

## Feature Review

## Open Access

# Integrating Genomic Selection and Machine Learning for Predicting Maize Yield Under Drought

Jiayi Wu, Huijuan Xu, Qian Li ✉

Modern Agricultural Research Center, Cuixi Academy of Biotechnology, Zhuji, 311800, Zhejiang, China

✉ Corresponding author: [qian.li@cuixi.org](mailto:qian.li@cuixi.org)Maize Genomics and Genetics, 2025, Vol.16, No.3 doi: [10.5376/mgg.2025.16.0014](https://doi.org/10.5376/mgg.2025.16.0014)

Received: 13 Apr., 2025

Accepted: 24 May, 2025

Published: 16 Jun, 2025

**Copyright** © 2025 Wu 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:**

Wu J.Y., Xu H.J., and Li Q., 2025, Integrating genomic selection and machine learning for predicting maize yield under drought, Maize Genomics and Genetics, 16(3): 139-148 (doi: [10.5376/mgg.2025.16.0014](https://doi.org/10.5376/mgg.2025.16.0014))

**Abstract** Drought is one of the most severe abiotic stresses faced by maize (*Zea mays* L.) production worldwide, which seriously restricts the stability of crop yield. Traditional breeding methods have limited adaptability in the context of complex climate change, and more efficient prediction methods are urgently needed. This study integrates genomic selection (GS) and machine learning (ML) methods, and uses large-scale genotype, phenotype and environmental data to improve the accuracy of maize yield prediction under drought conditions. This article systematically reviews the latest progress in genomic prediction of drought resistance traits, analyzes typical machine learning algorithms suitable for crop modeling, and proposes a strategy for integrating GS and ML and a hybrid model framework construction method. The feasibility and practicality of this method are verified through actual cases such as the CIMMYT drought-resistant maize project and Chinese maize hybrids. At the same time, the model's portability and robustness in different ecological environments are also evaluated. This study provides a theoretical basis and technical path for AI-driven precision breeding, which has important guiding significance for the cultivation of new maize stress-resistant varieties under drought conditions.

**Keywords** Maize; Genomic selection; Machine learning; Drought stress; Yield

## 1 Introduction

Corn is often encountered with drought when it is planted. This situation is very common and will also cause a reduction in yield, which also affects global food security. Corn itself is afraid of water shortage, so many breeding experts and researchers are studying how to make it more drought-resistant (Amadu et al., 2025). In the past, to predict corn yield, people basically relied on its appearance, that is, whether it grew well, and then combined it with some simple statistical analysis. But this method is not very accurate. Because the trait of drought resistance is too complex, it is not determined by one gene, but by several genes working together. And the relationship between genes and the environment is also difficult to explain clearly. In addition, the drought situation is different from year to year, and it is difficult to predict accurately using the old method, which also affects the speed of breeding new drought-resistant varieties (Shikha et al., 2017; Dias et al., 2018; Fernandes et al., 2024).

Now the situation is different. Genomic technology and high-throughput phenotyping analysis are developing rapidly, and scientists can collect more and more detailed data. These new technologies also allow us to use better methods to predict yields. For example, genomic selection (GS) can use genome-wide markers to estimate whether a variety is worth breeding. Machine learning (ML) can process these complex data, build models that adapt to different environments and genotypes, and make more accurate predictions (Saleh et al., 2023). There are many benefits to combining GS and ML. It can not only analyze the complex relationships between genes, but also take into account the impact of the environment. And it can use data from different channels. In this way, we can more accurately predict corn yield performance in drought conditions (Azrai et al., 2024; Wu et al., 2024).

This study systematically reviewed the research on combining genomic selection with machine learning to predict maize yield under drought stress, explored the background and challenges of maize drought stress, the shortcomings of traditional prediction models, and the emerging potential of the genomic selection-machine

learning (GS-ML) framework, focused on methodological progress, and emphasized the importance of these integrated methods for accelerating the breeding of drought-tolerant maize varieties and ensuring sustainable crop production under the background of climate change. This study hopes to promote the development of drought-resistant maize varieties and better cope with agricultural production under the challenges of climate change.

## **2 Progress in Drought-Tolerance-Oriented Genomic Prediction**

### **2.1 Advances in genetic mapping for drought traits**

In order to figure out how corn resists drought, scientists used two common methods: genome-wide association analysis (GWAS) and quantitative trait loci (QTL). These two methods have found many gene locations related to drought resistance. Some improved GWAS models have also found hundreds of nucleotide variations (QTNs) related to grain yield and flowering time, many of which are related to transcription factors such as AP2-EREBP and TCP (Li et al., 2016; Yuan et al., 2019; Zhang et al., 2023; Amadu et al., 2025). Later, researchers combined high-throughput phenotyping analysis with GWAS, and found thousands of gene locations related to drought resistance traits. These results have given us a deeper understanding of how corn copes with drought (Wu et al., 2021; Li et al., 2024).

### **2.2 Molecular markers used in drought tolerance selection**

When breeding drought-resistant maize varieties, scientists use some molecular markers, such as SNP, SilicoDArT, RFLP, SSR and AFLP (Hao et al., 2011; Zhang et al., 2022; Chen et al., 2024). Among them, SNP markers are the most commonly used because they are numerous and contain a lot of information. They can help us discover useful genetic variants and provide a reliable basis for seed selection (Wang et al., 2019). Some studies have also combined QTL analysis with transcriptome data to further narrow the location range of drought-related genes, so that breeding goals are clearer (Marino et al., 2009; Li et al., 2024).

### **2.3 Limitations of traditional genomic prediction methods**

Old genomic prediction methods such as RR-BLUP perform generally well in predicting drought resistance. This is because drought resistance is complex in itself, involving not only many genes but also environmental influences (Amadu et al., 2025). Moreover, these methods usually cannot accurately capture the interactions between genes, and it is difficult to deal with differences caused by environmental changes (Dias et al., 2018; Zhang et al., 2022). Although we can now try to add some markers related to the trait, or use models to consider the interaction between genotype and environment, the effect is still limited. To improve the accuracy of predictions, especially in areas where drought is more severe, we must rely on stronger algorithms and more advanced models.

## **3 Machine Learning Approaches for Yield Prediction**

### **3.1 Typical ML models used in crop prediction (RF, XGBoost, ANN)**

When predicting crop yield, the three commonly used machine learning methods are: random forest (RF), extreme gradient boosting (XGBoost) and artificial neural network (ANN). Many studies have found that XGBoost and RF usually perform better than other models. In particular, XGBoost often gives higher  $R^2$  values, which means that the prediction is more accurate and the error is smaller (Dhaliwal et al., 2022; Shawon et al., 2023; Gharakhanlou and Perez, 2024). Artificial neural networks are also very popular, especially when complex relationships need to be handled. However, it has higher requirements for data and computing (Van Klompenburg et al., 2020; Malphedwar et al., 2024). Sometimes, combining several models together, such as making a hybrid model or an integrated model, can further improve the prediction accuracy (Oikonomidis et al., 2022).

### **3.2 Data normalization and overfitting prevention**

Before starting modeling, it is important to do some data processing. For example, standardize the values or do some feature processing, so that the model will be smoother during training and learn faster (Abbasi et al., 2025). We can first scale the original data, such as unifying the values into a similar range, or adding some new indicators, such as "soil fertility index". In addition, weather, soil and field management data can be combined and used together (Nossam et al., 2024). In order to prevent the model from "remembering too much", which is the

so-called "overfitting", some methods can be used. Commonly used methods include cross-validation, regularization, or optimization through parameter adjustment. Sometimes the amount of data is too small, we can also create some "synthetic data" to supplement it. In addition, using an integrated model or only selecting the features most relevant to yield for modeling can also effectively reduce the problem of overfitting. This can make the model more stable and more reliable (Manjunath and Palayyan, 2023; Razavi et al., 2024).

### 3.3 Model interpretability and reliability analysis

Nowadays, people pay more and more attention to whether the model can explain how it makes judgments. This is called "interpretability". Tools like SHAP and LIME can tell us what data the model uses to make predictions (Figure 1) (Nurcahyo et al., 2023; Paudel et al., 2023; Pant et al., 2025). For example, they can analyze whether weather, soil, or planting methods have the greatest impact on yield. In this way, farmers and researchers will be more willing to trust the model after seeing the results. In addition to these, we can also use some methods to test whether the model is reliable. For example, do sensitivity analysis to see if the model will be chaotic when different variables change. You can also evaluate the "uncertainty" of the prediction, that is, whether the model is confident when making predictions. In addition, using some new data to test the model can also help us determine how it performs in real scenarios (Hu et al., 2023).

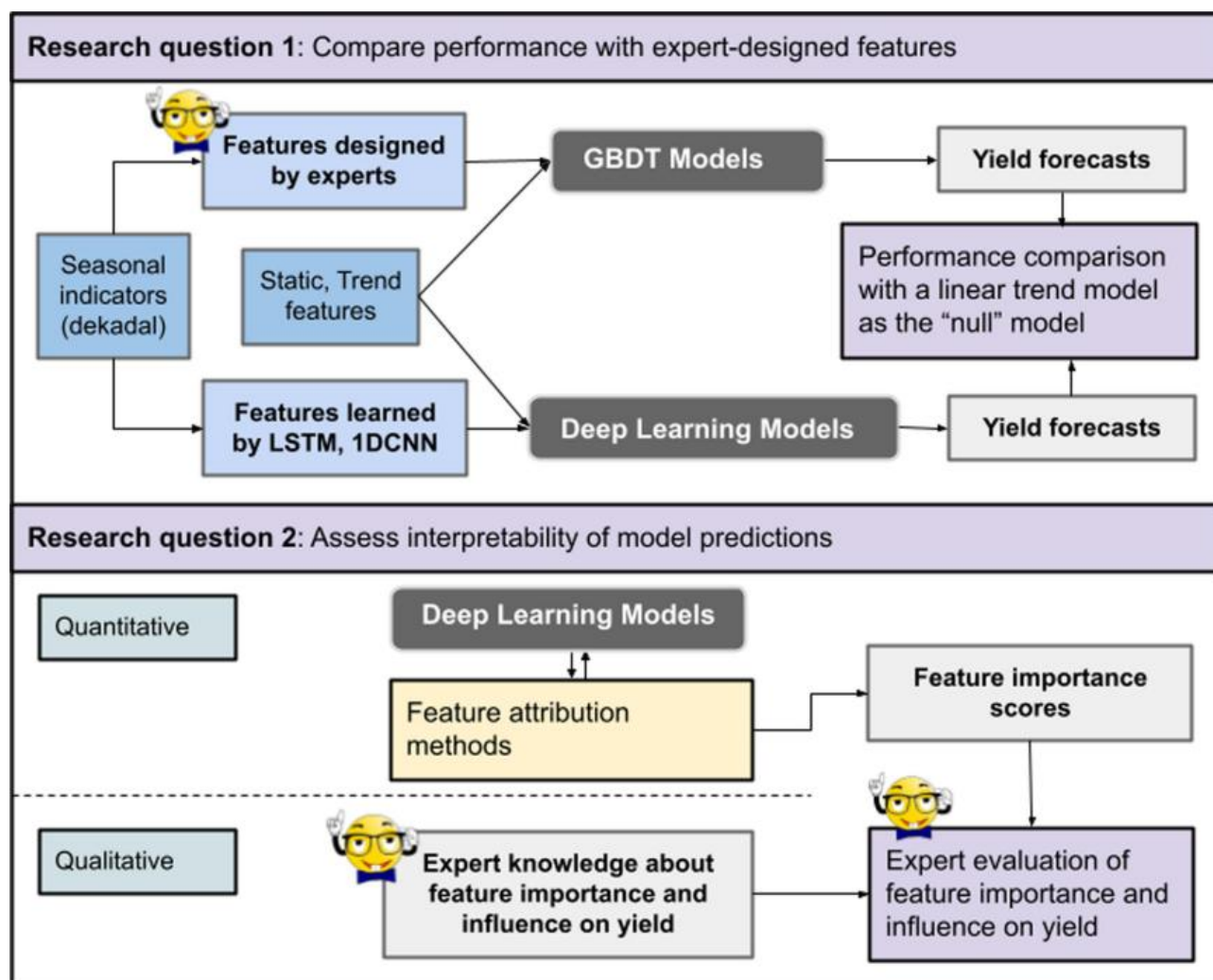


Figure 1 Framework to assess performance and interpretability of deep learning models (Adopted from Paudel et al., 2023)

## 4 Integration Strategies of Genomic Selection and Machine Learning

### 4.1 Fusion of genotype, phenotype, and environmental data

Under drought conditions, to more accurately predict corn yield, we cannot just look at one type of data. It will be more effective to analyze genotype, phenotype and environmental data together. Studies have found that if genetic

markers (such as SNPs), phenotypic measurement data, and processed environmental variables (such as climate, soil, etc.) are added to the model, the prediction will be more accurate. For example, not only using genetic data, but also adding markers related to traits and environmental information related to developmental stages, the prediction accuracy can be increased by 14% to 28% (He et al., 2025). It is critical to feature process the original environmental data before using it in the model, which can make the machine learning model understand the meaning of the data better (Fernandes et al., 2024).

#### **4.2 Architecture of GS+ML hybrid predictive models**

The GS+ML model combines traditional genomic selection methods (such as GBLUP, Bayes B) and modern machine learning methods (such as random forest, neural network, XGBoost). This combination can better handle the complex relationship between genes and environment. In terms of model structure, genes and environment can be combined in an "additive" way (G+E) or a "multiplicative" way (GEI). Additive models are fast to calculate and easy to use; machine learning methods such as tree models can automatically discover the relationship between genes and environment without us setting it in advance (Fernandes et al., 2024). Now we can also use automated machine learning platforms to integrate these models, which is more labor-saving and can quickly test multiple schemes (Saleh et al., 2023).

#### **4.3 Optimization of model pipelines for drought scenarios**

In order to more accurately predict corn yield under drought conditions, some optimization methods can be used. Multi-environment modeling is to train the model by putting data from different regions or different years together. In this way, data from other experiments can be used to fill in some missing parts, which helps to improve the prediction effect (Bhandari et al., 2018; Dias et al., 2018). Genetic markers and environmental variables that are closely related to drought resistance or environmental factors should be selected. This can reduce the interference of useless information in the model and allow the model to focus more on learning important parts (He et al., 2025). After the model is trained, it is necessary to do several rounds of verification and debug and optimize the parameters. This can prevent the model from "memorizing" the training data and not being able to use it under different conditions. Optimized models usually maintain relatively good results under different drought environments (Saleh et al., 2023; Fernandes et al., 2024). "Hybrid model" and "dimensionality reduction" techniques can also be used to reduce the pressure on the model during operation. Because when faced with a large amount of data, hybrid models can combine the advantages of multiple algorithms, and dimensionality reduction can simplify the number of variables and make the model run faster (Jighly et al., 2021). After optimization using these methods, the model is not only more accurate, but also can cope with various drought scenarios. This also helps us to more quickly select those corn varieties that are truly drought-resistant.

### **5 Model Evaluation and Cross-Environment Transferability**

#### **5.1 Cross-validation and external dataset testing**

Cross-validation is a common method to check whether a model is useful. K-fold cross-validation and leave-one-out-of-the-box (LOOCV) are two of them. Their approach is to repeatedly split the data into training sets and test sets, and then train and validate them in turn. This can help us see the predictive ability of the model and reduce the problem of the model "memorizing" the training data (Yates et al., 2022; Qiu, 2024). However, it is not enough to rely on these "internal data" for verification. Sometimes, the model may just remember the original data and it will not work in a different environment. Therefore, many researchers now pay more attention to "external validation", that is, testing the model with data from other places or under different conditions. This can show whether the model is easy to generalize, and can also find some problems that cannot be seen in internal testing, such as whether the model is overfitting or whether it is only applicable to a specific data distribution (Ho et al., 2020; Cabitza et al., 2021; Eertink et al., 2022; Riley et al., 2024).

#### **5.2 Evaluation across multiple drought scenarios**

The most direct way to know if a model is accurate in drought conditions is to test it in different drought environments. Droughts can be long or short, severe in some places or mild in others, and the climate conditions in some places are different. All of these will affect the performance of the model. We can test it in several ways.

For example, we can use data from past droughts or simulate some droughts that may be encountered in the future. This can give us a more comprehensive view of whether the model is reliable (Rahmati et al., 2020; Fooladi et al., 2021; Ahmad et al., 2024). In addition, we need to see if the model can detect changes in the time and location of droughts. This is also critical. Most importantly, it must be able to predict core indicators such as yields relatively accurately under these different drought conditions (Fooladi et al., 2021; Prodhon et al., 2021; Zhang and Xu, 2024).

### **5.3 Robustness and generalization in diverse environments**

Robustness means that the model can still maintain good results in different data, environments or populations. The best way to evaluate its generalization ability is to combine internal and external cross-validation, external data testing, and sensitivity analysis (Takada et al., 2021). Research has shown that if the data used for training and testing comes from different regions and backgrounds, such a model is more likely to adapt to the new environment and will not be "boxed" by a certain type of data. Therefore, it is very important to train with multiple types of data (Ho et al., 2020; Adkinson et al., 2024). A good evaluation method should not only look at whether the model is accurate, but also consider whether it is stable and has no deviations, and have clear evaluation criteria, such as the similarity and difference between data (Cabitza et al., 2021).

## **6 Case Studies**

### **6.1 Application in CIMMYT's drought-resilient maize breeding**

The International Maize and Wheat Improvement Center (CIMMYT) has done a lot of research on drought-resistant maize. They have used two methods, marker-assisted recurrent selection (MARS) and genomic selection (GS), to breed a number of drought-resistant maize varieties in sub-Saharan Africa (Figure 2). These new methods are more effective than traditional breeding methods and can select more stress-resistant varieties more quickly. To make breeding more efficient, CIMMYT combines QTL mapping (finding gene loci associated with important traits), high-throughput phenotyping, and some molecular tools. In this way, not only drought resistance is improved, but also nitrogen use efficiency and disease resistance are improved (Masuka et al., 2017; Prasanna, 2023). In the past 15 years, more than 300 climate-resistant maize varieties have been bred in sub-Saharan Africa and South Asia. Seeds of these varieties have been widely promoted, helping millions of small farmers (Semagn et al., 2015; Prasanna et al., 2021; Bm, 2022).

### **6.2 Model deployment in Chinese hybrid maize lines**

Corn is grown in many parts of China. Areas such as the northern plains and the Loess Plateau often encounter droughts, which affects corn yields. In order to solve this problem, some agricultural universities in China and local breeding units have cooperated in research. They used machine learning methods to build a prediction model and evaluated more than 300 corn hybrids. These varieties were tested under water and without water. The researchers used two models: support vector regression (SVR) and deep neural network (DNN). They analyzed the genetic data of SNPs and soil moisture conditions together to see which varieties were more drought-resistant. In the end, it was found that this method can more accurately select good varieties and provide a lot of useful information for breeding. In this way, those high-quality drought-resistant corn varieties can also be promoted to drought-resistant areas more quickly (Prasanna et al., 2021; Bm, 2022).

### **6.3 Regional application in sub-Saharan Africa**

In sub-Saharan Africa, CIMMYT and partner institutions have developed and tested hundreds of drought-tolerant maize varieties using GS, MARS and multi-environment testing methods. These varieties include hybrids and open-pollinated varieties. Studies have found that these newly developed varieties perform well under drought conditions, and have higher yields than old varieties both in controlled trials and in natural environments. In particular, the performance of new varieties is more obvious in some low-yield areas. With the joint efforts of governments, enterprises and seed systems, these drought-resistant maize varieties have been widely promoted, covering millions of hectares of land. This has greatly helped to improve the food security and risk resistance of small farmers (Worku et al., 2016; Manigben et al., 2024).



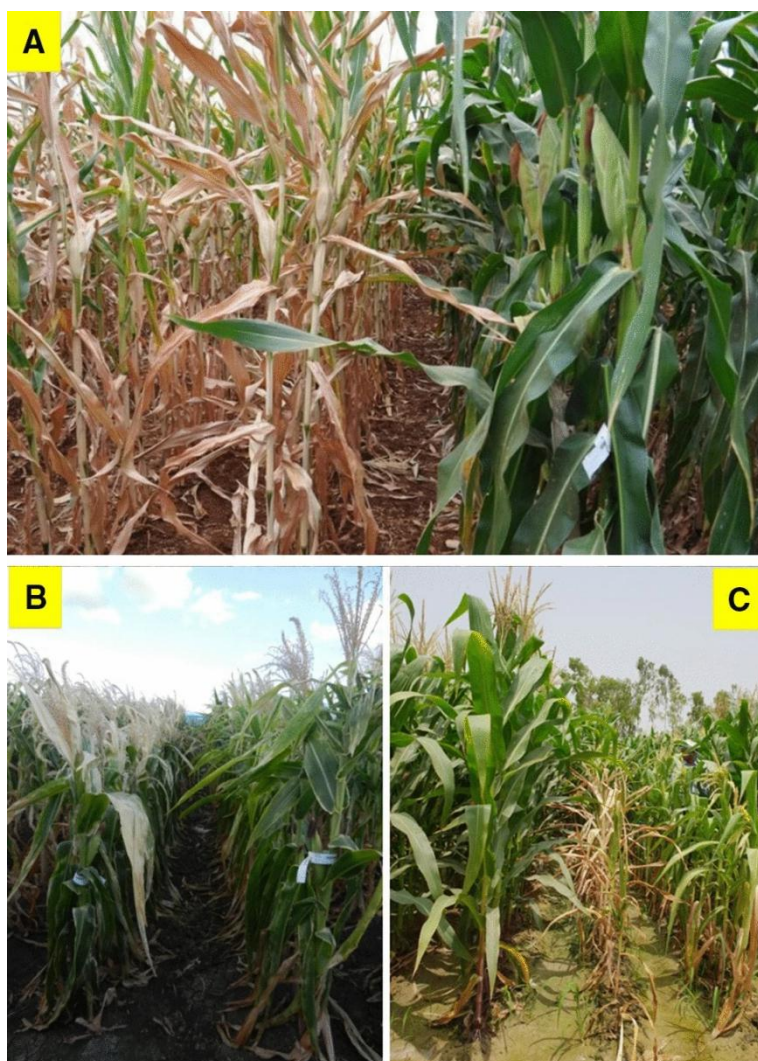


Figure 2 Phenotypic contrast of maize hybrids under managed drought stress (a), managed heat stress (b), and managed waterlogging stress (c) screening (Adopted from Prasanna et al., 2021)

## 7 Concluding Remarks

Combining genotyping (GS) and machine learning (ML), using high-throughput gene and trait data, plus environmental information, can more accurately predict complex traits such as yield and drought resistance. Some commonly used machine learning methods, such as random forests, support vector machines, and neural networks, can help us analyze complex relationships between genes and the environment, especially those that are not easily discovered by traditional statistical methods. The study also found that the combined framework built by these methods can effectively improve the prediction accuracy of corn and other crops, and is also helpful for judging stress resistance. In this way, breeding work can be done faster and more accurately. Integrating multiple omics data and adding some carefully selected key features can further improve the prediction effect.

GS plus ML is a more flexible approach, especially for dealing with climate change issues such as drought. It can help shorten breeding time and find climate-suitable genotypes more quickly, so that new high-yield and stress-resistant varieties can be cultivated more quickly. Nowadays, many artificial intelligence tools, high-throughput trait analysis technologies, and automated data processing methods have become more and more common. These technologies are gradually driving agriculture to become smarter, and more and more data is used. In this way, precision breeding is not only simpler, but also easier to promote to large-scale planting. However, these new technologies are not without problems. For example, how to combine different data, whether there are enough computing resources, and how to ensure that small farmers can afford and use them are all things we have to consider.

Next, research may focus more on how to make AI and machine learning algorithms more stable. People also hope that they can make it easier to explain the results and more convenient to use in different regions. At the same time, there must be unified standards for how to collect data, and it is best to have a better digital platform. In this way, breeding, data, and technical teams can communicate and cooperate more conveniently. There are also some issues that cannot be ignored, such as how to protect data privacy and how to make policies fair so that these new technologies can be promoted more reasonably and responsibly. In the future, as the use of AI in breeding becomes more and more mature, it may become a very useful tool to help us select new crops that are more drought-resistant and adaptable to climate change. In this way, it will also be of great help to ensure food security and promote sustainable agricultural development.

### Acknowledgments

We thank Mr J. Wu from the Institute of Life Science of Jiyang College of Zhejiang A&F University for his reading and revising suggestion.

### 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

- Abbasi M., Váz P., Silva J., and Martins P., 2025, Machine learning approaches for predicting maize biomass yield: leveraging feature engineering and comprehensive data integration, *Sustainability*, 17(1): 256.  
<https://doi.org/10.3390/su17010256>
- Adkinson B., Rosenblatt M., Dadashkarimi J., Tejaviyulya L., Jiang R., Noble S., and Scheinost D., 2024, Brain-phenotype predictions of language and executive function can survive across diverse real-world data: dataset shifts in developmental populations, *Developmental Cognitive Neuroscience*, 70: 101464.  
<https://doi.org/10.1016/j.dcn.2024.101464>
- Ahmad S., Batool A., and Ali Z., 2024, Spatial predictive analysis of drought duration in relation to climate change using interpolation techniques, *Stochastic Environmental Research and Risk Assessment*, 39: 639-656.  
<https://doi.org/10.1007/s00477-024-02886-x>
- Amadu M., Beyene Y., Chaikam V., Tongoon P., Danquah E., Ifie B., Burgueño J., Prasanna B., and Gowda M., 2025, Genome-wide association mapping and genomic prediction analyses reveal the genetic architecture of grain yield and agronomic traits under drought and optimum conditions in maize, *BMC Plant Biology*, 25: 135.  
<https://doi.org/10.1186/s12870-025-06135-3>
- Azrai M., Aqil M., Andayani N., Efendi R., Suarni, Suwardi, Jihad M., Zainuddin B., Salim, Bahtiar, Muliadi A., Yasin M., Hannan M., Rahman, and Syam A., 2024, Optimizing ensembles machine learning, genetic algorithms, and multivariate modeling for enhanced prediction of maize yield and stress tolerance index, *Frontiers in Sustainable Food Systems*, 8: 1334421.  
<https://doi.org/10.3389/fsufs.2024.1334421>
- Bhandari A., Bartholomé J., Cao T., Kumari N., Frouin J., Kumar A., and Ahmadi N., 2018, Selection of trait-specific markers and multi-environment models improve genomic predictive ability in rice, *PLoS ONE*, 14(5): e0208871.  
<https://doi.org/10.1371/journal.pone.0208871>
- Bm P., 2022, Breeding and deploying multiple stress-tolerant maize varieties in the tropics, *Journal of Rice Research*, 15: 59-63.  
<https://doi.org/10.58297/ojpn7450>
- Cabrita F., Campagner A., Soares F., Guadiana-Romualdo L., Challa F., Sulejmani A., Seghezzi M., and Carobene A., 2021, The importance of being external. methodological insights for the external validation of machine learning models in medicine, *Computer Methods and Programs in Biomedicine*, 208: 106288.  
<https://doi.org/10.1016/j.cmpb.2021.106288>
- Chen Q., Ying Q.H., Lei K.Z., Zhang J.M., and Liu H.Z., 2024, The integration of genetic markers in maize breeding programs, *Bioscience Methods*, 15(5): 226-236.  
<https://doi.org/10.5376/bm.2024.15.0023>
- Dhaliwal J., Panday D., Saha D., Lee J., Jagadamma S., Schaeffer S., and Mengistu A., 2022, Predicting and interpreting cotton yield and its determinants under long-term conservation management practices using machine learning, *Computers and Electronics in Agriculture*, 199: 107107.  
<https://doi.org/10.1016/j.compag.2022.107107>
- Dias K., Gezan S., Guimarães C., Nazarian A., Da Costa E Silva L., Parentoni S., De Oliveira Guimarães P., De Oliveira Anoni C., Pádua J., De Oliveira Pinto M., Noda R., Ribeiro C., De Magalhães J., Garcia A., De Souza J., Guimarães L., and Pastina M., 2018, Improving accuracies of genomic predictions for drought tolerance in maize by joint modeling of additive and dominance effects in multi-environment trials, *Heredity*, 121: 24-37.  
<https://doi.org/10.1038/s41437-018-0053-6>

- Eertink J., Heymans M., Zwezerijnen G., Zijlstra J., De Vet H., and Boellaard R., 2022, External validation: a simulation study to compare cross-validation versus holdout or external testing to assess the performance of clinical prediction models using PET data from DLBCL patients, *EJNMMI Research*, 12: 58.  
<https://doi.org/10.1186/s13550-022-00931-w>
- Fernandes I., Vieira C., Dias K., and Fernandes S., 2024, Using machine learning to combine genetic and environmental data for maize grain yield predictions across multi-environment trials, *Theoretical and Applied Genetics*, 137: 189.  
<https://doi.org/10.1007/s00122-024-04687-w>
- Fooladi M., Golmohammadi M., Safavi H., and Singh V., 2021, Fusion-based framework for meteorological drought modeling using remotely sensed datasets under climate change scenarios: resilience, vulnerability, and frequency analysis, *Journal of Environmental Management*, 297: 113283.  
<https://doi.org/10.1016/j.jenvman.2021.113283>
- Gharakhanlou N., and Perez L., 2024, From data to harvest: leveraging ensemble machine learning for enhanced crop yield predictions across Canada amidst climate change, *The Science of the Total Environment*, 951: 175764.  
<https://doi.org/10.1016/j.scitotenv.2024.175764>
- Hao Z., Li X., Xie C., Weng J., Li M., Zhang D., Liang X., Liu L., Liu S., and Zhang S., 2011, Identification of functional genetic variations underlying drought tolerance in maize using SNP markers, *Journal of Integrative Plant Biology*, 53(8): 641-652.  
<https://doi.org/10.1111/j.1744-7909.2011.01051.x>
- 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/advs.202412423>
- Ho S., Phua K., Wong L., and Goh W., 2020, Extensions of the external validation for checking learned model interpretability and generalizability, *Patterns*, 1(8): 100129.  
<https://doi.org/10.1016/j.patter.2020.100129>
- Hu T., Zhang X., Bohrer G., Liu Y., Zhou Y., Martin J., Li Y., and Zhao K., 2023, Crop yield prediction via explainable AI and interpretable machine learning: Dangers of black box models for evaluating climate change impacts on crop yield, *Agricultural and Forest Meteorology*, 109458: 447-469.  
<https://doi.org/10.1016/j.agrformet.2023.109458>
- Jighly A., Hayden M., and Daetwyler H., 2021, Integrating genomic selection with a genotype plus genotype x environment (GGE) model improves prediction accuracy and computational efficiency, *Plant, Cell & Environment*, 336: 109458.  
<https://doi.org/10.1111/pce.14145>
- Li C., Sun B., Li Y., Liu C., Wu X., Zhang D., Shi Y., Song Y., Buckler E., Zhang Z., Wang T., and Li Y., 2016, Numerous genetic loci identified for drought tolerance in the maize nested association mapping populations, *BMC Genomics*, 17: 894.  
<https://doi.org/10.1186/s12864-016-3170-8>
- Li R., Wang Y., Li D., Guo Y., Zhou Z., Zhang M., Zhang Y., Würschum T., and Liu W., 2024, Meta-Quantitative trait loci analysis and candidate gene mining for drought tolerance-associated traits in maize (*Zea mays* L.), *International Journal of Molecular Sciences*, 25(8): 4295.  
<https://doi.org/10.3390/ijms25084295>
- Ma L., Niu W., Li G., Du Y., Sun J., and Siddique K., 2024, Crop Yield prediction based on bacterial biomarkers and machine learning, *Journal of Soil Science and Plant Nutrition*, 24: 2798-2814  
<https://doi.org/10.1007/s42729-024-01705-0>
- Malphedwar L., Adsul A., Nagare S., Nimse Y., Nimble S., and Pakhle S., 2024, Crop yield prediction using machine learning, *International Journal of Advanced Research in Science, Communication and Technology*, 4(2): 395-398.  
<https://doi.org/10.48175/ijarsct-22172>
- Manigben K., Beyene Y., Chaikam V., Tongoono P., Danquah E., Ifie B., Aleri I., Chavangi A., Prasanna B., and Gowda M., 2024, Testcross performance and combining ability of intermediate maturing drought tolerant maize inbred lines in Sub-Saharan Africa, *Frontiers in Plant Science*, 15: 1471041.  
<https://doi.org/10.3389/fpls.2024.1471041>
- Manjunath M., and Palayyan B., 2023, An efficient crop yield prediction framework using hybrid machine learning model, *Revue d'Intelligence Artificielle*, 370428: 1057-1067.  
<https://doi.org/10.18280/ria.370428>
- Marino R., Ponnaiah M., Krajewski P., Fropa C., Gianfranceschi L., Pè M., and Sari-Gorla M., 2009, Addressing drought tolerance in maize by transcriptional profiling and mapping, *Molecular Genetics and Genomics*, 281: 163-179.  
<https://doi.org/10.1007/s00438-008-0401-y>
- Masuka B., Atlin G., Olsen M., Magorokosho C., Labuschagne M., Crossa J., Bänziger M., Pixley K., Vivek B., Biljon A., Macrobert J., Alvarado G., Prasanna B., Makumbi D., Tarekge A., Das B., Zaman-Allah M., and Cairns J., 2017, Gains in maize genetic improvement in Eastern and Southern Africa: I. CIMMYT hybrid breeding pipeline, *Crop Science*, 57: 168-179.  
<https://doi.org/10.2135/CROPSCI2016.05.0343>
- Nossam S., Katakam R., Pulastya G., and Venugopalan M., 2024, Enhanced crop yield prediction using machine learning techniques, 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), 10724901: 1-6.  
<https://doi.org/10.1109/ICCCNT61001.2024.10724901>
- Nurcahyo A., Heryadi Y., Lukas, Suparta W., and Sonata I., 2023, Interpretable machine learning for multi-class crop yield prediction, 2023 3rd International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA), 10428914: 194-200.  
<https://doi.org/10.1109/ICICyTA60173.2023.10428914>



- Oikonomidis A., Catal C., and Kassahun A., 2022, Hybrid deep learning-based models for crop yield prediction, *Applied Artificial Intelligence*, 36(1): 2031822.  
<https://doi.org/10.1080/08839514.2022.2031823>
- Pant H., Joshi G., Rawat B., Goyal H., Joshi Y., and Bohra C., 2025, Comparative study of crop yield prediction using explainable AI and interpretable machine learning techniques, 2025 Fifth International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), 10958878: 1-7.  
<https://doi.org/10.1109/ICAECT63952.2025.10958878>
- Paudel D., De Wit A., Boogaard H., Marcos D., Osinga S., and Athanasiadis I., 2023, Interpretability of deep learning models for crop yield forecasting, *Computers and Electronics in Agriculture*, 206: 107663.  
<https://doi.org/10.1016/j.compag.2023.107663>
- Prasanna B., 2023, Breeding and deploying climate resilient maize varieties in the tropics, *Indian Journal of Ecology*, 6: 1895-1899.  
<https://doi.org/10.55362/ije/2023/4153>
- 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, *Theoretical and Applied Genetics*, 134: 1729-1752.  
<https://doi.org/10.1007/s00122-021-03773-7>
- Prodhan F., Zhang J., Yao F., Shi L., Sharma T., Zhang D., Cao D., Zheng M., Ahmed N., and Mohana H., 2021, Deep learning for monitoring agricultural drought in South Asia using remote sensing data, *Remote Sensing*, 13(9): 1715.  
<https://doi.org/10.3390/rs13091715>
- Qiu J., 2024, An analysis of model evaluation with cross-validation: techniques, applications, and recent advances, *Advances in Economics, Management and Political Sciences*, 99: 69-72.  
<https://doi.org/10.54254/2754-1169/99/2024ox0213>
- Rahmati O., Falah F., Dayal K., Deo R., Mohammadi F., Biggs T., Moghaddam D., Naghibi S., and Bui D., 2020, Machine learning approaches for spatial modeling of agricultural droughts in the south-east region of Queensland Australia, *The Science of the Total Environment*, 699: 134230.  
<https://doi.org/10.1016/j.scitotenv.2019.134230>
- Razavi M., Nejadhashemi A., Majidi B., Razavi H., Kpodo J., Eeswaran R., Ciampitti I., and Prasad P., 2024, Enhancing crop yield prediction in Senegal using advanced machine learning techniques and synthetic data, *Artificial Intelligence in Agriculture*, 14: 99-114.  
<https://doi.org/10.1016/j.aiia.2024.11.005>
- Riley R., Archer L., Snell K., Ensor J., Dhiman P., Martin G., Bonnett L., and Collins G., 2024, Evaluation of clinical prediction models (part 2): how to undertake an external validation study, *The BMJ*, 384: e074820.  
<https://doi.org/10.1136/bmj-2023-074820>
- Saleh A., Moniruzzaman M., Islam S., Ahmed K., Rahaman M., Hossain S., and Manik T., 2023, Integrating genomic selection and machine learning: a data-driven approach to enhance corn yield resilience under climate change, *Journal of Environmental and Agricultural Studies*, 4(2): 20-27.  
<https://doi.org/10.32996/jeas.2023.4.2.6>
- Semagn K., Beyene Y., Babu R., Nair S., Gowda M., Das B., Tarekegne A., Mugo S., Mahuku G., Worku M., Warburton M., Olsen M., and Prasanna B., 2015, Quantitative trait loci mapping and molecular breeding for developing stress resilient maize for Sub-Saharan Africa, *Crop Science*, 55: 1449-1459.  
<https://doi.org/10.2135/CROPSCI2014.09.0646>
- Shawon S., Ema F., Mahi A., and Raihan M., 2023, Crop yield prediction: robust machine learning approaches for precision agriculture, 2023 26th International Conference on Computer and Information Technology (ICCIT), 10441634: 1-6.  
<https://doi.org/10.1109/ICCIT60459.2023.10441634>
- Shikha M., Kanika A., Rao A., Mallikarjuna M., Gupta H., and Nepolean T., 2017, Genomic selection for drought tolerance using genome-wide SNPs in maize, *Frontiers in Plant Science*, 8: 550.  
<https://doi.org/10.3389/fpls.2017.00550>
- Takada T., Nijman S., Denaxas S., Snell K., Uijl A., Nguyen T., Asselbergs F., and Debray T., 2021, Internal-external cross-validation helped to evaluate the generalizability of prediction models in large clustered datasets, *Journal of Clinical Epidemiology*, 137: 83-91.  
<https://doi.org/10.1016/j.jclinepi.2021.03.025>
- Van Klompenburg T., Kassahun A., and Catal C., 2020, Crop yield prediction using machine learning: a systematic literature review, *Computers and Electronics in Agriculture*, 177: 105709.  
<https://doi.org/10.1016/j.compag.2020.105709>
- Wang N., Liu B., Liang X., Zhou Y., Song J., Yang J., Yong H., Weng J., Zhang D., Li M., Nair S., Vicente F., Hao Z., Zhang X., and Li X., 2019, Genome-wide association study and genomic prediction analyses of drought stress tolerance in China in a collection of off-PVP maize inbred lines, *Molecular Breeding*, 39: 113.  
<https://doi.org/10.1007/s11032-019-1013-4>
- Worku M., Makumbi D., Beyene Y., Das B., Mugo S., Pixley K., Bänziger M., Owino F., Olsen M., Asea G., and Prasanna B., 2016, Grain yield performance and flowering synchrony of CIMMYT's tropical maize (*Zea mays* L.) parental inbred lines and single crosses, *Euphytica*, 211: 395-409.  
<https://doi.org/10.1007/s10681-016-1758-3>
- 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 X., Feng H., Wu D., Yan S., Zhang P., Wang W., Zhang J., Ye J., Dai G., Fan Y., Li W., Song B., Geng Z., Yang W., Chen G., Qin F., Terzaghi W., Stitzer M., Li L., Xiong L., Yan J., Buckler E., Yang W., and Dai M., 2021, Using high-throughput multiple optical phenotyping to decipher the genetic architecture of maize drought tolerance, *Genome Biology*, 22: 185.  
<https://doi.org/10.1186/s13059-021-02377-0>
- Yates L., Aandahl Z., Richards S., and Brook B., 2022, Cross validation for model selection: a review with examples from ecology, *Ecological Monographs*, 93(1): e1557.  
<https://doi.org/10.1002/ecm.1557>
- Yuan Y., Cairns J., Babu R., Gowda M., Makumbi D., Magorokosho C., Zhang A., Liu Y., Wang N., Hao Z., Vicente S., Olsen M., Prasanna B., Lu Y., and Zhang X., 2019, Genome-wide association mapping and genomic prediction analyses reveal the genetic architecture of grain yield and flowering time under drought and heat stress conditions in maize, *Frontiers in Plant Science*, 9: 1919.  
<https://doi.org/10.3389/fpls.2018.01919>
- Zhang X., and Xu M.L., 2024, Adaptation of maize to various climatic conditions: genetic underpinnings, *Bioscience Evidence*, 14(3): 122-130.  
<https://doi.org/10.5376/be.2024.14.0014>
- Zhang A., Chen S., Cui Z., Liu Y., Guan Y., Yang S., Qu J., Nie J., Dang D., Li C., Dong X., Fan J., Zhu Y., Zhang X., Crossa J., Cao H., Ruan Y., and Zheng H., 2022, Genomic prediction of drought tolerance during seedling stage in maize using low-cost molecular markers, *Euphytica*, 218: 154.  
<https://doi.org/10.1007/s10681-022-03103-y>
- Zhang N., Liu B., Fan Y., Chang J., Zhou Y., Wang Y., Zhang W., Zhang X., Shutu X., and Xue J., 2023, Molecular mechanisms of drought resistance using genome-wide association mapping in maize (*Zea mays* L.), *BMC Plant Biology*, 23: 468.  
<https://doi.org/10.1186/s12870-023-04489-0>



#### 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.

---