Journal of Experimental Botany, Vol. 00, No. 0 pp. 1–31, 2025
https://doi.org/10.1093/jxb/eraf498 Advance Access Publication 14 November 2025

REVIEW PAPER

Artificial intelligence in plant salt stress research: from predictive models to multi-omics integration

Javier Santos del Río[1,2], Alicia Talavera[1], Noé Fernández-Pozo[1], Francisco J. Veredas[3], and M. Gonzalo Claros[1,2,4,*]

[1] Institute for Mediterranean and Subtropical Horticulture 'La Mayora' (UMA-CSIC), Malaga 29010, Spain
[2] Department of Molecular Biology and Biochemistry, Universidad de Málaga, Malaga 29071, Spain
[3] Department of Computer Science and Programming Languages, Universidad de Málaga, Malaga 29071, Spain
[4] CIBER de Enfermedades Raras (CIBERER) U741, Malaga 29071, Spain

* Correspondence: claros@uma.es

Received 21 July 2025; Accepted 5 November 2025

Editor: Antonio Serrato, Estacion Experimental del Zaidin, Spain

Abstract

Salinity is a chronic environmental stressor causing irreversible damage to plants and resulting in significant economic losses. Early bioinformatics analyses on mono-omics data relying on predictive methods were highly effective in shedding light on the mechanisms of adaptation to salt stress. The incorporation of artificial intelligence has enabled analysis of multi-omics datasets combined with molecular, physiological, and morphological parameters relating to salt stress, and made it possible to perform high-throughput phenotyping using satellite snapshots and hyperspectral imaging to estimate soil salinization, predict salt stress in crops, and assess plant growth. Additionally, the arrival of transformers and the elaboration of large language models based on protein and nucleic acid sequences enabled identification of complex patterns underlying the 'language of life'. These generative models offer innovative hypotheses and experiments, particularly for understudied species or complex biological processes like salt stress tolerance. Protein language models also provided satisfactory results in identifying salt stress-related post-translational modifications. Predictive agro-climatic models are proving beneficial to the crop agriculture sector: they are expected to increase yields and reduce the time and costs involved in development or identification of commercially viable salt-tolerant cultivars. In conclusion, artificial intelligence is stimulating the discovery of novel facets of plant responses to salt stress, which is opening new frontiers in salinity research and contributing to previously unimaginable achievements.

Keywords: Artificial intelligence, bioinformatics, deep learning, high-throughput phenotyping, large language models, post-translational modification, salinization, salt stress.

Abbreviations: ABA, abscisic acid; AI, artificial intelligence; ANN, artificial neural network; AUC-ROC, area under the receiver operating characteristic curve; AUPRC, area under the precision–recall curve; BERT, bidirectional encoder representations from transformers; DL, deep learning; GABA, γ-aminobutyric acid; GEO, Gene Expression Omnibus; gLM, genomic language model; GPT, generative pre-trained transformer; GWAS, genome-wide association study; LLM, large language model; MCC, Matthews correlation coefficient; ML, machine learning; NLP, natural language processing; pLM, protein language model; PTM, post-translational modification; RAG, retrieval-augmented generation; RF, random forest; RMSE, root mean squared error; ROS, reactive oxygen species; SNP, single nucleotide polymorphism; SVM, support vector machine; T2T, telomere-to-telomere; UAV, unmanned aerial vehicles; XAI, explainable artificial intelligence.

Introduction

Salt stress, an important threat

Salinity, a major environmental stressor, severely affects plant morphology, physiology (flowering delay, less flowers, more flower abortions, less pollen production, etc), and biochemistry, and disrupts signalling pathways, reducing development, growth, agricultural production, and ecosystem stability (Atta *et al.*, 2023). Most crop species are glycophytes, that is they are highly vulnerable to salt stress. Unfortunately, plant saline stress is exacerbated by climate change factors (increased evaporation, reduced rainfall in arid and semi-arid regions) and poor management practices (irrigation with brackish water, inadequate drainage, and overuse of fertilizers) (Hassani *et al.*, 2021). Unlike other stresses, salinity is a chronic stress that can cause irreversible damage and entails important economic and food security issues (Muhammad *et al.*, 2024). Understanding salt tolerance mechanisms is crucial for developing methods to increase crop (N. Muhammad *et al.*, 2025) and forest (M. Zhang *et al.*, 2021) yields in salinized conditions.

Extremophyte responses

Some plants can tolerate high salinity by sequestering harmful ions inside their cells or blocking their entry. The extremophyte *Schrenkiella parvula* (formerly known as *Thellungiella parvula*), which is endemic to saline, drought, and cold habitats (Dassanayake *et al.*, 2011), can grow at 400 mM (Hajiboland *et al.*, 2018) by excluding Na^+ from leaves; the recretohalophyte *Limonium bicolor* contains a high-performance salt gland, controlled by a NAC transcription factor, that excretes Na^+ to avoid salt damage (Zhao *et al.*, 2023); the high-salt-tolerant mangrove *Avicennia officinalis* also has salt glands and an efficient salt filtration mechanism at the roots combined with crosstalk between Ca^{2+}, auxin, and ethylene signalling that seems to operate in an abscisic acid (ABA)-independent pathway (Krishnamurthy *et al.*, 2017).

However, most salt stress research has been focused on a few species, including *Arabidopsis thaliana* (Kumar *et al.*, 2023) and major cereals crops such as wheat, maize, and rice, for which soil salinization is a major threat (Zhang *et al.*, 2018; P. Hu *et al.*, 2021; Javid *et al.*, 2022; Kumar *et al.*, 2022).

Main transcription factors

Extensive research has revealed that plants respond to salt stress through a sophisticated signal transduction network that integrates various components, including protein kinases—such as mitogen-activated protein kinases (MAPKs) and cyclin-dependent protein kinases (CDPKs)—phospholipid-based signals, reactive oxygen species (ROS), calcium levels, phytohormones—including ABA, brassinosteroid, ethylene, gibberellin, salicylic acid, and jasmonic acid—and transcription factors (Zhang *et al.*, 2022; Xiao and Zhou, 2023). Transcription factors involved in salt stress responses are typically involved in other abiotic stresses, but specific salt-related effects have been identified (Bhoite *et al.*, 2025) such as for members of the WRKY, NAC, MYB, and bZIP families.

Members of the WRKY family are induced in response to salt stress to modulate growth, development, and root morphology to allocate more energy toward salt tolerance (Khoso *et al.*, 2022; Wagan *et al.*, 2024; L. Yang *et al.*, 2025a). Members of the NAC family play crucial roles in leaf senescence, osmotic regulation through an increase in proline, ROS scavenging, and phytohormone regulation (Han *et al.*, 2023; Xiong *et al.*, 2025). Multiple MYBs are involved in salt response and tolerance by maintaining ion and redox homoeostasis and leaf water potential, regulating alternative respiration, increasing leaf cuticles and ABA levels, and alleviating the degradation of soluble proteins (D. Zhang *et al.*, 2025). Several bZIPs can regulate ABA signalling pathway, chlorophyll stabilization, ROS homoeostasis, balancing the K^+/Na^+ ratio, modulating soluble sugar content, and influencing the expression of stress-related genes for salt response (Y. Yang *et al.*, 2025).

Main signalling pathways

Most of the signalling cascades affected by salt stress are required to reinforce the following biological processes for salt tolerance. (i) Osmotic adjustment through accumulation of specific osmolytes (e.g. proline, polyols such as mannitol or sorbitol, fructans, trehalose, glycine betaine, or flavonoids). (ii) Ion detoxification to avoid photosynthesis impairment, which involves sequestering Na^+ and Cl^- ions in root or leaf vacuoles and maintaining high K^+/Na^+ and Ca^{2+}/Na^+ ratios. (iii) Antioxidant defence (enzymatic and non-enzymatic) (H. Li *et al.*, 2023; Claros *et al.*, 2025; N. Muhammad *et al.*, 2025) since ROS elevation is detrimental to plants, causing oxidative damage to cellular biomolecules (Sachdev *et al.*, 2021). However, since ROS are also signalling molecules that regulate, among other processes, abiotic stress adaptation (Mittler *et al.*, 2022; Wang *et al.*, 2024), a delicate balance between signalling and damage must be maintained (Xu *et al.*, 2025) by means of stress- and tissue-specific isoforms of antioxidant enzymes (Huang *et al.*, 2019). (iv) Cell wall thickening, particularly in root cells, by deposition of cellulose, pectins, hemicelluloses, lignin, suberin, and waxes (M. Zhang *et al.*, 2021; Zhang *et al.*, 2022; Balasubramaniam *et al.*, 2023; Xiao and Zhou, 2023; H. Zhang *et al.*, 2025). Leaf thickness is also observed. (v) An increase of epitranscriptomic modifications in seedling and shoot mRNAs, mainly by introducing N^6-methyladenosine in transcripts involved in salt tolerance, plant growth, and stress response to enhance their RNA stability, has been demonstrated during stress adaptation (J. Hu et al., 2021; Zheng *et al.*, 2021; W. Wang *et al.*, 2022; Y. Wang *et al.*, 2022). (vi) Tolerance to salt stress also seems to involve specific post-translational modifications (PTMs) (Martí-Guillén *et al.*, 2022; L. Wei *et al.*, 2024;

Claros *et al.*, 2025), programmed cell death (Claros *et al.*, 2025), and autophagy (Claros *et al.*, 2025; Julian *et al.*, 2025).

Strategies to improve salt tolerance

In order to overcome the threat of soil salinization, a number of approaches have been adopted in research and field studies (Melino and Tester, 2023). One is the application of chemical additives to rehabilitate soil degradation, which is expensive and poses a risk for secondary salinization (Stigter *et al.*, 2006). Another approach involves the modification of microbiota to enhance salt tolerance (Kumar *et al.*, 2020; Muhammad *et al.*, 2024; M. Muhammad *et al.*, 2025; Zeng *et al.*, 2025) through the synthesis of biostimulants or the modification of plant gene expression (Claros *et al.*, 2025), or even the application of exogenous biomolecules to alter the internal levels of osmolytes that alleviate the salt stress (T. Liu *et al.*, 2025). Very interesting findings are provided by the induction of PTMs that are known to improve salt tolerance (Najar, 2024; Tibesigwa *et al.*, 2025) by means of chemical priming (Zulfiqar *et al.*, 2022). Finally, the considerable variation in salt tolerance within natural populations and between crops and their wild ancestor species encourages the selection of salt-tolerant varieties, cultivars, or rootstocks (Dag *et al.*, 2015; Díaz-Rueda *et al.*, 2020; Melino and Tester, 2023) because the use of advanced biotechnological tools to develop salt-tolerant plant varieties, although reporting promising results (Kotula *et al.*, 2020; Hualpa-Ramirez *et al.*, 2024), is subject to severe legal restrictions in many countries (Bruetschy, 2019; Ahmar *et al.*, 2024).

Artificial intelligence in salt stress research

The studies on salt stress have been greatly facilitated by the development of high-throughput omics techniques, which generate large amounts of data that require advanced computational methods for analysis and interpretation. In this context, different artificial intelligence (AI) methods have emerged as powerful tools that can extract meaningful patterns and insights from plant salt stress (Miolane, 2025). Hence, the present review is focused on clarifying what can be considered AI, how it has been incorporated to analyse and extract valuable knowledge about salt stress, and how it has opened up new avenues for predicting crop yields or alleviating salt stress.

Primer on AI for plant research

Every plant cell holds an overwhelming volume of biological information (thousands of genes, millions of molecular interactions, billions of DNA base pairs). Unlocking that information surpasses the capability of any human mind, requiring computers to grasp this information, where predictive approaches based on machine learning (ML; Fig. 1) have been widely used for the detection and molecular analysis of salt stress in plants (Dwivedi *et al.*, 2021). One of the main advantages of AI lies in its ability to analyse vast amounts of complex data, identify relevant and specific patterns, and generate predictive models. These models can then be adapted and improved as new data become available. Thanks to these capabilities, the study of complex traits, such as the response to salt stress, is now becoming more affordable.

However, we must ask, what does AI really mean? In 1955, John McCarthy gave the original definition of AI as 'the science and engineering of making intelligent machines'. More recently, AI has been defined as 'a system's ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation' (Kaplan and Haenlein, 2019). AI is nowadays a multidisciplinary field centred on the emulation of human cognitive abilities by means of different learning paradigms (Box 1; Fig. 1) implemented in methods (Box 1) and deep architectures (Box 2) that can be subsequently applied to different domains, such as computer vision and sequence modelling (Box 3) to perform specific tasks (Vrana and Singh, 2021; Khan *et al.*, 2022). The increasing complexity and capabilities of these algorithmic systems are steering such rapid development that some methods and architectures can be located in different paradigms depending on the report (Singla *et al.*, 2024), as occurs with transformers in Fig. 1. The following sections outline important AI concepts commonly used in AI-driven salt stress research.

Predictive and generative AI

AI systems can be divided into predictive or generative AI based on their resulting model. Predictive AI, sometimes referred to as narrow AI or discriminative AI, is designed to identify patterns and relationships within data—by means of ML methods, artificial neural networks (ANNs), or deep architectures—to perform informed classification, regression, or decision-making tasks (Yelmen and Jay, 2023; Miller, 2025). In contrast, generative AI can create new content—text, images, or audio—by learning the underlying patterns in training data (Bengesi *et al.*, 2024). Relying on deep ANN architectures, such as generative adversarial networks (GANs), diffusion models (Harshvardhan *et al.*, 2020; Guo *et al.*, 2024), and transformers (Box 2; Fig. 1; Vaswani *et al.*, 2017), it requires greater computational resources than predictive AI (Bandi *et al.*, 2023). In a few words, predictive AI makes analytical inferences while generative AI is creative.

Generative AI has enabled powerful new capabilities in natural language processing (NLP), computer vision, and the arts that resemble human-generated work (Harshvardhan *et al.*, 2020; Yelmen and Jay, 2023). The popularity of generative models such as ChatGPT, Gemini, or Claude explains why AI is often associated exclusively with generative AI. Unfortunately, generative AI presents some risks in science (Noorden, 2022): biased results and hallucinations—plausible-sounding but incorrect and inaccurate outputs—due to the

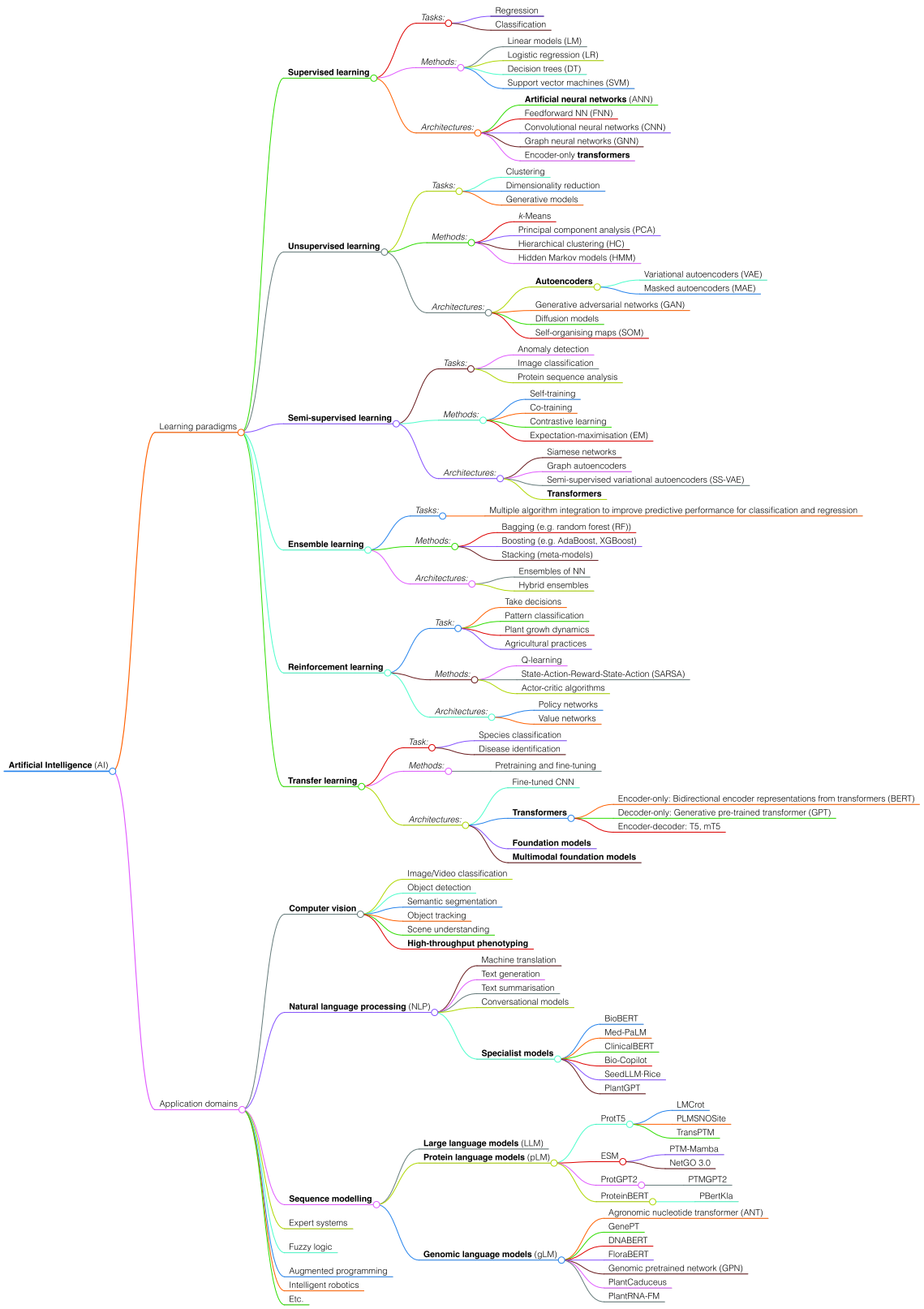

Fig. 1. Coherent hierarchy of AI learning paradigms and applications. Each paradigm contains the tasks in which it should be used, the main methods found in the literature (most of them considered ML algorithms; see Box 1), and the main deep architectures (Box 2) that are revolutionizing scientific research. Items in bold are explained in detail in the main text.

Box 1. ML workflow and learning paradigms

ML refers to the ability of computer systems to learn without being explicitly programmed. It occurs through the following distinct steps.

1. Feature engineering: the process of creating new input features (variables, such as transcripts, weight, and chlorophyll content) from raw data to improve the performance of ML models. It involves transforming, combining, or extracting information from existing data to better represent the underlying patterns relevant to the task at hand (Olaoye and Fajinmi, 2025, Preprint).
2. Feature selection: the process of identifying and retaining the most relevant features in a dataset while removing those that are redundant or irrelevant. This step helps reduce model complexity, mitigate overfitting, and enhance interpretability, often leading to more robust and generalizable models (Saeys *et al.*, 2007; Bermingham *et al.*, 2015; X. Zhang *et al.*, 2016).
3. Predictive modelling: the previous features are used to build a statistical model that can forecast future events or classify unseen data.
4. Model validation: this involves assessing the performance and generalizability of an ML model using both internal and external validation strategies. Internal validation usually relies on cross-validation or bootstrapping on the training dataset, while external validation evaluates the model on an entirely independent dataset.

Initially described as different ways of conducting ML algorithms, the learning paradigms shown in Fig. 1 are actually applicable to all the AI approaches, where typical ML algorithms are cited as methods and DL approaches as architectures. Briefly:

- Supervised learning is a process by which computers are able to learn from clearly labelled datasets by mapping a set of input features to their corresponding target output (Fig. 1). This approach is especially effective when a well-defined input–output relationship exists. Supervised learning encompasses two main tasks: (i) classification [naive Bayes, *k*-nearest neighbours, support vector machines (SVMs), decision trees, logistic regression, etc], and (ii) regression (linear, polynomial, Poisson, Bayesian, quantile, etc). Classifier performances seem to depend on the type of data that were analysed.
- Unsupervised learning is a process that enables computers to derive insights from unlabelled data, thereby facilitating the identification of latent patterns, the discernment of underlying structures, and the exposition of relationships within the data, all without the necessity for explicit instructions. It is particularly useful for exploring large datasets whose structure is not obvious. Common applications of unsupervised learning include (Fig. 1) clustering (e.g. hierarchical clustering, *k*-means, hidden Markov models, fuzzy C-means, and Gaussian mixture models) and dimensionality reduction [e.g. principal component analysis, *t*-distributed stochastic neighbour embedding (*t*-SNE), or independent component analysis]. These methods are widely used for pattern discovery, feature extraction, and data visualization, particularly in fields such as genomics, image analysis, and systems biology (Yan and Wang, 2022). For example, unsupervised learning has been used to relate chlorophyll fluorescence to the impact of low-temperature stress on plants (Lu *et al.*, 2023).
- Semi-supervised learning (Fig. 1) is a hybrid ML approach that combines a small amount of labelled data with a larger pool of unlabelled data during training. It is particularly advantageous when labelling is expensive, time-consuming, or requires expert knowledge (Yan and Wang, 2022). Semi-supervised learning can be used for anomaly detection, image classification, and protein sequence analysis. Popular techniques include Siamese networks, contrastive learning, and expectation-maximization algorithms (Sahito *et al.*, 2019).
- Ensemble learning combines multiple supervised models to improve robustness and accuracy, prioritizing performance over interpretability (Fig. 1). One of the most widely used ensemble techniques is bagging (bootstrap aggregation), where multiple models are trained independently on re-sampled datasets to reduce variance. A prominent example is the random forest (RF) algorithm, which aggregates the predictions of multiple decision trees through majority voting to reduce overfitting and variance (Lan *et al.*, 2018). Ensemble approaches to reduce bias include boosting, which builds models sequentially by focusing on errors from previous iterations—as implemented in gradient boosting machines (GBMs), AdaBoost, and XGBoost—making them prone to overfitting. Finally, stacking helps to reduce both variance and bias by constructing a high-level meta-model combining the predictions of several base models (Khan *et al.*,

Box 1. Continued.

2024). In the context of biological datasets, the use of ensemble methods is particularly advantageous when high dimensionality (too many features or genes), limited sample sizes, and data noise are challenging (Pirooznia *et al.*, 2008). Ensemble classifiers have been successfully applied in plant disease detection tasks (Pudumalar and Muthuramalingam, 2024), achieving 98.4% accuracy in cotton (Shahid *et al.*, 2024).

- Reinforcement learning (Fig. 1) is a paradigm in which an agent learns to make decisions by interacting with an environment and receiving feedback in the form of rewards or penalties. Through repeated interactions, the agent gradually learns optimal strategies that maximize cumulative reward over time (Murphy, 2025, Preprint). It excels in complex decision-making tasks where learning through exploration is essential, such as playing board games and autonomous vehicle control. In the field of plant sciences, reinforcement learning has been employed to simulate and optimize plant growth dynamics (Hitti *et al.*, 2024; Nasti *et al.*, 2024), as well as to enhance agricultural practices, including waste-water reuse strategies (Aponte-Rengifo *et al.*, 2023) and the implementation of digital twin systems for precision agriculture (Goldenits *et al.*, 2024).
- Transfer learning (Fig. 1) uses knowledge from pre-trained models on one task to enhance performance on a related task, such as adapting models trained on one species to another species for applications such as plant species classification (Kaya *et al.*, 2019) and disease identification (Paymode and Malode, 2022; Zhao *et al.*, 2022). Since models are not trained from scratch, transfer learning saves time and computational resources. It is particularly useful when there is a lack of annotated data for a specific plant species or when computational resources are limited.

probabilistic nature of the models and the nature of training data (Özer, 2024; L. Huang *et al.*, 2025).

Machine learning and data mining

ML and data mining are sometimes used as synonyms (Lan *et al.*, 2018), although they are not interchangeable. Data mining is a broader term that encompasses a wide range of techniques—including ML, but also statistical analysis and database systems—for extracting ('mining') valuable information buried in data, while ML is a specific approach within data mining that focuses on predictive modelling. The main advantage of ML is that it can deal with collinearity, non-linearity, and interactions better than other statistical approaches, and can prioritize among features (Box 1). In scientific research, data mining is often used as a first step to extract features and patterns from data (Box 1), which are then exploited by ML to identify informative groupings, extract subtle patterns (X. Zhang *et al.*, 2016; Greener *et al.*, 2022; Asnicar *et al.*, 2024), or anticipate future events (Ersöz and Ersöz, 2022; Collins *et al.*, 2024). Model performance is, however, absolutely dependent on data quality and sometimes requires human interaction to ensure accurate results (Hinton and Salakhutdinov, 2006).

Since a relevant strength of ML is the integration of different types of data, ML becomes very useful in plant biology to analyse genomic, transcriptomic, proteomic, phenotypic, metabolomic, and environmental data in order to identify relevant genes, genomic patterns, expression levels, metabolites, etc, associated with specific traits or responses to environmental stressors for a more comprehensive understanding of plant biology (Singh *et al.*, 2016; Silva *et al.*, 2019). ML also facilitates the precise breeding of salt-tolerant plants or rootstocks (Yan and Wang, 2023; Saleem *et al.*, 2025).

Deep learning for complex patterns

In contrast to ML, which typically relies on semi-manual feature extraction (Box 1) from structured relatively small datasets, deep learning (DL) is not a separate paradigm but a subfield of ML using deep, multilayered neural networks (Lan *et al.*, 2018; Dargan *et al.*, 2020) that can draw hierarchical representations directly from very large datasets. This is why in Fig. 1 most ML algorithms are in the Methods branches while the Architectures branches often contain DL approaches. Due to its elaborated architectures, DL is suitable to model intricate patterns where traditional ML methods struggle to scale (Zou *et al.*, 2019), as is the case of designing regulatory DNAs, which can be envisaged using GANs (Zrimec *et al.*, 2022). When DL is based on autoencoders (Box 2), it can detect plant leaf diseases with a remarkable accuracy (95.3%) (Abinaya *et al.*, 2023) and extract gene expression signatures (Tan *et al.*, 2017). DL models are then well suited to predict and simulate changes in methylation patterns of histones, genomes, and transcriptomes in stressed plants (Dobránszki *et al.*, 2025).

This ability of DL models to extract new insights from genomics data makes DL the method of choice for modelling transcription factor-binding sites, predicting the DNA accessibility and the splicing isoforms, and forecasting the impact of genetic variation on gene regulatory mechanisms and diseases (Eraslan *et al.*, 2019; Peng and Rajjou, 2024; Avsec *et al.*, 2025, Preprint). Assisted by plant phenotyping (see below), DL models can predict the occurrence of plant diseases and stresses

(Kamilaris and Prenafeta-Boldú, 2018; Navarro *et al.*, 2022; Murphy *et al.*, 2024).

Large language models

Among DL architectures, the class of large language models (LLMs) (Fig. 1) represents a significant advancement in sequence modelling (Box 3). They are now built upon the transformer architecture of Vaswani *et al.* (2017) (Box 2) and trained through self-supervised learning on vast text corpora to configure billions of parameters simultaneously rather than sequentially. LLM require high-performance computing infrastructures to process the text corpus and generate human-like complex and nuanced text (Steyvers *et al.*, 2025). This has placed LLMs at the core of modern language-based AI systems, powering everything from chatbots and virtual assistants to advanced reasoning agents (Bengesi *et al.*, 2024).

However, general-purpose chatbots often struggle to provide accurate and contextually relevant responses to specialized scientific queries, which is especially problematic in research, where accuracy and reproducibility are critical (L. Huang *et al.*, 2025; Steyvers *et al.*, 2025). In particular, they present a kind of 'plant blindness' (Lam *et al.*, 2024; L. Yang *et al.*, 2025b) because plant-specific scientific queries often contain factual inaccuracies (Geitmann and Bidhendi, 2023). Consequently, LLMs trained or fine-tuned on biomedical information, such as BioBERT, ClinicalBERT, Bio-Copilot, or Med-PaLM, are now available (Singhal *et al.*, 2023; Y. Liu *et al.*, 2025; Tong *et al.*, 2025). Specialist models for plant data have also begun to appear, such as SeedLLM·Rice trained with nearly 98.24% of published rice research which can address complex research questions, revealing unprecedented discoveries for climate adaptation of rice (F. Yang et al., 2025). Also, very recently, PlantGPT (R. Zhang *et al.*, 2025) has compiled the abstracts of >60 000 plant research articles to become a virtual expert in Arabidopsis phenotype–gene research that can be a blueprint for functional genomics research in food crops.

Since amino acid sequences, nucleotide sequences, and human languages display remarkable parallels in the course of their evolution (Searls, 2002), proteins and nucleic acids can be conceptualized as the biological languages of cells, enabling the application of LLMs to these biological sequences to identify the complex patterns underlying the 'language of life' (Box 3) (Lam *et al.*, 2024). As a result, protein language models (pLMs) and genomic language models (gLMs) such as GenePT, ESM, or DNABERT-2 were developed (Fig. 1) (Simon *et al.*, 2024; Benegas *et al.*, 2025; F. Guo *et al.*, 2025). The different foundation models for biological sequences enabled the generation of newly designed proteins or drugs, protein–ligand interaction modelling, regulatory region and DNA methylation forecasting, as well as structure and function prediction. The resulting essential message is that LLMs, while still in nascent stages, hold promise for hypothesis generation and experimental design, particularly in understudied species lacking extensive annotation databases, or in complex biological processes such as salt stress tolerance (Koh *et al.*, 2024a).

Computer vision

A significant application of AI is computer vision (Fig. 1) to enable machines to interpret and understand visual information from imaging sensors, including digital cameras in unmanned aerial vehicles (UAVs, commonly known as 'drones') (Walsh *et al.*, 2024). Applications for facial recognition, medical image analysis, autonomous vehicles, and surveillance systems have been clearly improved by the recent incorporation of DL, transformers and LLMs to image or video analysis. In plants, computer vision is exploited to capture phenotypic traits for the analysis and characterization of agricultural crops and the further quantification of crop responses to stress (Al-Tamimi *et al.*, 2022; Matese *et al.*, 2024). Spectral imaging sensors have become a preferred choice for many plant researchers because they provide better insight into plant physiology, have the potential to identify stresses during their initial phases, and are less vulnerable to external variations (Walsh *et al.*, 2024). While traditional methods relying on manual observation and analysis were subjective and prone to human error, AI-based techniques can accurately predict (up to 97.3%) early diseases in crops (Di Nisio *et al.*, 2020; Rajpoot *et al.*, 2023; Yu *et al.*, 2023; Jafar *et al.*, 2024), model plant growth (Goldenits *et al.*, 2024; Hitti *et al.*, 2024), analyse stressed conditions (Negrão *et al.*, 2017; Islam *et al.*, 2024), predict plant biomass accumulation (D. Chen *et al.*, 2018), or identifyolive tree cultivars combining molecular markers with DL-based image processing (Sesli *et al.*, 2020).

Data sources

A huge amount of data are available from omics approaches. Databases focused on salt stress were developed in recent decades, such as RiceMetaSys—a rice-specific resource centred on gene expression salinity stress (Sandhu *et al.*, 2017)—and the Salinity Tolerant Poplar Database [STPD; including information about genomic sequence, genes, functional information, non-coding RNAs, transposable elements, and polymorphisms (Ma *et al.*, 2015)]. Unfortunately, they are sometimes no longer accessible or maintained (López-Gómez *et al.*, 2026) and it is necessary to turn to resources integrating a broader group of plants and experimental conditions, such as ArrayExpress (https://www.ebi.ac.uk/biostudies/arrayexpress, accessed 2 July 2025) for high-throughput functional genomics experiments, including single-cell data, Expression Atlas for plants (https://www.ebi.ac.uk/gxa/home#plants, accessed 2 July 2025) to search by species and biological conditions, Gene Expression Omnibus (GEO) (Clough *et al.*, 2024) for high-throughput gene expression data, including microarray, RNA-seq, and single-cell sequencing datasets, Phytozome (Goodstein *et al.*, 2012), Ensembl Plants (Contreras-Moreira *et al.*, 2022), and even UniProt. There are also species-specific

Box 2. Common architectures for AI-driven bioinformatics applications

Artificial neural networks (ANNs) are widely used for classification and regression (Fig. 1), and are the basic architecture for pattern recognition. ANNs, inspired by the structure and function of brains, consist of interconnected processing units, or neurons (perceptrons), arranged in layered architectures that can capture complex non-linear relationships in data (Krogh, 2008). Despite their predictive power, ANNs are traditionally regarded as difficult-to-interpret black boxes, a situation that is changing nowadays with efforts on XAI, since interpretability is essential for generating hypotheses and validating biological findings in genomics (Karim *et al.*, 2023; Novakovsky *et al.*, 2023).

Autoencoders (Fig. 1) are a type of unsupervised learning architecture that consist of ANNs that are trained to map their unlabelled input data to a feature space and re-construct on their own the input from feature vectors. Two main components are required: an encoder, which compresses the input data into a lower dimensional latent representation (also called codings or embeddings), and a decoder, which attempts to reconstruct the original input from this compressed representation with minimal loss of information (Lan *et al.*, 2018; Bengesi *et al.*, 2024). The hidden layers between encoder and decoder constrain the learned representation and force the model to capture the most salient features of the data. Although they are computationally expensive and memory intensive, autoencoders are suitable for dimensionality reduction, feature extraction, and data reconstruction, making them particularly effective for tasks such as anomaly detection and image denoising (Bengesi *et al.*, 2024). The SSAE (sparse supervised autoencoder) model (Truchi *et al.*, 2024) is able to decipher weak perturbations in single-cell transcriptomic data that optimize the analysis of fine-tuned transcriptomic regulations. The GenoDrawing model (Jurado-Ruiz *et al.*, 2023) is based on autoencoders to predict and retrieve image-based phenotypes in apples from a small set of SNPs, and could be potentially useful to predict other genetically controlled visual phenotypes in other organs and species, as well as traits that are difficult to define.

Transformers are a type of deep neural network architecture (Fig. 1) originally developed for supervised language processing tasks, but that can be trained under different paradigms (e.g. self-/semi-supervised pre-training, supervised fine-tuning, transfer learning; Fig. 1), to dynamically analyse relationships between all words in a sequence, regardless of their distance from each other (Vaswani *et al.*, 2017), and are especially effective at next-token prediction. Tokens refer to the basic units of input data, such as words, sub-words, or characters, that are obtained by a tokenizer. The input text is first tokenized and the resulting tokens are then converted into dense vector representations (embeddings) via self-supervised pre-training, which leverages the self-attention mechanism to assign context-dependent weights to each token. This allows transformers to capture long-range dependencies and contextual (semantic) relationships more effectively than previous architectures such as recurrent neural networks or convolutional neural networks, which process texts sequentially. Transformer fundamental blocks are stacked encoders (which process the input sequence and produce a set of context-aware embeddings), decoders (which generate the output sequence based on the encoder's embeddings), or both (token embeddings including position embeddings are added to understand each token) (Cho *et al.*, 2024, Preprint). Based on how transformers use encoders and decoders, their architectures fall into three main categories (Fig. 1).

- Encoder-only models [e.g. bidirectional encoder representations from transformers (BERT)], which are effective for classification, semantic search, and information extraction (Peng and Rajjou, 2024).
- Decoder-only models [e.g. generative pre-trained transformers (GPTs)], which excel at text generation because they can predict the next sequence of words or characters (Peng and Rajjou, 2024).
- Encoder–decoder models (e.g. T5), which are suited for tasks such as translation and summarization (Bengesi *et al.*, 2024; Peng and Rajjou, 2024). In this group, LLMs built on autoregressive generative transformer architectures—often referred to as GPT-like, as most chatbots—excel at realistic text generation, dialogue systems, and summarization (Chen *et al.*, 2024; Rissom *et al.*, 2025).

Transformers have found widespread use in improving NLP (Fig. 1; Box 3), and they have also demonstrated significant potential in computational biology, particularly in analysing proteins and nucleotide sequences, such as the agronomic nucleotide transformer (ANT) (Mendoza-Revilla *et al.*, 2024).

Foundation models (Fig. 1) are large-scale, general-purpose models that are pre-trained on massive corpora of labelled and unlabelled data. They typically use transformer architectures to learn broad patterns in language, vision, graphs, or other modalities, encoding these patterns into high-dimensional embeddings (Chen *et al.*, 2024; F. Guo *et al.*, 2025). Foundation models can be then adapted to downstream tasks through fine-tuning using supervised learning or transfer learning strategies. Some of them (e.g. GPT-4, PaLM-2, and LLaMA-2) have become the backbone of many modern AI systems, enabling rapid advancements in sequence modelling (Box 3) (Rissom *et al.*, 2025). Foundation models have also been successfully used in bioinformatics for biomarker discovery, enzyme design, disease diagnosis, and omics

Box 2. Continued.

analysis, where they are able to identify the inherent relationships in amino acid and nucleotide sequences (F. Guo *et al.*, 2025). A new perspective is ongoing in molecular cell biology: the multimodal foundation models, which are pre-trained on vast and varied omics datasets using self-supervised and contrastive learning techniques to reduce overfitting, capture fundamental biological knowledge, and understand the complex language of cells (Cui *et al.*, 2025).

databases, such as The Arabidopsis Information Resource (TAIR; https://www.arabidopsis.org, accessed 10 June 2025) for *A. thaliana* that facilitate the discovery of, for example, salt stress-tolerant genes in other plants (Yaschenko *et al.*, 2025). The *Zea mays* pan-genome database MaizeGDB (Woodhouse *et al.*, 2021) has revealed gene families associated with salinity stress (Hayford *et al.*, 2024). Indeed, MaizeGDB together with TAIR and Gramene (Tello-Ruiz *et al.*, 2022) have allowed detection of 16 genes involved in the tolerance to salinity or water deficiency in maize (Luo *et al.*, 2021), highlighting the advantages of combining several genomic resources. Other genomics portals such as the Sol Genomics Network (SGN) (Fernandez-Pozo *et al.*, 2015) for *Solanaceae* species have also contributed to the study of wild species and mutants to improve salt tolerance in breeding programs (Rivero *et al.*, 2014; Razali *et al.*, 2018; Capel *et al.*, 2020; Molitor *et al.*, 2024; Sadder *et al.*, 2025). Finally, to build AI models about salt stress, large datasets in open-source repositories are also necessary (L. Yang *et al.*, 2025b), such as the generalist sites Kaggle (https://www.kaggle.com, accessed 29 May 2025), U.S. Government's Open Data (https://data.gov, accessed 29 May 2025), TERRA_REF [https://terraref.org, accessed 28 May 2025 (LeBauer *et al.*, 2021)], and Google Earth Engine (Kabiraj *et al.*, 2022). More specific datasets for plant diseases are PlantVillage (Hughes and Salathe, 2016, Preprint), PlantDoc (Singh *et al.*, 2020), or Cassava Leaf Disease Classification (available at Kaggle).

High-throughput phenotyping

Plant phenotyping is required for accurate and precise trait collection and use of genetic tools to improve plant performance. Common phenotyping procedures (e.g. fresh weight, chlorophyll levels, and oligoelement determinations) are error-prone, time-consuming, and labour-intensive, often destructive, requiring a large population of plants (Negrão *et al.*, 2017; Rai, 2024). The arrival of computer vision allowed for the quantitative and time-series assessment of plant performance based on satellite or UAV images, and visible light sensors (Kamilaris and Prenafeta-Boldú, 2018; Di Nisio *et al.*, 2020; Al-Tamimi *et al.*, 2022; Murphy *et al.*, 2024), opening the door to the high-throughput phenotyping—characterization of many phenotypes by non-destructively capturing plant traits assisted by computers and robotics—(Nabwire *et al.*, 2021; Al-Tamimi *et al.*, 2022). ML methods identified whether *Camelina sativa* seeds come from plants grown in high salt conditions (Vello *et al.*, 2024). Subsequent use of DL methods allowed the comprehensive integration of phenotypes, biochemical changes, and physiological determinations to detect plant stress symptoms and comprehend the impact of salt stress (Singh *et al.*, 2016; Al-Tamimi *et al.*, 2022; Khalifani *et al.*, 2022; Walsh *et al.*, 2024). DL was also able to demonstrate that hyperspectral imaging—a record of spectral signatures across hundreds of wavelengths—can forecast, and even substitute, physiological and biochemical assays (Feng *et al.*, 2020; Murphy *et al.*, 2024), and identify specific patterns in leaves to differentiate canopy and non-canopy regions in individual plant images (Zhou *et al.*, 2022), classify biotic and abiotic stress damage in plant leaves (Khalifani *et al.*, 2022), and discriminate non-stressed from stressed plants with a precision >0.84 (Alhnaity *et al.*, 2020; Del Cioppo *et al.*, 2024). A more successful DL model is ResNet50, that can classify plant stress levels with >98.6% accuracy based on chlorophyll fluorescence imaging (Deng *et al.*, 2024).

The brief take-home message is that different fields of AI can be successfully used for phenotyping of plants under salt stress based on images or spectra, with minimal laboratory burden. The success of such approaches would enable real-time monitoring of crop health in the field, which is crucial for precision agriculture to optimize at least growth efficiency and productivity.

Soil salinization

Prediction of soil salinization was explored early using ML models, but the advent of ANN and then DL provided high (usually >0.9) values of accuracy, area under the receiver operating characteristic curve (AUC-ROC), or R^2 (An *et al.*, 2023), preferably using multifeature combination (Xie *et al.*, 2025) after feature selection (Wu *et al.*, 2018; Sirpa-Poma *et al.*, 2023; Abd Elaziz *et al.*, 2024). Other studies were more focused on soil salinity patterns, revealing that higher altitudes discharged a large amount of salt to the surroundings farmlands (Wang *et al.*, 2021), or that 74% of the Satkhira district in Bangladesh has moderate to high soil salinity concentrations (Sarkar *et al.*, 2023). The model of An *et al.* (2023) supported the conjecture that subterranean CO_2 sequestration could reduce salinization effects in deserts: the abiotic CO_2 absorption by saline–alkali soils is a result of CO_2 reaction with soil salts and moisture. The comprehensive model of Shokri *et al.* (2021) integrated electrical conductivity with

Box 3. The language of life

The concept that the order of elements matters in texts, speeches, time-series data, or biological sequences, and that each element in a sequence is often dependent on those that preceded it, led to the development of sequence modelling (Fig. 1). This application of AI was initially based on recurrent neural networks, but now transformer-based LLMs are mainly used since they outperformed previous models (Merkx and Frank, 2021).

Natural language processing (NLP) is focused on enabling machines to understand, interpret, translate, generate, and manipulate human language in both written and spoken forms. Although the foundation for NLP and current chatbots started with ELIZA in 1966 (Miller, 2025), the burst of NLP was in the 21st century, when significant advancements in computational power and efficient ANNs (Dwivedi *et al.*, 2021; Samant *et al.*, 2022; Miller, 2025) have emerged, particularly transformers (Box 2; Samant *et al.*, 2022). The present conversational agents (chatbots) based on NLP can process user queries and respond in natural language, being capable of holding coherent conversations. The capabilities of NLP are now able to improve patient outcomes, reduce physician workload (Arivazhagan and Van Vleck, 2023), extract interactions between diseases, drugs, genes, and proteins from the scientific literature (Gachloo *et al.*, 2019), predict plant–pathogen relationships (Lei *et al.*, 2024), and even identify nucleotide sequences in sequencing experiments without prior mapping (Strzoda *et al.*, 2024). The increasing integration of generative AI models into NLP has given rise to hallucinations, which are now being harnessed through a combination of fine-tuning, RAGs, and specialized LLMs, such as PlantGPT (R. Zhang *et al.*, 2025).

Genomic language models (gLMs) are LLM-based sequence models trained on DNA sequences that have been successfully applied to predict gene expression levels, the functional impact of genetic variants, regulatory element activity, RNA processing, promoter/terminator strength, and design of novel DNA sequences (Benegas *et al.*, 2023, 2025; Mendoza-Revilla *et al.*, 2024). gLMs enable a comprehensive understanding of the entire genome in spite of the abundance of repetitive sequences and the variability of non-coding regions (Zhai *et al.*, 2025), especially in plants (Claros *et al.*, 2012). gLMs also have a promising use in the prediction of plant gene functions based on published results concerning gene/phenotype/environment interactions, as well as to uncover relationships previously unreported (Sunil *et al.*, 2024). Notable examples for plants are as follows (Fig. 1).

- ANT (agronomic nucleotide transformer) was developed for edible plant species and can predict regulatory elements, RNA processing patterns, and gene expression, enabling the identification of promoter and enhancer regions in orphan crops (Mendoza-Revilla *et al.*, 2024).
- FloraBERT is able to predict reliable gene expression in a single plant species (Levy *et al.*, 2022, Preprint).
- GPN (genomic pre-trained network) learns gene structure and DNA motifs in order to predict the functional effects of genetic variants across the entire *A. thaliana* genome (Benegas *et al.*, 2023). Based on the odds ratio, GPN outperforms predictors based on popular conservation scores such as phyloP and phastCons.
- PlantCaduceus was pre-trained on a curated dataset of 16 angiosperm genomes and was able to outperform previous models concerning translation initiation/termination sites, deleterious mutation identification, and splice donor and acceptor sites, demonstrating a high transferability between species (Zhai *et al.*, 2025).
- PlantRNA-FM is a specialized foundation model pre-trained on plant RNA sequences and structures from 1124 distinct plant species (H. Yu *et al.*, 2024). It enabled the prediction of RNA functions and the identification of biologically functional motifs in RNA sequences and structures across transcriptomes that were experimentally validated. More importantly, the results highlighted the importance of the position information of these functional RNA motifs in genic regions, demonstrating its potential for advancing our understanding of plant biology and improving crop breeding strategies.

Protein language models (pLMs) can efficiently capture in its embeddings the key aspects of the amino acid grammar underlying protein sequences, such as structural, functional, and evolutionary information (Barrios-Núñez *et al.*, 2024; Schmirler *et al.*, 2024; Shrestha *et al.*, 2024; H. Huang *et al.*, 2025; Q. Zhang *et al.*, 2025) as well as PTM predictions. Since pLMs make more accurate predictions than ML methods and the classical multiple sequence alignments, and foundation pLMs tend to consume far fewer resources than multiple sequence alignments (Weissenow and Rost, 2025), pLMs save efforts and resources, and avoid overfitting. Examples of remarkable pLMs (Fig. 1) are:

- ProtT5, a T5-based model fine-tuned on UniRef50 to accurately predict protein secondary structure and function (Elnaggar *et al.*, 2022).
- ProteinBERT, a BERT-inspired foundation model to study protein structure, PTMs, and biophysical attributes in protein sequences (Brandes *et al.*, 2022).
- ProtGPT2, an autoregressive GPT-inspired pLM that generates *de novo* protein sequences following the principles of natural sequences and allows the study of PTMs (Ferruz *et al.*, 2022).

Box 3. Continued.

- ESM (evolutionary scale modelling), a family of transformer-based pLMs that can predict protein structure, function, mutational effects, and interactions (Rives *et al.*, 2021).
- NetGO 3.0, an ESM-based pLM combined with logistic regression for automated function prediction of unannotated proteins that outperforms homology-based annotators (Wang *et al.*, 2023).

environmental information (global circulation models, soil, crop, topographic, climatic, vegetative, and landscape properties) to generate spatio-temporal predictions of soil salinity. Their results indicated that Brazil, Peru, Sudan, Colombia, and Namibia had a rapidly increasing total area of salinized soils, and that dry lands of South America, southern and western Australia, Mexico, south-west of the USA, and South Africa are salinization hotspots.

As essential findings, it should be retained that ANN and DL methods can successfully predict soil salinity levels by combining multiple features, preferably after feature selection. The resulting models are providing new insights into the patterns and causes of soil salinization, particularly in regions where the combined effect of a remote location and a difficult socio-economic context prevents frequent field sampling. Accurate predictions of soil salinization are vital for effective management strategies, and can support decision-making processes for farmers, policymakers, and environmental managers.

Studying plant responses to salt stress

Studies with simple ML methods were extensively used to identify genes involved in salt stress or even reveal some kind of coordination between genes (Sodini *et al.*, 2022; Sonsungsan *et al.*, 2024). The arrival of more computing capability, new AI methods, and high-throughput determinations in the field and laboratory enabled the integration of different omics data, deepening our understanding of how plants respond to salt stress.

From genomes to pan-genomes

Plant genome sequences are indispensable for the future of agriculture and the environment since they can contribute to safeguarding food security, and are a source of genotypes and genes for breeding and studying diversity to adapt crops to soil salinization and salt stress. Reference genomes for many plants have been generated using short-read sequencing methods (Garg *et al.*, 2024), but recent advances in single-molecule sequencing technologies have greatly increased the accuracy (to 95–99%) and length (to 20–100 kb) of sequencing reads (Sahu and Liu, 2023). These long reads simplified genome reconstruction so much that more plant genomes (2373 genomes from 1031 plant species) were sequenced between 2021 and 2023 than in the previous 20 years (1144 genomes from 782 plant species) (Xie *et al.*, 2024). When gapless sequences of entire chromosomes are obtained, those genomes are called telomere-to-telomere (T2T) assemblies (Rautiainen *et al.*, 2023).

T2T assemblies of crops including, for example, apple (Su *et al.*, 2024), olive tree (Lv *et al.*, 2024), and avocado (Yang *et al.*, 2024) are now available [recently reviewed by Garg *et al.* (2024)]. T2T genomes have paved the way for genetic diversity studies and genomics-assisted breeding in crops (Ellegren, 2014; López-Gómez *et al.*, 2026). The availability of gapless genomes has fostered the construction of plant pan-genomes [>110 are available; reviewed and compiled by He *et al.* (2025)] that provide a more comprehensive representation of genetic information within a plant population compared with a single reference genome. Pan-genomes have a very positive impact on mapping of allelic variants, discovery of candidate genes, development of molecular markers, and the introgression of traits related to tolerance to salt and other abiotic stresses (Varshney *et al.*, 2024). Notable findings in plant salt stress revealed by pan-genomes are loci associated with salt tolerance in natural grape populations (Cochetel *et al.*, 2023), identification of a gene involved in maintaining Na^+/K^+ homoeostasis in rice in a major locus for salt tolerance known as *qSTS5* (H. Wei *et al.*, 2024), or insights into salt adaptations in 11 species of the *Fraxinus* genus (J.N. Liu *et al.*, 2025).

The essential findings due to genomes, gap-less T2T genomes, and pan-genomes are that new opportunities for the research of plant stress responses are available, with the identification of new genes, alleles, and loci related to salt stress. They may also help with the design of genomics-assisted breeding programmes.

Genome-wide association studies

High-quality genome references facilitated the use of genome-wide association studies (GWASs) based on single nucleotide polymorphisms (SNPs) to identify relevant alleles or regions throughout a genome related to relevant plant traits (Bakshi, 2024), including abiotic stresses (Javid *et al.*, 2022). GWAS is suitable for studying plant salt stress responses since these are the result of many genes each having a small effect. An early GWAS in *A. thaliana* identified genes putatively responding to various osmotic stresses (Li *et al.*, 2008). Then authors utilized ANN to combine these genes and their *cis*-regulatory elements to identify other functionally relevant genes. Concerning salt stress, Kobayashi *et al.* (2016) discovered in *A. thaliana* that the most significant SNPs were linked to the genes of ATP-binding cassette B10, vacuole proton ATPase A2, potassium channel KAT1, and

calcium sensor SOS3, in addition to previously uncharacterized genes for salt tolerance related to pectin metabolism and protein trafficking. Using 550 lines of cotton (Abdelraheem *et al.*, 2021) or 445 maize natural accessions (X. Luo *et al.*, 2019), >50 genes associated with responses to salt tolerance were identified. The identification of 27 SNP markers associated with salt tolerance in *Medicago truncatula* (Medina *et al.*, 2020) was later used to train ML models to estimate the breeding values of the training population under salt stress. Support vector machine (SVM)- and random forest (RF)-based models were the more accurate [Pearson correlation of 0.793 and root mean square error (RMSE) of 0.353]. When the model was weighted and extended to other polyploid crops, the prediction accuracy of plant yield (in terms of biomass and plant growth vigour) was increased in alfalfa and potato from 50% to >80% (Medina *et al.*, 2021).

Therefore, the essential findings are that (i) GWAS combined with AI allowed the identification of relevant genes for salt stress tolerance in plants, and (ii) the SNPs extracted with this methodology with good breeding values were promising for predicting plant yield under salt stress.

Proteomics

Proteomics has also contributed to the identification of proteins regulating metabolic processes, hormonal crosstalk, and signalling pathways involved in salinity tolerance (Mansour and Hassan, 2022). For instance, the differences between salt-sensitive and -tolerant rice cultivars rely on 206 differentially expressed proteins associated with photosynthesis, energy metabolism, glutathione metabolism, nitrogen metabolism, and stress defences. The mechanism depended on Rubisco, chlorophyll *a*-*b*-binding protein, phosphoglycerate kinase, cytochrome *c* oxidase subunit 5C, glutamine synthetase, glutathione *S*-transferase, peroxidases, and thioredoxin (Frukh *et al.*, 2020). In the case of chickpea, photosynthesis, energy metabolism, stress responsiveness, and protein synthesis and degradation explained the variation in salinity tolerance between genotypes, with proteins similar to those identified in rice (chlorophyll *a*-*b*-binding protein, oxygen-evolving enhancer protein, and Rubisco subunits) and also ATP synthase, carbonic anhydrase, fructose-bisphosphate aldolase, ascorbate peroxidase, heat shock proteins, and late embryogenesis abundant (LEA) proteins (Arefian *et al.*, 2019). Combining metabolomics and proteomics analyses after a temporal exposure to salt stress to study the cell processes involved in the halophytes *Sesuvium portulacastrum*, *Salicornia brachiata*, and *Salicornia maritima* (Benjamin *et al.*, 2020) demonstrated that Ca^{2+}, ABA, and jasmonate signalling coordinately regulated salt tolerance to promote Na^{+} sequestration into the vacuole, and maintained ROS homoeostasis, and that photosynthesis, heat shock proteins, peroxidases, and expansins were also affected by salinity (Benjamin *et al.*, 2020; Cao *et al.*, 2023).

Specific AI methods have been designed for proteomics. For example, DeepAProt (Ahmed *et al.*, 2023) is an innovative DL-based method for classifying unknown abiotic stress (cold, drought, heat, and salinity) protein sequences that predicted new salt-responsive proteins in *Poaceae* with 81.69% accuracy. DeepGOPlus (Kulmanov and Hoehndorf, 2020) was designed to predict protein function from the sequence combining DL models with sequence similarity. A total of 16 *HKT* (high-affinity K^{+} transporter) genes were identified in *Spartina alterniflora*, most of them highly expressed in salt stress (Yang *et al.*, 2023).

There are some essential findings related to proteomics: confirmation of protein functions, signalling, and cell processes involved in the physiological response of plant to salt stress obtained using other approaches. DL methods based on proteomics can predict new salt-responsive proteins and their possible functions, which may be useful for breeding new salt-tolerant varieties.

Transcriptomics

The most relevant methodology applied in the study of salt stress in plants is high-throughput transcriptomics. Microarray data from the Gene Expression Omnibus (GEO) were recently used to perform a hybrid ML model (Nazari *et al.*, 2023) that identified 42 predictive genes involved in stress tolerance in *A. thaliana*, with XGBoost and RF models providing the best classifiers (accuracies of 0.991 and 0.985, respectively). The authors selected three loci (AT5G44050, AT2G47180, and AT1G70700) as promising biomarkers for salt-tolerant crops. However, the main source of transcriptomic information is RNA-seq, which allowed—to cite only a few examples—the identification of several transporters as gene candidates for salt tolerance in citrus rootstocks (Asins *et al.*, 2023), the selection of 196 rice genes responding to biotic and abiotic stresses (interestingly, many genes down-regulated in abiotic stresses were up-regulated in biotic stresses) (Shaik and Ramakrishna, 2014), the disclosure of brassinolides as alleviating salt stress (J.S. Wu *et al.*, 2025), the identification of very relevant genes (*EXPA4*, *HKT*, *SOS*, *TVP*, and *NHX*) in response to salt stress in wheat (Gholizadeh *et al.*, 2024) and cotton (Q. Liu *et al.*, 2025), the role of alternative splicing under salt stress in wild cotton (*Gossypium davidsonii*) (F. Zhang *et al.*, 2016), or the discovery of 18 miRNAs involved in plant adaptation to salt stress (X. Chen *et al.*, 2025).

Some specific AI models have also been developed for transcriptomics. 'mlDNA' is an ML method that outperforms general-purpose algorithms at identifying stress-related genes (Ma *et al.*, 2014). Uygun *et al.* (2019) constructed an ML-based pipeline to predict moderately well (AUC-ROC <0.71) whether an *A. thaliana* gene would be up-regulated or non-responsive for each of the cell types of interest under salt stress. Since extremophytes deploy divergent routes to achieve nutrient balance, osmotolerance, boost antioxidant capacity, and modify ion transport and specific signalling pathways, all of them absent in *A. thaliana* (Tran *et al.*, 2022, Preprint), Huang *et al.* (2024) took advantage of the pLM 'NetGO 3.0' (Wang *et al.*, 2023) to annotate gene functions in RNA-seq data of *S. alterniflora* (a extremophyte monocot living in estuarine salt marshes). They

identified substantial changes impacting gene transcription, alternative splicing, ion transport, and ROS metabolism during salt stress. Surprisingly, differential expression and alternative splicing were mutually exclusive.

In a few words, transcriptomics-based AI models have been the most widely used approach to discover, study, and classify genes, non-coding RNAs, and processes involved in plant stress response, as well as the proposal of salt tolerance biomarkers. Salt stress-specific AI models for differential gene expression and functional annotation have been developed, whose application revealed the mechanisms of some specific behaviours in extremophytes.

Multi-omics

The multifaceted nature of salt stress response can be better deciphered combining two or more omics approaches (i.e. 'multi-omics') rather than with the typical mono-omics analyses described above (Roychowdhury *et al.*, 2023). A more profound understanding of molecular pathways, regulatory networks, and molecular markers is obtained, which can reveal plants with genetic potential for future breeding programmes (Subramanian *et al.*, 2020; Ullah *et al.*, 2022; Roychowdhury *et al.*, 2023; Zou and Xu, 2025). Training and testing data can be obtained from scratch or with the practice of 'data reuse', which is strongly recommended in plant research (Hafner *et al.*, 2025), or from databases where multi-omics data for several individuals are already consolidated (Raza *et al.*, 2025a; Q. Wang et al., 2025; C. Zhang *et al.*, 2025). The advances in computer performance and AI algorithms have paved the way for integrative tools such as Omics Fusion (Brink *et al.*, 2016), mixOmics (Rohart *et al.*, 2017), or the LLM-based tool AutoBA (Zhou *et al.*, 2024).

In plants, the Omics Fusion platform was used to identify genes, proteins, metabolites, and pathways associated with salinity stress in oil palm (*Elaeis guineensis*) (Bittencourt *et al.*, 2022). A pioneer study that integrated advanced imaging technologies and multi-omics data (including GWAS) together with salt concentration was able to identify a cluster associated with the early growth decline in salt stress and another cluster regulating the longer term ionic stress (Campbell *et al.*, 2015). A study integrating high-throughput phenotyping—to measure plant growth, morphology, colour, and photosynthetic activity—with GWAS confirmed those loci in *A. thaliana*, with the early response genes corresponding to the osmotic phase of salt stress, and the long-term responding genes with the ionic phase (Awlia *et al.*, 2021). Mueller *et al.* (2024), merging proteomics with transcriptomics data, revealed that the adaptation mechanisms of *Phoenix dactylifera*—based on the efficient sodium and chloride exclusion at the roots, osmotic adjustment, ROS scavenging in leaves, and remodelling of the ribosome-associated proteome—show a remarkably high degree of convergence between gene expression and protein abundance. Zhu *et al.* (2025) combined transcriptomic and metabolomic data in the recretohalophyte *Limonium bicolor* through co-expression network analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis to identify that low salt treatment (100 mM) increased plant growth, photosynthesis efficiency, and antioxidant enzyme activity, while high salt treatments (300–400 mM) provoked severe osmotic and oxidative stress that led to a significant decrease in plant growth, photosynthesis efficiency, and antioxidant enzyme activity, accompanied by a significant increase in the content of organic soluble substances and ROS. More encouraging results were obtained with the recent combination of molecular markers and phenotype traits in GenPhenML, which enabled the prediction of barley genotypes resistant and sensitive to drought and salt stresses, with an accuracy >97% and R^2 >0.99, with ANNs outperforming other ML models such as SVM and RF (Akbari *et al.*, 2024). Integration of metabolomics and transcriptomics using ML, ensemble learning (XGBoost), and ANN in KANMB (Kolmogorov–Arnold network for identifying metabolic biomarkers) in the halophyte *S. alterniflora* revealed 226 new metabolic biomarkers, many of them involved in flavonoid biosynthesis (S. Chen *et al.*, 2025). The XGBoost model also outperformed the SVM and RF models in accuracy (92% versus 86%) and AUC-ROC (0.97 versus 0.95). Very recently, the ensemble learning model SaGP (saline–alkali genes prediction) was trained with gene sequences and protein features to uncover novel salt–alkali tolerance genes based on the protein sequence (Qiao *et al.*, 2025). The accuracy of 98.7%, Matthews correlation coefficient (MCC) of 0.598, AUC-ROC of 0.941, and area under the product–recall curve (AUPRC) of 0.602 indicates that this model outperformed all other similar classifiers previously published.

The essential findings that can be extracted from the previous examples are that multi-omic approaches provide more comprehensive insights into the complex trait of plant salt stress responses (Soltabayeva *et al.*, 2021), accelerating the knowledge of non-canonical stress adaptation strategies (Ahmed *et al.*, 2023; Zhang *et al.*, 2024), revealing novel targets for breeding stress-resilient crops (Koh *et al.*, 2024b, Preprint), and boosting the capacity to deduce interactions between genes with higher accuracy (Rico-Chávez *et al.*, 2022). However, the success of AI-driven multi-omics largely depends on adequate experimental design and the volume and quality of omics data (Box 4). When enough omics data are available to avoid overfitting, ANN or DL usually outperform good ML methods such as RF and SVM.

Predicting plant stress

For translational purposes, plant stress detection represents a critical aspect of modern agriculture. For example, *bHLH119* and the E3 ubiquitin protein ligase gene *DRIP2* were proposed as useful biomarkers for barley (*Hordeum vulgare*) resilience to stress after analysing 515 RNA-seq profiles across 18 independent studies (Panahi, 2024). The stress a plant is suffering can be predicted by StressGenePred (Kang *et al.*, 2019), which is based on

Box 4. How to conduct reliable AI studies using omics data

Since biological data have unique characteristics that require careful consideration when applying AI, models are expected to meet the following minimal quality criteria to produce reliable results that can be replicated and reproduced by other research groups (van Royen *et al.*, 2023).

(i) Complete and transparent reporting is required to warrant replicability, reproducibility, and thereby credibility. This implies that code and software should be open and available, and that the data used for training and testing should be accessible.
(ii) The intended use of the published model must be carefully defined.
(iii) Rigorous internal and external validation is required to determine the estimates of predictive performance and avoid overoptimism.
(iv) The sample size for splitting into the training and testing sets must be large enough to ensure adequate representation and avoid overfitting, depending on the learning paradigm. Ideally, the model should be trained and tested using thousands of elements, rather than just a few dozens.

Other particular pitfalls noted by Whalen *et al.* (2022) and Collins *et al.* (2024) are as follows.

- The inherent biological structure of omics data has an impact on model performance when training and testing sets were extracted from different contexts. For example, models trained on *in vitro* data often perform poorly on *in vivo* data.
- Omics data are often interconnected or even correlated, which violates the independence assumption of many AI algorithms. As a consequence, performance estimates are inflated or inconsistent.
- Experiments involving biological entities usually contain confounding factors—unmeasured variables that can create or mask associations. This is well known in GWAS, where genotype–phenotype relationships have an impact on the population structure.
- When the biological problem involves a small proportion of positive cases (e.g. genes responding to one transcription factor across the genome), the training and testing sets will be unevenly distributed, leading models to overfit the majority class. This flaw would be reduced using re-sampling techniques.
- Biological data are often noisy, biased, and high-dimensional (multifaceted), which can lead to overfitting and poor generalization to new data. This issue is exacerbated when model performance is based only on accuracy.
- Quality of data used for training AI models, as well as the identification of potential confounding factors (batch effects, age, variety or cultivar, study site, etc), is crucial since poor-quality data can lead to inaccurate predictions and unreliable results.
- The reproducibility of AI models on omics data cannot rely only on cross-validation: they must be validated on independent datasets to ensure robustness and generalizability.
- The integration of multi-omics data is a complex task that requires careful consideration of the different data types and their relationships. AI models must be designed to handle the complexity and heterogeneity of multi-omics data to provide meaningful insights.

ANN trained with time-series data of heat, cold, salt, and drought stresses. It provides an accuracy of 96.3%, very close to RF (96.1%) and SVM (94.5%). Another AI model is ASRpro (Meher *et al.*, 2024), that combined stress-specific protein sequences and six types of abiotic stresses (cold, drought, heat, light, oxidative, and salt). It was able to identify the type of stress with modest accuracy (75–86%), with SVM being the most accurate method. The SVM-based model of Mohammadi and Asefpour Vakilian (2023) demonstrated that miRNAs were more reliable indicators of plant abiotic stresses at early stages in cucumber (R^2=0.99) than conventional morpho-physiological (R^2=0.64) and biochemical (R^2=0.82) features at early stages of abiotic stresses in cucumber. They propose that electrochemical miRNA biosensors would enable farmers to detect the salt and drought stress with higher accuracy compared with conventional methods. Interestingly, they found that miRNA-477b and miRNA-399g exhibit divergent roles in stress response. AsmiR was another ML-based tool for predicting miRNA associated with four abiotic stresses (cold, drought, heat, and salt) by utilizing only sequence information (Pradhan *et al.*, 2023). The SVM-based models achieved the highest cross-validation accuracy in all four abiotic stress conditions (accuracy of 84.3% for salt stress versus 50–75% for the other models).

Computer vision technology based on high-resolution images, hyperspectral (better than multispectral) imaging, and thermal imaging can capture distinct stress signatures based on

physiological responses (e.g. leaf discoloration and temperature variations). These data, combined with ML (mainly SVM) or DL, although including the extraction of meaningful features, were able to rapidly detect stress symptoms and severity with a 99% accuracy (Islam *et al.*, 2024; Kaur and Kaur, 2024). The main stress indicators were leaf area, plant height, chlorophyll content, and reflectance patterns. For example, drought stress reduces leaf area and biomass, while nutrient stress alters leaf colour or texture. However, models based on images and spectra present some limitations: (i) early-stage detection of stress was not possible; (ii) the same morphological changes can be produced by different stresses; (iii) every model is dependent on the species analysed, hampering the differentiation of the stress type; and (iv) deploying machine vision systems on extensive agricultural fields can be logistically challenging and may require significant investment (Islam *et al.*, 2024).

The deduced essential finding is that AI models can successfully predict salt stress in plants, with SVM and ANN providing suitable models. The incorporation of miRNA data clearly improves the result, probably because in miRNAs are more significant than changes in genes under abiotic stres (Al-Tamimi *et al.*, 2022). Combined with computer vision, the main stress indicators have been revealed, and the combination of images and AI is currently opening up new methods of crop management for sustainable agriculture.

Predicting plant growth

Effective plant growth and yield prediction is an essential task for agriculture. Combinations of ML and DL algorithms to predict fruit yield (Vidhya *et al.*, 2024) and stem growth across different scenarios (Alhnaity *et al.*, 2020; Jeevan Nagendra Kuma, *et al.*, 2024) have been developed, with promising results. For example, sunflower grain yield under normal and salinity conditions has been modelled with an R^2=0.914, revealing that head diameter was the most influential parameter to predict growth under salt stress (Khalifani *et al.*, 2022).

Biostimulant treatments are being used to mitigate the negative impacts of stressful conditions on plant growth. γ-Aminobutyric acid (GABA) is known to accumulate in cells under stressful conditions and becomes involved in physiological and biochemical functions for survival, explaining the advantageous results of the exogenous application of GABA under abiotic stresses (Zarbakhsh and Shahsavar, 2022). The ANN-based model of Zarbakhsh and Shahsavar (2022) can predict the increase of physiological activity of pomegranate (*Punica granatum*) plants after GABA treatment. For example, it predicted that best values of crown diameter (18.42 cm), plant height (151.82 cm), leaf length index (5.67 cm), leaf width index (1.76 cm), and leaf area index (13.82 cm) could be achieved by applying 10.57 mM GABA on the 'Atabaki' cultivar under non-stress condition after 20.8 d (Zarbakhsh and Shahsavar, 2022).

Invasive plants can cause serious problems to native environments and local economies. Unfortunately, their adaptation to stressful environments is one of the mechanisms enhancing their invasiveness. ANNs were applied to soil parameters (pH, electrical conductivity, water content, temperature, and organic content) to accurately forecast growth based on plant height. This model successfully predicted the invasiveness of *Alternanthera philoxeroides* under salt stress, with a mean absolute percentage error (MAPE) as low as 2.2% (Javed *et al.*, 2022).

The final take-home message is that head diameter was the most influential parameter to predict growth under salt stress. An ANN-based model could predict GABA treatment to optimize pomegranate growth. ANN models were also helpful in the management of invasive species in stressful environments based on soil parameters.

Post-translational modifications

PTMs are enzymatic (e.g. phosphorylation, acetylation, methylation, glycosylation, palmitoylation, ubiquitinylation, or SUMOylation) or non-enzymatic (e.g. glycation and nitrosylation) covalent attachments of specific chemical groups to amino acid side chains in proteins (Schwartz *et al.*, 2009; Martí-Guillén *et al.*, 2022). PTMs alter the biophysical properties of the target proteins, rapidly affecting their function, stability, localization, and interactions. This results in an increase in the functional diversity of the cell proteome, beyond that encoded by the genome. PTMs are also involved in the regulation of plant responses to various external stimuli (Schwartz *et al.*, 2009; Hashiguchi and Komatsu, 2016). The plant-specific repository is Plant PTM Viewer, whose last update (Willems *et al.*, 2024) contains >334 255 sites for 33 types of PTMs from eight different plant species. Nowadays, PTMs are considered early regulators of cellular processes by dynamically tailoring protein behaviour to environmental cues.

It is well established that carbonylation, *S*-nitrosylation, Tyr-nitration, phosphorylation, methylation, ubiquitinylation, lipidation, and SUMOylation play a role in salt stress (Martí-Guillén *et al.*, 2022; Mata-Pérez *et al.*, 2023; Najar, 2024; L. Wei *et al.*, 2024). Additionally, crotonylation seems to alleviate salt stress in the chloroplast by modifying proteins involved in photosynthesis, ATP synthesis, and protein folding (D. Zhu *et al.*, 2023). Salt stress also increases *N*-glycosylation, altering cell wall organization, protein folding and modification, amino acid biosynthesis, and skeleton organization (Qin *et al.*, 2024). Interestingly, a global increase in Lys-deacetylation has been recently observed during salt stress in *Phyla nodiflora* (L. Wang *et al.*, 2025); in particular, deacetylation of glutathione *S*-transferase enhances its ROS-scavenging activity, and deacetylation of other proteins involved in the ABA signalling pathway, and flavonoid, auxin, and brassinosteroid biosyntheses have also been demonstrated in salt stress (L. Wang *et al.*, 2025). This contrasts with the already known association of histone acetylation with up-regulation of cell wall-loosening genes (Li *et al.*, 2014) and the maintenance of cell wall integrity in *A. thaliana* (Zheng *et al.*, 2019).

Table 1. Available PTM predictors based on pLMs

PTM type	Tool	Technique	Website	Reference
Crotonlyation	LMCrot	DL+pLM	https://github.com/KCLabMTU/LMCrot	Pratyush *et al.* (2024)
S-Nitrosylation	PLMSNOSite	DL+pLM embedding	https://github.com/KCLabMTU/PLMSNOSite	Pratyush *et al.* (2023)
Acetylation	TransPTM	pLM	https://github.com/TransPTM/TransPTM	Meng *et al.* (2024)
Phosphorylation (S, T, Y)	LMPhosSite	DL+pLM embedding	https://github.com/KCLabMTU/LMPhosSite	Pakhrin *et al.* (2023b)
Glycosylation	LMNglyPred	DL+pLM embedding	https://github.com/KCLabMTU/LMNglyPred	Pakhrin *et al.* (2023a)
Glycosylation	Stack-OglyPred-PLM	ANN+pLM embedding	https://github.com/PakhrinLab/Stack-OglyPred-PLM	Pakhrin *et al.* (2024)
Glycosylation	EMNGly	Supervised learning +pLM embedding	https://github.com/StellaHxy/EMNgly	Hou *et al.* (2023)
Lactylation[a]	PBertKla	DL+pLM	https://github.com/laihongyan/PBertKla; https://zenodo.org/records/15107500	Lai *et al.* (2025)
Succinylation[a]	LMSuccSite	DL+pLM	https://github.com/KCLabMTU/LMSuccSite	Pokharel *et al.* (2022)
Succinylation[a]	CBiLSuccSite	DL+pLM	https://github.com/nuinvtnu/CBiLSuccSite	Tran *et al.* (2025)
Multiple	PTM-Mamba	pLM	https://huggingface.co/ChatterjeeLab/PTM-Mamba; https://github.com/programmablebio/ptm-mamba	Peng *et al.* (2025)
Multiple	PTMGPT2	pLM	https://github.com/pallucs/PTMGPT2; https://zenodo.org/records/11362322; https://doi.org/10.5281/zenodo.11371883	Shrestha *et al.* (2024)

[a]These PTMs have not yet been described as being involved in salt tolerance.

The biological significance of PTMs and the difficulty in the experimental identification of such modifications led to the development of tools to predict PTM sites or motifs based on supervised ML completed with heuristics (see Box 5). Regarding plants, PlantNh-Kcr (Jiang *et al.*, 2024) was a DL model recently developed to predict non-histone lysine crotonylation sites in wheat, tobacco rice, peanut, and papaya based on ANN and self-attention mechanisms. The authors claim that having high sensitivity, specificity, and accuracy (all ranging from 81.0% to 83.8% in cross-validation and independent tests), the model outperformed species-specific models. Perhaps the most groundbreaking advance within the field of PTM prediction has been the application of pLMs (Table 1). For example, LMCrot (Pratyush *et al.*, 2024) is a ProtT5-based crotonylation predictor that requires embedding computations for the entire sequence context, outperforming previous computational predictors (Table 2, in Box 5) in all metrics (MCC, AUPRC, AUC-ROC, and F1-score). *S*-Nitrosylation is a non-enzymatic PTM that can also be predicted using pLMs. One is PLMSNOSite (Pratyush *et al.*, 2023), a robust ProtT5-based predictor that performs better (accuracy of 76.9%, specificity of 77.3%, MCC of 0.34) than previous AI approaches. TransPTM is also a ProtT5-based transformer that predicts non-histone acetylation sites (accuracy of 88%, MCC of 0.45) and provides a theoretical basis for developing drug targets (Meng *et al.*, 2024). PTM-Mamba (Peng *et al.*, 2025) uses ESM-2 pLM embeddings (Lin *et al.*, 2023) to predict plausible PTMs, the effects of PTMs on protein–protein interactions, and their probable role in diseases (accuracy of 83%, MCC of 0.65), providing new opportunities for modelling and designing PTM-specific protein sequences. However, the most promising tool for several PTM prediction is perhaps PTMGPT2 (Shrestha *et al.*, 2024), which is based on ProtGPT2 (Ferruz *et al.*, 2022) and can predict up to 19 PTMs (with MCCs ranging from 0.22 to 0.89), including PTMs related to salt tolerance (ubiquitinylation, SUMOylation, phosphorylation, and *S*-nitrosylation). Even without hyperparameter optimization procedures, the model obtained an average 5.45% improvement over previous PTM predictors. Concerning PTMs not yet known to be involved in salt stress (Table 3, in Box 5), PBertKla (Lai *et al.*, 2025) was developed to predict lysine lactylation by fine-tuning of ProteinBERT with peptide sequences and Gene Ontology (GO) annotations. It outperforms (accuracy of 89%, MCC of 0.61) previous prediction methods such as the DL-based methods DeepKla (accuracy of 85%, MCC of 0.52) (Lv *et al.*, 2022) and DeepKlapred (Guan *et al.*, 2024) (accuracy of 71%, MCC of 0.41), as well as the transformer-based Auto-Kla (accuracy of 72%, MCC of 0.48) (Lai and Gao, 2023). PBertKla is interesting because it may shed light on the role of lysine lactylation in plants, which has been underinvestigated but is at least critically involved in regulation during seed germination in wheat (J. Zhu *et al.*, 2023).

Box 5. PTM predictors based on computational models

The number of different PTM types is continuously increasing (> 400 types reviewed by Ramazi and Zahiri, 2021), although only a few of them are well characterized. This constant increase propelled the creation of dedicated databases [e.g. dbPTM (Lee, 2006), SysPTM (Li *et al.*, 2009), or PTMD 2.0 (X. Huang *et al.*, 2025)] and the inclusion of information about PTMs in UniProt. As a result, dbPTM includes >2243 million experimentally validated PTM sites (Chung *et al.*, 2025).

Among the first attempts to predict PTMs, FindMod (Wilkins *et al.*, 1999) was able to predict 22 different PTMs based on MS. Later on, NetPhos (Blom *et al.*, 1999) was dedicated to predict phosphorylation using ANNs, CSS-Palm for palmitoylation using clustering (ML) (Ren *et al.*, 2008), and SUMOsp for SUMOylation using iterative statistics (Xue *et al.*, 2006). ANNs were also used for *O*-glycosylation (Gupta *et al.*, 1999). MetODeep exploited DL and transfer learning to predict methionine oxidation sites (López-García *et al.*, 2019). Tools that could predict a wide variety of PTMs began to appear, such as GPS-lipid (Xie *et al.*, 2016), that predicted four lipidation classes (namely *S*-palmitoylation, *N*-myristoylation, *S*-farnesylation, and *S*-geranylgeranylation). The original MusiteDeep (Gao *et al.*, 2010) was improved by means of DL models (Wang *et al.*, 2017) to predict human phosphorylation sites, and enabled its customization to predict other types of PTM sites. CapsNet (Wang *et al.*, 2019), also based on DL, was able to predict phosphorylation, *N*-glycosylation, N^6-acetyl-lysine, methylarginine, *S*-palmitoylation, pyrrolidone-carboxylic-acid, and SUMOylation sites. It outperformed MusiteDeep and other tools when small amounts of training data were used, providing a comparable performance under large training sample conditions. CapsNet also demonstrated strong discriminant power in distinguishing kinase substrates from different kinase families. Authors published a web version for multiple PTMs (D. Wang *et al.*, 2020).

The burst of PTM prediction models prompted the review of the accuracy of the plethora of phosphorylation predictors (Trost and Kusalik, 2011) and other PTM predictors (Ramazi and Zahiri, 2021). The primary conclusion drawn from these reviews was that the early PTM predictors were influenced by the imbalance in the existing experimental datasets. As more accurate PTM predictors were published, older and less accurate tools were disappearing. The following tables gather only PTM predictors that are still available nowadays based on computational statistics (mainly ML and DL) and heuristics, but not in LLMs. Table 2 contains predictors for PTMs that are known to have a role in salt stress, while Table 3 lists predictors for PTMs that have not yet been linked to salt stress.

Table 2. List of PTM predictors for PTMs that are known to have a role in salt stress

PTM type	Tool	Technique	Website	Reference
Multiple	MusiteDeep	DL	https://www.musite.net/; https://github.com/duolinwang/MusiteDeep	D. Wang *et al.* (2020)
Multiple	CapsNet_PTM	DL	https://github.com/duolinwang/CapsNet_PTM	Wang *et al.* (2019)
Phosphorylation (K)	GPS 6.0	Transfer learning	https://gps.biocuckoo.cn/; https://github.com/BioCUCKOO/GPS6.0	Chen *et al.* (2023)
Phosphorylation (K)	NetPhosPan 1.0	DL	https://services.healthtech.dtu.dk/services/NetPhospan-1.0/	Fenoy *et al.* (2019)
Phosphorylation (S, T, Y)	NetPhos 3.1	ANN	https://services.healthtech.dtu.dk/services/NetPhos-3.1/	Blom *et al.* (2004)
Phosphorylation (S, T, Y)	DeepPhos	DL	https://github.com/USTC-HIlab/DeepPhos	F. Luo *et al.* (2019)
Phosphorylation (S, T, Y)	PhosIDN	DL	https://github.com/ustchangyuanyang/PhosIDN	Yang *et al.* (2021)
Glycosylation	NetOGlyc 4.0	ANN	https://services.healthtech.dtu.dk/services/NetOGlyc-4.0/	Hansen *et al.* (1998)
Acetylation	DeepAcet	DL	https://github.com/Lab-Xu/DeepAcet	Wu *et al.* (2019)
Acetylation	DNNAce	DL	https://github.com/QUST-AIBBDRC/DNNAce	Yu *et al.* (2020)
Ubiquitinylation	DeepUbi	DL	https://github.com/Lab-Xu/DeepUbi	Fu *et al.* (2019)
Ubiquitinylation	DeepTL-Ubi	DL+transfer learning	https://github.com/USTC-HIlab/DeepTL-Ubi	Liu *et al.* (2021)
Methylation	SSMFN	DL+ANN	https://github.com/bharuno/SSMFN-Methylation-Analysis	Lumbanraja *et al.* (2021)
Methylation	DeepRMethylSite	DL	https://github.com/dukkakc/DeepRMethylSite	Chaudhari *et al.* (2020)
Crotonylation	Deep-Kcr	DL	https://github.com/linDing-group/Deep-Kcr	Lv *et al.* (2021)
Crotonylation	DeepCap-Kcr	DL+ANN	https://github.com/Jhabindra-bioinfo/DeepCap-Kcr	Khanal *et al.* (2022)
Crotonylation	PlantNh-Kcr	DL	https://github.com/jiangyanming-individual/PlantNh-Kcr	Jiang *et al.* (2024)
SUMOylation	SUMOgo	Supervised learning	http://predictor.nchu.edu.tw/SUMOgo/	Chang *et al.* (2018)
SUMOylation	JASSA	Supervised learning	http://www.jassa.fr/	Beauclair *et al.* (2015)
SUMOylation	GPS-SUMO 2.0	DL	https://sumo.biocuckoo.cn/	Gou *et al.* (2024)
SUMOylation	SUMO-LMNet	DL		Ho *et al.* (2025)
S-Nitrosylation	PreSNO	Supervised learning	http://kurata14.bio.kyutech.ac.jp/PreSNO/prediction.php	Hasan *et al.* (2019)
S-Nitrosylation	DeepNitro	DL	http://deepnitro.renlab.org/	Xie *et al.* (2018)
Carbonylation	PreCar_Deep	DL	https://github.com/QUST-SHULI/PreCar_Deep/	Song *et al.* (2021)
Carbonylation	iCarPS	Ensemble learning	http://lin-group.cn/server/iCarPS/	D. Zhang *et al.* (2021)
S-Palmitoylation	GPS-Palm	DL	https://gpspalm.biocuckoo.cn/	Ning *et al.* (2021)

Box 5. Continued.

Table 3. List of PTM predictors for PTMs not yet associated with salt stress in the literature

PTM type	Tool	Technique	Website	Reference
Succinylation	DeepSuccinylSite	DL	https://github.com/dukkakc/DeepSuccinylSite	Thapa *et al.* (2020)
Malonylation	DeepMal	DL	https://github.com/QUST-AIBBDRC/DeepMal	M. Wang *et al.* (2020)
Malonylation	Lemp	DL+ensemble learning	https://github.com/XuhanLiu/lemp	Z. Chen *et al.* (2018)
Malonylation	RF-MaloSite & DL-MaloSite	DL+ensemble learning	https://github.com/dukkakc/DL-MaloSite-and-RF-MaloSite	Al-Barakati *et al.* (2020)
Lactylation	DeepKla	DL	http://lin-group.cn/server/DeepKla/; https://github.com/linDing-group/DeepKla	Lv *et al.* (2022)
Lactylation	Auto-Kla	DL (transformer-based)	http://origin.tubic.org/Kla/public/index.php; https://github.com/tubic/Auto-Kla	Lai and Gao (2023)
Lactylation	DeepKlapred	DL (transformer-based)	https://github.com/GGCL7/DeepKlapred	Guan *et al.* (2024)

Limitations, cautions, and risks

As developing predictive or generative models becomes accessible to most research laboratories, it is crucial to be aware of the limitations of AI in omics and plant biology to avoid flawed or misleading models.

Statistical shortcomings

To enhance comparability across AI-based predictive studies, reporting standardized metrics is required (Miller *et al.*, 2024). The following recommendations align with the best practices in the literature, which highlight the importance of comprehensive reporting of performance metrics to strengthen the robustness and interpretability of predictive modelling studies (Emmert-Streib, 2022; Miller *et al.*, 2024).

Classification tasks require reporting popular validation metrics, such as accuracy, precision, recall, F1-score, MCC, threat score (also called the critical success index), and the AUC-ROC (Hicks *et al.*, 2022). Regression analysis should report R^2, RMSE, and mean absolute error (MAE). Although not always reported, as in some studies cited in this review, MCC is one of the most suitable (Chicco and Jurman, 2020; Chicco *et al.*, 2021) since it increases when all categories in the confusion matrix reflect good results. In clinical settings, a complex combined metric is proposed to compare more holistically the performance of different models (Park *et al.*, 2023); perhaps plant research models could benefit from using this combined metric in addition to MCC. In limited or imbalanced datasets, which are common in plant stress experiments, reliance on accuracy alone can be misleading, while the AUPRC, the MCC, or the F1-score perform an honest performance assessment (Bland and Altman, 2015). Metrics to evaluate LLMs and other generative models are barely standardized and tend to be obscure and also misleading (Abbasian *et al.*, 2024). Moreover, these metrics, being grouped into four categories (accuracy, trustworthiness, empathy, and computing performance), are difficult to develop and implement yet.

Transparent reporting of the validation strategy is critical to avoid overestimation of model performance. Internal validation is always required in omics data through cross-validation or bootstrapping on the training dataset. External validation is highly recommended since it evaluates the model on an entirely independent dataset that was not used during training or model selection (Molinaro *et al.*, 2005). In fact, external validation provides a more rigorous assessment of how generalizable a model is to new populations or settings, but was not performed in some of the models cited in the review. Probably, the lack of external validation can explain why DL models that have successfully identified many crop diseases with performances at ~73–99.5% decrease the performance to 25–35% when external datasets are tested (Danilevicz *et al.*, 2025).

High-dimensional omics data

Training data in omics analyses should generally scale from hundreds to low thousands of samples for classical ML, but only a few tens are usually available. The DL requirements are more difficult to fulfil, as many thousands of samples are needed, unless strong regularization, feature selection, or transfer/self-supervised pre-training is used. A pragmatic starting point for supervised ML is to ensure adequate information per predictor ($n>>p$, where n is the sample size—independent observations—and p is the number of predictors—variables associated with each independent observation) and to combine penalization with dimensionality reduction to control optimism. Worryingly, in plant salt stress studies, omics data rely on very small cohorts (n often <20) containing an exceptionally large number of input variables (p is often in the range of several thousand to tens of thousands of genes, proteins, or other features). In other words, each sample has high dimensional data. This pronounced disproportion ($p>>n$)

poses major computational challenges during model training, overfitting issues, noise overload, model instability, low statistical power, and inflated false discovery (type I errors) (Guyon and Elisseeff, 2003; Konietschke *et al.*, 2021). As a result, ML models using high-dimensional omics datasets usually perform well on training data but perform poorly on new data (overfitting issue), indicating that innovative and efficient methodologies to extract meaningful insights are required (Morabito *et al.*, 2025). Nowadays, this drawback is mitigated with feature selection or dimensionality reduction (e.g. PCA, *t*-SNE, and autoencoders) (Han *et al.*, 2025), and extensive cross-validation (e.g. bootstrapping and leave-one-out), but reports such as the prediction of crop diseases mentioned above (Danilevicz *et al.*, 2025) add the dimensionality issue to the lack of external validation. For DL models, and particularly LLMs, it has been demonstrated that the model size and the model capacity (e.g. number of tokens) should be scaled equally to achieve optimal performance, suggesting that many LLMs are significantly undertrained (Hoffmann *et al.*, 2022, Preprint). This can explain the promising results reported for PTMGPT2 (Shrestha *et al.*, 2024) since this discards PTMs with <500 data points. The case of DL models failing to predict crop diseases on external data mentioned above (Danilevicz *et al.*, 2025) was also suffering a problem of dimensionality (far more predictors than samples).

When omics data for plant stress are analysed with AI, most ML and DL studies are 'self-validated', meaning that model predictability is tested only using a cross-validation scheme rather than by applying the trained model to an external dataset for validation (Yan and Wang, 2023). This means that n is very limited (tens to, in better cases, a few hundred samples), suggesting that only ML methods are applicable, preferably with robust feature selection. Alternatively, pre-training on large corpora and transfer learning are possible (Hoffmann *et al.*, 2022). Essentially, DL approaches should be reserved for settings with larger n or where self-supervised pre-training can amortize data needs, always validating with stratified cross-validation and external cohorts to detect residual overfitting and dimensionality effects (Dai *et al.*, 2025).

Transparency and interpretability

It is well known that interpreting an AI model is a trade-off between model accuracy and interpretability: simple models (e.g. linear regression or decision trees) are more interpretable but less accurate. Complex AI models, being able to accurately predict a wide variety of phenomena and generate information from unseen data, are often opaque, like black-boxes, which makes it difficult to understand the reasoning behind their decisions (Karim *et al.*, 2023; Steyvers *et al.*, 2025). Explainable AI (XAI) has recently received considerable attention to provide transparency and obtain insights into cellular processes, since for genomics researchers, explanatory information would frequently be of greater value than the predictions themselves, as it can lead to new hypothesis and validate biological findings (Novakovsky *et al.*, 2023). Particularly interesting is the XAI applied to understanding plant phenotypes after high-throughput phenotyping (Danilevicz *et al.*, 2025). It is expected that the increase of interpretable AI models would overcome the reluctance of researchers to trust AI predictions, even if they do not completely understand the underlying mechanisms (Karim *et al.*, 2023; Novakovsky *et al.*, 2023).

Lack of repeatability and reproducibility leads to spurious discoveries

Establishing a transparent and comprehensive reporting framework would facilitate reproducibility and allow meaningful benchmarking across diverse studies in plant stress research. In many cases, the assumptions underlying statistical models and performance evaluations do not always hold true (Whalen *et al.*, 2022), and flaws in the design or data collection process may go unnoticed. As a result, some published reports lack sufficient details to allow independent reproduction of the results (Collins *et al.*, 2024), even though they passed the peer review process. More concerning is that some of these analyses produce spurious discoveries (Box 4) due to the vast array of possibilities that can be selected during high-dimensional multi-omics analyses (Fan and Zhou, 2016). This concern would be easily mitigated with a stringent testing framework, transparent reporting, and the complete set of statistic metrics (Emmert-Streib, 2022).

We should be aware that general-purpose LLMs available nowadays have not been published in peer-reviewed journals (with the recent exception of DeepSeek R1; D. Guo *et al.*, 2025), and their performance has not been independently validated. However, they are widely downloaded, used, and tested without the source code and training data. Only some specialized LLMs such as Med-PaLM (Singhal *et al.*, 2023), and the pLMs and gLMs (Box 3) mentioned in this review, are published after a peer review process. These articles are evidence that scientific publications do not harm intellectual properties but increase the confidence in the models. Unfortunately, spurious discoveries are so widespread that the International Society for Computational Biology (ISCB) has decided to publish 'good practices' rules for the use of LLMs in science and in computational biology (https://www.iscb.org/iscb-policy-statements/iscb-policy-for-acceptable-use-of-large-language-models, accessed 30 April 2025). The take-home message is that inadequate reports containing biased benchmarks that show a model in the top position are difficult to detect. The minimal quality criteria and primary weaknesses in the use and publication of AI-based omics analyses are summarized in Box 4.

Hallucinations

While transformers excel in understanding context, they still struggle with tasks requiring deep reasoning and commonsense knowledge. Hence, accuracy and reliability of generative

AI models, particularly LLMs, is an important issue since they generate plausible but incorrect information. They are called hallucinations and are especially problematic when accuracy and reproducibility are critical (L. Huang *et al.*, 2025; Steyvers *et al.*, 2025). Hallucinations occur because the training and evaluation procedures reward guessing and penalize uncertain responses (Kalai *et al.*, 2025, Preprint). If incorrect statements cannot be distinguished from facts, then hallucinations in pre-trained models will arise through statistical pressures and will persist due to the way most evaluations are graded. To surpass specialist models and avoid introducing additional hallucination evaluation, the incorporation of uncertainty estimates into the training and evaluation of AI models has recently been proposed to mitigate hallucinations (Kalai *et al.*, 2025, Preprint). In parallel, retrieval-augmented generation (RAG) has emerged as a promising strategy to reduce hallucinations, since grounding model outputs in curated, domain-specific, knowledge sources improves factual consistency and reproducibility (Lewis *et al.*, 2021, Preprint; R. Zhang *et al.*, 2025). However, other experts argue that hallucinations are an inevitable feature of LLMs due to their statistical nature: they predict the next word based on patterns rather than verified facts.

Specialist models were the first attempt to mitigate hallucinations, incorrect contents, and the plant 'blindnesses' mentioned above (Chen *et al.*, 2024; Simon *et al.*, 2024). Despite the promising results, the detection of hallucinated outputs in genomic and protein language models (Box 3) remains largely unsolved: LLMs disclose unknown findings difficult to evaluate (Rissom *et al.*, 2025, Preprint; Steyvers *et al.*, 2025). Therefore, the implementation of RAG strategies and the development of new methods for analysing, interpreting, and validating the features and predictions of biological LLMs are required.

Environmental impact

The carbon footprint and greenhouse gas emissions of training and using AI models should encourage the sustainable use of AI. It has been estimated that a typical life sciences laboratory probably generates >20 t of CO_2 annually, annual meetings generate up to 22 000 t of CO_2 per association, while a single phylogenetic analysis using >300 000 CPU hours can generate up to 4.3 t of CO_2 (Grealey *et al.*, 2022). Generative AI models have become ubiquitous with a carbon footprint three orders of magnitude above: they emit 10.67–102.6 Mt of CO_2 per year, and are expected to grow by 30–40% annually over the next decade, producing 18.21–245 Mt of CO_2 by 2035, with average energy consumption ranging from 405 TWh to 1050 TWh (Bashir *et al.*, 2024; Y. Yu *et al.*, 2024; Ding *et al.*, 2025). As the amount of electricity consumed by AI models is fixed, reducing their carbon footprint requires the use of renewable energy sources. Hence, to ensure the sustainable development of AI and minimize its environmental impact, research institutes, governments, and industries should prioritize optimizing their data centres, advancing renewable energy, and reducing the cost of AI training and usage.

Ethical issues

Being in its infancy, AI is rapidly transforming plant research. However, AI is not expected to replace human intelligence or become independent in the near future. Ethics in AI is not about coding politeness into machines but rather about considering what kind of society we want to build when AI-powered machines begin to make decisions that were previously made by humans (Samara, 2024). The potential for the harmful misuse of AI requires constant vigilance to ensure that these technologies benefit society as a whole. Ethical use of AI remains largely experimental and focused on technical aspects due to the predominance of computer scientists over ethicists in the field. The question of who is responsible if an AI model makes a wrong prediction resulting in crop failure or environmental harm or discrimination in crop choices remains unanswered, as does the question of how to make AI behave ethically without human involvement (Flatås *et al.*, 2025, Preprint). To guide responsible AI development and use, multidisciplinary teams should be set up for the development of ethical guidelines, such as the one mentioned above from the ISCB, the UNESCO's Recommendation on the Ethics of Artificial Intelligence (https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence; accessed 24 September 2025), or the the EU AI Act published in the Official Journal on 12 July 2024 (https://artificialintelligenceact.eu/the-act/; accessed 24 September 2025). Europe's approach is unique because it puts people first, aiming to develop AI that respects human rights, democracy, and the rule of law. Ethical guidelines should address not only data privacy, sharing, and ownership, but also the responsible use of AI in plant research (Walsh *et al.*, 2024).

Other weakenesses

The impressive advances in AI outlined here suggest that the following issues should be given more attention. (i) Abiotic stress displays more subtle impacts on traits than diseases, demanding a more complex examination of data to reliably interpret these responses. (ii) Besides the increasing amount of source data for AI, there is a lack of large, high-quality, and well-annotated datasets for training and validation of AI models for non-model plants and salt stress (Zarbakhsh and Shahsavar, 2022; Yan and Wang, 2023; Murmu *et al.*, 2024; Walsh *et al.*, 2024). (iii) More robust and user-friendly bioinformatic methods and tools are required for multi-omics data integration and harmonization when high-dimensional omics data are used (Morabito *et al.*, 2025; Zou and Xu, 2025). (iv) Related to the previous issue, researchers must stay up-to-date with the latest methods and tools to effectively utilize AI technologies (Walsh *et al.*, 2024). For example, the mastery of prompt engineering is becoming a key skill for developers and data

scientists working with LLMs, as it influences both the quality of the generated text and the task-specific efficiency of models. (v) The scarce transfer of AI models to non-model organisms, other stress conditions, or different environmental contexts slows the entry of plant research into the '5Gs' generation (Genome assembly, Germplasm characterization, Gene function identification, Genomic breeding, and Gene editing) (Varshney *et al.*, 2020; Yan and Wang, 2023).

Future application: plant breeding

It can be deduced in this review, and is widely accepted, that ML-optimized frameworks are suitable to reveal novel salt tolerance genes and regulatory pathways, DL models are superior in classifying stress levels, and LLMs excel in improving the detection of protein motifs and the plant genome structures governing gene expression. Additionally, agricultural traits such as salt tolerance are the product of gene synergisms, developmental stage, and environment, where non-destructive remote sensing is becoming indispensable (Xu *et al.*, 2021; Rai, 2022). Hence, AI-driven technologies would pave the way for developing and identifying resilient crop varieties by integrating genotyping (Li *et al.*, 2025), phenotyping, physiology, and multi-omics (Cheng and Wang, 2024; Raza *et al.*, 2025b), or to predict crop yields under stressed conditions (Choi *et al.*, 2025, Preprint), or understanding phenotypes in crop breeding (Danilevicz *et al.*, 2025). In fact, plant breeding can take advantage of AI to (i) identify and predict the most relevant genes and alleles for salt tolerance, (ii) expedite the discovery of agronomically useful genes and mutations, (iii) transform this biological knowledge into precision-designed, knowledge-driven, molecularly drawn plant breeding in the near future, and (iv) provide biologically meaningful interpretations of phenotypes and salt stress responses in plants, which would accelerate crop improvement programmes (Khan *et al.*, 2022; Rivero *et al.*, 2022; Yan and Wang, 2023; Danilevicz *et al.*, 2025; Saleem *et al.*, 2025).

The AI-driven analyses integrating many crop genomes, salt-tolerant individuals, qualitative trait loci related to salt tolerance, and even intelligent cyber-agricultural platforms at field and plant levels has the potential to build agro-climatic models to foresee salt tolerance (Sarkar *et al.*, 2024; Raza et al., 2025a , b; H. Wu *et al.*, 2025), to forecast genomic crossovers to allow the identification of genomic regions with elevated mutation rates (Farooq *et al.*, 2024), and to predict yield, planting times, and conditions, and detect diseases and weeds (Q.-C. Li *et al.*, 2023). Many traits, genes, or alleles have been proposed as candidates for plant breeding salt-tolerant crops (Roychowdhury *et al.*, 2023; Koh *et al.*, 2024b, Preprint), but supporting results are still awaited (Saleem *et al.*, 2025). An early example of this approach has been recently published for drought stress, where the IDSDS pipeline (Maji *et al.*, 2025) combines remote sensing with AI to support both scientific research and practical crop management. In conclusion, the development of commercial, yield-stable, salt-tolerant cultivars with a significant reduction of time and costs in large-scale breeding efforts no longer belongs to the realm of science fiction any more thanks to AI.

Author contributions

MGC: conceived the review; MGC, FJV, NFP, AT, and JSR: wrote and revised the manuscript. All authors read and approved the final manuscript.

Conflict of interest

The authors declare no conflict of interest.

Funding

This research was funded by Ministerio de Ciencia e Innovación and Agencia Española de Investigación from Spain, co-funded with European Regional Development Funds from European Union (grants TED2021-130015B-C21, PID2020-113324GB-I00, PID2020-116898RB-I00, RED2022-134072-T, and the Computational Biology and Bioinformatics Connection 2304210057) to MGC, FJV, and JSR. NFP is grateful for funding by Ministerio de Ciencia e Innovación and Agencia Española de Investigación from Spain and by 'ERDF/EU' (PID2021-125805OA-I00), #NextGenerationEU from the Spanish Plan de Recuperación, Transformación y Resiliencia (CNS2023-144643), Consejo Superior de Investigaciones Científicas (CSIC) (20224AT004), and by the #NextGenerationEU funds from the European Union, through the Momentum CSIC Programme: Develop Your Digital Talent (MMT24-IHSM-01). AT was hired under the Generation D initiative, promoted by Red.es, an organization attached to the Ministry for Digital Transformation and the Civil Service, for the attraction and retention of talent through grants and training contracts, financed by the #NextGenerationEU funds from the Recovery, Transformation and Resilience Plan of the European Union (MMT24-IHSM-01). Funding for open access charge: Universidad de Málaga/CBUA.

Data availability

This review does not contribute any new data.

References

Abbasian M, Khatibi E, Azimi I, *et al.* 2024. Foundation metrics for evaluating effectiveness of healthcare conversations powered by generative AI. NPJ Digital Medicine **7**, 82.

Abd Elaziz AEA, Soliman KG, Abdel Hamed EMW, Metwally MS, Abu-Hashim MSD. 2024. Improving soil salinity prediction in semi-arid areas using machine learning models. Zagazig Journal of Agricultural Research **51**, 505–517.

Abdelraheem A, Thyssen GN, Fang DD, Jenkins JN, McCarty JC, Wedegaertner T, Zhang J. 2021. GWAS reveals consistent QTL for drought and salt tolerance in a MAGIC population of 550 lines derived from intermating of 11 upland cotton (*Gossypium hirsutum*) parents. Molecular Genetics and Genomics **296**, 119–129.

Abinaya S, Kumar KU, Alphonse AS. 2023. Cascading autoencoder with attention residual u-net for multi-class plant leaf disease segmentation and classification. IEEE Access **11**, 98153–98170.

Ahmar S, Usman B, Hensel G, Jung K-H, Gruszka D. 2024. CRISPR enables sustainable cereal production for a greener future. Trends in Plant Science **29**, 179–195.

Ahmed B, Haque MA, Iquebal MA, Jaiswal S, Angadi UB, Kumar D, Rai A. 2023. DeepAProt: deep learning based abiotic stress protein sequence classification and identification tool in cereals. Frontiers in Plant Science **13**, 1008756.

Akbari M, Sabouri H, Sajadi SJ, Yarahmadi S, Ahangar L. 2024. Classification and prediction of drought and salinity stress tolerance in barley using GenPhenML. Scientific Reports **14**, 17420.

Al-Barakati H, Thapa N, Hiroto S, Roy K, Newman RH, Kc D. 2020. RF-MaloSite and DL-Malosite: methods based on random forest and deep learning to identify malonylation sites. Computational and Structural Biotechnology Journal **18**, 852–860.

Alhnaity B, Pearson S, Leontidis G, Kollias S. 2020. Using deep learning to predict plant growth and yield in greenhouse environments. Acta Horticulturae **1296**, 425–432.

Al-Tamimi N, Langan P, Bernád V, Walsh J, Mangina E, Negrão S. 2022. Capturing crop adaptation to abiotic stress using image-based technologies. Open Biology **12**, 210353.

An P, Wang W, Chen X, Zhuang Z, Cui L. 2023. Machine learning brings new insights for reducing salinization disaster. Frontiers in Earth Science **11**, 1130070.

Aponte-Rengifo O, Francisco M, Vilanova R, Vega P, Revollar S. 2023. Intelligent control of wastewater treatment plants based on model-free deep reinforcement learning. Processes **11**, 2269.

Arefian M, Vessal S, Malekzadeh-Shafaroudi S, Siddique KHM, Bagheri A. 2019. Comparative proteomics and gene expression analyses revealed responsive proteins and mechanisms for salt tolerance in chickpea genotypes. BMC Plant Biology **19**, 300.

Arivazhagan N, Van Vleck TT. 2023. Natural language processing basics. Clinical Journal of the American Society of Nephrology **18**, 400–401.

Asins MJ, Bullones A, Raga V, *et al.* 2023. Combining genetic and transcriptomic approaches to identify transporter-coding genes as likely responsible for a repeatable salt tolerance QTL in citrus. International Journal of Molecular Sciences **24**, 15759.

Asnicar F, Thomas AM, Passerini A, Waldron L, Segata N. 2024. Machine learning for microbiologists. Nature Reviews. Microbiology **22**, 191–205.

Atta K, Mondal S, Gorai S, *et al.* 2023. Impacts of salinity stress on crop plants: improving salt tolerance through genetic and molecular dissection. Frontiers in Plant Science **14**, 1241736.

Avsec Ž, Latysheva N, Cheng J, *et al.* 2025. AlphaGenome: advancing regulatory variant effect prediction with a unified DNA sequence model. bioRxiv doi: 10.1101/2025.06.25.661532. [Preprint]

Awlia M, Alshareef N, Saber N, Korte A, Oakey H, Panzarová K, Trtílek M, Negrão S, Tester M, Julkowska MM. 2021. Genetic mapping of the early responses to salt stress in *Arabidopsis thaliana*. The Plant Journal **107**, 544–563.

Bakshi S. 2024. Genome-wide association studies: unraveling the genetic basis of complex traits. International Research Journal of Plant Science **15**, 34.

Balasubramaniam T, Shen G, Esmaeili N, Zhang H. 2023. Plants' response mechanisms to salinity stress. Plants **12**, 2253.

Bandi A, Adapa PVSR, Kuchi YEVPK. 2023. The power of generative AI: a review of requirements, models, input–output formats, evaluation metrics, and challenges. Future Internet **15**, 260.

Barrios-Núñez I, Martínez-Redondo GI, Medina-Burgos P, Cases I, Fernández R, Rojas AM. 2024. Decoding functional proteome information in model organisms using protein language models. NAR Genomics and Bioinformatics **6**, lqae078.

Bashir N, Donti P, Cuff J, Sroka S, Ilic M, Sze V, Delimitrou C, Olivetti E. 2024. The climate and sustainability implications of generative AI. An MIT Exploration of Generative AI. Cambridge, MA: MIT.

Beauclair G, Bridier-Nahmias A, Zagury J-F, Saïb A, Zamborlini A. 2015. JASSA: a comprehensive tool for prediction of SUMOylation sites and SIMs. Bioinformatics **31**, 3483–3491.

Benegas G, Batra SS, Song YS. 2023. DNA language models are powerful predictors of genome-wide variant effects. Proceedings of the National Academy of Sciences, USA **120**, e2311219120.

Benegas G, Yke C, Albors C, Li JC, Song YS. 2025. Genomic language models: opportunities and challenges. Trends in Genetics **41**, 286–302.

Bengesi S, El-Sayed H, Sarker MK, Houkpati Y, Irungu J, Oladunni T. 2024. Advancements in generative AI: a comprehensive review of GANs, GPT, autoencoders, diffusion model, and transformers. IEEE Access **12**, 69812–69837.

Benjamin JJ, Miras-Moreno B, Araniti F, Salehi H, Bernardo L, Parida A, Lucini L. 2020. Proteomics revealed distinct responses to salinity between the halophytes *Suaeda maritima* (L.) Dumort and *Salicornia brachiata* (Roxb). Plants **9**, 227.

Bermingham ML, Pong-Wong R, Spiliopoulou A, *et al.* 2015. Application of high-dimensional feature selection: evaluation for genomic prediction in man. Scientific Reports **5**, 10312.

Bhoite R, Onyemaobi O, Halder T, Shankar M, Sharma D. 2025. Transcription factors–insights into abiotic and biotic stress resilience and crop improvement. Current Plant Biology **41**, 100434.

Bittencourt CB, Carvalho da Silva TL, Rodrigues Neto JC, Vieira LR, Leão AP, de Aquino Ribeiro JA, Abdelnur PV, de Sousa CAF, Souza MT Jr. 2022. Insights from a multi-omics integration (MOI) study in oil palm (*Elaeis guineensis* Jacq.) Response to abiotic stresses: part one–salinity. Plants **11**, 1755.

Bland JM, Altman DG. 2015. Statistics notes: bootstrap resampling methods. BMJ **350**, h2622.

Blom N, Gammeltoft S, Brunak S. 1999. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. Journal of Molecular Biology **294**, 1351–1362.

Blom N, Sicheritz-Pontén T, Gupta R, Gammeltoft S, Brunak S. 2004. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. PROTEOMICS **4**, 1633–1649.

Brandes N, Ofer D, Peleg Y, Rappoport N, Linial M. 2022. ProteinBERT: a universal deep-learning model of protein sequence and function. Bioinformatics **38**, 2102–2110.

Brink BG, Seidel A, Kleinbölting N, Nattkemper TW, Albaum SP. 2016. Omics fusion–a platform for integrative analysis of omics data. Journal of Integrative Bioinformatics **13**, 296.

Bruetschy C. 2019. The EU regulatory framework on genetically modified organisms (GMOs). Transgenic Research **28**, 169–174.

Campbell MT, Knecht AC, Berger B, Brien CJ, Wang D, Walia H. 2015. Integrating image-based phenomics and association analysis to dissect the genetic architecture of temporal salinity responses in rice. Plant Physiology **168**, 1476–1489.

Cao D, Zhang W, Yang N, Li Z, Zhang C, Wang D, Ye G, Chen J, Wei X. 2023. Proteomic and metabolomic analyses uncover integrative mechanisms in *Sesuvium portulacastrum* tolerance to salt stress. Frontiers in Plant Science **14**, 1277762.

Capel C, Albaladejo I, Egea I, *et al.* 2020. The *res* (*restored cell structure by salinity*) tomato mutant reveals the role of the DEAD-box RNA helicase SIDEAD39 in plant development and salt response. Plant, Cell & Environment **43**, 1722–1739.

Chang C-C, Tung C-H, Chen C-W, Tu C-H, Chu Y-W. 2018. SUMOgo: prediction of sumoylation sites on lysines by motif screening models and the effects of various post-translational modifications. Scientific Reports **8**, 15512.

Chaudhari M, Thapa N, Roy K, Newman RH, Saigo H BKCD. 2020. DeepRMethylSite: a deep learning based approach for prediction of arginine methylation sites in proteins. Molecular Omics **16**, 448–454.

Chen D, Shi R, Pape J-M, Neumann K, Arend D, Graner A, Chen M, Klukas C. 2018. Predicting plant biomass accumulation from image-derived parameters. GigaScience **7**, 1–13.

Chen M, Zhang W, Gou Y, *et al.* 2023. GPS 6.0: an updated server for prediction of kinase-specific phosphorylation sites in proteins. Nucleic Acids Research **51**, W243–W250.

Chen S, Zhang H, Gao S, He K, Yu T, Gao S, Wang J, Li H. 2025. Unveiling salt tolerance mechanisms in plants: integrating the KANMB machine learning model with metabolomic and transcriptomic analysis. Advanced Science **12**, e2417560.

Chen X, Chen Z, Watts R, Luo H. 2025. Non-coding RNAs in plant stress responses: molecular insights and agricultural applications. Plant Biotechnology Journal **23**, 3195–3233.

Chen Z, He N, Huang Y, Qin WT, Liu X, Li L. 2018. Integration of a deep learning classifier with a random forest approach for predicting malonylation sites. Genomics, Proteomics & Bioinformatics **16**, 451–459.

Chen Z, Wei L, Gao G. 2024. Foundation models for bioinformatics. Quantitative Biology **12**, 339–344.

Cheng Q, Wang X. 2024. Machine learning for AI breeding in plants. Genomics, Proteomics & Bioinformatics **22**, qzae051.

Chicco D, Jurman G. 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics **21**, 6.

Chicco D, Tötsch N, Jurman G. 2021. The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. BioData Mining **14**, 13.

Cho A, Kim GC, Karpekov A, Helbling A, Wang ZJ, Lee S, Hoover B, Chau DH. 2024. Transformer explainer: interactive learning of text-generative models. arXiv doi: 10.48550/arXiv.2408.04619. [Preprint].

Choi JW, Hidayat MS, Cho SB, Hwang W-H, Lee H, Cho B-K, Kim MS, Baek I, Kim G. 2025. Recent trends in machine learning, deep learning, ensemble learning, and explainable artificial intelligence techniques for evaluating crop yields under abnormal climate conditions. Plants **14**, 2841.

Chung C-R, Tang Y, Chiu Y-P, *et al.* 2025. dbPTM 2025 update: comprehensive integration of PTMs and proteomic data for advanced insights into cancer research. Nucleic Acids Research **53**, D377–D386.

Claros MG, Bautista R, Guerrero-Fernández D, Benzerki H, Seoane P, Fernández-Pozo N. 2012. Why assembling plant genome sequences is so challenging. Biology **1**, 439–459.

Claros MG, Bullones A, Castro AJ, *et al.* 2025. Multi-omic advances in olive tree (*Olea europaea* subsp. *europaea* L.) under salinity: stepping towards 'smart oliviculture'. Biology **14**, 287.

Clough E, Barrett T, Wilhite SE, *et al.* 2024. NCBI GEO: archive for gene expression and epigenomics data sets: 23-year update. Nucleic Acids Research **52**, D138–D144.

Cochetel N, Minio A, Guarracino A, *et al.* 2023. A super-pangenome of the north American wild grape species. Genome Biology **24**, 290.

Collins GS, Moons KGM, Dhiman P, *et al.* 2024. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ **385**, e078378.

Contreras-Moreira B, Naamati G, Rosello M, Allen JE, Hunt SE, Muffato M, Gall A, Flicek P. 2022. Scripting analyses of genomes in Ensembl Plants. Methods in Molecular Biology **2443**, 27–55.

Cui H, Tejada-Lapuerta A, Brbić M, Saez-Rodriguez J, Cristea S, Goodarzi H, Lotfollahi M, Theis FJ, Wang B. 2025. Towards multimodal foundation models in molecular cell biology. Nature **640**, 623–633.

Dag A, Ben-Gal A, Goldberger S, Yermiyahu U, Zipori I, Or E, David I, Netzer Y, Kerem Z. 2015. Sodium and chloride distribution in grapevines as a function of rootstock and irrigation water salinity. American Journal of Enology and Viticulture **66**, 80–84.

Dai Y, Wu D, Carroll I, Zou F, Zou B. 2025. High-dimensional biomarker identification for interpretable disease prediction via machine learning models. Bioinformatics **41**, btaf266.

Danilevicz MF, Upadhyaya SR, Batley J, Bennamoun M, Bayer PE, Edwards D. 2025. Understanding plant phenotypes in crop breeding through explainable AI. Plant Biotechnology Journal **23**, 4200–4213.

Dargan S, Kumar M, Ayyagari MR, Kumar G. 2020. A survey of deep learning and its applications: a new paradigm to machine learning. Archives of Computational Methods in Engineering **27**, 1071–1092.

Dassanayake M, Oh D-H, Haas JS, *et al.* 2011. The genome of the extremophile crucifer *Thellungiella parvula*. Nature Genetics **43**, 913–918.

Del Cioppo G, Scalabrino S, Scippa GS, Trupiano D. 2024. Opportunities and limits of image-based plant stress phenotyping: detecting plant salt stress status using machine learning techniques. Botanical Journal of the Linnean Society **207**, 253–265.

Deng Y, Xin N, Zhao L, Shi H, Deng L, Han Z, Wu G. 2024. Precision detection of salt stress in soybean seedlings based on deep learning and chlorophyll fluorescence imaging. Plants **13**, 2089.

Di Nisio A, Adamo F, Acciani G, Attivissimo F. 2020. Fast detection of olive trees affected by *Xylella fastidiosa* from UAVs using multispectral imaging. Sensors **20**, 4915.

Díaz-Rueda P, Franco-Navarro JD, Messora R, *et al.* 2020. SILVOLIVE, a germplasm collection of wild subspecies with high genetic variability as a source of rootstocks and resistance genes for olive breeding. Frontiers in Plant Science **11**, 629.

Ding Z, Wang J, Song Y, Zheng X, He G, Chen X, Zhang T, Lee W-J, Song J. 2025. Tracking the carbon footprint of global generative artificial intelligence. The Innovation **6**, 100866.

Dobránszki J, Vassileva V, Agius DR, Moschou PN, Gallusci P, Berger MMJ, Farkas D, Basso MF, Martinelli F. 2025. Gaining insights into epigenetic memories through artificial intelligence and omics science in plants. Journal of Integrative Plant Biology **67**, 2320–2349.

Dwivedi YK, Hughes L, Ismagilova E, *et al.* 2021. Artificial intelligence (AI): multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management **57**, 101994.

Ellegren H. 2014. Genome sequencing and population genomics in non-model organisms. Trends in Ecology & Evolution **29**, 51–63.

Elnaggar A, Heinzinger M, Dallago C, *et al.* 2022. ProtTrans: toward understanding the language of life through self-supervised learning. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**, 7112–7127.

Emmert-Streib F. 2022. Severe testing with high-dimensional omics data for enhancing biomedical scientific discovery. NPJ Systems Biology and Applications **8**, 40.

Eraslan G, Avsec Ž, Gagneur J, Theis FJ. 2019. Deep learning: new computational modelling techniques for genomics. Nature Reviews. Genetics **20**, 389–403.

Ersöz T, Ersöz F. 2022. Data mining and machine learning approaches in data science: predictive modeling of traffic accident causes. International Journal of 3D Printing Technologies and Digital Industry **6**, 530–539.

Fan J, Zhou W-X. 2016. Guarding against spurious discoveries in high dimensions. Journal of Machine Learning Research **17**, 1–34.

Farooq MA, Gao S, Hassan MA, Huang Z, Rasheed A, Hearne S, Prasanna B, Li X, Li H. 2024. Artificial intelligence in plant breeding. Trends in Genetics **40**, 891–908.

Feng X, Zhan Y, Wang Q, Yang X, Yu C, Wang H, Tang Z, Jiang D, Peng C, He Y. 2020. Hyperspectral imaging combined with machine learning as a tool to obtain high-throughput plant salt-stress phenotyping. The Plant Journal **101**, 1448–1461.

Fenoy E, Izarzugaza JMG, Jurtz V, Brunak S, Nielsen M. 2019. A generic deep convolutional neural network framework for prediction of receptor–ligand interactions–NetPhosPan: application to kinase phosphorylation prediction. Bioinformatics **35**, 1098–1107.

Fernandez-Pozo N, Menda N, Edwards JD, *et al.* 2015. The Sol Genomics Network (SGN)–from genotype to phenotype to breeding. Nucleic Acids Research **43**, D1036–D1041.

Ferruz N, Schmidt S, Höcker B. 2022. ProtGPT2 is a deep unsupervised language model for protein design. Nature Communications **13**, 4348.

Flatås BA, Nordsteien A, Eide H, Lysdahl KB, Sanchez VG, Stendal K, Turk E, Eide T. 2025. Building ethics into artificial intelligence: a cross-disciplinary systematic review. SJIS Preprints https://aisel.aisnet.org/sjis_preprints/12/. [Preprint].

Frukh A, Siddiqi TO, Khan MIR, Ahmad A. 2020. Modulation in growth, biochemical attributes and proteome profile of rice cultivars under salt stress. Plant Physiology and Biochemistry: PPB **146**, 55–70.

Fu H, Yang Y, Wang X, Wang H, Xu Y. 2019. DeepUbi: a deep learning framework for prediction of ubiquitination sites in proteins. BMC Bioinformatics **20**, 86.

Gachloo M, Wang Y, Xia J. 2019. A review of drug knowledge discovery using BioNLP and tensor or matrix decomposition. Genomics & Informatics **17**, e18.

Gao J, Thelen JJ, Dunker AK, Xu D. 2010. Musite, a tool for global prediction of general and kinase-specific phosphorylation sites. Molecular & Cellular Proteomics **9**, 2586–2600.

Garg V, Bohra A, Mascher M, Spannagl M, Xu X, Bevan MW, Bennetzen JL, Varshney RK. 2024. Unlocking plant genetics with telomere-to-telomere genome assemblies. Nature Genetics **56**, 1788–1799.

Geitmann A, Bidhendi AJ. 2023. Plant blindness and diversity in AI language models. Trends in Plant Science **28**, 1095–1097.

Gholizadeh F, Mirmazloum I, Janda T. 2024. Genome-wide identification of *HKT* gene family in wheat (*Triticum aestivum* L.): insights from the expression of multiple genes (*HKT*, *SOS*, *TVP* and *NHX*) under salt stress. Plant Stress **13**, 100539.

Goldenits G, Mallinger K, Raubitzek S, Neubauer T. 2024. Current applications and potential future directions of reinforcement learning-based digital twins in agriculture. Smart Agricultural Technology **8**, 100512.

Goodstein DM, Shu S, Howson R, *et al.* 2012. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research **40**, D1178–D1186.

Gou Y, Liu D, Chen M, *et al.* 2024. GPS-SUMO 2.0: an updated online service for the prediction of SUMOylation sites and SUMO-interacting motifs. Nucleic Acids Research **52**, W238–W247.

Grealey J, Lannelongue L, Saw W-Y, Marten J, Méric G, Ruiz-Carmona S, Inouye M. 2022. The carbon footprint of bioinformatics. Molecular Biology and Evolution **39**, msac034.

Greener JG, Kandathil SM, Moffat L, Jones DT. 2022. A guide to machine learning for biologists. Nature Reviews. Molecular Cell Biology **23**, 40–55.

Guan J, Xie P, Dong D, Liu Q, Zhao Z, Guo Y, Zhang Y, Lee T-Y, Yao L, Chiang Y-C. 2024. DeepKlapred: a deep learning framework for identifying protein lysine lactylation sites via multi-view feature fusion. International Journal of Biological Macromolecules **283**, 137668.

Guo D, Yang D, Zhang H, *et al.* 2025. DeepSeek-R1 incentivizes reasoning in LLMs through reinforcement learning. Nature **645**, 633–638.

Guo F, Guan R, Li Y, Liu Q, Wang X, Yang C, Wang J. 2025. Foundation models in bioinformatics. National Science Review **12**, nwaf028.

Guo Z, Liu J, Wang Y, Chen M, Wang D, Xu D, Cheng J. 2024. Diffusion models in bioinformatics and computational biology. Nature Reviews. Bioengineering **2**, 136–154.

Gupta R, Jung E, Gooley AA, Williams KL, Brunak S, Hansen J. 1999. Scanning the available *Dictyostelium discoideum* proteome for *O*-linked GlcNAc glycosylation sites using neural networks. Glycobiology **9**, 1009–1022.

Guyon I, Elisseeff A. 2003. An introduction to variable and feature selection. Journal of Machine Learning Research **3**, 1157–1182.

Hafner A, DeLeo V, Deng CH, *et al.* 2025. Data reuse in agricultural genomics research: challenges and recommendations. GigaScience **14**, giae106.

Hajiboland R, Bahrami-Rad S, Akhani H, Poschenrieder C. 2018. Salt tolerance mechanisms in three Irano-Turanian Brassicaceae halophytes relatives of *Arabidopsis thaliana*. Journal of Plant Research **131**, 1029–1046.

Han E, Kwon H, Jung I. 2025. A review on multi-omics integration for aiding study design of large scale TCGA cancer datasets. BMC Genomics **26**, 769.

Han K, Zhao Y, Sun Y, Li Y. 2023. NACs, generalist in plant life. Plant Biotechnology Journal **21**, 2433–2457.

Hansen JE, Lund O, Tolstrup N, Gooley AA, Williams KL, Brunak S. 1998. NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. Glycoconjugate Journal **15**, 115–130.

Harshvardhan GM, Gourisaria MK, Pandey M, Rautaray SS. 2020. A comprehensive survey and analysis of generative models in machine learning. Computer Science Review **38**, 100285.

Hasan M, Manavalan B, Khatun M, Kurata H. 2019. Prediction of *S*-nitrosylation sites by integrating support vector machines and random forest. Molecular Omics **15**, 451–458.

Hashiguchi A, Komatsu S. 2016. Impact of post-translational modifications of crop proteins under abiotic stress. Proteomes **4**, 42.

Hassani A, Azapagic A, Shokri N. 2021. Global predictions of primary soil salinization under changing climate in the 21st century. Nature Communications **12**, 6663.

Hayford RK, Haley OC, Cannon EK, Portwood JL 2nd, Gardiner JM, Andorf CM, Woodhouse MR. 2024. Functional annotation and meta-analysis of maize transcriptomes reveal genes involved in biotic and abiotic stress. BMC Genomics **25**, 533.

He W, Li X, Qian Q, Shang L. 2025. The developments and prospects of plant super-pangenomes: demands, approaches, and applications. Plant Communications **6**, 101230.

Hicks SA, Strümke I, Thambawita V, Hammou M, Riegler MA, Halvorsen P, Parasa S. 2022. On evaluation metrics for medical applications of artificial intelligence. Scientific Reports **12**, 5979.

Hinton GE, Salakhutdinov RR. 2006. Reducing the dimensionality of data with neural networks. Science **313**, 504–507.

Hitti Y, Buzatu I, Verme MD, Lefsrud M, Golemo F, Durand A. 2024. GrowSpace: a reinforcement learning environment for plant architecture. Computers and Electronics in Agriculture **217**, 108613.

Ho C-H, Chu Y-W, Huang L-Y, Chen C-W. 2025. SUMO-LMNet: lossless mapping network for predicting SUMOylation sites in SUMO1 and SUMO2 using high-dimensional features. Computational and Structural Biotechnology Journal **27**, 1048–1059.

Hoffmann J, Borgeaud S, Mensch A, *et al.* 2022. Training compute-optimal large language models. arXiv doi: 10.48550/arXiv.2203.15556. [Preprint].

Hou X, Wang Y, Bu D, Wang Y, Sun S. 2023. EMNGly: predicting N-linked glycosylation sites using the language models for feature extraction. Bioinformatics **39**, btad650.

Hu J, Cai J, Park SJ, Lee K, Li Y, Chen Y, Yun J-Y, Xu T, Kang H. 2021. N6-methyladenosine mRNA methylation is important for salt stress tolerance in arabidopsis. The Plant Journal **106**, 1759–1775.

Hu P, Zheng Q, Luo Q, Teng W, Li H, Li B, Li Z. 2021. Genome-wide association study of yield and related traits in common wheat under salt-stress conditions. BMC Plant Biology **21**, 27.

Hualpa-Ramirez E, Carrasco-Lozano EC, Madrid-Espinoza J, Tejos R, Ruiz-Lara S, Stange C, Norambuena L. 2024. Stress salinity in plants: new strategies to cope with in the foreseeable scenario. Plant Physiology and Biochemistry **208**, 108507.

Huang H, Shi X, Lei H, Hu F, Cai Y. 2025. ProtChat: an AI multi-agent for automated protein analysis leveraging GPT-4 and protein language model. Journal of Chemical Information and Modeling **65**, 62–70.

Huang L, Yu W, Ma W, *et al.* 2025. A survey on hallucination in large language models: principles, taxonomy, challenges, and open questions. ACM Transactions on Information Systems **43**, 1–55.

Huang X, Feng Z, Liu D, Gou Y, Chen M, Tang D, Han C, Peng J, Peng D, Xue Y. 2025. PTMD 2.0: an updated database of disease-associated post-translational modifications. Nucleic Acids Research **53**, D554–D563.

Huang Y, Cao H, Yang L, *et al.* 2019. Tissue-specific respiratory burst oxidase homolog-dependent H_2O_2 signaling to the plasma membrane H^+-ATPase confers potassium uptake and salinity tolerance in Cucurbitaceae. Journal of Experimental Botany **70**, 5879–5893.

Huang Z, Chen S, He K, Yu T, Fu J, Gao S, Li H. 2024. Exploring salt tolerance mechanisms using machine learning for transcriptomic insights: case study in *Spartina alterniflora*. Horticulture Research **11**, uhae082.

Hughes DP, Salathe M. 2016. An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv doi: 10.48550/arXiv.1511.08060. [Preprint].

Islam S, Reza MN, Samsuzzaman S, Ahmed S, Cho YJ, Noh DH, Chung S-O, Hong S. 2024. Machine vision and artificial intelligence for plant growth stress detection and monitoring: a review. Precision Agriculture Science and Technology **6**, 33–57.

Jafar A, Bibi N, Naqvi RA, Sadeghi-Niaraki A, Jeong D. 2024. Revolutionizing agriculture with artificial intelligence: plant disease detection methods, applications, and their limitations. Frontiers in Plant Science **15**, 1356260.

Javed Q, Azeem A, Sun J, Ullah I, Du D, Imran MA, Nawaz M, Chattha H. 2022. Growth prediction of *Alternanthera philoxeroides* under salt stress by application of artificial neural networking. Plant Biosystems **156**, 61–67.

Javid S, Bihamta MR, Omidi M, Abbasi AR, Alipour H, Ingvarsson PK. 2022. Genome-wide association study (GWAS) and genome prediction of seedling salt tolerance in bread wheat (*Triticum aestivum* L). BMC Plant Biology **22**, 581.

Jeevan Nagendra Kuma Y, Chandan R, Harsh Somanini S, Vadtya S, Ram Lohit Pranay Y, Mohammed KA, Chandrashekar R, Kansal L, Kalra R. 2024. Predictive modeling for enhanced plant cultivation in greenhouse environment. E3S Web of Confeferences **507**, 01066.

Jiang Y, Yan R, Wang X. 2024. PlantNh-Kcr: a deep learning model for predicting non-histone crotonylation sites in plants. Plant Methods **20**, 28.

Julian J, Gao P, Del Chiaro A, *et al.* 2025. ATG8ylation of vacuolar membrane protects plants against cell wall damage. Nature Plants **11**, 321–339.

Jurado-Ruiz F, Rousseau D, Botía JA, Aranzana MJ. 2023. GenoDrawing: an autoencoder framework for image prediction from SNP markers. Plant Phenomics **5**, 0113.

Kabiraj S, Jayanthi M, Vijayakumar S, Duraisamy M. 2022. Comparative assessment of satellite images spectral characteristics in identifying the different levels of soil salinization using machine learning techniques in Google Earth Engine. Earth Science Informatics **15**, 2275–2288.

Kalai AT, Nachum O, Vempala SS, Zhang E. 2025. Why language models hallucinate. arXiv doi: 10.48550/arXiv.2509.04664. [Preprint].

Kamilaris A, Prenafeta-Boldú FX. 2018. Deep learning in agriculture: a survey. Computers and Electronics in Agriculture **147**, 70–90.

Kang D, Ahn H, Lee S, Lee C-J, Hur J, Jung W, Kim S. 2019. StressGenePred: a twin prediction model architecture for classifying the stress types of samples and discovering stress-related genes in *Arabidopsis*. BMC Genomics **20**, 949.

Kaplan A, Haenlein M. 2019. Siri, siri, in my hand: who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Business Horizons **62**, 15–25.

Karim MR, Islam T, Shajalal M, Beyan O, Lange C, Cochez M, Rebholz-Schuhmann D, Decker S. 2023. Explainable AI for bioinformatics: methods, tools and applications. Briefings in Bioinformatics **24**, bbad236.

Kaur M, Kaur U. 2024. Deep learning-based plant stress diagnosis: an optimized generative augmentation model approach. In: Chouhan SS, Saxena A, Singh UP, Jain S, eds. Artificial intelligence techniques in smart agriculture. Singapore: Springer Nature Singapore, 115–128.

Kaya A, Keceli AS, Catal C, Yalic HY, Temucin H, Tekinerdogan B. 2019. Analysis of transfer learning for deep neural network based plant classification models. Computers and Electronics in Agriculture **158**, 20–29.

Khalifani S, Darvishzadeh R, Azad N, Seyed Rahmani R. 2022. Prediction of sunflower grain yield under normal and salinity stress by RBF, MLP and, CNN models. Industrial Crops and Products **189**, 115762.

Khan AA, Chaudhari O, Chandra R. 2024. A review of ensemble learning and data augmentation models for class imbalanced problems: combination, implementation and evaluation. Expert Systems with Applications **244**, 122778.

Khan MHU, Wang S, Wang J, Ahmar S, Saeed S, Khan SU, Xu X, Chen H, Bhat JA, Feng X. 2022. Applications of artificial intelligence in climate-resilient smart-crop breeding. International Journal of Molecular Sciences **23**, 11156.

Khanal J, Tayara H, Zou Q, To Chong K. 2022. DeepCap-Kcr: accurate identification and investigation of protein lysine crotonylation sites based on capsule network. Briefings in Bioinformatics **23**, bbab492.

Khoso MA, Hussain A, Ritonga FN, *et al.* 2022. WRKY transcription factors (TFs): molecular switches to regulate drought, temperature, and salinity stresses in plants. Frontiers in Plant Science **13**, 1039329.

Kobayashi Y, Sadhukhan A, Tazib T, *et al.* 2016. Joint genetic and network analyses identify loci associated with root growth under NaCl stress in *Arabidopsis thaliana*. Plant, Cell & Environment **39**, 918–934.

Koh E, Sunil RS, Lam HYI, Mutwil M. 2024a. Confronting the data deluge: how artificial intelligence can be used in the study of plant stress. Computational and Structural Biotechnology Journal **23**, 3454–3466.

Koh E, Sunil RS, Lam HYI, Mutwil M. 2024b. Harnessing big data and artificial intelligence to study plant stress. arXiv doi: 10.48550/arXiv.2404.15776. [Preprint].

Konietschke F, Schwab K, Pauly M. 2021. Small sample sizes: a big data problem in high-dimensional data analysis. Statistical Methods in Medical Research **30**, 687–701.

Kotula L, Garcia Caparros P, Zörb C, Colmer TD, Flowers TJ. 2020. Improving crop salt tolerance using transgenic approaches: an update and physiological analysis. Plant, Cell & Environment **43**, 2932–2956.

Krishnamurthy P, Mohanty B, Wijaya E, Lee D-Y, Lim T-M, Lin Q, Xu J, Loh C-S, Kumar PP. 2017. Transcriptomics analysis of salt stress tolerance in the roots of the mangrove *Avicennia officinalis*. Scientific Reports **7**, 10031.

Krogh A. 2008. What are artificial neural networks? Nature Biotechnology **26**, 195–197.

Kulmanov M, Hoehndorf R. 2020. DeepGOPlus: improved protein function prediction from sequence. Bioinformatics **36**, 422–429.

Kumar A, Singh S, Gaurav AK, Srivastava S, Verma JP. 2020. Plant growth-promoting bacteria: biological tools for the mitigation of salinity stress in plants. Frontiers in Microbiology **11**, 1216.

Kumar N, Kaur G, Devi S, Lata C, Dasila H, Sanwal SK, Kumar A, Mann A. 2023. Advancement of omics approaches in understanding the mechanism of salinity tolerance in legumes. In: Kumar N, Dhansu P, Mann A, eds. Salinity and drought tolerance in plants: physiological perspectives. Singapore: Springer Singapore, 275–293.

Kumar P, Choudhary M, Halder T, *et al.* 2022. Salinity stress tolerance and omics approaches: revisiting the progress and achievements in major cereal crops. Heredity **128**, 497–518.

Lai F-L, Gao F. 2023. Auto-kla: a novel web server to discriminate lysine lactylation sites using automated machine learning. Briefings in Bioinformatics **24**, bbad070.

Lai H, Luo D, Yang M, *et al.* 2025. PBertKla: a protein large language model for predicting human lysine lactylation sites. BMC Biology **23**, 95.

Lam HYI, Ong XE, Mutwil M. 2024. Large language models in plant biology. Trends in Plant Science **29**, 1145–1155.

Lan K, Wang D, Fong S, Liu L, Wong KKL, Dey N. 2018. A survey of data mining and deep learning in bioinformatics. Journal of Medical Systems **42**, 139.

LeBauer D, Burnette M, Fahlgren N, Kooper R, McHenry K, Stylianou A. 2021. What does TERRA-REF's high resolution, multi sensor plant sensing public domain data offer the computer vision community? Proceedings of the IEEE/CVF international conference on computer vision (ICCV) workshops. Montreal, Canada, 1409–1415.

Lee T-Y. 2006. dbPTM: an information repository of protein post-translational modification. Nucleic Acids Research **34**, D622–D627.

Lei C, Zhou K, Zheng J, Zhao M, Huang Y, He H, Yang S, Zhang Z. 2024. AraPathogen2.0: an improved prediction of plant–pathogen protein–protein interactions empowered by the natural language processing technique. Journal of Proteome Research **23**, 494–499.

Levy B, Xu Z, Zhao L, Kremling K, Altman R, Wong P, Tanner C. 2022. FloraBERT: Cross-species transfer learning with attention-based neural networks for geneexpression prediction. Research Square doi: 10.21203/rs.3.rs-1927200/v1 [Preprint].

Lewis P, Perez E, Piktus A, *et al.* 2021. Retrieval-augmented generation for knowledge-intensive NLP tasks. arXiv doi:10.48550/arXiv.2005.11401. [Preprint].

Li F, Gates DJ, Buckler ES, *et al.* 2025. Environmental data provide marginal benefit for predicting climate adaptation. PLoS Genetics **21**, e1011714.

Li H, Xing X, Ding G, Li Q, Wang C, Xie L, Zeng R, Li Y. 2009. SysPTM: a systematic resource for proteomic research on post-translational modifications. Molecular & Cellular Proteomics **8**, 1839–1849.

Li H, Yan S, Zhao L, Tan J, Zhang Q, Gao F, Wang P, Hou H, Li L. 2014. Histone acetylation associated up-regulation of the cell wall related genes is involved in salt stress induced maize root swelling. BMC Plant Biology **14**, 105.

Li H, Duijts K, Pasini C, *et al.* 2023. Effective root responses to salinity stress include maintained cell expansion and carbon allocation. New Phytologist **238**, 1942–1956.

Li Q-C, Xu S-W, Zhuang J-Y, Liu J-J, Zhou Y, Zhang Z-X. 2023. Ensemble learning prediction of soybean yields in China based on meteorological data. Journal of Integrative Agriculture **22**, 1909–1927.

Li Y, Zhu Y, Liu Y, Shu Y, Meng F, Lu Y, Bai X, Liu B, Guo D. 2008. Genome-wide identification of osmotic stress response gene in *Arabidopsis thaliana*. Genomics **92**, 488–493.

Lin Z, Akin H, Rao R, *et al.* 2023. Evolutionary-scale prediction of atomic-level protein structure with a language model. Science **379**, 1123–1130.

Liu JN, Yan L, Chai Z, *et al.* 2025. Pan-genome analyses of 11 *Fraxinus* species provide insights into salt adaptation in ash trees. Plant Communications **6**, 101137.

Liu Q, Li P, Umer MJ, *et al.* 2025. Identification of EXPA4 as a key gene in cotton salt stress adaptation through transcriptomic and coexpression network analysis of root tip protoplasts. BMC Plant Biology **25**, 65.

Liu T, Qu J, Fang Y, Yang H, Lai W, Pan L, Liu J-H. 2025. Polyamines: the valuable bio-stimulants and endogenous signaling molecules for plant development and stress response. Journal of Integrative Plant Biology **67**, 582–595.

Liu Y, Li A, Zhao X-M, Wang M. 2021. DeepTL-Ubi: a novel deep transfer learning method for effectively predicting ubiquitination sites of multiple species. Methods **192**, 103–111.

Liu Y, Shen R, Zhou L, Xiao Q, Yuan J, Li Y. 2025. A data-intelligence-intensive bioinformatics copilot system for large-scale omics research and scientific insights. Briefings in Bioinformatics **26**, bbaf312.

López-García G, Jerez JM, Urda D, Veredas FJ. 2019. MetODeep: a deep learning approach for prediction of methionine oxidation sites in proteins. 2019 International joint conference on neural networks (IJCNN). Budapest, 1–8.

López-Gómez E, Bullones A, Santos-del Río J, Serrano-García V, Fernández-Pozo N. 2026. Challenges and opportunities for genomic data resources. In: Garg V, Henry R, Varshney R, eds. DNA of sustainability: genomics insights into food security challenges. Springer Nature, in press.

Lu M, Gao P, Hu J, Hou J, Wang D. 2023. A classification method of stress in plants using unsupervised learning algorithm and chlorophyll fluorescence technology. Frontiers in Plant Science **14**, 1202092.

Lumbanraja FR, Mahesworo B, Cenggoro TW, Sudigyo D, Pardamean B. 2021. SSMFN: a fused spatial and sequential deep learning model for methylation site prediction. PeerJ: Computer Science **7**, e683.

Luo F, Wang M, Liu Y, Zhao X-M, Li A. 2019. DeepPhos: prediction of protein phosphorylation sites with deep learning. Bioinformatics **35**, 2766–2773.

Luo M, Zhang Y, Li J, *et al.* 2021. Molecular dissection of maize seedling salt tolerance using a genome-wide association analysis method. Plant Biotechnology Journal **19**, 1937–1951.

Luo X, Wang B, Gao S, Zhang F, Terzaghi W, Dai M. 2019. Genome-wide association study dissects the genetic bases of salt tolerance in maize seedlings. Journal of Integrative Plant Biology **61**, 658–674.

Lv H, Dao F-Y, Guan Z-X, Yang H, Li Y-W, Lin H. 2021. Deep-Kcr: accurate detection of lysine crotonylation sites using deep learning method. Briefings in Bioinformatics **22**, bbaa255.

Lv H, Dao F-Y, Lin H. 2022. DeepKla: an attention mechanism-based deep neural network for protein lysine lactylation site prediction. iMeta **1**, e11.

Lv J, Jiang C, Wu W, *et al.* 2024. The gapless genome assembly and multi-omics analyses unveil a pivotal regulatory mechanism of oil biosynthesis in the olive tree. Horticulture Research **11**, uhae168.

Ma C, Xin M, Feldmann KA, Wang X. 2014. Machine learning-based differential network analysis: a study of stress-responsive transcriptomes in arabidopsis. The Plant Cell **26**, 520–537.

Ma Y, Xu T, Wan D, Ma T, Shi S, Liu J, Hu Q. 2015. The salinity tolerant poplar database (STPD): a comprehensive database for studying tree salt-tolerant adaption and poplar genomics. BMC Genomics **16**, 205.

Maji AK, Das S, Marwaha S, Kumar S, Dutta S, Choudhury MR, Arora A, Ray M, Perumal A, Chinusamy V. 2025. Intelligent decision support for drought stress (IDSDS): an integrated remote sensing and artificial intelligence-based pipeline for quantifying drought stress in plants. Computers and Electronics in Agriculture **236**, 110477.

Mansour MMF, Hassan FAS. 2022. How salt stress-responsive proteins regulate plant adaptation to saline conditions. Plant Molecular Biology **108**, 175–224.

Martí-Guillén JM, Pardo-Hernández M, Martínez-Lorente SE, Almagro L, Rivero RM. 2022. Redox post-translational modifications and their interplay in plant abiotic stress tolerance. Frontiers in Plant Science **13**, 1027730.

Mata-Pérez C, Sánchez-Vicente I, Arteaga N, Gómez-Jiménez S, Fuentes-Terrón A, Oulebsir CS, Calvo-Polanco M, Oliver C, Lorenzo Ó. 2023. Functions of nitric oxide-mediated post-translational modifications under abiotic stress. Frontiers in Plant Science **14**, 1158184.

Matese A, Prince Czarnecki JM, Samiappan S, Moorhead R. 2024. Are unmanned aerial vehicle-based hyperspectral imaging and machine learning advancing crop science? Trends in Plant Science **29**, 196–209.

Medina CA, Hawkins C, Liu X-P, Peel M, Yu L-X. 2020. Genome-wide association and prediction of traits related to salt tolerance in autotetraploid alfalfa (*Medicago sativa* L.). International Journal of Molecular Sciences **21**, 3361.

Medina CA, Kaur H, Ray I, Yu L-X. 2021. Strategies to increase prediction accuracy in genomic selection of complex traits in alfalfa (*Medicago sativa* L.). Cells **10**, 3372.

Meher PK, Sahu TK, Gupta A, Kumar A, Rustgi S. 2024. ASRpro: a machine-learning computational model for identifying proteins associated with multiple abiotic stress in plants. The Plant Genome **17**, e20259.

Melino V, Tester M. 2023. Salt-tolerant crops: time to deliver. Annual Review of Plant Biology **74**, 671–696.

Mendoza-Revilla J, Trop E, Gonzalez L, *et al.* 2024. A foundational large language model for edible plant genomes. Communications Biology **7**, 835.

Meng L, Chen X, Cheng K, Chen N, Zheng Z, Wang F, Sun H, Wong K-C. 2024. TransPTM: a transformer-based model for non-histone acetylation site prediction. Briefings in Bioinformatics **25**, bbae219.

Merkx D, Frank SL. 2021. Human sentence processing: recurrence or attention? Proceedings of the workshop on cognitive modeling and computational linguistics. Association for Computational Linguistics.

Miller C, Portlock T, Nyaga DM, O'Sullivan JM. 2024. A review of model evaluation metrics for machine learning in genetics and genomics. Frontiers in Bioinformatics **4**, 1457619.

Miller M. 2025. Using artificial intelligence. Absolute beginner's guide. Hoboken, NJ: Pearson Education, Inc.

Miolane N. 2025. The fifth era of science: artificial scientific intelligence. PLoS Biology **23**, e3003230.

Mittler R, Zandalinas SI, Fichman Y, Van Breusegem F. 2022. Reactive oxygen species signalling in plant stress responses. Nature Reviews. Molecular Cell Biology **23**, 663–679.

Mohammadi P, Asefpour Vakilian K. 2023. Machine learning provides specific detection of salt and drought stresses in cucumber based on miRNA characteristics. Plant Methods **19**, 123.

Molinaro AM, Simon R, Pfeiffer RM. 2005. Prediction error estimation: a comparison of resampling methods. Bioinformatics **21**, 3301–3307.

Molitor C, Kurowski TJ, Fidalgo de Almeida PM, Kevei Z, Spindlow DJ, Chacko Kaitholil SR, Iheanyichi JU, Prasanna HC, Thompson AJ, Mohareb FR. 2024. A chromosome-level genome assembly of *Solanum chilense*, a tomato wild relative associated with resistance to salinity and drought. Frontiers in Plant Science **15**, 1342739.

Morabito A, De Simone G, Pastorelli R, Brunelli L, Ferrario M. 2025. Algorithms and tools for data-driven omics integration to achieve multilayer biological insights: a narrative review. Journal of Translational Medicine **23**, 425.

Mueller HM, Franzisky BL, Messerer M, *et al.* 2024. Integrative multi-omics analyses of date palm (*Phoenix dactylifera*) roots and leaves reveal how the halophyte land plant copes with sea water. The Plant Genome **17**, e20372.

Muhammad M, Waheed A, Wahab A, Majeed M, Nazim M, Liu Y-H, Li L, Li W-J. 2024. Soil salinity and drought tolerance: an evaluation of plant growth, productivity, microbial diversity, and amelioration strategies. Plant Stress **11**, 100319.

Muhammad M, Wahab A, Waheed A, Hakeem KR, Mohamed HI, Basit A, Toor MD, Liu Y-H, Li L, Li W-J. 2025. Navigating climate change: exploring the dynamics between plant–soil microbiomes and their impact on plant growth and productivity. Global Change Biology **31**, e70057.

Muhammad N, Dong Q, Luo T, Zhang X, Song M, Wang X, Ma X. 2025. New developments in understanding cotton's physiological and molecular responses to salt stress. Plant Stress **15**, 100742.

Murmu S, Sinha D, Chaurasia H, Sharma S, Das R, Jha GK, Archak S. 2024. A review of artificial intelligence-assisted omics techniques in plant defense: current trends and future directions. Frontiers in Plant Science **15**, 1292054.

Murphy K. 2025. Reinforcement learning: an overview. arXiv doi: 10.48550/arXiv.2412.05265. [Preprint].

Murphy KM, Ludwig E, Gutierrez J, Gehan MA. 2024. Deep learning in image-based plant phenotyping. Annual Review of Plant Biology **75**, 771–795.

Nabwire S, Suh H-K, Kim MS, Baek I, Cho B-K. 2021. Review: application of artificial intelligence in phenomics. Sensors **21**, 4363.

Najar MA. 2024. Updates on protein post-translational modifications for modulating response to salinity. In: Ganie SA, Wani SH, eds. Genetics of salt tolerance in plants. Wallingford, UK: CABI Publishing, 96–118.

Nasti L, Vecchiato G, Heuret P, Rowe NP, Palladino M, Marcati P. 2024. A reinforcement learning approach to study climbing plant behaviour. Scientific Reports **14**, 18222.

Navarro A, Nicastro N, Costa C, Pentangelo A, Cardarelli M, Ortenzi L, Pallottino F, Cardi T, Pane C. 2022. Sorting biotic and abiotic stresses on wild rocket by leaf-image hyperspectral data mining with an artificial intelligence model. Plant Methods **18**, 45.

Nazari L, Ghotbi V, Nadimi M, Paliwal J. 2023. A novel machine-learning approach to predict stress-responsive genes in Arabidopsis. Algorithms **16**, 407.

Negrão S, Schmöckel SM, Tester M. 2017. Evaluating physiological responses of plants to salinity stress. Annals of Botany **119**, 1–11.

Ning W, Jiang P, Guo Y, Wang C, Tan X, Zhang W, Peng D, Xue Y. 2021. GPS-Palm: a deep learning-based graphic presentation system for the prediction of *S*-palmitoylation sites in proteins. Briefings in Bioinformatics **22**, 1836–1847.

Noorden RV. 2022. How language-generation AIs could transform science. Nature **605**, 21.

Novakovsky G, Dexter N, Libbrecht MW, Wasserman WW, Mostafavi S. 2023. Obtaining genetics insights from deep learning via explainable artificial intelligence. Nature Reviews. Genetics **24**, 125–137.

Olaoye G, Fajinmi J. 2025. Role of feature engineering in improving machine learning predictions of diabetes mellitus in healthcare data. Preprints.org doi: 10.20944/preprints202501.1739.v1 [Preprint].

Özer M. 2024. Is artificial intelligence hallucinating? Turk psikiyatri dergisi = Turkish Journal of Psychiatry **35**, 333–335.

Pakhrin SC, Pokharel S, Aoki-Kinoshita KF, Beck MR, Dam TK, Caragea D, Kc DB. 2023a. LMNglyPred: prediction of human *N*-linked glycosylation sites using embeddings from a pre-trained protein language model. Glycobiology **33**, 411–422.

Pakhrin SC, Pokharel S, Pratyush P, Chaudhari M, Ismail HD, Kc DB. 2023b. LMPhosSite: a deep learning-based approach for general protein phosphorylation site prediction using embeddings from the local window sequence and pretrained protein language model. Journal of Proteome Research **22**, 2548–2557.

Pakhrin SC, Chauhan N, Khan S, Upadhyaya J, Beck MR, Blanco E. 2024. Prediction of human O-linked glycosylation sites using stacked generalization and embeddings from pre-trained protein language model. Bioinformatics **40**, btae643.

Panahi B. 2024. Transcriptome signature for multiple biotic and abiotic stress in barley (*Hordeum vulgare* L.) identifies using machine learning approach. Current Plant Biology **40**, 100416.

Park SH, Sul A-R, Han K, Sung YS. 2023. How to determine if one diagnostic method, such as an artificial intelligence model, is superior to another: beyond performance metrics. Korean Journal of Radiology **24**, 601–605.

Paymode AS, Malode VB. 2022. Transfer learning for multi-crop leaf disease image classification using convolutional neural network VGG. Artificial Intelligence in Agriculture **6**, 23–33.

Peng FZ, Wang C, Chen T, Schussheim B, Vincoff S, Chatterjee P. 2025. PTM-Mamba: a PTM-aware protein language model with bidirectional gated Mamba blocks. Nature Methods **22**, 945–949.

Peng S, Rajjou L. 2024. Advancing plant biology through deep learning-powered natural language processing. Plant Cell Reports **43**, 208.

Pirooznia M, Yang JY, Yang MQ, Deng Y. 2008. A comparative study of different machine learning methods on microarray gene expression data. BMC Genomics **9**, S13.

Pokharel S, Pratyush P, Heinzinger M, Newman RH, Kc DB. 2022. Improving protein succinylation sites prediction using embeddings from protein language model. Scientific Reports **12**, 16933.

Pradhan UK, Meher PK, Naha S, Rao AR, Kumar U, Pal S, Gupta A. 2023. ASmiR: a machine learning framework for prediction of abiotic stress-specific miRNAs in plants. Functional & Integrative Genomics **23**, 92.

Pratyush P, Pokharel S, Saigo H, Kc DB. 2023. pLMSNOSite: an ensemble-based approach for predicting protein S-nitrosylation sites by integrating supervised word embedding and embedding from pre-trained protein language model. BMC Bioinformatics **24**, 41.

Pratyush P, Bahmani S, Pokharel S, Ismail HD, Kc DB. 2024. LMCrot: an enhanced protein crotonylation site predictor by leveraging an interpretable window-level embedding from a transformer-based protein language model. Bioinformatics **40**, btae290.

Pudumalar S, Muthuramalingam S. 2024. Hydra: an ensemble deep learning recognition model for plant diseases. Journal of Engineering Research **12**, 781–792.

Qiao B, Gao W, Zhang X, *et al.* 2025. SaGP: identifying plant saline–alkali tolerance genes based on machine learning techniques. Frontiers in Plant Science **16**, 1629794.

Qin S, Zhang Y, Tian Z. 2024. Quantitative *N*-glycoproteomics characterization of differential *N*-glycosylation in *Sorghum bicolor* under salinity stress. Biochemical and Biophysical Research Communications **737**, 150509.

Rai KK. 2022. Integrating speed breeding with artificial intelligence for developing climate-smart crops. Molecular Biology Reports **49**, 11385–11402.

Rai KK. 2024. Stress phenotyping in plants using artificial intelligence and machine learning. Journal of Agriculture and Livestock Farming **1**, 1–2.

Rajpoot V, Tiwari A, Jalal AS. 2023. Automatic early detection of rice leaf diseases using hybrid deep learning and machine learning methods. Multimedia Tools and Applications **82**, 36091–36117.

Ramazi S, Zahiri J. 2021. Post-translational modifications in proteins: resources, tools and prediction methods. Database **2021**, baab012.

Rautiainen M, Nurk S, Walenz BP, Logsdon GA, Porubsky D, Rhie A, Eichler EE, Phillippy AM, Koren S. 2023. Telomere-to-telomere assembly of diploid chromosomes with Verkko. Nature Biotechnology **41**, 1474–1482.

Raza A, Li Y, Prakash CS, Hu Z. 2025a. Panomics to manage combined abiotic stresses in plants. Trends in Plant Science **30**, 1079–1084.

Raza A, Zaman QU, Shabala S, Tester M, Munns R, Hu Z, Varshney RK. 2025b. Genomics-assisted breeding for designing salinity-smart future crops. Plant Biotechnology Journal **23**, 3119–3151.

Razali R, Bougouffa S, Morton MJL, *et al.* 2018. The genome sequence of the wild tomato *Solanum pimpinellifolium* provides insights into salinity tolerance. Frontiers in Plant Science **9**, 1402.

Ren J, Wen L, Gao X, Jin C, Xue Y, Yao X. 2008. CSS-Palm 2.0: an updated software for palmitoylation sites prediction. Protein Engineering Design and Selection **21**, 639–644.

Rico-Chávez AK, Franco JA, Fernandez-Jaramillo AA, Contreras-Medina LM, Guevara-González RG, Hernandez-Escobedo Q. 2022. Machine learning for plant stress modeling: a perspective towards hormesis management. Plants **11**, 970.

Rissom PF, Sarmiento PY, Safer J, Coley CW, Renard BY, Heyne HO, Iqbal S. 2025. Decoding protein language models: insights from embedding space analysis. bioRxiv doi: 10.1101/2024.06.21.600139. [Preprint].

Rivero RM, Mestre TC, Mittler R, Rubio F, Garcia-Sanchez F, Martinez V. 2014. The combined effect of salinity and heat reveals a specific physiological, biochemical and molecular response in tomato plants. Plant, Cell & Environment **37**, 1059–1073.

Rivero RM, Mittler R, Blumwald E, Zandalinas SI. 2022. Developing climate-resilient crops: improving plant tolerance to stress combination. The Plant Journal **109**, 373–389.

Rives A, Meier J, Sercu T, *et al.* 2021. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. Proceedings of the National Academy of Sciences, USA **118**, e2016239118.

Rohart F, Gautier B, Singh A, Lê Cao K-A. 2017. mixOmics: an R package for 'omics feature selection and multiple data integration. PLoS Computational Biology **13**, e1005752.

Roychowdhury R, Das SP, Gupta A, Parihar P, Chandrasekhar K, Sarker U, Kumar A, Ramrao DP, Sudhakar C. 2023. Multi-omics pipeline and omics-integration approach to decipher plant's abiotic stress tolerance responses. Genes **14**, 1281.

Sachdev S, Ansari SA, Ansari MI, Fujita M, Hasanuzzaman M. 2021. Abiotic stress and reactive oxygen species: generation, signaling, and defense mechanisms. Antioxidants **10**, 277.

Sadder MT, Ali AAM, Alsadon AA, Wahb-Allah MA. 2025. Long-term salinity-responsive transcriptome in advanced breeding lines of tomato. Plants **14**, 100.

Saeys Y, Inza I, Larrañaga P. 2007. A review of feature selection techniques in bioinformatics. Bioinformatics **23**, 2507–2517.

Sahito A, Frank E, Pfahringer B. 2019. Semi-supervised learning using Siamese networks. In: Liu J, Bailey J, eds. AI 2019: advances in artificial intelligence. Cham: Springer International Publishing, 586–597.

Sahu SK, Liu H. 2023. Long-read sequencing (method of the year 2022): the way forward for plant omics research. Molecular Plant **16**, 791–793.

Saleem MH, Noreen S, Ishaq I, *et al.* 2025. Omics technologies: unraveling abiotic stress tolerance mechanisms for sustainable crop improvement. Journal of Plant Growth Regulation **44**, 4165–4187.

Samant RM, Bachute MR, Gite S, Kotecha K. 2022. Framework for deep learning-based language models using multi-task learning in natural language understanding: a systematic literature review and future directions. IEEE Access **10**, 17078–17097.

Samara B. 2024. Artificial intelligence and ethics in healthcare. Pakistan Journal of Life and Social Sciences **22**, 23670–23679.

Sandhu M, Sureshkumar V, Prakash C, Dixit R, Solanke AU, Sharma TR, Mohapatra T Sv AM. 2017. RiceMetaSys for salt and drought stress responsive genes in rice: a web interface for crop improvement. BMC Bioinformatics **18**, 432.

Sarkar S, Ganapathysubramanian B, Singh A, *et al.* 2024. Cyber-agricultural systems for crop breeding and sustainable production. Trends in Plant Science **29**, 130–149.

Sarkar SK, Rudra RR, Sohan AR, Das PC, Ekram KMM, Talukdar S, Rahman A, Alam E, Islam MK, Islam ARMT. 2023. Coupling of machine learning and remote sensing for soil salinity mapping in coastal area of Bangladesh. Scientific Reports **13**, 17056.

Schmirler R, Heinzinger M, Rost B. 2024. Fine-tuning protein language models boosts predictions across diverse tasks. Nature Communications **15**, 7407.

Schwartz D, Chou MF, Church GM. 2009. Predicting protein post-translational modifications using meta-analysis of proteome scale data sets. Molecular & Cellular Proteomics **8**, 365–379.

Searls DB. 2002. The language of genes. Nature **420**, 211–217.

Sesli M, Yegenoğlu ED, Altıntaa V. 2020. Determination of olive cultivars by deep learning and ISSR markers. Journal of Environmental Biology **41**, 426–431.

Shahid MF, Khanzada TJS, Aslam MA, Hussain S, Baowidan SA, Ashari RB. 2024. An ensemble deep learning models approach using image analysis for cotton crop classification in AI-enabled smart agriculture. Plant Methods **20**, 104.

Shaik R, Ramakrishna W. 2014. Machine learning approaches distinguish multiple stress conditions using stress-responsive genes and identify candidate genes for broad resistance in rice. Plant Physiology **164**, 481–495.

Shokri N, Hassani A, Azapagic A. 2021. Soil salinization under different climate change scenarios: a global scale analysis. EGU General Assembly Conference Abstracts **EGU21**, 13668.

Shrestha P, Kandel J, Tayara H, Chong KT. 2024. Post-translational modification prediction via prompt-based fine-tuning of a GPT-2 model. Nature Communications **15**, 6699.

Silva J, Teixeira R, Silva F, Brommonschenkel S, Fontes E. 2019. Machine learning approaches and their current application in plant molecular biology: a systematic review. Plant Science **284**, 37–47.

Simon E, Swanson K, Zou J. 2024. Language models for biological research: a primer. Nature Methods **21**, 1422–1429.

Singh A, Ganapathysubramanian B, Singh AK, Sarkar S. 2016. Machine learning for high-throughput stress phenotyping in plants. Trends in Plant Science **21**, 110–124.

Singh D, Jain N, Jain P, Kayal P, Kumawat S, Batra N. 2020. Plantdoc: a dataset for visual plant disease detection. CoDS COMAD 2020. Proceedings of the 7th ACM IKDD CoDS and 25th COMAD. New York: Association for Computing Machinery, 249–253.

Singhal K, Azizi S, Tu T, *et al.* 2023. Large language models encode clinical knowledge. Nature **620**, 172–180.

Singla A, Nehra A, Joshi K, Kumar A, Tuteja N, Varshney RK, Gill SS, Gill R. 2024. Exploration of machine learning approaches for automated crop disease detection. Current Plant Biology **40**, 100382.

Sirpa-Poma J, Satgé F, Resongles E, Zolá R, Molina-Carpio J, Colque M, Ormachea M, Mollinedo P, Bonnet M-P. 2023. Towards the improvement of soil salinity mapping in a data-scarce context using sentinel-2 images in machine-learning models. Sensors **23**, 9328.

Sodini M, Astolfi S, Francini A, Sebastiani L. 2022. Multiple linear regression and linear mixed models identify novel traits of salinity tolerance in *Olea europaea* L. Tree Physiology **42**, 1029–1042.

Soltabayeva A, Ongaltay A, Omondi JO, Srivastava S. 2021. Morphological, physiological and molecular markers for salt-stressed plants. Plants **10**, 243.

Song L, Xu Y, Wang M, Leng Y. 2021. PreCar_Deep: a deep learning framework for prediction of protein carbonylation sites based on borderline-SMOTE strategy. Chemometrics and Intelligent Laboratory Systems **218**, 104428.

Sonsungsan P, Suratanee A, Buaboocha T, Chadchawan S, Plaimas K. 2024. Identification of salt-sensitive and salt-tolerant genes through weighted gene co-expression networks across multiple datasets: a centralization and differential correlation analysis. Genes **15**, 316.

Steyvers M, Tejeda H, Kumar A, Belem C, Karny S, Hu X, Mayer LW, Smyth P. 2025. What large language models know and what people think they know. Nature Machine Intelligence **7**, 221–231.

Stigter TY, Ribeiro L, Carvalho Dill AMM. 2006. Evaluation of an intrinsic and a specific vulnerability assessment method in comparison with groundwater salinisation and nitrate contamination levels in two agricultural regions in the south of Portugal. Hydrogeology Journal **14**, 79–99.

Strzoda T, Cruz-Garcia L, Najim M, Badie C, Polanska J. 2024. A mapping-free natural language processing-based technique for sequence search in nanopore long-reads. BMC Bioinformatics **25**, 354.

Su Y, Yang X, Wang Y, *et al.* 2024. Phased telomere-to-telomere reference genome and pangenome reveal an expansion of resistance genes during apple domestication. Plant Physiology **195**, 2799–2814.

Subramanian I, Verma S, Kumar S, Jere A, Anamika K. 2020. Multi-omics data integration, interpretation, and its application. Bioinformatics and Biology Insights **14**, 1177932219899051.

Sunil RS, Lim SC, Itharajula M, Mutwil M. 2024. The gene function prediction challenge: large language models and knowledge graphs to the rescue. Current Opinion in Plant Biology **82**, 102665.

Tan J, Huyck M, Hu D, Zelaya RA, Hogan DA, Greene CS. 2017. ADAGE signature analysis: differential expression analysis with data-defined gene sets. BMC Bioinformatics **18**, 512.

Tello-Ruiz MK, Jaiswal P, Ware D. 2022. Gramene: a resource for comparative analysis of plants genomes and pathways. Methods in Molecular Biology **2443**, 101–131.

Thapa N, Chaudhari M, McManus S, Roy K, Newman RH, Saigo H, Kc DB. 2020. DeepSuccinylSite: a deep learning based approach for protein succinylation site prediction. BMC Bioinformatics **21**, 63.

Tibesigwa DG, Zhuang W, Matola SH, Zhao H, Li W, Yang L, Ren J, Liu Q, Yang J. 2025. Molecular insights into salt stress adaptation in plants. Plant, Cell & Environment **48**, 5604–5615.

Tong R, Xu T, Ju X, Wang L. 2025. Progress in medical AI: reviewing large language models and multimodal systems for diagonosis. AI Medicine **1**, 5.

Tran K-N, Wang G, Oh D-H, Larkin JC, Smith AP, Dassanayake M. 2022. Multiple paths lead to salt tolerance–pre-adaptation vs dynamic responses from two closely related extremophytes. bioRxiv doi: 10.1101/2021.10.23.465591. [Preprint].

Tran T-X, Khanh Le NQ, Nguyen V-N. 2025. Integrating CNN and bi-LSTM for protein succinylation sites prediction based on natural language processing technique. Computers in Biology and Medicine **186**, 109664.

Trost B, Kusalik A. 2011. Computational prediction of eukaryotic phosphorylation sites. Bioinformatics **27**, 2927–2935.

Truchi M, Lacoux C, Gille C, *et al.* 2024. Detecting subtle transcriptomic perturbations induced by lncRNAs knock-down in single-cell CRISPRi screening using a new sparse supervised autoencoder neural network. Frontiers in Bioinformatics **4**, 1340339.

Ullah MA, Abdullah-Zawawi M-R, Zainal-Abidin R-A, Sukiran NL, Uddin MI, Zainal Z. 2022. A review of integrative omic approaches for understanding rice salt response mechanisms. Plants **11**, 1430.

Uygun S, Azodi CB, Shiu S-H. 2019. Cis-regulatory code for predicting plant cell-type transcriptional response to high salinity. Plant Physiology **181**, 1739–1751.

Van Royen FS, Asselbergs FW, Alfonso F, Vardas P, van Smeden M. 2023. Five critical quality criteria for artificial intelligence-based prediction models. European Heart Journal **44**, 4831–4834.

Varshney RK, Sinha P, Singh VK, Kumar A, Zhang Q, Bennetzen JL. 2020. 5Gs for crop genetic improvement. Current Opinion in Plant Biology **56**, 190–196.

Varshney RK, Barmukh R, Bentley A, Nguyen HT. 2024. Exploring the genomics of abiotic stress tolerance and crop resilience to climate change. The Plant Genome **17**, e20445.

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. 2017. Attention is all you need. NIPS'17. Proceedings of the 31st international conference on neural information processing systems. Red Hook, NY: Curran Associates Inc., 6000–6010.

Vello E, Letourneau M, Aguirre J, Bureau TE. 2024. Integrated web portal for non-destructive salt sensitivity detection of *Camelina sativa* seeds using fluorescent and visible light images coupled with machine learning algorithms. Frontiers in Plant Science **14**, 1303429.

Vidhya V, Kiran A, Raj MS, Kamesh D, Mishra S. 2024. Greenhouse environments yield prediction for plant growth using deep learning. 2024 International conference on advancements in smart, secure and intelligent computing (ASSIC). Bhubaneswar, India, 1–6.

Vrana J, Singh R. 2021. NDE 4.0–a design thinking perspective. Journal of Nondestructive Evaluation **40**, 8.

Wagan S, Ali M, Khoso MA, Alam I, Dinislam K, Hussain A, Brohi NA, Manghwar H, Liu F. 2024. Deciphering the role of WRKY transcription factors in plant resilience to alkaline salt stress. Plant Stress **13**, 100526.

Walsh J, Mangina E, Negrão S. 2024. Advancements in imaging sensors and AI for plant stress detection: a systematic literature review. Plant Phenomics **6**, 0153.

Wang D, Zeng S, Xu C, Qiu W, Liang Y, Joshi T, Xu D. 2017. MusiteDeep: a deep-learning framework for general and kinase-specific phosphorylation site prediction. Bioinformatics **33**, 3909–3916.

Wang D, Liang Y, Xu D. 2019. Capsule network for protein post-translational modification site prediction. Bioinformatics **35**, 2386–2394.

Wang D, Liu D, Yuchi J, He F, Jiang Y, Cai S, Li J, Xu D. 2020. MusiteDeep: a deep-learning based webserver for protein post-translational modification site prediction and visualization. Nucleic Acids Research **48**, W140–W146.

Wang J, Peng J, Li H, Yin C, Liu W, Wang T, Zhang H. 2021. Soil salinity mapping using machine learning algorithms with the sentinel-2 MSI in arid areas, China. Remote Sensing **13**, 305.

Wang L, Liu N, Zhou Y, Zheng F, Jian S, Liu X. 2025. Multi-omics analyses provide insights into the molecular basis for salt tolerance of *Phyla nodiflora*. The Plant Journal **123**, e70325.

Wang M, Cui X, Li S, Yang X, Ma A, Zhang Y, Yu B. 2020. DeepMal: accurate prediction of protein malonylation sites by deep neural networks. Chemometrics and Intelligent Laboratory Systems **207**, 104175.

Wang P, Liu W-C, Han C, Wang S, Bai M-Y, Song C-P. 2024. Reactive oxygen species: mmultidimensional regulators of plant adaptation to abiotic stress and development. Journal of Integrative Plant Biology **66**, 330–367.

Wang Q, Luo S, Yang Y, *et al.* 2025. WP-MOD: a multi-omics and taxonomy database for woody plants. Plant Communications **6**, 101290.

Wang S, You R, Liu Y, Xiong Y, Zhu S. 2023. NetGO 3.0: protein language model improves large-scale functional annotations. Genomics, Proteomics & Bioinformatics **21**, 349–358.

Wang W, Li W, Cheng Z, *et al.* 2022. Transcriptome-wide N6-methyladenosine profiling of cotton root provides insights for salt stress tolerance. Environmental and Experimental Botany **194**, 104729.

Wang Y, Du F, Li Y, Wang J, Zhao X, Li Z, Xu J, Wang W, Fu B. 2022. Global N6-methyladenosine profiling revealed the tissue-specific epitranscriptomic regulation of rice responses to salt stress. International Journal of Molecular Sciences **23**, 2091.

Wei H, Wang X, Zhang Z, *et al.* 2024. Uncovering key salt-tolerant regulators through a combined eQTL and GWAS analysis using the super pangenome in rice. National Science Review **11**, nwae043.

Wei L, Liao W, Zhong Y, Tian Y, Wei S, Liu Y. 2024. NO-mediated protein s-nitrosylation under salt stress: role and mechanism. Plant Science **338**, 111927.

Weissenow K, Rost B. 2025. Are protein language models the new universal key? Current Opinion in Structural Biology **91**, 102997.

Whalen S, Schreiber J, Noble WS, Pollard KS. 2022. Navigating the pitfalls of applying machine learning in genomics. Nature Reviews. Genetics **23**, 169–181.

Wilkins MR, Gasteiger E, Gooley AA, *et al.* 1999. High-throughput mass spectrometric discovery of protein post-translational modifications. Journal of Molecular Biology **289**, 645–657.

Willems P, Sterck L, Dard A, Huang J, De Smet I, Gevaert K, Van Breusegem F. 2024. The plant PTM viewer 2.0: in-depth exploration of plant protein modification landscapes. Journal of Experimental Botany **75**, 4611–4624.

Woodhouse MR, Cannon EK, Portwood JL, Harper LC, Gardiner JM, Schaeffer ML, Andorf CM. 2021. A pan-genomic approach to genome databases using maize as a model system. BMC Plant Biology **21**, 385.

Wu H, Han R, Zhao L, Liu M, Chen H, Li W, Li L. 2025. AutoGP: an intelligent breeding platform for enhancing maize genomic selection. Plant Communications **6**, 101240.

Wu J-S, Mu D-W, Feng N-J, Zheng D-F, Sun Z-Y, Khan A, Zhou H, Song Y-W, Liu J-X, Luo J-Q. 2025. Integrated analyses reveal the physiological and molecular mechanisms of brassinolide in modulating salt tolerance in rice. Plants **14**, 1555.

Wu M, Yang Y, Wang H, Xu Y. 2019. A deep learning method to more accurately recall known lysine acetylation sites. BMC Bioinformatics **20**, 49.

Wu W, Zucca C, Muhaimeed A, Al-Shafie W, Al-Quraishi AF, Nangia V, Zhu M, Liu G. 2018. Soil salinity prediction and mapping by machine learning regression in Central Mesopotamia, Iraq. Land Degradation & Development **29**, 4005–4014.

Xiao F, Zhou H. 2023. Plant salt response: perception, signaling, and tolerance. Frontiers in Plant Science **13**, 1053699.

Xie J, Shi C, Liu Y, Wang Q, Zhong Z, He S, Wang X. 2025. Soil salinization prediction through feature selection and machine learning at the irrigation district scale. Frontiers in Earth Science **12**, 1488504.

Xie L, Gong X, Yang K, *et al.* 2024. Technology-enabled great leap in deciphering plant genomes. Nature Plants **10**, 551–566.

Xie Y, Zheng Y, Li H, *et al.* 2016. GPS-Lipid: a robust tool for the prediction of multiple lipid modification sites. Scientific Reports **6**, 28249.

Xie Y, Luo X, Li Y, *et al.* 2018. DeepNitro: prediction of protein nitration and nitrosylation sites by deep learning. Genomics, Proteomics & Bioinformatics **16**, 294–306.

Xiong H, He H, Chang Y, Miao B, Liu Z, Wang Q, Dong F, Xiong L. 2025. Multiple roles of NAC transcription factors in plant development and stress responses. Journal of Integrative Plant Biology **67**, 510–538.

Xu L, Liu H, Mittler R, Shabala S. 2025. Useful or merely convenient? On the issue of a suitability of enzymatic antioxidant activity as a proxy for abiotic stress tolerance. Journal of Experimental Botany **76**, 1524–1533.

Xu Y, Liu X, Cao X, *et al.* 2021. Artificial intelligence: a powerful paradigm for scientific research. The Innovation **2**, 100179.

Xue Y, Zhou F, Fu C, Xu Y, Yao X. 2006. SUMOsp: a web server for sumoylation site prediction. Nucleic Acids Research **34**, W254–W257.

Yan J, Wang X. 2022. Unsupervised and semi-supervised learning: the next frontier in machine learning for plant systems biology. The Plant Journal **111**, 1527–1538.

Yan J, Wang X. 2023. Machine learning bridges omics sciences and plant breeding. Trends in Plant Science **28**, 199–210.

Yang F, Kong H, Ying J, *et al.* 2025. SeedLLM·Rice: a large language model integrated with rice biological knowledge graph. Molecular Plant **18**, 1118–1129.

Yang H, Wang M, Liu X, Zhao X-M, Li A. 2021. PhosIDN: an integrated deep neural network for improving protein phosphorylation site prediction by combining sequence and protein–protein interaction information. Bioinformatics **37**, 4668–4676.

Yang L, Fang S, Liu L, Zhao L, Chen W, Li X, Xu Z, Chen S, Wang H, Yu D. 2025a. WRKY transcription factors: hubs for regulating plant growth and stress responses. Journal of Integrative Plant Biology **67**, 488–509.

Yang L, Wang H, Zou M, Chai H, Xia Z. 2025b. Artificial intelligence-driven plant bio-genomics research: a new era. Tropical Plants **4**, e015.

Yang M, Chen S, Huang Z, *et al.* 2023. Deep learning-enabled discovery and characterization of HKT genes in *Spartina alterniflora*. The Plant Journal **116**, 690–705.

Yang T, Cai Y, Huang T, Yang D, Yang X, Yin X, Zhang C, Yang Y, Yang Y. 2024. A telomere-to-telomere gap-free reference genome assembly of avocado provides useful resources for identifying genes related to fatty acid biosynthesis and disease resistance. Horticulture Research **11**, uhae119.

Yang Y, Xu Y, Feng B, Li P, Li C, Zhu C-Y, Ren S-N, Wang H-L. 2025. Regulatory networks of bZIPs in drought, salt and cold stress response and signaling. Plant Science **352**, 112399.

Yaschenko AE, Alonso JM, Stepanova AN. 2025. Arabidopsis as a model for translational research. The Plant Cell **37**, koae065.

Yelmen B, Jay F. 2023. An overview of deep generative models in functional and evolutionary genomics. Annual Review of Biomedical Data Science **6**, 173–189.

Yu B, Yu Z, Chen C, Ma A, Liu B, Tian B, Ma Q. 2020. DNNAce: prediction of prokaryote lysine acetylation sites through deep neural networks with multi-information fusion. Chemometrics and Intelligent Laboratory Systems **200**, 103999.

Yu H, Yang H, Sun W, Yan Z, Yang X, Zhang H, Ding Y, Li K. 2024. An interpretable RNA foundation model for exploring functional RNA motifs in plants. Nature Machine Intelligence **6**, 1616–1625.

Yu S, Xie L, Huang Q. 2023. Inception convolutional vision transformers for plant disease identification. Internet of Things **21**, 100650.

Yu Y, Wang J, Liu Y, Yu P, Wang D, Zheng P, Zhang M. 2024. Revisit the environmental impact of artificial intelligence: the overlooked carbon emission source? Frontiers of Environmental Science & Engineering **18**, 158.

Zarbakhsh S, Shahsavar AR. 2022. Artificial neural network-based model to predict the effect of γ-aminobutyric acid on salinity and drought responsive morphological traits in pomegranate. Scientific Reports **12**, 16662.

Zeng Q, Hu H-W, Ge A-H, Xiong C, Zhai C-C, Duan G-L, Han L-L, Huang S-Y, Zhang L-M. 2025. Plant–microbiome interactions and their impacts on plant adaptation to climate change. Journal of Integrative Plant Biology **67**, 826–844.

Zhai J, Gokaslan A, Schiff Y, *et al.* 2025. Cross-species modeling of plant genomes at single nucleotide resolution using a pre-trained DNA language model. Biophysics and Computational Biology **122**, e2421738122.

Zhang C, Cui Y, Yuan C, *et al.* 2025. Rice3kGS: a powerful web platform and database for large-scale genome selection. Plant Communications **6**, 101369.

Zhang D, Xu Z-C, Su W, Yang Y-H, Lv H, Yang H, Lin H. 2021. iCarPS: a computational tool for identifying protein carbonylation sites by novel encoded features. Bioinformatics **37**, 171–177.

Zhang D, Zhou H, Zhang Y, Zhao Y, Zhang Y, Feng X, Lin H. 2025. Diverse roles of MYB transcription factors in plants. Journal of Integrative Plant Biology **67**, 539–562.

Zhang F, Zhu G, Du L, Shang X, Cheng C, Yang B, Hu Y, Cai C, Guo W. 2016. Genetic regulation of salt stress tolerance revealed by RNA-Seq in cotton diploid wild species, *Gossypium davidsonii*. Scientific Reports **6**, 20582.

Zhang H, Zhu J, Gong Z, Zhu J-K. 2022. Abiotic stress responses in plants. Nature Reviews. Genetics **23**, 104–119.

Zhang H, Yu C, Zhang Q, Qiu Z, Zhang X, Hou Y, Zang J. 2025. Salinity survival: molecular mechanisms and adaptive strategies in plants. Frontiers in Plant Science **16**, 1527952.

Zhang M, Cao Y, Wang Z, Wang Z, Shi J, Liang X, Song W, Chen Q, Lai J, Jiang C. 2018. A retrotransposon in an HKT1 family sodium transporter causes variation of leaf Na^+ exclusion and salt tolerance in maize. New Phytologist **217**, 1161–1176.

Zhang M, Liu Y, Han G, Zhang Y, Wang B, Chen M. 2021. Salt tolerance mechanisms in trees: research progress. Trees **35**, 717–730.

Zhang Q, Chen W, Qin M, *et al.* 2025. Integrating protein language models and automatic biofoundry for enhanced protein evolution. Nature Communications **16**, 1553.

Zhang R, Wang Y, Yang W, *et al.* 2025. PlantGPT: an arabidopsis-based intelligent agent that answers questions about plant functional genomics. Advanced Science **12**, e03926.

Zhang X, Acencio ML, Lemke N. 2016. Predicting essential genes and proteins based on machine learning and network topological features: a comprehensive review. Frontiers in Physiology **7**, 75.

Zhang X, Ibrahim Z, Khaskheli MB, Raza H, Zhou F, Shamsi IH. 2024. Integrative approaches to abiotic stress management in crops: combining bioinformatics educational tools and artificial intelligence applications. Sustainability **16**, 7651.

Zhao B, Zhou Y, Jiao X, Wang X, Wang B, Yuan F. 2023. Bracelet salt glands of the recretohalophyte *Limonium bicolor*: distribution, morphology, and induction. Journal of Integrative Plant Biology **65**, 950–966.

Zhao X, Li K, Li Y, Ma J, Zhang L. 2022. Identification method of vegetable diseases based on transfer learning and attention mechanism. Computers and Electronics in Agriculture **193**, 106703.

Zheng H, Sun X, Li J, Song Y, Song J, Wang F, Liu L, Zhang X, Sui N. 2021. Analysis of N6-methyladenosine reveals a new important mechanism regulating the salt tolerance of sweet sorghum. Plant Science **304**, 110801.

Zheng M, Liu X, Lin J, *et al.* 2019. Histone acetyltransferase GCN 5 contributes to cell wall integrity and salt stress tolerance by altering the expression of cellulose synthesis genes. The Plant Journal **97**, 587–602.

Zhou C, Ye H, Sun D, Yue J, Yang G, Hu J. 2022. An automated, high-performance approach for detecting and characterizing broccoli based on UAV remote-sensing and transformers: a case study from Haining, China. International Journal of Applied Earth Observation and Geoinformation **114**, 103055.

Zhou J, Zhang B, Li G, *et al.* 2024. An AI agent for fully automated multi-omic analyses. Advanced Science **11**, 2407094.

Zhu D, Liu J, Duan W, Sun H, Zhang L, Yan Y. 2023. Analysis of the chloroplast crotonylome of wheat seedling leaves reveals the roles of crotonylated proteins involved in salt-stress responses. Journal of Experimental Botany **74**, 2067–2082.

Zhu J, Guo W, Lan Y. 2023. Global analysis of lysine lactylation of germinated seeds in wheat. International Journal of Molecular Sciences **24**, 16195.

Zhu Z, Zhou Y, Liu X, Meng F, Xu C, Chen M. 2025. Integrated transcriptomic and metabolomic analyses uncover the key pathways of *Limonium bicolor* in response to salt stress. Plant Biotechnology Journal **23**, 715–730.

Zou J, Huss M, Abid A, Mohammadi P, Torkamani A, Telenti A. 2019. A primer on deep learning in genomics. Nature Genetics **51**, 12–18.

Zou Y, Xu X. 2025. Multi-omics analysis reveals key regulatory defense pathways in *Ruppia sinensis* in response to water salinity fluctuations. BMC Plant Biology **25**, 174.

Zrimec J, Fu X, Muhammad AS, *et al.* 2022. Controlling gene expression with deep generative design of regulatory DNA. Nature Communications **13**, 5099.

Zulfiqar F, Nafees M, Chen J, *et al.* 2022. Chemical priming enhances plant tolerance to salt stress. Frontiers in Plant Science **13**, 946922.