

Methods

Automated and accurate segmentation of leaf venation networks via deep learning

Hao Xu^{1*}, Benjamin Blonder^{2,3*} , Miguel Jodra², Yadvinder Malhi²  and Mark Fricker^{4*} 

¹Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Old Road Campus Research Building, Oxford, OX3 7DQ, UK; ²Environmental Change Institute, School of Geography and the Environment, University of Oxford, South Parks Road, Oxford, OX1 3QY, UK; ³Department of Environmental Science, Policy, and Management, University of California, 120 Mulford Hall, Berkeley, CA 94720, USA; ⁴Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, UK

Author for correspondence:
Mark Fricker
Email: mark.fricker@plants.ox.ac.uk

Received: 20 July 2020
Accepted: 24 August 2020

New Phytologist (2021) 229: 631–648
doi: 10.1111/nph.16923

Key words: biological network analysis, convolutional neural network, deep learning, hierarchical loop decomposition, leaf trait, leaf venation network, network scaling, spatial transportation network.

Summary

- Leaf vein network geometry can predict levels of resource transport, defence and mechanical support that operate at different spatial scales. However, it is challenging to quantify network architecture across scales due to the difficulties both in segmenting networks from images and in extracting multiscale statistics from subsequent network graph representations.
- Here we developed deep learning algorithms using convolutional neural networks (CNNs) to automatically segment leaf vein networks. Thirty-eight CNNs were trained on subsets of manually defined ground-truth regions from >700 leaves representing 50 southeast Asian plant families. Ensembles of six independently trained CNNs were used to segment networks from larger leaf regions (c. 100 mm²). Segmented networks were analysed using hierarchical loop decomposition to extract a range of statistics describing scale transitions in vein and areole geometry.
- The CNN approach gave a precision-recall harmonic mean of 94.5% ± 6%, outperforming other current network extraction methods, and accurately described the widths, angles and connectivity of veins. Multiscale statistics then enabled the identification of previously undescribed variation in network architecture across species.
- We provide a LEAFVEINCNN software package to enable multiscale quantification of leaf vein networks, facilitating the comparison across species and the exploration of the functional significance of different leaf vein architectures.

Introduction

Plant leaves are structured by a vein network that ranges in geometry from dendritic fans with few branches and no loops (e.g. *Ginkgo biloba*) to hierarchical forms with many loops (e.g. *Acer saccharum*) (Trivett & Pigg, 1996; Roth-Nebelsick *et al.*, 2001). The vein network has multiple roles that include transport of water, nutrients and sugars through the xylem and phloem tissues (Brodrribb *et al.*, 2010; Carvalho *et al.*, 2018; Katifori, 2018), controlled deformation during bud burst (Niklas, 1999), and mechanical support and resistance to damage in the mature leaf (Sack & Scoffoni, 2013; Sharon & Sahaf, 2018). Measurements of vein architecture have provided insights into leaf development (Kang & Dengler, 2004) and evolution (Boyce *et al.*, 2009; Brodrribb & Feild, 2010), prediction of leaf carbon and water fluxes (Sack & Frole, 2006; Brodrribb *et al.*, 2007; Brodrribb *et al.*, 2010), prediction of environmental and stress tolerances of species (de Boer *et al.*, 2012; Blonder & Enquist, 2014; Brodrribb

et al., 2016) and reconstruction of paleo-environments from leaf fossils (Manze, 1967; Uhl & Mosbrugger, 1999; Blonder *et al.*, 2014). Venation networks are also relevant to test theories of optimal branching structure and transportation efficiency for different network architectures (Pelletier & Turcotte, 2000; Price *et al.*, 2012; Price *et al.*, 2014; Price & Weitz, 2014), particularly in response to fluctuating loads or robustness to damage (Dodds, 2010; Katifori *et al.*, 2010; Katifori, 2018).

Prediction of function is based on statistics estimated from images of vein networks (Roth-Nebelsick *et al.*, 2001). Networks have veins of different orders, varying from primary veins (attached to the petiole), to secondary (attached to the primary veins), and continuing until the ultimate veins are reached. The vein density (vein length per unit area) at each order, or across all orders, is therefore a key statistic (Uhl & Mosbrugger, 1999; Sack & Scoffoni, 2013). Measurement of the distribution of vein radii is required to determine the vein order (Price *et al.*, 2012), which is itself challenging. Alternatively, vein orders can be extracted using hierarchical vein classification methods (Gan *et al.*, 2019). If vein radii can be extracted reliably, they can also be used to

*These authors contributed equally to this work.

estimate network construction costs (John *et al.*, 2017) or to predict maximum fluxes and resource flow rates (Brodribb *et al.*, 2007; McKown *et al.*, 2010). Many networks also contain loops that partition the lamina into increasingly small areoles. For higher vein orders, qualitative terms to describe patterning, such as craspedodromous or camptodromous, have traditionally been used (Hickey, 1979; Ellis *et al.*, 2009). Quantitatively, the number of areoles per unit area provides a statistic for the amount of 'loopiness' (Blonder *et al.*, 2011). More detailed analysis of the nesting of loops within higher-order loops can be achieved using hierarchical loop decomposition (HLD), which constructs a binary branching tree representing the order in which adjacent areoles become connected as the intervening veins are removed (Katifori *et al.*, 2010; Katifori & Magnasco, 2012; Ronellenfitsch *et al.*, 2015; Katifori, 2018). HLD provides a range of metrics such as the Horton–Strahler index, bifurcation ratio and subtree asymmetry by analogy to river network analysis (Katifori & Magnasco, 2012; Mileyko *et al.*, 2012; Ronellenfitsch *et al.*, 2015; Katifori, 2018). Some venation networks contain freely ending veins (FEVs) that do not form anastomoses with other veins, but which may help with the efficient supply of water (Fiorin *et al.*, 2016), or possibly the transport of sugars (Carvalho *et al.*, 2018). As such, density estimates of branching end-points or areole perimeter/area ratios (Kang & Dengler, 2004; Blonder *et al.*, 2018) are also useful. Finally, the distribution of branching angles and radius or length ratios at vein branching points (Bohn *et al.*, 2002) feed into theories and models of flow efficiency, network development and architecture (Pelletier & Turcotte, 2000; Price *et al.*, 2012, 2014).

Calculating these metrics requires accurate segmentation of the venation network over a range of scales. The finest veins are typically around 10–20 µm diameter, whilst the largest veins can reach mm width in a lamina up to m² in area. To capture the fine veins, networks are typically imaged after chemical clearing and staining using light microscopy or flat-bed scanning (Pérez-Harguindeguy *et al.*, 2013). X-ray imaging can also achieve high resolution with no tissue preparation, but requires access to suitable equipment (Wing, 1992; Blonder *et al.*, 2012; Schneider *et al.*, 2018). Regardless of the imaging method, accurate segmentation of the network from the image has been challenging. Images often have limited or uneven contrast, or contain unavoidable artefacts caused by other tissues and cell types such as trichomes or glands. Furthermore, field material often includes damaged regions. As a result, simple intensity-thresholding is rarely adequate to yield useful segmentations. Several computer programs have addressed this challenge, typically using local or global enhancement filters tuned to pick out vein-like structures, either with or without human supervision. These include PHENOVEIN (Bühler *et al.*, 2015), NET (Lasser & Katifori, 2017), LEAFGUI (Price *et al.*, 2011), LIMANI (Dhondt *et al.*, 2012), as well as other studies that have not provided code (see for example Bohn *et al.*, 2002; Mounsef & Karam, 2012; Larese *et al.*, 2014; Parsons-Wingerter *et al.*, 2014; Gan *et al.*, 2019). Other packages include machine learning methods to perform segmentation, for example NEFI (Rother *et al.*, 2004; Dirnberger *et al.*, 2015; Grinblat *et al.*, 2016), although these require extensive user interaction to

iteratively define and correct foreground and background partitioning, and have only been applied to simple leaf vein networks (Carvalho *et al.*, 2017).

However, in our experience, none of these segmentation algorithms is robust in that it can be accurately applied across a wide range of species and image qualities, particularly for field-collected specimens, without considerable expert human guidance. Most methods are either highly tuned for specific applications, for example *Arabidopsis thaliana* (Bühler *et al.*, 2015), or require the adjustment of many parameters to produce an acceptable output (e.g. Price *et al.*, 2011). Even in the best case scenarios, when these algorithms correctly classify the majority of pixels, they may fail to accurately reconstruct the topology of the network, falsely connecting or disconnecting veins, or producing incorrect vein radii. The difficulties with automated segmentation are well recognised with most studies preferring hand-traced images (e.g. Rolland-Lagan *et al.*, 2009). Indeed, current methods handbooks only recommend expert hand-tracing to estimate venation network properties (Pérez-Harguindeguy *et al.*, 2013). Nevertheless, whilst hand-tracing is accurate, it is slow and time intensive, limiting the throughput of samples and restricting the area of leaf analysed.

Here we developed an alternative deep learning method to segment minor vein networks from high-resolution cleared leaf images. The approach was motivated by our observation that humans can trace venation networks consistently and accurately after minimal training, despite low contrast and the presence of many artefacts in the underlying images. We therefore describe an approach for training an ensemble of convolutional neural networks (CNNs) to learn how to carry out this segmentation task based on ground-truth (GT) tracings previously made by humans from field-collected leaves of tropical rainforest trees. The trained CNNs are provided within the LEAFVEINCNN software package, which can be applied to any high-resolution cleared leaf image.

Materials and Methods

Leaf vein imaging

A set of calibration leaves was sampled from eight permanent forest plots in Sabah state, Borneo, Malaysia (Riutta *et al.*, 2018), and which were characterised by a mixed dipterocarp lowland forest. In total, 727 samples were obtained from 295 species (50 families). The sampling protocols and nonvein datasets have been described previously (Both *et al.*, 2019) and the complete set of vein images published (Blonder *et al.*, 2019).

A 1 cm² sample was cut from the middle of the lamina, excluding any primary veins, chemically cleared, stained, and slide-mounted following standard protocols (Pérez-Harguindeguy *et al.*, 2013; Blonder *et al.*, 2018). Each sample was imaged using a compound microscope (Olympus, BX43) with a ×2 apochromat objective and an Olympus SC100 colour camera (3840 × 2748 pixel resolution). Here, 9–16 overlapping image fields were stitched together to obtain a complete image of the sampled area using the IMAGE CAPTURE software. Final images

had a resolution of 595 pixels mm⁻¹ and a typical extent of *c.* 7000 × 7000 pixels.

Images were preprocessed by retaining the green channel (which had the greatest contrast), and applying contrast-limited adaptive histogram equalisation (CLAHE; Zuiderveld, 1994), with a 400 × 400 pixel window and a contrast limit of 0.01.

GT tracing

All veins within a polygonal region-of-interest (ROI) of *c.* 700 × 700 pixels were traced manually at full width for each image, using a digitising tablet (Cintiq 22HD, Wacom) and image processing software (GIMP). In addition, any veins in the complete image with width >200 µm were manually traced, as these veins were not routinely included in the GT-ROIs used to train the CNNs. Inclusion of manual large veins avoided over-segmentation of texture within the larger veins, which were occasionally recognised by the CNN as separate veins. Alternatively, we used down-sampled images with the same CNN to automatically segment larger veins at full width. Lastly, a separate mask was traced manually to exclude damaged or background regions from the full image.

Each tracing required *c.* 30 min. of human time. Each image layer in GIMP (i.e. the ROI, large vein or mask) was drawn by a single person and was reviewed and subsequently edited by two or three other people until consensus was reached, to give a final set of 727 GT images.

CNN segmentation using U-Net

CNNs are a deep learning method that allows automated extraction of features from images (Krizhevsky *et al.*, 2012; LeCun *et al.*, 2015). They differ from classic machine learning methods in that they synthesise information at multiple scales via a sequence of convolution, pooling and up-sampling operations. Convolution–deconvolution networks, such as U-Net (Ronneberger *et al.*, 2015), use a set of rectangular kernels that picks up different features in the image at each scale, and that are then pooled and combined in each network layer to ultimately yield pixel-level classifications. ‘Deep’ neural networks include multiple convolutional layers whilst ‘learning’ occurs via a backpropagation algorithm (Werbos, 1974) that optimises the weights of the network by comparing the output result to the GT.

We used a CNN based on the U-net architecture (Ronneberger *et al.*, 2015). The first convolutional layer comprised 32 small (3 × 3) kernels that were convolved with the image to give 32 feature maps (Fig. 1). The feature maps were batch-normalised to give a mean of zero and variance of 1 to speed up training (Ioffe & Szegedy, 2015), followed by nonlinear rectification (ReLU) that sets negative inputs to zero to improve training performance (Glorot *et al.*, 2011). A second convolution using another set of 32 small (3 × 3) kernels was used to increase the receptive field in this layer (Simonyan & Zisserman, 2014), followed by batch-normalisation and ReLU (Fig. 1). The size and resolution of the input feature maps and the output feature maps were preserved when passing through the convolutional blocks,

with padding applied during convolutions. To generate the next layer, the output from each convolution block was subject to nonlinear down-sampling using the maximum from a (2 × 2) pixel window with a stride of 2, effectively halving the resolution of the feature maps in both dimensions. This gave five levels of feature maps in the down-sampling arm of the network that provided increasingly higher-level representation of the input image patches.

We also followed the approach of U-Net and Fully Convolutional Networks (FCN) (Long *et al.*, 2015), and kept the same number of feature maps for layers with the same resolution, whilst doubling the number of feature maps when the resolution was halved by down-sampling, and halving the number of feature maps when resolution was doubled in the up-sampling arm. The feature maps at each down-sampled layer were also copied and concatenated with those at the same resolution in the up-sampling arm, to give feed-forward shortcut connections that retained the high-resolution information required to yield a per-pixel output (Long *et al.*, 2015). The activation function of the final fully connected output layer was sigmoid and gave a 1 : 1 pixel classification of the vein probability. This resulted in nine convolutional blocks in total (Fig. 1).

Data augmentation and training data sets

The manually digitised GT images were used to train and evaluate the CNN. Each GT region was resampled to augment the training data and to prevent over-fitting. In total, 32 samples per image, fully contained within the GT region, were extracted with randomised variation in the box size (256 × 256 ± 20% pixels), rotation angle (between 0 and 2π radians), (*x*, *y*) shift in centroid coordinates, and randomised image reflection. Augmented samples were then resized to 256 × 256 pixels (Fig. 2).

Network training

The CNN was initialised using the Glorot uniform initialiser with random weights drawn from a uniform distribution (Glorot & Bengio, 2010). At the start of each training run, 120 nonoverlapping images were reserved as the test data set. For each training epoch, 480 leaf images were randomly selected from the remaining images and the manually digitised ROIs used as training samples. The remaining 120 images were used to validate the network performance. The trained CNN was then used to segment the set of 120 full-size test images that had not been used in training or validation. This required seven iterations to segment all of the original full-size images. The process was repeated to give an ensemble of six independently trained networks for each image (Hansen & Salamon, 1990; Sollich & Krogh, 1996). A flow diagram of the process is shown in Fig. 2. Of the 727 samples in the dataset, seven had digitised regions that were too small for image patch extraction, and which were excluded from training, but were still used for validation. At each training iteration, a mini-batch of eight 256 × 256 patches, constrained by the memory limit of each graphical processing units (GPU), was sampled and fed to the network in sequence to ensure robust convergence

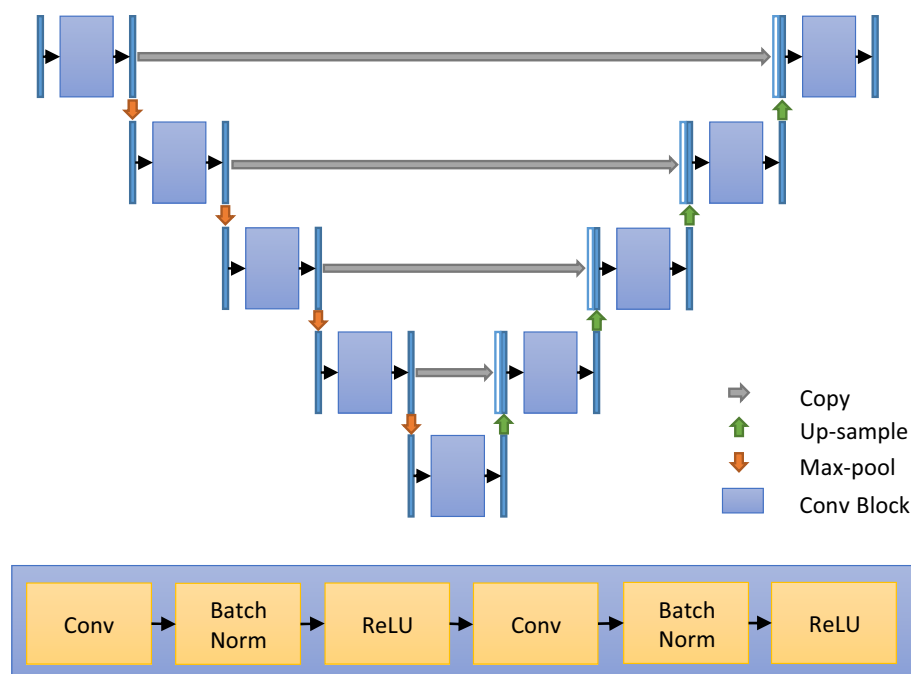


Fig. 1 Schematic diagram of the U-Net convolutional neural network (CNN) architecture. Each layer in the CNN is built from two concatenated blocks comprising a convolution layer, followed by batch-normalisation and nonlinear rectification (ReLU). This is followed by 2×2 max pooling to reduce the scale in the down-sampling arm (left). In addition, a copy of the output of the convolution block is appended to the up-sampling arm (right) at the same level that retains the high-resolution information required to yield per-pixel outputs. The final layer is fully connected and provides a probability classification into vein or nonvein pixels using a sigmoid function.

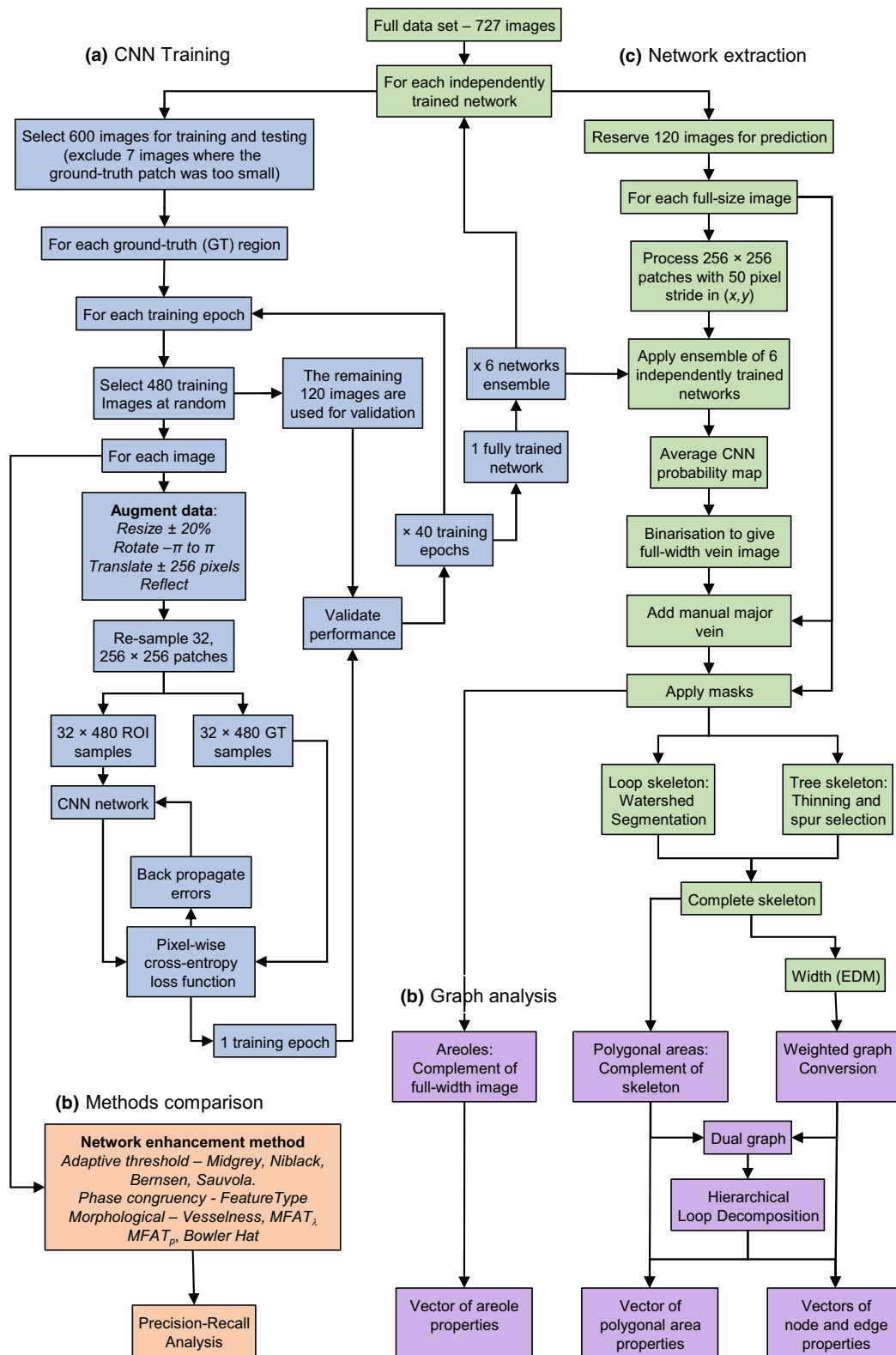
by avoiding local minima. The pixel-wise cross-entropy was used as the loss function during training, and the Dice similarity metric was used for network selection. Networks were typically trained for 20–40 epochs, in which an epoch was a complete run through one set of 480×32 samples, with a new training dataset used at each epoch (giving $480 \times 32 \times 40$ augmented training samples) until the error asymptotically approached a stable value with $< 0.1\%$ variation.

The U-Net algorithm was run on one of two clusters, either with $2 \times$ Intel E5-2609v2 processors (total eight cores at 2.5

GHz, 128 GB RAM), or $2 \times$ Intel E5-2609v3 processors (total 12 cores at 1.9 GHz, 128 GB memory). Image processing occurred in parallel on three GPUs (nVidia, GeForce GTX 1080 Ti, each with 3580 cores and 11 GB memory). All CNN code was implemented in Keras with TensorFlow using Python 3. Training the CNN on the calibration data required 8 min per epoch using three GPUs and one central processing unit (CPU).

The trained networks were also imported into MATLAB (Mathworks, Natick, MA, USA) for incorporation into a graphical user interface (GUI) interface (software and manual available from

Fig. 2 Flow diagram of the major steps in convolutional neural network (CNN) training, segmentation and analysis. At each iteration, one set of 120 nonoverlapping images is reserved for prediction and network extraction (green boxes). At each training epoch (blue boxes), 480 images were selected at random for training from the remainder, and a further 120 reserved for validation (a). For each training image, a region-of-interest (ROI) and the associated manually delineated full-width ground-truth (GT) were augmented by scaling, translation, rotation and reflection to give $32, 256 \times 256$ image patches. The patches were fed to the CNN, and the output compared with the augmented GT patches using the Dice Similarity Coefficient (F_1 statistic). The errors are back-propagated through the network to adjust the internal weights, and the process repeated with another training image. Once the network has been trained on all 480 initial images, termed an epoch, the output is tested against the validation set. Training continues for up to 40 epochs to give a fully trained network, and then repeated to give an ensemble of six independently trained networks. These are used to predict the vein architecture in the set of reserved test images. Complete prediction for the complete data set requires seven iterations. (b) The performance of the CNN approach is compared with a range of different network enhancement and segmentation methods (orange boxes) including local adaptive thresholding (Midgrey, Niblack, Bernsen and Sauvola), or multiscale ridge enhancement methods (Vesselness, MFAT₁, MFAT_p, Bowler Hat and phase-congruency), using Precision-Recall (PR) analysis. (c) To extract the vein network, the full-sized images are processed through the ensemble of independently trained CNN networks to give an average probability map of the vein network (green boxes). The probability map is segmented to give a full-width binary image (FWB) and then skeletonised in a two-step process to give loops and trees. The width of the veins is estimated from the Euclidean Distance Map (EDM) of the FWB image. (d) The skeleton is converted to a weighted graph representation with nodes at junctions linked by edges along the veins, which are both associated with a vector of properties (purple boxes). The areoles are extracted from the complement of the FWB image, whilst the polygonal regions, including the area of the veins up to the skeleton are extracted from the complement of the skeleton. In addition, a weighted dual graph is constructed that links adjacent areoles (nodes) with edges weighted by the width of the intervening vein. Edges are removed in sequence to fuse the parent areoles and reveal the nested loop structure of the network by hierarchical loop decomposition (HLD). Network metrics are calculated at each step of the HLD to track the vein architecture at different scales.



<https://doi.org/10.5281/zenodo.4007731>). This required replacement of the Keras *Upsampling2D* layers with MATLAB *transposedConv2dLayer* layers using a stride of 2 and uniform weights equal to 1.

Vein prediction on the full-size images

For vein prediction over the full 10 mm × 10 mm region, images were processed as overlapping 256 × 256 pixel patches. At each

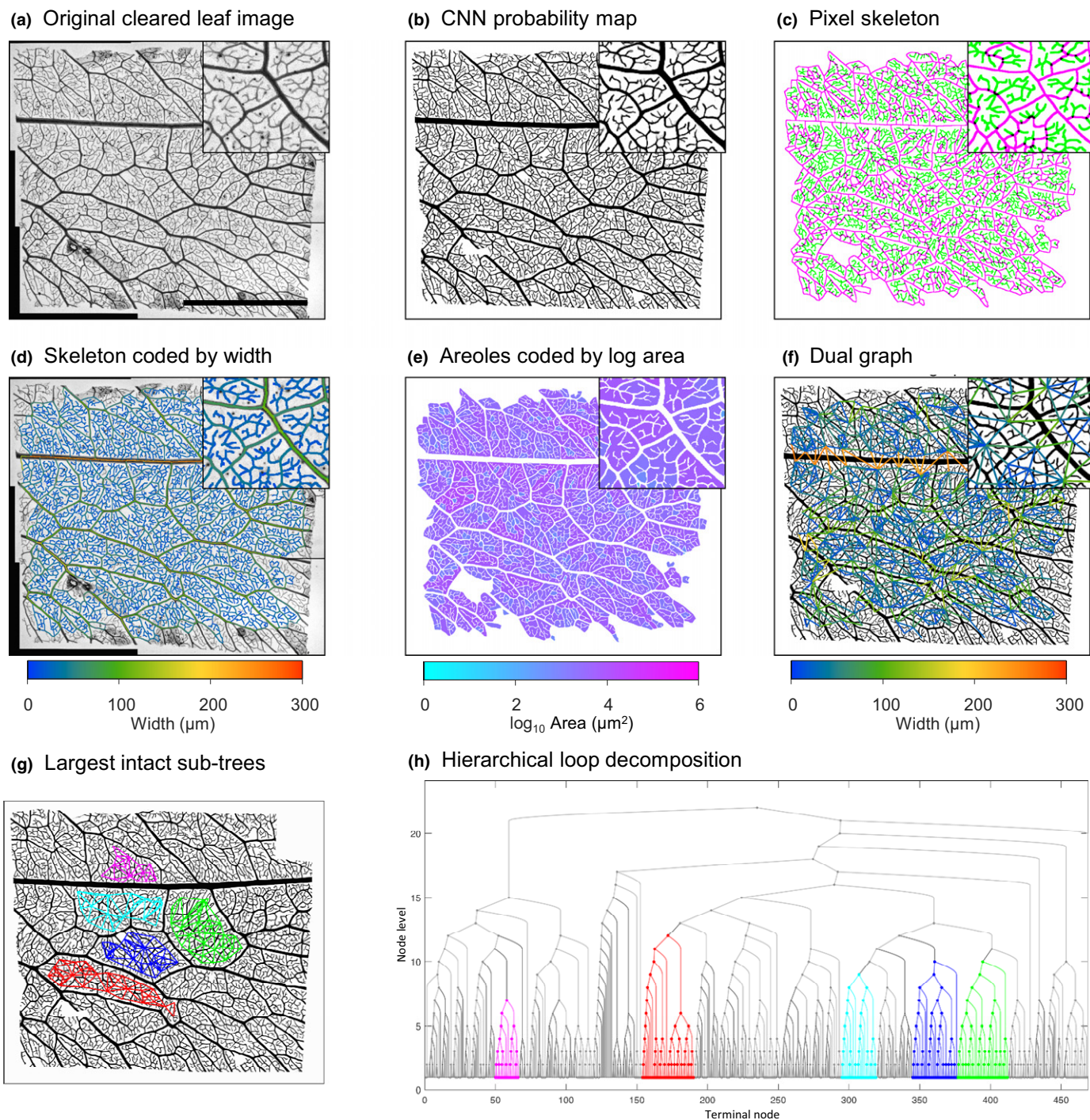


Fig. 3 Processing steps in convolutional neural network (CNN) training, segmentation and analysis. (a) Original image of a typical cleared leaf sample following CLAHE enhancement, with inset showing the central region at $\times 6$ zoom. Bar, 5 mm. (b) Average ensemble CNN probability map. (c) Pixel skeleton with loops in magenta and trees in green. (d) Pseudocolour-coded vein width skeleton superimposed on the original image. (e) Areoles, pseudocolour-coded by \log_{10} area. (f) Dual graph connecting centroids of adjacent areoles with the edge weight given by the average width of the intervening vein. (g) Overlay of the dual graph for the five largest intact subtrees in the full-size image. (h) The complete hierarchical loop decomposition (HLD) of the dual graph including the five largest intact subtrees following the colour scheme in (g).

iteration, a mini-batch of eight patches was sampled at random without replacement, with a separation (stride) of 50 pixels in both x and y to give >25 -fold over-sampling, and a small overlap between patches. The results were assembled back to the original

image size to give a probability map of predicted values, with each pixel classified at least 25 times within different image contexts due to the stride overlap. Prediction of the CNN on a full-sized image took ≈ 8 min on the GPU clusters. The analysis code

running on a single CPU took *c.* 20 min depending on the complexity of the network and amount of down-sampling. The typical processing speed was *c.* 60 000 pixels s^{-1} .

Conversion of the CNN probability map to a weighted network

For each full-sized leaf image (Fig. 3a), the six CNN probability maps were averaged to generate a mean probability map (Fig. 3b), down-sampled by a factor of 2 to reduce subsequent computational costs, and binarised using a threshold based on the average F_β statistic from Precision-Recall analysis (to be described later). Only the largest connected component was retained. Damaged and background regions that were defined manually for each image were masked from analysis.

The resultant full-width binary (FWB) image was thinned to a single-pixel wide skeleton (Supporting Information Fig. S1b). Simple thinning algorithms gave an irregular zig-zag backbone around junctions. A better centreline was realised using a hybrid skeletonisation protocol that combined watershed segmentation to find all closed loops (Fig. 3c, magenta), with thinning to extract tree-like branches (Fig. 3c, green). In brief, watershed segmentation was applied to the Euclidean Distance Map (EDM) from the FWB image to find all closed loops (Fig. 3c, magenta). Tree-like branches were extracted as the difference between the complete thinned skeleton (Zhang & Suen, 1984), and the skeleton after removal of spurs (Fig. 3c, green). The 'trunks' of the tree-like skeleton were then re-connected to the nearest pixel in the watershed loop skeleton, determined from the nearest-neighbour transform of the EDM, and the pixel positions written back into the image using the Bresenham line algorithm (Bresenham, 1965). This hybrid approach had the benefit that loops and tree-like portions of the skeleton were automatically separated and could be analysed separately (Fig. 3c).

The pixel skeleton was converted into separate vein segments using the MATLAB *edgeline* algorithm (Kovesi, 2000), which provided a set of pixel-coordinates for each vein in the network. The vein segments were converted to a graph representation (Fig. S1c) with nodes (N) assigned to each junction or free end connected by edges (E) (Roth-Nebelsick *et al.*, 2001).

The width at each pixel of the skeleton was calculated from the EDM. Pixel width estimates were averaged to give an initial width estimate for each vein. However, this over-estimated the width for small veins branching from larger veins, as it included values where the smaller vein overlapped the larger one (Bohn *et al.*, 2002). Thus, the width for the smallest vein at each junction was refined by excluding pixels in the overlap region, before taking the average of the remaining pixels.

To visualise the segmented network, the average centre-width of each edge was mapped back into the pixel skeleton, colour coded on a perceptually uniform rainbow scale, and superimposed on the original leaf image (Fig. 3d). The areole regions between the veins (i.e. excluding the veins themselves) and the area of the polygonal region up to the pixel skeleton were determined from the complement of the FWB image or the skeleton, respectively (Fig. 3e).

Single-scale network metrics

Multiple metrics were extracted from the graph representations for each vein (Fig. S1d; Table S1), node (Fig. S1e; Table S2), areole, and polygonal area (Fig. S1f; Tables S3, S4, respectively). In addition, a set of summary statistics describing the distribution of each metric were also calculated (Table S5), adopting a comparable terminology to Larese *et al.* (2014), with prefixes *V* - veins, *N* - nodes, *A* - areoles and *P* - polygonal areas. Vein metrics were further subdivided into the total vein network (*Tot*), veins forming loops (*Loop*), and veins forming trees (*Tree*). The summary statistics were drawn from mean (*av*), median (*md*), min (*mn*), max (*mx*), standard deviation (*sd*), or skewness (*sk*) of each metric distribution. In the case of angular metrics, values were summarised using circular statistics (Berens, 2009). Each summary statistic was then given a three-letter code.

The revised 'centre-weighted' width value (three-letter code, *Wid*), along with the corresponding length (*Len*), excluding vein overlap and measured as the sum of the Euclidean distance between edge pixels, surface area (*SAr*), volume (*Vol*), tortuosity (*Tor*), and orientation (*Ori*) were added to the vector of edge properties, along with the average edge intensity (*Int*) and average CNN probability (*Prb*) (Fig. S1d). A vector of properties was also calculated for each node that included the node degree (*K*), diameter of the parent branch (*BD0*), and daughter branches (*BD1* and *BD2*), their orientation (*OD0*, *OD1* and *OD2*), and branching angles (*A10*, *A20* and *A21*). The angular metrics were calculated for straight-line segments joining the node and the midpoint of each vein (Fig. S1e).

Areole and polygonal area metrics

Several morphological features were automatically calculated for each areole (*A*) and polygonal region (*P*) (Fig. S1f; Tables S3, S4, respectively), along with summary statistics labelled with a three-letter code, including the area (*Are*), convex area (*CnA*), eccentricity (*Ecc*), equivalent diameter (*EqD*), perimeter (*Per*), major (*Maj*) and minor (*Min*) axes, orientation (*Ori*), solidity (*Sld* = Are/CnA), elongation (*Elg* = Maj/Min), circularity (*Cir* = $4 \times \pi \times Are/Per^2$) and roughness (*Rgh* = Per^2/Are), maximum (*Dmx*) and mean (*Dav*) distance to the nearest vein.

Dual-graph representation and HLD

A dual graph of the vein skeleton was calculated that joined adjacent polygons, with nodes placed at the centroid of the polygons, and the connecting edges weighted using the width of the intervening vein segment (Fig. 3f). The dual graph was converted to a nested branching tree using HLD (Katifori & Magnasco, 2012) by progressively removing the thinnest edges in sequence and fusing the adjacent polygons until only a single polygon remained (Fig. 3g,h). This process determined how vein loops are nested within larger vein loops across spatial scales. A subset of vein and polygonal area metrics was calculated at each step, allowing the analysis of network metrics across the scale of the fusion events (Table S6). Analysis was restricted to fully bounded regions (Fig.

3g) to ensure that metrics were not influenced by incomplete veins or areoles truncated at the boundary, although this also restricted the number of scales available in each sample.

Precision-recall analysis

The performance of the CNN classification and selection of the optimum threshold were determined by Precision-Recall (P-R) analysis. The number of true positives (T_p), true negatives (T_N), false positives (F_p) and false negatives (F_N) was calculated against the manually defined GT image. For comparisons of skeleton extraction, a tolerance of three pixels (*c.* 10 μm for the down-sampled images) was allowed (Lopez-Molina *et al.*, 2013). P-R analysis was used in preference to Receiver Operating Characteristic (ROC) plots, as the former is better suited to imbalanced data sets, where T_N from the background is expected to be much greater than T_p from the skeleton (Saito & Rehmsmeier, 2015). Precision (P) was calculated as $P = \frac{T_p}{T_p + F_p}$, recall (R) as $R = \frac{T_p}{T_p + F_n}$ and overall performance assessed with the Dice similarity metric (F_1) score as the harmonic mean of precision and recall, $F_1 = 2 \times \frac{P \times R}{P + R}$ or the more generalised F_β score, where $F_\beta = (1 + \beta^2) \cdot \frac{P \times R}{\beta^2 \cdot P + R}$, in this case using $\beta = 2$, which weights recall higher than precision. The threshold that gave the highest F_1 or $F_{\beta 2}$ score was used to segment the image.

Comparison with other vein segmentation methods

To compare the performance to other methods, the GT-ROI was analysed using a range of standard local adaptive thresholding methods including Midgrey, Niblack (Niblack, 1985), Sauvola (Sauvola & Pietikäinen, 2000) and Bernsen (Bernsen, 1986) algorithms, effectively extending the local adaptive threshold approach used in *LeafGUI* (Price *et al.*, 2011; Green *et al.*, 2014). Each method was applied to an inverted version of the original image, with a disc-shaped local neighbourhood with a radius of 45 pixels, slightly greater than the largest vein in the down-sampled GT images. The offset constant was varied from -0.5 to 0.5 (for Midgrey and Niblack) or 0 to 1 for the other algorithms, in 0.05 increments, and the optimum threshold value selected using the F_1 or $F_{\beta 2}$ metric from the P-R analysis.

To emulate the multiscale Hessian-based Vesselness approach used in *phenoVein* (Bühler *et al.*, 2015), the MATLAB fibremetric implementation of the Frangi 'Vesselness' filter (Frangi *et al.*, 1998) was used with vein thickness between 4 and 9 pixels, applied over four scales using a Gaussian image pyramid (Burt & Adelson, 1983) to cover the largest veins. We also tested improved Hessian-based enhancement techniques using Multiscale Fractional Anisotropy Tensors (MFAT), in both their eigenvalue-based (MFAT_v) and probability-based (MFAT_p) form (Alhasson *et al.*, 2019), and the intensity-independent, multiscale phase-congruency enhancement developed by Kovsi (Kovsi, 1999, 2000), which was previously used to segment fungal (Obara *et al.*, 2012), slime mould (Fricker *et al.*, 2017) and ER networks (Pain *et al.*, 2019). We used the normalised local weighted mean phase angle ('Feature Type') to give intensity-

independent enhancement of vein structures, initially calculated over 3 – 5 scales and six orientations, and then applied to an image pyramid to cover larger scales. In both the vesselness and phase-congruency enhancement, noise was suppressed by setting all variation <0.1 – 0.2 in an extended minimum transformation to zero. As an alternative morphology-based approach, we also tested the multiscale *Bowler Hat* algorithm (Sazak *et al.*, 2019), which was designed to be more robust at junctions and less sensitive to interference from blob-like inclusions. In all cases, the resultant multiscale representation was collapsed to a single image using a maximum projection and segmented over varying thresholds from 0 to 1 in 0.05 increments to find the optimum performance as judged by the maximum F_1 or $F_{\beta 2}$ statistic in Precision-Recall comparison with GT. A full set of processing parameters for these methods is given in Table S7.

As the vein enhancement approaches may be better suited to extraction of the pixel skeleton, rather than a binary image covering the full width of the vein, the P-R analysis was also run following conversion of the binary image at each threshold value to a single-pixel wide skeleton.

Data and algorithm availability

A MATLAB App or the standalone LEAFVEINCNN GUI software package, including the trained networks, and manual (Fig. S2) are openly available from <https://doi.org/10.5281/zenodo.4007731>. The original image dataset is openly available (Blonder *et al.*, 2019). The down-sampled CNN predictions, ground truths, and the MATLAB scripts used for the Precision-Recall (PR) analysis and calculation of network metrics are openly available from <https://doi.org/10.5281/zenodo.4008614>. All results are openly available from <https://doi.org/10.5281/zenodo.4008361>.

Results

CNNs provided high accuracy vein network segmentation

To illustrate the accuracy of the ensemble CNN approach, we show weighted networks for six species (*Artocarpus odoratissimus*, *Dryobalanops lanceolata*, *Lophopetalum javanicum*, *Macaranga pearsonii*, *Pentace laxiflora*, and *Terminalia citrina*) covering a range of different vein architectures from dense loops to open trees (Fig. 4). In each case the centreline of each vein was correctly identified, the topology of the complete network was segmented accurately and the width corresponded to the vein thickness.

To quantify ensemble CNN performance, we examined a masked ROI from each full-sized image (Fig. 5a) containing the manually delineated GT (Fig. 5a'). The ensemble CNN applied to this ROI provided a smooth, high-contrast probability map of vein identity (Fig. 5b) that was thresholded to give a FWB image (Fig. 5b') and compared with the GT using P-R analysis (Fig. 5l). The value used to threshold the CNN probability map was varied systematically to generate a P-R curve, and the optimum value determined from the maximum F_1 or $F_{\beta 2}$ metrics (see Fig.

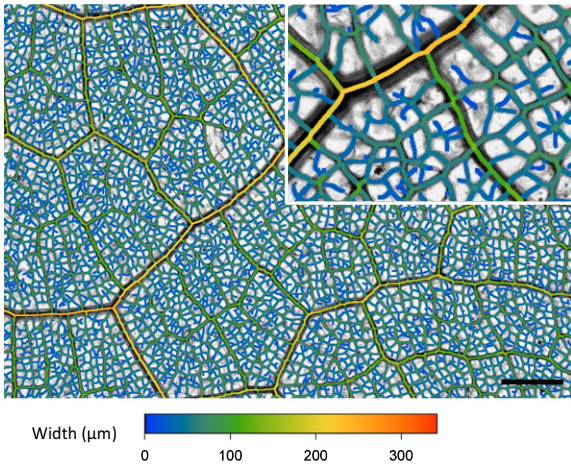
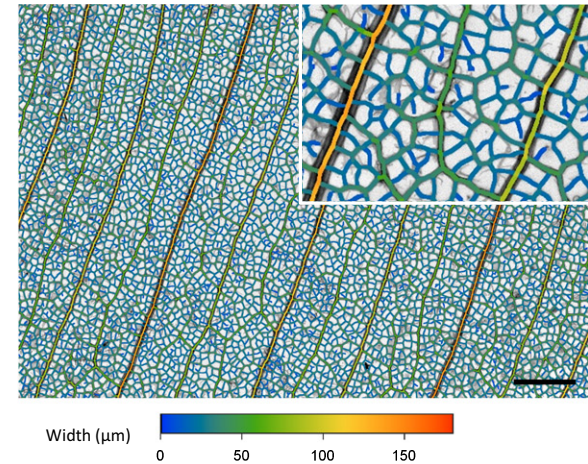
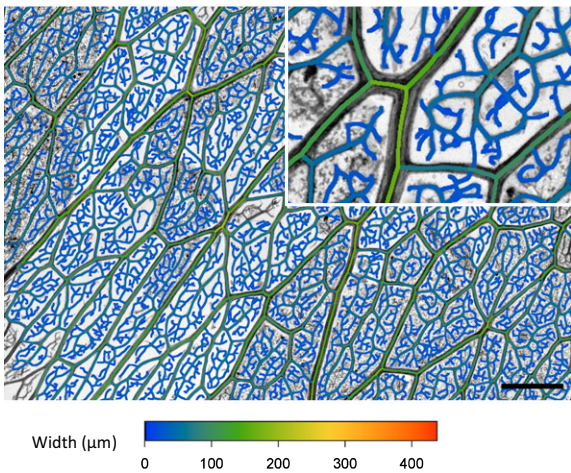
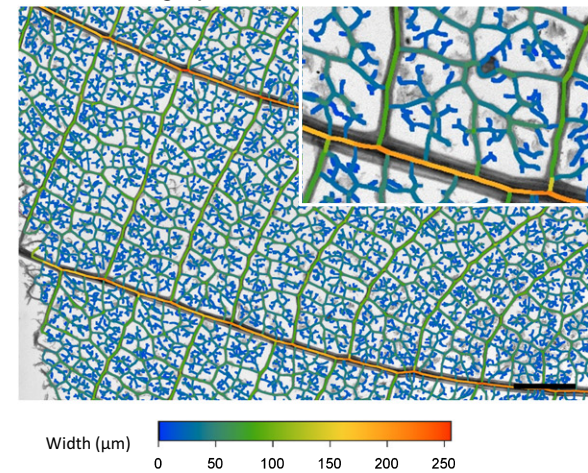
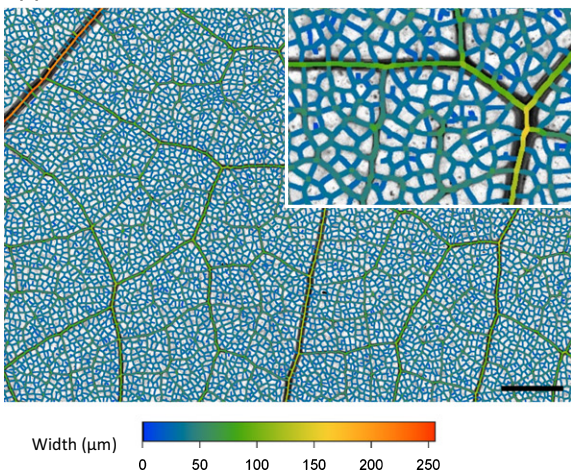
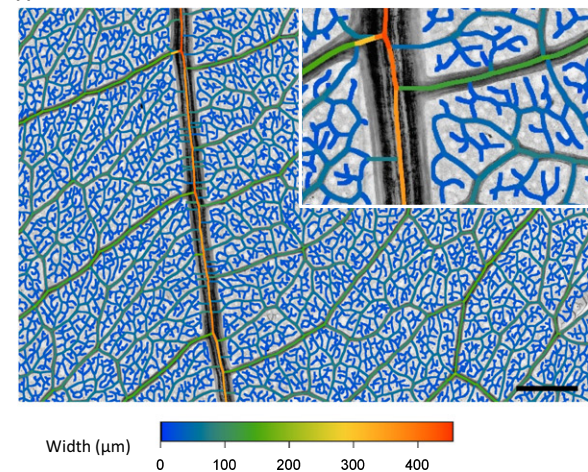
(a) *Artocarpus odoratissimus*(b) *Dryobalanops lanceolata*(c) *Lophopetalum javanicum*(d) *Macaranga pearsonii*(e) *Pentace laxiflora*(f) *Terminalia citrina*

Fig. 4 Typical convolutional neural network (CNN) segmentation results. Six examples of leaf venation networks (grey scale images) overlaid with the extracted network skeleton pseudocolour-coded to represent vein thickness on a rainbow scale (blue, thin; red, thick). Inset panels show a three-fold zoom of centre-located region-of-interest (ROI). Bars, 10 mm.

5l, circled and asterisk points, respectively). The P-R image constructed using the $F_{\beta 2}$ threshold value (Fig 5b') showed that the majority of the segmentation matched the GT (green), whilst the

very tips of the FEVs were slightly clipped (shown as F_N in red), and the vein widths slightly over-estimated with some F_P (shown in blue).

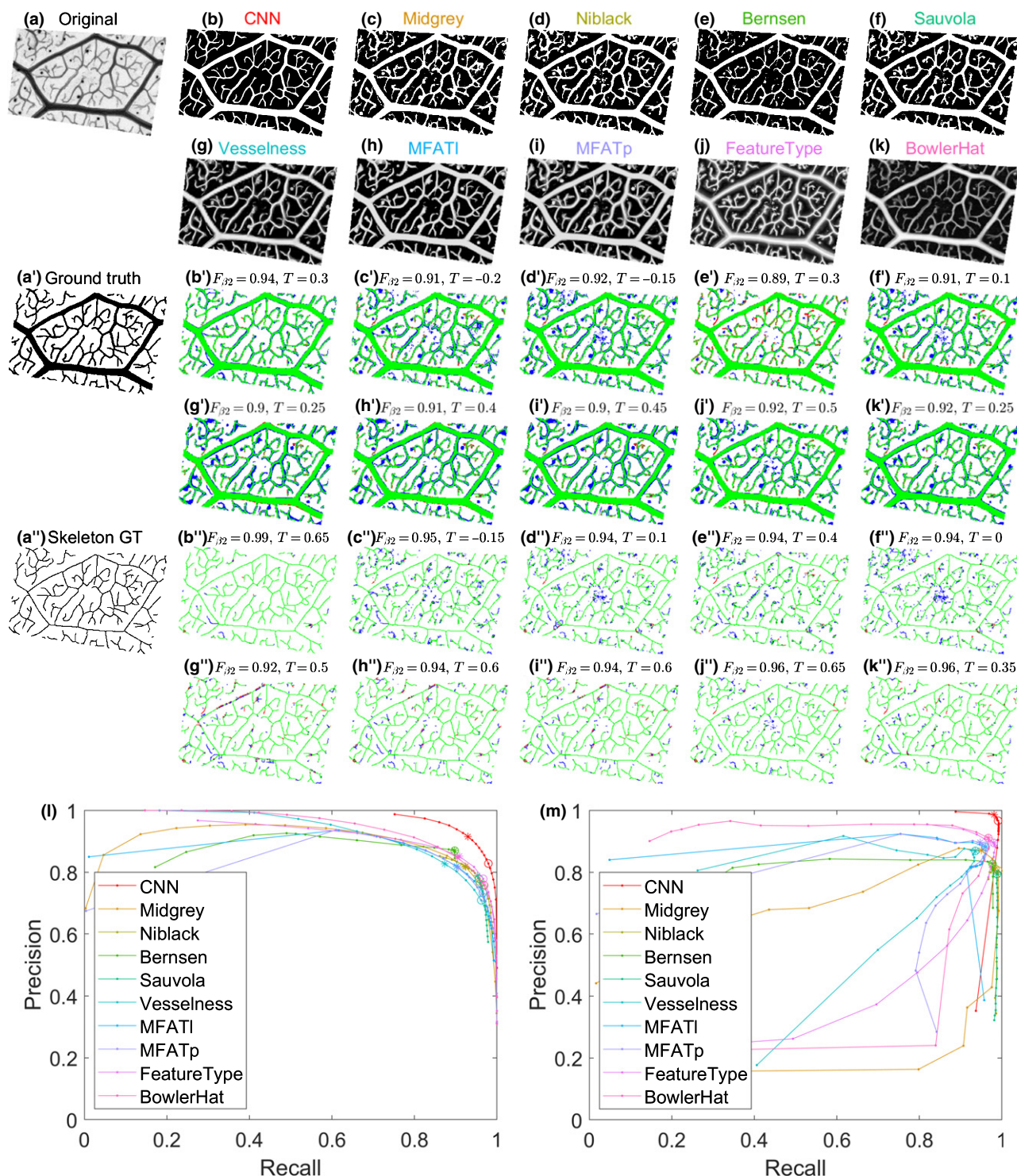


Fig. 5 Comparison of different network enhancement and segmentation methods. A small region-of-interest (ROI) from the full-size image (a), was processed using different enhancement algorithms including the trained convolutional neural network (CNN) (b); local adaptive thresholding using Midgrey (c), Niblack (d), Bernsen (e) or Sauvola (f) algorithms; Hessian-based vessel enhancement using Vesselness (g), MFAT_h (h) and MFAT_p (i); Phase-congruency Feature Type (j), and Bowler Hat morphological filtering (k). The performance of each method was evaluated against a manually defined FWB-GT image (a') or the resultant single-pixel skeleton (a''). In each case, true positives (T_P) are coded in green, false negatives (F_N) in red, and false positives (F_P) in blue for the FWB image (b'–k'), and the single-pixel skeleton (b'–k'). Precision-Recall (P-R) plots illustrate the performance of each method as the threshold is varied for the full-width comparison (l) and following skeletonisation (m). For each algorithm the optimum threshold was selected using the maximum value of the Dice Similarity Coefficient (F_1 , asterisk), or $F_{\beta 2}$ statistic (open circle) from the P-R plots. Values are shown above the corresponding image for the $F_{\beta 2}$ statistic, and the optimal threshold value T . The CNN gave the best performance for both the FWB image ($F_{\beta 2} = 0.943$) and the skeleton ($F_{\beta 2} = 0.99$) comparison. In addition, the P-R plots showed the CNN method performed well over a wide range of segmentation thresholds, whilst the other algorithms were extremely sensitive to the precise threshold used.

Ensemble CNN performance was affected by the contrast range present in the image initially (Fig. S3), but this could be completely recovered using CLAHE to enhance contrast (Fig. S4). Furthermore, performance was stable within a wide range of CLAHE settings (Fig. S5), suggesting that most critical feature is to cover the full contrast range, rather than the fine tuning the local adaptive features. Increasing levels of blur degraded CNN performance (Fig. S6), but only at levels that would give unacceptable out-of-focus in the original image. Possibly the strongest effect on CNN performance was the pixel resolution of the original image, where fine veins were progressively lost as the resolution decreased (Fig. S7). Nevertheless, performance was almost completely recovered by simple image interpolation back to the standard pixel size (Fig. S8).

CNNs performed better than other network extraction algorithms

P-R curves for the other enhancement methods were worse than the CNN result (Fig. 5l). All the local adaptive thresholding approaches tested (Midgrey, Niblack, Bernsen and Sauvola, Fig. 5c–f, respectively) captured much of the main vein structures, but missed or fragmented fine veins (Fig. 5c'–f', red), and included some noise or nonvein structures (Fig. 5c'–f', blue). The Hessian-based techniques ('Vesselness', MFAT_λ and MFAT_p; Fig. 5g–i, respectively) showed slightly greater selectivity for the veins over background compared to adaptive thresholding methods, but had errors in the width estimate (Fig. 5g'–i'). The 'Vesselness' retained more of the intensity information from the original image and tended to give low intensities at junctions (Fig. 5g.g'), whilst MFAT_λ and MFAT_p gave higher contrast and better resolution of junctions (Fig. 5h,i, h',i'). The phase-congruency Feature Type enhancement gave high contrast and even intensity for all veins, irrespective of their size, and captured the topology quite well (Fig. 5j), but gave numerous small errors along the boundary of the binarised veins, and included discrete round structures that we infer to be glands (Fig. 5j'). The Bowler Hat enhancement (Fig. 5k) improved the selection of veins over nonveins, but was still dominated by the original intensity information, making subsequent threshold selection difficult (Fig. 5k').

As several of these enhancement methods were originally designed to facilitate extraction of the skeleton, rather than FWB image, we also compared their performance against the GT skeleton (Fig. 5a''–k''). The performance of the CNN was even better in this comparison ($F_1 = 0.98$, $F_{\beta 2} = 0.99$), and was robust to threshold selection (Fig. 5m), with the optimal threshold across all species of 0.534 ± 0.202 SD for F_1 , 0.378 ± 0.197 SD for $F_{\beta 2}$. The other methods lagged behind the CNN result, but typically showed better P-R performance for the skeleton compared with the full width. However, they also showed much greater sensitivity to threshold selection, with very rapid fall-off in performance moving away from the optimal threshold (Fig. 5m). Without a GT for each image, such automatic optimisation of the threshold value for any of the algorithms would not be possible, so methods that are highly sensitive to the threshold value are

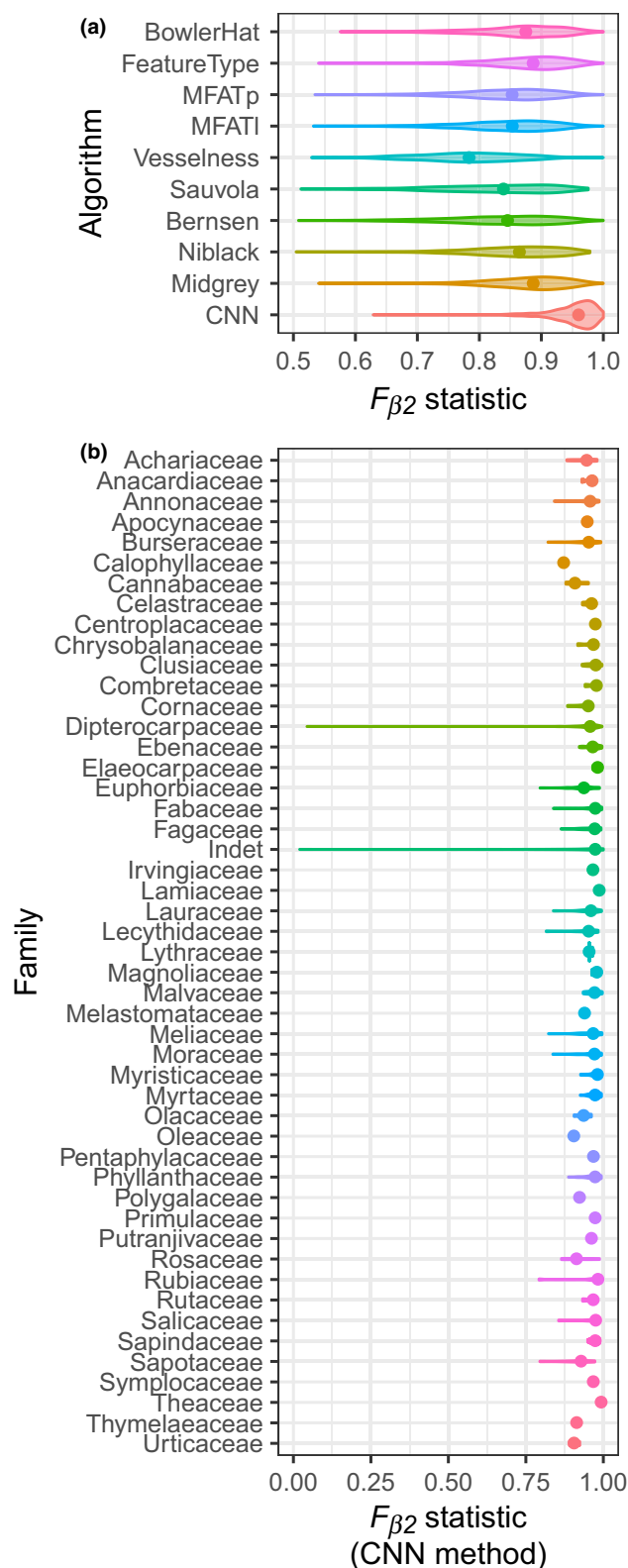


Fig. 6 Relative convolutional neural network (CNN) performance for pixel-based PR metrics. (a) The $F_{\beta 2}$ statistic from P-R analysis against FWB-GT tracings was measured using different enhancement algorithms for all 727 specimens using the optimised threshold selection for each method to ensure the best possible performance. (b) The ensemble CNN $F_{\beta 2}$ was also compared across plant families with different network architectures.

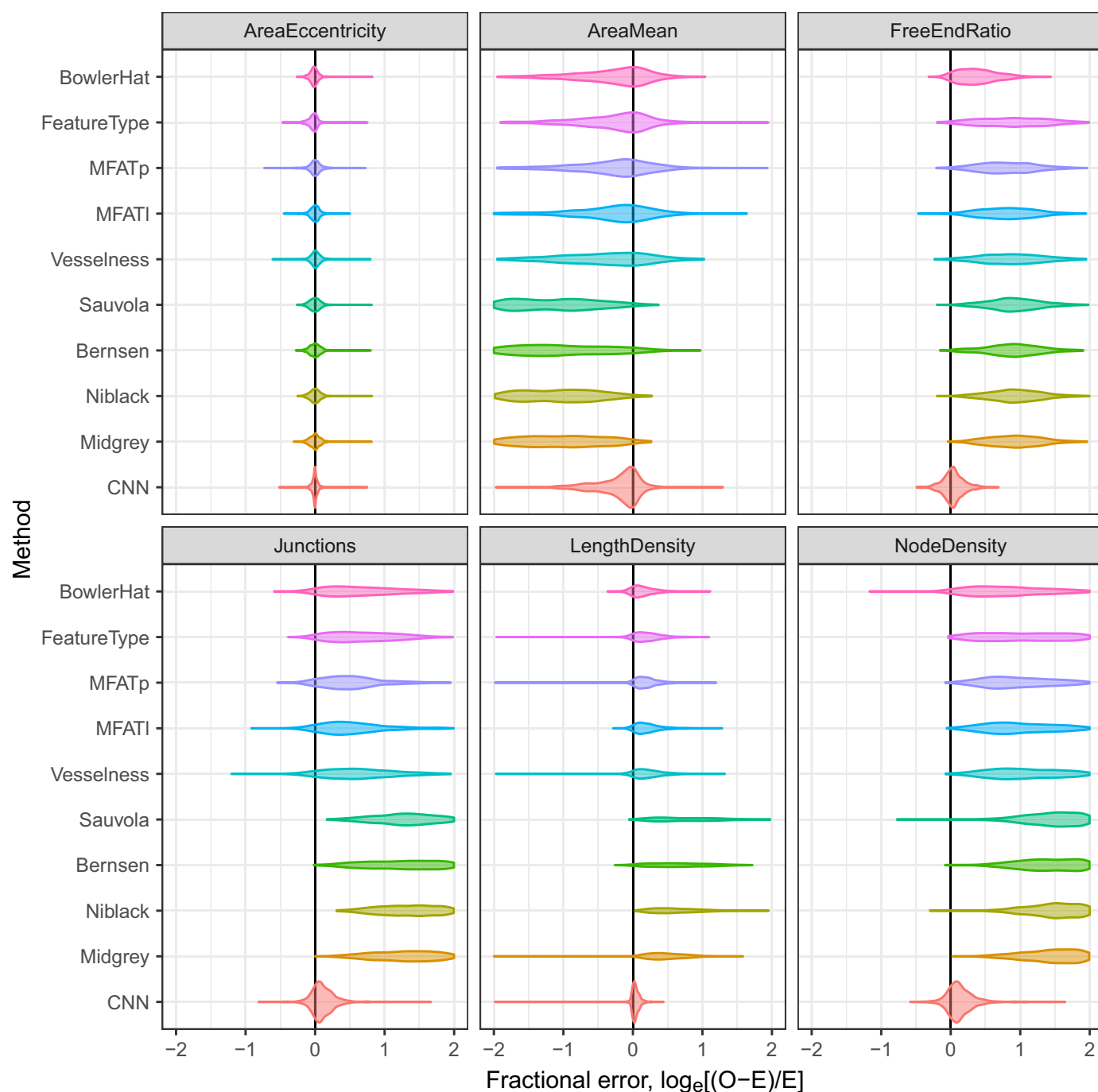


Fig. 7 Relative convolutional neural network (CNN) performance for network metrics. The performance of each network enhancement and segmentation method was compared for a set of network metrics, and presented as the log fractional error ($\log_e[(O-E)/E]$, where O is the observed value and E is the expected value) from the GT skeleton.

likely to perform much worse when used for typical datasets that lack GTs.

When the above procedure was applied to all leaves in the dataset, the F_{p2} statistic values for the CNN were higher than for all other algorithms: 0.945 vs 0.837, $t = -36.463$, $df = 1062.7$, $P < 10^{-16}$ (Fig. 6a). Thus, CNN enhancement gave accurate and robust segmentation of these leaf vein networks and outperformed the currently available network enhancement and

segmentation methods tested. Performance was also consistent across plant families, and by inference, across diverse network architectures (Fig. 6b).

As the best PR performance based on pixel classification may not equate to the optimal network extraction, we also examined the impact of varying thresholds on a set of basic network metrics including *Area eccentricity* (Fig. 7a), *Area mean* (Fig. 7b), *FEV ratio* (Fig. 7c), *No. junctions* (Fig. 7d), *Length density* (Fig. 7e),

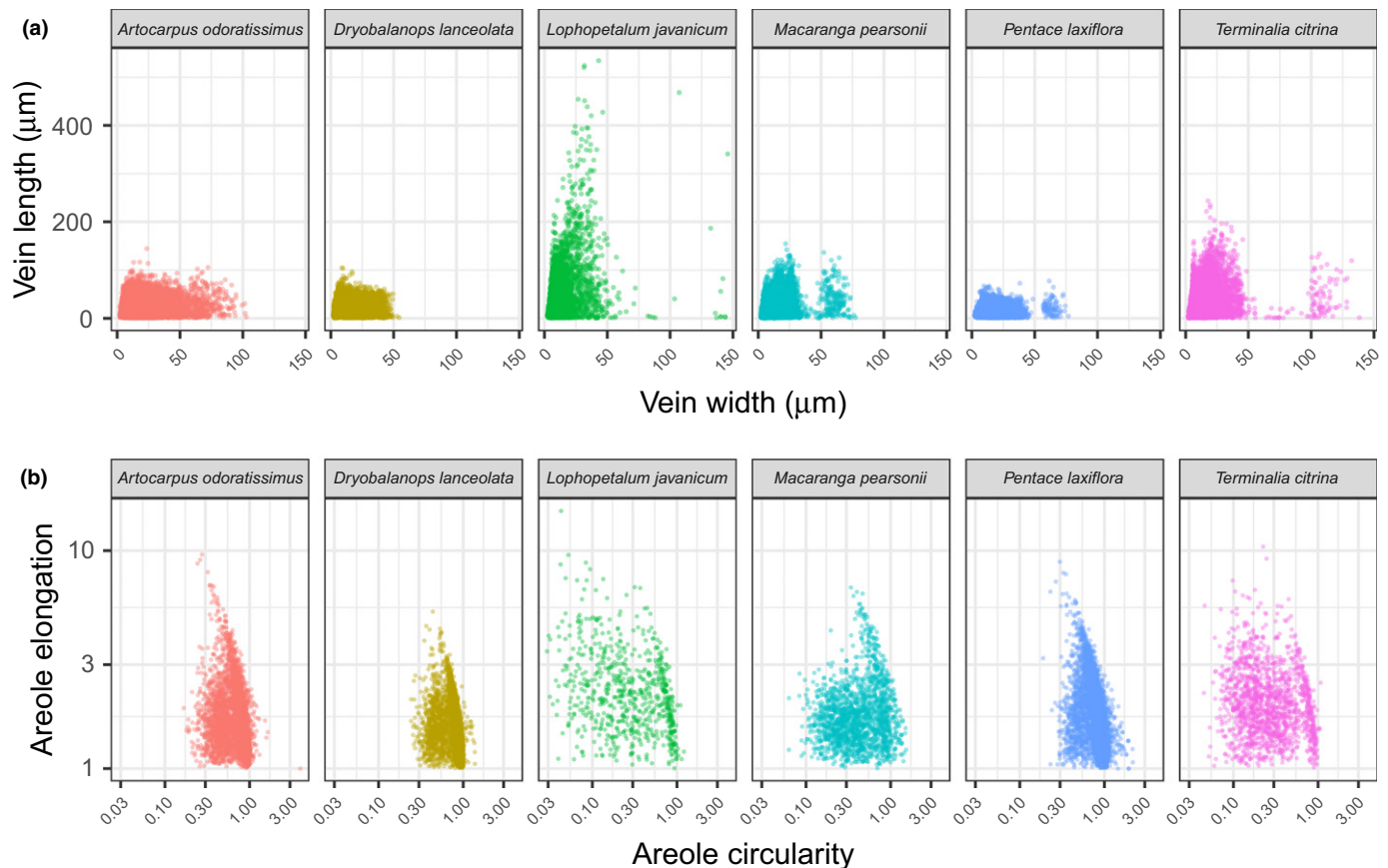


Fig. 8 Example single-scale metrics for leaf venation networks. (a) Bivariate relationship between vein width and vein length for the six species shown in Fig. 4. Arrows indicate the occurrence of discrete minor vein size classes in some species. (b) Bivariate relationship between areole circularity and areole elongation. Points represent each unique vein segment (a) or areole (b).

and *Node density* (Fig. 7f). For all metrics, the CNN approach showed the lowest fractional error from the GT skeleton, followed by the BowlerHat, Phase-congruency and Hessian-based techniques. The adaptive threshold methods performed very poorly, with fractional errors 1–2 \log_e units adrift.

Venation comparisons using single-scale metrics

A range of metrics was calculated following CNN network extraction from the full-size images for the veins, areoles and polygonal regions (Tables S2–S5). Each metric had distribution characteristic of the species. For example, the relationship between vein width and length (Fig. 8a) or between loop circularity and elongation (Fig. 8b), showed markedly different patterns between the six species illustrated. For example, in some species (*A. odoratissimus* and *D. lanceolata*), the vein width showed little separation into distinct vein orders, whilst for the others, discrete vein classes were readily identifiable (Fig. 8a), but with markedly different numbers in each class.

Venation comparisons using HLD and multiscale metrics

Given the limitations of single-scale metrics, we explored analysis using HLD to progressively group regions of the leaf into clusters

at different scales to generate multiscale statistics for each metric (Fig. 3g,h). For example, the spatial scaling of vein density (*VTotLD*) changed as a function of the fusion vein width (*Wid*) (Fig. 9a). At small scale, some species had a much higher vein density (e.g. *P. laxiflora*). However, this reduced rapidly once a specific size class of veins was removed (50 μm in this case; Fig. 8a). In some other species (e.g. *T. citrina*), as the scale increased (or equivalently as the size class of the intervening vein increased), the vein density was maintained to a much greater extent.

The different vein sizes and overall vein density reflected different partitioning of the leaf lamina (Fig. 9b). Thus, the median loop area was an order of magnitude larger for *L. javanicum*, *M. pearsonii* and *T. citrina*, relative to other species at small scale, indicating a lower areole density in these species. By contrast, at intermediate scales, other species like *D. lanceolata* or *P. laxiflora* transitioned rapidly to low levels of looping from a dense initial value, whilst *A. odoratissimus* maintained a more consistent response across scales, reflecting the relative investment in loops and redundant flow pathways at different spatial scales.

The level of redundancy was also reflected in the amount of network branching across spatial scales measured by the minimum spanning tree (MST) ratio (Fig. 9c). The MST ratio increased with vein size class fastest for *D. lanceolata* and *P. laxiflora*, and slowest for *L. javanicum*. In all cases there were

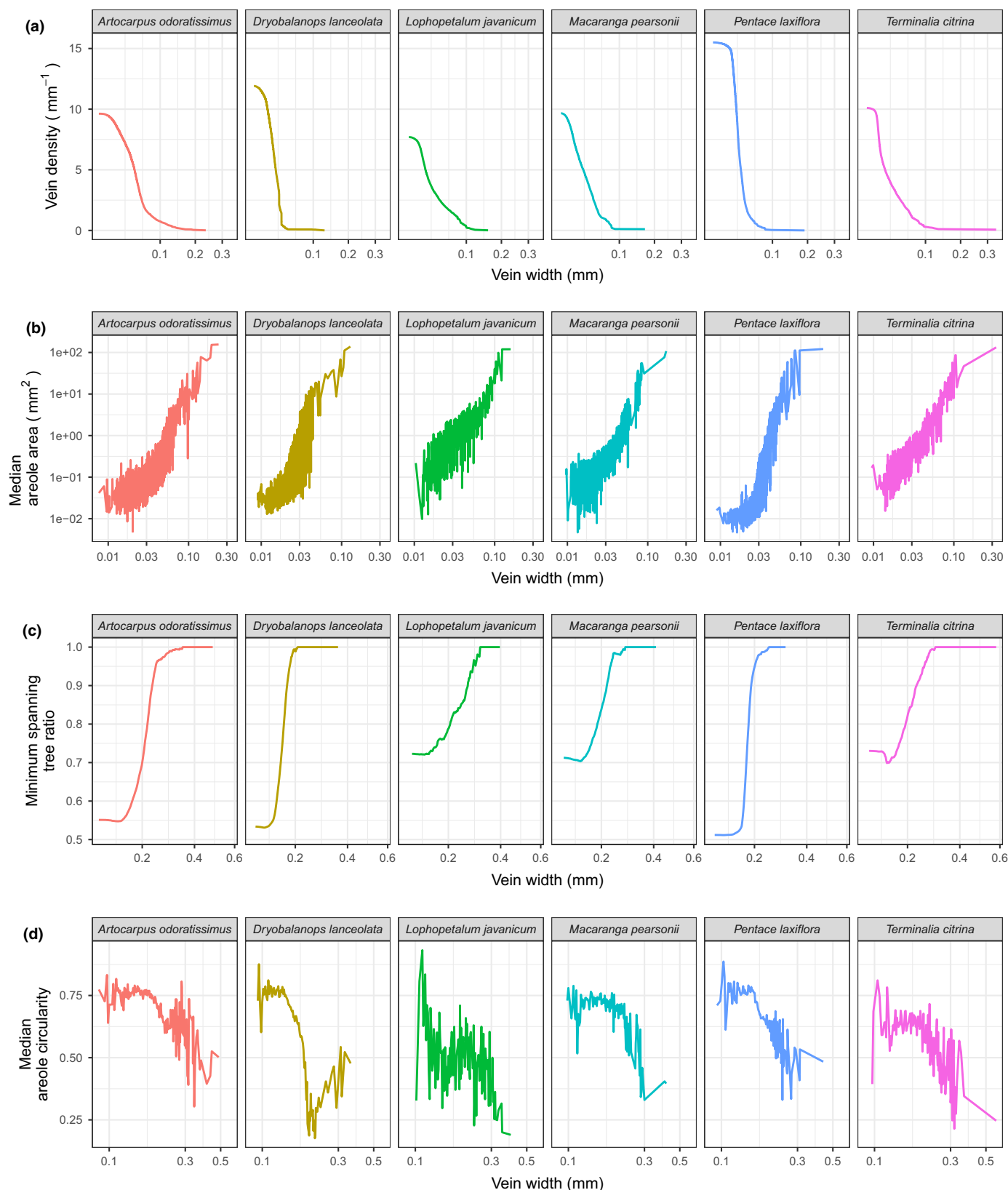


Fig. 9 Example multiscale metrics for leaf venation networks. (a) Changes in vein density with increasing vein width for the six species shown in Fig. 4 following successive vein removal during hierarchical loop decomposition (HLD). (b) Changes in median areole area with increasing vein width removal. (c) Changes in minimum spanning tree ratio with increasing vein width removal. (d) Changes in median areole circularity with increasing areole area following fusion.

shifts in gradient corresponding to scale transitions between veins. *T. citrina* and *M. pearsonii* also showed nonmonotonic changes in the MST ratio, consistent with early pruning of FEV and tree-like branches to give a minimum at an intermediate spatial scale. Thus, this multiscale statistic indicates how investment in veins of different sizes yields different branching organisation.

The space-filling geometry of the network also changed across spatial scales (Fig. 9d). Thus, the median areole circularity decreased with vein size class in all species, but at different rates. Areole circularity was similar between species at small spatial scales, but at larger spatial scales. *D. lanceolata* had the least circular areoles due to fusing regions becoming more elongated with a more parallel vein architecture, whilst *A. odoratissimus* maintained the most circular areoles.

Discussion

CNNs improve network segmentation accuracy and robustness

Ensemble CNNs provide an alternative approach for leaf vein network segmentation with high accuracy, as judged both by P-R analysis against GT tracings using pixel classification and a set of network measures. Ensemble CNNs also provided consistently better segmentation of leaf vein networks from field-collected leaves across many different species and leaf vein architectures than other currently available network extraction algorithms. In addition, the CNN probability map produced smooth, full-width segmentation of veins, even in the presence of image artefacts such as differential clearing, air bubbles or trichomes. This in turn led to improved robustness, precision and accuracy in graph extraction, vein width estimation and subsequent quantitative metrics.

We noted that this deep learning approach contrasted with previous neural network approaches for leaf architecture (Grinblat *et al.*, 2016), which used simpler network architectures and more limited training data, with a goal to develop a classification system rather than extract the venation network itself.

Ensemble CNNs allowed rapid characterisation of minor vein networks at multiple spatial scales with minimal manual intervention, removing a major bottleneck of manual tracing. For comparison, hand-tracing the equivalent area predicted by the CNNs would have required >36 000 person-hours to complete. Nevertheless, the results presented here are truncated by the leaf sample extents, which prevented exploration of larger scales than $c. 10 \times 10 \text{ mm}^2$ areas.

The trained ensemble CNNs used input images with high pixel resolution (1.68 μm per pixel) and contrast adjusted to cover the full dynamic range following CLAHE, and could be used without modification to analyse other leaf images of comparable contrast and resolution. Other leaf vein images at lower spatial resolution could still be segmented well using up-sampling with interpolation, to ensure that the smallest veins are $c. 5$ pixels wide. A typical example is shown in Fig. S9 for a leaf originally collected at 6.7 μm per pixel using the LeafVeinCNN GUI (Fig. S2). Nevertheless, the performance of the ensemble CNNs was

driven by the quality of the GT and the input images. For example, the LeafVeinCNN may currently have limited performance on lower-resolution images that are typical for digitised historic cleared leaf collections, such as the Smithsonian/ Wolfe collection (Lobet *et al.*, 2013; Das *et al.*, 2014), without additional training, or using approaches such as generative adversarial networks (GANs) to refine segmentations and avoid unlikely predictions that have large blotches or disconnected areas.

We are expanding the underlying CNN models through refinements in network architecture and training with additional GT data, including for whole leaves. Readers should visit <https://doi.org/10.5281/zenodo.4007731> to obtain updated trained CNNs to use with the program as these become available.

Network analysis across multiple scales

Most studies have primarily focused on descriptors of the network at single scales, for example the density of major and minor veins (Sack *et al.*, 2012), or metrics for the shape of areoles (Blonder *et al.*, 2011; Price & Weitz, 2014). At multiple scales, other studies have suggested that the hierarchical nesting of veins and areoles may distinguish species or have functional significance (Katifori *et al.*, 2010; Katifori & Magnasco, 2012; Ronellenfitsch *et al.*, 2015; Brodribb *et al.*, 2016; Fiorin *et al.*, 2016). Existing systematic manuals recognise a broad diversity of leaf venation forms and associated terminology that are not well captured by extant quantitative metrics (Ellis *et al.*, 2009). Additionally, prior efforts to leverage HLD yielded a set of statistics, such as Strahler number, derived originally from analysis of river basins, which are quite difficult to interpret in terms of biological processes (Pelletier & Turcotte, 2000; Price *et al.*, 2011; Ronellenfitsch *et al.*, 2015; Lasser & Katifori, 2017). Here, we also used HLD to determine scaling relationships, but retained a set of more conventional metrics to compare network architecture across scales.

The multiscale statistics based on HLD are likely to provide additional insights into how network architecture influences transport, support, defence, embolism resistance, and resilience to herbivory (Sack *et al.*, 2008; Katifori *et al.*, 2010; Brodribb *et al.*, 2016; Blonder *et al.*, 2018). Many of the metrics we explored here are empirically coordinated with different leaf functions. Using venation networks extracted from this dataset (Blonder *et al.*, 2020), we recently showed that the MST ratio, and circularity ratio are linked to mechanical strength, whilst the elongation ratio and circularity ratio are linked to secondary defences. Further analyses of multiscale statistics may enable a more quantitative and richer description of network architecture that supplements the utility of qualitative classifications.

Acknowledgements

We thank David Ford for assistance with the GPU cluster. This work was supported by the Human Frontier Science Program (RGP0053/2012, MDF), the Leverhulme Trust (RPG-2015-437, MDF), the UK Natural Environment Research Council (NERC) (NE/M019160/1, BB), the US National Science Foundation (DEB-2025282, DEB-1840209, BB), and an ERC


Advanced Investigator Award (GEM-TRAIT 321131, YM). The field work was funded by the UK NERC BALI consortium (NE/K016253/1). The SAFE Project was funded by the Sime Darby Foundation and the UK NERC.

Author contributions

HX developed the CNN training algorithms. MDF developed the network extraction and analysis algorithms, and GUI interface. MJ chemically cleared and imaged the leaves. BB carried out the fieldwork and the statistical analysis. YM and BB secured funding for the field work and laboratory work, and computing infrastructure. HX, MDF and BB contributed equally to the manuscript.

ORCID

Benjamin Blonder  <https://orcid.org/0000-0002-5061-2385>

Mark Fricker  <https://orcid.org/0000-0002-8942-6897>

Yadvinder Malhi  <https://orcid.org/0000-0002-3503-4783>

References

- Alhasson HF, Alharbi SS, Obara B. 2019. 2D and 3D vascular structures enhancement via multiscale fractional anisotropy tensor. *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*. Berlin, Germany: Springer, 365–374.
- Berens P. 2009. CircStat: a MATLAB toolbox for circular statistics. *Journal of Statistical Software* 31: 1–21.
- Bernsen J. 1986. Dynamic thresholding of gray-level images. *Proceedings of the Eighth International Conference on Pattern Recognition*, Paris, France, 1251–1255.
- Blonder B, Both S, Jodra M, Majalap N, Burslem D, Teh YA, Malhi Y. 2019. Leaf venation networks of Bornean trees: images and hand-traced segmentations. *Ecology* 100: e02844.
- Blonder B, Both S, Jodra M, Xu H, Fricker MD, Matos I, Majalap-Lee N, Burslem D, Teh YA, Malhi Y. 2020. Linking functional traits to multiscale statistics of leaf venation networks. *New Phytologist* 228: 1796–1810.
- Blonder B, Carlo F, Moore J, Rivers M, Enquist BJ. 2012. X-ray imaging of leaf venation networks. *New Phytologist* 196: 1274–1282.
- Blonder B, Enquist BJ. 2014. Inferring climate from angiosperm leaf venation networks. *New Phytologist* 204: 116–126.
- Blonder B, Royer DL, Johnson KR, Miller I, Enquist BJ. 2014. Plant ecological strategies shift across the cretaceous-paleogene boundary. *PLoS Biology* 12: e1001949.
- Blonder B, Salinas N, Patrick Bentley L, Shenkin A, Chambi Porroa PO, Valdez Teixeira Y, Boza Espinoza TE, Goldsmith GR, Enrico L, Martin R. 2018. Structural and defensive roles of angiosperm leaf venation network reticulation across an Andes-Amazon elevation gradient. *Journal of Ecology* 106: 1683–1699.
- Blonder B, Violle C, Bentley LP, Enquist BJ. 2011. Venation networks and the origin of the leaf economics spectrum. *Ecology Letters* 14: 91–100.
- Bohn S, Andreotti B, Douady S, Munzinger J, Couder Y. 2002. Constitutive property of the local organization of leaf venation networks. *Physical Review E* 65: 061914.
- Both S, Riutta T, Paine CT, Elias DM, Cruz R, Jain A, Johnson D, Kritzer UH, Kuntz M, Majalap-Lee N. 2019. Logging and soil nutrients independently explain plant trait expression in tropical forests. *New Phytologist* 221: 1853–1865.
- Boyce CK, Brodribb T, Feild TS, Zwieniecki MA. 2009. Angiosperm leaf vein evolution was physiologically and environmentally transformative. *Proceedings of the Royal Society B: Biological Sciences* 276: 1771–1776.
- Bresenham JE. 1965. Algorithm for computer control of a digital plotter. *IBM Systems Journal* 4: 25–30.
- Brodribb 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.
- Brodribb T, Feild TS. 2010. Leaf hydraulic evolution led a surge in leaf photosynthetic capacity during early angiosperm diversification. *Ecology Letters* 13: 175–183.
- Brodribb T, Feild T, Jordan G. 2007. Leaf maximum photosynthetic rate and venation are linked by hydraulics. *Plant Physiology* 144: 1890.
- Brodribb TJ, Feild TS, Sack L. 2010. Viewing leaf structure and evolution from a hydraulic perspective. *Functional Plant Biology* 37: 488–498.
- 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.
- Burt P, Adelson E. 1983. The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications* 31: 532–540.
- Carvalho MR, Losada JM, Niklas KJ. 2018. Phloem networks in leaves. *Current Opinion in Plant Biology* 43: 29–35.
- Carvalho MR, Turgeon R, Owens T, Niklas KJ. 2017. The hydraulic architecture of Ginkgo leaves. *American Journal of Botany* 104: 1285–1298.
- Das A, Bucksch A, Price CA, Weitz JS. 2014. ClearedLeavesDB: an online database of cleared plant leaf images. *Plant Methods* 10: 8.
- de Boer HJ, Eppinga MB, Wassen MJ, Dekker SC. 2012. A critical transition in leaf evolution facilitated the Cretaceous angiosperm revolution. *Nature Communications* 3: 1221.
- 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.
- Dirnberger M, Kehl T, Neumann A. 2015. NEFI: network extraction from images. *Scientific Reports* 5: 15669.
- Dodds PS. 2010. Optimal form of branching supply and collection networks. *Physical Review Letters* 104: 048702.
- Ellis B, Daly DC, Hickey LJ, Johnson KR, Mitchell JD, Wilf P, Wing SL. 2009. *Manual of leaf architecture*. Ithaca, NY, USA: Cornell University Press.
- Fiorini L, Brodribb TJ, Anfodillo T. 2016. Transport efficiency through uniformity: organization of veins and stomata in angiosperm leaves. *New Phytologist* 209: 216–227.
- Frangi A, Niessen W, Vincken K, Viergever M. 1998. Multiscale vessel enhancement filtering. In: Wells W, Colchester A, Delp S, eds. *Medical image computing and computer-assisted intervention*. Berlin/Heidelberg, Germany: Springer, 130–137.
- Fricker MD, Akita D, Heaton LLM, Jones N, Obara B, Nakagaki T. 2017. Automated analysis of *Physarum* network structure and dynamics. *Journal of Physics D* 50: 254005.
- 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.
- Glorot X, Bengio Y. 2010. Understanding the difficulty of training deep feedforward neural networks. *Proceedings of Machine Learning Research: PMLR* 9: 249–256.
- Glorot X, Bordes A, Bengio Y. 2011. Deep sparse rectifier neural networks. *Proceedings of Machine Learning Research: PMLR* 15: 315–323.
- Green WA, Little SA, Price CA, Wing SL, Smith SY, Kotrc B, Doria G. 2014. Reading the leaves: a comparison of leaf rank and automated areole measurement for quantifying aspects of leaf venation. *Applications in Plant Sciences* 2: apps.1400006.
- 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.
- Hansen LK, Salamon P. 1990. Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12: 993–1001.
- Hickey LJ. 1979. A revised classification of the architecture of dicotyledonous leaves. In: Metcalfe CR, Chalk L eds. *Anatomy of the Dicotyledons. vol. 1, systematic anatomy of the leaf and stem*. Oxford, UK: Clarendon Press, 25–39.

- Ioffe S, Szegedy C. 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift. In Francis B, David B, eds. *Proceedings of the 32nd International Conference on Machine Learning. Proceedings of Machine Learning Research, PMLR* 37: 448–456.
- John GP, Scoffoni C, Buckley TN, Villar R, Poorter H, Sack L. 2017. The anatomical and compositional basis of leaf mass per area. *Ecology Letters* 20: 412–425.
- Kang J, Dengler N. 2004. Vein pattern development in adult leaves of *Arabidopsis thaliana*. *International Journal of Plant Sciences* 165: 231–242.
- Katifori E. 2018. The transport network of a leaf. *Comptes Rendus Physique* 19: 244–252.
- Katifori E, Magnasco MO. 2012. Quantifying loopy network architectures. *PLoS ONE* 7: e37994.
- Katifori E, Szollosi GJ, Magnasco MO. 2010. Damage and fluctuations induce loops in optimal transport networks. *Physical Review Letters* 104: 048704.
- Kovesi PD. 1999. Image features from phase congruency. *Videre* 1: 1–26.
- Kovesi PD. 2000. *MATLAB and Octave functions for computer vision and image processing*. [WWW document] URL <http://www.csse.uwa.edu.au/~pk/research/htmlabfn/> [accessed 28 October 2020].
- Krizhevsky A, Sutskever I, Hinton GE. 2012. Imagenet classification with deep convolutional neural networks. In: Pereira, F, Burges, CJC, Bottou, L, Weinberger, KQ, eds. *Advances in neural information processing systems, vol. 25*. Red Hook: Curran Associates, 1097–1105.
- Larese MG, Námias R, Cravioito RM, Arango MR, Gallo C, Granitto PM. 2014. Automatic classification of legumes using leaf vein image features. *Pattern Recognition* 47: 158–168.
- 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.
- LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. *Nature* 521: 436–444.
- Lobet G, Draye X, Périlleux C. 2013. An online database for plant image analysis software tools. *Plant Methods* 9: 1–8.
- Long J, Shelhamer E, Darrell T. 2015. *Fully convolutional networks for semantic segmentation*. Proceedings of the IEEE conference on computer vision and pattern recognition. 3431–3440.
- Lopez-Molina C, De Baets B, Bustince H. 2013. Quantitative error measures for edge detection. *Pattern Recognition* 46: 1125–1139.
- Manze U. 1967. *Die Nervaturdichte der Blätter als Hilfsmittel der Paläoklimatologie*. Köln, Germany: Geologisches Institut der Universität zu Köln.
- McKown AD, Cochard H, Sack L. 2010. Decoding leaf hydraulics with a spatially explicit model: principles of venation architecture and implications for its evolution. *American Naturalist* 175: 447–460.
- Mileyko Y, Edelsbrunner H, Price CA, Weitz JS. 2012. Hierarchical ordering of reticular networks. *PLoS ONE* 7: e36715.
- Mounsef J, Karam L. 2012. Fully automated quantification of leaf venation structure. Proceedings of the 2012 International Conference on Artificial Intelligence, ICAI 2012. 820–825.
- Niblack W. 1985. *An introduction to digital image processing*. Topstyrkkt, Denmark: Strandberg.
- Niklas KJ. 1999. A mechanical perspective on foliage leaf form and function. *New Phytologist* 143: 19–31.
- Obara B, Grau V, Fricker MD. 2012. A bioimage informatics approach to automatically extract complex fungal networks. *Bioinformatics* 28: 2374–2381.
- Pain C, Kriebbaum V, Kittelmann M, Hawes C, Fricker M. 2019. Quantitative analysis of plant ER architecture and dynamics. *Nature Communications* 10: 984.
- Parsons-Wingerter P, Vickerman MB, Paul A-L, Ferl RJ. 2014. Mapping by VESGEN of leaf venation patterning in Arabidopsis with bioinformatic dimensions of gene expression. *Gravitational and Space Research* 2: 68–81.
- Pelletier JD, Turcotte DL. 2000. Shapes of river networks and leaves: are they statistically similar? *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 355: 307–311.
- Pérez-Harguindeguy N, Díaz S, Garnier E, Lavorel S, Poorter H, Jaureguiberry P, Bret-Harte M, Cornwell W, Craine J, Gurvich D. 2013. New handbook for standardised measurement of plant functional traits worldwide. *Australian Journal of Botany* 61: 167–234.
- Price CA, Knox S-JC, Brodribb TJ. 2014. The Influence of branch order on optimal leaf vein geometries: murray's law and area preserving branching. *PLoS ONE* 8: e85420.
- 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.
- Price CA, Weitz JS. 2014. Costs and benefits of reticulate leaf venation. *BMC Plant Biology* 14: 234.
- Price CA, Wing S, Weitz JS. 2012. Scaling and structure of dicotyledonous leaf venation networks. *Ecology Letters* 15: 87–95.
- Riutta T, Malhi Y, Kho LK, Mathews TR, Huaraca Huasco W, Khoo M, Tan S, Turner E, Reynolds G, Both S. 2018. Logging disturbance shifts net primary productivity and its allocation in Bornean tropical forests. *Global Change Biology* 24: 2913–2928.
- Rolland-Lagan A-G, Amin M, Pakulska M. 2009. Quantifying leaf venation patterns: two-dimensional maps. *The Plant Journal* 57: 195–205.
- Ronellenfisch 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.
- Ronneberger O, Fischer P, Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation. In: In Navab N, Hornegger J, Wells WM, Frangi AF, eds. *Medical image computing and computer-assisted intervention – MICCAI 2015*. Cham, Switzerland: Springer International, 234–241.
- Rother C, Kolmogorov V, Blake A. 2004. Grabcut: Interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics (TOG)* 23: 309–314.
- Roth-Nebelsick A, Uhl D, Mosbrugger V, Kerp H. 2001. Evolution and function of leaf venation architecture: a review. *Annals of Botany* 87: 553–566.
- Sack L, Dietrich EM, Streeter CM, Sánchez-Gómez D, Holbrook NM. 2008. Leaf palmate venation and vascular redundancy confer tolerance of hydraulic disruption. *Proceedings of the National Academy of Sciences, USA* 105: 1567–1572.
- Sack L, Frole K. 2006. Leaf structural diversity is related to hydraulic capacity in tropical rain forest trees. *Ecology* 87: 483–491.
- 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, Scoffoni C, McKown AD, Frole K, Rawls M, Havran JC, Tran H, Tran T. 2012. Developmentally based scaling of leaf venation architecture explains global ecological patterns. *Nature Communications* 3: 837.
- Saito T, Rehmsmeier M. 2015. The Precision-Recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE* 10: e0118432.
- Sauvola J, Pietikäinen M. 2000. Adaptive document image binarization. *Pattern Recognition* 33: 225–236.
- Sazak Ç, Nelson CJ, Obara B. 2019. The multiscale bowler-hat transform for blood vessel enhancement in retinal images. *Pattern Recognition* 88: 739–750.
- 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: 7.
- Sharon E, Sahaf M. 2018. The mechanics of leaf growth on large scales. In: Geitmann A, Gril J, eds. *Plant biomechanics: from structure to function at multiple scales*. Cham, Switzerland: Springer International, 109–126.
- Simonyan K, Zisserman A. 2014. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556v6.
- Sollich P, Krogh A. 1996. Learning with ensembles: How overfitting can be useful. *Advances in Neural Information Processing Systems* 8: 190–196.
- Trivett ML, Pigg KB. 1996. A survey of reticulate venation among fossil and living land plants. In: Taylor DW, Hickey LJ, eds. *Flowering plant origin, evolution & phylogeny*. Boston, MA, USA: Springer, 8–31.
- Uhl D, Mosbrugger V. 1999. Leaf venation density as a climate and environmental proxy: a critical review and new data. *Palaeogeography, Palaeoclimatology, Palaeoecology* 149: 15–26.
- Werbos P. 1974. *Beyond regression: new tools for prediction and analysis in the behavioral sciences*. Cambridge, MA, USA: Harvard University.
- Wing SL. 1992. High-resolution leaf X-radiography in systematics and paleobotany. *American Journal of Botany* 79: 1320–1324.

Zhang TY, Suen CY. 1984. A fast parallel algorithm for thinning digital patterns. *Communications of the ACM* 27: 236–239.

Zuiderveld K. 1994. Contrast limited adaptive histogram equalization. *Graphics Gems* 474–485.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Schematic illustration of the network metrics extracted.

Fig. S2 The LeafVeinCNN graphical user interface (GUI).

Fig. S3 Sensitivity analysis for varying levels of contrast, with no correction.

Fig. S4 Sensitivity analysis for varying levels of contrast followed by CLAHE.

Fig. S5 Sensitivity analysis for varying levels of contrast-limited adaptive histogram equalisation (CLAHE).

Fig. S6 Effect of increasing blur on ensemble CNN performance.

Fig. S7 Effect of image resolution on ensemble CNN performance.

Fig. S8 Effect of up-sampling to correct for low resolution input images.

Fig. S9 Application of LeafVeinCNN to an intact leaf.

Table S1 Vein metrics.

Table S2 Node metrics.

Table S3 Areole metrics.

Table S4 Polygonal region metrics.

Table S5 Summary statistics for vein, node, areole and polygon metrics.

Table S6 HLD metrics.

Table S7 Image enhancement processing parameters.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.