



OPEN ACCESS

EDITED BY

Daojun Yuan,
Huazhong Agricultural University, China

REVIEWED BY

Xiujun Zhang,
Chinese Academy of Sciences (CAS), China
Dong-Liang Huang,
Guangxi Academy of Agricultural Sciences,
China

*CORRESPONDENCE

Youxiong Que
✉ queyouxiong@126.com

RECEIVED 18 July 2023

ACCEPTED 13 September 2023

PUBLISHED 03 October 2023

CITATION

Wang X, Zeng H, Lin L, Huang Y, Lin H and
Que Y (2023) Deep learning-empowered
crop breeding: intelligent, efficient
and promising.
Front. Plant Sci. 14:1260089.
doi: 10.3389/fpls.2023.1260089

COPYRIGHT

© 2023 Wang, Zeng, Lin, Huang, Lin and
Que. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Deep learning-empowered crop breeding: intelligent, efficient and promising

Xiaoding Wang¹, Haitao Zeng¹, Limei Lin¹, Yanze Huang²,
Hui Lin¹ and Youxiong Que^{3,4*}

¹Fujian Provincial Key Lab of Network Security & Cryptology, College of Computer and Cyber Security, Fujian Normal University, Fuzhou, China, ²School of Computer Science and Mathematics, Fujian Provincial Key Laboratory of Big Data Mining and Applications, Fujian University of Technology, Fuzhou, China, ³Key Laboratory of Sugarcane Biology and Genetic Breeding, Ministry of Agriculture and Rural Affairs, Fujian Agriculture and Forestry University, Fuzhou, China, ⁴National Key Laboratory for Tropical Crop Breeding, Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, Hainan, China

Crop breeding is one of the main approaches to increase crop yield and improve crop quality. However, the breeding process faces challenges such as complex data, difficulties in data acquisition, and low prediction accuracy, resulting in low breeding efficiency and long cycle. Deep learning-based crop breeding is a strategy that applies deep learning techniques to improve and optimize the breeding process, leading to accelerated crop improvement, enhanced breeding efficiency, and the development of higher-yielding, more adaptive, and disease-resistant varieties for agricultural production. This perspective briefly discusses the mechanisms, key applications, and impact of deep learning in crop breeding. We also highlight the current challenges associated with this topic and provide insights into its future application prospects.

KEYWORDS

crop breeding, deep learning, smart breeding, challenge, prospect

Introduction

Crop quality has always been a focal point in human cultivation, and crop breeding, as a primary approach to increasing crop yield and improving crop quality, is one of the oldest agricultural activities, equivalent to human civilization (Shen et al., 2022). Emerging as the times require, crop breeding is the process of artificially selecting and cultivating plant varieties to improve their agronomic traits and economic benefits (Herath et al., 2021). In the early days, farmers preserved and planted the seeds of the best-performing plants to grow crops in the next season, a natural selection process that facilitated the accumulation of favorable traits (Ibe, 2022). Over time, people gradually realized the importance of specific traits for crop yield, quality, disease resistance, and adaptability, and began consciously to select and breed plants with these characteristics. With the development of technology and improvement of living standards, higher demands have been placed on crop yield and quality, necessitating continuous innovation in breeding techniques,

methods, and applications to provide strong support (Wallace et al., 2018; Jiang et al., 2020). This has significant impacts and effects on agriculture and the economy, promoting sustainable agricultural development.

During the entire history of crop breeding technology, it has roughly gone through three major stages, and it is now advancing towards the fourth stage. The first stage is conventional breeding (Breeding 1.0), which mainly relies on visual observation of crop phenotypes and subjective selection of crops that meet predetermined requirements. Generally, wild species are gradually domesticated into cultivated varieties with improved qualities through multiple rounds of artificial selection (Khoshbakht and Hammer, 2008; Moose and Mumm, 2008). However, this stage primarily relies on natural variation and the subjective experience of breeders, resulting in slow progress, low efficiency, and high uncertainty. In the late 19th century, with the rapid development of genetics, genetic breeding (Breeding 2.0) emerged as the mainstream, bringing breeding into the realm of science. During this stage, significant success was achieved in crop breeding for crops like wheat, rice, maize, greatly improving yields. Unfortunately, there were still shortcomings such as long breeding cycles, low efficiency in genetic improvement, and high field costs (Zhang et al., 2014; Abdallah et al., 2015). At the end of the 20th century, genetic engineering propelled the development of modern molecular biology, ushering in the era of molecular breeding (Breeding 3.0). The gradual application of technologies such as transgenic techniques, molecular markers, genomic selection, and gene editing provided more efficient, precise, and targeted breeding methods. Nevertheless, high costs and complexity remain limiting factors for the application of molecular breeding (Jing et al., 2021). Breeding scientists sincerely hope that the integration of new generation information technologies such as big data and artificial intelligence with biotechnology will propel crop breeding into the era of Breeding 4.0, which also terms as Intelligent breeding (Wang et al., 2020) and is marked by Deep Learning-Empowered Crop Breeding (Yang et al., 2020; Wang et al., 2023).

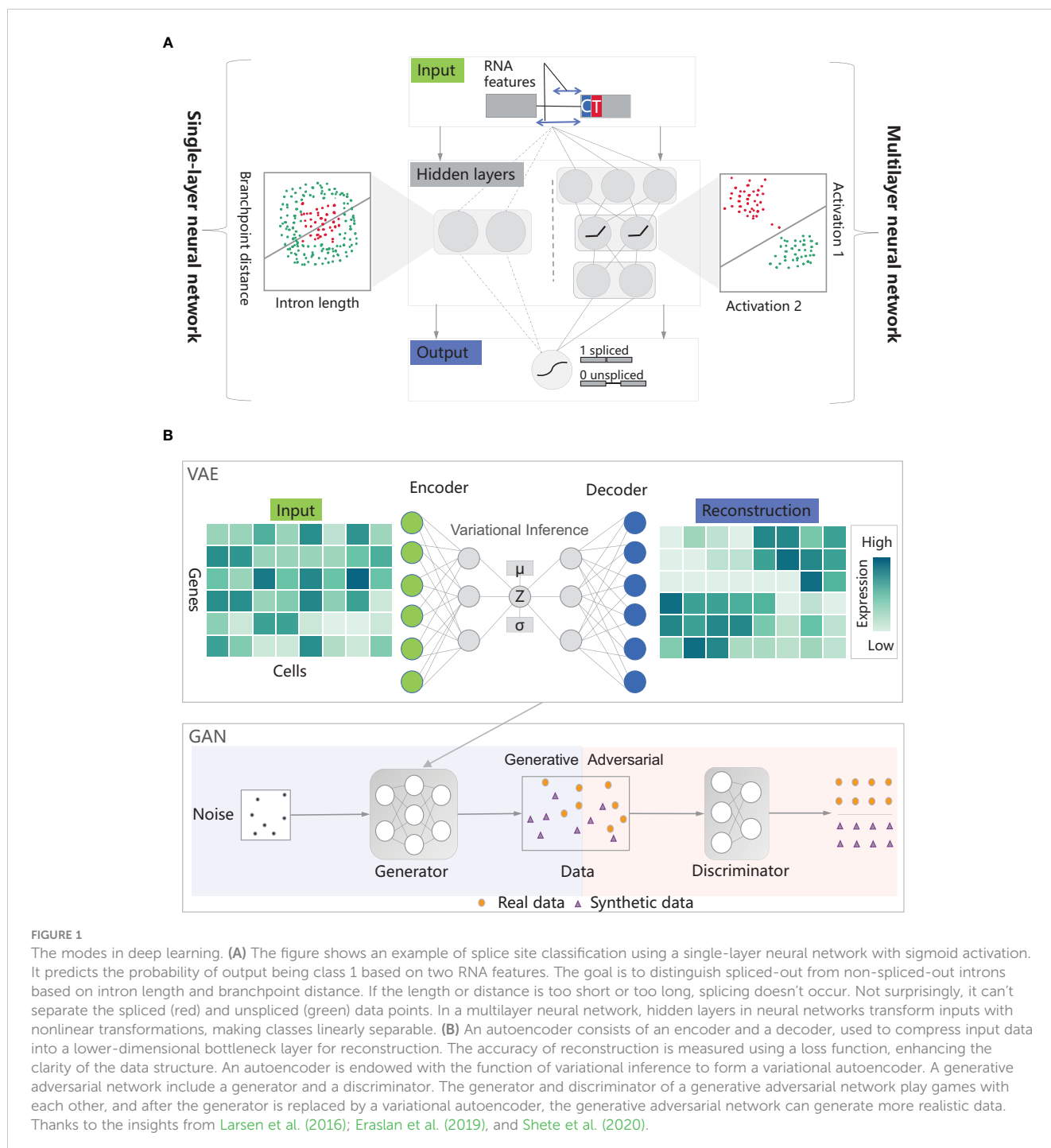
With the push of large-scale datasets, powerful computing capabilities, and algorithmic improvements, deep learning has made breakthrough progress in multiple fields (Khan et al., 2019). Deep learning is a machine learning method that revolves around the idea of building multi-layer neural network models to simulate the neural networks of the human brain, enabling the learning and pattern recognition abilities of data, which can be further applied to tasks such as classification, prediction, and generation (LeCun et al., 2015). Depending on whether the training data has label information, deep learning can be divided into two learning modes (Cunningham et al., 2008): (1) Supervised Learning: It relies on labeled training data, where labels correspond to the expected outputs or categories for each input sample. In this mode, explicit supervision signals are provided to the model, enabling it to learn the mapping relationship between inputs and outputs. The algorithms include Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Graph Convolutional Neural Networks (GCNs) (Yan and Wang, 2022). Neural networks are models used to capture nonlinear

dependencies. They transform inputs through hidden layers, mapping them to a space where classes can be linearly separated. In the example of splice site classification, a single-layer neural network employs logistic regression for prediction but fails to accurately differentiate spliced and unspliced data points. Surprisingly, by utilizing neural networks with intermediate layers, more complex nonlinear transformations can be performed, enabling effective discrimination between spliced and unspliced data points (Figure 1A). Deep neural networks, on the other hand, are neural network architectures that consist of multiple hidden layers (Miikkulainen et al., 2019). (2) Unsupervised Learning: It utilizes unlabeled training data. In this case, no explicit output or category information is given, and the goal is to discover hidden structures, patterns, or features from the data through the model's own learning process (Hastie et al., 2009). Unsupervised learning is commonly used for tasks such as clustering, dimensionality reduction, anomaly detection, and generative modeling (Fan et al., 2020). The algorithms include Autoencoders, Generative Adversarial Networks (GANs) and Variational AutoEncoders (VAEs). An autoencoder consists of an encoder and a decoder, used for data compression and reconstruction (Bank et al., 2020). The encoder compresses input data into lower dimensions and stores it in the bottleneck layer, while the decoder attempts to reconstruct the original input from the compressed data in the bottleneck layer. A generative adversarial network consists of a generator and a discriminator, trained together to generate realistic samples and perform discrimination (Creswell et al., 2018). The discriminator is responsible for distinguishing between real and synthetic data, while the generator aims to deceive the discriminator by generating more realistic synthetic samples (Figure 1B).

Deep learning-empowered breeding is a method that applies deep learning techniques to improve and optimize the breeding process. It utilizes deep learning models to analyze and process agricultural and genetic data, in order to predict and optimize the agricultural characteristics and genetic traits of crops (Uzal et al., 2018). Deep learning-assisted breeding can enhance breeding efficiency, accelerate the improvement process of crops or animals, and provide higher-yielding, more adaptive, and disease-resistant varieties for agricultural production, through steps such as data collection and preprocessing, model construction and training, as well as genetic parameter optimization and selection (Liu and Wang, 2017). The aim of this perspective is to provide an overview of the latest developments in deep learning in the field of crop breeding, analyze current challenges, and highlight its potential as a promising technology for crop breeding.

Principles of deep learning-empowered crop breeding

Deep learning solves complex problems by processing large-scale data. Currently, images remain the main data format for phenotypic selection in crop breeding (Araus and Cairns, 2014). The application of deep learning in plant phenotyping image processing is rapidly advancing, especially with the impressive



performance of CNN in analyzing phenotype big data ([Chang et al., 2016](#)). It possesses powerful feature extraction and modeling capabilities, providing new approaches for overcoming challenges in data analysis. The workflow for crop breeding based on deep learning generally includes six steps: (1) Data collection: Gather agricultural and genetic data relevant to the target crops, including phenotypic traits, environmental factors, genetic markers, and other related information ([Crossa et al., 2010](#)). (2) Data preprocessing: Clean and preprocess the collected data, involving data normalization, feature extraction, missing data handling, and data augmentation techniques ([SChadt et al., 2001](#)). (3) Model

construction: Build a deep learning model suitable for the specific breeding task, selecting appropriate neural network architectures based on the data nature and breeding objectives. (4) Model training: Train the deep learning model using preprocessed data, optimizing model parameters to minimize differences between predicted outputs and observed values. Training typically involves techniques like backpropagation and gradient descent to update model weights ([Zhou, 2018](#)). (5) Genetic parameter optimization and selection: Utilize the trained deep learning model to predict and evaluate agricultural characteristics and genetic traits of crops. Optimize genetic parameters and select suitable individuals for

further breeding based on the predictions. (6) Iterative improvement: Repeat the training, evaluation, and selection steps, iteratively improving the deep learning model and breeding process (Ni et al., 2019). This establishes an effective breeding plan, enabling the offspring of parental generations to approach the desired phenotypes.

Deep learning-empowered breeding is built upon important identified genes, integrating multiomics, next-generation biotechnologies, and novel information technologies such as artificial intelligence and big data. The identification of important genes using genetics and transgenic methods forms the foundation of deep learning-based breeding. New technologies, including multiomics, artificial intelligence, and big data, expedite the breeding process through plant phenotypic analysis and high-throughput phenotyping platforms, facilitate the evaluation of plant materials, discovery of specific genes, and accelerated breeding (Pan, 2015; Banerjee et al., 2020). Integration of multiomics data, encompassing genomics, metabolomics, phenomics, proteomics, and transcriptomics, aids in analyzing biological changes and regulatory processes, identifying key genes

and regulatory elements, and driving plant breeding (Yang et al., 2021). What's even more exciting is that Telomere-to-Telomere (T2T) complete genome and T2T whole-genome analysis serve as representative markers for accurately identifying genetic diversity and enhancing functional genomics and genetic improvement (Deng et al., 2022). Additionally, gene editing techniques have also contributed to breeding advancements (Li et al., 2018). Intelligent breeding strategies driven by big data and artificial intelligence, enable targeted breeding, such as through comprehensive genomic and environmental prediction (iGEP) based on genomics and population-environment interactions (Yin et al., 2008). Deep learning frameworks support automatic differentiation, enabling efficient implementation of these scores with just a few lines of code. They will assist in handling extensive multidimensional big data of genotype-phenotype-environment, facilitating efficient selection and cultivation of high-quality, disease-resistant new varieties (see Table 1). It should be pointed out that extensive genetic experiments with correlated phenotypic and environmental data are necessary (Parmley et al., 2019; Xu

TABLE 1 Deep learning algorithms and models used in intelligent breeding.

Categories (model)	Data	Application	
CNN (DeepBind)	DNA sequences	DNA and Gene Characteristics (Alipanahi et al., 2015)	
CNN (DeepSEA)	DNA sequences	DNA and Gene Characteristics (Zhou and Troyanskaya, 2015)	
Bi-GRU and LSTM (DeepCpG)	DNA sequences	DNA and Gene Characteristics (Angermueller et al., 2017)	
CNN and RNN (BiRen)	DNA sequences	DNA and Gene Characteristics (Yang et al., 2017)	
CNN (Basset)	DNA sequences	DNA and Gene Characteristics (Kelley et al., 2018)	
RNN	DNA sequences, DNA probes	DNA and Gene Characteristics (Zhang et al., 2021a)	
ResNet and LSTM (DeepcycP)	DNA sequences	DNA and Gene Characteristics (Li et al., 2022a)	
DBN	Gene sequences	DNA and Gene Characteristics (Rachmatia et al., 2017)	
GCN	Genetic diagram	interaction	DNA and Gene Characteristics (Dutil et al., 2018)
CNN and LSTM (DeepNovo)	Peptide spectrum	sequences,	Protein Characteristics (Tran et al., 2017)
BiLSTM and SVM-rank (pNovo3)	Peptide spectrum	sequences,	Protein Characteristics (Yang et al., 2019)
GRU	spectrum		Protein Characteristics (Gessulat et al., 2019)
Stacked AutoEncoder and DNN	protein sequences		Protein Characteristics (Zhang and Kabuka, 2019)
RCNN	protein sequences		Protein Characteristics (Xu et al., 2021)
Evoformer and transformer (AlphaFold)	protein sequences		Protein Characteristics (Jumper et al., 2021)
GCN (Decagon)	protein-protein interaction network		Protein Characteristics (Zitnik et al., 2018)
GAN	DNA sequences		Protein Characteristics (Gupta and Zou, 2018)
Multilayer Perceptron and autoencoder	SNP		Genomics Variations (Xie et al., 2017)
GAN	RNA sequences		Genomics Variations (Ghahramani et al., 2018)
DCNN	DNA sequences		Genomics Variations (Kelley, 2020)
CNN (AMBER)	gene sequences		Genomics Variations (Zhang et al., 2021a)
RCNN	RGB image		Genomics Variations (Li et al., 2022b)
Multilayer Perceptron Classifier	DNA sequences		Genomics Variations (Chen et al., 2023)

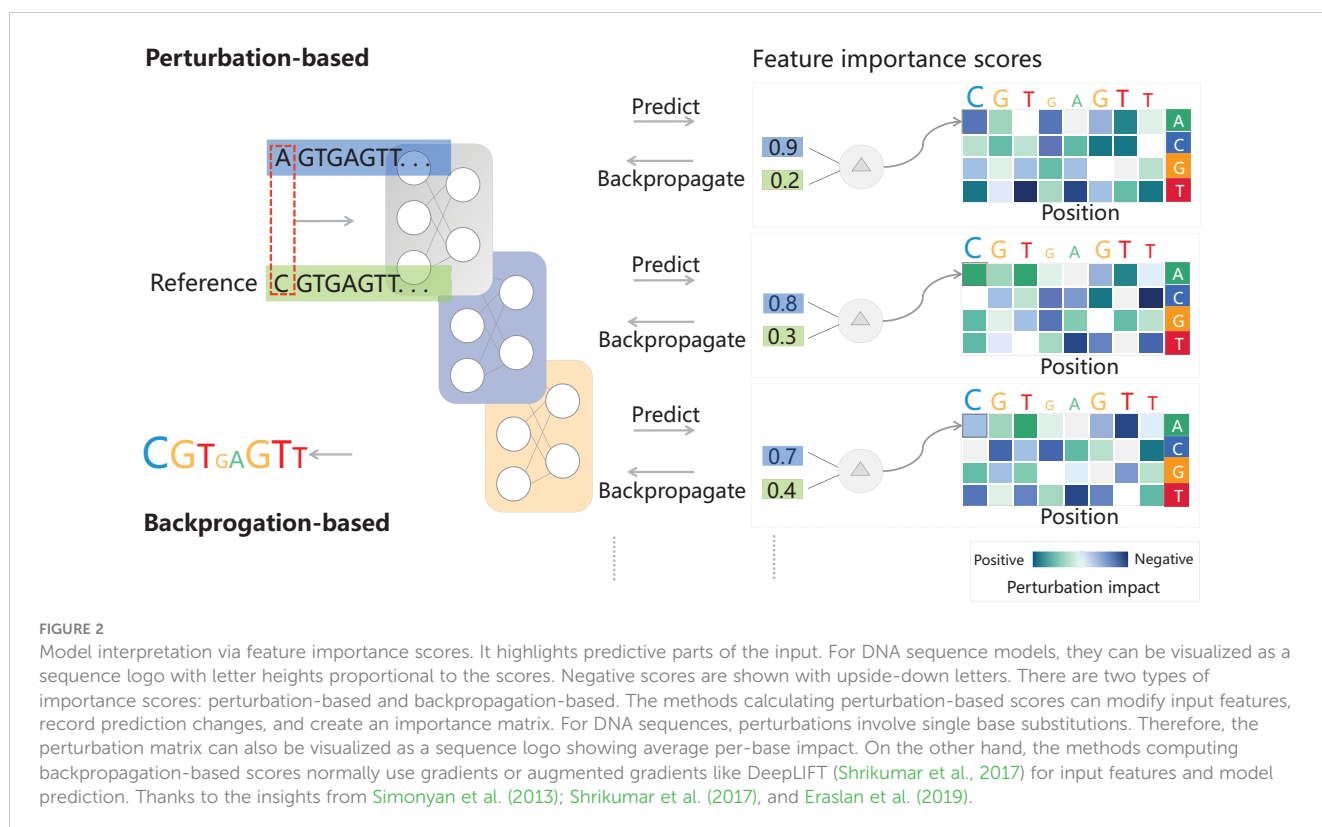
et al., 2021). It is also important, especially in complex models, to indirectly examine parameters by inspecting the input-output relationships for each predicted example. Feature importance scores highlight the most influential parts of a given input for model predictions, helping to explain why such predictions are made. In DNA sequence-based models, feature importance scores highlight sequence motifs and are widely used in genomics (Alipanahi et al., 2015; Kelley et al., 2016; Kelley et al., 2018). They can also be used to explore more complex epistatic interactions (Greenside et al., 2018). Feature importance scores can be divided into two categories: perturbation-based and backpropagation-based (Figure 2). The former perturbs input features and observe changes in the output, but it is computationally expensive. On the other hand, the latter calculates the importance scores for all input features using a single backpropagation pass, making them computationally efficient. The simplest backpropagation-based importance scores are saliency maps (Simonyan et al., 2013) and input-masked gradients (Shrikumar et al., 2017).

Applications of deep learning in crop breeding

DNA and gene characteristics

The shape of DNA plays an important role in the specificity of transcription factor (TF)-DNA binding (Lai et al., 2019), and deep learning models can utilize various types of data for analysis

(Zampieri et al., 2019). Understanding the sequence specificity of DNA and RNA binding proteins is crucial for biological system regulation models and pathogenic variant identification (Wang et al., 2020). There are currently several deep learning-based methods for predicting TF binding properties. DeepBind (Alipanahi et al., 2015), Basset (Kelley et al., 2018), and DeepSEA (Zhou and Troyanskaya, 2015) were among the earliest convolutional neural networks (CNNs) applied to genomic data analysis. DeepBind trained multiple single-task models to predict the binding affinities of transcription factors, while DeepSEA compiled a large set of chromatin maps for non-coding variants to study chromatin features, and Basset predicted DNA accessibility features. The impact of functional non-coding variants was evaluated in DeepSEA, DFIM (Greenside et al., 2018), and DeFine (Wang et al., 2018). This has always been considered a challenge to identify critical genomic regulatory regions in species with abundant repetitive elements and broad intergenic regions. To address this challenge, efficient and accurate annotation of regulatory regions in maize was achieved using methods based on natural language processing, such as k-mer grammar (Qin and Feng, 2017). These methods have played an important role in the prediction of functional non-coding variants, regulatory region annotation, and transcription factor binding site (TFBS) prediction. Machine learning models have proven to be efficient in plant biology, capable of being trained on various types of sequencing data while incorporating additional information, such as DNase I hypersensitivity data, to improve the prediction of *in vivo* transcription factor binding sites (Qin and Feng, 2017). In summary, CNNs have been widely applied in predicting molecular



phenotypes from DNA sequences and have become advanced models. They have been used for classifying transcription factor binding sites (Wang et al., 2018), chromatin function (Kelley et al., 2018), DNA contact mapping (Schreiber et al., 2017), DNA methylation (Angermueller et al., 2017; Zeng and Gifford, 2017), gene expression (Zhou et al., 2018), and RNA binding protein (Pan and Shen, 2017). Additionally, CNNs have been successfully applied to tasks such as RNA specificity prediction (Kim et al., 2018) and enhancing Hi-C data resolution (Zhang et al., 2018). Not surprisingly, CNNs can model long-range dependencies in the genome and improve the accuracy of predicting molecular phenotypes from linear DNA sequences through dilated convolutions (Zeng and Gifford, 2017). Interestingly, in addition to the CNN model, several other deep learning models are also used to analyze genetic characteristics. For instance, Angermueller et al. (2017) designed the DeepCpG model based on RNN, which can predict single-cell methylation states from local DNA sequences and observed neighboring methylation states. Zhang et al. (2021a) constructed a deep learning model to predict the depth of next generation sequencing according to the DNA probe sequences. Enhancer elements are non coding fragments of DNA that play a crucial role in controlling gene expression programs. Yang et al. (2017) proposed a hybrid BiRen architecture based deep learning, which only used DNA sequences to predict enhancers. Li et al. (2022b) constructed a deep model called DeepcycP that combines the Inception ResNet structure and LSTM layer, which can predict intrinsic DNA cyclization with high fidelity. Rachmatia et al. (2017) designed a deep learning algorithm DBN, which used whole-genome single-nucleotide polymorphism (SNP) as training and testing data to construct a genome prediction model. The results showed that the DBN algorithm had a correlation of 0.579 within the range of $[-1,1]$ with non additive features. Dutil et al. (2018) studied gene expression by deep learning and applied bias to the model using gene interaction maps, which has advantages in specific tasks within a low data range.

Protein characteristics

There is a close relationship between the function and structure of proteins. The function of a protein is determined by its tertiary structure, which can be revealed through comprehensive analysis of various protein characteristics. To extract important amino acid features from primary peptide sequences, DeepNovo (Tran et al., 2017) was developed using the CNN method. pNovo3 (Yang et al., 2019) utilizes a learning-to-rank framework to differentiate similar peptide candidates for each spectrum. It employs three metrics to measure the similarity between experimental and theoretical spectra, with the theoretical spectra precisely predicted through deep learning using the pDeep algorithm. In mass spectrometry-based proteomics, identification and quantification of peptides and proteins heavily rely on database searching and spectrum matching. The lack of accurate models for predicting fragment ion intensities limits the practicality of these methods. By expanding the ProteomeTools synthetic peptide library and training the deep neural network Prosit, the prediction accuracy of

chromatographic retention time and fragment ion intensities has been significantly improved (Gessulat et al., 2019). Gupta and Zou (2018) used GAN to generate DNA sequences for proteins with variable coding lengths, which have ideal biophysical properties. Protein-protein interactions are crucial for understanding biological processes and disease mechanisms. Researchers have explored various methods to predict protein-protein interactions, including sequence-based prediction techniques (Hashemifar et al., 2018) and deep learning models (Mirabello and Wallner, 2018). One approach involves unsupervised derivation of novel protein features from the “protein-protein” interaction network, followed by using these features to predict protein functions in different tissues (Zitnik and Leskovec, 2017). Zitnik et al. (2018) proposed the graph convolutional neural network model Decagon, and used this model to construct multimodal graphs of protein protein interactions, drug protein target interactions, and multi drug side effects. Some of these methods also incorporate physicochemical features of proteins and topological features of protein-protein interaction (PPI) networks to enhance predictive performance through multimodal supervised deep representation learning (Zhang and Kabuka, 2019). A novel residue representation method called Res2vec has been designed to represent protein sequences, combining effective feature embedding and powerful deep learning techniques, providing a universal computational pipeline for inferring “protein-protein” interactions (Longwell and Shimko, 2022). The confidence score of a protein sequence pair can be regarded as a measure of PPI. Therefore, a deep learning framework (Xu et al., 2021), namely the ordinal regression and recursive convolutional neural network approach, has been introduced to predict PPI from the perspective of confidence. Analysis of co-variation in homologous sequences aids in predicting protein structures. AlphaFold is an algorithm that predicts protein structures using deep learning methods, training neural networks to predict distances between residues and generate protein structures (Jumper et al., 2021). AlphaFold2 is an improved version of AlphaFold, greatly enhancing the accuracy of protein structure prediction by introducing new neural network architectures and training procedures (Tunyasuvunakool et al., 2021).

Genomics variations

Despite the presence of numerous genetic variations in natural populations, it is possible to train deep learning models on a small subset of these variations to predict the effects across the entire spectrum of mutations (Killamsetty et al., 2021). For instance, models trained on certain genes can be used to predict the outcomes of other genes. These models encompass various types of mutations, including common alleles as well as rare and low-frequency variants, regardless of their impact on gene function. Xie et al. (2017) constructed a deep automatic encoder model to evaluate the impact of genetic variation on gene expression changes. Li et al. (2022b) developed an image-based wheat spike counter using the Faster R-CNN algorithm, revealing significant differences between genotypes. ExPecto is a deep learning

framework that accurately predicts the transcriptional effects of mutations in DNA sequences, including rare or unobserved mutations (Zhou et al., 2018). This framework enables initial predictions for exploring the evolutionary constraints on gene expression and the effects of mutational diseases. Furthermore, models trained in one species can be directly applied to closely related species (Kelley, 2020), due to the conservation of molecular processes in closely related species. Chen et al. (2023) proposed an unsupervised clustering method and developed a deep learning model accordingly to predict gene mutations. Ghahramani et al. (2018) used GAN to simulate gene expression and predict perturbations in single cells, thereby identifying biological state determining genes and ultimately inferring gene regulatory relationships. A biologically-informed automated modeling framework, known as AMBER (Zhang et al., 2021b), has been proposed. It is a fully automated framework that efficiently designs and applies CNNs to genomic sequences. AMBER utilizes state-of-the-art neural architecture search to design optimal models for specified biological questions. Applying AMBER to modeling tasks of genomic regulatory features has demonstrated significantly more accurate predictions compared to non-neural architecture search baseline models, matching or even surpassing expert-designed models. In summary, deep learning models have the potential to greatly advance our understanding of genomic variations in relation to the ultimate phenotypes.

The impact on crop quality and yield of deep learning-empowered crop breeding

Climate change is seriously hindering the development of agricultural productivity globally, with significant impacts on crop yield and quality (Praveen and Sharma, 2019). Analyzing and identifying crop images using deep learning models can aid in rapidly identifying superior plants with target traits, thereby accelerating the process of crop breeding and selecting high-yielding, disease-resistant, and other desirable varieties. Deep learning models can also recognize crop performance under adverse environmental conditions such as drought and salinity stress, helping to cultivate more resilient crop varieties (Sun et al., 2019). Specifically, the identification and classification algorithm for corn leaf blight achieved high accuracy using the CNN algorithm, which is of significant importance for rapid detection of crop diseases and improving crop yields (Abdullahi et al., 2017). The solutions for crop disease identification and diagnosis were provided using different deep learning methods (Ferentinos, 2018). In terms of abiotic stress, the extraction of time-series chlorophyll fluorescence features using the SAE neural network algorithm provided an effective means for identifying chlorophyll fluorescence fingerprints in *Arabidopsis thaliana*, offering new insights for improving crop drought resistance (Sun et al., 2019). By using a large number of soybean leaf images for deep learning classification, the identification and classification of soybean symptoms under non-biological stress was achieved, enabling

rapid detection of non-biological stress in soybeans (Ghosal et al., 2018). The good correlation between the classification of corn freeze damage based on spectral features of multiple genotypes and the discrimination results based on chemical values was demonstrated using a CNN model (Yang et al., 2019). Employing various deep learning methods for diagnosing pumpkin leaf diseases helps farmers detect crop damage early (Nirmala and Gomathy, 2019). Using an integrated classifier based on a deep convolutional neural network for identifying citrus pests has effectively enhanced the quality and yield of citrus fruits (Khanramaki et al., 2021). Developing a model to estimate the number of leaves and plant age for watermelon plants, and classifying them under normal and low-temperature stress, facilitates growth monitoring and improves water and sugar content in watermelons (Nabwire et al., 2022). Training deep learning models to classify coffee leaf images and determine if they are infected with leaf rust disease can aid in early detection of diseases and enables timely measures to protect coffee crop yields and quality (Shao et al., 2022). Furthermore, deep learning can analyze the correlation between phenotypic genomic, facilitating precise selection and optimization of genomic combinations as well as gene editing techniques to improve crop yields and quality. For instance, deep learning-based genomic selection models (GS) have shown outstanding performance in predicting wheat terminal quality traits, advancing the deployment of superior genotypes into broader grain yield trials (Sandhu et al., 2021). Therefore, deep learning can analyze massive amounts of data, build intelligent breeding decisions, and rapidly create superior inbred lines, effectively shortening breeding cycles, improving breeding efficiency, reducing costs, and enhancing crop yields and quality.

Challenges and prospects

In the past 20 years, machine learning has achieved significant success in the field of agriculture. In recent years, deep learning, as a branch of machine learning, has represented the most advanced technology in smart agriculture (Kamilaris and Prenafeta-Boldú, 2018). As an integral part of agriculture, deep learning has been widely applied to various plant phenomics, such as image classification (Ramcharan et al., 2017), object detection (Ghosal et al., 2019), and semantic segmentation (Aich and Stavness, 2017). Consequently, it has tremendous potential in predicting plant growth, estimating yield, detecting maturity, and perceiving biotic/abiotic stresses. However, deep learning algorithms require a large amount of labeled data (Cordts et al., 2016), and data acquisition comes at a high cost, especially when identifying numerous categories (Tong et al., 2022) or subtle differences between categories (Taghavi Namin et al., 2018). Furthermore, collecting phenotype data faces additional obstacles of severe occlusion and various lighting conditions (Scharf et al., 2016), which increase the time required for obtaining the necessary annotations. Genotypic, phenotypic, and environmental big data form the core of intelligent breeding design (Talbot et al., 2017). Nevertheless, there is a significant shortage of accumulated phenotype data, and the problems with traditional manual

detection are increasingly prominent, necessitating a balanced consideration of accuracy, throughput, and cost (Liu and Wang, 2021). It is anticipated that breakthroughs and innovations in next-generation sensors and robotics will serve as underlying driving technologies to accelerate the acquisition of crop phenotype big data (Sony et al., 2019). By utilizing bio-sensors and agricultural robots, continuous detection of multiple traits is achieved, leading to improved detection accuracy, but the development of sensors and robots also faces certain challenges. Firstly, the working mechanisms and conditions may vary significantly. Even within the same species and variety, robot components may require adjustments or replacements (De Preter et al., 2018), reducing the universality of robots. Most studies only report simulations, experiments, preliminary results, and specifications related to hardware/software design. In contrast, only a few studies discuss commercial solutions (Bagagiolo et al., 2022). Additionally, the efficiency of the sensors and robots used is not high. If local labor is inexpensive, there is an unacceptable risk associated with using sensors and robots. Currently, a better solution may be to enable collaboration between workers and robots (Bragança et al., 2019).

The availability of massive big data enables informed decision-making, however the adaptability of deep learning models across different crops and environmental conditions may be limited (Khaki and Wang, 2019). Due to the differences in crop genetics and environmental factors, model transfer from one crop to another may require additional adjustments and optimization (Ghazi et al., 2017). This poses challenges for the application of deep learning in intelligent analysis and interpretation. Big data includes plant phenotypes, genetic genotypes, environmental parameters, diseases, pest conditions, and more. What's frustrating is that, the acquisition and processing of plant data lag behind research needs, limiting the development of intelligent breeding and functional plant genomics. Furthermore, they are not organically integrated, which hinders informed decision-making. In the future, researchers should collectively strive to establish a large-scale database, and interdisciplinary collaboration and data sharing can unlock greater potential for deep learning in breeding applications, benefiting more people from big data (Kim, 2019). In addition, transfer learning (Pan and Yang, 2009) and few-shot learning (Snell et al., 2017) will be effective approaches to alleviate the deep learning models' dependency on massive datasets. Transfer learning aims to transfer knowledge accumulated from a source task with ample labeled data to a new or similar target task, particularly when training data is limited. Notably, when the source and target domains exhibit strong similarity, transfer learning can provide a more economical and expedited solution to address the constraints of scarce training data (Sun et al., 2018). The most distinct characteristic of few-shot learning is its capacity for "learning to learn", achieved by emulating human-level concept learning, meaning that acquiring new concepts requires only a small number of labeled examples (Chen et al., 2019). Approaches like data augmentation (Shorten and Khoshgoftaar, 2019), image segmentation (Minaee et al., 2021), and attention mechanisms (Niu et al., 2021) can be used to solve the problem of severe occlusion in collected phenotype data, and improve the performance of deep learning models when facing such problems.

Deep reinforcement learning is the process of making intelligent decisions through reinforcement learning on the basis of deep learning (Shaikh et al., 2022). By using deep reinforcement learning to plan the robot's path and make decisions on its actions during its journey, the robot can efficiently assist farmers in crop data collection, crop picking, transportation, watering, and fertilization operations. No doubt, addressing key issues related to accurate collection, intelligent analysis of crop deep phenotype, and intelligent decision-making for precision agriculture on this basis will be of significant importance to the research of intelligent breeding.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Ethics statement

Written informed consent was obtained from the individual(s) for the publication of any identifiable images or data included in this article.

Author contributions

XW: Formal Analysis, Software, Visualization, Writing – original draft. HZ: Formal Analysis, Software, Visualization, Writing – original draft. LL: Visualization, Writing – original draft. YH: Visualization, Writing – original draft. HL: Conceptualization, Methodology, Supervision, Visualization, Writing – original draft. YQ: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

Funding

The authors declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Key Research and Development Program of China (2022YFD2301100 and 2019YFD1000503), the Special Fund for Science and Technology Innovation of Fujian Agriculture and Forestry University (CXZX2020081A), and the Agriculture Research System of China (CARS-17).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Abdallah, N. A., Prakash, C. S., and McHughen, A. G. (2015). Genome editing for crop improvement: challenges and opportunities. *GM Crops & Food* 6, 183–205.
- Abdullahi, H. S., Sheriff, R., and Mahieddine, F. (2017). "Convolution neural network in precision agriculture for plant image recognition and classification," in *2017 Seventh International Conference on Innovative Computing Technology (INTECH)*, Piscataway, New Jersey. Vol. 10. 256–272 (Ieee).
- Aich, S., and Stavness, I. (2017). "Leaf counting with deep convolutional and deconvolutional networks," in *Proceedings of the IEEE international conference on computer vision workshops*. (Piscataway: IEEE), 2080–2089.
- Alipanahi, B., Delong, A., Weirauch, M. T., and Frey, B. J. (2015). Predicting the sequence specificities of dna-and rna-binding proteins by deep learning. *Nat. Biotechnol.* 33, 831–838. doi: 10.1038/nbt.3300
- Angermueller, C., Lee, H. J., Reik, W., and Stegle, O. (2017). Deepcp: accurate prediction of single-cell dna methylation states using deep learning. *Genome Biol.* 18, 1–13. doi: 10.1186/s13059-017-1189-z
- Araus, J. L., and Cairns, J. E. (2014). Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci.* 19, 52–61. doi: 10.1016/j.tplants.2013.09.008
- Bagagiolo, G., Matrangola, G., Cavallo, E., and Pampuro, N. (2022). Greenhouse robots: Ultimate solutions to improve automation in protected cropping systems, a review. *Sustainability* 14, 6436. doi: 10.3390/su14116436
- Banerjee, B. P., Joshi, S., Thoday-Kennedy, E., Pasam, R. K., Tibbits, J., Hayden, M., et al. (2020). Highthroughput phenotyping using digital and hyperspectral imaging-derived biomarkers for genotypic nitrogen response. *J. Exp. Bot.* 71, 4604–4615. doi: 10.1093/jxb/eraa143
- Bank, D., Koenigstein, N., and Giry, R. (2020). Autoencoders. *arXiv preprint arXiv. 2003.05991*, 353–374. 2003.05991. doi: 10.1007/978-3-031-24628-9_16
- Bragança, S., Costa, E., Castellucci, I., and Arezes, P. M. (2019). A brief overview of the use of collaborative robots in industry 4.0: Human role and safety. *Occup. Environ. Saf. Health* 202, 641–650. doi: 10.1007/978-3-030-14730-3_68
- Chang, L., Deng, X.-M., Zhou, M.-Q., Wu, Z.-K., Yuan, Y., Yang, S., et al. (2016). Convolutional neural networks in image understanding. *Acta Automatica Sin.* 42, 1300–1312. doi: 10.16383/j.aas.2016.c150800
- Chen, Z., Li, X., Yang, M., Zhang, H., and Xu, X. S. (2023). Optimization of deep learning models for the prediction of gene mutations using unsupervised clustering. *J. Pathol.: Clin. Res.* 9, 3–17. doi: 10.1002/cjp.2302
- Chen, W.-Y., Liu, Y.-C., Kira, Z., Wang, Y.-C. F., and Huang, J.-B. (2019). A closer look at few-shot classification. *arXiv preprint arXiv. 1904.04232*, 1–17. doi: 10.48550/arXiv.1904.04232
- Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., et al. (2016). "The cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (Los Alamitos, California: IEEE Computer Society), 3213–3223.
- Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., and Bharath, A. A. (2018). Generative adversarial networks: An overview. *IEEE Signal Process. Mag.* 35, 53–65. doi: 10.1109/MSP.2017.2765202
- Crossa, J., de Los Campos, G., Pérez, P., Gianola, D., Burgueno, J., Araus, J. L., et al. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics* 186, 713–724. doi: 10.1534/genetics.110.118521
- Cunningham, P., Cord, M., and Delany, S. J. (2008). "Supervised learning," in *Machine learning techniques for multimedia: case studies on organization and retrieval* (Berlin: Springer), 21–49.
- Deng, Y., Liu, S., Zhang, Y., Tan, J., Li, X., Chu, X., et al. (2022). A telomere-to-telomere gap-free reference genome of watermelon and its mutation library provide important resources for gene discovery and breeding. *Mol. Plant* 15, 1268–1284. doi: 10.1016/j.molp.2022.06.010
- De Preter, A., Anthonis, J., and De Baerdemaeker, J. (2018). Development of a robot for harvesting strawberries. *IFAC-PapersOnLine* 51, 14–19. doi: 10.1016/j.ifacol.2018.08.054
- Dutil, F., Cohen, J. P., Weiss, M., Derevyanko, G., and Bengio, Y. (2018). Towards gene expression convolutions using gene interaction graphs. *arXiv preprint arXiv. 1806.06975*, 1–5. doi: 10.48550/arXiv.1806.06975
- Eraslan, G., Avsec, Ž., Gagneur, J., and Theis, F. J. (2019). Deep learning: new computational modelling techniques for genomics. *Nat. Rev. Genet.* 20, 389–403. doi: 10.1038/s41576-019-0122-6
- Fan, H., Zhang, F., Wang, R., Xi, L., and Li, Z. (2020). "Correlation-aware deep generative model for unsupervised anomaly detection," in *Advances in Knowledge Discovery and Data Mining: 24th Pacific-Asia Conference, PAKDD 2020*, Singapore, May 11–14, 2020, Proceedings, Part II 24. (Cham: Springer), 688–700.
- Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* 145, 311–318. doi: 10.1016/j.compag.2018.01.009
- Gessulat, S., Schmidt, T., Zolg, D. P., Samaras, P., Schnatbaum, K., Zerweck, J., et al. (2019). Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nat. Methods* 16, 509–518. doi: 10.1038/s41592-019-0426-7
- Ghahramani, A., Watt, F. M., and Luscombe, N. M. (2018). Generative adversarial networks simulate gene expression and predict perturbations in single cells. *BioRxiv*, 262501. doi: 10.1101/262501
- Ghazi, M. M., Yanikoglu, B., and Aptoula, E. (2017). Plant identification using deep neural networks via optimization of transfer learning parameters. *Neurocomputing* 235, 228–235. doi: 10.1016/j.neucom.2017.01.018
- Ghosal, S., Blystone, D., Singh, A. K., Ganapathysubramanian, B., Singh, A., and Sarkar, S. (2018). An explainable deep machine vision framework for plant stress phenotyping. *Proc. Natl. Acad. Sci.* 115, 4613–4618. doi: 10.1073/pnas.1716999115
- Ghosal, S., Zheng, B., Chapman, S. C., Potgieter, A. B., Jordan, D. R., Wang, X., et al. (2019). A weakly supervised deep learning framework for sorghum head detection and counting. *Plant Phenom.* 1–14. doi: 10.34133/2019/1525874
- Greenside, P., Shimko, T., Fordyce, P., and Kundaje, A. (2018). Discovering epistatic feature interactions from neural network models of regulatory dna sequences. *Bioinformatics* 34, i629–i637. doi: 10.1093/bioinformatics/bty575
- Gupta, A., and Zou, J. (2018). Feedback gan (fbgan) for dna: A novel feedback-loop architecture for optimizing protein functions. *arXiv preprint arXiv. 1804.01694*, 1–15. doi: 10.48550/arXiv.1804.01694
- Hashemifar, S., Neyshabur, B., Khan, A. A., and Xu, J. (2018). Predicting protein-protein interactions through sequence-based deep learning. *Bioinformatics* 34, i802–i810. doi: 10.1093/bioinformatics/bty573
- Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Tibshirani, R., and Friedman, J. (2009). "Unsupervised learning," in *The elements of statistical learning: Data mining, inference, and prediction* (New York: Springer), 485–585.
- Herath, H. N., Rafii, M., Ismail, S., Jj, N., and Ramlee, S. (2021). Improvement of important economic traits in chilli through heterosis breeding: a review. *J. Hortic. Sci. Biotechnol.* 96, 14–23. doi: 10.1080/14620316.2020.1780162
- Ibe, C. N. (2022). Democratizing plant genomics to accelerate global food production. *Nat. Genet.* 54, 911–913. doi: 10.1038/s41588-022-01122-y
- Jiang, S., Cheng, Q., Yan, J., Fu, R., and Wang, X. (2020). Genome optimization for improvement of maize breeding. *Theor. Appl. Genet.* 133, 1491–1502. doi: 10.1007/s00122-019-03493-z
- Jing, H., Tian, Z., Chong, K., and Li, J. (2021). Progress and perspective of molecular design breeding. *Sci. China Life Sci.* 51, 1356. doi: 10.1360/SSV-2021-0214
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with alphafold. *Nature* 596, 583–589. doi: 10.1038/s41586-021-03819-2
- Kamilaris, A., and Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Comput. Electron. Agric.* 147, 70–90. doi: 10.1016/j.compag.2018.02.016
- Kelley, D. R. (2020). Cross-species regulatory sequence activity prediction. *PLoSComputationalBiology* 16, e1008050. doi: 10.1371/journal.pcbi.1008050
- Kelley, D. R., Reshef, Y. A., Bileschi, M., Belanger, D., McLean, C. Y., and Snoek, J. (2018). Sequential regulatory activity prediction across chromosomes with convolutional neural networks. *Genome Res.* 28, 739–750. doi: 10.1101/gr.227819.117
- Kelley, D. R., Snoek, J., and Rinn, J. L. (2016). Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks. *Genome Res.* 26, 990–999. doi: 10.1101/gr.200535.115
- Khaki, S., and Wang, L. (2019). Crop yield prediction using deep neural networks. *Front. Plant Sci.* 10, 621. doi: 10.3389/fpls.2019.00621
- Khan, M., Jan, B., Farman, H., Ahmad, J., Farman, H., and Jan, Z. (2019). "Deep learning methods and applications," in *Deep learning: convergence to big data analytics* (Singapore: Springer), 31–42.
- Khanramaki, M., Asli-Ardeh, E. A., and Kozegar, E. (2021). Citrus pests classification using an ensemble of deep learning models. *Comput. Electron. Agric.* 186, 106192. doi: 10.1016/j.compag.2021.106192

- Khoshbakht, K., and Hammer, K. (2008). How many plant species are cultivated? *Genet. Resour. Crop Evol.* 55, 925–928. doi: 10.1007/s10722-008-9368-0
- Killamsetty, K., Sivasubramanian, D., Ramakrishnan, G., and Iyer, R. (2021). “Glist: Generalization based data subset selection for efficient and robust learning,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, (Menlo Park, Calif: AAAI Press), 35, 8110–8118.
- Kim, J. (2019). Overview of disciplinary data sharing practices and promotion of open data in science. *Sci. Ed* 6, 3–9. doi: 10.6087/kcse.149
- Kim, H. K., Min, S., Song, M., Jung, S., Choi, J. W., Kim, Y., et al. (2018). Deep learning improves prediction of crispr–cpf1 guide rna activity. *Nat. Biotechnol.* 36, 239–241. doi: 10.1038/nbt.4061
- Lai, X., Stigliani, A., Vachon, G., Carles, C., Smaczniak, C., Zubieta, C., et al. (2019). Building transcription factor binding site models to understand gene regulation in plants. *Mol. Plant* 12, 743–763. doi: 10.1016/j.molp.2018.10.010
- Larsen, A. B. L., Sønderby, S. K., Larochelle, H., and Winther, O. (2016). “Autoencoding beyond pixels using a learned similarity metric,” in *International conference on machine learning (PMLR)*. (Cambridge: JMLR), 1558–1566.
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature* 521, 436–444. doi: 10.1038/nature14539
- Li, K., Carroll, M., Vafabakhsh, R., Wang, X. A., and Wang, J.-P. (2022a). Dnacycp: a deep learning tool for dna cyclizability prediction. *Nucleic Acids Res.* 50, 3142–3154. doi: 10.1093/nar/gkac162
- Li, L., Hassan, M. A., Yang, S., Jing, F., Yang, M., Rasheed, A., et al. (2022b). Development of imagebased wheat spike counter through a faster r-cnn algorithm and application for genetic studies. *Crop J.* 10, 1303–1311. doi: 10.1016/j.cj.2022.07.007
- Li, T., Yang, X., Yu, Y., Si, X., Zhai, X., Zhang, H., et al. (2018). Domestication of wild tomato is accelerated by genome editing. *Nat. Biotechnol.* 36, 1160–1163. doi: 10.1038/nbt.4273
- Liu, Y., and Wang, D. (2017). “Application of deep learning in genomic selection,” in *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. (Piscataway: IEEE), 2280.
- Liu, J., and Wang, X. (2021). Plant diseases and pests detection based on deep learning: a review. *Plant Methods* 17, 1–18. doi: 10.1186/s13007-021-00722-9
- Longwell, S., and Shimko, T. (2022). Res2vec: Amino acid vector embeddings from 3d-protein structure. *THRESHOLD* 30, 22–34.
- Miikkulainen, R., Liang, J., Meyerson, E., Rawal, A., Fink, D., Francon, O., et al. (2019). Evolving deep neural networks. *Artif. Intell. age Neural Networks Brain computing*, 293–312. doi: 10.1016/B978-0-12-815480-9.00015-3
- Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., and Terzopoulos, D. (2021). Image segmentation using deep learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 44, 3523–3542. doi: 10.1109/TPAMI.2021.3059968
- Mirabello, C., and Wallner, B. (2018). rawmsa: Proper deep learning makes protein sequence profiles and feature extraction obsolete. *bioRxiv*, 394437. doi: 10.1101/394437
- Moose, S. P., and Mumm, R. H. (2008). Molecular plant breeding as the foundation for 21st century crop improvement. *Plant Physiol.* 147, 969–977. doi: 10.1104/pp.108.118232
- Nabwire, S., Wakholi, C., Faqeerzada, M. A., Arief, M. A. A., Kim, M. S., Baek, I., et al. (2022). Estimation of cold stress, plant age, and number of leaves in watermelon plants using image analysis. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.847225
- Ni, P., Huang, N., Zhang, Z., Wang, D.-P., Liang, F., Miao, Y., et al. (2019). DeepSignal: detecting dna methylation state from nanopore sequencing reads using deep-learning. *Bioinformatics* 35, 4586–4595. doi: 10.1093/bioinformatics/btz276
- Nirmala, V., and Gomathy, B. (2019). Diagnosis of leaf disease in cucurbita gourd family using machine learning algorithms. *Int. J. Recent Technol. Eng.* 8, 6800–6804. doi: 10.35940/ijrte.C5232.098319
- Niu, Z., Zhong, G., and Yu, H. (2021). A review on the attention mechanism of deep learning. *Neurocomputing* 452, 48–62. doi: 10.1016/j.neucom.2021.03.091
- Pan, Y. (2015). Analysis of concepts and categories of plant phenome and phenomics. *Acta Agronom. Sin.* 41, 175–186. doi: 10.3724/SP.J.1006.2015.00175
- Pan, X., and Shen, H.-B. (2017). Rna-protein binding motifs mining with a new hybrid deep learning based cross-domain knowledge integration approach. *BMC Bioinf.* 18, 1–14. doi: 10.1186/s12859-017-1561-8
- Pan, S. J., and Yang, Q. (2009). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22, 1345–1359. doi: 10.1109/TKDE.2009.191
- Parmley, K. A., Higgins, R. H., Ganapathysubramanian, B., Sarkar, S., and Singh, A. K. (2019). Machine learning approach for prescriptive plant breeding. *Sci. Rep.* 9, 17132. doi: 10.1038/s41598-019-53451-4
- Praveen, B., and Sharma, P. (2019). A review of literature on climate change and its impacts on agriculture productivity. *J. Public Affairs* 19, e1960. doi: 10.1002/pa.1960
- Qin, Q., and Feng, J. (2017). Imputation for transcription factor binding predictions based on deep learning. *PLoS Comput. Biol.* 13, e1005403. doi: 10.1371/journal.pcbi.1005403
- Rachmatia, H., Kusuma, W., and Hasibuan, L. (2017). Prediction of maize phenotype based on whole-genome single nucleotide polymorphisms using deep belief networks. *J. Phys.: Conf. Ser. (IOP Publishing)* 835, 012003. doi: 10.1088/1742-6596/835/1/012003
- Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J., and Hughes, D. P. (2017). Deep learning for image-based cassava disease detection. *Front. Plant Sci.* 8, 1852. doi: 10.3389/fpls.2017.01852
- Sandhu, K. S., Aoun, M., Morris, C. F., and Carter, A. H. (2021). Genomic selection for end-use quality and processing traits in soft white winter wheat breeding program with machine and deep learning models. *Biology* 10, 689. doi: 10.3390/biology10070689
- Schadt, E. E., Li, C., Ellis, B., and Wong, W. H. (2001). Feature extraction and normalization algorithms for high-density oligonucleotide gene expression array data. *J. Cell. Biochem.* 84, 120–125. doi: 10.1002/jcb.10073
- Scharr, H., Minervini, M., French, A. P., Klukas, C., Kramer, D. M., Liu, X., et al. (2016). Leaf segmentation in plant phenotyping: a collation study. *Mach. Vision Appl.* 27, 585–606. doi: 10.1007/s00138-015-0737-3
- Schreiber, J., Libbrecht, M., Bilmes, J., and Noble, W. S. (2017). Nucleotide sequence and dnase1 sensitivity are predictive of 3d chromatin architecture. *BioRxiv*, 103614. doi: 10.1101/103614
- Shaikh, T. A., Rasool, T., and Lone, F. R. (2022). Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming. *Comput. Electron. Agric.* 198, 107119. doi: 10.1016/j.compag.2022.107119
- Shao, B., Hou, Y., Huang, N., Wang, W., Lu, X., and Jing, Y. (2022). “Deep learning based coffee beans quality screening,” in *2022 IEEE International Conference on e-Business Engineering (ICEBE)*. (Piscataway: IEEE), 271–275.
- Shen, Y., Zhou, G., Liang, C., and Tian, Z. (2022). Omics-based interdisciplinarity is accelerating plant breeding. *Curr. Opin. Plant Biol.* 66, 102167. doi: 10.1016/j.cpb.2021.102167
- Shete, S., Srinivasan, S., and Gonsalves, T. A. (2020). Tasselgan: An application of the generative adversarial model for creating field-based maize tassel data. *Plant Phenom.* 2020, 1–15. doi: 10.34133/2020/8309605
- Shorten, C., and Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *J. big Data* 6, 1–48. doi: 10.1186/s40537-019-0197-0
- Shrikumar, A., Greenside, P., and Kundaje, A. (2017). “Learning important features through propagating activation differences,” in *International conference on machine learning*. (Cambridge: PMLR), 3145–3153.
- Simonyan, K., Vedaldi, A., and Zisserman, A. (2013). Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv. 1312.6034*, 1–7. doi: 10.48550/arXiv.1312.6034
- Snell, J., Swersky, K., and Zemel, R. (2017). Prototypical networks for few-shot learning. *Adv. Neural Inf. Process. Syst.* 30, 4080–4090. doi: 10.48550/arXiv.1703.05175
- Sony, S., Laventure, S., and Sadhu, A. (2019). A literature review of next-generation smart sensing technology in structural health monitoring. *Struct. Control Health Monit.* 26, e2321. doi: 10.1002/stc.2321
- Sun, X., Nasrabadi, N. M., and Tran, T. D. (2018). “Supervised deep sparse coding networks,” in *2018 25th IEEE International Conference on Image Processing (ICIP)*. (Piscataway: IEEE), 346–350.
- Sun, D., Zhu, Y., Xu, H., He, Y., and Cen, H. (2019). Time-series chlorophyll fluorescence imaging reveals dynamic photosynthetic fingerprints of sos mutants to drought stress. *Sensors* 19, 2649. doi: 10.3390/s19122649
- Taghavi Namin, S., Esmailzadeh, M., Najafi, M., Brown, T. B., and Borevitz, J. O. (2018). Deep phenotyping: deep learning for temporal phenotype/genotype classification. *Plant Methods* 14, 1–14. doi: 10.1186/s13007-018-0333-4
- Talbot, B., Chen, T.-W., Zimmerman, S., Joost, S., Eckert, A. J., Crow, T. M., et al. (2017). Combining genotype, phenotype, and environment to infer potential candidate genes. *J. Heredity* 108, 207–216. doi: 10.1093/jhered/esw077
- Tong, Y.-S., Tou-Hong, L., and Yen, K.-S. (2022). Deep learning for image-based plant growth monitoring: A review. *Int. J. Eng. Technol. Innovation* 12, 225. doi: 10.46604/ijeti.2022.8865
- Tran, N. H., Zhang, X., Xin, L., Shan, B., and Li, M. (2017). De novo peptide sequencing by deep learning. *Proc. Natl. Acad. Sci.* 114, 8247–8252. doi: 10.1073/pnas.1705691114
- Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Židek, A., et al. (2021). Highly accurate protein structure prediction for the human proteome. *Nature* 596, 590–596. doi: 10.1038/s41586-021-03828-1
- Uzal, L. C., Grinblat, G. L., Namias, R., Lares, M. G., Bianchi, J. S., Morandi, E. N., et al. (2018). Seed-per-pod estimation for plant breeding using deep learning. *Comput. Electron. Agric.* 150, 196–204. doi: 10.1016/j.compag.2018.04.024
- Wallace, J. G., Rodgers-Melnick, E., and Buckler, E. S. (2018). On the road to breeding 4.0: unraveling the good, the bad, and the boring of crop quantitative genomics. *Annu. Rev. Genet.* 52, 421–444. doi: 10.1146/annurev-genet-120116-024846
- Wang, K., Abid, M. A., Rasheed, A., Crossa, J., Hearne, S., and Li, H. (2023). Dnngp, a deep neural network-based method for genomic prediction using multi-omics data in plants. *Mol. Plant* 16, 279–293. doi: 10.1016/j.molp.2022.11.004
- Wang, H., Cimen, E., Singh, N., and Buckler, E. (2020). Deep learning for plant genomics and crop improvement. *Curr. Opin. Plant Biol.* 54, 34–41. doi: 10.1016/j.cpb.2019.12.010
- Wang, M., Tai, C., Weinan, E., and Wei, L. (2018). Define: deep convolutional neural networks accurately quantify intensities of transcription factor-dna binding and facilitate evaluation of functional noncoding variants. *Nucleic Acids Res.* 46, e69. doi: 10.1093/nar/gky215

- Xie, R., Wen, J., Quitadamo, A., Cheng, J., and Shi, X. (2017). A deep auto-encoder model for gene expression prediction. *BMC Genomics* 18, 39–49. doi: 10.1186/s12864-017-4226-0
- Xu, W., Gao, Y., Wang, Y., and Guan, J. (2021). Protein–protein interaction prediction based on ordinal regression and recurrent convolutional neural networks. *BMC Bioinf.* 22, 1–20. doi: 10.1186/s12859-021-04369-0
- Yan, J., and Wang, X. (2022). Machine learning bridges omics sciences and plant breeding. *Trends Plant Sci.* 28, 199–211. doi: 10.1016/j.tplants.2022.08.018
- Yang, H., Chi, H., Zeng, W.-F., Zhou, W.-J., and He, S.-M. (2019). p novo 3: precise *de novo* peptide sequencing using a learning-to-rank framework. *Bioinformatics* 35, i183–i190. doi: 10.1093/bioinformatics/btz366
- Yang, Z., Gao, S., Xiao, F., Li, G., Ding, Y., Guo, Q., et al. (2020). Leaf to panicle ratio (lpr): a new physiological trait indicative of source and sink relation in japonica rice based on deep learning. *Plant Methods* 16, 1–15. doi: 10.1186/s13007-020-00660-y
- Yang, B., Liu, F., Ren, C., Ouyang, Z., Xie, Z., Bo, X., et al. (2017). Biren: predicting enhancers with a deep-learning-based model using the dna sequence alone. *Bioinformatics* 33, 1930–1936. doi: 10.1093/bioinformatics/btx105
- Yang, Y., Saand, M. A., Huang, L., Abdelaal, W. B., Zhang, J., Wu, Y., et al. (2021). Applications of multi-omics technologies for crop improvement. *Front. Plant Sci.* 12, 563953. doi: 10.3389/fpls.2021.563953
- Yin, J., Huo, L., Guo, L., and Hu, J. (2008). “Short-term load forecasting based on improved gene expression programming,” in *2008 7th World Congress on Intelligent Control and Automation*. (Piscataway: IEEE), 5647–5650.
- Zampieri, G., Vijayakumar, S., Yaneske, E., and Angione, C. (2019). Machine and deep learning meet genome-scale metabolic modeling. *PLoS Comput. Biol.* 15, e1007084. doi: 10.1371/journal.pcbi.1007084
- Zeng, H., and Gifford, D. K. (2017). Predicting the impact of non-coding variants on dna methylation. *Nucleic Acids Res.* 45, e99. doi: 10.1093/nar/gkx177
- Zhang, Y., An, L., Xu, J., Zhang, B., Zheng, W. J., Hu, M., et al. (2018). Enhancing hi-c data resolution with deep convolutional neural network hicplus. *Nat. Commun.* 9, 750. doi: 10.1038/s41467-018-03113-2
- Zhang, D., and Kabuka, M. (2019). Multimodal deep representation learning for protein interaction identification and protein family classification. *BMC Bioinf.* 20, 1–14. doi: 10.1186/s12859-019-3084-y
- Zhang, Z., Park, C. Y., Theesfeld, C. L., and Troyanskaya, O. G. (2021b). An automated framework for efficiently designing deep convolutional neural networks in genomics. *Nat. Mach. Intell.* 3, 392–400. doi: 10.1038/s42256-021-00316-z
- Zhang, F., Wen, Y., and Guo, X. (2014). Crispr/cas9 for genome editing: progress, implications and challenges. *Hum. Mol. Genet.* 23, R40–R46. doi: 10.1093/hmg/ddu125
- Zhang, J. X., Yordanov, B., Gaunt, A., Wang, M. X., Dai, P., Chen, Y.-J., et al. (2021a). A deep learning model for predicting next-generation sequencing depth from dna sequence. *Nat. Commun.* 12, 4387. doi: 10.1038/s41467-021-24497-8
- Zhou, X. (2018). Understanding the convolutional neural networks with gradient descent and backpropagation. *J. Phys.: Conf. Ser. (IOP Publishing)* 1004, 012028. doi: 10.1088/1742-6596/1004/1/012028
- Zhou, J., Theesfeld, C. L., Yao, K., Chen, K. M., Wong, A. K., and Troyanskaya, O. G. (2018). Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. *Nat. Genet.* 50, 1171–1179. doi: 10.1038/s41588-018-0160-6
- Zhou, J., and Troyanskaya, O. G. (2015). Predicting effects of noncoding variants with deep learning–based sequence model. *Nat. Methods* 12, 931–934. doi: 10.1038/nmeth.3547
- Zitnik, M., Agrawal, M., and Leskovec, J. (2018). Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics* 34, i457–i466. doi: 10.1093/bioinformatics/bty294
- Zitnik, M., and Leskovec, J. (2017). Predicting multicellular function through multi-layer tissue networks. *Bioinformatics* 33, i190–i198. doi: 10.1093/bioinformatics/btx252