

Research Insight

Open Access

Big Data Analytics in Enhancing Maize Breeding Programs

Xian Zhang, Jiamin Wang, Yunchao Huang ✉

Hainan Provincial Key Laboratory of Crop Molecular Breeding, Sanya, 572025, Hainan, China

✉ Corresponding email: yunchao.huang@hitar.orgBioscience Evidence, 2025, Vol.15, No.5 doi: [10.5376/be.2025.15.0025](https://doi.org/10.5376/be.2025.15.0025)

Received: 26 Aug., 2025

Accepted: 30 Sep., 2025

Published: 15 Oct., 2025

Copyright © 2025 Zhang et al., This is an open access article published under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preferred citation for this article:

Zhang X., Wang J.M., and Huang Y.C., 2025, Big data analytics in enhancing maize breeding programs, Bioscience Evidence, 15(5): 249-259 (doi: [10.5376/be.2025.15.0025](https://doi.org/10.5376/be.2025.15.0025))

Abstract With the development of high-throughput omics, remote sensing and artificial intelligence, big data is transforming corn breeding. Research shows that the combination of machine learning and multi-omics can better predict and screen the yield and stress resistance of corn, and also accelerate the breeding speed of new varieties. The emergence of unmanned aerial vehicle (UAV) sensors, deep learning, and federated learning has made high-throughput phenotyping, early yield prediction, and multi-party collaborative breeding work easier to achieve. Meanwhile, the multi-genome database of corn and the intelligent analysis platform have also laid the foundation for the integration and sharing of global resources. Of course, this process also poses many challenges, such as different data sources, the complexity of biological issues themselves, and the influence of socio-economic factors. Overall, however, big data has become an important force driving corn breeding to be more intelligent, precise and sustainable. Next, it is necessary to strike a balance between technological innovation and green development and enhance cooperation. Our research objective is to explore how these new methods can be utilized to help corn breeding serve global food security more efficiently.

Keywords Corn breeding; Big data analysis; Machine learning; High-throughput phenotype; Intelligent breeding

1 Introduction

Corn, as a major global food crop, high-yield, stress-resistant and highly adaptable varieties are of great significance to food security and sustainable agricultural development. Traditional corn breeding methods, such as population selection and hybrid breeding, were of great help to crop improvement in the past. However, these methods have deficiencies in accuracy and efficiency, and have been difficult to meet the increasing demand for food and cope with the complex environmental challenges (Andorf et al., 2019; Bhuiyan et al., 2025).

In recent years, high-throughput omics and information technology have developed rapidly, and agriculture has entered the era of big data. Big data contains information in multiple aspects such as genomics, phenotypes and the environment. Algorithms such as artificial intelligence (AI) and machine learning (ML) can be used to analyze and predict the complex relationship between genes and the environment (Nepolean et al., 2018; Najafabadi et al., 2023; Crossa et al., 2024; Farooq et al., 2024; Wu et al., 2024; Zhu et al., 2024). These technologies have made trait prediction in corn breeding more accurate, accelerated the cultivation of new varieties, and also promoted the emergence of intelligent and precise breeding (Jiang et al., 2019; Fritsche-Neto et al., 2021).

This study will introduce the progress of big data analysis in corn breeding, with a focus on the integration of multi-omics data and the application of AI and ML in gene mining, phenotypic prediction, and genomic selection. At the same time, the role of these methods in enhancing efficiency, improving decision-making and addressing future challenges will also be discussed. Finally, the existing problems will also be pointed out and the directions for future research will be proposed. Through interdisciplinary cooperation and technological innovation, big data-driven corn breeding is expected to provide support for global food security and sustainable development.

2 The Role of Big Data in Modern Breeding

2.1 Sources of big data in maize research

Modern corn breeding cannot do without various types of big data. These data provide a basis for trait prediction and genetic improvement. The main sources include:

Genomic data: As sequencing technology becomes increasingly advanced, the whole genome and pan-genome data of corn are constantly increasing. These data contain information such as sequences, genetic models, structural variations and transposon elements of different varieties, subspecies and wild relatives (Woodhouse et al., 2021; Sen et al., 2023).

Multi-omics data: including transcriptome, epigenome, proteome and metabolome, etc. This information can reveal the regulation of gene expression and metabolic pathways, and also explain the molecular basis of phenotypic differences (Sen et al., 2023; Wu et al., 2024).

High-throughput phenotypic data: By using technologies such as unmanned aerial vehicles, orbital platforms, remote sensing and image analysis, phenotypic data can be obtained at different growth stages, such as plant height, flowering period and leaf area (Meng et al., 2021; Guo et al., 2022; Adak et al., 2023; Li et al., 2023; Wu et al., 2024).

Environmental and management data: including climate, soil, fertilization and irrigation, etc., this information is important for understanding the interaction between genes and the environment and optimizing breeding methods (Meng et al., 2021; Guo et al., 2022; Li et al., 2023).

Database and resource platforms: such as MaizeGDB and Maize Feature Store, integrate and manage various types of data to facilitate the use and analysis by researchers (Woodhouse et al., 2021; Sen et al., 2023).

2.2 Integration of heterogeneous data

There are many types of data for corn breeding, including structured, semi-structured and unstructured data. How to effectively combine these data is the key to improving efficiency.

Multi-omics data fusion: Integrating genomic, phenotypic and metabolomic data, and then using machine learning and artificial intelligence methods to improve the accuracy of trait prediction. For example, the combination of SNP, image traits and metabolites can significantly improve yield prediction (Adak et al., 2023; Sen et al., 2023; Wu et al., 2024).

Database and platform integration: Databases like MaizeGDB, which adopt the pan-genome framework and link genomic, expression, methylation and variation data from different sources, can support cross-species and cross-environment comparisons (Woodhouse et al., 2021; Sen et al., 2023).

Semantic and ontological integration: Through semantic framework and ontological methods, the problem of inconsistent data meaning can be solved, enabling intercommunication and unified retrieval of different data sources.

Distributed and cloud computing platforms: By leveraging Hadoop, Spark and cloud platforms, large-scale data can be efficiently stored and processed, and real-time integration can also be achieved.

Intelligent Algorithms and Data Cleaning: Facing multi-source complex data, methods such as intelligent clustering and anomaly detection can be used to improve data quality and integration efficiency.

Although big data integration has promoted the intelligence of corn breeding, there are still many problems, such as inconsistent standards, different data meanings, unstable quality and privacy security. In the future, it is necessary to strengthen cross-domain cooperation, improve data standards and sharing mechanisms, and enhance the integration capabilities of real-time, cross-domain and unstructured data.

3 Analytical Tools and Approaches

3.1 Machine learning and AI in breeding

Machine learning (ML) and artificial intelligence (AI) have become two important tools in current corn breeding. These two methods mainly automatically extract information from a large amount of genomic, phenotypic and environmental data to achieve the purpose of improving the accuracy and efficiency of trait prediction (Esposito et al., 2019; Xu et al., 2022; Yan and Wang, 2022; Crossa et al., 2024; Wu et al., 2024; He et al., 2025) (Figure 1).

The commonly used specific methods at present include support vector machines, random forests, gradient boosting and deep neural networks, etc. These algorithms and deep learning (DL) models have been widely applied in various techniques such as genotype-phenotype prediction, gene mining, and multitrait improvement (Galli et al., 2021; Yan et al., 2021; Kudiyarasudevi and Suresh, 2024; Wu et al., 2025). Platforms similar to AutoGP combine multiple ML and DL models, allowing users to select the most appropriate algorithm for faster and more accurate genomic selection. meta-ensemble learning integrates different algorithms to make multi-trait prediction more stable and better adaptable to different situations

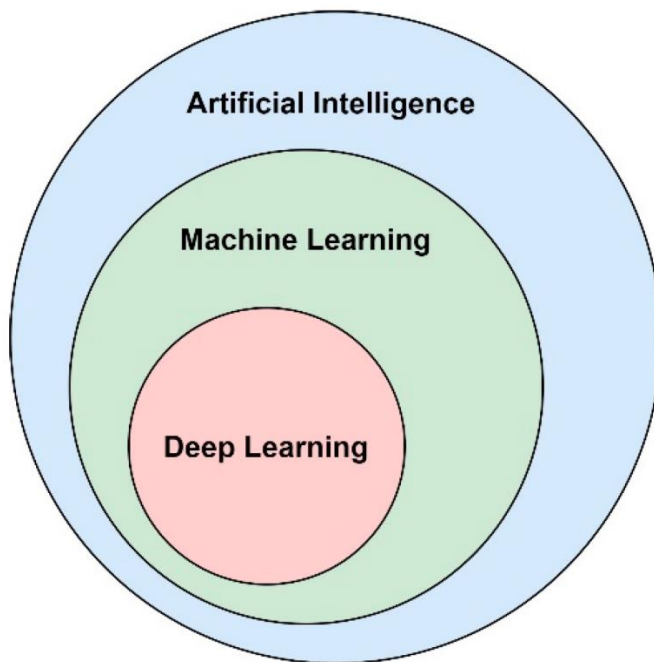


Figure 1 Relationship between AI, ML, and DL (Adopted from Cravero et al., 2022)

3.2 Genomic selection and predictive breeding

Genomic selection (GS) uses whole-genome markers and statistical or machine learning models to predict an individual's breeding values. This can shorten the breeding time and also improve genetic progress (Tong and Nikoloski, 2020; Xu et al., 2022; Mora-Poblete et al., 2023; Barreto et al., 2024; Kudiyarasudevi and Suresh, 2024; Wu et al., 2024; He et al., 2025; Wu et al., 2025). Nowadays, some new methods, such as deep learning, automated machine learning (AutoML), and ensemble learning, can handle high-dimensional and multi-type data. These methods have significantly improved the prediction accuracy of complex traits such as yield and stress resistance. For example, after the AutoML framework combines the dimensionally reduced environmental parameters and feature labels, the prediction accuracy can be increased by 14% to 28%, which is helpful for the development of climate-adaptive corn varieties (He et al., 2025). Meanwhile, deep learning models with multiple traits and multiple environments perform better than traditional Bayesian models and linear hybrid models in the prediction of complex traits such as flowering period (Mora-Poblete et al., 2023).

3.3 Data visualization and decision-support systems

With the increasing volume and complexity of data, the role of data visualization and decision support systems (DSS) in corn breeding is becoming more and more important. Modern intelligent breeding platforms (such as AutoGP, CropGBM, etc.) all come with visual interfaces. They can present multi-dimensional data such as genotype, phenotype and environment in an interactive way, helping researchers intuitively understand the prediction results and key factors (Yan et al., 2021; Wu et al., 2025). These systems not only improve the efficiency of data interpretation, but also provide references for breeding decisions, such as selecting parental combinations and screening superior strains. In addition, the decision support system driven by big data can also automatically recommend the best breeding plan, promoting intelligent and precise breeding (Esposito et al., 2019).

4 Applications in Maize Breeding

4.1 Trait improvement examples

Nowadays, corn breeding mostly utilizes big data and new biotechnologies. Researchers have made breakthroughs in multiple aspects of corn, including yield, drought resistance, disease resistance and nutritional quality, through techniques and methods such as genomic selection (GS), multi-omics data and machine learning models. Lorenzo et al. (2022) utilized the BREEDIT platform and CRISPR/Cas9 technology to simultaneously improve 48 growth-related genes in corn. The improved corn material has seen a 5% to 10% increase in leaf length and a 20% increase in leaf width. Molecular marker-assisted selection and genomic selection have also been widely applied in drought and disease resistance improvement, significantly enhancing the yield and stability of maize under adverse conditions (Gedil and Menkir, 2019; Liu and Qin, 2021; Prasanna et al., 2021; He et al., 2024).

4.2 Accelerated breeding cycles

Big data and new technologies have significantly shortened the time required for corn breeding. Ploidy doubling (DH) and genome editing techniques (such as the IMGE system) can obtain homozygous superior lines within two generations, which is much faster than traditional methods (Nepolean et al., 2018; Wang et al., 2019; Prasanna et al., 2021). Meanwhile, high-throughput genotyping and phenotypic analysis, combined with automated data processing platforms (such as AutoGP), can quickly obtain and analyze genotype-phenotypic data, improving the selection efficiency. Furthermore, facilities such as "speed breeding" and digital greenhouses can enable corn to complete multiple generations of cycles in a year by controlling photoperiod and growth environment, accelerating the breeding and promotion of new varieties (Singh et al., 2020).

4.3 Digital platforms and breeding networks

Digital platforms and global breeding networks have also made cooperation and innovation in corn breeding smoother. Platforms like AutoGP, which combine genotype extraction, phenotype extraction, GS models and multiple machine learning algorithms, can provide users with one-stop intelligent breeding tools and lower the threshold for using complex models (Wu et al., 2025). International maize improvement projects (such as IITA, CIMMYT) have promoted cross-institutional and cross-regional cooperation and variety sharing through digital data management, decision support systems and high-throughput phenotypic platforms (Gedil and Menkir, 2019; Prasanna et al., 2021; Liu et al., 2025). These platforms have not only improved the utilization rate of data, but also accelerated the adaptation and promotion of new varieties in different regions, promoting the modernization of global corn breeding.

5 Case Study: Big Data in Action for Maize Breeding

5.1 Background of the program

With the development of high-throughput omics and automated phenotypic technologies, corn breeding has entered the stage of big data. Traditional methods have limited efficiency and accuracy when dealing with complex traits and multi-environment data. To better predict yield and improve traits, some international and regional breeding projects have begun to introduce multi-omics data, environmental information and artificial intelligence tools to make maize improvement smarter and more data-dependent (Beyene et al., 2021; Fritsche-neto et al., 2021; Sarzaeim et al., 2022; Zhang et al., 2023; Wu et al., 2024).

5.2 Implementation of big data analytics

In 2024, Wu's team integrated multi-omics data from 156 maize recombinant inbred lines (including 2 496 SNP markers, 46 image traits across 16 growth stages, and 133 major metabolites), all collected through an automated phenotypic platform. The team used methods such as partial least squares, random forests, and Gaussian process regression to predict its yield, and compared different data types and feature screening methods. Research has confirmed that the use of multi-omics data combined with the random forest model can significantly improve the prediction accuracy (Wu et al., 2024). In 2021, Beyene et al. carried out international projects such as CIMMYT. In tropical corn breeding, they compared different genomic selection strategies by leveraging the genotype and phenotype data accumulated over many years and optimized the predictions through cross-validation.

5.3 Outcomes

After the combination of multi-omics and machine learning, the accuracy rate of corn yield prediction increased from 0.32 to 0.43, which was significantly better than that of single data (Wu et al., 2024).

Genomic selection has achieved early screening of new strains in tropical maize, with the highest prediction accuracy reaching 0.71, saving a large amount of experiments and resources (Beyene et al., 2021).

Federated learning methods enable different institutions to collaborate and improve model performance and prediction accuracy without sharing raw data (Zhang et al., 2023).

Deep learning combined with high-throughput phenotypes has also improved the early prediction of complex traits such as yield and flowering period, accelerating the progress of breeding (Sarzaeim et al., 2022; Shu et al., 2022).

5.4 Lessons learned

Integrating multi-omics and multi-source data can improve the prediction of complex traits, but it also poses higher requirements for data quality, feature screening and model selection (Beyene et al., 2021; Sarzaeim et al., 2022; Wu et al., 2024).

Cross-institutional and cross-regional data collaboration, such as federated learning, can enhance the generalization ability of models while protecting privacy, and is suitable for large-scale breeding networks (Fritsche-Neto et al., 2021; Zhang et al., 2023).

New technologies such as high-throughput phenotyping and deep learning need to be combined with traditional breeding experience in order to truly achieve the transition from "data-driven" to "decision support" (Beyene et al., 2021; Sarzaeim et al., 2022; Shu et al., 2022).

Continuous improvement of data collection, cleaning and feature engineering processes is the basis for the success of big data breeding projects.

6 Challenges and Limitations

6.1 Technical barriers

There are many technical challenges in big data analysis in corn breeding. First of all, the volume of data is extremely large and its sources are complex, including genomic, phenotypic, environmental and management information, etc. This puts a lot of pressure on data storage, transmission and processing (Kamilaris et al., 2017; Nepolean et al., 2018; Cravero et al., 2022; Xu et al., 2022). Data cleaning, integration and standardization are also very difficult, especially for unstructured data such as images, text and sensor data, which are not easy to automatically process and extract effective information (Onsongo et al., 2022). In addition, machine learning and deep learning models are highly dependent on high-quality, well-labeled data. However, agricultural data often have missing parts, noise and inconsistent labels, which can reduce the prediction accuracy and generalization ability of the model (Govaichelvan et al., 2023; Crossa et al., 2024; Kudiyarasudevi and Suresh, 2024; Wu et al., 2024) (Figure 2). Meanwhile, the shortage of high-performance computing resources and professional data talents also limits the application of big data in breeding (Lassoued et al., 2021).

6.2 Biological complexity

The traits of corn are influenced by many factors, including genotype, environment, and the interaction between genes and the environment ($G \times E$). This makes the breeding problem highly biologically complex (Nepolean et al., 2018; Xu et al., 2022; Crossa et al., 2024). Integrating multi-dimensional omics data can improve the prediction effect, but the model will become more complex and more difficult to interpret (Wu et al., 2024). Complex traits, such as yield and stress resistance, are often jointly determined by polygenic, epigenetic and metabolic networks, and vary greatly in different environments, which also limits the migration and generalization ability of the model. In addition, the genetic basis of some traits is still unclear, and there is a lack of reliable functional annotation and biological verification, which poses obstacles to precision breeding (Nepolean et al., 2018; Xu et al., 2022).

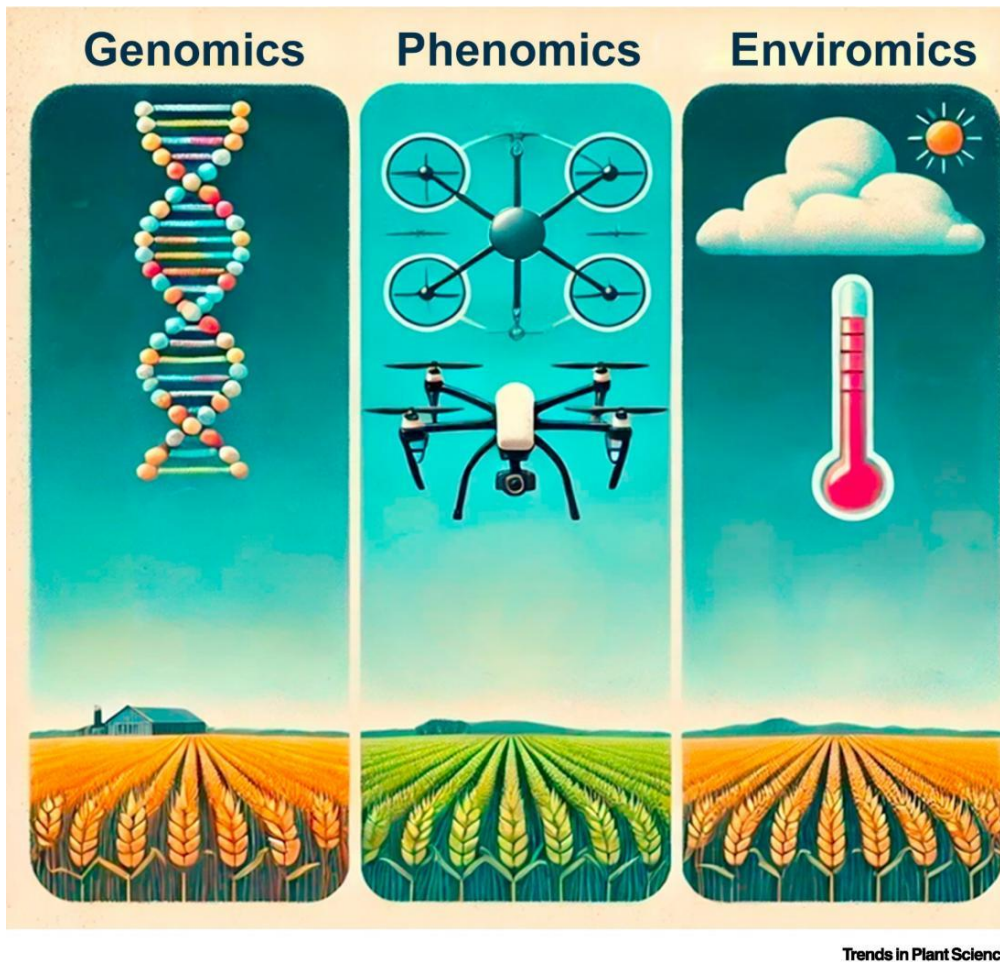


Figure 2 Components of modern plant breeding include not only phenotypic data collected from observed field cultivar trials but also genomic (molecular markers), phenomic (images from drones, airplanes, satellites), and enviromic (temperature, sun radiation, precipitation, soil humidity) data (Adopted from Crossa et al., 2024)

6.3 Socioeconomic and ethical considerations

The promotion of big data breeding is also influenced by social and ethical aspects. First of all, data collection, storage and analysis all require high costs and long-term infrastructure support. Developing countries and small and medium-sized breeding institutions often have difficulty affording these costs, which can easily lead to the digital divide (Bhat and Huang, 2021; Lassoued et al., 2021). Secondly, issues of data privacy, security and intellectual property rights are prominent. Many institutions are reluctant to share data, which will affect large-scale cooperation and innovation (Lassoued et al., 2021; Xu et al., 2022). In addition, the governance, standards and ethical norms of agricultural big data are still not perfect. Sensitive issues such as farmers' rights and interests, data ownership and algorithm transparency require multi-party cooperation and policy promotion.

7 Future Perspectives

7.1 Next-generation breeding with big data

In the future, corn breeding will enter the "intelligent breeding" stage. It will rely on big data, artificial intelligence (AI) and multi-omics ensemble prediction (such as iGEP) to combine the information of genotype, phenotype and environment (G-P-E) (Xu et al., 2022; Zhu et al., 2024; Liu et al., 2025). AI and machine learning will drive the automation of the entire process, making it more efficient and accurate from gene discovery, functional gene mining to complex trait prediction (Farooq et al., 2024; Wu et al., 2024; Zhu et al., 2024) (Figure 3). Intelligent Precision Design Breeding (IPDB) proposes an open, collaborative and shared platform to promote cooperation among biologists, informatics experts, breeders and farmers. Meanwhile, the combination of cross-species prediction, pan-genome and environmental omics will also bring new ideas for the improvement of complex traits and adaptive breeding.

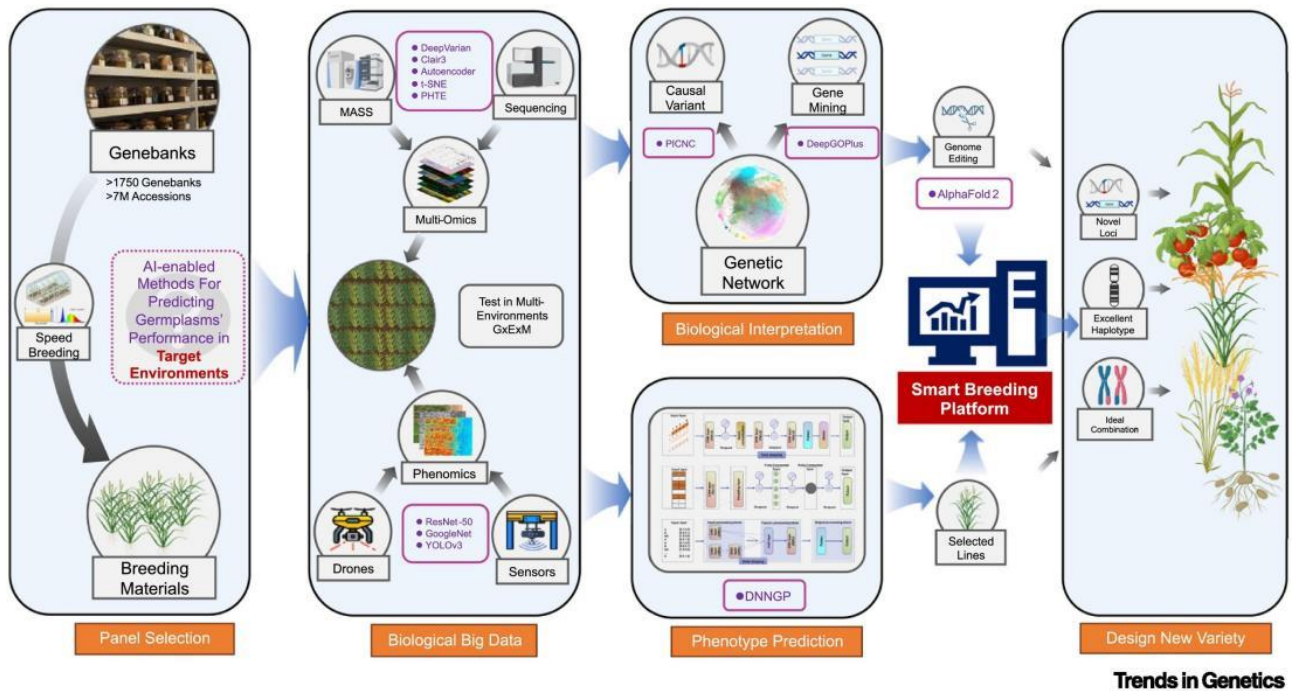


Figure 3 A roadmap of artificial intelligence (AI)-enabled plant breeding (Adopted from Farooq et al., 2024)

7.2 Scalability

As data becomes larger and more complex, big data analysis in corn breeding requires better scalability. Cloud computing, distributed storage and high-performance computing platforms will become the basis for processing large-scale data (Xu et al., 2022; Govaichelvan et al., 2023; Zhu et al., 2024). Data sharing and joint analysis among different institutions and regions (such as federated learning) are expected to break data silos and enable models to be better applied globally (Zhao et al., 2021). Meanwhile, improving the algorithm, automating feature screening and an efficient data cleaning process will also enhance the efficiency and practicality of the analysis (Wu et al., 2024). In the future, the popularization of low-cost genotyping and high-throughput phenotypic technologies will lower the threshold, enabling breeding projects of different scales and resource conditions to use big data methods.

7.3 Towards sustainable maize production

Big data combined with artificial intelligence-driven smart breeding (emphasizing the use of computers and algorithms to improve breeding) is providing new support for sustainable corn production. By more accurately predicting and optimizing the interaction between genotypes and the environment, new varieties with high yield, strong stress resistance and higher resource utilization rate can be bred, thereby enhancing food security and environmental adaptability (Xu et al., 2022; Farooq et al., 2024; Liu et al., 2025). Meanwhile, digital decision support systems and intelligent management platforms also play a role in precision agriculture, such as more scientific management of fertilizer and water, prevention and control of pests and diseases, and reduction of carbon emissions. All these can promote the development of green agriculture (emphasizing ecological protection and low-carbon development) (Wolfert et al., 2017; Lassoued et al., 2021). Furthermore, the open and shared breeding platform and the sound data governance system will further promote international cooperation and technological innovation, and provide assistance in addressing the challenges brought about by climate change and population growth (Zhu et al., 2024).

8 Conclusion

With the help of big data analysis in combination with machine learning and artificial intelligence (common methods in intelligent breeding), the prediction process of corn breeding has become faster and more accurate. After combining data from different sources such as genomics, phenotypic information and environmental conditions, the algorithm can automatically select key features and improve the model, thereby making the

prediction of complex traits such as yield and stress resistance more reliable. The combination of multi-omics data (such as genes, transcriptomics, metabolomics and other multi-level data) with ensemble learning and deep learning methods has significantly improved the accuracy of yield prediction, with some studies even achieving an increase of over 12%. Meanwhile, new approaches such as high-throughput technology, unmanned aerial vehicle (UAV) remote sensing, and multimodal data fusion have also made the identification and rapid screening of early traits more efficient, thereby accelerating the progress of breeding. In addition, methods like federated learning can achieve data sharing and joint modeling among different institutions without directly exchanging raw data. This not only enhances the performance of the model but also improves the efficiency of resource utilization.

Big data has become an important driving force for corn breeding. It has brought about three changes: First, it has shifted breeding decisions from relying on experience to relying on data, achieving more automation and intelligence; Second, through large-scale data integration, the genetic analysis and prediction accuracy of complex traits have been enhanced. Third, it has promoted cross-disciplinary and cross-institutional cooperation and facilitated the development of new concepts such as Intelligent Precision Design Breeding (IPDB). Big data not only accelerates the breeding and promotion of new varieties, but also provides support for addressing climate change and food security.

Looking ahead, corn breeding needs to strike a balance among technological innovation, open sharing and sustainable development. Technically, it is necessary to continue developing efficient algorithms, low-cost high-throughput technologies and intelligent decision-making platforms to enhance the efficiency and practicality of big data analysis. In terms of inclusiveness, efforts should be made to promote data sharing and the construction of open platforms, reduce the digital divide, and enable different regions to have fair access to global breeding resources. In terms of sustainability, intelligent breeding should serve green agriculture, cultivate new varieties that are high-yielding, stress-resistant and have high resource utilization rates, and help the global food system transform towards sustainability.

Acknowledgments

Thank you to the project team for the careful guidance and strong support.

Conflict of Interest Disclosure

The authors affirm that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Adak A., Kang M., Anderson S., Murray S., Jarquín D., Wong R., and Katzfuss M., 2023, Phenomic data-driven biological prediction of maize through field-based high throughput phenotyping integration with genomic data, *Journal of Experimental Botany*, 74(17): 5307-5326.
<https://doi.org/10.1093/jxb/erad216>
- Andorf C., Beavis W., Hufford M., Smith S., Suza W., Wang K., Woodhouse M., Yu J., and Lübberstedt T., 2019, Technological advances in maize breeding: past, present and future, *Theoretical and Applied Genetics*, 132: 817-849.
<https://doi.org/10.1007/s00122-019-03306-3>
- Barreto C., Dias K., De Sousa I., Azevedo C., Nascimento A., Guimarães L., Guimarães C., Pastina M., and Nascimento M., 2024, Genomic prediction in multi-environment trials in maize using statistical and machine learning methods, *Scientific Reports*, 14: 1062.
<https://doi.org/10.1038/s41598-024-51792-3>
- Beyene Y., Gowda M., Pérez-Rodríguez P., Olsen M., Robbins K., Burgueño J., Prasanna B., and Crossa J., 2021, Application of genomic selection at the early stage of breeding pipeline in tropical maize, *Frontiers in Plant Science*, 12: 685488.
<https://doi.org/10.3389/fpls.2021.685488>
- Bhat S., and Huang N., 2021, Big data and ai revolution in precision agriculture: survey and challenges, *IEEE Access*, 9: 110209-110222.
<https://doi.org/10.1109/ACCESS.2021.3102227>
- Bhuiyan M., Noman I., Aziz M., Rahaman M., Islam M., Manik M., and Das K., 2025, Transformation of plant breeding using data analytics and information technology: innovations, applications, and prospective directions, *Frontiers in Bioscience*, 17(1): 27936.
<https://doi.org/10.31083/FBE27936>
- Cravero A., Pardo S., Sepúlveda S., and Muñoz L., 2022, Challenges to use machine learning in agricultural big data: a systematic literature review, *Agronomy*, 12(3): 748.
<https://doi.org/10.3390/agronomy12030748>

- Crossa J., Montesinos-López O., Costa-Neto G., Vitale P., Martini J., Runcie D., Fritsche-Neto R., Montesinos-López A., Pérez-Rodríguez P., Gerard G., Dreisigacker S., Crespo-Herrera L., Pierre C., Lillemo M., Cuevas J., Bentley A., and Ortiz R., 2024, Machine learning algorithms translate big data into predictive breeding accuracy, *Trends in Plant Science*, 30(2): 167-184.
<https://doi.org/10.1016/j.tplants.2024.09.011>
- Esposito S., Carputo D., Cardi T., and Tripodi P., 2019, Applications and trends of machine learning in genomics and phenomics for next-generation breeding, *Plants*, 9(1): 34.
<https://doi.org/10.3390/plants9010034>
- Farooq M., Gao S., Hassan M., Huang Z., Rasheed A., Hearne S., Prasanna B., Li X., and Li H., 2024, Artificial intelligence in plant breeding, *Trends in genetics: TIG*, 40(10): 891-908.
<https://doi.org/10.1016/j.tig.2024.07.001>
- Fritsche-Neto R., Galli G., Borges K., Costa-Neto G., Alves F., Sabadin F., Lyra D., Morais P., De Andrade L., Granato Í., and Crossa J., 2021, Optimizing genomic-enabled prediction in small-scale maize hybrid breeding programs: a roadmap review, *Frontiers in Plant Science*, 12: 658267.
<https://doi.org/10.3389/fpls.2021.658267>
- Galli G., Sabadin F., Yassue R., De Souza C., Carvalho H., and Fritsche-Neto R., 2021, Automated machine learning: a case study of genomic “image-based” prediction in maize hybrids, *Frontiers in Plant Science*, 13: 845524.
<https://doi.org/10.3389/fpls.2022.845524>
- Gedil M., and Menkir A., 2019, An integrated molecular and conventional breeding scheme for enhancing genetic gain in maize in Africa, *Frontiers in Plant Science*, 10: 1430.
<https://doi.org/10.3389/fpls.2019.01430>
- Govaichelvan K., Pathmanathan D., Abidin R., and Abu A., 2023, Machine learning for major food crops breeding: applications, challenges, and ways forward, *Agronomy Journal*, 116(3): 1112-1125.
<https://doi.org/10.1002/agj.2.1393>
- Guo Y., Zhang X., Chen S., Wang H., Jayavelu S., Cammarano D., and Fu Y., 2022, Integrated UAV-based multi-source data for predicting maize grain yield using machine learning approaches, *Remote. Sens.*, 14: 6290.
<https://doi.org/10.3390/rs14246290>
- He B., Pan S., Zhao J., Zou X., Liu X., and Wu S., 2024, Maize improvement based on modern breeding strategies: progress and perspective, *ACS Agricultural Science & Technology*, 4(3): 274-282.
<https://doi.org/10.1021/acsagstech.3c00427>
- He K., Yu T., Gao S., Chen S., Li L., Zhang X., Huang C., Xu Y., Wang J., Prasanna B., Hearne S., Li X., and Li H., 2025, Leveraging automated machine learning for environmental data-driven genetic analysis and genomic prediction in maize hybrids, *Advanced Science*, 12(17): 2412423.
<https://doi.org/10.1002/adv.202412423>
- Jiang S., Cheng Q., Yan J., Fu R., and Wang X., 2019, Genome optimization for improvement of maize breeding, *Theoretical and Applied Genetics*, 133:1491-1502.
<https://doi.org/10.1007/s00122-019-03493-z>
- Kamilaris A., Kartakoullis A., and Prenafeta-Boldú F., 2017, A review on the practice of big data analysis in agriculture, *Comput. Electron. Agric.*, 143: 23-37.
<https://doi.org/10.1016/j.compag.2017.09.037>
- Kudiyarasudevi C., and Suresh S., 2024, Enhanced genomic prediction for maize breeding using deep learning and k-mer-based sequence encoding, 2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS), 933-940.
<https://doi.org/10.1109/ICICNIS64247.2024.10823173>
- Lassoued R., Macall D., Smyth S., Phillips P., and Hessel H., 2021, Expert insights on the impacts of, and potential for, agricultural big data, *Sustainability*, 13(5): 2521.
<https://doi.org/10.3390/SU13052521>
- Li Y., Wen W., Fan J., Gou W., Gu S., Lu X., Yu Z., Wang X., and Guo X., 2023, Multi-source data fusion improves time-series phenotype accuracy in maize under a field high-throughput phenotyping platform, *Plant Phenomics*, 5: 43.
<https://doi.org/10.34133/plantphenomics.0043>
- Liu H., Liu J., Zhai Z., Dai M., Tian F., Wu Y., Tang J., Lu Y., Wang H., Jackson D., Yang X., Qin F., Xu M., Fernie A., Zhang Z., and Yan J., 2025, Maize2035: A decadal vision for intelligent maize breeding, *Molecular Plant*, 18(2): 313-332.
<https://doi.org/10.1016/j.molp.2025.01.012>
- Liu S., and Qin F., 2021, Genetic dissection of maize drought tolerance for trait improvement, *Molecular Breeding: New Strategies in Plant Improvement*, 41: 8.
<https://doi.org/10.1007/s11032-020-01194-w>
- Lorenzo C., Debray K., Herwegh D., Develtere W., Impens L., Schaumont D., Vandeputte W., Aesaert S., Coussens G., De Boe Y., Demuyne K., Van Hautegeem T., Pauwels L., Jacobs T., Ruttink T., Nelissen H., and Inzé D., 2022, BREEDIT: a multiplex genome editing strategy to improve complex quantitative traits in maize, *The Plant Cell*, 35(1): 218-238.
<https://doi.org/10.1093/plcell/koac243>
- Meng L., Liu H., Ustin S., and Zhang X., 2021, Predicting maize yield at the plot scale of different fertilizer systems by multi-source data and machine learning methods, *Remote. Sens.*, 13: 3760.
<https://doi.org/10.3390/rs13183760>

- Mora-Poblete F., Maldonado C., Henrique L., Uhdre R., Scapim C., and Mangolim C., 2023, Multi-trait and multi-environment genomic prediction for flowering traits in maize: a deep learning approach, *Frontiers in Plant Science*, 14: 1153040.
<https://doi.org/10.3389/fpls.2023.1153040>
- Najafabadi M., Hesami M., and Eskandari M., 2023, Machine learning-assisted approaches in modernized plant breeding programs, *Genes*, 14(4): 777.
<https://doi.org/10.3390/genes14040777>
- Nepolean T., Kaul J., Mukri G., and Mittal S., 2018, Genomics-enabled next-generation breeding approaches for developing system-specific drought tolerant hybrids in maize, *Frontiers in Plant Science*, 9: 361.
<https://doi.org/10.3389/fpls.2018.00361>
- Onsongo G., Fritsche S., Nguyen T., Belemlihi A., Thompson J., and Silverstein K., 2022, ITALLIC: A tool for identifying and correcting errors in location based plant breeding data, *Comput. Electron. Agric.*, 197: 106947.
<https://doi.org/10.1016/j.compag.2022.106947>
- Prasanna B., Cairns J., Zaidi P., Beyene Y., Makumbi D., Gowda M., Magorokosho C., Zaman-Allah M., Olsen M., Das A., Worku M., Gethi J., Vivek B., Nair S., Rashid Z., Vinayan M., Issa A., Vicente S., Dhliwayo T., and Zhang X., 2021, Beat the stress: breeding for climate resilience in maize for the tropical rainfed environments, *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, 134: 1729-1752.
<https://doi.org/10.1007/s00122-021-03773-7>
- Sarzaeim P., Muñoz-Arriola F., and Jarquin D., 2022, Climate and genetic data enhancement using deep learning analytics to improve maize yield predictability, *Journal of Experimental Botany*, 73(15): 5336-5354.
<https://doi.org/10.1093/jxb/erac146>
- Sen S., Woodhouse M., Portwood J., and Andorf C., 2023, Maize feature store: a centralized resource to manage and analyze curated maize multi-omics features for machine learning applications, *Database: the Journal of Biological Databases and Curation*, 2023: baad078.
<https://doi.org/10.1093/database/baad078>
- Shu M., Fei S., Zhang B., Yang X., Guo Y., Li B., and Ma Y., 2022, Application of UAV multisensor data and ensemble approach for high-throughput estimation of maize phenotyping traits, *Plant Phenomics*, 2022: 9802585.
<https://doi.org/10.34133/2022/9802585>
- Singh R., Prasad A., Muthamilarasan M., Parida S., and Prasad M., 2020, Breeding and biotechnological interventions for trait improvement: status and prospects, *Planta*, 252: 54.
<https://doi.org/10.1007/s00425-020-03465-4>
- Tong H., and Nikoloski Z., 2020, Machine learning approaches for crop improvement: Leveraging phenotypic and genotypic big data, *Journal of Plant Physiology*, 257: 153354.
<https://doi.org/10.1016/j.jplph.2020.153354>
- Wang B., Zhu L., Zhao B., Zhao Y., Xie Y., Zheng Z., Li Y., Sun J., and Wang H., 2019, Development of a haploid-inducer mediated genome editing system for accelerating maize breeding, *Molecular Plant*, 12(4): 597-602.
<https://doi.org/10.1016/j.molp.2019.03.006>
- Wolfert S., Ge L., Verdouw C., and Bogaardt M., 2017, Big data in smart farming – a review, *Agricultural Systems*, 153: 69-80.
<https://doi.org/10.1016/j.agsy.2017.01.023>
- Woodhouse M., Cannon E., Portwood J., Harper L., Gardiner J., Schaeffer M., and Andorf C., 2021, A pan-genomic approach to genome databases using maize as a model system, *BMC Plant Biology*, 21: 385.
<https://doi.org/10.1186/s12870-021-03173-5>
- Wu C., Luo J., and Xiao Y., 2024, Multi-omics assists genomic prediction of maize yield with machine learning approaches, *Molecular Breeding*, 44: 1-17.
<https://doi.org/10.1007/s11032-024-01454-z>
- Wu H., Han R., Zhao L., Liu M., Chen H., Li W., and Li L., 2025, AutoGP: An intelligent breeding platform for enhancing maize genomic selection, *Plant Communications*, 6(4): 101240.
<https://doi.org/10.1016/j.xplc.2025.101240>
- Xu Y., Zhang X., Li H., Zheng H., Zhang J., Olsen M., Varshney R., Prasanna B., and Qian Q., 2022, Smart breeding driven by big data, artificial intelligence and integrated genomic-enviromic prediction, *Molecular Plant*, 15(11): 1664-1695.
<https://doi.org/10.1016/j.molp.2022.09.001>
- Yan J., and Wang X., 2022, Machine learning bridges omics sciences and plant breeding, *Trends in Plant Science*, 28(2): 199-210.
<https://doi.org/10.1016/j.tplants.2022.08.018>
- Yan J., Xu Y., Cheng Q., Jiang S., Wang Q., Xiao Y., Ma C., Yan J., and Wang X., 2021, LightGBM: accelerated genomically designed crop breeding through ensemble learning, *Genome Biology*, 22: 271.
<https://doi.org/10.1186/s13059-021-02492-y>
- Zhang Q., Zhao X., Han Y., Yang F., Pan S., Liu Z., Wang K., and Zhao C., 2023, Maize yield prediction using federated random forest, *Comput. Electron. Agric.*, 210: 107930.
<https://doi.org/10.1016/j.compag.2023.107930>
- Zhang Z., Wang X., Zhang Y., Zhou K., Yu G., Yang W., Li F., Guan X., Zhang X., Yang Z., Xu C., and Xu Y., 2025, SPDC-HG: An accelerator of genomic hybrid breeding in maize, *Plant Biotechnology Journal*, 23: 1847-1861.
<https://doi.org/10.1111/pbi.70011>

- Zhao Y., Thorwarth P., Jiang Y., Philipp N., Schulthess A., Gils M., Boeven P., Longin C., Schacht J., Ebmeyer E., Korzun V., Mirdita V., Dörnte J., Avenhaus U., Horbach R., Cöster H., Holzapfel J., Ramgraber L., Kühnle S., Varenne P., Starke A., Schürmann F., Beier S., Scholz U., Liu F., Schmidt R., and Reif J., 2021, Unlocking big data doubled the accuracy in predicting the grain yield in hybrid wheat, *Science Advances*, 7(24): eabf9106.
<https://doi.org/10.1126/sciadv.abf9106>
- Zhu W., Li W., Zhang H., and Li L., 2024, Big data and artificial intelligence-aided crop breeding: Progress and prospects, *Journal of Integrative Plant Biology*, 67: 722-739.
<https://doi.org/10.1111/jipb.13791>

**Disclaimer/Publisher's Note**

The statements, opinions, and data contained in all publications are solely those of the individual authors and contributors and do not represent the views of the publishing house and/or its editors. The publisher and/or its editors disclaim all responsibility for any harm or damage to persons or property that may result from the application of ideas, methods, instructions, or products discussed in the content. Publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
