

Review

Explainable Artificial Intelligence for Genomic Prediction in Plants: A Critical Review of Methods, Biological Insights, and Practical Challenges

Agata Głuchowska ¹, Muhammad Hafeez Ullah Khan ² and Magdalena Pawełkowicz ^{1,*} 

¹ Department of Plant Genetics, Breeding and Plant Biotechnology, Institute of Biology, Warsaw University of Life Sciences–SGGW, Nowoursynowska 159, 02-776 Warsaw, Poland; agata_gluchowska@sggw.edu.pl

² Henan Collaborative Innovation Center of Modern Biological Breeding, Henan Institute of Science and Technology, Xinxiang 453003, China; hafeezbiotech1989@gmail.com

* Correspondence: magdalena_pawelkowicz@sggw.edu.pl

Abstract

Machine learning has become an important tool in plant genomic prediction for modeling complex genotype–phenotype relationships and improving breeding decisions. However, many high-performing models, particularly ensemble and deep learning approaches, remain difficult to interpret, limiting their biological applicability. This review summarizes major machine learning methods and explainable artificial intelligence (XAI) approaches used in plant genomics, including SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-Agnostic Explanations), attention mechanisms, permutation importance, tree-based feature importance, and gradient-based attribution methods. XAI can help identify influential SNPs, genomic regions, candidate genes, regulatory elements, and omics features associated with complex traits. For example, SHAP analysis in an almond germplasm collection identified a genomic region associated with shelling fraction, illustrating how XAI can generate testable hypotheses for further validation. The review further discusses applications in trait prediction, breeding, functional genomics, and multi-omics integration. Importantly, we emphasize major limitations, including data bias, model instability, correlated genomic markers, limited model transferability, and the common misconception that feature importance implies biological causality. We recommend integrating XAI with linkage disequilibrium pruning, stability assessment, biological annotation, and experimental validation before prioritizing candidate genes. Overall, XAI should be considered a framework for model interpretation, feature prioritization, and hypothesis generation rather than a replacement for experimental validation in plant genomics and breeding.

Keywords: explainable artificial intelligence (XAI); plant genomics; machine learning; snp markers; genomic selection; SHAP; LIME; plant breeding; genotype–phenotype prediction



Academic Editor: Dirk Tischler

Received: 29 May 2026

Revised: 14 July 2026

Accepted: 17 July 2026

Published: 21 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Recent advances in high-throughput sequencing, genotyping platforms, and digital phenotyping have substantially transformed the analysis and use of plant genetic resources in breeding. Contemporary plant genomics now routinely generates large-scale datasets, including dense single-nucleotide polymorphism (SNP) profiles, which can be integrated with phenotypic and environmental information to improve the prediction of complex agronomic traits. This development has strengthened the role of genomic prediction in

crop improvement, particularly for traits such as yield, stress tolerance, disease resistance, flowering time, and biomass, which are typically controlled by many loci of small effect and are strongly influenced by environmental conditions.

The conceptual foundation of genomic prediction was established by Meuwissen et al. (2001), who demonstrated that genome-wide dense molecular markers could be used to predict genetic value more efficiently than approaches based only on phenotypic observations or a limited number of significant loci [1]. In breeding programs, regulatory approval, crossing decisions, and intellectual property claims increasingly require transparent justification of model-based selections, making interpretability not merely an academic concern but a practical necessity. Since then, genomic prediction has become an important component of modern plant breeding, supporting selection decisions in populations where direct phenotypic evaluation is costly, time-consuming, or strongly affected by environmental variation [2–4]. Recent scoping reviews also indicate that genomic selection is increasingly applied in crop improvement, including fruit and vegetable breeding, although its performance depends strongly on training-population design, trait architecture, validation strategy, and environmental context [5].

However, the increasing availability of high-dimensional genomic data has also introduced important analytical challenges. Plant traits are rarely determined by single genetic effects; instead, they often arise from complex interactions among loci, regulatory mechanisms, population structure, linkage disequilibrium, and genotype-by-environment interactions [6–8]. These characteristics make genomic prediction a suitable area for machine learning approaches, which can model nonlinear relationships and high-dimensional patterns that may be difficult to capture using classical linear models alone.

Machine learning methods, including random forests, support vector machines, gradient boosting, and deep neural networks, have therefore attracted growing interest in plant genomic selection [9–12]. Their use is particularly relevant when genomic data are combined with phenotypic, environmental, pangenomic, or multi-omics information. Recent methodological tools, such as ShinyGS, an interactive graphical platform that enables users to benchmark statistical, machine learning, and genomic selection models without extensive programming expertise, further show that plant breeders increasingly require accessible workflows that compare statistical and machine learning models under different genomic selection scenarios, though they give little insight into the genomic features, candidate genes, or molecular mechanisms that drive those predictions [13].

This lack of interpretability remains a major limitation for the broader use of machine learning in plant genomics and breeding. In applied breeding contexts, prediction alone is often insufficient; researchers and breeders also need to assess whether biologically plausible signals support model outputs. This is especially relevant for SNP-based datasets, where correlated markers, linkage disequilibrium, population structure, and dataset-specific biases may produce feature-importance patterns that are difficult to interpret. Explainable artificial intelligence (XAI) has emerged as a methodological response to this problem by providing tools to examine how complex models generate predictions and which input features contribute most strongly to model outputs [14–16].

In plant breeding, recent studies have begun to apply XAI directly to genotype-to-phenotype prediction. The use of explainable artificial intelligence in an almond germplasm collection shows how model explanations can support the interpretation of genotype–phenotype relationships in a breeding-relevant context [17]. Such studies indicate that XAI may help connect predictive modeling with candidate genomic regions, biological hypotheses, and more transparent breeding decisions. However, explanations derived from machine learning models must be interpreted cautiously. Feature importance does

not necessarily imply causality, and highly ranked SNPs may reflect linkage, correlation, or population structure rather than direct biological effects [18,19].

Despite the growing use of machine learning in plant genomic prediction and the increasing interest in XAI, the connection between these two areas remains insufficiently synthesized. No prior review has systematically compared how different XAI methods perform on SNP-based versus multi-omics data in plants, nor proposed a standardized validation workflow for XAI outputs [20,21]. Existing studies often focus on specific algorithms, crops, traits, or datasets, whereas fewer works critically examine how explainability methods can support the biological interpretation of genomic prediction models in plants.

This review focuses primarily on SNP-based genomic prediction because SNPs remain the most widely available and extensively used marker type in genomic selection studies across crop species [2,3]. Their high genome-wide density, compatibility with modern genotyping platforms, and extensive representation in public datasets make them the dominant data source for current machine learning applications [2,4]. Nevertheless, many concepts discussed in this review are also applicable to other marker systems, including SSRs, InDels, structural variants, haplotypes, pangenomic markers, and multi-omics features. These marker systems are briefly discussed where relevant, particularly in the context of emerging multi-omics and pangenomic approaches [4,22].

Therefore, the objective of this review is to provide a focused overview of explainable artificial intelligence methods applied to genomic prediction in plants. Specifically, this review aims to: (i) summarize major machine learning approaches used in plant genomic prediction; (ii) discuss key XAI methods and their relevance for genomic and multi-omics data; (iii) evaluate how these methods may support the identification of biologically meaningful genomic features; and (iv) critically examine current limitations, including data bias, instability of explanations, correlated markers, and the risk of interpreting feature importance as biological causality. By integrating these perspectives, this review aims to clarify how XAI can contribute to more transparent, reproducible, and biologically informative genomic prediction in plant breeding and functional genomics.

2. Machine Learning in Genomic Prediction

Machine learning (ML) is a branch of artificial intelligence that enables computational models to learn patterns from data and use these patterns to make predictions or support decision-making. Unlike traditional rule-based programming, where relationships must be explicitly defined in advance, ML models infer relationships from examples. In genomic prediction, this means that a model is trained on individuals for which both genomic marker data and phenotypic values are available, and then applies the learned genotype–phenotype relationships to predict traits in new, untested genotypes [23].

This distinction is particularly important in plant genomics, where biological relationships are often too complex to be described by a small number of predefined rules. A typical genomic prediction dataset may include thousands or millions of single-nucleotide polymorphisms (SNPs), whereas the number of genotypes, accessions, breeding lines, or cultivars is often comparatively limited. This creates a high-dimensional, low-sample-size setting, in which many markers must be analyzed simultaneously despite a restricted number of observations. Such data structures are common in plant breeding and represent one of the central statistical challenges of genomic prediction.

The basic aim of genomic prediction is to estimate the genetic potential of individuals using genome-wide marker information. The conceptual foundation of this approach was established through the demonstration that dense molecular markers distributed across the genome could be used to predict total genetic value [1]. Since then, genomic selection has become an important strategy for increasing genetic gain and shortening breeding cycles,

particularly for complex traits that are expensive, time-consuming, or environmentally sensitive to measure directly [5,24]. Recent reviews confirm that genomic selection is increasingly used in crop improvement, including fruit and vegetable breeding, although its practical performance depends on population size, marker density, trait architecture, training-population design, and validation strategy [5].

Classical genomic prediction methods, such as genomic best linear unbiased prediction (GBLUP), Bayesian ridge regression, BayesA, BayesB, and reproducing kernel Hilbert space regression, remain widely used because they are statistically well established and computationally efficient [1,25]. These methods have provided a strong foundation for genomic selection in both animal and plant breeding. However, many of them rely on assumptions about additive genetic effects or relatively simple covariance structures. As a result, they may not fully capture nonlinear genotype–phenotype relationships, epistasis, genotype-by-environment interactions, or complex interactions among genomic, phenotypic, environmental, and omics-level variables.

Machine learning methods offer a more flexible modeling framework. Instead of assuming a predefined linear relationship between markers and traits, ML algorithms can learn complex patterns directly from data. This flexibility is useful for traits such as yield, flowering time, drought tolerance, disease resistance, biomass accumulation, and nutrient-use efficiency, which are typically controlled by many loci of small effect and may be influenced by regulatory networks, environmental conditions, and interactions among genomic regions. In this context, ML has gained increasing attention as a tool for genomic prediction, especially when genomic data are combined with phenotypic, environmental, pangenomic, or multi-omics information [9,13,24,26]. Recent benchmarking tools such as ShinyGS also show that plant breeders increasingly need accessible workflows that combine statistical and machine learning models for genomic selection [13]. ShinyGS is an interactive R Shiny-based platform developed to facilitate genomic selection analyses through a user-friendly graphical interface. The tool integrates multiple genomic prediction approaches, including conventional statistical methods (e.g., GBLUP and Bayesian models) as well as machine learning algorithms such as random forests, support vector machines, and deep learning models. In addition, ShinyGS enables model benchmarking, cross-validation, prediction accuracy assessment, and comparative evaluation of genomic selection strategies, thereby supporting reproducible and accessible genomic prediction workflows for plant breeding applications [13].

2.1. Classical Machine Learning Models

Classical machine learning models remain widely used in genomic prediction because they provide a practical balance between predictive performance, computational cost, and relative interpretability. Among the most frequently applied approaches are random forests, support vector machines, and gradient boosting methods.

Random forests are ensemble models composed of many decision trees. Each tree is trained on a subset of observations and predictors, and the final prediction is obtained by averaging or voting across trees. This structure allows random forests to model nonlinear relationships and interactions among SNP markers while reducing the risk of overfitting compared with a single decision tree [27]. In genomic prediction, random forests are useful because they can handle high-dimensional data and provide internal measures of variable importance. However, impurity-based importance measures from random forests can be biased toward correlated predictors, which is particularly relevant in SNP datasets where linkage disequilibrium creates strong correlations among neighboring markers [28]. Researchers should therefore use conditional permutation importance or SHAP values as alternatives when interpreting SNP effects.

Support vector machines (SVMs) are another class of supervised learning models that have been applied to high-dimensional biological data. Their main strength lies in the use of kernel functions, which allow complex relationships to be modeled by transforming the input data into a higher-dimensional feature space [29]. In genomic prediction, this can be useful when genotype–phenotype relationships are nonlinear. However, SVM performance depends strongly on kernel choice, parameter tuning, and data scaling. A practical limitation of SVMs in plant genomic prediction is their poor scalability to datasets with >10,000 samples, as the kernel matrix becomes computationally prohibitive.

Gradient boosting methods, including XGBoost, build models sequentially. Each new tree attempts to correct the errors made by previous trees, which often leads to strong predictive performance in structured tabular data [30]. In genomic prediction, boosting methods may be useful when the target trait is influenced by nonlinear effects or interactions among markers. Nevertheless, boosting algorithms can overfit if tuning parameters, training-set size, and cross-validation procedures are not carefully controlled.

Overall, classical ML models are valuable alternatives or complements to linear genomic prediction models. They may improve prediction when the relationship between markers and traits is nonlinear or when interactions contribute meaningfully to trait variation. At the same time, these models do not automatically provide biological insight. A model can predict accurately while still leaving unclear which genomic regions, genes, or marker combinations are driving the prediction.

2.2. Deep Learning Approaches

Deep learning (DL) is a subset of machine learning based on artificial neural networks with multiple layers. These models can learn hierarchical representations of data: lower layers may capture simple patterns, while deeper layers combine them into more complex features [31]. In genomics, this is attractive because biological signals are often distributed across many genomic regions and may involve nonlinear interactions that are difficult to specify in advance.

Several deep learning architectures have been explored in genomic prediction. Multi-layer perceptrons can model complex relationships among SNP markers. Convolutional neural networks may be useful when local genomic structure, marker order, or sequence context is relevant. Recurrent neural networks and attention-based models have also been considered for genomic or sequence-based prediction, particularly when long-range dependencies or ordered genomic information is important [9,32].

In plant genomics, deep learning is promising but not universally superior to classical statistical or machine learning methods. Its performance depends on the number of training samples, marker density, trait architecture, population structure, and validation design. Many plant breeding datasets remain relatively small compared with the number of genomic features, which increases the risk of overfitting [9,12]. Deep learning models also require more computational resources, careful hyperparameter tuning, and transparent reporting of model architecture and training procedures [9]. For this reason, recent reviews emphasize that DL should not be treated as automatically better than simpler models; rather, its usefulness must be evaluated empirically for each crop, trait, dataset, and prediction scenario [9,26].

Another important limitation of deep learning is interpretability. Neural networks may identify predictive patterns that improve accuracy, but it is often difficult to determine which SNPs, genes, genomic intervals, or molecular processes contribute most strongly to the prediction. This limitation is particularly relevant in plant breeding, where model outputs are often used not only to rank genotypes but also to guide biological interpretation, candidate-gene prioritization, and future crossing strategies.

2.3. Model Evaluation and Practical Use in Breeding

The usefulness of machine learning in genomic prediction depends not only on the algorithm itself, but also on how the model is trained, validated, and interpreted. In breeding applications, the most important question is not whether a model performs well on the training data, but whether it can predict accurately in relevant future scenarios. These scenarios may involve new breeding lines, different years, different environments, related or unrelated populations, or genotype-by-environment interactions. Robust validation is therefore essential. Cross-validation is commonly used, but its design must reflect the intended breeding application. Random cross-validation may overestimate prediction accuracy when training and test sets contain closely related individuals. More demanding validation schemes, such as across-population, across-family, across-environment, or across-year prediction, may provide a more realistic estimate of model performance [2,33]. Recent tools and benchmarking studies emphasize that model comparison should include multiple traits, populations, and validation schemes, because no single model is optimal across all genomic prediction problems [5,13]. Table 1 summarizes common validation strategies for genomic prediction, including their appropriate use cases, main limitations, and suitability for XAI-based interpretation.

Table 1. Validation strategies for genomic prediction.

Validation Scheme	When to Use	Main Limitation	Recommended for XAI?	Reference
Random k-fold cross-validation (CV)	Initial model evaluation and preliminary screening of prediction performance	May overestimate predictive accuracy when closely related individuals are present in both training and testing sets	Yes, suitable for initial screening and exploratory analysis	[2,34]
Leave-one-family-out validation	Assessing model generalization across unrelated breeding families	Requires large and well-structured populations with clear pedigree information	Yes, highly recommended for breeding relevance and family-level transferability	[34,35]
Leave-one-environment-out validation	Evaluating prediction stability across locations or environmental conditions	May underestimate prediction accuracy within a single environment	Yes, recommended for assessing environmental robustness and stability	[2]
Leave-one-year-out validation	Testing temporal robustness and model consistency across growing seasons	Requires multi-year phenotypic and genomic datasets	Yes, recommended for evaluating temporal generalization	[2,33]
Across-population validation	Assessing transferability of models across genetically distinct populations	Requires diverse germplasm and strong population diversity	Yes, important for evaluating model generalizability and broad applicability	[34,35]
Leave-one-genotype-out validation	Evaluating prediction accuracy for completely unseen genotypes	Highly conservative and computationally demanding	Yes, recommended for stringent model validation and reliability assessment	[36,37]

Model choice should also be guided by the biological and practical context. Linear models may be sufficient for traits dominated by additive effects and well-represented training populations. Kernel methods, random forests, or boosting algorithms may be useful when nonlinear relationships are expected. Deep learning may be most appropriate when datasets are large, multimodal, or sequence-based. However, more complex models should be preferred only when they provide clear advantages in prediction, robustness, or interpretability.

2.4. Motivation for Explainability

Machine learning can improve genomic prediction, but prediction alone is not always sufficient. In plant breeding and functional genomics, researchers often need to understand whether model predictions are based on biologically meaningful signals or on artifacts such as population structure, relatedness, unbalanced datasets, or environment-specific correlations. This distinction is particularly important in SNP-based datasets, where many markers are correlated and where a highly predictive marker may not be causal but simply linked to a causal variant.

Explainable artificial intelligence (XAI) addresses this problem by providing methods for examining how models make predictions. In genomic prediction, XAI can help identify influential SNPs, genomic intervals, or molecular features; compare feature-importance patterns across models; and evaluate whether predictive signals are stable across populations, environments, or traits. Previous studies have suggested that interpretable machine learning should be used for hypothesis generation rather than confirmation, and that model-derived importance scores should be treated as candidates for further testing rather than as definitive biological evidence [38].

However, explainability should not be confused with causality. A feature that receives a high importance score may be predictive because it is correlated with a causal locus, because it captures population structure, or because it reflects a dataset-specific pattern. Therefore, XAI outputs should be interpreted as hypotheses rather than definitive biological conclusions. In plant genomic prediction, the most useful role of XAI is to connect predictive modeling with biological reasoning, reproducibility checks, and experimental validation.

Taken together, machine learning and explainable AI should be viewed as complementary components of modern genomic prediction. Machine learning provides flexible tools for modeling complex genotype–phenotype relationships, while XAI helps evaluate whether these predictions are interpretable, stable, and biologically meaningful. This combination is particularly important for plant breeding, where the long-term goal is not only to predict trait performance, but also to support more informed, transparent, and biologically grounded crop improvement. A key distinction often overlooked is between global interpretability (understanding the model’s overall behavior) and local interpretability (understanding a single prediction). Both are valuable but serve different purposes: global explanations support population-level discovery, while local explanations aid individual genotype evaluation.

3. Explainable Artificial Intelligence Methods

Building on the interpretability problem outlined in the Introduction, this section reviews the main explainable artificial intelligence (XAI) approaches used to examine machine learning models in plant genomic prediction. In this context, XAI methods are not treated as tools for proving biological causality, but as approaches for analyzing model behavior, prioritizing predictive genomic features, and generating hypotheses for further biological validation. This distinction is particularly important in plant genomic datasets, where feature-importance patterns may be influenced by linkage disequilibrium, correlated

SNPs, population structure, genotype-by-environment interactions, or dataset-specific biases [17,38,39].

The general workflow of XAI-supported genomic prediction is presented in Figure 1. Genomic data, typically represented by SNP markers, are first used as input for machine learning models, including classical algorithms and deep learning architectures. These models generate predictions for target plant traits, which can then be examined using XAI methods such as SHAP, LIME, attention mechanisms, permutation importance, tree-based feature importance, and gradient-based attribution. The resulting explanations can support the interpretation of important variants, candidate genes, functional regions, and trait-related pathways, thereby linking predictive modeling with biological interpretation and breeding decisions.

Explainable AI Framework for Plant Genomic Prediction and Breeding

From genomic data to interpretable insights and improved crops

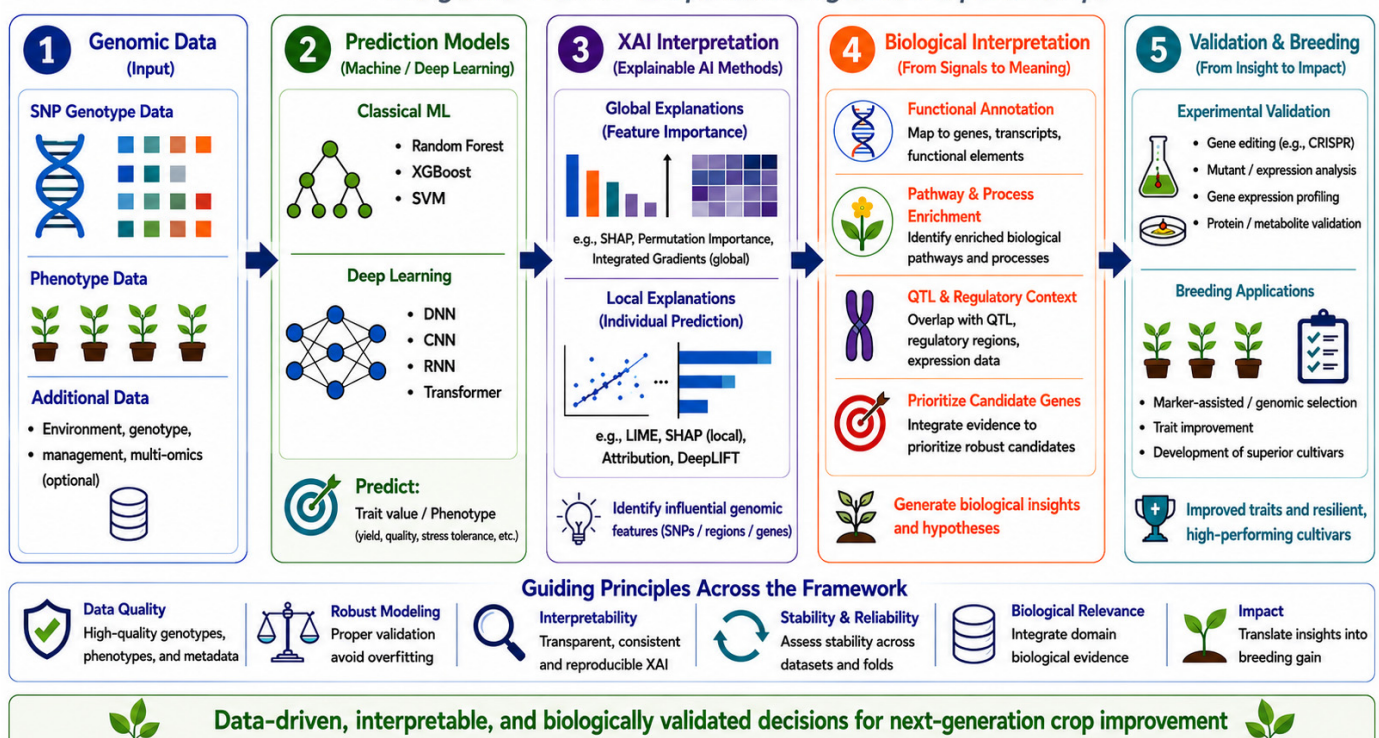


Figure 1. Explainable artificial intelligence (XAI) framework for plant genomic prediction and breeding. The figure illustrates the overall workflow from genomic and phenotypic data through machine learning and deep learning models to XAI-based interpretation, biological insight generation, experimental validation, and practical breeding applications. It highlights how explainable AI supports the identification of influential genomic features, facilitates biological interpretation, and enables transparent and biologically informed decision-making in modern plant breeding.

XAI methods can be classified according to the type of model they explain and the level at which the explanation is produced [40]. A first distinction separates model-specific and model-agnostic approaches. Model-specific methods are designed for particular model architectures, such as impurity-based or permutation-based feature importance in tree-based models, gradient-based attribution in neural networks, or attention weights in attention-based architectures. By contrast, model-agnostic methods, including SHAP and LIME, can be applied across different predictive models because they explain the relationship between input features and model outputs without depending on the internal structure of a given algorithm.

A second distinction concerns global and local explanations. Global explanations summarize model behavior across the dataset and can help identify genomic regions or omics features that repeatedly contribute to predictions. Local explanations focus on individual predictions, for example by showing which SNPs contributed to the predicted trait value of a particular genotype or breeding line. In plant genomic prediction, both levels are useful: global explanations support population-level interpretation and candidate-region prioritization, whereas local explanations can assist in the evaluation of individual genotypes. However, all XAI outputs require cautious interpretation, because feature importance may reflect predictive association rather than direct biological effect [17,38,39].

The following subsections discuss the main XAI approaches relevant to plant genomic prediction, including SHAP, LIME, attention-based mechanisms, permutation and tree-based feature importance, and gradient-based attribution methods.

3.1. SHAP: SHapley Additive exPlanations

SHAP is one of the most widely used XAI methods for interpreting machine learning models. It is based on Shapley values from cooperative game theory and estimates how much each input feature contributes to a model prediction [41]. In a genomic prediction setting, each SNP, genomic marker, or omics feature can be treated as an input feature, and SHAP values can be used to estimate how strongly that feature increases or decreases the predicted trait value.

A key advantage of SHAP is that it can provide both local and global interpretations. At the local level, SHAP can explain the prediction for an individual genotype by showing which markers contributed most strongly to its predicted phenotype. At the global level, SHAP values can be aggregated across individuals to rank genomic features according to their overall contribution to model predictions. This makes SHAP particularly useful in genomic prediction, where researchers often want to move from prediction accuracy toward the identification of candidate genomic regions.

In plant breeding, SHAP has recently been applied directly to genotype-to-phenotype prediction. The use of XAI in an almond germplasm collection demonstrated that SHAP analysis can highlight genomic regions associated with shelling fraction, including a region located in a gene potentially involved in seed development [17]. This type of application illustrates the value of SHAP as a hypothesis-generating tool: it can point researchers toward genomic regions that may deserve further biological investigation. However, such results should not be interpreted as proof of causality. A highly ranked SNP may be important for prediction because it is linked to a causal variant, captures population structure, or reflects correlations specific to the training dataset. For datasets with >50,000 SNPs, TreeSHAP (for tree-based models such as random forests or XGBoost) is recommended over KernelSHAP, as the latter scales poorly with feature dimension. When exact SHAP is computationally infeasible, a practical alternative is to apply SHAP to genomic windows (e.g., 10–50 kb bins) rather than individual SNPs, aggregating importance across biologically meaningful units.

SHAP is also useful because its visual outputs are relatively intuitive. Summary plots can show which features are most influential across the dataset, dependence plots can illustrate how feature values relate to prediction changes, and force plots can explain individual predictions. These visualizations can support communication between data scientists, geneticists, and breeders. This is important because genomic prediction is often used in interdisciplinary contexts, where model interpretation must be understandable not only to machine learning specialists but also to plant scientists and breeding practitioners.

Nevertheless, SHAP has limitations. First, it can be computationally expensive, particularly for very large genomic datasets with thousands or millions of markers. Although optimized versions such as TreeSHAP are efficient for tree-based models, model-agnostic

SHAP implementations may be difficult to scale to whole-genome datasets [42]. Second, SHAP explanations can be affected by correlated features. This is a major issue in genomic data, where linkage disequilibrium creates strong correlations among neighboring markers. When several SNPs carry similar information, SHAP may distribute importance among them in a way that is statistically reasonable but biologically difficult to interpret. Third, SHAP explanations depend on the trained model. If the model learns biased or unstable patterns, the explanation will reflect those patterns rather than correct them. A specific technical note: SHAP values can be affected by the choice of background dataset. In plant genomic prediction, the background should ideally represent the population distribution (e.g., 100–500 randomly selected individuals), rather than a single reference, to avoid biased importance estimates [43,44].

For these reasons, SHAP should be used as part of a broader interpretive workflow. In plant genomic prediction, SHAP results are most informative when combined with genomic annotation, linkage disequilibrium analysis, independent validation populations, pathway information, and, where possible, experimental validation.

3.2. LIME: Local Interpretable Model-Agnostic Explanations

LIME is another widely used model-agnostic XAI method. Unlike SHAP, which is based on Shapley values, LIME explains individual predictions by approximating a complex model locally with a simpler, interpretable surrogate model [45]. The method works by perturbing the input data around a selected instance and observing how the model prediction changes. A local linear model is then fitted to approximate the behavior of the original model in that small region of the feature space. In plant genomic prediction, LIME has been used less frequently than SHAP, primarily due to two limitations: (1) perturbation instability in SNP data, where random perturbations may generate biologically impossible genotypes that violate linkage disequilibrium patterns, and (2) lack of global explanations, as LIME only explains individual predictions. Researchers should therefore consider LIME as a complementary local explainer rather than a replacement for global methods like SHAP or permutation importance.

In genomic prediction, LIME can be used to explain why a specific genotype receives a given predicted trait value. For example, if a breeding line is predicted to have high drought tolerance or early flowering, LIME can identify which markers contributed most strongly to that individual prediction. This can be useful in case-specific analyses, especially when breeders are interested in understanding particular candidate lines.

The main strength of LIME is its simplicity and flexibility. It can be applied to almost any predictive model, including random forests, support vector machines, neural networks, and gradient boosting models. Its outputs are usually easy to understand because they are based on a simple surrogate explanation. This makes LIME attractive for communicating model behavior to non-specialist users.

However, LIME has important limitations in genomic applications. Because it relies on local perturbations, the quality of the explanation depends on how those perturbations are generated. In SNP data, random perturbation may create unrealistic genotypes that do not reflect biological constraints, allele frequencies, linkage disequilibrium, or population structure. As a result, LIME explanations may become unstable or difficult to interpret. Repeated runs may produce different feature rankings, especially when many markers are correlated or when the local neighborhood around an instance is poorly defined [44,46].

Another limitation is that LIME provides local rather than global explanations. It can help explain a single prediction, but it does not automatically indicate whether an SNP or genomic region is important across the whole population. Therefore, LIME may be more useful for inspecting individual predictions than for identifying candidate genomic

regions at the population level. In plant genomic prediction, LIME should therefore be used cautiously and preferably alongside more global approaches such as SHAP, permutation importance, or stability analyses across validation folds. A recommended practice is to run LIME multiple times (e.g., 10–20 repetitions) on the same instance and report the stability of feature rankings (e.g., standard deviation of importance scores). Rank correlations below 0.7 indicate unstable explanations.

3.3. Attention-Based Mechanisms in Deep Learning

Attention mechanisms were originally developed to improve the performance of sequence models by allowing them to assign different weights to different parts of the input [32]. In deep learning, attention can help a model focus on specific input regions that are more relevant for a given prediction. In genomics, this idea is attractive because biological signals may be distributed across distant genomic regions, regulatory elements, or sequence motifs.

In plant genomics, attention-based models may be useful for sequence-based prediction, multi-omics integration, and genotype-to-phenotype modeling. For example, attention layers can be used to assign weights to genomic regions, sequence windows, markers, transcripts, or environmental variables. In principle, these weights may help identify regions or features that the model uses when making predictions. This makes attention appealing not only as a modeling mechanism but also as a possible interpretive tool.

However, attention should not be treated as a straightforward explanation. Several studies have shown that attention weights do not always correspond directly to feature importance or causal relevance [47]. In plant genomics, attention mechanisms have been applied to genotype-by-environment prediction, although it has been explicitly noted that attention weights should not be interpreted as causal evidence [48]. Evidence from sequence-based models suggests that attention patterns learned from promoter sequences can recover known transcription factor binding motifs, indicating that attention may be informative when models are appropriately constrained and carefully validated [49]. Nevertheless, a model may assign high attention to a feature because it is useful for internal representation, but this does not necessarily mean that the feature is the main driver of the final prediction. Other studies have argued that attention can sometimes provide meaningful explanations, but only under specific modeling assumptions and with careful validation [50]. Therefore, attention-based explanations should be interpreted as model-derived indicators rather than direct biological evidence.

This distinction is especially important in plant genomic prediction. A genomic region with high attention may be biologically meaningful, but it may also reflect marker density, population structure, sequence composition, or correlations with other features. For this reason, attention-based interpretations should be compared with other explanation methods and biological evidence. When attention is used for interpretation, authors should report how attention weights were calculated, whether they are stable across model runs, and whether highlighted regions are supported by genomic annotation or independent evidence. To avoid overinterpretation, attention-based explanations should be compared with perturbation-based methods (e.g., leave-one-out or SHAP) to assess whether high-attention regions truly affect model outputs. If removing a high-attention SNP does not change predictions, the attention weight may reflect internal representation rather than causal contribution.

3.4. Permutation Importance and Tree-Based Feature Importance

In addition to SHAP, LIME, and attention-based approaches, several simpler feature-importance methods are commonly used in genomic prediction. One of the most widely

applied is permutation importance. This method measures how much model performance decreases when the values of a feature are randomly shuffled. If shuffling an SNP strongly reduces prediction accuracy, the SNP is considered important for the model.

Permutation importance is attractive because it is conceptually simple and model-agnostic. It can be applied to many types of models and provides an intuitive measure of how much a feature contributes to predictive performance. In plant genomics, it may be used to rank SNPs, markers, genes, or environmental covariates according to their contribution to trait prediction.

However, permutation importance has limitations in high-dimensional genomic data. When features are strongly correlated, shuffling one SNP may have little effect because neighboring SNPs contain similar information. In this case, the importance of a genomic region may be underestimated. Conversely, if a marker captures population structure or technical variation, it may appear important even if it is not biologically causal. These issues are particularly relevant in plant breeding populations, where linkage disequilibrium, relatedness, and stratification are common.

Tree-based models, such as random forests and gradient boosting, often provide built-in feature-importance scores. These scores are easy to obtain and can be useful for exploratory analysis. However, classical impurity-based importance measures can be biased toward variables with more categories, higher variability, or stronger correlations with other predictors [28]. In SNP datasets, this means that marker importance should not be interpreted without considering genomic structure and correlation among features.

A practical approach is to evaluate feature importance at the level of genomic windows or annotated regions rather than relying only on individual SNP rankings. Aggregating importance scores across neighboring markers may provide a more stable biological interpretation, especially when linkage disequilibrium makes it difficult to distinguish between nearby correlated variants.

3.5. Gradient-Based Attribution and Saliency Methods

Gradient-based attribution methods are mainly used with neural networks. These methods estimate how changes in input features affect model outputs by calculating gradients with respect to the input. In sequence-based genomics, saliency maps, integrated gradients, DeepLIFT, and related methods can be used to identify nucleotides, motifs, or sequence regions that contribute to a prediction [51,52].

In plant genomics, gradient-based methods may be particularly relevant for models that use DNA sequence, gene expression profiles, methylation data, or other omics inputs. For example, if a deep learning model predicts gene expression from promoter or flanking sequence, attribution methods can help identify sequence regions that influence the predicted expression level. This can support the discovery of regulatory motifs or candidate cis-regulatory elements.

However, these methods also require caution. Gradients can be noisy, sensitive to model architecture, and difficult to interpret when input features are correlated. Attribution scores may identify regions that influence the model output, but they do not prove that those regions are biologically causal. As with other XAI methods, the strongest conclusions arise when attribution patterns are stable, replicated across models or datasets, and supported by biological annotation or experimental evidence.

3.6. Summary

XAI methods provide an important interpretive layer for machine learning-based genomic prediction in plants. SHAP offers a theoretically grounded framework for local and global feature attribution, while LIME provides flexible local explanations for individual

predictions. Attention-based mechanisms and gradient-based attribution methods are particularly relevant for deep learning and sequence-based models, whereas permutation importance and tree-based importance scores remain useful for exploratory analyses.

Despite these advantages, XAI outputs must be interpreted cautiously. Feature importance does not imply causality, and explanations can be affected by correlated markers, linkage disequilibrium, population structure, model instability, and dataset bias. For this reason, XAI should be integrated with biological annotation, robust validation, stability analysis, and experimental follow-up. In plant genomic prediction, its strongest role is not to provide final biological proof, but to make predictive models more transparent and to generate testable hypotheses for breeding and functional genomics. To assist researchers in applying and validating explainable artificial intelligence in plant genomic prediction, Figure 2 presents a stepwise framework that integrates model development, XAI method selection, feature interpretation, stability assessment, biological annotation, and experimental validation.

Framework for Applying and Validating Explainable AI in Plant Genomics

A stepwise pathway from methodological decisions to biologically meaningful and validated insights

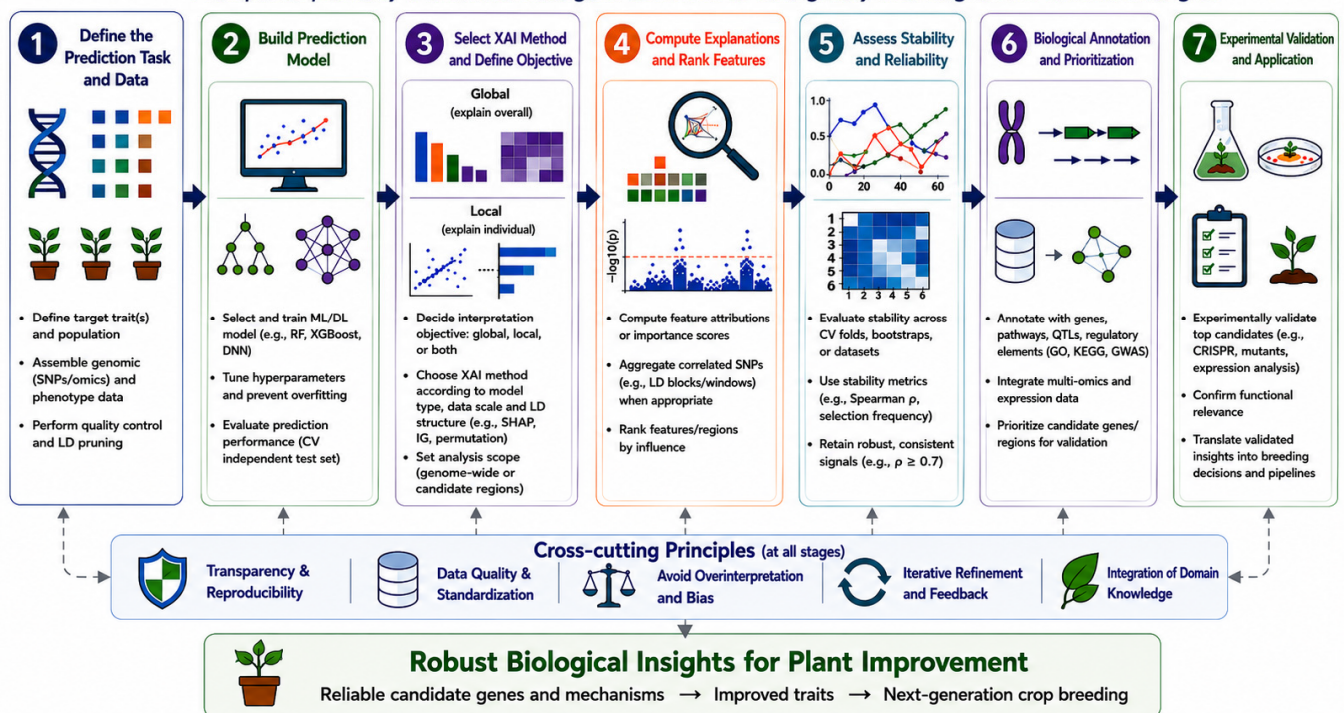


Figure 2. Recommended workflow for applying and validating explainable artificial intelligence (XAI) in plant genomic prediction. The framework presents the sequential steps for implementing XAI in genomic prediction studies, including data preparation, model development, selection of appropriate explanation methods, feature ranking, stability assessment, biological annotation, and experimental validation. Cross-cutting principles—including transparency, reproducibility, data quality, robustness, and integration of domain knowledge—are emphasized throughout the workflow to ensure reliable biological interpretation and support the identification of robust candidate genes for crop improvement.

4. Applications of Explainable AI in Plant Genomics

The integration of explainable artificial intelligence (XAI) into plant genomics has changed the role of genomic prediction from a largely predictive workflow into a more interpretable framework for biological discovery and breeding support. In conventional genomic selection, the main goal is often to maximize prediction accuracy for complex

traits using marker data [53]. However, many modern machine learning and deep learning models are difficult to interpret, even when they outperform simpler statistical models in relation to explainable AI in genetics [54]. This creates a practical and scientific problem in plant genomics, because researchers and breeders need to know not only which genotype is predicted to perform well, but also why the model made that prediction and whether the genomic signals it relies on are biologically credible. Recent reviews in genomics and crop science emphasize that explainability is becoming essential because plant traits are typically shaped by polygenic effects, nonlinear relationships, epistasis, regulatory variation, and genotype-by-environment interactions, all of which are difficult to represent and interpret with traditional linear methods alone [55,56].

In this context, XAI refers to methods that make the internal behavior of machine learning systems more transparent. These include feature-attribution methods such as SHAP and related importance-scoring approaches, local explanation methods that clarify why a specific genotype received a particular prediction, and intrinsically interpretable architectures that are designed to expose informative regions, embeddings, or attention patterns [54]. In plant genomics, these tools are especially valuable because biological usefulness depends on whether model explanations can be connected to loci, genes, pathways, regulatory regions, or molecular processes already known, or probably involved, in a target trait [57]. The literature increasingly treats XAI not as an improving addition to predictive modeling, but as a bridge between statistical learning and plant biology, enabling genomic models to generate experimentally testable hypotheses and more transparent breeding decisions [58].

4.1. Prediction of Complex Traits

One of the primary applications of XAI in plant genomics is the prediction of complex traits such as yield, flowering time, stress tolerance, nutrient-use efficiency, and disease resistance. These traits are difficult to predict because they are usually controlled by many loci of small effect, often interact across pathways, and are strongly modulated by environmental conditions [53]. Reviews in plant breeding and genomics consistently note that machine learning is attractive in this setting because it can model nonlinear genotype-to-phenotype relationships and leverage large genomic, phenotypic, and environmental datasets more flexibly than many classical approaches. At the same time, those same reviews emphasize that purely predictive gains are insufficient if the underlying model remains opaque, especially when the model is intended to support selection decisions or biological interpretation [56,59,60].

XAI improves complex-trait prediction by showing which genomic or molecular features drive model outputs. This is particularly important in plant datasets, where good predictive performance can sometimes arise from confounding structure, relatedness, or noncausal correlations [54]. By revealing feature-importance patterns, local contributions, and cross-feature dependencies, XAI enables researchers to assess whether a model is relying on biologically plausible information rather than a spurious signal. That interpretive layer is what makes machine learning more useful in plant genomics than a simple black-box predictor. In other words, XAI adds a biological auditing function to genomic prediction, making it easier to judge whether the model's behavior aligns with what is known about trait architecture and plant physiology [24].

A strong recent example comes from multi-omics prediction in *Arabidopsis thaliana*. Recent studies have shown that genomic, transcriptomic, and methylomic data can all support prediction of complex plant traits, and that explainability analyses can recover biologically relevant genes and molecular interactions underlying those predictions. These findings are important because they demonstrate that interpretability is not limited to

marker ranking; it can also clarify which omics layer carries useful signal for a given trait and how integrating layers improves prediction. The results suggest that multi-omics integration can reveal molecular mechanisms underlying trait variation, which is exactly the broader promise of XAI in plant genomics: combining predictive power with biological insight [55,57].

The value of this approach becomes even clearer when traits are developmentally dynamic or environmentally conditional. Recent work on dynamic genomic prediction showed that plant phenotypes change substantially across developmental stages and that prediction from markers can benefit from explicitly modeling temporal trajectories rather than static end-point traits alone. This has a direct implication for XAI: explanations in plant genomics should increasingly address not only which features matter, but also when they matter during development. Such time-aware interpretability is especially relevant for traits like growth, morphology, pigmentation, and stress response, where gene effects can vary substantially across plant ontogeny [55,61].

4.2. Identification of Important Genomic Regions

A central strength of XAI in plant genomics is its capacity to identify important genomic regions associated with phenotypic traits. In conventional genomic prediction, a model may achieve high accuracy without making it easy to determine which markers, haplotypes, or genomic intervals were most influential. XAI methods address this by assigning global or local importance to features, thereby allowing researchers to identify genomic segments that repeatedly contribute to prediction across samples or environments. This is especially valuable in plants because causal variation is often embedded within large linkage blocks, structural variation, or complex regulatory neighborhoods that cannot be fully characterized using isolated single-marker tests alone [54].

This ability to prioritize influential regions makes XAI highly complementary to genome-wide association studies. GWAS remains a powerful tool for detecting marker-trait associations, but its classical implementations are more sensitive to additive effects than to complex nonlinear interactions or distributed weak signals [62]. By contrast, interpretable machine learning can aggregate signal across many correlated features and can surface markers or sequence regions that matter jointly, even when no single SNP stands out strongly by itself. The recent crop-genomics literature has therefore argued for integrating GWAS with richer machine learning pipelines and pangenomic representations so that variant prioritization is informed both by statistical association and by model-based explanatory importance. A recent study clearly proposed combining pangenome analysis, GWAS, and interpretable machine learning to improve prioritization of agronomically relevant variants in crops, highlighting the practical role of XAI in post-GWAS discovery workflows [63]. XAI and GWAS identify different SNP sets in 30–50% of cases, suggesting high complementarity. GWAS is more sensitive to additive effects of individual SNPs, while XAI can capture interactions and nonlinear effects. A practical workflow is to use GWAS for initial screening, followed by XAI on the reduced marker set to identify interaction effects and prioritize candidates for validation.

XAI is also valuable because it can illuminate interactions among genomic features rather than simply ranking them individually. In biological systems, phenotype often emerges from combinations of loci, regulatory elements, and contextual dependencies. Recent work on interpretable deep learning in genomics shows that perturbation-based and attribution-based methods can recover interacting sequence features and regulatory dependencies from deep neural networks [54]. Although much of that work has been developed in broader genomics rather than plant-specific settings, the underlying principle applies directly to plant genomes, where epistasis, gene-regulatory interactions, and

structural complexity are common. For plant genomics, this means that XAI can help identify not just “important SNPs,” but clusters of features that likely participate in shared regulatory or developmental processes [54,64].

Sequence-based regulatory modeling provides a particularly strong example of this transition from simple feature importance to mechanistic interpretation. Interpretable deep learning models have been developed to predict gene expression from flanking genomic sequences in *Arabidopsis*, tomato, sorghum, and maize. These models not only achieved strong predictive performance but also identified informative regulatory regions, including untranslated regions and species-specific versus conserved sequence features. Such findings demonstrate that XAI can recover biologically meaningful genomic regions at the level of regulatory architecture, not just at the level of association markers, which substantially broadens its importance for plant genomics [49].

Another important advantage is that XAI can be used to assess the stability of genomic signals across populations and environments. Global feature-importance profiles can indicate whether a region remains influential across datasets or whether the model’s attention shifts unpredictably depending on training context [65]. This matters for plant breeding because stable explanatory regions are often more useful than highly variable ones, even when predictive accuracy is similar. More stable explanations can support prioritization of candidate loci for functional validation, marker development, and breeding deployment, especially when decisions must generalize across breeding populations and target environments [48,55].

4.3. Implications for Plant Breeding

The implications of XAI for plant breeding are substantial because breeding decisions increasingly rely on genomic prediction models. In genomic selection, models are used to rank genotypes, guide parent choice, and reduce the time and cost of field-based selection. However, breeders do not operate in a purely statistical setting. They must justify selection decisions, manage uncertainty, and often integrate model outputs with biological knowledge and breeding experience. As a result, interpretability is not just academically desirable; it has direct operational value in breeding pipelines [66]. Recent studies argue that AI for crop improvement will be most effective when it supports transparent decision-making, integrates diverse data types, and aligns predictive outputs with usable biological information for crop design and selection [55,60].

By exposing the genomic features driving a prediction, XAI can support marker-assisted selection and candidate-gene prioritization. Instead of selecting lines only because they received a high predicted breeding value, breeders can also ask whether those predictions are driven by loci or genomic patterns that make biological and practical sense. This can improve confidence in early-generation selection, support more targeted crossing strategies, and help distinguish robust signals from artifacts caused by population structure or data imbalance. In effect, XAI provides a rationale for selection decisions, which is particularly valuable when breeding programs are deciding among many lines with similar predicted performance [54,65].

An interpretable deep learning framework for crop genomic prediction has recently been developed and evaluated across five major crop species (maize, wheat, rice, soybean, cotton). Beyond achieving strong predictive performance, the framework demonstrated potential for identifying important genomic features and enabling genomic design [48]. In other words, XAI is beginning to support actionable breeding tasks such as line selection, trait dissection, and the design of more informed genomic selection strategies [65].

The breeding relevance of XAI becomes even greater in multi-environment and climate-responsive breeding. Traits such as yield stability, drought adaptation, and disease resis-

tance are strongly context-dependent, meaning that the importance of a genomic feature may vary across environments. Recent models such as GEFormer and MeNet were developed to improve prediction under genotype-by-environment complexity, showing gains in cross-environment prediction across major crops. From an XAI perspective, these studies reinforce an important conceptual point: explanations in breeding must increasingly be conditional explanations, revealing which features matter under which environmental scenarios, not simply which features matter on average. This is critical for climate-resilient breeding, where the goal is often to match genotypes to target environments rather than to identify universally superior lines [48,55].

XAI also helps build trust in model deployment. In applied breeding, models that cannot be interrogated are harder to validate, harder to debug, and harder to defend in collaborative decision-making environments. Transparent explanation can reveal whether a model is over-relying on a few correlated markers, ignoring known biology, or behaving inconsistently across germplasm groups [54]. This auditing function can improve reproducibility and facilitate adoption by breeders who need interpretable rather than purely technical outputs. For this reason, interpretability is increasingly viewed as a practical requirement for integrating machine learning into real breeding programs [54,55].

4.4. Biological Interpretation of XAI Outputs in Genomic Prediction

Before discussing specific applications of XAI to functional genomics, it is essential to clarify how XAI outputs should be interpreted biologically. The main value of XAI lies not in ranking features per se, but in helping researchers connect predictive signals to biological knowledge. The main value of XAI in plant genomic prediction lies not only in ranking features, but in helping researchers interpret predictive signals biologically. A list of important SNPs is rarely sufficient on its own. To become biologically useful, XAI outputs should be connected to genomic position, linkage disequilibrium structure, nearby genes, regulatory annotations, known pathways, trait biology, and, where possible, experimental validation.

For example, if SHAP identifies a set of SNPs as important for flowering time, the next step should be to examine whether these SNPs are located near genes involved in flowering regulation, photoperiod response, hormone signaling, or developmental timing. Similarly, if important markers are associated with drought tolerance, their interpretation should consider stress-response pathways, root architecture, stomatal regulation, osmotic adjustment, or other relevant biological processes.

This step is essential because XAI explanations are model explanations, not direct biological explanations. They describe how a trained model uses input features to generate predictions. They do not, by themselves, demonstrate that a feature causes a phenotype. In genomic data, this distinction is particularly important because predictive features may reflect linkage, population structure, or correlations among omics layers. Therefore, XAI should be treated as a method for prioritization and hypothesis generation rather than as a substitute for genetic mapping, functional genomics, or experimental validation.

Recent plant-focused studies illustrate this potential. Applications of SHAP to genotype-to-phenotype prediction have demonstrated its ability to support biological interpretation by highlighting genomic regions associated with important traits, including shelling fraction in almond [17]. More broadly, interpretable machine learning has been proposed as an important direction for genetics because it can help researchers connect complex predictive models to biological mechanisms while still acknowledging the limitations of model-derived explanations [38,39].

4.5. Integration with Functional Genomics

Beyond breeding and prediction, XAI has important applications in functional genomics because it can generate biologically grounded hypotheses about gene function, regulation, and molecular mechanisms. When an explainable model assigns consistently high importance to a set of SNPs, transcripts, methylation features, or sequence regions, those features can be mapped onto genes, promoters, enhancers, untranslated regions, or broader regulatory intervals [57]. This makes XAI a bridge between computational inference and experimental follow-up, allowing model outputs to inform gene expression studies, mutant analysis, CRISPR targeting, and pathway-level investigation [49,57].

A recent multi-omics study provides a good example of this bridge between predictive modeling and biological interpretation. By integrating genomic, transcriptomic, and methylomic data and then interpreting the resulting models, the researchers were able to identify genes and interactions with plausible mechanistic relevance to flowering-time biology. This is significant because it shows that XAI can contribute not only to ranking genotypes, but also to identifying candidate genes and regulatory relationships for experimental validation. In plant functional genomics, this can accelerate the transition from pattern detection to mechanism-oriented research by helping prioritize which features are most worth testing in downstream experiments [57].

The contribution of XAI to regulatory genomics is even more explicit in sequence-based models. Interpretable deep learning approaches have been shown to identify informative sequence regions associated with gene expression across multiple plant species. These findings suggest that deep models can be used not only to predict transcriptional output but also to recover regulatory determinants embedded in genomic sequences. That makes XAI especially valuable for dissecting cis-regulatory logic in plants, where promoter architecture, untranslated regions, and other proximal sequence features can strongly influence gene expression and downstream phenotypes. In this way, XAI contributes directly to functional genomics by pointing researchers toward candidate regulatory elements and sequence modules that merit experimental study [49,53].

Foundation models are likely to extend this role. AgroNT, a DNA-based foundational model trained on 48 plant species, was developed to predict molecular phenotypes across plant genomes [67]. Models of this kind may be powerful because they learn broad genomic representations that transfer across species and tasks, but they also intensify the need for interpretability. If such models are to contribute meaningfully to plant functional genomics, researchers must understand which latent representations correspond to regulatory activity, sequence semantics, or functional variation. XAI will therefore become increasingly important not only for explaining single predictions, but also for interpreting learned plant genome representations at scale [55,67].

At the same time, the literature repeatedly warns that importance is not equivalent to causality. Features identified by SHAP, saliency, attention, or other attribution methods may be predictive because they are correlated with causal factors rather than causal themselves. This problem is especially serious in genomics, where linkage disequilibrium, redundant markers, and correlated omics layers are common. A recent study found that interpretable AI can aid inference of causal molecular relationships, but only when explanation is situated in an explicit causal framework rather than treated as proof by visualization. For plant functional genomics, that means XAI outputs should usually be understood as high-value hypotheses for validation, not as definitive demonstrations of gene function or causal mechanisms [58,68].

This is why biologically informed model design is becoming so important. Recent studies have shown that biologically informed neural networks may produce explanations that are more faithful and scientifically useful because pathway structure, molecular hierar-

chy, or prior biological knowledge is embedded directly into the model architecture. For plant genomics, this is a promising direction because many agronomic traits are associated with known developmental, metabolic, hormonal, or stress-response pathways [58]. More broadly, recent crop-improvement research suggests that the value of XAI in functional genomics will grow as plant AI becomes more multimodal and more deeply integrated with omics technologies. The integration of omics data, phenotyping, genome editing and AI-enabled tools has been identified as a key direction for future crop improvement. Within this framework, explainability is likely to become increasingly important because it provides the interpretive layer needed to connect large-scale predictive systems to molecular mechanism and experimental design [55].

Overall, the application of XAI in plant genomics extends far beyond making machine learning models easier to inspect. It enhances complex-trait prediction by clarifying which genomic and molecular features drive model outputs, supports discovery of important genomic regions that may be missed or under-prioritized by conventional association approaches, improves the transparency and usefulness of genomic selection for plant breeding, and helps connect predictive modeling with functional genomics and experimental validation. Across the recent literature, the recurring message is that XAI becomes most valuable when it links predictive success to biologically interpretable signal and when explanations are stable enough to support scientific or breeding decisions [54,55,57,65].

The broader direction of the field is toward multimodal, biologically informed, and context-aware AI systems. Recent work on multi-omics prediction, interpretable crop genomic prediction, dynamic trait modeling, genotype-by-environment modeling, and plant foundation models all points to the same conclusion: explainability is becoming a core requirement rather than an optional add-on. In plant genomics, the true promise of XAI lies in its ability to transform accurate but opaque prediction into transparent, biologically useful, and experimentally actionable knowledge. That is why XAI is now increasingly viewed as a key component of both advanced plant genomics research and next-generation crop improvement [55,56,61,67].

4.6. XAI in SNP-Based and Multi-Omics Prediction

The effectiveness and interpretability of explainable artificial intelligence (XAI) methods depend strongly on the structure, dimensionality, and biological context of the underlying data. Although SNP markers remain the most widely used genomic features in plant genomic prediction, increasing availability of transcriptomic, metabolomic, proteomic, epigenomic, and pangenomic datasets has stimulated the development of multi-omics prediction frameworks. Consequently, the suitability of different XAI approaches should be evaluated in relation to the specific characteristics of SNP-based and multi-omics datasets.

In SNP-based genomic prediction, models are typically trained on tens of thousands to millions of genetic markers distributed across the genome. These datasets are characterized by high dimensionality, strong linkage disequilibrium (LD), and extensive correlations among neighboring markers. As a result, feature-attribution methods such as SHAP, permutation importance, and integrated gradients may identify clusters of correlated SNPs rather than individual causal variants. Feature-importance scores can therefore be influenced by marker redundancy, population structure, and genotype relatedness, requiring careful biological interpretation [19,69].

Despite these limitations, XAI methods can provide valuable insights into genomic regions contributing to model predictions. In breeding applications, feature-importance profiles may help prioritize candidate loci for subsequent validation through genome-wide association studies (GWASs), quantitative trait locus (QTL) analysis, gene expression studies, or functional experiments. Global explanation methods are particularly useful for

identifying genomic regions consistently associated with trait variation across populations and environments.

Multi-omics prediction presents a different set of opportunities and challenges. In addition to genomic markers, these datasets may include transcriptomic, metabolomic, proteomic, epigenomic, phenomic, and environmental variables. Such data capture multiple layers of biological regulation and may improve prediction accuracy for complex traits compared with genomic information alone [4]. However, multi-omics integration substantially increases model complexity because features originate from different biological scales, measurement platforms, and statistical distributions.

In multi-omics settings, XAI methods can be used not only to rank individual features but also to identify interactions between biological layers (Table 2). For example, explainability analyses may reveal relationships between genomic variants, gene expression profiles, metabolic pathways, and phenotypic outcomes. Deep learning interpretation methods, including attention mechanisms, integrated gradients, DeepLIFT, and attention-based transformer models, are particularly attractive for multi-omics applications because they can capture hierarchical and nonlinear relationships across multiple data types [40,49].

Table 2. Comparison of XAI methods for SNP-based and multi-omics datasets.

XAI Method	SNP-Based Datasets	Multi-Omics Datasets	Main Limitation	Reference
SHAP	Excellent for ranking influential SNPs and genomic regions	Effective across omics layers	Sensitive to correlated features	[14,17]
Permutation Importance	Simple and model-agnostic	Applicable to integrated datasets	Can underestimate correlated variables	[70]
LIME	Useful for local prediction explanations	Applicable but less stable in very high-dimensional data	Local instability	[15,20]
Integrated Gradients	Moderate utility in SNP data	Highly suitable for deep learning multi-omics models	Baseline dependence	[49,71]
DeepLIFT	Limited use in classical genomic prediction	Well-suited for neural networks and multi-omics integration	Reference dependence	[72]
Attention-based methods	Moderate interpretability	Particularly useful for transformers and multi-omics architectures	Attention does not necessarily imply biological importance	[48,73]

Nevertheless, interpretation of multi-omics models remains challenging. Feature-importance scores may be affected by batch effects, missing data, unequal feature representation, and strong correlations both within and between omics layers. Furthermore, highly ranked features do not necessarily indicate causal biological mechanisms. Consequently, XAI outputs should be considered as hypothesis-generating tools that require independent validation through functional genomics, gene editing, transcriptomic analyses, or experimental breeding studies.

Overall, SNP-based and multi-omics datasets present distinct interpretability challenges. SNP-based analyses primarily require disentangling linkage disequilibrium and marker redundancy, whereas multi-omics analyses require interpretation across multiple biological layers and data modalities. In both cases, robust validation and biological corroboration remain essential for translating explainable machine learning outputs into biologically meaningful knowledge and practical breeding applications.

5. Challenges and Limitations

Despite the growing use of explainable artificial intelligence (XAI) in plant genomics, several methodological and biological limitations remain unresolved. These limitations affect not only predictive performance, but also the reliability and biological validity of model explanations. This distinction is important because XAI methods explain the behavior of a trained model rather than the causal architecture of a trait [38,54,74].

5.1. Data Bias and Representativeness

One of the main challenges in genomic prediction is the limited representativeness of the underlying data. Plant genomic datasets are often generated from specific breeding populations, cultivars, environments, or experimental conditions, which may not capture the full genetic and environmental diversity relevant to broader breeding applications [17,59].

Such bias can lead to models that perform well within the original dataset but generalize poorly to new genotypes, populations, years, or environments. This problem is particularly relevant when XAI is used for biological interpretation, because explanation methods may highlight features that are important within a biased training set but not stable across broader germplasm. In SNP-based data, feature importance may also reflect allele-frequency differences, linkage disequilibrium, genetic background, or genotype-by-environment interactions rather than broadly relevant trait biology [17,38,54].

Addressing this limitation requires diverse training populations, multi-environment evaluation, careful control of population structure, and external validation. Where possible, XAI outputs should be compared across independent datasets to determine whether the same genomic regions or molecular features remain important under different biological and environmental conditions [17,59].

5.2. Model Instability and Reproducibility

Another important limitation is the instability of model predictions and feature-importance estimates. In high-dimensional genomic datasets, the number of markers often greatly exceeds the number of genotypes. Under these conditions, different algorithms, data partitions, hyperparameter settings, or random seeds may produce different predictions and different explanations [54,74].

This instability directly affects the reliability of XAI outputs. If a marker is ranked as important in one cross-validation fold but not in another, it is difficult to determine whether it reflects a robust biological signal or a model-specific artifact. This issue is especially relevant for flexible models such as random forests, gradient boosting methods, and neural networks, where small changes in training data or model configuration can influence feature-attribution patterns [54,65].

Improving reproducibility requires repeated cross-validation, independent test sets, external validation populations, sensitivity analyses, and comparison of multiple explanation methods. In plant genomic prediction, useful XAI results should not only identify important features, but also show whether these features remain important across model runs, populations, environments, and validation strategies [17,54].

5.3. Overinterpretation of XAI Results

A major risk in the use of XAI is the overinterpretation of model explanations. Methods such as SHAP, LIME, permutation importance, and attention-based mechanisms provide information about how a trained model uses input features, but they do not demonstrate biological causality [38,54,75].

In plant genomics, this distinction is essential because markers are often correlated. A highly ranked SNP may be predictive because it is linked to a causal variant, captures

population structure, or reflects a dataset-specific correlation. Similarly, attention weights should not automatically be interpreted as feature importance, because attention distributions do not always correspond to the actual contribution of input features to model outputs [47,50].

For this reason, XAI outputs should be treated as hypotheses for further investigation rather than final biological conclusions. Strong biological claims require additional evidence, such as co-localization with known genes, functional annotation, expression analysis, independent validation populations, mutant analysis, genome editing, or other experimental assays [17,38,54].

5.4. Limitations of Current XAI Methods

Current XAI methods also have important methodological limitations. SHAP provides a theoretically grounded framework for feature attribution, but exact computation can be expensive for large genomic datasets, and SHAP values may be difficult to interpret when predictors are highly correlated. LIME is flexible and model-agnostic, but its explanations can be sensitive to perturbation strategy, local sampling, and model choice [41,45,74].

Feature dependence is a particularly important issue in genomic data. Many explanation methods are easier to interpret when features are independent, but genomic markers are rarely independent because of linkage disequilibrium, haplotype structure, population history, and epistatic interactions. When correlated features contain similar information, importance may be distributed across multiple markers or assigned to one representative marker, which can complicate biological interpretation [28,74,76].

Tree-based feature-importance measures also require caution. Random forests and gradient boosting models often provide built-in importance scores, but these scores may be biased under certain data structures, particularly when predictors differ in scale, variability, or correlation patterns. This limitation has been demonstrated for random forest variable-importance measures and remains relevant for high-dimensional biological datasets [28,76,77].

Therefore, XAI results in plant genomics should be interpreted at multiple levels: individual markers, genomic windows, genes, pathways, and biological processes. Aggregating feature importance across genomic regions may often provide a more stable and biologically meaningful interpretation than focusing on single SNPs.

5.5. Representative Applications of XAI in Genotype-to-Phenotype Prediction

Although the application of explainable artificial intelligence in plant genomic prediction remains relatively recent, several studies have demonstrated its potential to improve the biological interpretation of genotype–phenotype relationships. Beyond early applications in almond germplasm, XAI approaches have increasingly been incorporated into genomic and multi-omics prediction workflows to identify biologically meaningful features and candidate genes [17].

In *Arabidopsis thaliana*, the integration of genomic, transcriptomic, and methylomic information has been used to predict complex developmental traits, while model interpretation methods have enabled the identification of genes and regulatory interactions associated with flowering-time regulation. These findings demonstrate that explainability analyses can recover biologically relevant signals from multi-omics prediction models and generate experimentally testable hypotheses [57].

Recent interpretable deep learning studies have also applied attribution methods to regulatory genomics. Interpretable neural network approaches have been used to identify cis-regulatory sequence determinants of gene expression across multiple plant species. These studies have shown that attribution-based methods can reveal sequence motifs and

regulatory features associated with transcriptional activity, extending explainability beyond marker-based prediction [49].

Furthermore, attention-based transformer architectures have recently emerged as promising tools for interpretable genomic prediction. Transformer-based frameworks have demonstrated that attention scores can highlight genomic regions contributing to predictive performance while maintaining competitive accuracy across multiple crop datasets [48]. Similarly, foundation-model approaches have shown that biologically meaningful representations can be learned from large-scale plant genomic datasets and subsequently interpreted to support downstream biological analyses [78].

Collectively, these studies suggest that XAI can facilitate biological interpretation across diverse data types, including SNP markers, regulatory sequences, transcriptomic profiles, and integrated multi-omics datasets [17,48,49,57,78]. However, important limitations remain. Most studies rely on feature-attribution methods that identify predictive signals rather than causal biological mechanisms. Explanations may also be affected by linkage disequilibrium, correlated markers, feature redundancy, and dataset-specific biases. Consequently, XAI outputs should generally be viewed as hypothesis-generating tools that require independent validation through genetic mapping, expression analyses, mutant studies, or genome-editing experiments [38,79,80].

5.6. Data Constraints in Plant Genomics

Plant genomic datasets often combine a very large number of predictors with a relatively small number of observations. This “large p, small n” structure increases the risk of overfitting and can reduce the stability of both predictions and explanations. Missing genotypes, uneven marker density, imputation uncertainty, population imbalance, and variable sequencing or genotyping quality further complicate model training and interpretation [54,76]. Phenotypic data introduce additional challenges. Many agronomic traits are affected by measurement error, developmental stage, management conditions, and environmental variation. If phenotype data are noisy or inconsistently collected, even a well-performing model may learn patterns that do not generalize beyond the original experiment. This is particularly relevant for traits such as yield, stress tolerance, flowering time, and disease resistance, where genotype-by-environment interactions can be substantial [17,59].

Multi-environment and multi-omics datasets may help address some of these limitations by providing richer biological context. However, they also introduce new challenges, including missing data, batch effects, heterogeneous data scales, cross-layer correlations, and increased model complexity. XAI methods applied to such datasets must therefore distinguish biologically meaningful cross-layer signals from artifacts introduced by preprocessing, dataset structure, or model design [54,65].

XAI provides valuable tools for interpreting machine learning models in plant genomics, but its outputs must be evaluated critically. The main limitations concern data bias, limited representativeness, model instability, correlated genomic features, methodological assumptions of XAI techniques, and the risk of treating feature importance as biological causality [38,54,74].

To improve reliability, future studies should combine robust validation, diverse training populations, stability analysis, independent test sets, and biological validation. XAI-based insights are most useful when they are reproducible across models and datasets and when they can be connected to genomic annotation, known trait biology, or experimental evidence. For a quick reference comparing the properties, advantages, and limitations of each XAI method discussed in this review, see Table 3.

Table 3. Summary of explainable artificial intelligence (XAI) methods for plant genomic prediction.

XAI Method	Explanation Type	Model Requirement	Computational Cost	Stability Under Linkage Disequilibrium (LD)	Main Limitation
SHAP (KernelSHAP)	Global and local explanations	Model-agnostic	Very high ($O(2^M)$)	Low	Computationally expensive and scales poorly with large SNP datasets
SHAP (TreeSHAP)	Global and local explanations	Tree-based models only	Moderate ($O(M \log M)$)	Moderate	Feature importance may be influenced by tree structure and correlated splits
LIME	Local explanations only	Model-agnostic	Moderate	Low	Sensitive to perturbation strategy and unstable in high-dimensional SNP data
Attention mechanisms	Global weighting patterns	Neural networks and transformers	Low	Moderate	Attention weights do not necessarily represent biological importance
Permutation importance	Global explanations	Model-agnostic	Moderate	High	Performance can be biased by correlated genomic markers
Tree-based feature importance	Global explanations	Tree-based models only	Low	Low	Subject to impurity and selection bias toward highly variable features
Integrated Gradients	Local explanations	Neural network models only	Moderate	Moderate	Attribution may be affected by gradient saturation and noise
DeepLIFT	Local explanations	Neural network models only	Moderate	Moderate	Results depend strongly on the selected reference baseline

6. Future Perspectives

The future development of explainable artificial intelligence (XAI) in plant genomic prediction will depend on whether these methods can move from exploratory model interpretation toward reproducible and biologically useful workflows. Current studies show that XAI can help identify predictive genomic or molecular features, but its practical value depends on the stability of explanations, external validation, biological interpretability, and usability in breeding contexts [17,38,54]. In plant science more broadly, machine learning is increasingly used to integrate genomic, phenotypic, and environmental data, but the

reliability of such models still depends strongly on data quality, validation design, and biological context [59].

Recent plant-focused examples illustrate both the promise and the limitations of this direction. Applications of XAI to genotype-to-phenotype prediction have demonstrated that SHAP values can identify genomic regions associated with important traits, including shelling fraction in almond germplasm [17]. Similarly, the integration of genomic, transcriptomic, and methylomic data has enabled the prediction of complex traits in *Arabidopsis thaliana*, while model interpretation has facilitated the identification of genes and molecular interactions relevant to flowering time [57]. These studies suggest that future work should focus not only on prediction accuracy, but also on interpretability, validation, scalability, and biological usefulness [17,57,65].

6.1. Integration of XAI into Breeding Pipelines

One of the most important future directions is the integration of XAI into operational plant breeding pipelines. Genomic prediction models are already used to support selection decisions, but their practical adoption may be limited when model outputs are difficult to interpret or communicate to breeders. In breeding programs, a predicted value is useful only when it can be evaluated together with field performance, biological knowledge, target environments, and the breeder's understanding of the germplasm [13,24,59].

XAI can support this process by helping breeders identify which genomic features contribute to a prediction and whether these features are consistent with known trait biology. For example, feature-attribution results may help prioritize markers for marker-assisted selection, identify candidate regions for further validation, or support crossing decisions when several lines show similar predicted performance [17,38]. However, these explanations should be used as decision-support information rather than as direct evidence of causality, because a predictive SNP may reflect linkage disequilibrium, population structure, or correlation with a causal variant rather than a direct biological effect [38,54].

For XAI to become useful in routine breeding, future studies should develop workflows that are accessible to plant breeders and not only to machine learning specialists. Recent work has introduced a graphical toolkit that allows users to compare statistical and machine learning models for genomic selection through a point-and-click interface [13]. Although these tools are not specifically designed for XAI, they illustrate the type of user-oriented infrastructure that will also be needed for explainable genomic prediction workflows [13].

6.2. Advances in Explainability for Deep Learning

Deep learning models are increasingly applied in genomics because they can represent nonlinear relationships, sequence patterns, and interactions among features. However, their interpretability remains a major limitation, especially when predictions are used to support biological inference. In genomics, explanation methods can help identify sequence motifs, regulatory regions, genes, or molecular interactions that contribute to model outputs, but these explanations depend on model architecture, data structure, and the assumptions of the explanation method [53,54].

Future research should therefore focus on explanation methods that are better adapted to genomic and sequence-based deep learning models. Methods such as gradient-based attribution, saliency maps, perturbation-based approaches, attention mechanisms, and interpretable neural network architectures may help identify genomic regions that influence model outputs. However, these methods require careful validation because model-derived importance does not automatically correspond to biological function or causal relevance [54,81]. It has been emphasized that XAI can help extract biological insight from

deep learning models, but also that interpretation must account for assumptions, limitations, and the specific structure of high-throughput biological datasets [54].

Recent plant genome foundation models make this issue even more important. DNA-based foundation models trained on genomes from multiple plant species have been developed for applications such as regulatory annotation, promoter and terminator strength prediction, tissue-specific expression prediction, and functional variant prioritization [67]. Models of this scale may support transfer across species and prediction tasks, but they also intensify the need for robust interpretability methods that explain what learned representations capture and how they relate to regulatory function or phenotypic variation [67].

6.3. Multimodal and Multi-Omics Integration

Another key future direction is the integration of genomic, transcriptomic, methylomic, metabolomic, phenotypic, and environmental data. Complex plant traits are rarely determined by DNA sequence variation alone. They often arise from interactions among genetic variation, gene expression, epigenetic regulation, development, management conditions, and environmental stress. Multimodal and multi-omics models may therefore provide a more complete representation of genotype–phenotype relationships than models based on SNP markers alone [57,59].

XAI can play an important role in these models by helping identify which data layers contribute to prediction and how different biological signals interact. The integration of genomic, transcriptomic, and methylomic data has been shown to support prediction of complex traits in *Arabidopsis thaliana* [57]. Furthermore, model interpretation has enabled the identification of genes and molecular interactions relevant to flowering-time regulation, including experimentally validated genes. These findings suggest that XAI applied to multi-omics models can support biological discovery as well as prediction [57].

However, multi-omics integration also introduces methodological challenges. These include missing data, batch effects, heterogeneous data scales, cross-layer correlations, differences in sampling time, and increased model complexity. Future work should therefore develop interpretable multimodal frameworks that can distinguish robust biological signals from artifacts introduced by preprocessing, data imbalance, or model architecture. In this context, XAI should be combined with biological annotation, pathway information, independent validation, and, where possible, experimental follow-up [38,54,57].

6.4. Addressing Bias and Improving Generalization

Bias and limited generalization remain central challenges in genomic prediction. Many plant genomic datasets are derived from specific breeding populations, environments, or experimental designs. A model trained on such data may perform well within the original dataset but generalize poorly to other populations, years, or target environments. This problem is not solved by XAI; in some cases, explanations may simply reveal biases already present in the training data [17,38,59].

Future research should therefore focus on improving both predictive generalization and explanation robustness. This may include the use of diverse training populations, multi-environment trials, external validation sets, uncertainty quantification, and bias-aware learning strategies. In addition, XAI methods could be used to audit models by identifying whether predictions rely on biologically plausible features or on confounding signals such as population structure, relatedness, or dataset-specific marker patterns [17,54].

This is especially important for plant breeding, where models are often expected to support decisions beyond the conditions under which they were trained. Explanations should therefore be evaluated not only within a single dataset, but also across populations, environments, traits, and validation schemes. Only stable and reproducible explana-

tions should be considered strong candidates for biological interpretation or breeding application [38,54,81].

6.5. Standardization and Reproducibility

A major limitation in the current use of XAI in plant genomics is the lack of standardized practices for model training, explanation, interpretation, and reporting. Studies often differ in how they split data, tune models, calculate feature importance, aggregate markers, and validate explanations. This makes it difficult to compare results across studies and to determine whether identified features are robust or dataset-specific [13,54,81].

Future research should establish clearer reporting standards for XAI-based genomic prediction. At minimum, studies should report the model architecture or algorithm, data preprocessing steps, validation design, hyperparameter tuning strategy, explanation method, number of repeated runs, stability of feature importance, and biological interpretation procedure. Where possible, authors should also provide code, data, trained models, and reproducible workflows [13,81].

Standardized benchmarks will be particularly important. The AgroNT study introduced the Plants Genomic Benchmark as a resource for evaluating deep learning-based plant genomic models, which may support more comparable model development in the future [67]. Similar benchmarking efforts are needed specifically for XAI in plant genomic prediction, including datasets with known or partially validated biological signals, a multi-environment structure, and reproducible evaluation protocols [13,67].

6.6. Bridging Computational Predictions and Experimental Validation

The long-term value of XAI in plant genomics depends on whether model explanations can guide experimental biology. A feature identified by SHAP, LIME, attention, or gradient-based attribution should generally be interpreted as a candidate for further study, not as a confirmed biological mechanism. Therefore, future research should strengthen the link between computational interpretation and experimental validation [38,54,57].

Potential validation strategies include co-localization with known genes or QTLs, functional annotation, pathway enrichment, expression analysis, mutant screening, genome editing, and evaluation in independent populations. For example, if an XAI method identifies markers near genes associated with flowering time, these candidates should be compared with known flowering-time regulators and, where possible, tested experimentally. Recent studies have demonstrated the value of this approach by combining multi-omics prediction with experimental validation of model-derived candidate genes, thereby strengthening the biological interpretation of the computational results [57].

The increasing availability of genome-editing technologies provides new opportunities to experimentally validate candidate genes and genomic regions prioritized by XAI methods. In particular, CRISPR/Cas-based approaches allow targeted modification of genes, promoters, or regulatory elements identified through feature-attribution analyses. Candidate loci consistently highlighted by SHAP, integrated gradients, attention-based methods, or other XAI techniques may therefore serve as promising targets for functional validation. Such experiments can help determine whether highly ranked genomic features directly contribute to trait variation or merely reflect linkage disequilibrium, population structure, or correlated markers. The integration of explainable machine learning with genome editing has the potential to accelerate the translation of computational predictions into biologically validated mechanisms and practical breeding applications [82–84].

This connection between prediction and validation is important because XAI explanations can be affected by linkage disequilibrium, correlated markers, population structure, or model instability. Without validation, there is a risk that feature-importance outputs will

be overinterpreted. Future studies should therefore treat XAI as part of a broader discovery workflow: prediction first, interpretation second, and biological validation before strong mechanistic conclusions are drawn [17,38,54].

Future progress in XAI for plant genomic prediction will depend on the development of methods that are accurate, interpretable, robust, and useful in real breeding and biological contexts. Key priorities include integrating XAI into breeding pipelines, improving explainability for deep learning and foundation models, developing interpretable multi-omics frameworks, addressing bias and generalization, standardizing reporting practices, and linking model explanations with experimental validation [54,57,59,67].

The most promising role of XAI is not to replace biological experiments or breeder expertise, but to make predictive models more transparent and to help prioritize hypotheses for further study. When combined with robust validation and biological evidence, XAI can contribute to more reproducible, interpretable, and biologically informative genomic prediction in plant breeding and functional genomics [17,38,57].

6.7. Towards Causal Inference with XAI

A growing research direction at the intersection of machine learning and plant genomics is the integration of explainable artificial intelligence (XAI) with causal discovery algorithms, aiming to move beyond predictive association toward more mechanistically informed inference. While SHAP values and other feature-attribution methods can effectively highlight which genomic markers are most influential for model predictions, they do not and cannot imply causation on their own [38,41]. Nevertheless, XAI can serve as a powerful feature selection step prior to applying more rigorous causal structure learning methods such as the PC algorithm, ICALiNGAM, or causal forests [85–87]. A promising hybrid workflow involves four sequential steps: first, using SHAP to identify a reduced set of candidate SNPs (e.g., the top 100 most important markers); second, applying linkage disequilibrium pruning to remove redundant markers that are highly correlated; third, running causal discovery algorithms to infer potential directions of effects and causal networks among the remaining variants; and fourth, validating the most promising causal relationships using orthogonal approaches such as Mendelian randomization, population segregation analysis, or direct genome editing via CRISPR/Cas systems. Genome editing provides an especially valuable framework for testing whether genomic regions prioritized by XAI correspond to true causal determinants of phenotypic variation. [88,89].

However, this workflow remains particularly challenging in plant genomics due to typically small sample sizes, complex population structures, and extensive linkage disequilibrium that can obscure true causal signals [24,59]. It has been formally demonstrated that Shapley values alone cannot identify causal features without additional structural assumptions or interventional data [75]. Therefore, any XAI-based causal claim must always be accompanied by rigorous experimental validation [68]. For plant breeding applications, causal XAI is especially valuable for identifying high-priority targets for gene editing: an SNP that is both highly predictive and plausibly causal—for instance, a missense variant located within a known metabolic or developmental pathway represents a far stronger candidate for CRISPR validation than a predictive SNP residing in a non-coding region with no functional annotation or evident biological context [55,58]. Ultimately, integrating XAI with causal discovery offers a promising but still nascent pathway toward more mechanism-aware genomic prediction and breeding decisions [48,65].

7. Conclusions

Machine learning has become an important approach for genomic prediction in plants, particularly for modeling complex genotype–phenotype relationships that are difficult

to capture using traditional linear methods alone. However, the increasing use of high-performing machine learning and deep learning models also raises an important interpretability problem. In plant genomics and breeding, accurate prediction is valuable, but it is often insufficient. Researchers and breeders also need to understand which genomic features contribute to model outputs and whether these features are biologically plausible.

This review discussed the role of explainable artificial intelligence (XAI) in addressing this limitation. We summarized major machine learning approaches used in plant genomic prediction and reviewed key XAI methods, including SHAP, LIME, attention-based mechanisms, permutation importance, tree-based feature importance, and gradient-based attribution methods. These approaches can help identify influential SNPs, genomic intervals, genes, regulatory regions, or omics features that contribute to predicted traits. In this way, XAI can support the interpretation of predictive models and help generate hypotheses for further biological validation.

At the same time, XAI outputs require careful interpretation. Feature importance does not provide direct evidence of causality, and model explanations may be affected by linkage disequilibrium, correlated markers, population structure, dataset bias, model instability, and validation design. Therefore, XAI should be treated as a tool for model inspection, feature prioritization, and hypothesis generation rather than as a substitute for genetic mapping, functional genomics, or experimental validation.

Future progress in this field will depend on the development of robust, reproducible, and biologically informed workflows. In particular, further work is needed to improve explanation stability, integrate multi-omics and environmental data, standardize reporting practices, and connect computational predictions with experimental validation. If applied cautiously, XAI can strengthen the transparency and biological usefulness of genomic prediction models and contribute to more informed decision-making in plant breeding and functional genomics.

Author Contributions: Conceptualization, M.P.; writing—original draft: M.P., M.H.U.K. and A.G.; writing—review and editing, M.P., M.H.U.K. and A.G.; visualization, M.P.; supervision, M.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT by OpenAI, based on GPT-5.5, for language editing, improvement of text clarity and structure, and support in drafting graphical elements used to illustrate the conceptual workflow. The authors reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Meuwissen, T.H.; Hayes, B.J.; Goddard, M.E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **2001**, *157*, 1819–1829. [[CrossRef](#)] [[PubMed](#)]
2. Crossa, J.; Perez-Rodriguez, P.; Cuevas, J.; Montesinos-Lopez, O.; Jarquin, D.; de Los Campos, G.; Burgueno, J.; Gonzalez-Camacho, J.M.; Perez-Elizalde, S.; Beyene, Y.; et al. Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends Plant Sci.* **2017**, *22*, 961–975. [[CrossRef](#)] [[PubMed](#)]
3. Voss-Fels, K.P.; Cooper, M.; Hayes, B.J. Accelerating crop genetic gains with genomic selection. *Theor. Appl. Genet.* **2019**, *132*, 669–686. [[CrossRef](#)] [[PubMed](#)]

4. Varshney, R.K.; Bohra, A.; Yu, J.; Graner, A.; Zhang, Q.; Sorrells, M.E. Designing Future Crops: Genomics-Assisted Breeding Comes of Age. *Trends Plant Sci.* **2021**, *26*, 631–649. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Lee, A.M.J.; Foong, M.Y.M.; Song, B.K.; Chew, F.T. Genomic selection for crop improvement in fruits and vegetables: A systematic scoping review. *Mol. Breed.* **2024**, *44*, 60. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Mackay, T.F.; Stone, E.A.; Ayroles, J.F. The genetics of quantitative traits: Challenges and prospects. *Nat. Rev. Genet.* **2009**, *10*, 565–577. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Millet, E.J.; Kruijer, W.; Coupel-Ledru, A.; Alvarez Prado, S.; Cabrera-Bosquet, L.; Lacube, S.; Charcosset, A.; Welcker, C.; van Eeuwijk, F.; Tardieu, F. Genomic prediction of maize yield across European environmental conditions. *Nat. Genet.* **2019**, *51*, 952–956. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Cooper, M.; Messina, C.D.; Podlich, D.; Totir, L.R.; Baumgarten, A.; Hausmann, N.J.; Wright, D.; Graham, G. Predicting the future of plant breeding: Complementing empirical evaluation with genetic prediction. *Crop Pasture Sci.* **2014**, *65*, 311–336. [\[CrossRef\]](#)
9. Montesinos-Lopez, O.A.; Montesinos-Lopez, A.; Perez-Rodriguez, P.; Barron-Lopez, J.A.; Martini, J.W.R.; Fajardo-Flores, S.B.; Gaytan-Lugo, L.S.; Santana-Mancilla, P.C.; Crossa, J. A review of deep learning applications for genomic selection. *BMC Genom.* **2021**, *22*, 19. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Bellot, P.; de Los Campos, G.; Perez-Enciso, M. Can Deep Learning Improve Genomic Prediction of Complex Human Traits? *Genetics* **2018**, *210*, 809–819. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Washburn, J.D.; Mejia-Guerra, M.K.; Ramstein, G.; Kremling, K.A.; Valluru, R.; Buckler, E.S.; Wang, H. Evolutionarily informed deep learning methods for predicting relative transcript abundance from DNA sequence. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 5542–5549. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Azodi, C.B.; Bolger, E.; McCarren, A.; Roantree, M.; de Los Campos, G.; Shiu, S.H. Benchmarking Parametric and Machine Learning Models for Genomic Prediction of Complex Traits. *G3* **2019**, *9*, 3691–3702. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Yu, L.; Dai, Y.; Zhu, M.; Guo, L.; Ji, Y.; Si, H.; Cheng, L.; Zhao, T.; Zan, Y. ShinyGS—a graphical toolkit with a serial of genetic and machine learning models for genomic selection: Application, benchmarking, and recommendations. *Front. Plant Sci.* **2024**, *15*, 1480902. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Van Hilten, A.; Katz, S.; Saccenti, E.; Niessen, W.J.; Roshchupkin, G. V Designing interpretable deep learning applications for functional genomics: A quantitative analysis. *Brief. Bioinform.* **2024**, *25*, bbae449. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Ribeiro, M.T.; Singh, S.; Guestrin, C. “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. *arXiv* **2016**, arXiv:1602.04938.
16. Rudin, C.; Chen, Z.; Huang, H.; Semenova, L.; Zhong, C. Interpretable Machine Learning: Fundamental Principles and 10 Grand Challenges. *arXiv* **2021**, arXiv:2103.11251.
17. Novielli, P.; Romano, D.; Pavan, S.; Losciale, P.; Stellacci, A.M.; Diacono, D.; Bellotti, R.; Tangaro, S. Explainable artificial intelligence for genotype-to-phenotype prediction in plant breeding: A case study with a dataset from an almond germplasm collection. *Front. Plant Sci.* **2024**, *15*, 1434229. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Hooker, S.; Erhan, D.; Kindermans, P.J.; Kim, B. A Benchmark for Interpretability Methods in Deep Neural Networks. *arXiv* **2018**, arXiv:1806.10758.
19. Molnar, C. General Pitfalls of Model-Agnostic Interpretation Methods for Machine Learning Models. In *xxAI—Beyond Explainable AI*; Springer: Cham, Switzerland, 2022.
20. Arrieta, A.B.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; García, S.; Gil-López, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. *arXiv* **2019**, arXiv:1910.10045.
21. Carriero, A.; de Hond, A.; Cappiers, B.; Paulovich, F.; Abeln, S.; Moons, K.G.; van Smeden, M. Explainable AI in healthcare: To explain, to predict, or to describe? *Diagn. Progn. Res.* **2025**, *9*, 29. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Tao, Y.; Zhao, X.; Mace, E.; Henry, R.; Jordan, D. Exploring and Exploiting Pan-genomics for Crop Improvement. *Mol. Plant* **2019**, *12*, 156–169. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Jordan, M.I.; Mitchell, T.M. Machine learning: Trends, perspectives, and prospects. *Science* **2015**, *349*, 255–260. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Bayer, P.E.; Petereit, J.; Danilevich, M.F.; Anderson, R.; Batley, J.; Edwards, D. The application of pangenomics and machine learning in genomic selection in plants. *Plant Genome* **2021**, *14*, e20112. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Heslot, N.; Yang, H.P.; Sorrells, M.E.; Jannink, J.L. Genomic Selection in Plant Breeding: A Comparison of Models. *Crop Sci.* **2012**, *52*, 146–160. [\[CrossRef\]](#)
26. Yoosefzadeh-Najafabadi, M.; Xavier, A.; Eskandari, M.; Hesami, M. Machine learning after a decade: Is it still a missing keystone in genomic-based plant breeding? *Artif. Intell. Rev.* **2025**, *58*, 260. [\[CrossRef\]](#)
27. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
28. Strobl, C.; Boulesteix, A.L.; Zeileis, A.; Hothorn, T. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinform.* **2007**, *8*, 25. [\[CrossRef\]](#) [\[PubMed\]](#)

29. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
30. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794. [[CrossRef](#)]
31. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 434–444. [[CrossRef](#)] [[PubMed](#)]
32. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. [[CrossRef](#)]
33. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillera-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schröder, B.; Thuiller, W.; et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **2017**, *40*, 913–929. [[CrossRef](#)]
34. Werner, C.R.; Gaynor, R.C.; Gorjanc, G.; Hickey, J.M.; Kox, T.; Abbadi, A.; Leckband, G.; Snowdon, R.J.; Stahl, A. How Population Structure Impacts Genomic Selection Accuracy in Cross-Validation: Implications for Practical Breeding. *Front. Plant Sci.* **2020**, *11*, 592977. [[CrossRef](#)] [[PubMed](#)]
35. Isidro, J.; Jannink, J.L.; Akdemir, D.; Poland, J.; Heslot, N.; Sorrells, M.E. Training set optimization under population structure in genomic selection. *Theor. Appl. Genet.* **2015**, *128*, 145–158. [[CrossRef](#)] [[PubMed](#)]
36. Dekkers, J.C.M.; Su, H.; Cheng, J. Predicting the accuracy of genomic predictions. *Genet. Sel. Evol.* **2021**, *53*, 55. [[CrossRef](#)] [[PubMed](#)]
37. Ould Estaghirou, S.B.; Ogutu, J.O.; Schulz-Streeck, T. Evaluation of approaches for estimating the accuracy of genomic prediction in plant breeding. *BMC Genom.* **2013**, *14*, 860. [[CrossRef](#)] [[PubMed](#)]
38. Azodi, C.B.; Tang, J.; Shiu, S.H. Opening the Black Box: Interpretable Machine Learning for Geneticists. *Trends Genet.* **2020**, *36*, 442–455. [[CrossRef](#)] [[PubMed](#)]
39. Watson, D.S. Interpretable machine learning for genomics. *Hum. Genet.* **2022**, *141*, 1499–1513. [[CrossRef](#)]
40. Mieth, B.; Rozier, A.; Rodriguez, J.A.; Höhne, M.M.; Görnitz, N.; Müller, K.R. DeepCOMBI: Explainable artificial intelligence for the analysis and discovery in genome-wide association studies. *NAR Genom. Bioinform.* **2021**, *3*, lqab065. [[CrossRef](#)] [[PubMed](#)]
41. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4765–4774.
42. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.I. From Local Explanations to Global Understanding with Explainable AI for Trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [[CrossRef](#)] [[PubMed](#)]
43. Molnar, C. *Interpretable Machine Learning*, 2nd ed.; Lulu.com: Durham, NC, USA, 2024.
44. Zhao, Y.; Shao, J.; Asmann, Y.W. Assessment and optimization of explainable machine learning models applied to transcriptomic data. *Genom. Proteom. Bioinform.* **2022**, *20*, 899–911. [[CrossRef](#)] [[PubMed](#)]
45. Withnell, E.; Zhang, X.; Sun, K.; Guo, Y. XOMiVAE: An interpretable deep learning model for cancer classification using high-dimensional omics data. *Brief. Bioinform.* **2021**, *22*, bbab315. [[CrossRef](#)] [[PubMed](#)]
46. Mostafa, S.; Mondal, D.; Panjvani, K.; Kochian, L.; Stavness, I. Explainable deep learning in plant phenotyping. *Front. Artif. Intell.* **2023**, *6*, 1203546. [[CrossRef](#)] [[PubMed](#)]
47. Jain, S.; Wallace, B.C. Attention is not explanation. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*; Association for Computational Linguistics: Minneapolis, MN, USA, 2019; pp. 3543–3556. [[CrossRef](#)]
48. Yao, Z.; Yao, M.; Wang, C.; Li, K.; Guo, J.; Xiao, Y.; Yan, J.; Liu, J. GEFormer: A genotype-environment interaction-based genomic prediction method that integrates the gating multilayer perceptron and linear attention mechanisms. *Mol. Plant* **2025**, *18*, 527–549. [[CrossRef](#)] [[PubMed](#)]
49. Peleke, F.F.; Zumkeller, S.M.; Gultas, M.; Schmitt, A.; Szymanski, J. Deep learning the cis-regulatory code for gene expression in selected model plants. *Nat. Commun.* **2024**, *15*, 3488. [[CrossRef](#)] [[PubMed](#)]
50. Wiegrefe, S.; Pinter, Y. Attention is not not explanation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*; Association for Computational Linguistics: Hong Kong, China, 2019; pp. 11–20. [[CrossRef](#)]
51. Shrikumar, A.; Greenside, P.; Kundaje, A. Learning important features through propagating activation differences. In *Proceedings of the 34th International Conference on Machine Learning*; PMLR: Sydney, Australia, 2017; pp. 3145–3153.
52. Sundararajan, M.; Taly, A.; Yan, Q. Axiomatic attribution for deep networks. In *Proceedings of the 34th International Conference on Machine Learning*; PMLR: Sydney, Australia, 2017; pp. 3319–3328.
53. Eraslan, G.; Avsec, Z.; Gagneur, J.; Theis, F.J. Deep learning: New computational modelling techniques for genomics. *Nat. Rev. Genet.* **2019**, *20*, 389–403. [[CrossRef](#)] [[PubMed](#)]
54. Novakovsky, G.; Dexter, N.; Libbrecht, M.W.; Wasserman, W.W.; Mostafavi, S. Obtaining genetics insights from deep learning via explainable artificial intelligence. *Nat. Rev. Genet.* **2023**, *24*, 125–137. [[CrossRef](#)] [[PubMed](#)]
55. Li, G.; An, L.; Yang, W.; Yang, L.; Wei, T.; Shi, J.; Wang, J.; Doonan, J.H.; Xie, K.; Fernie, A.R.; et al. Integrated biotechnological and AI innovations for crop improvement. *Nature* **2025**, *643*, 925–937. [[CrossRef](#)] [[PubMed](#)]

56. Farooq, M.A.; Gao, S.; Hassan, M.A.; Huang, Z.; Rasheed, A.; Hearne, S.; Prasanna, B.; Li, X.; Li, H. Artificial intelligence in plant breeding. *Trends Genet.* **2024**, *40*, 891–908. [[CrossRef](#)] [[PubMed](#)]
57. Wang, P.; Lehti-Shiu, M.D.; Lotreck, S.; Segura, A.; Krysan, P.J.; Shiu, S.H. Prediction of plant complex traits via integration of multi-omics data. *Nat. Commun.* **2024**, *15*, 6856. [[CrossRef](#)] [[PubMed](#)]
58. Selby, D.A.; Sprang, M.; Ewald, J.; Vollmer, S.J. Beyond the black box with biologically informed neural networks. *Nat. Rev. Genet.* **2025**, *26*, 371–372. [[CrossRef](#)] [[PubMed](#)]
59. van Dijk, A.D.J.; Kootstra, G.; Kruijer, W.; de Ridder, D. Machine learning in plant science and plant breeding. *iScience* **2021**, *24*, 101890. [[CrossRef](#)] [[PubMed](#)]
60. Crossa, J.; Montesinos-Lopez, O.A.; Costa-Neto, G.; Vitale, P.; Martini, J.W.R.; Runcie, D.; Fritsche-Neto, R.; Montesinos-Lopez, A.; Perez-Rodriguez, P.; Gerard, G.; et al. Machine learning algorithms translate big data into predictive breeding accuracy. *Trends Plant Sci.* **2025**, *30*, 167–184. [[CrossRef](#)] [[PubMed](#)]
61. Hobby, D.; Tong, H.; Heuermann, M.; Mbebi, A.J.; Laitinen, R.A.E.; Dell’Acqua, M.; Altmann, T.; Nikoloski, Z. Predicting plant trait dynamics from genetic markers. *Nat. Plants* **2025**, *11*, 1018–1027. [[CrossRef](#)] [[PubMed](#)]
62. Ma, Y.; Li, X.; Ji, X.; Wang, C.; Zhang, D.; Zhai, T.; Wang, H.; Liu, P. DNAtwhisper: An Integrated Deep Learning Pyramidal Framework for Multi-Trait Genomic Prediction and Adaptive Marker Prioritisation. *Plant Biotechnol. J.* **2026**, *24*, 4005–4026. [[CrossRef](#)] [[PubMed](#)]
63. Wang, W.; Wu, T.; Fan, G.; Zhang, S.; Liu, S.; Jiang, S.; Cheng, Q.; Shang, M.; Xu, Y.; Zhang, W.; et al. Integrating pan-genome analysis, GWAS, and interpretable machine learning to prioritize trait-associated structural variations in *Setaria italica*. *Plant Commun.* **2026**, *7*, 101626. [[CrossRef](#)] [[PubMed](#)]
64. Toneyan, S.; Koo, P.K. Interpreting cis-regulatory interactions from large-scale deep neural networks. *Nat. Genet.* **2024**, *56*, 2517–2527. [[CrossRef](#)] [[PubMed](#)]
65. Wang, H.; Yan, S.; Wang, W.; Chen, Y.; Hong, J.; He, Q.; Diao, X.; Lin, Y.; Chen, Y.; Cao, Y.; et al. Cropformer: An interpretable deep learning framework for crop genomic prediction. *Plant Commun.* **2025**, *6*, 101223. [[CrossRef](#)] [[PubMed](#)]
66. Jaganathan, D.; Dutta, B.; Basu, S.; Bag, S.; Srinivasan, R.; Eybshitz, A.; Barchenger, D.W.; Caarls, L.; Simmons, A.M.; Nankar, A.N. Breeding vegetables for whitefly resistance: Past, present, and future in the AI era. *Front. Plant Sci.* **2025**, *16*, 1724403. [[CrossRef](#)] [[PubMed](#)]
67. 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.; et al. A foundational large language model for edible plant genomes. *Commun. Biol.* **2024**, *7*, 835. [[CrossRef](#)] [[PubMed](#)]
68. Dibaeinia, P.; Ojha, A.; Sinha, S. Interpretable AI for inference of causal molecular relationships from omics data. *Sci. Adv.* **2025**, *11*, eadk0837. [[CrossRef](#)] [[PubMed](#)]
69. Novielli, P.; Nazzicari, N.; Pavan, S.; Delvento, C.; Diacono, D.; Zoani, C.; Bellotti, R.; Tangaro, S. Leveraging Explainable Artificial Intelligence for Genotype-to-Phenotype Prediction: A Case Study in *Arabidopsis thaliana*. *Appl. Syst. Innov.* **2025**, *8*, 164. [[CrossRef](#)]
70. Altmann, A.; Tološi, L.; Sander, O.; Lengauer, T. Permutation importance: A corrected feature importance measure. *Bioinformatics* **2010**, *26*, 1340–1347. [[CrossRef](#)] [[PubMed](#)]
71. Toussaint, P.A.; Leiser, F.; Thiebes, S.; Schlesner, M.; Brors, B.; Sunyaev, A. Explainable Artificial Intelligence for Omics Data: A Systematic Mapping Study. *Brief. Bioinform.* **2024**, *25*, bbad453. [[CrossRef](#)] [[PubMed](#)]
72. van Hilten, A.; Kushner, S.A.; Kayser, M.; Ikram, M.A.; Adams, H.H.H.; Niessen, W.J.; Roshchupkin, G.V. GenNet Framework: Interpretable Deep Learning for Predicting Phenotypes from Genetic Data. *Commun. Biol.* **2021**, *4*, 1094. [[CrossRef](#)] [[PubMed](#)]
73. Eshkiki, H.; Tanhaei, F.; Caraffini, F.; Mora, B. A Survey of the Application of Explainable Artificial Intelligence in Biomedical Informatics. *Appl. Sci.* **2025**, *15*, 12934. [[CrossRef](#)]
74. Salih, A.M.; Raisi-Estabragh, Z.; Galazzo, I.B.; Radeva, P.; Petersen, S.E.; Lekadir, K.; Menegaz, G. A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME. *Adv. Intell. Syst.* **2024**, *7*, 2400304. [[CrossRef](#)]
75. Janzing, D.; Minorics, L.; Blöbaum, P. Feature relevance quantification in explainable AI: A causal problem. In *Proceedings of AISTATS*; PMLR: Palermo, Italy, 2020; Volume 108, pp. 2907–2916.
76. Strobl, C.; Boulesteix, A.L.; Kneib, T.; Augustin, T.; Zeileis, A. Conditional variable importance for random forests. *BMC Bioinform.* **2008**, *9*, 307. [[CrossRef](#)] [[PubMed](#)]
77. Loecher, M. Unbiased variable importance for random forests. *Commun. Stat.—Theory Methods* **2020**, *51*, 1413–1425. [[CrossRef](#)]
78. Li, C.; Zhang, Q.; Chen, H.; Lin, K.; Luo, C.; Yang, M.; Xu, W.; Yao, F.; Yan, J.; Yang, Q.Y.; et al. PlantGFM: A Genomic Foundation Model for Discovery and Creation of Plant Genes. *Adv. Sci.* **2026**, e75772. [[CrossRef](#)] [[PubMed](#)]
79. Ali, S.; Abuhmed, T.; El-Sappagh, S.; Muhammad, K.; Alonso-Moral, J.M.; Confalonieri, R.; Guidotti, R.; Del Ser, J.; Díaz-Rodríguez, N.; Herrera, F. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. *Inf. Fusion* **2023**, *99*, 101805. [[CrossRef](#)]

80. Wang, K.; Abid, M.A.; Rasheed, A.; Crossa, J.; Hearne, S.; Li, H. DNNGP, a deep neural network-based method for genomic prediction using multi-omics data in plants. *Mol. Plant* **2023**, *16*, 279–293. [[CrossRef](#)] [[PubMed](#)]
81. Roscher, R.; Bohn, B.; Duarte, M.F.; Garcke, J. Explainable Machine Learning for Scientific Insights and Discoveries. *IEEE Access* **2020**, *8*, 42200–42216. [[CrossRef](#)]
82. Chen, K.; Wang, Y.; Zhang, R.; Zhang, H.; Gao, C. CRISPR/Cas Genome Editing and Precision Plant Breeding in Agriculture. *Annu. Rev. Plant Biol.* **2019**, *70*, 667–697. [[CrossRef](#)] [[PubMed](#)]
83. Gao, C. Genome engineering for crop improvement and future agriculture. *Cell* **2021**, *184*, 1621–1635. [[CrossRef](#)] [[PubMed](#)]
84. Varshney, R.K.; Sinha, P.; Singh, V.K.; Kumar, A.; Zhang, Q.; Bennetzen, J.L. 5Gs for crop genetic improvement. *Curr. Opin. Plant Biol.* **2020**, *56*, 190–196. [[CrossRef](#)] [[PubMed](#)]
85. Spirtes, P.; Glymour, C.; Scheines, R. *Causation, Prediction, and Search*, 2nd ed.; MIT Press: Cambridge, MA, USA, 2000.
86. Shimizu, S.; Hoyer, P.O.; Hyvärinen, A.; Kerminen, A. A linear non-Gaussian acyclic model for causal discovery. *J. Mach. Learn. Res.* **2006**, *7*, 2003–2030.
87. Wager, S.; Athey, S. Estimation and inference of heterogeneous treatment effects using random forests. *J. Am. Stat. Assoc.* **2018**, *113*, 1228–1242. [[CrossRef](#)]
88. Smith, G.D.; Hemani, G. Mendelian randomization: Genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* **2014**, *23*, 89–98. [[CrossRef](#)]
89. Liu, Y.C.; Wang, T.; Dong, Q.; Li, H.; Song, C. CRISPR-based functional genomics in plants: Advances and opportunities. *Mol. Plant* **2020**, *13*, 1640–1655.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.