Quantitative Plant Biology

cambridge.org/qpb

Pattern recognition of stomatal distribution: A network automata approach to identify plasticity under environmental stress

Jeaneth Machicao[1], Humberto Antunes de Almeida Filho[2], Kallil Zielinski[3] and Odemir Martinez Bruno[2]

[1]Computer Engineering Department, Polytechnic School, Universidade de São Paulo, Brazil; [2]Instituto de Física de São Carlos, Universidade de São Paulo, Brazil; [3]Federal University of Technology of Paraná, Brazil

Original Research Article

Cite this article: J. Machicao et al. Pattern recognition of stomatal distribution: A network automata approach to identify plasticity under environmental stress. *Quantitative Plant Biology*, **7**:e16, 1–10
https://dx.doi.org/10.1017/qpb.2026.10051

Received: 15 February 2025 UTC
Revised: 30 June 2026 UTC
Accepted: 3 July 2026 UTC

Keywords:
LLNA; network automata; pattern recognition; plasticity; stomata.

Corresponding author:
Jeaneth Machicao;
Email: mj.machicao@gmail.com

Associate Editor:
Prof. Iain Johnston

Abstract

Plants exhibit highly plastic stomatal distributions on the leaf epidermis driven by environmental factors. However, methods to model this complex spatial organization as a global topological system remain limited, as most existing tools focus on segmentation and counting. This study proposes a graph-based classification framework to characterize these distributions. We use stomatal centroids to construct networks and employ Life-Like Network Automata (LLNA) to extract dynamical signatures from them. This approach simulates network emergent behaviours to derive feature vectors based on optimized transition rules. Compared against traditional network metrics and some graph-based algorithms, our results indicate that our proposed method can effectively classify environmental stressors across the three studied species. In these restricted data scenarios, the evaluated models achieved accuracies from 90.00 (±5.65)% to 95.47 (±4.61)%, outperforming the chosen traditional and graph-embedding techniques. We present a pattern recognition framework capable of distinguishing environmental conditions based on the stomatal topological distribution.

1. Introduction

Stomata are pores found on the surface of plant leaves that perform critical roles in gas exchange and transpiration (Cowan & Troughton, 1971; Willmer & Fricker, 1996). Generally, these structures are found on the leaf epidermis but can also be present on stems. Certain phenotypes related to stomata, such as stomatal density (number of stomata per unit area) and stomatal distribution, can adapt according to environmental variations (Hetherington & Woodward, 2003), including fluctuations in CO_2 concentration, humidity and light exposure (Cai & Cao, 2004), driving adaptations in photosynthetic capacity across different climates. This results in significant plasticity in the leaf epidermis, including changes in stomatal density.

Stomatal size and density vary between species and environmental conditions, ranging from 10 to 80 μm in length, while stomatal density can range from 5 to 1,000 per mm^2 of epidermis (Eckerson, 1908). Gas exchange simulations show an inverse correlation between stomatal size and density, indicating a compensatory relationship between CO_2 uptake and water transpiration (Beerling & Woodward, 1997). Additionally, light intensity and quality are critical for stomata plasticity, influencing the plant's ability to adapt the photosynthetic capacity to different lighting conditions (Murchie & Horton, 1997).

Since stomatal distribution phenotypes exhibit high plasticity, novel approaches are required to identify intrinsic characteristics based on environmental fluctuations. Despite the use of geometric morphometry in taxonomic classification and phenotypic plasticity analysis (Viscosi & Cardini, 2011), there is a lack of methods to characterize stomatal distribution as a parameter for quantifying plant adaptation. While other computational approaches have modelled stomata through cellular nonlinear networks to simulate analogue computation (Peak et al., 2023) or via non-steady-state dynamical equations to predict conductance under fluctuating conditions (Liu et al., 2024), these typically focus on physiological flux rather than spatial pattern recognition. Some works propose to evaluate stomatal distribution over the surface of purple abaxial epidermis in extant plants using interstomatic distances as a parameter; however, they do not measure the extent of adaptation under induced environmental changes (de Almeida Filho, 2018).

The geometry of stomatal distribution on leaf surfaces can provide quantitative parameters that quantify how environmental changes affect plants. Plants exhibit phenotypic plasticity by adjusting stomatal spatial distribution as a survival strategy against environmental stressors, which we approach from a network perspective. In this context, stomata represent vertices and their spatial relationships represent edges. Unlike studies aimed at species

© The Author(s), 2026. Published by Cambridge University Press in association with John Innes Centre. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

John Innes Centre
Unlocking Nature's Diversity
CAMBRIDGE UNIVERSITY PRESS

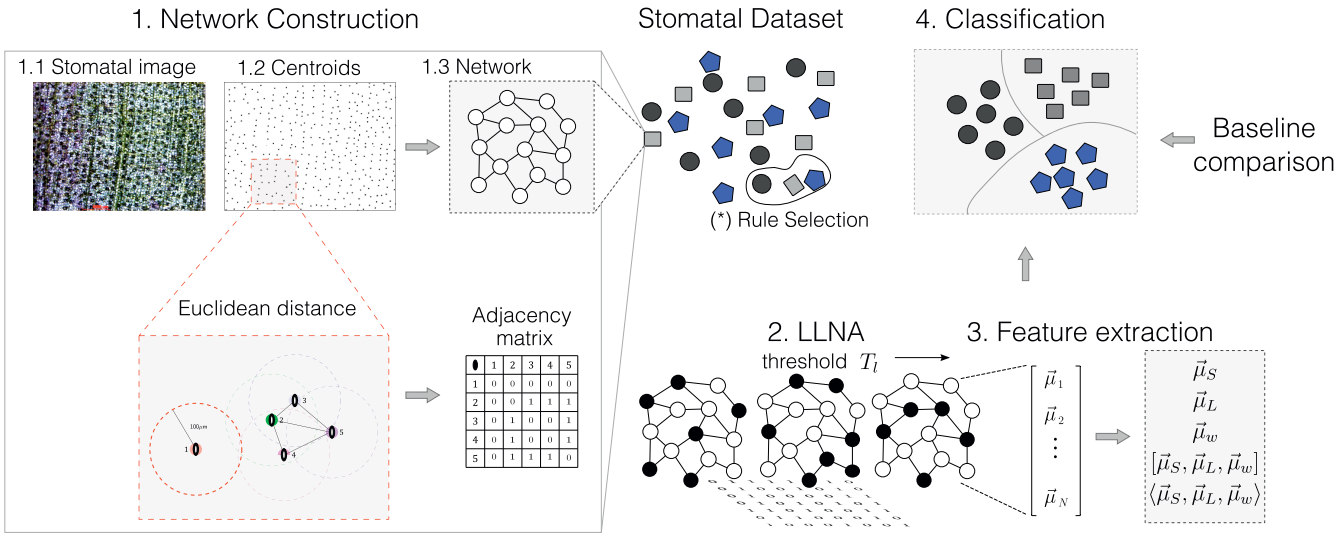


Figure 1. Proposed workflow to use LLNA as a pattern recognition method.

Table 1. Information of the dataset of networks studied

	Stomata dataset			Data split	
	Classes/environmental conditions		No. of images	No. of networks (training/validation)	No. of networks (rule selection)
Ctenanthe	2	L4h, L24h	36	36	–
Trandescantia	3	Environment, L4h, L24h	18	12	6
Callisia	3	Environment, $CO_2(\downarrow)$, $CO_2(\uparrow)$,	15	15	–

identification (da Silva et al., 2016), the goal here is to explore a pattern recognition method capable of separating environmental conditions where plasticity is fundamental.

Pattern recognition using complex networks has achieved success in various tasks (Costa et al., 2007, 2011; Miranda et al., 2018), where network properties serve as topological fingerprints for environmental conditions. Structural measurements, such as clustering coefficient, degree and average path length, are commonly employed to characterize topological properties and have demonstrated reasonable performance (Costa et al., 2007). More recently, graph-based methods have emerged, such as deep topological embedding (DTE) (Scabini et al., 2022), deterministic tourist walk-based (DTWB) (Merenda & Bruno, 2023) and deep graph convolutional neural networks (DGCNNs) (Zhang et al., 2018). However, alternative approaches, such as network automata, have shown interesting properties for complex tasks (Machicao et al., 2018; Miranda et al., 2016;Ribas et al., 2020; Zielinski et al., 2024). By simulating emergent behaviours, these methods provide a dynamical perspective on network evolution that can offer complementary insights when identifying subtle environmental adaptations.

In this study, we utilize pre-identified stomatal centroids to construct graphs and apply the Life-Like Network Automata (LLNA) method (Miranda et al., 2016) to accurately classify environmental conditions to which three plant species were subjected: *Ctenanthe oppenheimiana*, *Tradescantia zebrina* and *Callisia repens*. We adopted this method due to the high performance achieved in other classification tasks, such as authorship attribution (Machicao et al., 2018), and in various extended versions of the LLNA framework (Ribas et al., 2020; Zielinski et al., 2024).

2. Material and methods

2.1. Method workflow

An overview of the pattern recognition method is shown in Figure 1. First, we introduce the network model for constructing networks from stomatal imagery, including a data split that allocates a small portion for LLNA rule selection. Next, we apply the LLNA and its corresponding feature extraction. The metrics derived from the LLNA dynamics are then utilized to classify the environmental conditions of the stomata across various plant species. Finally, we compared these results with baseline methods.

2.2. Images dataset

We used the stomatal image dataset acquired from the Scientific Computing Group of the São Carlos Institute of Physics (IFSC) (Supplementary Material). A summary of the dataset is presented in Table 1.

This dataset contains leaf images obtained by optical microscopy from three plant species: *Ctenanthe oppenheimiana*, *Tradescantia zebrina* and *Callisia repens*. Each image includes corresponding segmented binary images, with the centroids of each stoma manually identified by a specialist. The plants were subjected to one of two different stress conditions: light exposure and injection of CO_2, as detailed in Section 2.3.

2.3. Crop, acclimation and microscopy

- *Ctenanthe oppenheimiana*: Plants, measuring 5–8 cm, were planted in standard Carolina substrate (Carolina soil, Brazil). Samples were acclimated in a controlled growth chamber

(Eletrolab EL212/4G-RS) with continuous 55% relative humidity, a photon flux of 100 µmol / m^2×s and a temperature of 25 °C. Two light/dark photoperiods were tested: 24 h (L24) and 4 h (L4) per day for 60 days, resulting in 18 images per condition. Samples are shown in Figure 2a.

- *Tradescantia zebrina*: Plants were acclimated in the same controlled growth chamber under identical conditions for 68 days. A third group was maintained under uncontrolled conditions in an external environment. A total of six images were obtained for each condition. This collection was previously described by Florindo et al. (2015). Samples are shown in Figure 2b.
- *Callisia repens*: Three experimental groups were used. The first was sealed in a bottle with Carolina soil and subjected to a CO_2 gas stream (1,500 ppm for 6 h/day over 30 days). The second group was treated similarly but without the CO_2 flow. In both bottled groups, relative humidity was maintained between 85% and 90% during the daily 6-h cycle. A third group was collected from a natural, uncontrolled environment in July 2015 (georeferenced at −22.007 S; −47.894 W). Five images were captured for each condition. Samples are shown in Figure 2c.

Regarding imaging specifications, images were captured at different magnifications depending on the species. For *Ctenanthe oppenheimiana*, a 100 × objective lens was used, resulting in a pixel scale of 0.347 µm/px and a resolution of 2,560 × 1,920 pixels. For *Tradescantia zebrina* and *Callisia repens*, images were captured at 15 × magnification, corresponding to 4.65935 µm/px and 1,388 × 1,038 pixels. These specific optical scales were used for all subsequent spatial measurements and pixel-to-micron conversions.

2.4. Stomatic network modelling

Figure 2 exemplifies the steps for the construction of the stomatal network model, as proposed by Backes et al. (2009) and Florindo et al. (2015). In this model, each stoma corresponds to a vertex in the lattice, and the rule for establishing the edges is based on the Euclidean distance between pairs of stomata governed by a threshold T_l.

Each image contains a different number of stomata (vertices) positioned at the centres of circles; thus, the threshold T_l defines the radius of connectivity of a stoma with its nearest neighbours. A distance matrix containing all distances between stomata pairs in each image is obtained through the Euclidean distances between each pair of vertices (W). This matrix is then transformed into several networks represented by the adjacency matrix A_{T_l} depending on the threshold T_l, as given by Equation (1). Specifically, if the distance between a pair of stomata is less than or equal to the threshold, then an edge is established between them:

$$A_{T_l} = \delta_{T_l}(W) = \forall w \in W \begin{cases} a_{ij} = 0 & \text{if } w_{ij} > T_l \\ a_{ij} = 1 & \text{if } w_{ij} \leq T_l. \end{cases} \quad (1)$$

Since the characterization of these networks is carried out through a series of transformations in which T_l is incremented between T_{ini} and T_Q (the initial and final values of the threshold), it allows for a multi-scale analysis of the topology (Backes et al., 2009). Since the centroid coordinates are provided in the real scale of the plant, the pixel-based image coordinates are converted using a specific scale factor for each species (as detailed in Section 2.3). For instance, in *Ctenanthe oppenheimiana*, a factor of 0.347 µm/px was applied. From top to bottom, Figure 2 presents samples of abaxial epidermis images of each species of plant and the corresponding environmental condition.

Figure 2a illustrates the network construction process following Equation (1). On the left side of the figure, the microscopic image of a leaf of *Ctenanthe oppenheimiana* is shown. The red circles represent the centroids of the stomata identified in the segmented binary image. On the right side, four resulting networks are presented, each corresponding to a specific radius threshold T_l. For illustration, various thresholds T_l are shown to demonstrate the topological evolution. Note that as T_l increases, more connections are established between stomata centroids (nodes), leading to higher network connectivity.

Considering that different networks can be obtained, the threshold T_l must be limited according to the size of the image. Note that in the networks of the species *Tradescantia zebrina* and *Callisia repens*, the thresholds are numerically higher because the image samples have lower magnification (15 × vs. 100 ×), which results in a different spatial distribution of centroids within the field of view.

In Figures 2a and b, it can be observed that the epidermis of plants exposed to a 24-h daily photoperiod has a higher stomata density than plants exposed to 4 h. Due to this fact, networks constructed based on threshold distances (T_l) become highly connected in images where the stomatal density is larger. Due to the high plasticity of the plant, the number of stomata (vertices) of plants exposed to 24-h daily illumination is proportionally higher than in plants exposed to 4 h for the same threshold range. Likewise, in Figure 2c, which corresponds to the species *Callisia repens*, it is observed that higher CO_2 injection correlates with an increase in stomatal density and therefore leads to higher network connectivity for the same threshold values.

2.5. Life-like network automata

The LLNA $\mathcal{R}$ can be defined by the tuple $\mathcal{R} = \langle \mathcal{T}, \mathbf{S}, s_0, \phi \rangle$ (Miranda et al., 2016). The tessellation $\mathcal{T}$ represents the network topology, where each vertex is represented by a cell v_i connected to its k_i neighbours. $\mathbf{S}$ is the set of binary states, where $s_i = 1$ (live) and $s_i = 0$ (dead). The $s(c_i, t)$ function indicates the state of the cell c_i at time t; thus, s_0 represents the initial configuration of all cells and ϕ is the transition function, that is, the rule governing the dynamics of the network automata, which defines how the cell states are updated.

The LLNA transition function (ϕ) is defined as

$$s(c_i, t+1) = \begin{cases} 1, & \text{if } s(c_i,t) = 0 \text{ and } x/r \leq \rho_i < (x+1)/r \Rightarrow \text{born (B)} \\ 1, & \text{if } s(c_i,t) = 1 \text{ and } y/r \leq \rho_i < (y+1)/r \Rightarrow \text{survive (S)} \\ 0, & \text{otherwise,} \end{cases} \quad (2)$$

where the neighbourhood density ρ_i of node i is defined as the proportion of active neighbours, that is:

$$\rho_i = \frac{1}{k_i} \sum_{j=1}^{N} A_{ij} s(c_j, t), \quad (3)$$

where A_{ij} is the adjacency matrix of the network and k_i is the number of neighbours (degree) of node i.

In Equation (2), the variables x and y represent the conditions of the Life-Like rule expressed in the form B $x_0 x_1 \ldots x_8$-S $y_0 y_1 \ldots y_8$, where B and S represent the 'born' and 'survive'

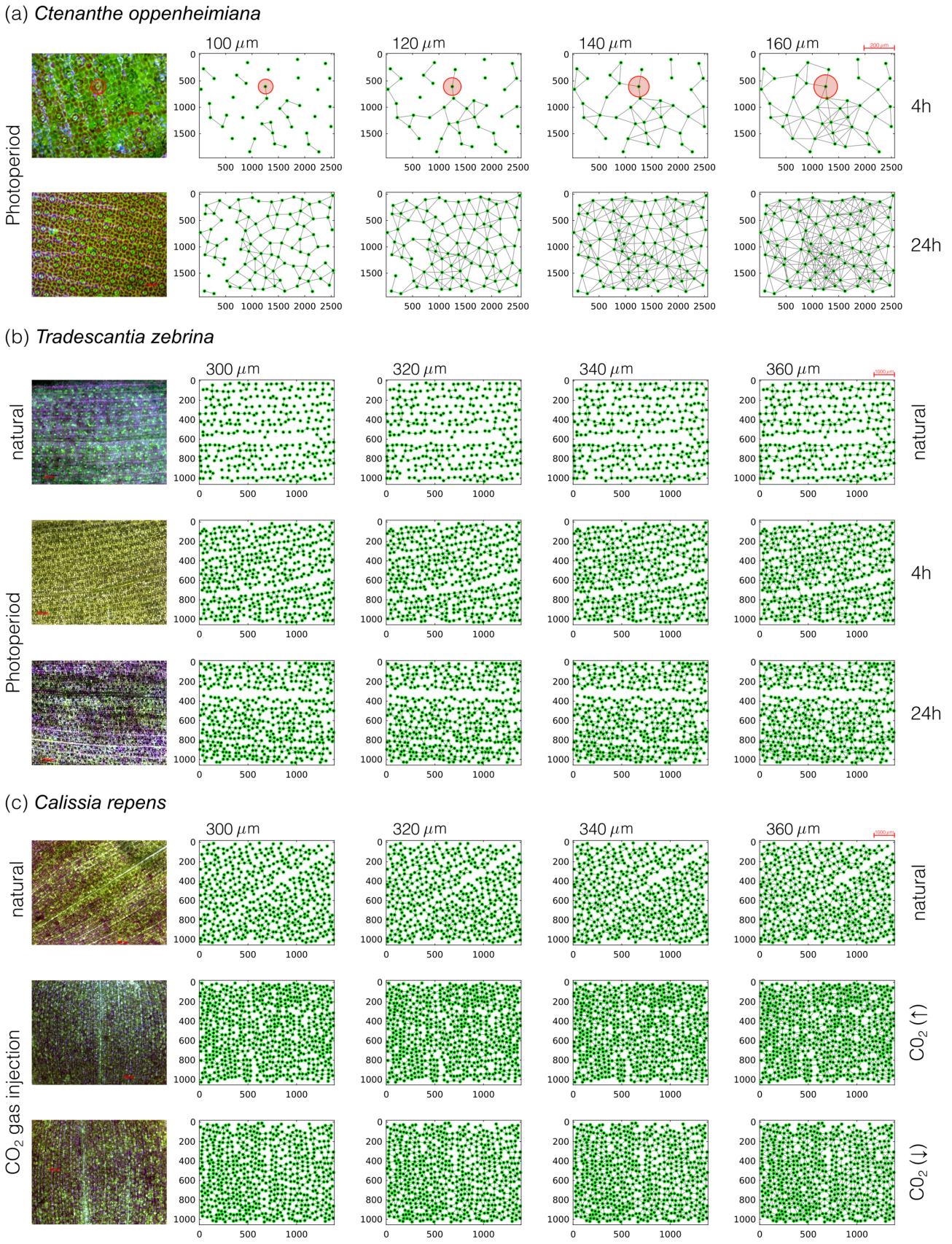


Figure 2. Stomatal network modelling. The first column shows microscopic images of leaf abaxial epidermis from species: (a) *Ctenanthe oppenheimiana* (4 h and 24 h photoperiods); (b) *Tradescantia zebrina* (natural environment and 4 h and 24 h photoperiods) and (c) *Callisia repens* (natural environment, bottled CO_2 flow and bottled control). Scale bars are provided to account for different magnifications (100 × for *Ctenanthe*; 15 × for others). Subsequent columns display the resulting networks across different threshold radii T_l (red circle). These thresholds are calibrated to each image scale to capture the topological transition from sparse to dense connectivity (vertices in green, edges in grey).

conditions, respectively, and $x_i, y_i \in \{0, \ldots, 8\}$. The LLNA establishes a correspondence between the transition function and the rules of the two-dimensional Life-Like CA family (*Bx-Sy*). While the Moore neighbourhood traditionally uses eight fixed neighbours ($r = 9$), the LLNA generalizes this to graphs using the neighbourhood density (Equation (3)), defined as the proportion of active neighbours. This ensures that the transition rules remain consistent regardless of the varying number of neighbours in the network. Therefore, there are a total of $2^{18} = 262,144$ transition rules available to find optimal solutions for a particular problem (Miranda et al., 2016).

2.6. Feature extraction

Using a given rule, the dynamics of a network automaton produce a time-evolution pattern (TEP) that represents the state sequence over time for each node. Each node can be analysed as a sequence of zeros and ones. From a TEP, various measurements can quantify individual node properties, allowing for network characterization. For feature extraction, we selected the suggested metrics: Shannon entropy (μ_S), word length (μ_W) and Lempel–Ziv complexity (μ_L). Since these LLNA metrics use the number of bins b as a parameter, we used the same ones as suggested by Miranda et al. (2016).

The first feature vector, μ_S, represents the distribution of Shannon entropy values ranging from 0 to 1. We calculated the Shannon entropy for each node and generated a histogram by dividing the interval $[0,1]$ into $b = 20$ bins, resulting in 20 descriptors reflecting their frequencies. The second feature vector (μ_W) captures the distribution of word lengths, defined as a sequence of ones surrounded by zeros. For instance, in $q = (00\mathbf{1111}0\mathbf{111}000)$, there is one word of length four and another of length three. The maximum word length is limited by the number of evolution steps; since words longer than 40 occur infrequently, we considered only shorter words. The histogram bins have a length of 2, yielding 20 descriptors for μ_W. The final feature vector (μ_L) consists of the Lempel–Ziv complexity distribution, also divided into $b = 20$ bins and normalized by the maximum value across the samples. Additionally, we created a composite vector by concatenating the three metrics $[\vec{\mu}_S, \vec{\mu}_W, \vec{\mu}_L]$, as well as the average values for these measures: $[\langle\mu_S\rangle, \langle\mu_W\rangle, \langle\mu_L\rangle]$.

2.7. Baseline methods

To characterize the topological properties of the stomatal networks and establish a robust comparative framework, we employed two sets of evaluation approaches. First, we calculated a set of traditional network measurements (Costa et al., 2011), including: average degree ($\langle k\rangle$), which reflects overall connectivity; average hierarchical degrees at levels 1 ($\langle H_{k_1}\rangle$) and 2 ($\langle H_{k_2}\rangle$), indicating immediate and neighbourhood connectivity; average clustering coefficient ($\langle cc\rangle$), quantifying node grouping; average path length ($\langle l\rangle$), providing insight into information flow efficiency; and the degree Pearson correlation (ρ_P), to identify assortative mixing patterns.

In addition to these structural metrics, we compared our LLNA approach against three selected graph-based classification methods to capture a broader spectrum of computational paradigms . The DTE method (Scabini et al., 2022) transforms network topology into high-dimensional tensor representations via tensor decomposition for classification. The DTWB approach (Merenda & Bruno, 2023) generates discriminative features by simulating deterministic walker movements to create histograms of trajectory lengths. Finally, the DGCNN (Zhang et al., 2018) utilizes hierarchical graph convolutions and TopK pooling to learn multi-scale graph representations for classification through fully connected layers. These specific methods were chosen because they operate effectively on unweighted whole-graph structures without requiring rich node or edge attribute matrices, which are absent in our network dataset. This constraint makes heavily parameterized or feature-dependent alternatives, such as standard graph transformers or dense graph neural networks (GNNs), less suitable for this specific task. By incorporating tensor decomposition, deterministic walks and deep learning, this selection provides a diverse, though non-exhaustive, framework to evaluate our approach.

2.8. Rule selection

To identify candidate rules capable of distinguishing environmental stressors, we adopted a generalization protocol using a subset of *Tradescantia zebrina*. As shown in Table 1, only two images per class were used for rule selection, while the remaining images and the other two species (*Ctenanthe* and *Callisia*) were reserved to evaluate the classifiers' performance. This protocol was designed to verify if the LLNA captures universal topological signatures of stomatal distribution that remain consistent across different species and stress types, without requiring re-optimization for each species.

We exhaustively evaluated each of the 262,144 rules of the Life-Like family in the set of stomata networks for rule selection. The networks were obtained using a threshold of $T_l = 512\,\mu$m (equivalent to 110 pixels). To characterize the LLNA dynamics, a feature vector was created by concatenating the Shannon entropy and the Lempel–Ziv $[\vec{\mu}_S, \vec{\mu}_L]$ distributions into 40 descriptors (Miranda et al., 2016). Both distributions were windowed into $b = 20$ bins and submitted to the k-nearest neighbours (kNN) classifier.

2.9. Classification task

We used the LLNA approach to extract intrinsic patterns from stomata networks aiming to distinguish between environmental conditions, aiding our understanding of plasticity. In the training phase, this technique extracts feature patterns for each stomata network (condition), which are then used to classify unseen instances during the classification phase.

After obtaining the feature vectors, we applied the kNN (for rule selection only) and support vector machine (SVM) algorithms, comparing our results with baseline methods. The SVM classification was performed using the sequential minimal optimization (SMO) algorithm with a polynomial kernel (exponent $E = 1.0$) and a complexity parameter $C = 1.0$ in Weka software (version 3.8), employing default settings to ensure a standardized comparison.

We conducted one experiment for each plant species, employing a k-fold cross-validation strategy (Bishop, 2006) to assess classification performance. To maximize the use of our dataset while ensuring statistical robustness, we set $k = 10$ for *Ctenanthe* and *Tradescantia*, and $k = 5$ for *Callisia* due to its sample size. The entire process was repeated across 100 independent experiments. This method splits the data into two sets: a training dataset for model training and a test set for validation. As these sets are mutually exclusive, the cross-validation method provides a reliable evaluation strategy. This procedure ensures that every sample is used for both training and testing in independent folds.

We further analysed the results using confusion matrices, which detail the relationship between predicted and real classes for each distinct environmental condition. A confusion matrix identifies performance patterns: α_{m_i,m_i} indicates instances correctly assigned to class m_i, while α_{m_i,m_j} shows instances of class m_i incorrectly

Table 2. Classification accuracy rates (%) and standard deviations for LLNA-based feature extraction

Rules	$\vec{\mu}_S$	$\vec{\mu}_W$	$\vec{\mu}_L$	$[\vec{\mu}_S, \vec{\mu}_W, \vec{\mu}_L]$	$[\langle\mu_S\rangle, \langle\mu_W\rangle, \langle\mu_L\rangle]$
Ctenanthe					
B57-S378	60.61 (± 5.42)	86.75 (± 4.12)	75.47 (± 1.71)	**95.03 (± 2.40)**	76.58 (± 3.97)
Tradescantia					
B12345-S04568	87.00 (± 6.70)	57.75 (± 10.62)	75.83 (± 8.12)	88.25 (± 6.47)	**90.00 (± 5.65)**
Callisia					
B014-S04567	**95.47 (± 4.61)**	23.93 (± 8.70)	74.87 (± 9.47)	65.67 (± 10.34)	70.67 (± 8.22)

Note: Feature vector dimensions are scaled according to the number of thresholds: 140, 420 and 21 descriptors for *Ctenanthe oppenheimiana*, and 320, 960 and 48 for *Tradescantia* and *Callisia*, respectively.

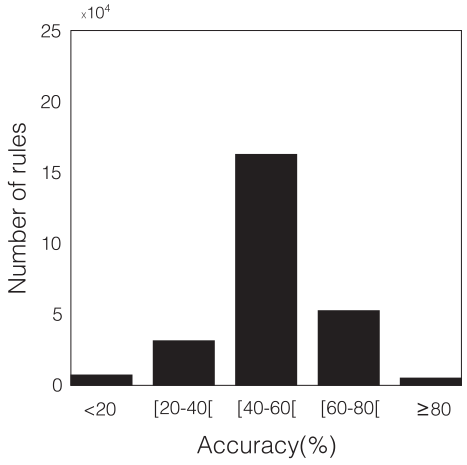


Figure 3. Histogram of the distribution of the 262,144 rules of the Life-Like family based on accuracy rates. The rule selection utilized stomata from the species *Tradescantia zebrina*. Shannon entropy and Lempel–Ziv complexity were considered only as feature vectors for the kNN classifier.

assigned to class m_j. This information is particularly useful for identifying samples that the proposed technique struggles to discriminate.

3. Results

The LLNA method was employed to characterize the phenotypic plasticity of the species *Tradescantia zebrina*, *Callisia repens* and *Ctenanthe oppenheimiana* when subjected to stress conditions. The dataset of networks used for validation is detailed in Table 1.

3.1. Configurations

3.1.1. Network threshold calibration. The distance threshold (T_l) range for network construction was tailored for each species to compensate for variations in optical magnification and stomatal density. For *Ctenanthe oppenheimiana*, we used seven thresholds ranging from T_{ini} = 140 to T_Q = 260 µm. For *Tradescantia zebrina* and *Callisia repens*, the feature vector was expanded to 16 thresholds from T_{ini} = 400 to T_Q = 640 µm. All pixel-to-micron conversions followed the optical scales defined in Section 2.3.

3.1.2. Life-like rules selection. Figure 3 shows the histogram distribution of the accuracy rates obtained for the rulespace of the Life-Like family. More than 5,000 rules achieved accuracy rates above 80%, confirming that several automata rules can effectively discriminate stomatal stress patterns. From this universe, we selected the 400 best-performing rules for the final classification stage. These candidate rules, originally identified using a subset of *Tradescantia zebrina*, were then used to train and test the models for all three species. This approach evaluates the generalization power of the rules across different plant species and environmental stressors.

3.2. Performance analysis

We adopted the strategy of concatenating descriptors obtained from the networks transformed by each value of δ_{T_l}, as proposed by Backes et al. (2009).

The feature vectors $\vec{\mu}_S$, $\vec{\mu}_W$ and $\vec{\mu}_L$ were constructed by concatenating the 20 descriptors calculated at each threshold level. For *Ctenanthe oppenheimiana*, seven thresholds were used, resulting in 140 descriptors. For *Tradescantia zebrina* and *Callisia repens*, we used 16 thresholds, resulting in 320 descriptors per vector. Additionally, we evaluated the combined vector $[\vec{\mu}_S, \vec{\mu}_W, \vec{\mu}_L]$ and the vector of means $[\langle\mu_S\rangle, \langle\mu_W\rangle, \langle\mu_L\rangle]$. The dimensionality of these vectors depends on the number of thresholds used: for *Ctenanthe* (seven thresholds), the vectors range from 21 to 420 descriptors, while for the other species (16 thresholds), they range from 48 to 960 descriptors (see Table S1 in the Supplementary Material).

Table 2 presents the results of the best rule for classifying the stomata networks of the three analysed species using the SVM classifier for each feature vector. The highest accuracy achieved for *Ctenanthe oppenheimiana* was 95.03 ± 2.40% using rule B57-S378 and the combined $[\vec{\mu}_S, \vec{\mu}_W, \vec{\mu}_L]$. *Tradescantia zebrina* was 90 ± 5.65% using LLNA rule B12345-S04568 and $[\langle\mu_S\rangle, \langle\mu_W\rangle, \langle\mu_L\rangle]$ as the feature vector. For *Callisia repens*, the highest accuracy was 95.47 ± 4.61% with rule B12345-S04568 and $\vec{\mu}_S$ as the feature vector. In all cases, the standard deviation inherent to the limited number of instances per class.

On the left side of Figure 4, the dimensional projection obtained from the canonical variable analysis (CVA) for the feature vectors with the highest accuracy for each species is presented, along with the confusion matrix on the right. From top to bottom, the vectors and rules used were: $[\vec{\mu}_S, \vec{\mu}_W, \vec{\mu}_L]$ with rule B57-S378 (*Ctenanthe oppenheimiana*), $[\langle\mu_S\rangle, \langle\mu_W\rangle, \langle\mu_L\rangle]$ with rule B12345-S0456 (*Tradescantia zebrina*) and $\vec{\mu}_S$ with rule B014-S04567 (*Callisia repens*). These figures clearly demonstrate a distinct separation among the three network models, supporting the high success rates. Additionally, the L24 class is linearly separable from the others, while the natural and L4 classes are often confused, as indicated by the confusion matrix, which aligns with the high similarity of these stomata images in Figure 2b.

3.3. Robustness analysis

To provide further intuition into the discriminative power of the LLNA, we analysed the features in the high-dimensional space.

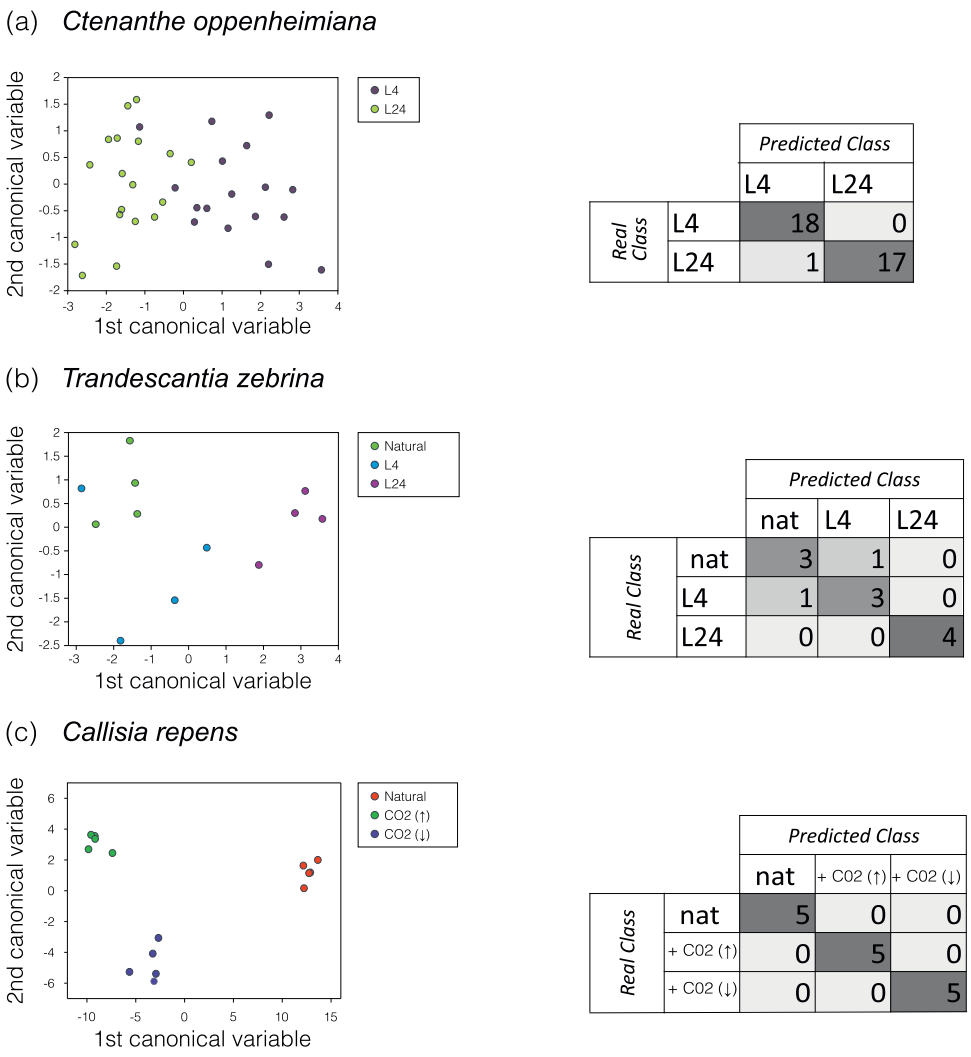


Figure 4. Two-dimensional projection obtained from the analysis of canonical variables and the corresponding confusion matrix for the highest accuracy rate. (a) CVA of the vector $[\vec{\mu}_S, \vec{\mu}_W, \vec{\mu}_L]$ with rule B57-S378 for the species *Ctenanthe oppenheimiana*. (b) CVA of the vector $[\langle\mu_S\rangle, \langle\mu_W\rangle, \langle\mu_L\rangle]$ with rule B12345-S0456 for classifying the species *Tradescantia zebrina*. (c) CVA of the vector $\vec{\mu}_S$ using rule B014-S04567 for the species *Callisia repens*. In the confusion matrix, each cell indicates the number of correctly predicted instances.

Figure 5 presents a pairplot of five representative descriptors for *Callisia repens*. Given the high dimensionality of the full feature vector (320 descriptors), only the first five were selected for this visualization. The diagonal plots show the estimated probability density for each individual descriptor, while the scatter plots illustrate how they combine to form well-separated clusters. Although the raw TEPs may appear stochastic and are not shown here – visual examples can be found in Miranda et al. (2016), Neiva and Bruno (2023), Ribas et al. (2020) and Zielinski et al. (2024) – these statistical measures confirm the separability of the classes, demonstrating that the LLNA signatures are highly sensitive to the topological patterns induced by environmental stress.

3.4. Baseline comparison

Table 3 compares the accuracy rates obtained by the LLNA with those from traditional structural network measurements and graph-based methods (Section 2.7).

Regarding traditional network measurements, for *Ctenanthe oppenheimiana*, LLNA showed a classification increment of 14.56 (±3.55) percentage points compared to the average degree ($\langle k \rangle$), which obtained 80.47 (±2.62)%. For *Callisia repens*, LLNA yielded a higher accuracy by 35.07 (±8.57) percentage points compared to the combination of measures that achieved 60.40 (±7.23)%. Finally, for *Tradescantia zebrina*, LLNA presented an absolute gain of 7.17 (±6.64) percentage points compared to the hierarchical degree ($\langle H_{k_2} \rangle$), which yielded 82.83 (±3.50)%. In these three scenarios, the results suggest that the dynamical signatures of the LLNA can capture topological nuances of phenotypic plasticity that may be less evident in static network metrics.

Regarding the three graph-based classification methods: DGCNN (Zhang et al., 2018), DTWB (Merenda & Bruno, 2023) and DTE (Scabini et al., 2022), the results in Table 3 indicate that LLNA shows competitive accuracy rates when evaluated against these graph-based methods across the three studied species . When compared to the DTE method, which was the most competitive among them in our tests , LLNA yielded higher accuracy margins . Specifically, LLNA presented a gain over DTE by 6.74 (±3.82) percentage points for *Ctenanthe oppenheimiana*, and by margins of 33.79 (±15.83) and 18.94 (±15.99) percentage points for *Tradescantia zebrina* and *Callisia repens*, respectively.

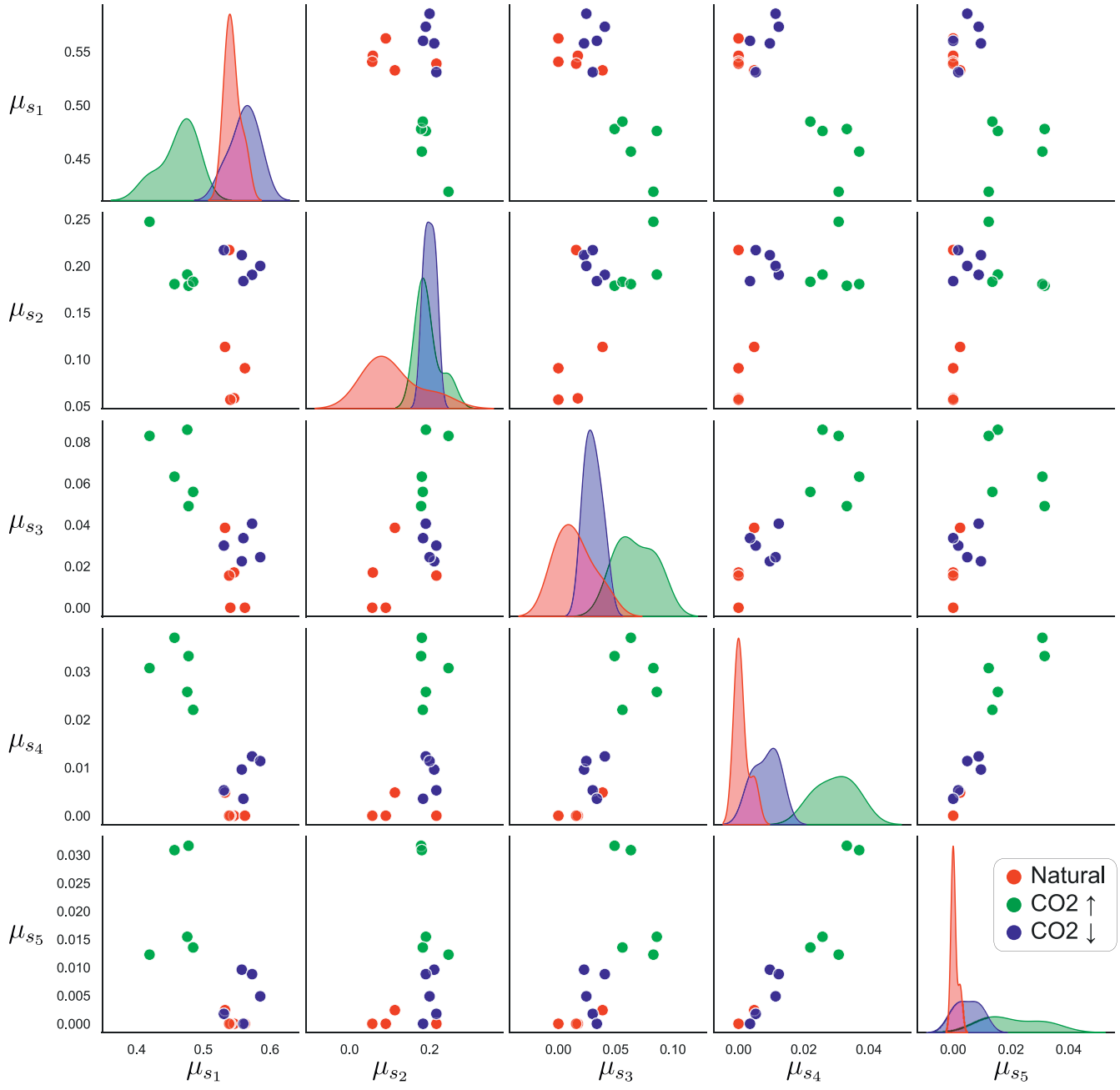


Figure 5. Pairplot analysis of the first five descriptors for *Callisia repens* using the best-performing rule (B014-S04567 with $\vec{\mu}_S$). The diagonal plots show the density distribution for each descriptor, while the scatter plots illustrate pairwise relationships.

The DGCNN and DTWB methods showed limited effectiveness in identifying the stress conditions, with accuracy rates often close to random guessing. The architectural complexity of deep frameworks like DGCNN easily leads to overfitting due to the severe imbalance between their large weight space and the small sample size available in this study. For DTWB, this performance might be explained by its original tuning on Watts–Strogatz networks, a specific topology that may not reflect the empirical distribution of our biological data. In contrast, while the DTE method performed better by using topological descriptors, it still exhibited high instability, as seen in the large standard deviations. The LLNA framework, by capturing the rules of network evolution as a deterministic feature extractor, appears to obtain stable patterns that are well-suited for these restricted data conditions.

4. Discussion

We analysed stomata distribution patterns by constructing stomata networks based on the distance matrix between pairs of stomata derived from microscopic images (Figure 2).

Regarding the network modelling, the connectivity of the network vertices (stomata) depends on the distance information and a threshold δ_{T_l}, with connectivity increasing as the threshold rises until the network becomes fully connected. The distance matrix for all pairs of stomata provides valuable insights into the geometry of the plant leaf, revealing correlations between the environment and its stomatic distribution structure (geometry). Notably, the specific threshold values used here do not limit the application of the method to other species or imaging setups. The fundamental

Table 3. Classification performance comparison

Method	Accuracy(%)	*Ctenanthe*	*Tradescantia*	*Callisia*
Network measures	$\langle k \rangle$	**80.47** (± 2.62)	82.67 (± 3.27)	46.60 (± 10.00)
	$\langle H_{k_2} \rangle$	77.72 (± 3.17)	**82.83** (± 3.50)	47.40 (± 10.45)
	$\langle H_{k_3} \rangle$	72.61 (± 4.23)	81.33 (± 4.58)	47.13 (± 10.08)
	$\langle cc \rangle$	73.97 (± 2.96)	66.83 (± 8.16)	47.93 (± 12.74)
	$\langle l \rangle$	72.42 (± 4.01)	38.75 (± 8.11)	35.53 (± 8.28)
	ρ_P	71.50 (± 3.98)	13.25 (± 7.18)	27.40 (± 10.19)
	combined	77.69 (± 2.62)	72.83 (± 8.38)	**60.40** (± 7.23)
Graph-based classification	DGCNN	59.26 (± 1.24)	34.89 (± 1.24)	37.54 (± 1.69)
	DTE	**88.29** (± 2.97)	**56.21** (± 14.79)	**76.53** (± 15.31)
	DTWB	63.06 (± 1.37)	37.49 (± 1.92)	34.46 (± 1.33)
Our	LLNA	**95.03** (± 2.40)	**90.00** (± 5.65)	**95.47** (± 4.61)
	Rule	(B57-S378)	(B12345-S04568)	(B014-S04567)

combined $[\langle k \rangle, \langle H_{k_2} \rangle, \langle H_{k_3} \rangle, \langle cc \rangle, \langle l \rangle, \rho_P]$
Note: Accuracy rates (%) and standard deviations for identifying phenotypic plasticity using traditional network structural measures, graph-based methods (DGCNN, DTE and DTWB) and the proposed LLNA approach (best-performing rules).

requirement for the LLNA is the coverage of a broad spectrum of connectivity – from a nearly disconnected state to a highly connected one. This multi-scale approach allows the automata dynamics to characterize the spatial complexity of the distribution; therefore, researchers can adapt these ranges to their specific resolution, provided that this topological progression is maintained.

While the exhaustive search for rules in the LLNA framework adds an initial computational layer, this selection is a one-time training phase. Once the optimal rules are identified, feature extraction and classification become computationally efficient. In our study, LLNA presented higher accuracy margins than the selected graph models (DGCNN, DTE and DTWB) under small-sample constraints, supporting the observation that while deep learning models often need very large datasets to work well, the LLNA can be a suitable alternative for biological data with few samples. Notably, the 400 optimal rules, derived from only two images of *Tradescantia zebrina*, generalized to the other two species without re-optimization. This indicates that these rules capture shared structural patterns related to stomatal distribution within the evaluated cases, instead of just learning the specific details of one dataset. Consequently, the proposed approach shows promise for small-sample biological datasets, where network dynamics can effectively capture relevant patterns despite data limitations.

From the perspective of pattern recognition, the method effectively captures the unique characteristics of the three species. The maximum accuracy achieved by LLNA consistently surpasses both structural network measures and graph-based classification methods. Specifically, for *Tradescantia zebrina*, we attained an accuracy of 90 (±5.65)%, despite the inherent difficulty in distinguishing natural and 4-h classes, as indicated by canonical analysis and the confusion matrix (Figure 4b). For *Ctenanthe oppenheimiana*, a similar high accuracy was observed, with only one sample misclassified. In the case of *Callisia repens*, we achieved excellent class separability.

From a biological standpoint, we observed that the stomata density (Figures 2a and b) is highly dependent on photoperiods, where longer exposure to light correlates with increased density. Similarly, for *Callisia repens*, the stress induced by CO_2 injection also influences stomata density, making the CO_2 (↑) and CO_2 (↓) classes challenging to differentiate using purely structural metrics.

Regarding plant plasticity, the LLNA approach helps to identify subtle changes in how stomata are organized. While traditional metrics often miss the complex spatial relationships between pores, the network automata can detect specific patterns caused by environmental stress. By measuring how these distributions change – from sparse to clustered or highly connected – the method provides a reliable way to evaluate these structural variations. This helps to show how plants adjust their leaf surface for gas exchange under different environmental conditions.

Such topological distribution analysis can be a valuable tool for monitoring environmental changes in natural ecosystems, as stomatal density alone is already used to model how plants sense and drive global changes (Cai & Cao, 2004; Hetherington & Woodward, 2003). Therefore, precise spatial frameworks like LLNA offer a non-destructive strategy to track human-induced climate impacts on biomes over time. This approach is particularly advantageous for live leaves of species with a purple abaxial epidermis – such as those studied here (de Almeida Filho, 2018) – since distance metrics can be preserved and analysed without prior histological processing. Additionally, this methodology can be extended to agricultural sciences to evaluate crop adaptation to different fertilization, irrigation and local climatic regimes.

Despite these achievements, the dataset's limitations must be acknowledged. Compiling a comprehensive biological dataset was inherently challenging due to the meticulous time constraints required for controlled plant cultivation and monitoring. It is worth noting that while the LLNA framework was originally designed with large datasets in mind (Miranda et al., 2016), evaluating its performance under these restricted sample sizes was a deliberate test of its versatility. The positive results observed here suggest a promising adaptability to data-scarce scenarios, although further extensive testing remains necessary. Future research should focus on automated methods for processing leaf images and expanding the datasets to include more species to further enhance the robustness and validate the generalizability of our approach.

Finally, the development of these computational tools remains essential for expanding our understanding of ecological and agricultural resilience. In this context, our findings suggest the feasibility of the LLNA as an effective pattern recognition tool to quantify the phenotypic plasticity of stomatal distribution in response to environmental stressors.

Open peer review. To view the open peer review materials for this article, please visit http://doi.org/10.1017/qpb.2026.10051.

Acknowledgements

We would like to thank the Scientific Computing Group of the São Carlos Institute of Physics for providing the stomatal image dataset.

Author contributions. All authors contributed synergistically to the different stages of the research project. More specifically:

Conceptualization: J.M., H.A.d.A.F. and O.M.B.; Data curation: J.M. and H.A.d.A.F.; Formal analysis: J.M.; Investigation: J.M. and H.A.d.A.F.; Methodology: J.M. and H.A.d.A.F.; Software: J.M. and K.Z.; Supervision: O.M.B.; Validation: J.M. and H.A.d.A.F.; Visualization: J.M.; Writing – original draft: J.M. and H.A.d.A.F.; Writing – review and editing: J.M. All authors have read and agreed to the published version of the manuscript.

Competing interests. There are no competing interests to declare.

Data availability statement. A zip file containing the images dataset and centroid files for each of the three species can be found in de Almeida Filho & Machicao (2026) https://doi.org/10.5281/zenodo.18567387. The source code for the LLNA feature extraction and the classification pipeline is now available at https://github.com/machicao/stomata-LLNA.

Funding statement. This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Disclosure statement. This manuscript has not been published nor is it under consideration by any other journal. There are no financial obligations with governmental or private organizations that can affect the content, results or conclusions of the present manuscript.

Declaration of generative AI in the writing process. During the preparation of this work, the authors used Gemini to enhance text clarity and flow. The authors thoroughly reviewed and edited the final content and take full responsibility for the published article.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/qpb.2026.10051.

References

Backes, A. R., **Casanova, D.**, & **Bruno, O. M.** (2009). A complex network-based approach for boundary shape analysis. *Pattern Recognition*, **42**(1), 54–67.

Beerling, D., & **Woodward, F.** (1997). Changes in land plant function over the phanerozoic: Reconstructions based on the fossil record. *Botanical Journal of the Linnean Society*, **124**(2), 137–153.

Bishop, C. M. (2006). *Pattern recognition and machine learning (information science and statistics)*. Springer-Verlag New York, Inc.

Cai, Z., & **Cao, K.** (2004). Response of stomatal characteristics and its plasticity to different light conditions in leave of seven tropical woody seedlings. *Chinese Journal of Applied Ecology*, **15**(2), 201–204.

Costa, L., **Oliveira, O. N.**, **Travieso, G.**, **Rodrigues, F. A.**, **Villas Boas, P. R.**, **Antiqueira, L.**, **Viana, M. P.**, & **da Rocha, L. E. C.** (2011). Analyzing and modeling real-world phenomena with complex networks: A survey of applications. *Advances in Physics*, **60**(3), 329–412.

Costa, L., **Rodrigues, F. A.**, **Travieso, G.**, & **Boas, P. V.** (2007). Characterization of complex networks: A survey of measurements. *Advances in Physics*, **56**(1), 167.

Cowan, I. R., & **Troughton, J. H.** (1971). The relative role of stomata in transpiration and assimilation. *Planta*, **97**(4), 325–336.

da Silva, N. R., **Oliveira, M. W. D. S.**, **Filho, H. A. D.**, **Pinheiro, L. F. S.**, **Rossatto, D. R.**, **Kolb, R. M.**, & **Bruno, O. M.** (2016). Leaf epidermis images for robust identification of plants. *Scientific Reports*, **6**, 25994, https://doi.org/10.1038/srep25994.

de Almeida Filho, H. A. (2018). Plants with purple abaxial leaves: A repository of metrics from stomata distribution. bioRxiv, 294553.

de Almeida Filho, H. A., & **Machicao, J.** (2026). [Dataset] Spatial centroids and network representations of stomata in three plant species under environmental stress. Zenodo. https://doi.org/10.5281/ZENODO.18567387

Eckerson, S. H. (1908). The number and size of the stomata. *Botanical Gazette*, **46**(3), 221–224.

Florindo, J. B., **Landini, G.**, **Almeida Filho, H.**, & **Bruno, O. M.** (2015). Analysis of stomata distribution patterns for quantification of the foliar plasticity of Tradescantia Zebrina. *Journal of Physics Conference Series*, **633**, 12113.

Hetherington, A. M., & **Woodward, F. I.** (2003). The role of stomata in sensing and driving environmental change. *Nature*, **424**(6951), 901–908.

Liu, K., **Wang, Y.**, **Magney, T. S.**, & **Frankenberg, C.** (2024). Non-steady-state stomatal conductance modeling and its implications: From leaf to ecosystem. *Biogeosciences*, **21**(6), 1501–1516. https://doi.org/10.5194/bg-21-1501-2024

Machicao, J., **Corrêa, E. A.**, **Miranda, G. H. B.**, **Amancio, D. R.**, & **Bruno, O. M.** (2018). Authorship attribution based on life-like network automata. *PLoS One*, **13**(3), e0193703.

Merenda, J. V., & **Bruno, O. M.** (2023). Using deterministic self-avoiding walks as a small-world metric on watts–strogatz networks. *Physica A: Statistical Mechanics and its Applications*, **621**, 128713.

Miranda, G. H., **Machicao, J.**, & **Bruno, O. M.** (2016). Exploring spatiotemporal dynamics of cellular automata for pattern recognition in networks. *Scientific Reports*, **6**, 37329, https://doi.org/10.1038/srep37329.

Miranda, G. H., **Machicao, J.**, & **Bruno, O. M.** (2018). An optimized shape descriptor based on structural properties of networks. *Digital Signal Processing*, **82**, 216–229.

Murchie, E. H., & **Horton, P.** (1997). Acclimation of photosynthesis to irradiance and spectral quality in British plant species: Chlorophyll content, photosynthetic capacity and habitat preference. *Plant, Cell &; Environment*, **20**(4), 438–448.

Neiva, M. B., & **Bruno, O. M.** (2023). Exploring ordered patterns in the adjacency matrix for improving machine learning on complex networks. *Physica A: Statistical Mechanics and its Applications*, **626**, 129086. https://doi.org/10.1016/j.physa.2023.129086

Peak, D., **Hogan, M. T.**, & **Mott, K. A.** (2023). Stomatal patchiness and cellular computing. *Proceedings of the National Academy of Sciences*, **120**(14). https://doi.org/10.1073/pnas.2220270120

Ribas, L. C., **Machicao, J.**, & **Bruno, O. M.** (2020). Life-like network automata descriptor based on binary patterns for network classification. *Information Sciences*, **515**, 156–168.

Scabini, L., **Ribas, L.**, **Ribeiro, E.**, & **Bruno, O.** (2022). Deep topological embedding with convolutional neural networks for complex network classification. In Ribeiro, P., Silva, F., Mendes, J.F., Laureano, R. (eds) *Network science*. NetSci-X 2022. Lecture Notes in Computer Science, vol 13197. (pp. 54–66). Springer, Cham, https://doi.org/10.1007/978-3-030-97240-0_5.

Viscosi, V., & **Cardini, A.** (2011). Leaf morphology, taxonomy and geometric morphometrics: A simplified protocol for beginners. *PLoS One*, **6**(10), e25630.

Willmer, C., & **Fricker, M.** (1996). *Stomata*. Springer. https://doi.org/10.1007/978-94-011-0579-8

Zhang, M., **Cui, Z.**, **Neumann, M.**, & **Chen, Y.** (2018). *An end-to-end deep learning architecture for graph classification*. AAAI Press.

Zielinski, K. M., **Ribas, L. C.**, **Machicao, J.**, & **Bruno, O. M.** (2024). A network classification method based on density time evolution patterns extracted from network automata. *Pattern Recognition*, **146**, 109946. https://doi.org/10.1016/j.patcog.2023.109946