**SINTECHCOM: Science, Technology and Communication Journal**
Vol. 6, No. 3, June 2026, pp. 173-188, DOI: 10.59190/stc.v6i3.368

# Systematic literature review using deep learning in plant genomic prediction

**Rizki Darmawan**
Department of Computer Science, Universitas Dian Nuswantoro, Semarang 50131, Indonesia

## ABSTRACT

Genomic prediction (GP) is a critical approach in plant breeding used to forecast agronomic traits based on genetic marker data. In recent years, machine learning and deep learning methods have been increasingly applied in genomic prediction to address the limitations of traditional methods such as genomic best linear unbiased prediction (GBLUP). This systematic literature review aims to evaluate research trends over the past decade (2015–2025) regarding the application of deep learning in genomic prediction for crops, encompassing publication trends, publication sources, dataset types, key research topics, methods/algorithms used, implementation schemes, and commonly used evaluation metrics. The study follows the PRISMA protocol, with literature searches conducted on the primary databases Scopus and SpringerLink. From thousands of identified articles, 43 studies meeting the inclusion criteria were selected. The data revealed a significant increase in publications since 2018, peaking in 2024. The majority of articles were published in open-access journals, with notable contributions in Frontiers in Plant Science, Scientific Reports, and G3. The research covers various crop types, including wheat, maize, rice, soybean, sugarcane, and others, with diverse deep learning schemes such as convolutional neural networks (CNN), Ensemble Stacking, and Multi-Task Learning, often compared to conventional methods. The synthesis indicates that deep learning frequently enhances the accuracy of genomic prediction, particularly for complex multi-environment or multi-trait data. However, in some cases, classical linear/Bayesian models remain competitive. The application of deep learning in plant genomic prediction is rapidly advancing and demonstrates significant potential for improving crop selection performance.

## ARTICLE INFO

**Article history:**

Received May 11, 2026
Revised May 20, 2026
Accepted May 21, 2026

**Keywords:**

Crops
Deep Learning
Genomic Prediction
Plant Breeding
Systematic Literature

**Corresponding Author**
E-mail address: P41202400082@mhs.dinus.ac.id

## 1. INTRODUCTION

World's population continuously growing and increasingly complex environmental challenges, the development of superior crop varieties has become an urgent priority [1]. In this context, genomic prediction (GP) has emerged as a revolutionary approach in modern plant breeding. GP enables faster and more accurate selection of superior genotypes by predicting an individual's genetic value based on their genomic data, even before field phenotypes are available. Traditional GP methods, such as Genomic Best Linear Unbiased Prediction (GBLUP) and Ridge Regression Best Linear Unbiased Prediction (RR-BLUP) [2], have been the gold standard for years. However, these methods have inherent limitations, particularly in handling complex traits governed by multiple genes with non-additive interactions and in managing the vast volume of genomic data. These limitations often hinder the ability to fully capture the intricate genetic architecture of key agronomic traits.

Deep learning (DL), a subfield of machine learning, has demonstrated remarkable capabilities in identifying complex patterns from large and diverse datasets. DL architectures, such as Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), Deep Neural Networks (DNN), and Transformers, are designed to process data through multiple layers of non-linear processing [3]. This capability allows DL to automatically learn hierarchical feature representations from genomic and phenotypic data, potentially overcoming the limitations of traditional GP methods. By capturing non-linear relationships and complex interactions between genotypes and phenotypes, DL models offer a new avenue for improving prediction accuracy, particularly for traits that are difficult to predict using linear approaches. These innovations open opportunities to accelerate genetic progress in plant breeding programs.

Genomic prediction (GP) was first utilized to leverage genotype markers across the genome to estimate breeding values for unobserved populations, enabling rapid identification of superior genotypes and accelerating the breeding process [4]. Extensive GP research has focused on optimizing marker density, training population size, family relationships, and model selection. Genomic Best Linear Unbiased Prediction (GBLUP) is a widely used GP model that relies on marker-based relationship matrices for predictions [5]. In contrast, Bayesian models incorporate prior distributions, requiring different models for distinct traits. For example, Bayes B employs a Gaussian mixture, assuming not all markers contribute to genetic variance. Bayesian lasso (BL) applies a double exponential prior for continuous shrinkage and variable selection, using a heavy-tailed Student-t distribution for marker effects [6]. However, the precise influence of individual SNPs remains elusive and may not align with specific distributions. Additionally, these parametric models often fail to capture complex SNP interactions, particularly for complex traits driven by epistasis.

The Deep Learning Genome-Wide Association Study (DLGWAS) is a dual-stream deep learning model [4]. It has demonstrated superior predictive performance compared to traditional statistical methods on simulated and soybean datasets. The DNNGP model integrates three convolutional neural network (CNN) layers, batch normalization (BN) to prevent overfitting, and two dropout layers [7]. This model efficiently processes complex omics data, outperforming commonly used GP methods such as GBLUP, LightGBM, SVR, DeepGS, and DLGWAS. SoyDNGP is a deep network consisting of 12 convolutional blocks and fully connected layers [8]. It incorporates a coordinate attention (CA) mechanism after the first and last convolutional layers to enhance spatial information extraction. SoyDNGP surpasses methods like AdaBoost, Decision Tree, Naive Bayes, and Random Forest in classification tasks and outperforms DeepGS and DNNGP in regression tasks, highlighting its versatility and strength in GP [9].

The main goals are to analyze publication trends in genomic prediction studies using deep learning, identify the primary publication sources (journals and conferences) publishing papers on this topic, Describe the datasets used in the research, including crop types and data availability, identify frequently discussed research topics and their relationship to genomic prediction using deep learning, analyze how deep learning methods, models, or algorithms are applied to optimize genomic prediction describe the schemes, flowcharts, or pseudocode implemented in genomic prediction methods, identify the evaluation metrics used to measure research outcomes.The expected contributions of this review include providing a comprehensive overview of DL applications in plant GP, identifying prevalent datasets and methodologies, analyzing performance metrics, and formulating future research directions. Thus, this review will serve as a valuable reference for researchers and practitioners in plant breeding and bioinformatics.

## 2. MATERIAL AND METHODS

### 2.1. Review Method

This section describes in detail the methodology used to conduct this systematic literature review, ensuring transparency, reproducibility, and adherence to established scientific standards. The systematic review methodology adopted in this study primarily follows the guidelines proposed by Kitchenham [10]. This approach emphasizes a structured and iterative process for planning, conducting, and reporting review results. By adhering to an established framework, this review aims to minimize bias and ensure comprehensive and relevant coverage of the literature.

### 2.2. Research Question (RQs)

The research questions (RQs) guiding this systematic literature review were formulated to systematically explore the application of deep learning in genomic prediction for crops. These RQs serve as an intellectual roadmap for the entire review process, directing data extraction and analysis of findings. Table 1 presents the complete list of formulated research questions.

Table 1. The PICOC definition.

| Population | Genomic datasets, plat species, plant varietes |
|---|---|
| Intervention | Deep Learning (CNN, RNN, DNNGP, etc.), Machine Learning (SVM, RF) |
| Comparison | Traditional methods (e.g., GBLUP, RR-BLUP) vs AI models |
| Outcomes | Prediction accuracy, yield prediction, disease resistance prediction |
| Context | Agriculture, Plant Breeding, Genomic Selection, Genomic Prediction |

This systematic review is guided by a set of precisely formulated research questions, ensuring a focused and comprehensive investigation of the topic. These questions serve as a framework for data extraction and synthesis.

RQ1: What are the published trends in genomic prediction studies using deep learning?
RQ2: What journal publications sources publish articles on genomic prediction using deep learning?
RQ3: Which datasets are used in the reviewed studies?
RQ4: What research topics are frequently discussed and how do they relate to genomic prediction using deep learning?
RQ5: How are genomic prediction methods, models, or algorithms applied to optimize deep learning in the proposed methods?
RQ6: How are genomic prediction schemes, flowcharts, or pseudocodes implemented in the proposed methods?
RQ7: What evaluation metrics have been employed to measure outcomes?

The formulation of research questions (RQ) follows the Population, Intervention, Comparison, Outcome, and Context (PICOC) criteria. Subsequently, the RQs were tailored to suit the research objectives within this systematic review. Specifically, RQ1 and RQ2 are formulated to obtain publication sources and trends; RQ3 to identify datasets used; RQ4 to explore frequently discussed research topics; RQ5 and RQ6 to determine types of algorithms and methods employed and developed in using deep learning for genomic prediction; and RQ7 to review the types of evaluation metrics used in assessing the performance of proposed methods. You can see the mapping in Table 2.

Table 2. Data extraction properties mapped to RQs.

| Property | Research question (RQ) |
|---|---|
| Publication Sources and Trends | RQ1, RQ2 |
| Datasets | RQ3 |
| Research topics | RQ4 |
| Methods, models, algorithms | RQ5 |
| Flowchart schematics | RQ6 |
| Evaluation metrics | RQ7 |

### 2.3. Search and Selection Strategy

The literature search was conducted across two major digital libraries to ensure broad coverage of scientific publications: SpringerLink and Scopus. A comprehensive search query was employed to identify the most relevant articles effectively. This query was designed to integrate key terms related to genomic prediction, deep learning, and plant species, thus ensuring focused search results at the intersection of these fields. The search query used was: ("genomic prediction" OR "genomic selection") AND ("deep learning" OR "CNN" OR "Convolutional Neural Network" OR "RNN" OR "Recurrent Neural Network" OR "LSTM" OR "GRU" OR "Transformer" OR "ANN" OR "DNN") AND ("crop" OR "crops" OR "plant" OR "plants" OR "agriculture" OR "maize" OR "rice" OR "wheat").

### 2.4. Data Extraction

The article selection process was governed by strict inclusion and exclusion criteria. Articles were considered eligible if published between 2015 and May 26 2025, written in English, classified as peer-reviewed research articles, and available as open access. Conversely, articles were excluded if they were non-peer-reviewed papers, other review articles, or studies not focused on plants (such as human or animal studies). These criteria ensured that this review concentrated on relevant and current empirical research within the defined context.

The article selection process followed the systematic Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) workflow, involving multiple stages to ensure transparent identification and inclusion of relevant primary studies. Initially, a search in Scopus yielded 3,749 articles, and SpringerLink contributed 377 articles. After removing duplicates and applying initial inclusion criteria (article type, publication year 2015-2025, English language, open access), the number of articles considered decreased to 1,294 from Scopus and 81 from SpringerLink. Further exclusion, targeting review articles or non-plant focused studies (e.g., animals or humans), eliminated 2,455 articles from Scopus and 296 from SpringerLink.

Subsequent screening based on titles and abstracts was conducted, followed by a full-text eligibility assessment of the screened articles. From this process, 26 articles from Scopus and 17 from SpringerLink were deemed eligible based on abstract review. Ultimately, after careful full text evaluation to confirm full relevance and methodological compliance, a total of 43 primary articles were included in the final systematic literature review.

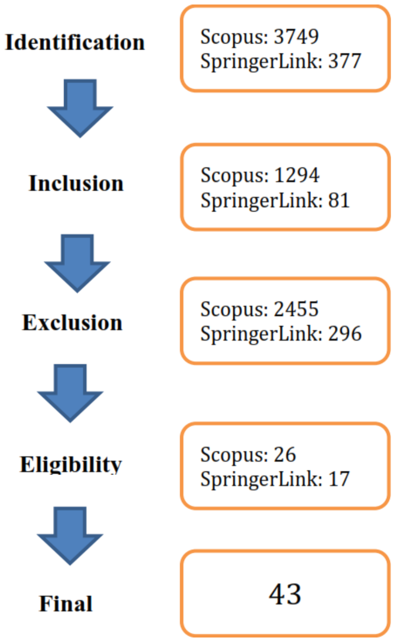

Figure 1. SLR Article Selection Flow (PRISMA Style Summary).

The significant reduction from an initially vast pool to just 43 final articles highlights the specialized nature of this research area. This considerable narrowing is not due to overly broad search queries or inefficient screening but instead underscores the highly specialized and interdisciplinary nature of research at the intersection of deep learning and genomic prediction specifically in plants. Numerous publications may exist on general genomic prediction (e.g., in livestock or humans, or using traditional statistical/machine learning methods) or deep learning in other domains. The stringent criteria applied in this review effectively isolated the highly specific niche targeted by this SLR. Consequently, the 43 identified papers represent a concentrated knowledge base, making the findings of this SLR extremely valuable for researchers targeting this precise intersection. It also emphasizes the critical importance of careful search and screening interdisciplinary reviews to avoid an overload of irrelevant literature. Figure 1 summarizes the article selection flow.

### 2.5. Study Quality Assessment and Data Synthesis

The assessment of the quality of selected primary studies was conducted to guide the synthesis of interpretations and to determine the strength of conclusions drawn from the findings. This involved evaluating methodological rigor and empirical evidence within each study. The primary goal of data synthesis was to compile evidence from selected primary studies to address the formulated RQs. Data and interpretations of findings from the selected primary studies were systematically collected and tabulated. These tabulated data were subsequently processed using Scopus: 3749 SpringerLink: 377 Scopus: 1294 SpringerLink: 81 Scopus: 2455 SpringerLink: 296 Scopus: 26 SpringerLink: 17 43 Identification Inclusion Exclusion Eligibility Final visualization tools to present the information through bar charts, pie charts, and tables, aiding in identifying significant patterns and trends.

### 2.6. Threats to Validity

Despite careful efforts to ensure rigor, every systematic literature review has potential threats to its validity. One possible threat is the potential omission of some papers due to limitations in the search queries or databases used. Although search queries were designed comprehensively, there is no guarantee that all relevant publications have been identified. Furthermore, publication bias, where studies with positive results are more likely to be published, may affect the literature's representativeness. Subjective interpretation during data extraction and synthesis is also a crucial consideration. Despite clear guidelines and structured processes, there remains a possibility of variability in interpretation among reviewers.

## 3. RESULT AND ANALYSIS

This core section presents the findings obtained from the systematic review, structured around the established research questions. It includes detailed analysis, interpretation, and synthesis of the extracted data, along with deeper implications.

### 3.1. Publication Trends (RQ1)

Analysis of the temporal distribution of publications on deep learning in plant genomic prediction reveals a very interesting trend. Among the 43 primary studies included in this review, there is a clear increase in the number of publications over the years within the period of 2015-2025. It can be seen in Table 3.

Tabel 1. Publication distribution by year.

| Years | Number of articles |
|---|---|
| 2018 | 3 |
| 2019 | 3 |
| 2020 | 2 |
| 2021 | 4 |
| 2022 | 7 |
| 2023 | 5 |
| 2024 | 12 |
| 2025 | 7 |

Annual mapping shows a steady increase over the past decade, with a clear acceleration starting in 2019 and a sharp spike in 2024 (12 articles), followed by sustained momentum through 2025 (7 articles). This pattern suggests that deep learning (DL) in genomic prediction has gone beyond proof-of-concept: it is rapidly gaining acceptance and application within the plant breeding and genomics communities. This acceleration is likely driven by three factors:\

1. The increasing availability of large-scale genomic and phenotypic datasets capable of supporting complex DL models.
2. Increased access to computing GPUs and cloud platforms which has lowered training costs; and
3. The methodological maturity of DL itself, including more efficient architectures and improved training/calibration techniques. Together, these conditions have shifted the field from exploratory demonstrations to operationally relevant studies (e.g., multi-site or multi-season evaluations).

At the same time, published results are still concentrated on major crops (corn, wheat, soybean), and only a small proportion of studies explicitly model multi-environment or multitrait scenarios. Evaluation standards and data partitioning also vary across papers, limiting direct comparisons. Overall, this trend is dynamic and promising for future research and practice, but it underscores the need for open benchmarking, clearer reporting (dataset characteristics, partitioning, leakage checks), and more consistent metrics to ensure comparable and actionable results for breeding decisions.

### 3.2. Publication Source (RQ2)

Identifying leading journals and research publishing papers on deep learning in plant genomics prediction reveals the strong interdisciplinary nature of the field. Journals such as Scientific Reports, Frontiers in Plant Science, G3: Genes, Genomes, Genetics, Theoretical and Applied Genetics, Plant Genome, and BMC Genomics frequently appear in the list of publication sources. Top publication sources can be seen in Table 4. The dominance of journals focused on genetics, plant science, and general scientific research, compared to computer science or artificial intelligence journals, means that deep learning methodologies are primarily adopted, adapted, and published in the traditional biological and agricultural research communities [11].

The gap in underrepresentation in AI journals suggests opportunities for work on reporting computational evaluation standards (timing/memory profiles), leakage checks, and reproducibility assets (code, data/model cards) to facilitate cross-study comparisons. Furthermore, some low-frequency signal venues have thematic niches, such as aquatic or commodity specific, that could be leveraged if studies target specific domain applications.

This suggests successful integration of computational techniques into domain-specific problems, rather than simply theoretical computer science applications. The relative paucity of pure computer science journals suggests that innovation is often driven by biological questions and validated in biological contexts. For researchers, this means that successful contributions require a solid understanding of both deep learning principles and the specific biological context of plant genomic predictions [7]. It also provides guidance on where to search for relevant literature and where to target publications for maximum impact within the plant breeding community.

Table 4. Top publication sources.

| No | Journal source | Number of articles |
|---|---|---|
| 1 | Frontiers in Plant Science | 9 |
| 2 | Scientific Reports | 6 |
| 3 | G3: Genes, Genomes, Genetics | 5 |
| 4 | Plant Genome | 4 |
| 5 | Theoretical and Applied Genetics | 3 |
| 6 | Genetics Selection Evolution | 2 |
| 7 | BMC Genomics | 2 |
| 8 | Biology | 1 |
| 9 | Aquaculture Reports | 1 |
| 10 | Nature Plants | 1 |
| 11 | BMC Bioinformatics | 1 |
| 12 | Agriculture (Switzerland) | 1 |

### 3.3. Dataset (RQ3)

A comprehensive review of the datasets used in the reviewed studies shows a clear focus on major cereal crops and significant heterogeneity in data types. The most frequently mentioned datasets in the “Dataset” column are those involving maize [12-19], wheat [20-22], rice [23, 24], and soybean [25, 26]. These crops are very important global food commodities, which explains why most studies focus on them [27]. It could be seen in Table 5.

While some papers refer to public genomic datasets many specifically mention soybean phenotypic and genotypic data, wheat genotypes with genomic and phenotypic traits, or maize + NIR + pedigree data. This implies a significant diversity of datasets, often private or specific to a particular breeding program or experiment. This contrasts sharply with the standard public datasets (such as

MNIST or CIFAR10) often used in general computer vision. The focus on maize, wheat, rice, and soybean illustrates the importance of food security and economics that drives research on these particular crops. However, the heterogeneity of specific datasets, often associated with specific breeding programs or experimental trials, presents significant challenges for direct comparative studies. This heterogeneity makes it difficult to rigorously compare the performance of different deep learning models across studies, as results are often context dependent. There is a need for more standardized, large-scale, publicly accessible benchmark datasets that cover a wider range of crop species and trait complexities. Such datasets would greatly facilitate comparative research, accelerate model development, and promote the generalization of deep learning approaches in crop breeding.

Table 5. Distribution of dataset usage by plant type and origin (Public/private).

| Plant type | Number of articles | Data source |
|---|---|---|
| Corn | 9 | Mixed (Public/private) |
| Wheat | 8 | Mixed (Public/private) |
| Soybeans | 5 | Mixed (Public/private) |
| Rice | 3 | Mixed (Public/private) |
| Sugarcane | 3 | Private |
| Coffee | 1 | Private |
| Rubber | 1 | Private |
| Beans | 1 | Private |
| Capsicum (Chili) | 2 | Private |
| White lupin | 1 | Private |
| Oat | 1 | Private |
| Potatoes | 1 | Private |
| Rapeseed | 1 | Private |
| Arabidopsis | 1 | Private |

The sizes vary widely: from small datasets (29 chile accessions, 80 bean cultivars) [11, 28] to large datasets (1,179 maize hybrids). Number of genetic markers: the majority of papers used 10,000–70,000+ SNPs per individual; some studies used other markers (ISSR, GBS, Sanger, etc.). High dimensionality (tens of thousands of SNPs) with limited samples can potentially lead to overfitting, making feature selection and regularization crucial. Almost all studies combined genotype (SNPs/markers) and phenotype (yield, disease resistance, protein, quality, etc.) data. About half of these studies demonstrate the importance of generalizing models to real-world conditions.

Few studies integrate GWAS results to enrich more informative and biologically interpretable features. Only 8 datasets (~19%) of all papers are publicly available (OpenICPSR, Hugging Face, etc.), the rest are private or anonymous. The main public datasets come from large initiatives such as OpenICPSR, Hugging Face, CIMMYT, and are sometimes linked through institutional DOIs (but often restricted). Wheat is the most dominant (10 studies), followed by maize, soybean, rice, sugar cane, coffee, potato, oat, lupine, pepper, beans. From prediction of yield, quality, disease resistance, to trait dynamics throughout the season. As a result, comparability & reproducibility are very limited. Universal model development and learning transfer between studies is difficult. Deep learning, ensemble learning, and classical ML topics are all limited by dataset openness. Major species such as wheat, maize, soybean, the distribution of public datasets is still very small. Animal and human studies have more open access datasets than crops.

The Public category is very dominant in almost all major species (especially in the rows Other, Cattle, Maize, Wheat, and Wheat, Maize). Other (likely mixed/multi-crop or benchmark studies) has 4 Public datasets, making it the category that uses open-access data the most. Cattle (including genomic studies in livestock) also uses a lot of Public datasets. Maize is fairly balanced, with 2 Private and 2 Public, indicating a combination of private breeding data and open trial data. Maize and Coffee are recorded as using Private datasets. This is common in commodity studies where breeding results are still sensitive/commercial. Sugarcane, Other, and Soybean, Bean also appear in the Private column, indicating access to data from specific projects or countries that are not widely shared. Wheat is prominent in the Unknown category (4 papers). This can happen if the paper does not explicitly state the status of the data (or the data is available on request). Many minor species (Bean, Lupine, Potato, Rice, etc.) also have the status Unknown, possibly because the datasets are from small experiments, internal breeding, or data are not explained in the paper.

Wheat, Maize, and Cattle are the three main species that are often used. Species combinations such as (Wheat, Maize), (Wheat, Maize, Rice), and (Wheat, Pig, Mice, Pine) illustrate the trend of cross-species benchmarking in some studies. Multi-species and multi-task studies (Other row and species combinations) tend to use Public datasets, for benchmarking and reproducibility purposes. Some strategic commodities such as maize, sugarcane, and coffee still keep their breeding data private, this is understandable considering the economic value and competition between breeding programs.

To improve reproducibility and collaboration, studies on species with Private/Unknown status are recommended to be directed to Public or open-access datasets if possible. Minor species or private breeding data can be an opportunity for collaboration between institutions/universities, in order to create new open benchmark datasets. Public datasets dominate, especially for benchmarking, multi-crop, and major species (Wheat, Maize, Cattle). Unknown status is common for minor species and some wheat potential for improved dataset reporting in the future. Private studies are still important for species with high economic value or specific breeding. It can be seen in Figure 2.

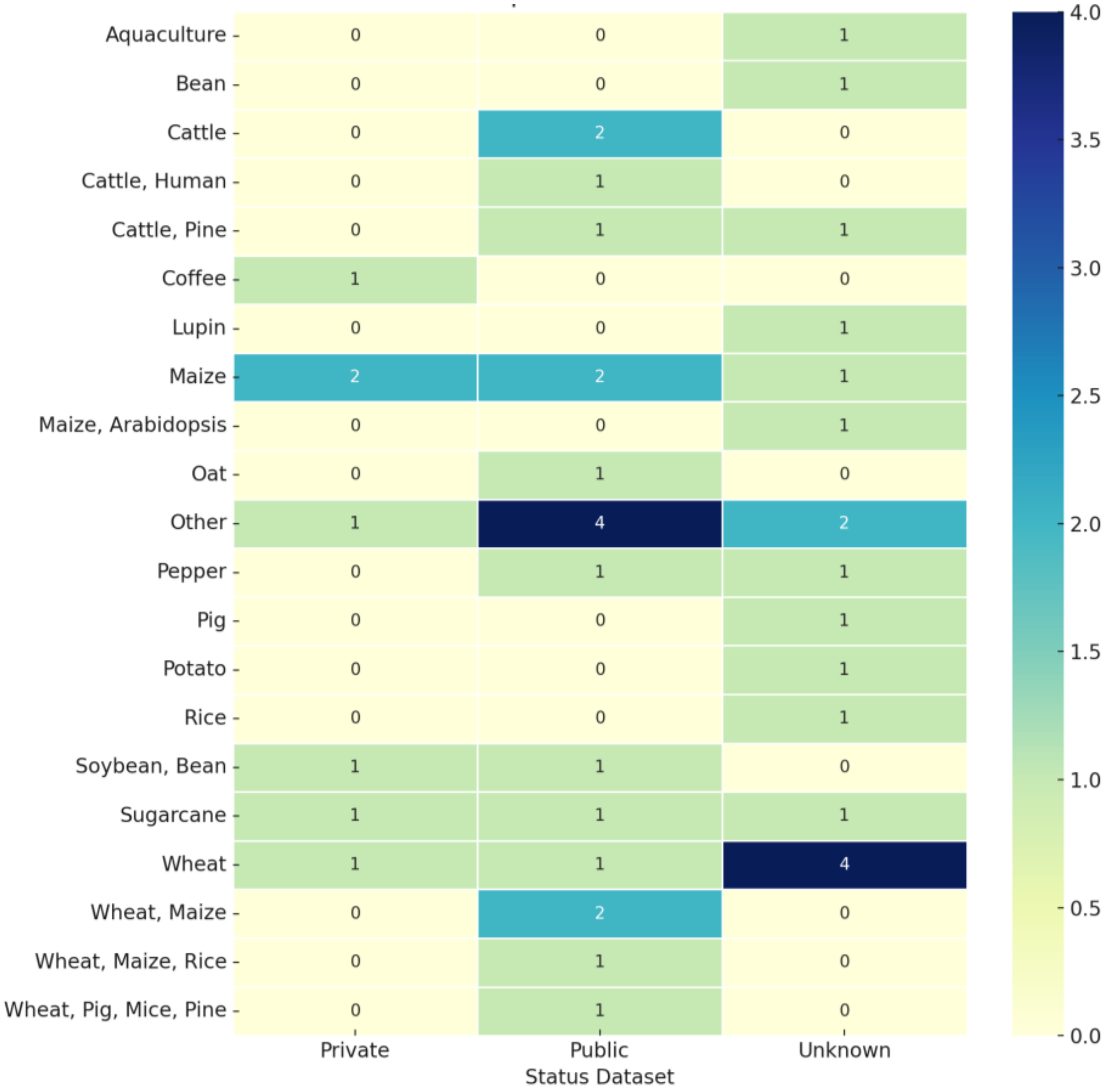

Figure 2. Species distribution vs dataset status.

### 3.4. Research Topic (RQ4)

The categorization of research topics frequently discussed in the reviewed literature shows a clear evolution from basic benchmarks towards more complex and realistic prediction scenarios. Initially, many studies focused on topics such as (DL vs SVM vs BLUP benchmarks) or (Model comparison), aiming to fundamentally demonstrate the effectiveness of deep learning compared to conventional statistical and machine learning methods for basic genomic prediction tasks (e.g., single trait, single environment). It can be seen on Table 6.

Table 6. Mapping of research topic clusters in genomic prediction studies.

| Topic clusters | Research topics | Number of articles |
|---|---|---|
| Model comparison & optimization | Benchmarking DL vs SVM vs BLUP, Model comparison, Automated model selection, Hyperparameter optimization, ML vs GBLUP | 10 |
| Complex trait prediction | Multivariate genomic prediction, Multienvironment genomic prediction, Multi-trait multi-environment prediction, Predicting plant trait dynamics, Genomic prediction for complex traits | 10 |
| Data & feature integration | LLM-based representation for genomic features, Combined genomic and phenomic data, Integration of ML and GWAS, Spectral information in genomic selection | 6 |
| DL/ML architecture & methods | Stacking Ensemble Learning, Hybrid deep kernel methods, ElasticNet + BERT + attention mechanism, Multimodal DL, Stacked kinship CNN, Shallow Quantum Neural Networks | 11 |
| Crop/trait specific prediction | Soybean branching prediction, Flowering traits in bean, Rice traits using genomic variants, Anthracnose resistance in sugarcane, Sugarcane single-stalk weight prediction | 6 |

However, there is a clear shift towards addressing more complex and realistic breeding challenges. Topics such as Multivariate genomic prediction, Multi-environment genomic prediction, Ensemble learning for multi-trait genomic prediction, Uncertainty-aware prediction, LLM-based representation for genomic features, and dynamicGP + Dynamic Mode Decomposition for time series prediction are gaining prominence [29]. This evolution indicates that the field has matured, moving from basic validation to addressing the complexities inherent in real-world crop breeding programs. This includes predicting multiple traits simultaneously (“multi-trait”), accounting for genotype-environment interactions (“multi-environment”), quantifying prediction uncertainty, and even exploring new data representations (e.g., using Large Language Models for DNA sequences) as well as dynamic trait prediction over time. This shift has significant implications [30]. Future research should increasingly prioritize these complex scenarios, as they have the greatest potential to make a practical impact in accelerating crop improvement. The exploration of advanced techniques such as Large Language Models (LLM) also marks a new frontier in integrating various biological information and computational paradigms. This shows that researchers are now focusing not only on whether DL can work, but also how DL can be optimized to address the most pressing challenges in plant breeding.

DL methods such as CNN, DNN, attention, and transformer have been at the forefront of cutting-edge research. These models excel in multi-trait prediction, big data, and multi-modal tasks. Multi-task CNN and stack kinship models are increasingly explored, especially for multitrait & multi-environment problems. Ensemble/Hybrid Stacking [27], LightGBM [31], and hybrid DL+GBLUP/ENet have become popular strategies for higher stability and accuracy, especially in real-world conditions [9]. Baseline & Classical ML such as BLUP (GBLUP/Bayes BLUP) and ridge/penalized regression are still used as benchmarks in almost all papers. SVM, Elastic Net, Random Forest, and XGBoost are commonly used as comparisons and baselines [32]. Multiple kernel learning (including deep kernel) is used in high-dimensional and complex trait cases, but its frequency is still small. PLS (Partial Least Squares) is effective in multidimensional/high trait data, but is relatively rare compared to DL/ensemble [33].

### 3.5. Tabulation of Genomic Prediction Methods/Models/Algorithms (RQ5)

New studies are starting to combine ML with GWAS results for marker selection and improved biological interpretation. The integration of phenomics (imagery, spectral) and environment (weather, soil) has increased in the last 2 years, in line with the availability of high throughput field & phenotype data. AutoML & Hyperparameter Optimization Still a major gap. Although mentioned in several papers, AutoML workflows or deep hyperparameter tuning have only been systematically applied [12]. Recommendations and Gaps are Hybrid & Multi-modal models have been shown to provide significant benefits, both in terms of statistics and biological/agronomic interpretation. AutoML & hyperparameter optimization should be more widely adopted for modern DL models. Comparisons between models are still constrained by limited public datasets which makes open benchmarking initiatives necessary for faster progress. New studies are advised to always compare at least BLUP, one classical ML, one DL model, and one ensemble model. All of these details can be seen in Table 7.

Tabel 7. Tabulation of genomic prediction methods/models/algorithms (RQ5).

| Category | Model/algorithm | Number of paper | Analysis |
|---|---|---|---|
| Deep learning | CNN, Autoencoder, Deep Kernel, DNN, LSTM, Transformer, Multi-task DL, Attention, Stacked kinship CNN | 14 | Dominant, used for nonlinear, multitrait/multienvironment relationships, often outperforms classical ML on big data |
| Ensemble/hybrid | Stacking, LightGBM, Random Forest, Hybrid (DL+GBLUP/ENet), Stacked ensemble parentage | 7 | Blending/stacking models often outperforms single models, stable for multi-trait & realworld data |
| Kernel methods | Multiple Kernel Learning, Deep Kernel, Sparse Kernels (SKM) | 3 | Focus on capturing nonlinearities & complex relationships, especially medium/small datasets |
| Partial least squares | PLS, Multi-trait PLS | 2 | Effective for dimensionality reduction & multitrait prediction on high-dimensional data, small-medium samples |
| GWAS integration | GWAS-feature selection, GWASinformed ML | 3 | Improve interpretability & direct ML to relevant biological markers |
| BLUP methods | GBLUP, Bayesian BLUP, Ridge regression | 6 | Classic baseline, robust for small samples, used as a mandatory comparison |
| Machine learning | SVM, Elastic Net, Lasso, Ridge, Random Forest, XGBoost | 7 | Classical ML is widely used for baseline & interpretable models. Often combined with feature selection/GWAS |
| AutoML/optimization | Hyperparameter tuning (TPE, Bayesian opt), Automated ML | 1 | Still rare, only a few studies have used AutoML pipelines/systematic tuning |
| Phenomic/enviromic | DL/ML with spectral, image, weather features | 4 | Multi-modal models are increasingly being used, increasing accuracy over SNP data alone |

### 3.6. Schematics, Flowcharts, or Pseudocode of Genomic Prediction (RQ6)

Presentations and discussions of the architectural and algorithmic workflows commonly used in the reviewed studies reveal a significant evolution towards integrated and specialized pipelines for genomic data. Descriptions of Flowcharts in the reviewed studies range from Standard ML Pipeline: Pre-processing → Model Training → Evaluation to more complex ones such as Pre-processing → Feature Selection → Model Training → SHAP Interpretation, Tokenization → Embedding → Transfer Learning, DL Pipeline for SNP encoding, training, validation, Joint feature extraction → model prediction, and Multi-input properties → Deep neural network → Multi-output prediction. It can be seen in Table 8.

Tabel 8. General genomic prediction pipeline stages with deep learning integration.

| Pipeline stages | General description | Implementation examples from the study |
|---|---|---|
| Data preprocessing | Cleaning, normalization, handling missing values, SNP encoding | SNP Encoding, Pre-processing → Feature Selection |
| Feature extraction / embedding | Transforming raw data (e.g., DNA sequences, image data) into learnable representations | BERT-style embedding for DNA sequences, Combined feature extraction |
| DL model architecture | Deep neural network design (e.g., number of layers, layer types, activation functions) | Deep neural network (DNN), CNN, MLP, Autoencoder, DBN, Multi-task deep learning (MtCro) |
| Model/data integration | Combining different models or data modalities for better predictions | Stacking (RF + SVM + DNN), Hybrid deep kernel methods, Multimodal DL, Stacked kinship CNN |
| Model training | Fitting models to training data, including hyperparameter optimization | Model training → Prediction, Hyperparameter optimization (TPE) |
| Evaluation & interpretation | Measuring model performance and understanding feature contributions | Evaluation, SHAP Interpretation |

While basic machine learning pipelines are a starting point, the field is clearly moving toward highly specialized and integrated workflows designed to address the unique characteristics of genomic data. These include sophisticated steps for SNP decoding, advanced feature selection (often integrating insights from Genome-Wide Association Studies (GWAS) [29], specialized embedding techniques for DNA sequences (e.g., BERT-style for DNA) [34], and multi-input/multi-output architectures to handle diverse data modalities and prediction tasks. The success of deep learning in genomic prediction is increasingly tied to the efficacy of this entire end-to-end pipeline, rather than just its core deep learning model. This highlights the need for researchers to focus not only on novel deep learning models but also on optimizing the entire data processing and modeling pipeline. It also suggests future directions involving the development of standardized, user-friendly computational frameworks that encapsulate these best practices for handling genomic data, making advanced genomic predictions more accessible to plant breeders.

### 3.7. Evaluation Metrics (RQ7)

To assess the efficacy of genomic prediction models, various metrics are used, reflecting different aspects of predictive performance. Pearson Correlation Coefficient (R), widely used to measure the correlation between predicted breeding values and observed phenotypes. It indicates the accuracy of prediction. Mean Squared Error (MSE) / Root Mean Squared Error (RMSE), measures the average of the squared differences between predicted and actual values, indicating the prediction error. The lowest MSE is often the target. Mean Absolute Error (MAE), measures the average of the absolute differences between predicted and actual values [35]. $R^2$, represents the proportion of variance in the dependent variable that can be predicted from the independent variables. Computational Time, a critical metric, especially for large-scale applications, indicates the efficiency of a DL model. The speed at which DL can perform computations is a major advantage. Other

Metrics: F1-score, recall, sensitivity, specificity, Brier Score, Misclassification Rate, and Normalized Root Mean Squared Error (NRMSE) are also used depending on the task [36]. Although accuracy (Pearson R, $R^2$, low MSE) remains paramount, frequent mention of computational load , computational cost, faster training, and speed in running computations alongside accuracy metrics show a shift. This suggests that for large-scale practical breeding programs, predictive power alone is not enough. Efficiency (time, resources) becomes an equally important evaluation criterion. This implies that future model development will increasingly need to balance predictive accuracy with computational feasibility [30], going beyond purely theoretical performance gains to address real-world operational constraints. It can be seen in Table 9.

Tabel 9. Evaluation metrics.

| Metric name | Description | Interpretation |
|---|---|---|
| Pearson correlation coefficient (R) | Measures the strength and direction of the linear relationship between predicted and actual values | Prediction accuracy; higher values indicate better predictions |
| Mean squared error (MSE) | The mean of the squares of the differences between predicted and actual values | Measure of prediction error; lower values indicate better model |
| Root mean squared error (RMSE) | The square root of the MSE | Measure of prediction error in the same units as the target variable; lower values indicate better model |
| Mean absolute error (MAE) | The mean of the absolute values of the differences between predicted and actual values | Measure of prediction error, less sensitive to outliers than MSE; lower values indicate better model |
| $R^2$ | The proportion of the variance in the dependent variable that can be predicted from the independent variables | How well the model fits the data; higher values indicate better fit |
| Computation Time | The time it takes a model to train or make predictions | Efficiency models; lower times indicate more efficient model |
| F1-score | The harmonic mean of precision and recall | Balance between precision and recall, important for imbalanced classes |
| Sensitivity (Recall) | The proportion of true positives correctly identified | The ability of the model to find all positive cases |
| Specificity | The proportion of true negatives correctly identified | The ability of the model to correctly identify negative cases |

## 4. CONCLUSIONS

This systematic literature review has comprehensively analyzed the deep learning landscape in crop genomics prediction. Findings show a significant increase in interest in this field, with an accelerated publication trend, especially from 2019 to 2025. Journals focused on genetics and crop sciences dominate the publications, underscoring the interdisciplinary nature of this research. The datasets used are mostly centered on major cereal crops such as maize, wheat, rice, and soybean, although with marked heterogeneity that hinders direct comparisons between studies. Research topics have evolved from basic benchmarks to more complex prediction scenarios, including multi-trait and multi-environment predictions. Methodologically, there is a clear trend towards hybrid and ensemble deep learning approaches, indicating an attempt to combine the strengths of different models and integrate diverse data modalities. Commonly used evaluation metrics are prediction accuracy, RMSE, R2, and MAE, with a growing recognition of the importance of computational efficiency and model interpretability. The findings of this review have broad implications for plant breeders, geneticists, and computational biologists.

Deep learning is progressively [37], offering new capabilities to accelerate crop improvement through more accurate and efficient predictions. The ability of DL models to capture non-linear relationships and complex interactions in genomic and phenotypic data represents a significant advance over traditional methods [38]. The successful integration of advanced computational techniques into real-world crop breeding programs underscores the transformative potential of this field. However, to maximize this impact, continued interdisciplinary collaboration is essential. The gap between advanced computational methodologies and practical breeding challenges needs to be bridged, ensuring that theoretical innovations can be translated into applicable and impactful solutions in the field [32]. Based on the analysis conducted, several open challenges and promising avenues for future work can be identified. Development of standardized data, there is an urgent need for more standardized, large-scale, and publicly accessible benchmark datasets. These datasets should cover a wider range of crop species and varying trait complexities to facilitate robust comparative studies and accelerate model development.

Complex trait and environment modeling future research should focus more on developing deep learning models for multi-trait [39], multi-environment [33], and timeseries prediction. This will enable addressing more realistic and complex breeding challenges, moving beyond single-trait and single-environment scenarios. Advanced model architectures exploration and development of intelligent hybrid and ensemble deep learning frameworks should continue. These architectures should be able to effectively integrate diverse data modalities (e.g., genomics, phenomics, environment, imaging) and leverage the complementary strengths of different models for optimal performance.

Research should be invested in optimizing the entire deep learning workflow, including novel genomic data encoding strategies, automated feature engineering, and efficient training protocols. The success of genomic prediction increasingly depends on the efficacy of the entire workflow. Improving model interpretability and uncertainty quantification in developing methods to improve the interpretability of deep learning models in genomic prediction is essential. Furthermore, quantifying the uncertainty of their predictions will build confidence and facilitate practical adoption by breeders. Computational efficiency and scalability addressing the computational demands of deep learning models for large-scale genomic datasets remains a priority. This may involve exploring distributed computing or quantum-inspired approaches, although the latter is beyond the immediate scope of this SLR, a longer-term consideration for computational efficiency. Novel data representations and further investigation into advanced data representations, such as the application of Large Language Models (LLMs) to genomic sequences, may unlock new predictive capabilities that have yet to be explored. Overall, the field of deep learning-powered plant genomic prediction is on a rapid growth trajectory, promising significant advances in crop breeding. By addressing the challenges outlined and pursuing these promising research directions, the scientific community can further harness the potential of deep learning for global food security.

## REFERENCES

[1] Montesinos-López, O. A., Martín-Vallejo, J., Crossa, J., Gianola, D., Hernández-Suárez, C. M., Montesinos-López, A., Juliana, P., & Singh, R. (2019). A benchmarking between deep learning, support vector machine and Bayesian threshold best linear unbiased prediction for predicting ordinal traits in plant breeding. *G3: Genes, Genomes, Genetics*, **9**(2), 601–618.

[2] Lozada, D. N., Sandhu, K. S., & Bhatta, M. (2023). Ridge regression and deep learning models for genome-wide selection of complex traits in New Mexican Chile peppers. *BMC Genomic Data*, **24**(1), 80.

[3] Montesinos-Lopez, A., Crespo-Herrera, L., Dreisigacker, S., Gerard, G., Vitale, P., Saint Pierre, C., Govindan, V., Tarekegn, Z. T., Flores, M. C., Pérez-Rodríguez, P., Ramos-Pulido, S., Lillemo, M., Li, H., Montesinos-López, O. A., & Crossa, J. (2024). Deep learning methods improve genomic prediction of wheat breeding. *Frontiers in Plant Science*, **15**, 1324090.

[4] Ji, L., Hou, W., Zhou, H., Xiong, L., Liu, C., Yuan, Z., & Li, L. (2025). EBMGP: a deep learning model for genomic prediction based on Elastic Net feature selection and bidirectional encoder representations from transformer's embedding and multi-head attention pooling. *Theoretical and Applied Genetics*, **138**(5), 103.

[5] Cuevas, J., Montesinos-López, O., Juliana, P., Guzmán, C., Pérez-Rodríguez, P., González-Bucio, J., Burgueño, J., Montesinos-López, A., & Crossa, J. (2019). Deep kernel for genomic and near infrared predictions in multi-environment breeding trials. *G3: Genes, Genomes, Genetics*, **9**(9), 2913–2924.

[6] Waldmann, P. (2018). Approximate Bayesian neural networks in genomic prediction. *Genetics Selection Evolution*, **50**(1), 70.

[7] Wang, K., Abid, M. A., Rasheed, A., Crossa, J., Hearne, S., & Li, H. (2023). DNNGP, a deep neural network-based method for genomic prediction using multi-omics data in plants. *Molecular Plant*, **16**(1), 279–293.

[8] Liang, M., An, B., Li, K., Du, L., Deng, T., Cao, S., Du, Y., Xu, L., Gao, X., Zhang, L., Li, J., & Gao, H. (2022). Improving genomic prediction with machine learning incorporating TPE for hyperparameters optimization. *Biology*, **11**(11), 1647.

[9] Liang, M., Chang, T., An, B., Duan, X., Du, L., Wang, X., Miao, J., Xu, L., Gao, X., Zhang, L. Li, J., & Gao, H. (2021). A stacking ensemble learning framework for genomic prediction. *Frontiers in Genetics*, **12**, 600040.

[10] Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering–a systematic literature review. *Information and Software Technology*, **51**(1), 7–15.

[11] Rosado, R. D. S., Cruz, C. D., Barili, L. D., de Souza Carneiro, J. E., Carneiro, P. C. S., Carneiro, V. Q., da Silva, J. T., & Nascimento, M. (2020). Artificial neural networks in the prediction of genetic merit to flowering traits in bean cultivars. *Agriculture*, **10**(12), 638.

[12] Galli, G., Sabadin, F., Yassue, R. M., Galves, C., Carvalho, H. F., Crossa, J., Montesinos-López, O. A., & Fritsche-Neto, R. (2022). Automated machine learning: a case study of genomic "image-based" prediction in maize hybrids. *Frontiers in Plant Science*, **13**, 845524.

[13] Barreto, C. A. V., das Graças Dias, K. O., de Sousa, I. C., Azevedo, C. F., Nascimento, A. C. C., Guimarães, L. J. M., Guimarães, C. T., Pastina, M. M., & Nascimento, M. (2024). Genomic prediction in multi-environment trials in maize using statistical and machine learning methods. *Scientific Reports*, **14**(1), 1062.

[14] Shayanowako, A. I. T., Shimelis, H., Laing, M. D., & Mwadzingeni, L. (2018). Genetic diversity of maize genotypes with variable resistance to Striga asiatica based on SSR markers. *Cereal Research Communications*, **46**(4), 668–678.

[15] Ramstein, G. P. & Buckler, E. S. (2022). Prediction of evolutionary constraint by genomic annotations improves functional prioritization of genomic variants in maize. *Genome Biology*, **23**(1), 183.

[16] 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., & Zheng, H. (2022). Genomic prediction of drought tolerance during seedling stage in maize using low-cost molecular markers. *Euphytica*, **218**(11), 154.

[17] Fernandes, I. K., Vieira, C. C., Dias, K. O., & Fernandes, S. B. (2024). Using machine learning to combine genetic and environmental data for maize grain yield predictions across multi-environment trials. *Theoretical and Applied Genetics*, **137**(8), 189.

[18] Zhang, T., Jiang, L., Ruan, L., Qian, Y., Liang, S., Lin, F., Lu, H., Dai, H., & Zhao, H. (2021). Heterotic quantitative trait loci analysis and genomic prediction of seedling biomass-related traits in maize triple testcross populations. *Plant Methods*, **17**(1), 85.

[19] Lopez-Cruz, M., Aguate, F. M., Washburn, J. D., De Leon, N., Kaeppler, S. M., Lima, D. C., Tan, R., Thompson, A., De La Bretonne, L. W., & de Los Campos, G. (2023). Leveraging data from the Genomes-to-Fields Initiative to investigate genotype-by-environment interactions in maize in North America. *Nature Communications*, **14**(1), 6904.

[20] Montesinos-López, O. A., Montesinos-López, A., Tuberosa, R., Maccaferri, M., Sciara, G., Ammar, K., & Crossa, J. (2019). Multi-trait, multi-environment genomic prediction of durum wheat with genomic best linear unbiased predictor and deep learning methods. *Frontiers in Plant Science*, **10**, 1311.

[21] Berkner, M. O., Schulthess, A. W., Zhao, Y., Jiang, Y., Oppermann, M., & Reif, J. C. (2022). Choosing the right tool: Leveraging of plant genetic resources in wheat (*Triticum aestivum L.*) benefits from selection of a suitable genomic prediction model. *Theoretical and Applied Genetics*, **135**(12), 4391–4407.

[22] Mróz, T., Shafiee, S., Crossa, J., Montesinos-Lopez, O. A., & Lillemo, M. (2024). Multispectral-derived genotypic similarities from budget cameras allow grain yield prediction and genomic selection augmentation in single and multi-environment scenarios in spring wheat. *Molecular Breeding*, **44**(1), 5.

[23] Ghosal, S., Quilloy, F. A., Casal Jr, C., Septiningsih, E. M., Mendioro, M. S., & Dixit, S. (2020). Trait-based mapping to identify the genetic factors underlying anaerobic germination of rice: Phenotyping, GXE, and QTL mapping. *BMC Genetics*, **21**(1), 6.

[24] Vourlaki, I. T., Ramos-Onsins, S. E., Pérez-Enciso, M., & Castanera, R. (2024). Evaluation of deep learning for predicting rice traits using structural and single-nucleotide genomic variants. *Plant Methods*, **20**(1), 121.

[25] Van der Laan, L., Parmley, K., Saadati, M., Pacin, H. T., Panthulugiri, S., Sarkar, S., Ganapathysubramanian, B., Lorenz, A., & Singh, A. K. (2025). Genomic and phenomic prediction for soybean seed yield, protein, and oil. *The Plant Genome*, **18**(1), e70002.

[26] Zhou, W., Yan, Z., & Zhang, L. (2024). A comparative study of 11 non-linear regression models highlighting autoencoder, DBN, and SVR, enhanced by SHAP importance analysis in soybean branching prediction. *Scientific Reports*, **14**(1), 5905.

[27] Chao, D., Wang, H., Wan, F., Yan, S., Fang, W., & Yang, Y. (2025). Mtcro: multi-task deep learning framework improves multi-trait genomic prediction of crops. *Plant Methods*, **21**(1), 12.

[28] Ortiz, R., Reslow, F., Montesinos-López, A., Huicho, J., Pérez-Rodríguez, P., Montesinos-López, O. A., & Crossa, J. (2023). Partial least squares enhance multi-trait genomic prediction of potato cultivars in new environments. *Scientific Reports*, **13**(1), 9947.

[29] Hobby, D., Tong, H., Heuermann, M., Mbebi, A. J., Laitinen, R. A., Dell'Acqua, M., Altmann, T., & Nikoloski, Z. (2025). Predicting plant trait dynamics from genetic markers. *Nature Plants*, **11**(5), 1018–1027.

[30] Mendoza-Revilla, J., Trop, E., Gonzalez, L., Roller, M., Dalla-Torre, H., de Almeida, B. P., Richard, G., Caton, J., Lopez Carranza, N., Skwark, M., Laterre, A., Beguir, K., Pierrot, T., & Lopez, M. (2024). A foundational large language model for edible plant genomes. *Communications Biology*, **7**(1), 835.

[31] Heilmann, P. G., Frisch, M., Abbadi, A., Kox, T., & Herzog, E. (2023). Stacked ensembles on basis of parentage information can predict hybrid performance with an accuracy comparable to marker-based GBLUP. *Frontiers in Plant Science*, **14**, 1178902.

[32] Lourenço, V. M., Ogutu, J. O., Rodrigues, R. A., Posekany, A., & Piepho, H. P. (2024). Genomic prediction using machine learning: a comparison of the performance of regularized regression, ensemble, instance-based and deep learning methods on synthetic and empirical data. *BMC Genomics*, **25**(1), 152.

[33] Montesinos-López, O. A., Montesinos-López, A., Bernal Sandoval, D. A., Mosqueda-Gonzalez, B. A., Valenzo-Jiménez, M. A., & Crossa, J. (2022). Multi-trait genome prediction of new environments with partial least squares. *Frontiers in Genetics*, **13**, 966775.

[34] Yan, J., Xu, Y., Cheng, Q., Jiang, S., Wang, Q., Xiao, Y., Ma, C., Yan, J., & Wang, X. (2021). LightGBM: accelerated genomically designed crop breeding through ensemble learning. *Genome Biology*, **22**(1), 271.

[35] Fan, Y. & Waldmann, P. (2024). Tabular deep learning: a comparative study applied to multi-task genome-wide prediction. *BMC Bioinformatics*, **25**(1), 322.

[36] Montesinos-López, O. A., Montesinos-López, A., Cano-Paez, B., Hernández-Suárez, C. M., Santana-Mancilla, P. C., & Crossa, J. (2022). A comparison of three machine learning methods for multivariate genomic prediction using the sparse kernels method (SKM) library. *Genes*, **13**(8), 1494.

[37] Abdollahi-Arpanahi, R., Gianola, D., & Peñagaricano, F. (2020). Deep learning versus parametric and ensemble methods for genomic prediction of complex phenotypes. *Genetics Selection Evolution*, **52**(1), 12.

[38] Maleki, H. H., Darvishzadeh, R., & Azad, N. (2025). Sweet pepper yield modeling via deep learning and selection of superior genotypes using GBLUP and MGIDI. *Scientific Reports*, **15**(1), 14718.

[39] Sandhu, K., Patil, S. S., Pumphrey, M., & Carter, A. (2021). Multitrait machine-and deep-learning models for genomic selection using spectral information in a wheat breeding program. *The Plant Genome*, **14**(3), e20119.