

REVIEW PAPER

Machine learning for image-based multi-omics analysis of leaf veins

Yubin Zhang¹, Ning Zhang¹, Xiujuan Chai^{1,*} and Tan Sun^{2,3,*}

¹ Agricultural Information Institute, Chinese Academy of Agricultural Sciences, No.12 Zhongguancun South St, Beijing 100081, China

² Key Laboratory of Agricultural Big Data, Ministry of Agriculture and Rural Affairs, Beijing, China

³ Chinese Academy of Agricultural Sciences, No.12 Zhongguancun South St, Beijing 100081, China

* Correspondence: chaixiujuan@caas.cn or suntan@caas.cn

Received 1 April 2023; Editorial decision 19 June 2023; Accepted 29 June 2023

Editor: Tracy Lawson, University of Essex, UK

Abstract

Veins are a critical component of the plant growth and development system, playing an integral role in supporting and protecting leaves, as well as transporting water, nutrients, and photosynthetic products. A comprehensive understanding of the form and function of veins requires a dual approach that combines plant physiology with cutting-edge image recognition technology. The latest advancements in computer vision and machine learning have facilitated the creation of algorithms that can identify vein networks and explore their developmental progression. Here, we review the functional, environmental, and genetic factors associated with vein networks, along with the current status of research on image analysis. In addition, we discuss the methods of venous phenotype extraction and multi-omics association analysis using machine learning technology, which could provide a theoretical basis for improving crop productivity by optimizing the vein network architecture.

Keywords: Deep learning, enviromics analysis, growth prediction model, image analysis, multi-omics analysis, phenotype omics, vein network.

Introduction

Plants have a variety of traits that reflect their ability to adapt to the environment, including leaf morphology, photosynthesis, quality, and nutrient element transmission (Messier *et al.*, 2017). The leaf phenotype and leaf vein systems are closely related and feature prominently in plant growth performance and functions (Violle *et al.*, 2007; Wang and Jiao, 2020). The veins are the main supporting structure of the leaf blade, acting as pathways for the transport of water, minerals, and photosynthates (Stewart *et al.*, 2018a). The efficiency of the transport of water and inorganic salts is closely related to the surface area, dry matter content, and nitrogen content of the leaves (Ma

et al., 2021), and so the traits of veins significantly affect the leaves' phenotypic characteristics.

The vein network structure of a leaf is a significant morphological feature of the vein system, as it shows the organization and distribution of the vein system on the leaf. As shown in Supplementary Fig. S1, vein density (VD), vein spacing, vein diameter, and venous closure are all key factors influencing the efficiency of the vein system. Photosynthesis, water uptake, and defense against interference are all affected by the growth structure of leaf veins (Pan *et al.*, 2022). Significant changes in leaf veins could characterize the growth state of

plants (Brodrribb *et al.*, 2016). For example, an increased VD can lead to a higher stomatal density, faster gas exchange between the leaf and the environment, and a more efficient transport of photosynthetic products (McAdam *et al.*, 2017). High VD is linked to leaf structural traits and mechanical stress, as evidenced by research indicating a positive correlation between tearing force and VD (Kawai and Okada, 2020; Duarte *et al.*, 2023). The anatomical structure of leaf veins is an important factor in determining the physiological properties of plants. Studies have shown that leaf VD affects the efficiency of water transport to the stomata, which consequently affects the rate of carbon absorption and the nitrogen and carbon content of the lamina (Carvalho *et al.*, 2017; Hua *et al.*, 2020). Moreover, the presence of plenty of free-end veins can lead to greater hydraulic efficiency and photosynthesis. The topology of veins varies between species (Fiorin *et al.*, 2016; Aritsara *et al.*, 2022, Preprint). Furthermore, the traits of veins, as well as the size of the leaves, are crucial components of plant development and have been influential in the evolution of plant species. Thus, the study of leaf vein patterning and its connection to leaf properties can provide insights into plant growth and contribute to the development of the leaf economic spectrum theory (Ji *et al.*, 2020).

Environmental and genetic factors have been shown to have considerable effects on the size, thickness, photosynthetic products, and content of water and inorganic salts in leaves. The changes in these phenotypic traits are caused by leaf vein conduction (Beerling and Franks, 2010). The architecture of the vein network in a leaf has been shown to affect leaf growth, resource transport, and mechanical stress in different ecological contexts (Brodrribb *et al.*, 2010). However, the mechanisms underlying the interactions between leaf vein phenotype and leaf traits, resource use, and environmental adaptation, as well as the role of genetic factors in these interactions, are incompletely understood. This will undoubtedly pose a persistent challenge, which demands the coordinated advancement of cross-scale measurement of leaf vein phenotype, reliable image-extraction techniques, and intelligent big data analysis technologies (Xu *et al.*, 2021). The integration of artificial intelligence technologies, such as deep learning (DL) and computer vision, with fields such as plant physiology and biochemistry, plant ecology, and multi-omics joint analysis would effectively advance the in-depth investigation of plant phenotypes.

The influence of enviromics on vein phenomics

The term ‘enviromics’ originated from the field of human medicine and has been subsequently extended to include the study of plants. By using statistical modeling techniques, researchers can correlate distinct phenotypes with environmental factors and evaluate the effects of complex environmental variables on plants’ growth processes (Resende *et al.*,

2021). Investigating the reaction of leaf veins to environmental changes and exploring the adaptation mechanism of plants to their developing environment through vein phenotypes remains a central topic in plant ecology. Light is a crucial factor in plant development, which involves a series of photoreceptor-mediated responses (Franklin, 2016). Light quality regulates the growth of organs, including leaf veins, by modulating a diverse array of light receptors. For example, red and far-red light affect the elongation of leaves and central veins through phytochromes (Shafiq *et al.*, 2021), and UV-B inhibits stem and leaf growth through UVR8 (Rai *et al.*, 2019). When blue light strikes the photoreceptors phot1 and phot2, they initiate signal transduction pathways that increase the acidity of sap in xylem of Arabidopsis. This acidification leads to enhanced permeability of the bundle sheath cells, increasing water mobility throughout the leaf (Grunwald *et al.*, 2022). In addition, aquaporin expression and activation in veins, which affects water transport efficiency, is regulated by the light environment (Cochard *et al.*, 2007). Besides light quality, light intensity affects phenotypic changes in leaf vein growth. Changes in light intensity lead to modifications in leaf tube characteristics, resulting in alterations in the transport efficiency of water and inorganic salts, as well as transpiration and photosynthetic rates (Stewart *et al.*, 2018b). Variations in available light cause differences in the lengths of minor veins when the plant is growing and developing. In low-light conditions, plant species or leaves tend to have fewer veins. *Toona ciliata* (Meliaceae) tropical plants growing under $50 \mu\text{mol m}^{-2} \text{s}^{-1}$ exhibited a significant (33%) decrease in VD compared with the VD under $1500 \mu\text{mol m}^{-2} \text{s}^{-1}$ (Carins Murphy *et al.*, 2012). The quality and quantity of light are crucial factors that modulate plant growth and development, affecting various processes such as photosynthesis, physiological metabolism, morphogenesis, and reproductive development. Although it is well understood that leaf veins serve as supporting structures for leaf growth and as channels for nutrient transport, there is a lack of systematic research on the impact of the light environment on their growth and development. By exploring the effect of light on leaf vein phenotype, we can unlock vital insights into environmental regulation and breeding within controlled agriculture, such as controlled environment cultivation.

The water exchange between plants and their environment is a crucial determinant of plant growth and phenotype. The geometric structure of the leaf vein network is pivotal in this process and is inevitably impacted by various environmental factors (Sack and Scoffoni, 2013). Leaf water transport efficiency is primarily governed by vein traits including VD, topological structure, vein length (VL), and xylem vessel radius (Venturas *et al.*, 2017). Among these, alterations in VD are critical to maintaining a balance between plant water supply and the demands of evaporation. Prolonged drought conditions can lead to reduced stomatal size and increased VD, enabling faster responses to bolster water regulation (Li *et al.*, 2021). Evidence supports the notion that an increase in VD serves as the foundation for water

transport, photosynthesis, and increased leaf developmental rate (Sack *et al.*, 2004). However, limiting VD is necessary as it is controlled by cuticular conductance, leading to excessive water expenditure during severe drought (Scoffoni *et al.*, 2011). Furthermore, water loss through stomata and the cuticle during periods of drought can impair survival. In this case, high VD and coarse veins provide sufficient pathways for water transport, enhancing leaf hydraulic efficiency and drought resistance (Fiorin *et al.*, 2016). Coarse veins maintain consistent growth over an extended period, whereas thin veins exhibit greater adaptability to environmental changes (Ellis *et al.*, 2009). Vein size structure is also a key factor in the diffusion of embolism damage caused by drought (Brodribb *et al.*, 2016). Redundant pathways enable water transport to avoid xylem embolisms induced by drought while protecting the hydraulic system from vein damage (Brodribb and Jordan, 2011).

Water transfer efficiency in veins is a pathway for light and drought regulation, playing a vital role in regulating the overall growth of leaves. Leaf veins are a crucial hub for the exchange of nutrient substances and moisture between plants and their environment (Sack and Scoffoni, 2013). Systematically exploring the relationship between leaf veins and climate, growth form, and habitat is of great significance for unraveling the code of plant growth, improving crop varieties, guiding controlled agricultural production, and understanding the cyclic flow of the ecosystem.

The influence of molecular omics on vein phenomics

Vein tissue is composed of vascular bundles and peripheral mechanical tissues. Several genes that affect leaf vein formation have been identified to date. The *DOF* gene family has a significant role in the early stages of vein development (Gardiner *et al.*, 2010). The *PIN* gene family drives vein tissue elongation (Scarpella *et al.*, 2006), while midrib growth is controlled by the *DROOPING LEAF* gene, a member of the *YABBY* gene family (Yamaguchi *et al.*, 2004). Mutations in the *WOX* gene family cause changes in vein patterns in some dicotyledons (Vandenbussche *et al.*, 2009; Tadege *et al.*, 2011). *WOX* in the primordium directs the growth of vein topologies by regulating the location and timing of cell proliferation. The Class III homologous domain leucine zipper (HD-ZipIII) is a key regulator of vascular development and has been shown to control vein specification, patterning, and differentiation (Ruonala *et al.*, 2017; Manuela and Xu, 2020). This regulation involves a tightly controlled positive feedback loop of auxin networks, cell division, and vascular bundle formation (Prigge *et al.*, 2005; Schuetz *et al.*, 2019). Additionally, *NAL1*, *NAL9*, *NRL1*, and *NRL2* regulate the number of veins. LMI1-like and KNOX1-like transcription factors (TFs) synergistically coordinate vein pattern formation, and *UNHINED*, *FORKED1*, *FORKED-LIKE1* (*FKL1*), *FKL2*, and *FKL3* have been identified as

potential regulators of VD (Dhakal *et al.*, 2021). *GmREV-L-1* and *GmHB14-L-2* affect cell differentiation in the xylem of veins (Gao *et al.*, 2022). Genes are involved in the development of leaf vein phenotype by regulating specific signaling and metabolic pathways. TFs play a crucial role as key regulators of gene expression, enabling precise temporal and spatial control that governs the formation and differentiation of leaf veins. Therefore, genes and TFs are vital components in the development of the leaf vein phenotype.

In addition to genes and TFs, hormones and proteins are also involved in the development of vein phenotype. The roles of hormones and proteins in the development of vein phenotype extend beyond the regulation of gene expression; they can also directly affect vein formation and differentiation at the cellular and tissue levels. Induction of ectopic cell division affects auxin responses and venation patterning in the procambium and ground tissues (Wenzel *et al.*, 2007). The PIN protein-mediated transport of auxin results in the suppression of procambial development in the phloem by KAN TFs (Scarpella *et al.*, 2006; Ilegems *et al.*, 2010). COV1, a transmembrane protein found in the Golgi apparatus, inhibits vein development (Shirakawa *et al.*, 2014). Leaf vein systems play vital roles in plant nutrient and hormone transport as well as photosynthesis. Overexpression of *AAO3* increases abscisic acid accumulation (Koiwai *et al.*, 2004; Shi *et al.*, 2021), while three different cytosolic glutamine synthetase isoforms work together to efficiently reuse nitrogen (Moison *et al.*, 2018). Chloroplasts present in veins are a major site of photosynthetic activity, with *OBV*-encoded C_2H_2L regulating chloroplast growth and development in tomato veins (Lu *et al.*, 2021; Liu *et al.*, 2022; Song *et al.*, 2022). Several proteins, including OsNUS1 (Kusumi *et al.*, 2011), Osarv14 (Wang *et al.*, 2016), bml (Akhter *et al.*, 2018), and yml (Guo *et al.*, 2022), regulate chlorophyll concentration and chloroplast growth, affecting vein coloration.

The identification of vein gene function plays a crucial role in elucidating fundamental biological principles underlying the adaptation mechanisms of veins to environmental changes, regulatory mechanisms governing growth and development, and metabolic substance synthesis in plants. Additionally, vein gene function-based approaches provide a rational basis for harnessing advances in plant hybridization breeding and transgenic technology development, thereby increasing the efficiency of genetic improvement and enhancing crop productivity. Although the relationship between some genes and leaf vein development has been analyzed, traditional methods of genetic methods suffer from issues such as high time cost, incomplete data, and poor accuracy. The continuous development of artificial intelligence technology has paved the way for a new research direction in multi-omics collaborative analysis. For example, by integrating bioinformatics, genetics, plant ecology, and epigenetics, more detailed and comprehensive models of leaf vein pathways can be established. Therefore, multi-omics collaborative analysis will prove to be a critical tool for future leaf vein pathway research, ultimately advancing this field.

Extraction methods of vein phenomics

The purpose of analyzing leaf vein phenotypes is to aid in the study of leaf structure development and investigate the metabolic regulatory route of related phenotypes, which requires precise measurements of leaf vein phenotypes to infer the genes involved in regulation (Gardiner *et al.*, 2010). Extracting and examining leaf vein traits is difficult due to the complexity of vein images and the scarcity of straightforward linear characteristics in image annotation data (Blonder *et al.*, 2020). Non-invasive methods of image analysis facilitate automated and high-throughput measurement of various phenotypes (Nguyen *et al.*, 2022).

Vein identification techniques require precise vein network segmentation for accurate trait calculation. Depending on image complexity and consistency, feature recognition and measurement can be manual, semi-automated, or completely automated. Automating this process is effective for uniform and standard samples, but complexity and variations in sample features can lead to reduced accuracy and increased time consumption (Xu *et al.*, 2021). Vein geometry analyses form the foundation for understanding leaf morphological and functional diversity, as well as carbon balance. Despite its potential, automated vein data extraction faces challenges related to reduced accuracy and increased time requirement. In response to the challenges of automating vein identification techniques, researchers have turned to digital image processing as a solution to standardize leaf vein network characteristics. Advances in this field have allowed predictions to be made of functional traits relevant to leaf economic spectra, providing valuable insights into the complex world of leaf physiology (Rolland-Lagan *et al.*, 2009; Blonder *et al.*, 2011). Researchers have also developed the ‘nesting tree’ pipeline for dicotyledonous plants that separates geometric information (precise edge and node locations) from the topological structure (connectivity and edge weight) (Katifori and Magnasco, 2012). However, the lack of a comprehensive understanding of venation and ambiguous correlations between hierarchical tree attributes and underlying principles governing leaf function impede the practical application of this approach (Ronellenfisch *et al.*, 2015). Achieving effective segmentation is challenging due to poor contrast, variable quality, and unavoidable artifacts caused by other interfering tissues, making simple intensity thresholding ineffective.

Numerous algorithms for extracting leaf vein phenotypes have aided investigations into the relationship between veins and leaves. Automated algorithms such as LIMANI, LEAFGUI, and NET have been used to segment and measure vein patterns, including VD, free-end VL, endpoints, branchpoints, and areole area (Price *et al.*, 2011; Dhondt *et al.*, 2012; Lasser and Katifori, 2017). Vein features are associated with photosynthetic processing and plant weight (Malinowski, 2013) and are useful for comparing venation patterns across species (Rolland-Lagan *et al.*, 2009). Fiorin *et al.* (2016) combined image and

environmental data to examine vein and stomata distribution. Convolutional neural networks (CNNs) and artificial neural networks (ANNs) have been used for plant recognition and classification based on vein patterns (Grinblat *et al.*, 2016; Gan *et al.*, 2019; Salve *et al.*, 2021; Du *et al.*, 2022; Balasundaram *et al.*, 2023). The specific vein traits and segmentation accuracy of these methods are listed in Table 1. These methods offer an effective approach for analyzing the intricate structure and composition of leaf veins, providing crucial insights for enhancing leaf vein architecture. However, evaluation of network architecture across different species is challenging due to difficulty in extracting multi-scale statistics from subsequent network graph representations and network segmentation from photographs. To address this issue, Li *et al.* (2022) developed CoRE-Net, a vein segmentation method that can hot-start with only a few labeled samples and accurately processes vein intersections, breakpoints, and blurred boundaries, with a comparative test accuracy of 94.88%. Xu *et al.* (2021) applied a CNN-based module to automatically segment leaf vein networks and quantify their width, angle, and connectivity; this method has potential applications in exploring the function of leaf vein structures across different plant species, with a precision-recall harmonic mean of $94.5 \pm 6\%$. Blonder *et al.* (2020) optimized multi-scale quantitative network architecture based on the vein networks of south-east Asian tree species to provide tools for measuring the role of vein networks in developing tissues. DL techniques hold promise for accurate analysis of intricate leaf vein networks and identification of crucial features, but further research is needed to improve their robustness and extend their application to diverse plant species.

Image segmentation plays a crucial role in the analysis of leaf vein phenotypes as it enables the extraction of a wealth of phenotype information, including, among others, color, shape, and texture. These features can be used to quantitatively analyze differences between various vein phenotypes (Ambarwari *et al.*, 2020). Additionally, dynamic information about leaf vein growth can be captured at different intervals using image segmentation technology. This further helps to reveal the ecological and genetic characteristics of plants and provides support for plant-breeding and ecological conservation efforts. However, vein image segmentation suffers from some notable shortcomings, such as susceptibility to noise and interference factors, leading to dislocations or incomplete segmentations (Xu *et al.*, 2021). Efficient algorithms and a significant amount of computational power are required to reduce processing time. Manual input from experts is also necessary to achieve an accurate segmentation (Du *et al.*, 2022). Although segmentation methods for vein networks have made significant progress, few of the algorithms are truly robust, making accurate segmentation of vein networks across different species with high precision a challenging task. Despite these challenges, vein image segmentation remains an indispensable technique in the field of vein phenotype analysis.

Table 1. Typical methods of analysis of several main vein phenotypes

Number of species studied	Data-set size	Is the dataset public?	Image pre-processing	Phenotype recognition methods	Research results	Evaluation of identification methods	Reference
5	Not stated	No	Chemically cleared imaging	image disposing tool	4 traits of veins	Highly correlated with the results of image analysis tools ($R^2=0.9996$).	(Bühler <i>et al.</i> , 2015)
53	271	No	Operator denoising imaging	Boruta and SVM	19 traits of veins	Average accuracy of 77.57%	(Ambarwari <i>et al.</i> , 2018)
43	1290	No	Operator denoising imaging	CNN-based method	62 traits of veins	Comparable testing accuracy of 94.88%	(Tan <i>et al.</i> , 2020)
200	1200	No	Concentrated light imaging	Directional morphological method	Leaf vein classification	Average deviation of <5 pixels; 54.28% average completeness	(Gan <i>et al.</i> , 2019)
6	300	Yes	Operator denoising imaging	SVM with RBF kernel	19 features traits of veins	Average precision of 84% and recall of 83%	(Ambarwari <i>et al.</i> , 2020)
295	727	Yes	Chemically cleared imaging	U-Net	Methods of vein measurement	Precision-recall harmonic mean of $94.5 \pm 6\%$	(Xu <i>et al.</i> , 2021)
15	60	Yes	Operator denoising imaging	Canny operator and VGG16 network	Length and curvature values of veins	Accuracy of 95%	(Pushpa and Lakshmi, 2022)
400	800	No	Images entered directly	U-Net	3 traits of veins	Mid-rib testing mean Intersection over Union (IoU) of 90.18%; other veins testing mean IoU of 70.09%	(Du <i>et al.</i> , 2022)
36s	4977	Yes	Images entered directly	CoRE-Net	Leaf vein segmentation	Mean IoU of 71.02% and mean Dice of 79.76%	(Li <i>et al.</i> , 2022)

Challenges and developments in vein phenomics

Construction of high-throughput acquisition platforms

The significance of leaf veins extends across several research domains, including but not limited to systematics, evolutionary biology, plant physiology, and investigations of climate change and paleoecology. Despite the increasing demand for data on vein traits, research is limited due to the lagging development of methods to rapidly generate large amounts of such data (Schneider *et al.*, 2018). The advances and expansion of biological sciences has been supported by the advancement of imaging technologies that have made it possible for researchers to examine the fundamental shifts that living systems undergo through time and space. The advantages and disadvantages of different imaging platforms and the corresponding leaf vein phenotypes are shown in Supplementary Table S1. High-throughput, field-based phenotypic analysis is anticipated to become achievable through the integration of microscopes and RGB cameras in areas such as ecology, physiology, and crop breeding, as their availability and performance improve (Liu *et al.*, 2021). However, the platform analysis of vein phenotype still requires laborious pre-processing (Zhao *et al.*, 2019). Although the primary purpose is to use two-dimensional images of veins, three-dimensional imaging is essential for understanding the mechanisms of light absorption, water transport, and biomechanics. In terms of their ability to obtain three-dimensional images of veins, many techniques, including confocal microscopy, nuclear magnetic resonance, computed tomography (CT), and laser ablation tomography (LAT), each have certain advantages and disadvantages. Confocal microscopy works well for three-dimensional imaging, although it has poor throughput and is limited by the opacity of the tissue being examined. LAT is a ‘label-free’ technique that can be used without dyes and other labeling methods, which provides a greater possibility of identifying tissue components. The development of LAT has filled in the gaps in the flux and three-dimensional scale of sample data that occurred with the current internal structure imaging techniques (Hall and Lanba, 2019). Despite its speed and its ability to provide three-dimensional information without being constrained by the opacity of tissues, one disadvantage of LAT is that it is damaging to samples (Levin *et al.*, 2021). Despite the non-destructive nature of nuclear magnetic resonance and CT (Brodersen and Roddy, 2016), the low throughput of these techniques limits their application to veins. The extremely lengthy (usually multi-hour) scan times per sample remain a downside of extremely high-resolution CT. Hyperspectral imaging enables the acquisition of vein-centric morphological measurements, including but not limited to VD, VL, vein diameter, and other related characteristics. At the same time, the reflectance data captured by hyperspectral imaging can be used for the purpose of differentiating between distinct areas of the leaf, thereby

facilitating selective vein information retrieval across diverse regions. These technological advances, taken together, have made it possible to quantify a wide variety of characteristics in vein traits (Grimm *et al.*, 2017; Liu *et al.*, 2018). Moreover, they enable remarkable measures to be made in stable samples, and spatial phenomics is being applied to simultaneously measure dozens of phenotypic species in plant organs and tissues while retaining spatial information (Moen *et al.*, 2019).

Selecting the appropriate imaging platform is crucial for accurate analysis of plant phenotype. The chosen platform must consider factors such as vein characteristics, plant volume, leaf size, and test site capacity. Portable instruments that integrate mobile phones have made optical and other sensors more applicable, improved connectivity and portability, and facilitated the collection of relevant phenotypic data both in laboratories and in the field (Yang *et al.*, 2020). With the development of imaging equipment, leaf vein research could obtain phenotypic information with higher resolution, multiple angles, and in three dimensions more quickly and with less effort, thereby enriching our knowledge on the mechanisms of growth and development of leaf veins. This advanced technology would be widely applicable in forestry, agriculture, and environmental protection, as it would allow a more comprehensive understanding of leaf vein morphology.

Development of vein phenomics segmentation

To ensure the reliability and robustness of DL estimations, it is imperative to gather enough data for ML model training (Yan and Wang, 2023). However, the process of labeling vein images is time-consuming and laborious, making it arduous to obtain enough markers for traits as intricate and nuanced as vein networks (van Dijk *et al.*, 2021). It is also necessary to extract omics data of leaf veins at different growth stages. Complicated data extraction is one of the difficulties that hinder the development of vein phenotype omics.

As illustrated in [Supplementary Fig. S2](#), leaf vein image phenotype technology can be broadly classified into three categories: digital image processing methods, ML-based image processing methods, and integrated image processing methods. Image semantic segmentation methods based on traditional ML techniques primarily include threshold-based methods (Zhao *et al.*, 2021), edge-based algorithms (Fang *et al.*, 2021), mathematical morphology-based methods (Heijmans, 2020), and region-based segmentation algorithms (Fang *et al.*, 2021). The methods exhibit computational efficiency; nevertheless, they suffer from inadequate automated feature extraction capacities, suboptimal segmentation accuracy, and limited generalization proficiency. Adequate annotated data are necessary for DL to effectively learn features. However, manual annotation of data poses challenges such as the requirement for professional domain knowledge and experience, the time- and labor-intensive processes involved, and difficulties in labeling certain data types such as hyperspectral and thermal imaging. The primary

algorithm models used in the fully supervised segmentation algorithm based on DL include FCN (Lu *et al.*, 2019), U-Net (Siddique *et al.*, 2021), SegNet (Deng *et al.*, 2022), PSPNet (Zhou *et al.*, 2019), and DeepLab (C. Liu *et al.*, 2019). Owing to the high cost of image annotation, several semi-supervised learning approaches have been developed for image segmentation, such as semi-supervised clustering, dimensionality reduction, regression, and classification (X. Yang *et al.*, 2022). Semi-supervised segmentation methods, however, still have issues regarding learning performance degradation and lack of robustness.

To overcome these challenges, researchers suggest including images of different scenarios in datasets, using crowdsourcing or artificial intelligence pre-labeling to expedite data annotation and reduce costs. Amazon's Mechanical Turk and Google's Fluid Annotation are examples of such technologies. This would enhance model reliability and reduce dependence on human labor, thereby saving time and costs (Andriluka *et al.*, 2018). In recent years, some researchers have proposed methods for combining Transformers (Vaswani *et al.*, 2017) and CNNs to achieve the fusion of global and local features, such as TransUNet (Chen *et al.*, 2021, Preprint) and DS-TransUNet (Lin *et al.*, 2022); these methods are based on the strong global context information representation ability of Transformers and the advantages of CNNs in local feature extraction. Another effective way to establish global feature association is to construct graph convolution networks (Dong *et al.*, 2017). Graph convolution networks turn image classification and segmentation into depth metric learning of neighboring points, and automatically learn node features and relationship information to improve image classification and segmentation. Generative adversarial networks can automatically learn the data distribution of the original actual sample set to supplement the training set with data that resembles genuine data (de Melo *et al.*, 2021). These methods are also suitable for the extraction of vein phenotypes.

Plant growth is a complex process influenced by metabolic reactions, physiological factors, and environmental conditions that lead to dynamic changes in plant organ growth. To track these changes, DL techniques such as CNNs and long short-term memory have been used to extract features and capture temporal relationships, respectively (Lee *et al.*, 2017; Sakurai *et al.*, 2019). Nevertheless, these methodologies encounter difficulties in encoding plant growth patterns when the time intervals are extensive or when the complexity of the plant and background is elevated. Further research is needed to develop more effective methods for plant growth analysis.

The extraction of leaf vein phenotype, while complex, is a crucial step in the development of a comprehensive vein analysis model once an appropriate imaging platform has been selected. Future research should focus on developing image-classification techniques and automation of vein data exploration. Automating and enhancing the efficiency of non-invasive plant phenotyping could greatly enhance vein-recognition practices by expediting the assessment process through repeated measurements

and reducing the cost of evaluating changes in traits (Nguyen *et al.*, 2022). Given the present level of image recognition, a combination of conventional segmentation techniques and ML methods is deemed to be a superior option for identifying leaf vein phenotypes. In addition, the development of leaf phenomics lags behind that of root phenomics, as leaf vein phenotypes are easier to observe and collect data on than the phenotypes of roots grown underground. Thus, methods for analyzing leaf vein phenotypes could be adapted for root phenotype analysis (Teramoto and Uga, 2020). The vascular traits are analogous to vein phenotypes. To address the issues of limited vascular pixels in human medical images such as retinal fundus vascular images, and the micro-vessels in segmentation being prone to breakage and interference from lesions and equipment noise, researchers have conducted numerous studies using U-Net, dense residual network, and general adversarial network models (Chen *et al.*, 2020; Guo *et al.*, 2020). The extraction of leaf vein phenotype data could also be based on the methods for extracting vascular phenotype data from medical images.

Use of machine learning techniques for analyzing multi-omics data

Multi-omics association analysis pipeline

Machine learning (ML) is poised to become increasingly vital in the realm of multi-omics analysis. With the proliferation of

high-throughput technologies such as genomics, transcriptomics, and proteomics, vast quantities of data are being generated and amassed. Yet, challenges persist in processing and analyzing these data, including unwieldy sample sizes, complex sample structures, and the emergence of multiple data types from the same analysis objects. In this context, ML can provide a means of solving complex data processing and analysis tasks by automatically identifying and extracting underlying features (Hu *et al.*, 2019).

At present, no multi-omics analysis method based on ML has been found for leaf vein phenotype. However, multi-omics association analysis is the most promising method for understanding plant growth mechanisms. In order to gain insight into plant phenotypes and associated metabolic processes and resource allocation patterns, the collection of plant ontology spatiotemporal omics data is becoming increasingly important (Munné-Bosch, 2022). To evaluate the change characteristics of the target phenotype throughout the growth phase, a large amount of time-related data is required (Montesinos-López *et al.*, 2018). In addition, a single omics dataset is not sufficient to analyze the regulatory mechanisms of complex physiological processes systematically and comprehensively (Azodi *et al.*, 2020). As shown in Fig. 1, multi-omics joint analysis can establish data relationships among various molecular levels, such as the genome, transcriptome, proteome, metabolic group, and phenotype group, by normalizing, comparative analysis, and

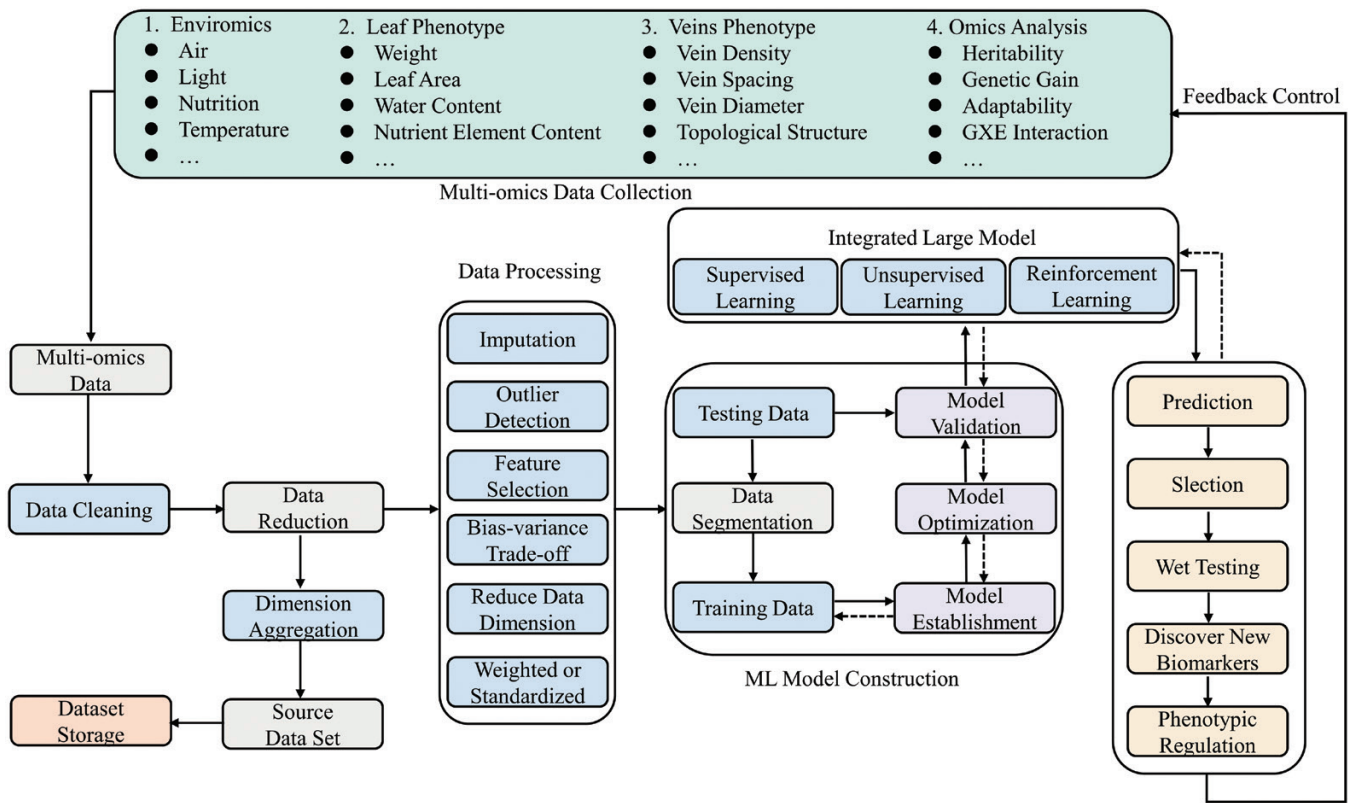


Fig. 1. Overview of multi-omics analysis using machine learning.

correlation analysis of batches of data (Omidbakhshfard *et al.*, 2021). This approach has been used to explore the potential regulatory network mechanism of networks, providing more evidence for the mechanisms underlying plant action. Finally, the productivity of plants can be improved by transgenic approaches or by regulating environmental group parameters.

The conundrum of heterogeneous multiple datasets from different high-throughput sources is a common issue, with transcriptomics and proteomics using distinct standardization and scaling procedures and metabolomics producing sparse data (Bersanelli *et al.*, 2016). Subsequently, the interpolation and outlier detection of each dataset should be independently assessed before they are integrated. The high-dimensional space generated by such multi-omics data often contains related features, potentially leading to erroneous algorithm training. Dimensionality reduction methods, such as feature extraction and feature selection, could be used to decrease the number of dimensions. Additionally, to manage the bias-variance trade-offs, regularization, hyperparameter adjustment, and cross-validation could be used to tackle class imbalances and overfitting problems (Zebari *et al.*, 2020). The lack of transparency and interpretability is the main obstacle to the ML method, which highlights the importance of solving this problem for multi-omics pathway analysis (Gunning *et al.*, 2019).

The different characteristics of the model should be combined into various structures according to the data attributes,

which helps the neural network to improve the precision of the model prediction. The model mainly includes the following characteristics: (i) it automatically analyzes the extent to which the phenotype is affected by environmental or gene editing, and the related effects of these two factors; (ii) it can process large amounts of data efficiently; (iii) the training model contains many hidden layers and can capture highly complex associations through multi-modal inputs (e.g., phenotypic images, temperature, humidity, wind speed and other environmental parameters) (Heslot *et al.*, 2012); (iv) allowing shared parameters and making use of pooling procedures are two ways to cut down on the number of parameters. As shown in Fig. 2, ML analysis of multi-omics data has the following steps (Reel *et al.*, 2021). The first step is data preparation. Raw data and corresponding phenotypic information including individual omics (e.g. environmental group, genomics, transcriptomics, metabolomics and phenomics) are directly input. The second step is to construct an omics model. There are two ways to do this: (i) develop a separate model for each omics; (ii) create distinct transformations, presented as graphical or nuclear relationships, for each omics. Each omics dataset is converted into a graph or matrix, and then all datasets are merged into a single dataset, and then the model is constructed. The third step is the integrated omics model. The multi-omics data from the first step are directly combined into a large data matrix. Another way is to aggregate the model, graph, or kernel

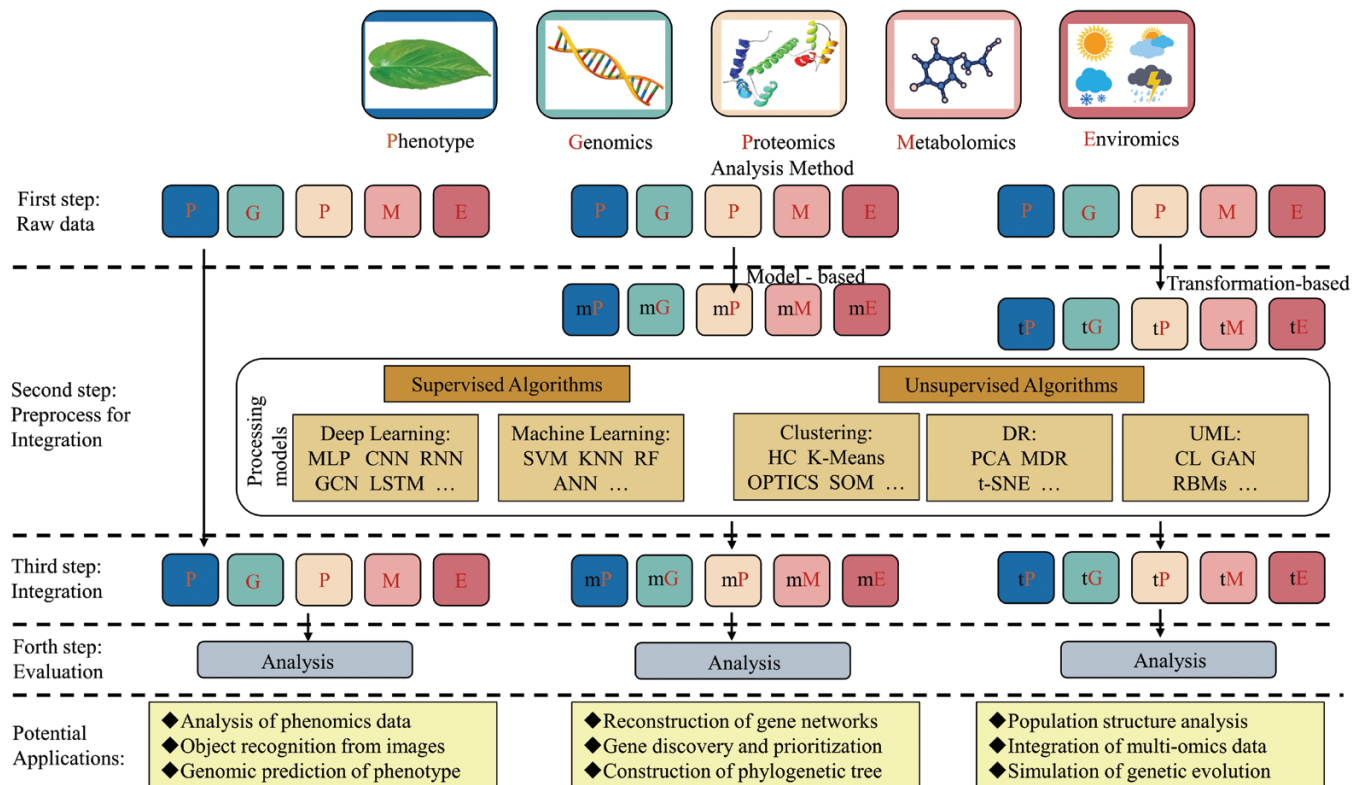


Fig. 2. System pipeline for machine learning methods for multi-omics analysis.

matrix from the second stage into a joint model or matrix. The fourth step is to conduct supervised or unsupervised analysis.

Through multi-omics joint analysis, ML algorithms can effectively integrate data from diverse omics domains while uncovering intricate interactions and feature associations between distinct sets of omics data (Costa-Neto *et al.*, 2021). Such comprehensive analyses offer crucial insights into the underlying mechanisms of biological systems, supporting research into areas such as, among others, cell signal transduction, gene localization, pathway analysis, and crop growth environment regulation (Boutsika *et al.*, 2022). However, it is important to note that there is still significant room for improving ML models in terms of data pre-processing, feature selection/extraction, dimensionality reduction, model construction, and training. The plant science and medical domains share fundamental biological principles, including biomolecular interactions and metabolic regulation. Thus, akin to their use in medical research, ML methods including DL, genetic algorithm, support vector machine (SVM), and so on, can be deployed in plant science. Leveraging ML approaches from medicine is expected to expedite, optimize, standardize, and deepen research efforts in the plant field, leading to novel ideas and techniques for addressing significant challenges in agriculture and biomedical research.

Supervised multi-omics analysis methods

The application of multiple omics datasets, such as genomics, transcriptomics, metabolomics, phenomics and enviromics, has become increasingly popular as a means to understand complex biological processes. In agriculture, the integration of these datasets is particularly useful for predicting phenotypic traits such as yield, plant height, and disease resistance, which are influenced by multiple genes and environmental factors. As is shown in [Supplementary Table S2](#), supervised learning has been widely used in phenotypic prediction by providing multi-omics data to classical ML algorithms such as ANNs, SVMs, and K-nearest neighbors (Lin and Lane, 2017; Ding *et al.*, 2018). However, these methods are limited in handling large data volumes and achieving high prediction accuracy. To address these challenges, DL techniques have been applied to model-based supervised learning (Poirion *et al.*, 2021) such as auto-encoding and hierarchical integration using stacked auto-encoders (Xu *et al.*, 2019; Lee *et al.*, 2020). These methodologies have demonstrated potential in amalgamating diverse omics data for prognostication of survival and acquiring sophisticated representations.

A recent study by K. Wang *et al.* (2023) introduced a novel approach called DNNGP, which uses extensive and heterogeneous genomic information to forecast agricultural phenotypic traits with a focus on exploring hidden non-additive effects and improving prediction accuracy. In a related study, W. Wang *et al.* (2023) used ML techniques to predict plant height in maize varieties using image-based traits. These resources and prediction models provide valuable tools for

molecular design breeding and predicting ideal plant structures for maize varieties. Other strategies for omics analysis, such as integrating phenotypic and omics traits and employing multi-variate DL techniques (W. Yang *et al.*, 2022), have also shown promise in improving predictions of plant performance and yield. However, the importance of larger datasets for achieving better predictions has been emphasized (Montesinos-López *et al.*, 2019). This suggests that increasing dataset sizes may be a critical factor in improving the accuracy of predicting plant performance and yield.

The integration of multiple omics datasets and ML techniques provides powerful tools for predicting phenotypic traits in agriculture. These approaches enable the identification of key genes and pathways that regulate traits of interest and facilitate the creation of new and improved plant varieties with desirable traits. Supervised learning is a powerful tool for analyzing metabolic pathways and predicting phenotypic outcomes by leveraging prior experience to inform output predictions. However, first, it should be noted that the performance of the established model depends on the size and quality of the dataset. Taking the phenotypic group as an example, data acquisition is difficult and small. Further research is needed to overcome the limitations of small datasets. Second, a major limitation of supervised learning is the high human resources and time costs of labeling associated with model training, especially for complex tasks. Therefore, it is essential to optimize the labeling strategy and use semi-supervised or unsupervised learning approaches where feasible to minimize costs while maintaining high accuracy. Finally, if the test dataset differs significantly from the training dataset, the supervised learning model might not accurately predict the output, highlighting the importance of model validation and generalization to diverse datasets.

Unsupervised multi-omics analysis methods

Multi-omics association analysis has emerged as a crucial area of investigation in recent years, with unsupervised methods playing an increasingly essential role in unraveling complex relationships between molecular features. Multi-omics analysis using ML in agriculture is a relatively unexplored research area. As such, this section primarily draws upon advanced omics analysis techniques employed in the field of human medicine to propose novel approaches for multi-omics analysis in agriculture. As shown in [Supplementary Table S3](#), among these methods, matrix factorization-based techniques, such as joint non-negative matrix factorization, have shown promise for integrating diverse omics data types (Zhang *et al.*, 2012). However, the high computational requirements and memory demands of this technique can limit its practicality. To address these challenges, subsequent methods, including iCluster+ and MoCluster, have been developed to enable the integration of a wider range of omics data types (Mo *et al.*, 2013; Meng *et al.*, 2016). Multi-omics factor analysis has also been introduced to

disentangle shared heterogeneity across multiple omics datasets (Argelaguet *et al.*, 2018). Additionally, iCluster Bayes offers a fully Bayesian latent variable model that can enhance statistical inference and computational speed (Mo *et al.*, 2018). These techniques represent valuable tools for uncovering patterns and sources of variability in multi-omics data, and are expected to continue driving progress in the field.

Some methods of analysis combining model and clustering have also been developed. This approach creates greater flexibility and enables omics data from different clusters to be integrated later. Like similarity network fusion, this method first constructs a pipeline network of single omics data and then integrates the data into a joint network for unified analysis (Wang *et al.*, 2014). Some methods of analysis combining model and clustering have also been developed. This approach creates greater flexibility and enables omics data from different clusters to be integrated later. PINSPlus is an innovative unsupervised phenotypic discovery technique that holds immense promise for the integration of diverse omics datasets into a cohesive and holistic analysis. This technique is distinguished by its exceptional robustness in the face of noise and its ability to make unstable quantitative determinations (Nguyen *et al.*, 2019). NEMO, the underlying algorithmic framework, can analyze certain datasets without the need for data interpolation, thereby reducing computational complexity and enhancing analytical efficiency. The method entails the initial construction of a similarity matrix for each omics, which is subsequently integrated into a composite matrix. Finally, the network is clustered together, enabling the discovery of new phenotypic associations and biomarkers (Rappoport and Shamir, 2019). Another technique, PAMOGK, combines multi-omics data with existing metabolic pathway knowledge. As a graphical kernel, it evaluates the similarity of samples based on a single molecular change type in the context of pathways (Tepeli *et al.*, 2021). Unsupervised multi-omics association analysis provides an effective method to detect complex interactions between genetic and environmental factors involved in various phenotypic changes.

Compared with supervised learning, unsupervised learning has the benefit of dispensing with manual data labeling. It can automatically learn data relationships, extract patterns and structures, and facilitate data comprehension. Furthermore, it can recognize latent or novel patterns and anomalous data points, thereby supporting data mining and exploratory analysis. Nonetheless, unsupervised learning does not incorporate prior knowledge to guide model training, rendering it vulnerable to noise and invalid information, which diminishes model accuracy and reliability. Training outcomes might also be subjective and less elucidating, with challenges in model optimization or enhancement. The development of unsupervised learning techniques based on DL has experienced an up-tick and these techniques are extensively used in pre-training, transfer learning, and other methodologies. In summary, unsupervised learning provides unique benefits in data processing and analysis. The direction of its development is mainly focusing on improving the accuracy and interpretability of the

model, expanding the field of application, and combining other learning methods.

Conclusions

This review aimed to analyze the structure and function of veins from a multi-omics perspective and with the help of image recognition technology. It also investigated the functional, environmental, and genetic factors associated with vein networks, as well as the current status of research based on image analysis. Furthermore, the methods of venous phenotype extraction and multi-omics association analysis using ML technology have been discussed. These findings could provide a theoretical basis for improving crop productivity by optimizing the vein network architecture.

Supplementary data

The following supplementary data are available at [JXB online](#).

Fig. S1. Schematic diagram of leaf vein phenotype and related functions.

Fig. S2. Leaf vein image phenotypic acquisition method and typical examples.

Table S1. Advantages and disadvantages of different phenotypic acquisition methods and the characteristics suitable for acquisition.

Table S2. Representative cases of multi-omics studies using different supervised ML methods.

Table S3. Representative cases of multi-omics studies using different unsupervised ML methods.

Acknowledgements

The authors thank the two anonymous reviewers for their insightful suggestions and constructive comments on a previous version of the manuscript.

Author contributions

YZ conceived and wrote the manuscript; NZ, XC, and TS contributed to funding acquisition, reviewing and editing, validation, and supervision.

Conflict of interest

The authors declare no conflict of interest.

Funding

This research was funded by the National key research and development program (2022YFD2002205) and the Agricultural Science and Technology Innovation Program of Chinese Academy of Agricultural

Sciences (CAAS-ASTIP-2023-AII, CAAS-ASTIP-2016-AII, CAAS-ZDRW202107).

References

- Akhter D, Qin R, Nath UK, Alamin M, Jin X, Shi C.** 2018. The brown midrib leaf (*bml*) mutation in rice (*Oryza sativa* L.) causes premature leaf senescence and the induction of defense responses. *Genes* **9**, 203.
- Ambarwari A, Adrian QJ, Herdiyeni Y, Hermadi I.** 2020. Plant species identification based on leaf venation features using SVM. *TELKOMNIKA (Telecommunication Computing Electronics and Control)* **18**, 726–732.
- Ambarwari A, Herdiyeni Y, Hermadi I.** 2018. Biometric analysis of leaf venation density based on digital image. *TELKOMNIKA (Telecommunication Computing Electronics and Control)* **16**, 1735–1744.
- Andriluka M, Uijlings JR, Ferrari V.** 2018. Fluid annotation: a human-machine collaboration interface for full image annotation. In: *MM '18: Proceedings of the 26th ACM international conference on Multimedia*, Seoul, Korea, 22–26 October 2018. Proceedings. New York: Association for Computing Machinery, 1957–1966.
- Argelaguet R, Velten B, Arnol D, Dietrich S, Zenz T, Marioni JC, Buettner F, Huber W, Stegle O.** 2018. Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. *Molecular Systems Biology* **14**, e8124.
- Aritsara ANA, Ni M-Y, Ramanantoandro T, Zhu S-D, Li Y-Z, Gleason S, Lawren S, Cao K-F.** 2022. Leaf vein topology confers water transport efficiency. *Research Square* doi: [10.21203/rs.3.rs-2344990/v1](https://doi.org/10.21203/rs.3.rs-2344990/v1). [Preprint]
- Azodi CB, Pardo J, VanBuren R, de los Campos G, Shiu S-H.** 2020. Transcriptome-based prediction of complex traits in maize. *The Plant Cell* **32**, 139–151.
- Balasundaram A, Dilip G, Ashokkumar S, Manickam M, Gurunathan K, Kothandaraman D.** 2023. Detecting true medicinal leaves among similar leaves using computer vision and CNN. *National Academy Science Letters* **46**, 257–262.
- Beerling DJ, Franks PJ.** 2010. The hidden cost of transpiration. *Nature* **464**, 495–496.
- Bersanelli M, Mosca E, Remondini D, Giampieri E, Sala C, Castellani G, Milanese L.** 2016. Methods for the integration of multi-omics data: mathematical aspects. *BMC Bioinformatics* **17**, 167–177.
- Blonder B, Both S, Jodra M, Xu H, Fricker M, Matos IS, Majalap N, Burslem DF, Teh YA, Malhi Y.** 2020. Linking functional traits to multiscale statistics of leaf venation networks. *New Phytologist* **228**, 1796–1810.
- Blonder B, Violle C, Bentley LP, Enquist BJ.** 2011. Venation networks and the origin of the leaf economics spectrum. *Ecology Letters* **14**, 91–100.
- Boutsika A, Tanou G, Xanthopoulou A, Samiotaki M, Nianiou-Obeidat I, Ganopoulos I, Mellidou I.** 2022. Insights and advances in integrating multi-omic approaches for potato crop improvement. *Scientia Horticulturae* **305**, 111387.
- Brodersen CR, Roddy AB.** 2016. New frontiers in the three-dimensional visualization of plant structure and function. *American Journal of Botany* **103**, 184–188.
- Brodrribb TJ, Bienaimé D, Marmottant P.** 2016. Revealing catastrophic failure of leaf networks under stress. *Proceedings of the National Academy of Sciences, USA* **113**, 4865–4869.
- Brodrribb TJ, Feild TS, Sack L.** 2010. Viewing leaf structure and evolution from a hydraulic perspective. *Functional Plant Biology* **37**, 488–498.
- Brodrribb TJ, Jordan GJ.** 2011. Water supply and demand remain balanced during leaf acclimation of *Nothofagus cunninghamii* trees. *New Phytologist* **192**, 437–448.
- Bühler J, Rishmawi L, Pflugfelder D, Huber G, Scharr H, Hülkamp M, Koornneef M, Schurr U, Jahnke S.** 2015. phenoVein—a tool for leaf vein segmentation and analysis. *Plant Physiology* **169**, 2359–2370.
- Carins Murphy MR, Jordan GJ, Brodrribb TJ.** 2012. Differential leaf expansion can enable hydraulic acclimation to sun and shade. *Plant, Cell & Environment* **35**, 1407–1418.
- Carvalho MR, Turgeon R, Owens T, Niklas KJ.** 2017. The hydraulic architecture of *Ginkgo* leaves. *American Journal of Botany* **104**, 1285–1298.
- Chen D, Ao Y, Liu S.** 2020. Semi-supervised learning method of U-net deep learning network for blood vessel segmentation in retinal images. *Symmetry* **12**, 1067.
- Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, Lu L, Yuille AL, Zhou Y.** 2021. TransUNet: Transformers make strong encoders for medical image segmentation. *arXiv* doi: [10.48550/arXiv.2102.04306](https://arxiv.org/abs/10.48550/arXiv.2102.04306). [Preprint].
- Cochard H, Venisse J-S, Barigah TS, Brunel N, Herbette S, Guillot A, Tyree MT, Sakr S.** 2007. Putative role of aquaporins in variable hydraulic conductance of leaves in response to light. *Plant Physiology* **143**, 122–133.
- Costa-Neto G, Fritsche-Neto R, Crossa J.** 2021. Nonlinear kernels, dominance, and envirotyping data increase the accuracy of genome-based prediction in multi-environment trials. *Heredity* **126**, 92–106.
- de Melo CM, Torralba A, Guibas L, DiCarlo J, Chellappa R, Hodgins J.** 2021. Next-generation deep learning based on simulators and synthetic data. *Trends in Cognitive Sciences* **26**, 174–187.
- Deng T, Fu B, Liu M, He H, Fan D, Li L, Huang L, Gao E.** 2022. Comparison of multi-class and fusion of multiple single-class SegNet model for mapping karst wetland vegetation using UAV images. *Scientific Reports* **12**, 13270.
- Dhawal S, Reiter JW, Laroche A, Schultz EA.** 2021. Leaf vein pattern response to heat and drought requires genes that influence PINFORMED1 localization and is mimicked by ABA treatment. *Environmental and Experimental Botany* **185**, 104426.
- Dhondt S, Van Haerenborgh D, Van Cauwenbergh C, Merks RMH, Philips W, Beemster GTS, Inzé D.** 2012. Quantitative analysis of venation patterns of *Arabidopsis* leaves by supervised image analysis. *The Plant Journal* **69**, 553–563.
- Ding MQ, Chen L, Cooper GF, Young JD, Lu X.** 2018. Precision oncology beyond targeted therapy: combining omics data with machine learning matches the majority of cancer cells to effective therapeutics assigning cancers to effective drugs with big data. *Molecular Cancer Research* **16**, 269–278.
- Dong H, Li T, Leng J, Kong L, Bai G.** 2017. GCN: GPU-based cube CNN framework for hyperspectral image classification. In: *46th International Conference on Parallel Processing (ICPP)*, Bristol, UK, 14–17 August 2017. Proceedings. Piscataway: IEEE, 41–49.
- Du J, Li B, Lu X, Yang X, Guo X, Zhao C.** 2022. Quantitative phenotyping and evaluation for lettuce leaves of multiple semantic components. *Plant Methods* **18**, 1–18.
- Duarte MA, Woo S, Hultine K, Blonder B, Aparecido LMT.** 2023. Vein network redundancy and mechanical resistance mitigate gas exchange losses under simulated herbivory in desert plants. *AoB Plants* **15**, plad002.
- Ellis B, Daly DC, Hickey LJ, Johnson KR, Mitchell JD, Wilf P, Wing SL.** 2009. *Manual of leaf architecture*. New York: Cornell University Press.
- Fang J, Liu H, Zhang L, Liu J, Liu H.** 2021. Region-edge-based active contours driven by hybrid and local fuzzy region-based energy for image segmentation. *Information Sciences* **546**, 397–419.
- Fiorin L, Brodrribb TJ, Anfodillo T.** 2016. Transport efficiency through uniformity: organization of veins and stomata in angiosperm leaves. *New Phytologist* **209**, 216–227.
- Franklin KA.** 2016. Photomorphogenesis: plants feel blue in the shade. *Current Biology* **26**, R1275–R1276.
- Gan Y, Rong Y, Huang F, Hu L, Yu X, Duan P, Xiong S, Liu H, Peng J, Yuan X.** 2019. Automatic hierarchy classification in venation networks using directional morphological filtering for hierarchical structure traits extraction. *Computational Biology and Chemistry* **80**, 187–194.
- Gao J, Chen J, Feng L, Wang Q, Li S, Tan X, Yang F, Yang W.** 2022. HD-Zip III gene family: identification and expression profiles during leaf vein development in soybean. *Plants* **11**, 1728.
- Gardiner J, Sherr I, Scarpella E.** 2010. Expression of *DOF* genes identifies early stages of vascular development in *Arabidopsis* leaves. *International Journal of Developmental Biology* **54**, 1389–1396.

- Grimm JB, Muthusamy AK, Liang Y, *et al.* 2017. A general method to fine-tune fluorophores for live-cell and *in vivo* imaging. *Nature Methods* **14**, 987–994.
- Grinblat GL, Uzal LC, Larese MG, Granitto PM. 2016. Deep learning for plant identification using vein morphological patterns. *Computers and Electronics in Agriculture* **127**, 418–424.
- Grunwald Y, Gosa SC, Torne-Srivastava T, Moran N, Moshelion M. 2022. Out of the blue: phototropins of the leaf vascular bundle sheath mediate the regulation of leaf hydraulic conductance by blue light. *The Plant Cell* **34**, 2328–2342.
- Gunning D, Stefik M, Choi J, Miller T, Stumpf S, Yang G-Z. 2019. XAI—explainable artificial intelligence. *Science Robotics* **4**, eaay7120.
- Guo C, Szemenyei M, Yi Y, Xue Y, Zhou W, Li Y. 2020. Dense residual network for retinal vessel segmentation. In: 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020. Proceedings. Piscataway: IEEE, 1374–1378.
- Guo JY, Liu BM, Yang HJ, Qin CQ, Ren Y, Jiang HR, Tao LZ, Ye YF, Wu YJ. 2022. Genetic analysis and gene mapping of the yellow midrib leaf mutant (*ym1*) in rice (*Oryza sativa* L.). *Acta Agronomica Sinica* **48**, 10.
- Hall B, Lanba A. 2019. Three-dimensional analysis of biological systems via a novel laser ablation technique. *Journal of Laser Applications* **31**, 022602.
- Heijmans HJAM. 2020. Morphological image operators. *Advances in Imaging and Electron Physics*, vol **216**. Academic Press.
- Heslot N, Yang HP, Sorrells ME, Jannink JL. 2012. Genomic selection in plant breeding: a comparison of models. *Crop Science* **52**, 146–160.
- Hu X, Xie W, Wu C, Xu S. 2019. A directed learning strategy integrating multiple omic data improves genomic prediction. *Plant Biotechnology Journal* **17**, 2011–2020.
- Hua L, He P, Goldstein G, Liu H, Yin D, Zhu S, Ye Q. 2020. Linking vein properties to leaf biomechanics across 58 woody species from a subtropical forest. *Plant Biology* **22**, 212–220.
- Ilegems M, Douet V, Meylan-Bettex M, Uyttewaald M, Brand L, Bowman JL, Stieger PA. 2010. Interplay of auxin, KANADI and Class III HD-ZIP transcription factors in vascular tissue formation. *Development* **137**, 975–984.
- Ji W, LaZerte SE, Waterway MJ, Lechowicz MJ. 2020. Functional ecology of congeneric variation in the leaf economics spectrum. *New Phytologist* **225**, 196–208.
- Katifori E, Magnasco MO. 2012. Quantifying loopy network architectures. *PLoS One* **7**, e37994.
- Kawai K, Okada N. 2020. Leaf vascular architecture in temperate dicotyledons: correlations and link to functional traits. *Planta* **251**, 17.
- Koiwai H, Nakaminami K, Seo M, Mitsuhashi W, Toyomasu T, Koshiba T. 2004. Tissue-specific localization of an abscisic acid biosynthetic enzyme, AAO3, in Arabidopsis. *Plant Physiology* **134**, 1697–1707.
- Kusumi K, Sakata C, Nakamura T, Kawasaki S, Yoshimura A, Iba K. 2011. A plastid protein NUS1 is essential for build-up of the genetic system for early chloroplast development under cold stress conditions. *The Plant Journal* **68**, 1039–1050.
- Lasser J, Katifori E. 2017. NET: a new framework for the vectorization and examination of network data. *Source Code for Biology and Medicine* **12**, 4.
- Lee SH, Chan CS, Mayo SJ, Remagnino P. 2017. How deep learning extracts and learns leaf features for plant classification. *Pattern Recognition* **71**, 1–13.
- Lee TY, Huang KY, Chuang CH, Lee CY, Chang TH. 2020. Incorporating deep learning and multi-omics autoencoding for analysis of lung adenocarcinoma prognostication. *Computational Biology and Chemistry* **87**, 107277.
- Levin KA, Tucker MR, Strock CF, Lynch JP, Mather DE. 2021. Three-dimensional imaging reveals that positions of cyst nematode feeding sites relative to xylem vessels differ between susceptible and resistant wheat. *Plant Cell Reports* **40**, 393–403.
- Li L, Hu W, Lu J, Zhang C. 2022. Leaf vein segmentation with self-supervision. *Computers and Electronics in Agriculture* **203**, 107352.
- Li S, Hamani AKM, Zhang Y, Liang Y, Gao Y, Duan A. 2021. Coordination of leaf hydraulic, anatomical, and economical traits in tomato seedlings acclimation to long-term drought. *BMC Plant Biology* **21**, 536.
- Lin A, Chen B, Xu J, Zhang Z, Lu G, Zhang D. 2022. DS-transUNet: dual swin transformer u-net for medical image segmentation. *IEEE Transactions on Instrumentation and Measurement* **71**, 1–15.
- Lin E, Lane H-Y. 2017. Machine learning and systems genomics approaches for multi-omics data. *Biomarker Research* **5**, 2.
- Liu C, Chen L-C, Schroff F, Adam H, Hua W, Yuille AL, Fei-Fei L. 2019. Auto-DeepLab: hierarchical neural architecture search for semantic image segmentation. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, USA, 15–20 June 2019. Proceedings. Piscataway: IEEE, 82–92.
- Liu H, Dong P, Ioannou MS, Li L, Shea J, Pasolli HA, Grimm JB, Rivlin PK, Lavis LD, Koyama M. 2018. Visualizing long-term single-molecule dynamics *in vivo* by stochastic protein labeling. *Proceedings of the National Academy of Sciences, USA* **115**, 343–348.
- Liu Y, Khan AR, Azhar W, Wong CE, Li Y, Huang Y, Cao X, Liu Z, Gan Y. 2022. Cys2/His2-Type zinc finger proteins regulate plant growth and development. *Critical Reviews in Plant Sciences* **41**, 351–363.
- Liu Y, Rollins AM, Levenson RM, Fereidouni F, Jenkins MW. 2021. Pocket MUSE: an affordable, versatile and high-performance fluorescence microscope using a smartphone. *Communications Biology* **4**, 334.
- Lu J, Pan C, Li X, *et al.* 2021. *OBV* (obscure vein), a C₂H₂ zinc finger transcription factor, positively regulates chloroplast development and bundle sheath extension formation in tomato (*Solanum lycopersicum*) leaf veins. *Horticulture Research* **230**, 8.
- Lu Y, Chen Y, Zhao D, Chen J. 2019. Graph-FCN for image semantic segmentation. In: Lu H, Tang H, Wang Z, eds. *Advances in neural networks – ISNN 2019*. Cham: Springer, 97–105.
- Manuela D, Xu M. 2020. Patterning a leaf by establishing polarities. *Frontiers in Plant Science* **11**, 568730.
- Ma J, Zhao Z-L, Lin S, Xie YM. 2021. Topology of leaf veins: experimental observation and computational morphogenesis. *Journal of the Mechanical Behavior of Biomedical Materials* **123**, 104788.
- Malinowski R. 2013. Understanding of leaf development—the science of complexity. *Plants* **2**, 396–415.
- McAdam SAM, Eléouët MP, Best M, *et al.* 2017. Linking auxin with photosynthetic rate via leaf venation. *Plant Physiology* **175**, 351–360.
- Meng C, Helm D, Frejno M, Kuster B. 2016. moCluster: identifying joint patterns across multiple omics data sets. *Journal of Proteome Research* **15**, 755–765.
- Messier J, McGill BJ, Enquist BJ, Lechowicz MJ. 2017. Trait variation and integration across scales: is the leaf economic spectrum present at local scales? *Ecography* **40**, 685–697.
- Mo Q, Shen R, Guo C, Vannucci M, Chan KS, Hilsenbeck SG. 2018. A fully Bayesian latent variable model for integrative clustering analysis of multi-type omics data. *Biostatistics* **19**, 71–86.
- Mo Q, Wang S, Seshan VE, Olshen AB, Schultz N, Sander C, Powers RS, Ladanyi M, Shen R. 2013. Pattern discovery and cancer gene identification in integrated cancer genomic data. *Proceedings of the National Academy of Sciences, USA* **110**, 4245–4250.
- Moen E, Bannon D, Kudo T, Graf W, Covert M, Van Valen D. 2019. Deep learning for cellular image analysis. *Nature Methods* **16**, 1233–1246.
- Moison M, Marmagne A, Dinant S, *et al.* 2018. Three cytosolic glutamine synthetase isoforms localized in different-order veins act together for N remobilization and seed filling in Arabidopsis. *Journal of Experimental Botany* **69**, 4379–4393.
- Montesinos-López OA, Montesinos-López A, Crossa J, Gianola D, Hernández-Suárez CM, Martín-Vallejo J. 2018. Multi-trait, multi-environment deep learning modeling for genomic-enabled prediction of plant traits. *G3: Genes, Genomes, Genetics* **8**, 3829–3840.
- Montesinos-López OA, Montesinos-López A, Tuberosa R, Maccaferri M, Sciara G, Ammar K, Crossa J. 2019. Multi-trait, multi-environment

genomic prediction of durum wheat with genomic best linear unbiased predictor and deep learning methods. *Frontiers in Plant Science* **10**, 1311.

Munné-Bosch S. 2022. Spatiotemporal limitations in plant biology research. *Trends in Plant Science* **27**, 346–354.

Nguyen H, Shrestha S, Draghici S, Nguyen T. 2019. PINSPPlus: a tool for tumor subtype discovery in integrated genomic data. *Bioinformatics* **35**, 2843–2846.

Nguyen VD, Sarić R, Burge T, Berkowitz O, Trtilek M, Whelan J, Lewsey MG, Čustović E. 2022. Noninvasive imaging technologies in plant phenotyping. *Trends in Plant Science* **27**, 316–317.

Omidbakhshfard MA, Sokolowska EM, Vittori VD, Perez de Souza L, Kuhalskaya A, Brotman Y, Alseekh S, Fernie AR, Skirycz A. 2021. Multi-omics analysis of early leaf development in *Arabidopsis thaliana*. *Patterns* **2**, 100235.

Pan L, George-Jaeggli B, Borrell A, Jordan D, Koller F, Al-Salman Y, Ghannoum O, Cano FJ. 2022. Coordination of stomata and vein patterns with leaf width underpins water-use efficiency in a C₄ crop. *Plant, Cell & Environment* **45**, 1612–1630.

Poirion OB, Jing Z, Chaudhary K, Huang S, Garmire LX. 2021. DeepProg: an ensemble of deep-learning and machine-learning models for prognosis prediction using multi-omics data. *Genome Medicine* **13**, 112.

Price CA, Symonova O, Mileyko Y, Hilley T, Weitz JS. 2011. Leaf extraction and analysis framework graphical user interface: segmenting and analyzing the structure of leaf veins and areoles. *Plant Physiology* **155**, 236–245.

Prigge MJ, Otsuga D, Alonso JM, Ecker JR, Drews GN, Clark SE. 2005. Class III homeodomain-leucine zipper gene family members have overlapping, antagonistic, and distinct roles in *Arabidopsis* development. *The Plant Cell* **17**, 61–76.

Pushpa BR, Lakshmi P. 2022. Deep learning model for plant species classification using leaf vein features. In: International Conference on Augmented Intelligence and Sustainable Systems (ICAIS), Trichy, India, 24–26 November 2022. Proceedings. Piscataway: IEEE, 238–243.

Rai N, Neugart S, Yan Y, et al. 2019. How do cryptochromes and UVR8 interact in natural and simulated sunlight? *Journal of Experimental Botany* **70**, 4975–4990.

Rappoport N, Shamir R. 2019. NEMO: cancer subtyping by integration of partial multi-omic data. *Bioinformatics* **35**, 3348–3356.

Reel PS, Reel S, Pearson E, Trucco E, Jefferson E. 2021. Using machine learning approaches for multi-omics data analysis: a review. *Biotechnology Advances* **49**, 107739.

Resende RT, Piepho H-P, Rosa GJM, Silva-Junior OB, e Silva FF, de Resende MDV, Grattapaglia D. 2021. Enviromics in breeding: applications and perspectives on envirotypic-assisted selection. *Theoretical and Applied Genetics* **134**, 95–112.

Rolland-Lagan A-G, Amin M, Pakulska M. 2009. Quantifying leaf venation patterns: two-dimensional maps. *The Plant Journal* **57**, 195–205.

Ronellenfitch H, Lasser J, Daly DC, Katifori E. 2015. Topological phenotypes constitute a new dimension in the phenotypic space of leaf venation networks. *PLoS Computational Biology* **11**, e1004680.

Ruonala R, Ko D, Helariutta Y. 2017. Genetic networks in plant vascular development. *Annual Review of Genetics* **51**, 335–359.

Sack L, Scoffoni C. 2013. Leaf venation: structure, function, development, evolution, ecology and applications in the past, present and future. *New Phytologist* **198**, 983–1000.

Sack L, Streeter CM, Holbrook NM. 2004. Hydraulic analysis of water flow through leaves of sugar maple and red oak. *Plant Physiology* **134**, 1824–1833.

Sakurai S, Uchiyama H, Shimada A, Taniguchi R. 2019. Plant growth prediction using convolutional LSTM. In: 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019), Prague, Czech Republic, 25–27 February 2019. Proceedings. Setúbal: SciTePress, 105–113.

Salve P, Sardesai M, Yannawar P. 2021. Combining multiple classifiers using hybrid votes technique with leaf vein angle, CNN and Gabor features

for plant recognition. In: Santosh KC, Gawali B, eds. Recent trends in image processing and pattern recognition. Third International Conference, RTIP2R 2020, Aurangabad, India, 3–4 January 2020. Proceedings. Singapore: Springer, 313–331.

Scarpella E, Marcos D, Friml J, Berleth T. 2006. Control of leaf vascular patterning by polar auxin transport. *Genes & Development* **20**, 1015–1027.

Schneider JV, Rabenstein R, Wesenberg J, Wesche K, Zizka G, Habersetzer J. 2018. Improved non-destructive 2D and 3D X-ray imaging of leaf venation. *Plant Methods* **14**, 1–15.

Schuetz M, Fidanza M, Mattsson J. 2019. Identification of auxin response factor-encoding genes expressed in distinct phases of leaf vein development and with overlapping functions in leaf formation. *Plants* **8**, 242.

Scoffoni C, Rawls M, McKown A, Cochard H, Sack L. 2011. Decline of leaf hydraulic conductance with dehydration: relationship to leaf size and venation architecture. *Plant Physiology* **156**, 832–843.

Shafiq I, Hussain S, Raza MA, et al. 2021. Crop photosynthetic response to light quality and light intensity. *Journal of Integrative Agriculture* **20**, 4–23.

Shi X, Tian Q, Deng P, Zhang W, Jing W. 2021. The rice aldehyde oxidase *OsAO3* gene regulates plant growth, grain yield, and drought tolerance by participating in ABA biosynthesis. *Biochemical and Biophysical Research Communications* **548**, 189–195.

Shirakawa M, Ueda H, Koumoto Y, Fuji K, Nishiyama C, Kohchi T, Hara-Nishimura I, Shimada T. 2014. CONTINUOUS VASCULAR RING (COV1) is a *trans*-Golgi network-localized membrane protein required for Golgi morphology and vacuolar protein sorting. *Plant and Cell Physiology* **55**, 764–772.

Siddique N, Paheding S, Elkin CP, Devabhaktuni V. 2021. U-net and its variants for medical image segmentation: a review of theory and applications. *IEEE Access* **9**, 82031–82057.

Song J, Shang L, Li C, Wang W, Wang X, Zhang C, Ai G, Ye J, Yang C, Li H. 2022. Variation in the fruit development gene *POINTED TIP* regulates protuberance of tomato fruit tip. *Nature Communications* **13**, 5940.

Stewart JJ, Adams W, Cohu C, Demmig-Adams B. 2018a. Tocopherols modulate leaf vein arrangement and composition without impacting photosynthesis. *Photosynthetica* **56**, 382–391.

Stewart JJ, Polutchnko SK, Demmig-Adams B, Adams WWI. 2018b. *Arabidopsis thaliana* Ei-5: minor vein architecture adjustment compensates for low vein density in support of photosynthesis. *Frontiers in Plant Science* **9**, 693.

Tadege M, Lin H, Bedair M, Berbel A, Wen J, Rojas CM, Niu L, Tang Y, Sumner L, Ratet P. 2011. *STENOFOLIA* regulates blade outgrowth and leaf vascular patterning in *Medicago truncatula* and *Nicotiana glauca*. *The Plant Cell* **23**, 2125–2142.

Tan JW, Chang S-W, Binti Abdul Kareem S, Yap HJ, Yong K-T. 2020. Deep learning for plant species classification using leaf vein morphometric. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **17**, 82–90.

Tepeli YI, Ünal AB, Akdemir FM, Tastan O. 2021. PAMOGK: a pathway graph kernel-based multiomics approach for patient clustering. *Bioinformatics* **36**, 5237–5246.

Teramoto S, Uga Y. 2020. A deep learning-based phenotypic analysis of rice root distribution from field images. *Plant Phenomics* **2**, 3194308.

Vandenbussche M, Horstman A, Zethof J, Koes R, Rijpkema AS, Gerats T. 2009. Differential recruitment of *WOX* transcription factors for lateral development and organ fusion in petunia and *Arabidopsis*. *The Plant Cell* **21**, 2269–2283.

van Dijk ADJ, Kootstra G, Kruijer W, de Ridder D. 2021. Machine learning in plant science and plant breeding. *iScience* **24**, 101890.

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. 2017. Attention is all you need. In: Guyon I, Von Luxburg U, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, eds. *Advances in Neural Information Processing Systems* 30 (NIPS 2017). https://papers.nips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

- Venturas MD, Sperry JS, Hacke UG.** 2017. Plant xylem hydraulics: what we understand, current research, and future challenges. *Journal of Integrative Plant Biology* **59**, 356–389.
- Violle C, Navas M, Vile D, Kazakou E, Fortunel C, Hummel I, Garnier E.** 2007. Let the concept of trait be functional! *Oikos* **116**, 882–892.
- Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, Haibe-Kains B, Goldenberg A.** 2014. Similarity network fusion for aggregating data types on a genomic scale. *Nature Methods* **11**, 333–337.
- Wang K, Abid MA, Rasheed A, Crossa J, Hearne S, Li H.** 2023. DNNGP, a deep neural network-based method for genomic prediction using multi-omics data in plants. *Molecular Plant* **16**, 279–293.
- Wang L, Xu J, Nian J, Shen N, Lai K, Hu J, Zeng D, Ge C, Fang Y, Zhu L.** 2016. Characterization and fine mapping of the rice gene *OsARVL4* regulating leaf morphology and leaf vein development. *Plant Growth Regulation* **78**, 345–356.
- Wang W, Guo W, Le L, et al.** 2023. Integration of high-throughput phenotyping, GWAS, and predictive models reveals the genetic architecture of plant height in maize. *Molecular Plant* **16**, 354–373.
- Wang Y, Jiao Y.** 2020. Keeping leaves in shape. *Nature Plants* **6**, 436–437.
- Wenzel CL, Schuetz M, Yu Q, Mattsson J.** 2007. Dynamics of *MONOPTEROS* and *PIN-FORMED1* expression during leaf vein pattern formation in *Arabidopsis thaliana*. *The Plant Journal* **49**, 387–398.
- Xu H, Blonder B, Jodra M, Malhi Y, Fricker M.** 2021. Automated and accurate segmentation of leaf venation networks via deep learning. *New Phytologist* **229**, 631–648.
- Xu J, Wu P, Chen Y, Meng Q, Dawood H, Dawood H.** 2019. A hierarchical integration deep flexible neural forest framework for cancer subtype classification by integrating multi-omics data. *BMC Bioinformatics* **20**, 527.
- Yamaguchi T, Nagasawa N, Kawasaki S, Matsuoka M, Nagato Y, Hirano H-Y.** 2004. The *YABBY* gene *DROOPING LEAF* regulates carpel specification and midrib development in *Oryza sativa*. *The Plant Cell* **16**, 500–509.
- Yan J, Wang X.** 2023. Machine learning bridges omics sciences and plant breeding. *Trends in Plant Science* **28**, 199–210.
- Yang W, Feng H, Zhang X, Zhang J, Doonan JH, Batchelor WD, Xiong L, Yan J.** 2020. Crop phenomics and high-throughput phenotyping: past decades, current challenges, and future perspectives. *Molecular Plant* **13**, 187–214.
- Yang W, Guo T, Luo J, Zhang R, Zhao J, Warburton ML, Xiao Y, Yan J.** 2022. Target-oriented prioritization: targeted selection strategy by integrating organismal and molecular traits through predictive analytics in breeding. *Genome Biology* **23**, 80.
- Yang X, Song Z, King I, Xu Z.** 2022. A survey on deep semi-supervised learning. *IEEE Transactions on Knowledge and Data Engineering*. doi: [10.1109/TKDE.2022.3220219](https://doi.org/10.1109/TKDE.2022.3220219).
- Zebari R, Abdulazeez A, Zeebaree D, Zebari D, Saeed J.** 2020. A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. *Journal of Applied Science and Technology Trends* **1**, 56–70.
- Zhang S, Liu C-C, Li W, Shen H, Laird PW, Zhou XJ.** 2012. Discovery of multi-dimensional modules by integrative analysis of cancer genomic data. *Nucleic Acids Research* **40**, 9379–9391.
- Zhao C, Zhang Y, Du J, Guo X, Wen W, Gu S, Wang J, Fan J.** 2019. Crop phenomics: current status and perspectives. *Frontiers in Plant Science* **10**, 714.
- Zhao D, Liu L, Yu F, Heidari AA, Wang M, Liang G, Muhammad K, Chen H.** 2021. Chaotic random spare ant colony optimization for multi-threshold image segmentation of 2D Kapur entropy. *Knowledge-Based Systems* **216**, 106510.
- Zhou J, Hao M, Zhang D, Zou P, Zhang W.** 2019. Fusion PSPnet image segmentation based method for multi-focus image fusion. *IEEE Photonics Journal* **11**, 1–12.