, phenomics is connected to AI and Genomics as well (Hong *et al.*, 2018).

4.2.1 Improving Genomic Prediction

Phenomic features improve the power to predict phenotypes by providing spatial, temporal and physiological context of the genomic data in GS models. Multi-trait GS models that used the traits of UAVs as examples were found to be better predictors of wheat grain yields compared to single-trait models ^[18] (Kaushal *et al.*, 2024).

In a different article, a transformer-based DL model that was trained on UAV and environmental time-series data could predict yield at the plot level with high accuracy proving the strength of AI to combine phenomic with environmental data to predict yields in real-time.

4.2.2 Multi-Task Deep Learning Architectures (e.g., MtCro)

Multi-task deep learning (MTDL) methods enable modeling

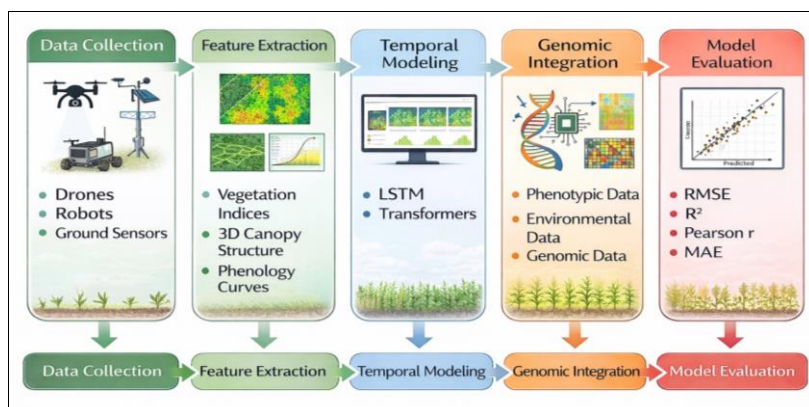
multiple characteristics based on shared neural architecture at once MtCro has been used on wheat and maize data in the recent past, and it has resulted in the 1-9% increase in prediction accuracy over the single-trait models ^[2] (Chao *et al.*, 2025). The models increase computational performance and interdependence of traits.

4.2.3 Artificial Intelligence-based DSSs

There are also new technologies that can guide breeders in real-time with the help of AI and simulate and chat bots. This new hybrid system ruptured UAV imagery, the WOFOST crop model, and a Transformer-based temporal fusion network and a big language model interface. This allowed the users to forecast the yield and perform a what-if analysis with interactive and chat-like inputs (Yang *et al.*, 2025).

4.3 The Workflow Model Performance and Integration

The general HTP and AI combination with genomic prediction workflow consists of:



Source: Made using canva

Fig 4: HTP and AI with genomic prediction workflow

- **Data Collection:** Collect high-resolution images and environmental data with drones, robots or ground sensors.
- **Feature Extraction:** Process raw sensor data to measure variables like vegetation indices, 3D canopy structure and phenology curves.
- **Temporal Modeling:** Model dynamic characteristics of traits over time using models like LSTMs or Transformers.
- **Genomic Integration:** Use both phenotypic and environmental data together with genomic data to predict using DL.
- **Model Evaluation:** The model evaluation involves the use of measures designed to evaluate performance such as RMSE, R 2, Pearson r, and MAE to provide robust and reliable predictions.

Table 3: HTP Platforms and Traits

Platform	Sensor Types	Key Traits Captured	Notable Study
UAV (RGB, multispectral)	NDVI, biomass, stress metrics	Grain yield, canopy traits	^[18] Kaushal <i>et al.</i> , 2024
Ground robot (TerraSentia)	RGB, LiDAR	Height, stem thickness, LAI	Nature Maize Study, 2025
Multi-task DL (MtCro)	Genomics + UAV features	Simultaneous trait prediction	^[2] Chao <i>et al.</i> , 2025

5. AI/GS Applications Across Traits in Major Crops

Table 4: Application of AI/GS in major crops

Crop	Trait Type	AI/GS Model Used	Details of Improvement / Performance	Reference
Wheat	Heat tolerance	DL + UAV-based NDVI and Canopy Temperature (CT) data	A DL model of aerial image and CT features enhanced prediction accuracy of the yield of grains in heat stress by 32% relative to conventional GBLUP models. This enabled the breeders to choose genotypes that are heat-resistant at early stages of growth.	^[18] Kaushal <i>et al.</i> (2024)
Maize	Drought tolerance	CNN-based DL with G×E interaction modeling	CNN models that included environmental covariates in more than one site showed a great enhancement in predicting the drought tolerance. Multi-environment trials improved the accuracy of prediction (R2 = 0.48, compared with R2 = 0.36, linear).	Montesinos López <i>et al.</i> (2023)
Rice	Grain quality traits	CNN) + SVs + SNPs	DL models based on both SVs and SNPs forecasted grain characteristics such as length, width, and weight with 20 percent more accuracy as compared to Bayesian models. This gave SVs added predictive value in 87 out of 100.	^[31] Ramos-Onsins <i>et al.</i> (2024)
Wheat	Stripe rust resistance	Deep Neural Networks (DNN) + Environmental Covariates	The use of a multi-layer DNN based on genomic measures and past rust severity measurements, together with local weather data, increased the prediction of rust severity by 35 percent as compared to the traditional Bayesian GS. Particularly efficient in the high disease pressure areas.	Zhou <i>et al.</i> (2023)
Rice	Blast disease resistance	CNN LSTM hybrid + Multi omics (transcriptomics, SNP, metabolomics)	The CNN LSTM models were trained using combined omics data and it was found that these models could classify resistant and susceptible rice lines with 92% accuracy in diagnosing the blast disease. Importance was also ranked as to functional markers.	Xie <i>et al.</i> (2025)
Maize	MLN resistance	RF Classifier using SNP and Phenotypic Severity Data	An RF model supervised on a dataset of about 10,000 maize hybrids obtained F1-score of 0.84 in a task of classifying levels of maize lethal necrosis (MLN) resistance. There was high performance despite a very high genotype x environment variability.	Mwangi <i>et al.</i> (2024)

Another new direction is the implementation of federated learning (FL) a decentralized model of machine learning in which models are trained on multiple datasets without sharing raw data. The breeding programs can jointly enhance prediction models using this approach without jeopardizing the privacy of data or ownership. Federated structures are especially useful in cases when data is isolated in national institutions or companies with stringent data-sharing policies ^[5] (Choudhury *et al.*, 2024). Another recent simulation of wheat trials in various countries also showed that FL models were equally accurate as centralized models and achieved better data security and local autonomy ^[3] (Chen *et al.*, 2025). This is particularly important when sensitive breeding databases with proprietary germplasm or government-sponsored breeding plans are involved. The other significant opportunity is to combine multi-environment trials (METs) and AI-based GS. Genomic prediction models which are traditionally used are not scalable to interactions with G×E. Nevertheless, the potential of utilizing improved predictions of heterogeneous agro-ecological zones will be significantly enhanced through the combination of the progress in enviromics and environmental characterization with the models of DL. More recent models, such as EnvRNN and GeoCNN, are being deployed to add environmental covariates

(e.g., soil type, rainfall, temperature) to genomic prediction pipelines, which helps to improve cross-location and cross-season transferability of models (Lopez *et al.*, 2025; Weber *et al.*, 2024). At the same time, an expanding outburst of pangenomics and epigenomics are providing more promising sources of predictive features than traditional single-reference genomes. Pangenomes offer a more in-depth perspective of diversity in genomes by incorporating SVs, presence-absence variation and new genes among accessions (Tao *et al.*, 2023). These factors are currently being integrated into GS pipelines in order to increase the prediction of traits that are affected by rare variants or lineage-specific ones, including disease resistance and abiotic stress tolerance. As an example, the accuracy of DL models using structural variant-based features of pangenomes has been found to be enhanced to predict resistance to stripe rust in wheat and blast resistance in rice (Zhang *et al.*, 2024). Predicting trends of DNA methylation and histone modification is known as epigenomics that is also on the move in AI-enabled breeding. Although in its early stages with many crops, epigenetic data can be used to improve traits with complicated gene regulation, including flowering time and stress memory. Graph-based neural networks or attention-based DL models

already exist in the development of integrative models that exploit genomic, epigenomic, and transcriptomic layers to give answers in Arabidopsis and rice^[20] (Li *et al.*, 2024).

The most significant change in GS and AI future of breeding is, perhaps, the democratization of AI in open-source platforms. Projects such as DeepCCR, BreedBase, and AgML are providing ready-to-use pipelines, easy-to-use interfaces, and everything as a service that reduce the technical barrier to AI adoption. These also come in handy where breeding programs are in developing countries and where such countries do not have the expertise of data science in-house. As an illustration, DeepCCR offers automated DL pipelines that have parameters that can be adjusted to ensure that small teams can undertake complex trait prediction without writing code (Ma *et al.*, 2024).

The availability of publicly available data like the Rice3K, T₃ (Triticeae Toolbox) and the CIMMYT wheat dataset also enables researchers and breeders everywhere. Open-source machine learning tools like TensorFlow, PyTorch, scikit-learn, and XGBoost are being tailored to genomic prediction tasks, and plug-ins to imputed genotype handling, multi-trait analysis and environmental metadata are in development (Fang *et al.*, 2025). The integration of explainable AI (XAI) tools into these platforms is also being pursued to enable breeders to understand AI decisions and enhance transparency and comply with regulatory expectations of advanced biotech crops.

In the future, it is possible to imagine that with the integration of federated learning, omics, and the democratization of AI, there will be a brought about change in the way plant breeding is done. The next generation breeding pipelines can be defined as one that is dynamically trained on the latest trials, inter-institutional cooperation without data silos, and local breeding decisions. Such developments are predicted to not only improve breeding efficiency and genetic gain, but also make the state of the art technology more equitable, scalable, and inclusive a necessary move towards global food and nutrition security.

Conclusion: Use of AI in GS is introducing radical change in plant breeding. Due to the growing strain on agriculture through climate change, the growing world population, and the dwindling supply of arable land, there is a sense of urgency in the development of breeding methodologies that are quicker, more precise, as well as data-intensive. Conventional methods like phenotype-based selection, and MAS have been helpful in improving crops but mostly fail in cases where there are complex G×E interactions with abilities. Conversely, AI and, in particular, DL has become a significant strategy that can identify non-linear genetic patterns, simulate the interaction of genes, and combine large amounts of data in genomics, phenomics, and environmental data. This transition of traditional breeding to information-driven approaches is a significant milestone of increasing the genetic yield faster and benefiting food and nutritional security in the world.

Recent years have witnessed encouraging outcomes of AI-based strategies in key crops, such as rice, wheat, maize, and barley. Deep learning networks (CNNs, LSTM networks, transformers, etc.) have been continually more effective than traditional linear models, particularly in non-additive genetic effects of traits. With the addition of HTP and multi-omics data, these models become even more accurate. As an example, a combination of aerial drone-based phenotypic imaging and genomic data has enabled accurate prediction of multifactorial dimensions including yield and stress response. Likewise the integration of structural variations, transcriptome information, and metabolomic profiles into DL systems have contributed to our

current knowledge of hereditary traits and have also resulted in improved selection approaches.

The other important change is the rising availability of new and more sophisticated phenotyping and AI-assisted breeding platforms. Phenotypic data over large field trials could be gathered in time-series, and in detail, through the use of drones, automated field equipment, and multispectral sensors. When incorporated into a DL model, these data can provide useful clues to the temporal nature of such phenomena as plant growth, stress adaptation and flowering dynamics. Moreover, techniques like multi-task learning and reinforcement learning are allowing breeders to forecast multiple attributes at the same time and to make long term optimal breeding choices. Such a convergence of multiple datasets genotypic, phenotypic, and environmental is in line with the concepts of Breeding 5.0, a futuristic model combining AI, big data, and the latest genome editing technologies to make crop improvement programs more efficient and sustainable.

Although these are occurring, there are still a number of constraints which limit the extensive adoption of AI in breeding. Data quality is one of the significant problems. Many of the techniques, such as GBS and low-coverage genome sequencing, are cheap but yield incomplete or noisy data, which requires more complex imputation and filtering approaches. The field phenotyping is also susceptible to errors brought about by environmental variables such as lighting and soil heterogeneity. To overcome these challenges, the standardized data collection protocols and advanced data-cleaning pipelines will be used. Moreover, the technical aspect of AI such as the requirement to use high-performance computing, expansive memory storage, and special knowledge is a challenge to many public and resource-constrained breeding programs. Nevertheless platforms like DeepCCR and BreedBase are gaining ground by providing easy to use cloud-based offerings that lower the barriers to entry in terms of technology.

The interpretability of AI models is another significant issue. Though the predictions made by the DL approaches tend to be very accurate, their internal mechanism is usually obscure, hence the breeders might be unable to see why certain things happen. This is particularly a problem where attributes that are subject to regulatory interest are involved. New tools such as SHAP, LIME, and integrated gradients are starting to provide information on the decisions of models, although further adoption of these tools into breeding pipelines remains immature. In the future, innovations like federated learning that enables joint model generation without the sharing of raw data and the integration of pangenomic and epigenomic data are very promising. In addition to these, long-term spending on computational infrastructure, education, and policy will be necessary to make AI-based breeding technologies inclusive and scalable. To sum up, AI and GS merger is redefining the future of plant breeding with more accurate, effective and strategic plant breeding solutions to an ever-increasing global challenge.

References

1. Cao L, *et al.* Integrating UAV phenotyping and deep learning to enhance genomic prediction of yield in winter wheat. *Frontiers in Plant Science*. 2025;15:1410249. DOI:10.3389/fpls.2024.1410249.
2. Chao X, *et al.* MtCro: A multi-task deep learning framework for crop trait prediction. *Computational Biology and Agriculture*. 2025.
3. Chen R, Han W, Zhang H, *et al.* Federated learning for privacy-preserving genomic prediction across breeding

- centers. *Plant Genome*. 2025;18(1):e20345. DOI:10.1002/tpg2.20345.
4. Cheng X, Wang D. Deep learning in plant breeding: Progress and challenges. *Theoretical and Applied Genetics*. 2024;137(3):421-437.
 5. Choudhury D, *et al.* Collaborative AI and federated architectures in plant phenotyping and breeding. *Trends in Biotechnology*. 2024;42(2):132-145. DOI:10.1016/j.tibtech.2023.10.005.
 6. Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, de Los Campos G, *et al.* Genomic selection in plant breeding: Methods, models, and perspectives. *Trends in Plant Science*. 2017;22(11):961-975. DOI:10.1016/j.tplants.2017.08.011.
 7. Crossa J, Pérez-Rodríguez P, Montesinos-López O, *et al.* Genomic selection in plant breeding: Methods, models, and perspectives. *Trends in Plant Science*. 2021;26(10):1035-1056. DOI:10.1016/j.tplants.2021.07.003.
 8. Cuevas Casado R, Montesinos López OA, Montesinos López A, Crossa J, Cuevas J. A comparison of multivariate genomic prediction methods for grain yield and protein content in wheat. *Genes*. 2023;13(8):1494. DOI:10.3390/genes13081494.
 9. Dong Y, Chen S, Zhang J. Learnable Group Transform: Enhancing genotype-to-phenotype prediction for rice breeding. *arXiv*. 2025;arXiv:2503.11180.
 10. Elshire RJ, *et al.* A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLOS ONE*. 2011;6(5):e19379. DOI:10.1371/journal.pone.0019379.
 11. Fang R, He Y, Zhang Z. Hybrid deep learning approaches for improved genomic prediction in crop breeding. *Agriculture*. 2025;15(11):1171. DOI:10.3390/agriculture15111171.
 12. Fang R, Zhang Z, He Y. Transformer-based k-mer tokenized models for genomic selection. *arXiv*. 2025. DOI:10.48550/arXiv.2405.09585.
 13. Hayes BJ, *et al.* Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science*. 2009;92(2):433-443. DOI:10.3168/jds.2008-1646.
 14. Heffner EL, Sorrells ME, Jannink JL. Genomic selection for crop improvement. *Crop Science*. 2009;49(1):1-12. DOI:10.2135/cropsci2008.08.0512.
 15. Hickey JM, Chiurugwi T, Mackay I, Powell W. Genomic prediction unifies animal and plant breeding programs. *Nature Genetics*. 2017;49(9):1297-1303. DOI:10.1038/ng.3920.
 16. Jannink JL, Lorenz AJ, Iwata H. Genomic selection in plant breeding: From theory to practice. *Briefings in Functional Genomics*. 2010;9(2):166-177. DOI:10.1093/bfpg/elq001.
 17. Kamenya SN, *et al.* Emerging technologies for the genetic improvement of orphan crops in Africa. *Frontiers in Plant Science*. 2020;11:219. DOI:10.3389/fpls.2020.00219.
 18. Kaushal S, Gill HS, Billah MM. Enhancing phenomic-genomic prediction in wheat using UAV and deep learning. *Frontiers in Plant Science*. 2024;15:1410249.
 19. Li R, *et al.* Hybrid deep learning for genomic prediction in agriculture. *Agriculture*. 2025;15(11):1171.
 20. Li X, Younis S, He J. Multi-omic integration in deep learning for trait prediction in crops. *BMC Bioinformatics*. 2024;25(3):99. DOI:10.1186/s12859-024-05683-0.
 21. Lisboa PJG, Taktak AFG, Duarte L. Making neural networks interpretable for clinical decision support. *Artificial Intelligence in Medicine*. 2023;142:102556.
 22. Lopez FA, Singh P, Jarquín D. Evaluating deep learning and GBLUP for genomic prediction across multi-environment plant breeding datasets. *Frontiers in Genetics*. 2025;16:1183210.
 23. Lopez MA, Montesinos-López OA, Crossa J. Environmental deep learning models for multi-environment genomic predictions in cereals. *G3: Genes, Genomes, Genetics*. 2025;15(1):jkaa099. DOI:10.1093/g3journal/jkaa099.
 24. Ma X, *et al.* DeepCCR: Large-scale genomic-based deep learning for rice improvement. *Plant Biotechnology Journal*. 2024;22(3):450-466.
 25. Ma Y, Zhang P, Xu C. DeepCCR: An open-source platform for genomic prediction in rice breeding. *Plant Methods*. 2024;20(2):45. DOI:10.1186/s13007-024-01245-9.
 26. Meuwissen TH, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 2001;157(4):1819-1829.
 27. Montesinos López OA, *et al.* Deep learning improves genomic prediction in maize. *G3: Genes, Genomes, Genetics*. 2022;12(3). DOI:10.1093/g3journal/jkac015.
 28. Montesinos López OA, *et al.* Multi-trait and multi-environment deep learning for maize drought prediction. *G3: Genes, Genomes, Genetics*. 2023;13(2).
 29. Mwangi JM, Otieno W, Badu-Apraku B. ML-based detection of maize lethal necrosis resistance. *BMC Plant Biology*. 2024;24:187.
 30. Okeke UG, Akdemir D, Jannink JL. Assessing artificial neural networks for genomic prediction. *Theoretical and Applied Genetics*. 2021;134(4):1145-1160.
 31. Ramos-Onsins SE, Fang Y, Li C. Deep learning and structural variants in rice genomic prediction. *Plant Methods*. 2024;20:121.
 32. Roorkiwal M, *et al.* Integrating genomics and speed breeding. *The Plant Genome*. 2022;15(2):e20185. DOI:10.1002/tpg2.20185.
 33. Scheben A, *et al.* Towards CRISPR/Cas crops. *New Phytologist*. 2017;216(3):682-698. DOI:10.1111/nph.14702.
 34. Spindel JE, McCouch SR. When more is better: How data sharing would accelerate genomic selection. *New Phytologist*. 2016;212(4):814-826. DOI:10.1111/nph.14174.
 35. Su R, Lv J, Xue Y, *et al.* Genomic selection in pig breeding. *Genetics Selection Evolution*. 2025;57:13.
 36. Tao Y, Zhao X, Zhang Z. Construction of rice pangenome. *Nature Genetics*. 2023;55:1098-1107. DOI:10.1038/s41588-023-01365-2.
 37. TerraSentia Robotic Field Phenotyping Report. *Nature Biotechnology Reports*. 2025.
 38. Wang G, *et al.* LSTM autoencoder-based DNN for barley prediction. *arXiv*. 2024;arXiv:2404.10231.
 39. Weber R, Ravel C, Michonneau C. AI-based multi-environment modeling for durum wheat. *Field Crops Research*. 2024;310:109943. DOI:10.1016/j.fcr.2024.109943.
 40. Xie Y, Chen R, Qian X. Multi-omics prediction of rice blast resistance. *Rice*. 2025;18(1):65.
 41. Xu Y, *et al.* Smart breeding driven by big data and AI. *Molecular Plant*. 2022;15(11):1664-1695.
 42. Yang G, Jin N, Ai W, *et al.* Integrating remote sensing, deep learning, and LLM. *arXiv*. 2025.
 43. Younis OG, *et al.* Breeding program optimization with reinforcement learning. *arXiv*. 2024;arXiv:2401.03910.
 44. Zhang T, Fang Y, Onwuegbuzie E. Structural variant-aware DL improves rust prediction. *The Plant Genome*. 2024;17(1):e20321.
 45. Zhou X, Singh D, Gupta P. DL-based prediction of stripe rust. *Theoretical and Applied Genetics*. 2023;136(5):1245-1258.