

Classification of *Camellia* (Theaceae) Species Using Leaf Architecture Variations and Pattern Recognition Techniques

Hongfei Lu^{1*}, Wu Jiang¹, M. Ghiassi², Sean Lee², Mantri Nitin³

1 College of Chemistry and Life Science, Zhejiang Normal University, Jinhua, China, **2** Santa Clara University, Santa Clara, California, United States of America, **3** School of Applied Sciences, Health Innovations Research Institute, RMIT University, Melbourne, Victoria, Australia

Abstract

Leaf characters have been successfully utilized to classify *Camellia* (Theaceae) species; however, leaf characters combined with supervised pattern recognition techniques have not been previously explored. We present results of using leaf morphological and venation characters of 93 species from five sections of genus *Camellia* to assess the effectiveness of several supervised pattern recognition techniques for classifications and compare their accuracy. Clustering approach, Learning Vector Quantization neural network (LVQ-ANN), Dynamic Architecture for Artificial Neural Networks (DAN2), and C-support vector machines (SVM) are used to discriminate 93 species from five sections of genus *Camellia* (11 in sect. *Furfuracea*, 16 in sect. *Paracamellia*, 12 in sect. *Tuberculata*, 34 in sect. *Camellia*, and 20 in sect. *Theopsis*). DAN2 and SVM show excellent classification results for genus *Camellia* with DAN2's accuracy of 97.92% and 91.11% for training and testing data sets respectively. The RBF-SVM results of 97.92% and 97.78% for training and testing offer the best classification accuracy. A hierarchical dendrogram based on leaf architecture data has confirmed the morphological classification of the five sections as previously proposed. The overall results suggest that leaf architecture-based data analysis using supervised pattern recognition techniques, especially DAN2 and SVM discrimination methods, is excellent for identification of *Camellia* species.

Citation: Lu H, Jiang W, Ghiassi M, Lee S, Nitin M (2012) Classification of *Camellia* (Theaceae) Species Using Leaf Architecture Variations and Pattern Recognition Techniques. PLoS ONE 7(1): e29704. doi:10.1371/journal.pone.0029704

Editor: Robert DeSalle, American Museum of Natural History, United States of America

Received: September 22, 2011; **Accepted:** December 2, 2011; **Published:** January 3, 2012

Copyright: © 2012 Lu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was supported by grants from the Science and Technology research Plan of Jinhua city, China (No. 2009-2-020). No additional external funding received for this study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: luhongfei63@yahoo.com.cn

These authors contributed equally to this work.

Introduction

Camellia is a large genus of family Theaceae with many species of significant economic and scientific value [1]. Some *Camellia* species are used to produce green tea, a popular beverage. It is estimated that more than 3.6 million tons of tea leaves are produced annually in 40 countries [2,3,4]. *Camellia* species offer a range of health benefits [5]. Some species are primarily cultivated as ornamental plants while the seeds of others are used as edible oils [6,7]. This wide usage of the *Camellia* species has resulted in extensive cultivation and production. In China alone, more than 3 million hectares of agricultural land is used to grow *Camellia* species to produce in excess of 164,000 tons of edible cooking oil [5].

Although *Camellia* is grown in many regions of the world, it is particularly prevalent in East and Southeast Asia and its identification and classification has been the subject of many studies [6,7,8,9]. Traditionally, professionals dealing with the production, distribution and sales of *Camellia* use their experience and intuition to classify the plants into categories with distinct economic values. Later, researchers developed different taxonomic and analytical methods for classification. In 1958, Sealy [8] reported 82 *Camellia* species that he classified

into 12 sections. More recently, Chang [10] grouped the native Chinese *Camellia* into four subgenera, 22 sections, and 280 species, whilst Ming [6] arranged them into two subgenera, 14 sections, and 119 species [11]. However, there is still disagreement in the interspecies relationship of the genus *Camellia* [5].

The aforementioned classifications were based on morphological approach. Recent studies suggest that classifications purely based on the traditional morphological characteristics are insufficient [12,13,14]. Therefore, alternative taxonomic methods were developed for classification of *Camellia* [15,16].

Contemporary advances in technology have resulted in new tools that allow classification based on alternative and innovative approaches. Lu et al. [12] used Fourier transform infrared spectroscopy (FTIR) on *Camellia* leaves to determine if they can be discriminated based on biochemical profiles. Chen et al. [3] and Yang et al. [17] used molecular approach based on genetic information for classification of *Camellia* species. Clearly, there is disagreement among researchers and no dominant method for this important classification problem has emerged. There are still many uncertainties about the relationships among species within sections and further taxonomic research on this section is necessary [13].

We acknowledge that although the flowers and the fruit are seasonal, the leaf lacks those limitations and their traits are more commonly used in plant taxonomic applications [18,19,20,21]. Especially, Lin et al. [22] and Lu et al. [12] successfully revised three sections of genus *Camellia* based on leaf anatomic characters. Pi et al. [13] have used leaf morphology and anatomical characters for delimitation of species. They report that “leaf features have been largely unexploited in taxonomic studies, resulting from a belief that they respond in a plastic manner to environmental forces.”

Although leaf morphology has been the subject of some studies, lack of standard definitions of leaf characteristics has caused confusion in interpreting the value of the resulting classifications [13]. Taxonomical classification of *Camellia* based on a more comprehensive description of leaf morphology (also referred to as leaf architecture) is, therefore required. Leaf architecture refers to the placement and form of various elements constituting the outward expression of leaf structure, including leaf shape, leaf size, marginal configuration, gland position and venation pattern [23]. The leaf architecture has been the subject of several studies to resolve taxonomic and evolutionary relationships [24]. However, little research has been performed utilizing leaf architecture of genus *Camellia* species [25,26,27,28].

The traditional analytical approaches employed by researchers to perform *Camellia* classification have included the principal component analysis, multivariate analysis, cluster analysis, and simulated annealing. Recently, some researchers have used supervised classification techniques in their studies. Supervised techniques are one of the most effective analysis tools in a variety of domains, such as information retrieval, remote sensing, and food bruise detection [29,30,31]. These tools apply available information about a category membership of samples to develop a model for classification of the genus. The classification model is developed using a training set with a priori defined categories and

the performance is appraised using samples from a test set by comparing predicted categories with their true categories, as defined by experts [32,33].

Artificial neural networks (ANN), as a pattern recognition tool, have been used for modeling complex systems [34,35,36,37,38]. Pandolfi et al. [15] discriminate and identify morphotypes of *Banksia integrifolia* by BP-ANN based on morphological and fractal parameters of leaves. Similarly, Pandolfi et al. [38] have used the BP-ANN approach to morphologically differentiate 17 Vietnamese tea plants. Support vector machine (SVM) is another supervised pattern recognition technology that has seen popularity of applications over the past several years [31,39,40,41,42]. This algorithm was developed in the machine learning community [43,44] and is capable of learning in high-dimensional feature spaces [45].

Although pattern recognition tools have been applied in variety of fields, to the best of our knowledge this approach has not been used for classification of genus *Camellia* using leaf architecture data. We have used two different ANN architectures (LVQ-ANN and DAN2) and the support vector machine (SVM) to model *Camellia* classification. As stated earlier, there is still disagreement in the interspecies relationship of the genus *Camellia* [5,12,13,14]. Researchers continue to use different taxonomical methods and analytical approaches to find more discriminating results. In this research, we combine the leaf architecture properties of genus *Camellia* with various pattern recognition tools, including a newly introduced method (DAN2), using a relatively large data set, to analyze the taxonomical classification of *Camellia* plants. The goal of the present work, therefore, is to classify *Camellia* species based on leaf architecture data. We present (1) results of using leaf morphological and venation characters of 93 species in five sections for *Camellia* classification, and (2) report the effectiveness of supervised pattern recognition techniques (LVQ-ANN, DAN2, and SVM) for such classifications and (3) compare their accuracy.

Table 1. Leaf architectural characters and related morphological characters [23,26].

Characteristic	Encoding number					Figure 1
	0	1	2	3	4	
1. Whole lamina shape	Lanceolate	Ovate	Oblong	Longly oblong	Broadly oblong	See A
2. Base only	Auriculate	Rounded	Cuneate			See B
3. Apex	Long acuminate	Short acuminate	Obtusely acuminate	Acute		See C
4. Abaxial surface	Hairy	Glabrous				/
5. Adaxial surface	Hairy	Glabrous				/
6. Reticulate veins	Slight	Unconspicuous	Obvious	Conspicuous		See G
7. Secondary veins shape	Bend	Zigzag				See D
8. Secondary veins balance	Uniform	Uneven				See E
9. Areoles development	Incompletely	Imperfect				See K
10. Margin shape	Entire	Spinose	Cassidate	Setaceous	Spherulate	See H
11. Margin spacing	Regular	Irregular				See I
12. ASVPV on upper part ¹	Nearly right angle	Sharp angle				/
13. ASVPV on middle part	Nearly right angle	Sharp angle				/
14. ASVPV on lower part	Nearly right angle	Sharp angle				/
15. VADSV ²	Nearly uniform	Upper secondary veins more obtuse than lower	Upper secondary veins more acute than lower			See F
16. Veinlets		Simple	1–2 times branched	More than 3 times branched		See J

¹ASVPV means angulation between secondary veins and primary veins;

²VADSV means variations in angle of divergence of secondary veins.

doi:10.1371/journal.pone.0029704.t001

Materials and Methods

Materials

In this research we use comprehensive leaf morphology and leaf architecture for taxonomical classification of *Camellia*. Healthy leaf samples, consisting of 11 species from sect. *Furfuracea*, 16 species from sect. *Paracamellia*, 12 species from sect. *Tuberculata*, 34 species from sect. *Camellia*, and 20 species from sect. *Theopsis*, for a total of 93 plants, are examined in this study (following Chang [10] taxonomic treatment, Table S1). Leaf samples were taken from the third mature leaves that were fully exposed to sunlight and were horizontally arranged on the 2-year-old branches of the plants in the garden. At least three different individual plants per species are selected. Plant materials are all collected from the International *Camellia* Garden in Jinhua, Zhejiang Province (29°07'N, 119°35'E, altitude 40 m). Voucher specimens for all species are deposited in

the Chemistry and Life Science College of Zhejiang Normal University (ZJNU) (see Appendix S1 for voucher details).

Leaf veins specimen preparation

In order to use information from leaf vein patterns, we produced leaf veins specimens. The method used for making leaf veins specimen follows the process used by Zhang and Xia [46]. Leaves were placed in a glass tube, 10% sodium hydroxide (NaOH) was added in sufficient quantity to cover the material at 70–80°C for 3–4 hours. Since the leaf texture may differ among species, thicker leaves were treated for longer time periods. Leaves were taken out when the epidermis and mesophyll showed sufficient segregation. The leaves were then gently brushed to remove the epidermis and mesophyll with a paint brush. They were next rinsed in water and bleached by 10% hydrogen peroxide (H₂O₂) for approximately 60 minutes until the specimen

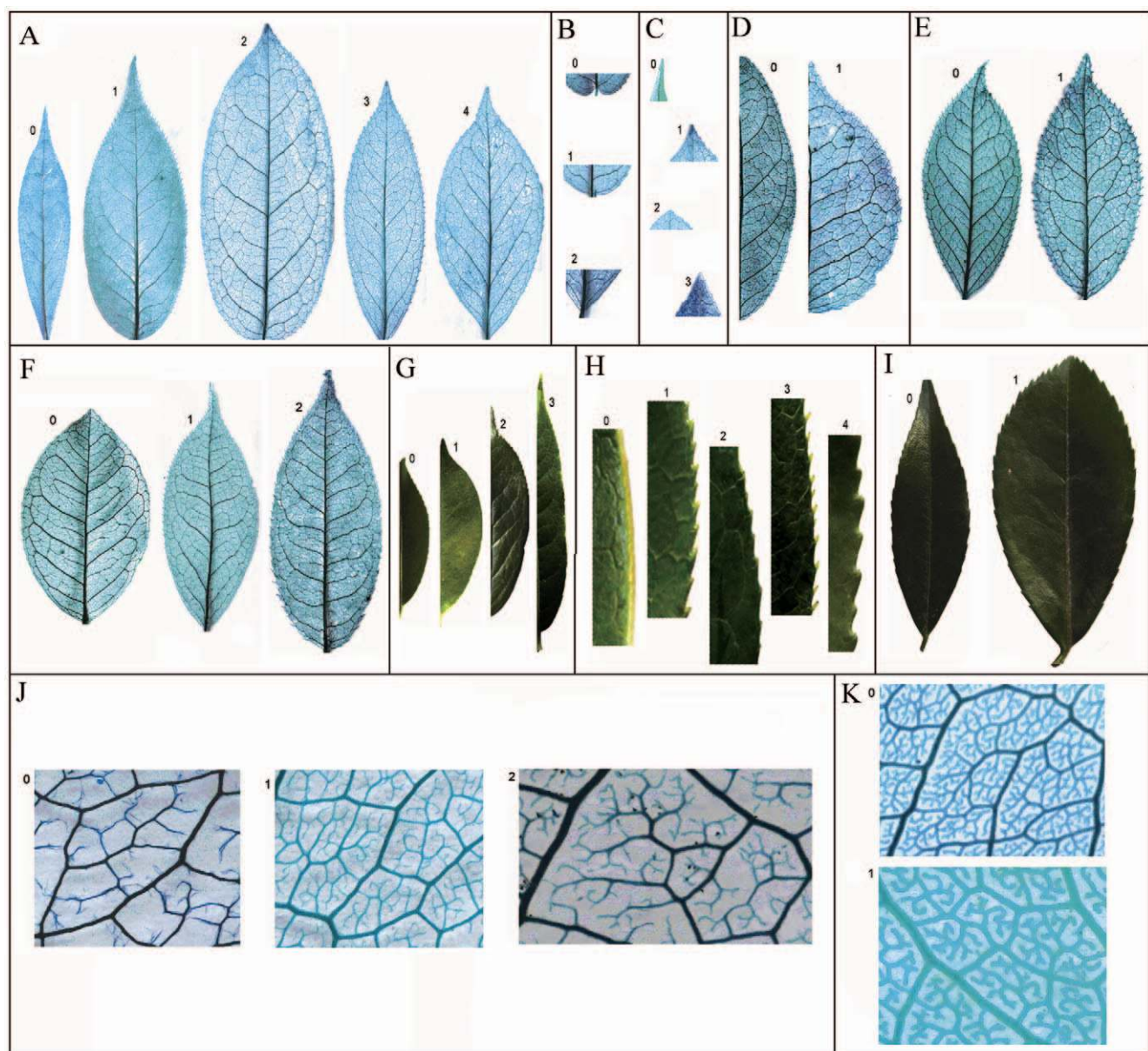


Figure 1. Leaf architectural characters and related morphological characters.

doi:10.1371/journal.pone.0029704.g001

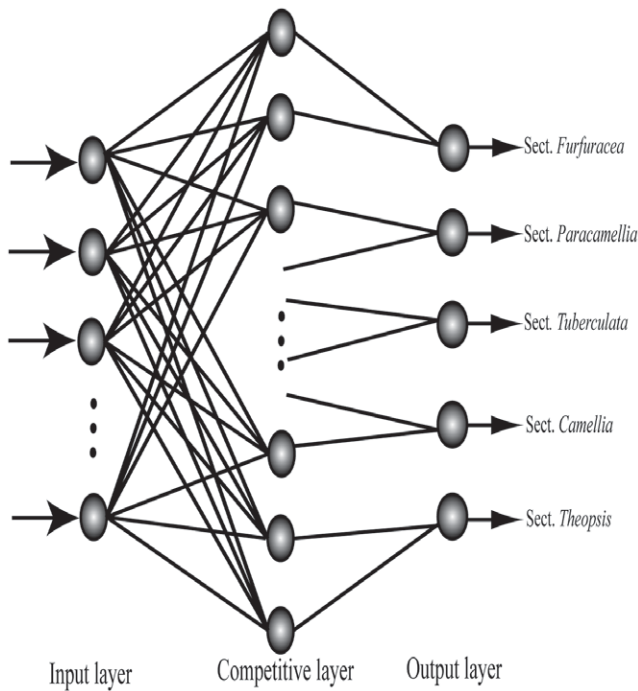


Figure 2. Schematic diagram of LVQ-ANN.
doi:10.1371/journal.pone.0029704.g002

was white. Bleached and cleared leaves were then washed in running water thoroughly. They were then fully stained by 0.5% methyl green for at least two hours. Subsequently, their pictures were taken with the Canon EOS 50D camera for further analysis.

Leaf architecture data collection

(Table 1) presents list of the most commonly used leaf characteristics from literature [23,26]. 31 characteristics of each leaf are collected and measured. All the test indexes were measured according to Hickey [23] specifications and guidelines. The leaf architecture data results are expressed as mean values. Below we describe the process in detail.

I. Leaf shape observation

Following earlier research [23,26], we selected 16 characteristics of leaf architecture and morphology that best describe leaf shapes for this research (Figure 1, Table 1). The same encoding values are used in all of the classification models.

II. Leaf size measurements

The leaves of each species were scanned by CanoScan 4400FF Canon scanner (resolution of 4800*9600 dpi) using the WinFOLIA system (Regent Instruments Inc., Canada). For each sample, we measure leaf area, perimeter, vertical length, horizontal width, leaf aspect ratio (width/length), and leaf form factor (LEF). The formula used for estimating the LEF is:

$$LEF = \frac{16 * A}{P^2} \tag{1}$$

Where *A* is the area and *P* is the perimeter. Other characteristics including number of secondary veins (pairs), petiole length, average value of entirely vein height (EVH), average value of leaf widest part height (LWPH), average ratio of EVH and leaf vertical length, average ratio of LWPH and leaf vertical length, serrulate length in upper part of leaf, serrulate length in middle part of leaf, and serrulate length in lower part of leaf, were measured with ImageJ Launcher (Broken Symmetry Software).

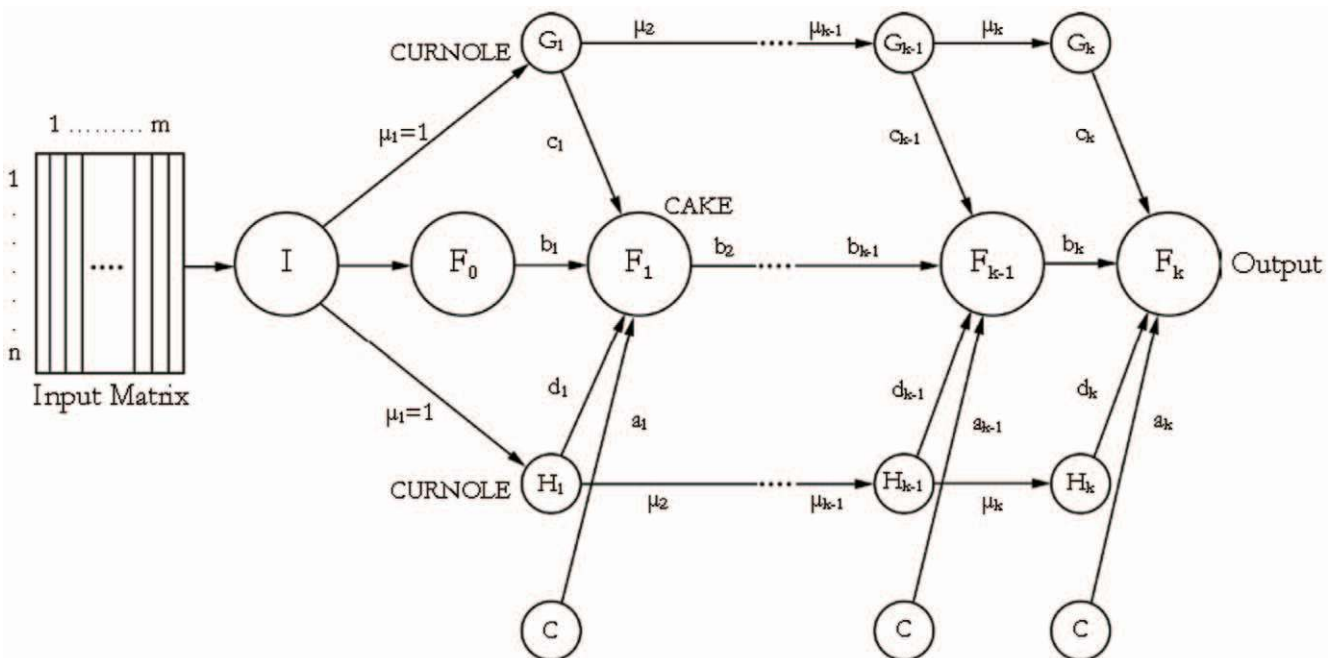


Figure 3. The DAN2 Network Architecture.
doi:10.1371/journal.pone.0029704.g003

Analysis methods

Traditional methods used for *Camellia* classification includes principal component analysis, multivariate analysis, cluster analysis, and simulated annealing. We examine the effectiveness of using various pattern recognition methods in this research. Specifically, we used a traditional artificial neural network (LVQ), a dynamic artificial neural network (DAN2), support vector machines (SVM), and cluster analysis for classification of the 93 samples. We used Chang (1998) classification data, presented in (Table S1), grouped into training and testing data sets, to measure and compare the accuracy of the three classification algorithms presented in this study. The training of these algorithms relies on (a) leaf characteristics data, and (b) class designation. The class designation used predefined Chang [10] classification.

LVQ-ANN classification model. The first ANN model used is the Learning Vector Quantization (LVQ). LVQ is a special case of ANN that uses the “winner-take-all Hebbian learning strategy” [47,48]. The network architecture consists of three layers: the input layer, the competitive layer (Kohonen layer) and the output

layer. The input layer represents properties of species while the output layer represents the number of classes. In the competitive layer, each unit corresponds to a cluster, with the center designated as the “codebook” vector. An input vector closest to the codebook vector (using the Euclidean distance measure) belongs to the corresponding cluster. The optimal number of neurons in the competitive layer is determined experimentally. In this study, 93 samples belong to five different sections (categories) were selected: 48 samples were used to generate the classification model input for the training set and the remaining 45 samples were used in the testing stage (Table S1). Each vector of the input layer includes the 31 feature attributes of leaf architecture mentioned earlier. The number of nodes in the competitive layer varied from 20 to 30, and their impact was assessed on the respective classification capabilities. The output layer contained five neurons representing specific sections (taxon), including sect. *Furfuracea*, sect. *Paracamellia*, sect. *Tuberculata*, sect. *Camellia*, and sect. *Theopsis*. LVQ-ANN topology used in our study is shown in (Figure 2). By computing the Euclidean distance between a

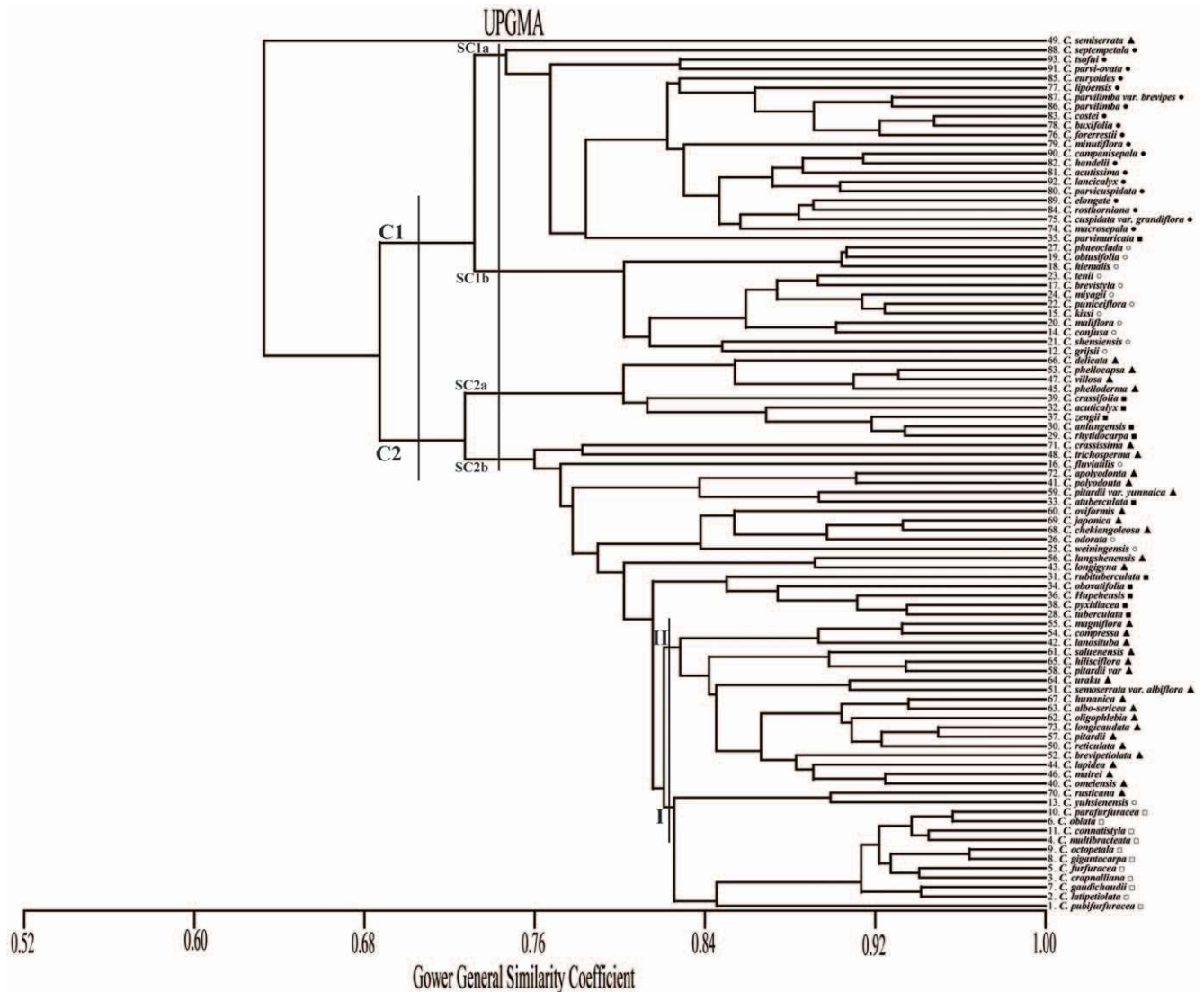


Figure 4. UPGMA dendrogram of genus *Camellia* based on leaf architectural characteristics. sect. *Camellia* (▲), sect. *Theopsis* (●), sect. *Tuberculata* (■), sect. *Paracamellia* (○), sect. *Furfuracea* (□). doi:10.1371/journal.pone.0029704.g004

training vector and the weight of each node, the nearest node (“winner node”) was generated. Winner nodes move towards the training vector when the winner nodes are in the same class, otherwise, they move away. The input vectors were then allocated to the category with the winning nodes. Training is complete when the mean square error (MSE) converges, or it is less than 0.1, or the number of training iterations reaches 1,000 epochs. We used two implementations of the LVQ algorithm: the LVQ1 and LVQ2. In LVQ1 a single best machine codebook vector is selected and moved closer or further for each data vector at each iteration, whereas in LVQ2 two sets of best machine codebook vectors are selected and only updated if one belongs to the desired class and one does not [49]. The LVQ-ANN modeling program was designed and programmed under MATLAB software (The Mathworks, Inc., Natick, MA, USA, version 7.9 R2009b).

DAN2 classification model. DAN2, (*A Dynamic Architecture for Artificial Neural Networks*), is a dynamic ANN model. It consists of input and output layers similar to LVQ and other ANNs. However, in DAN2 the number of hidden layers and hidden neurons are automatically and dynamically generated [50]. Two significant properties of DAN2 are: (1) its dynamic nature eliminates the need to experimentally define the number of hidden layers and hidden nodes, and (2) its architecture is fully scalable and can easily and effectively process any number of inputs. DAN2 is shown to be very effective in solving a variety of complex problems including classification problems [51,52]. (Figure 3) presents the overall DAN2 architecture. As shown in (Figure 3), each hidden layer is composed of four nodes. The first node is the bias or constant (e.g. 1) input node, referred to as the C node. The second node is a function that encapsulates the “Current Accumulated Knowledge Element” (CAKE node) during the previous training step. The third and fourth nodes represent the current residual (remaining) nonlinear component of the process via a transfer function of a weighted and normalized

sum of the input variables. Such nodes represent the “Current Residual Nonlinear Element” (CURNOLE nodes).

The scalability of DAN2 is a distinguishing strength of the approach from traditional artificial neural networks. In order to compare effectiveness of each technique, we use the exact same input vectors and the same training and testing data sets for the DAN2 model that were used in the LVQ models (Table S1) and report its results.

SVM classification model. SVM is based on statistical learning theory and structural risk minimization and was first proposed by Vapnik [44]. This approach generates hyperplanes to separate classes [53]. The boundaries of the hyperplane are represented by support vectors instead of a single boundary value. Support vectors run through the sample patterns which are the most difficult to classify and are closest to the actual class boundaries. Overfitting is prevented by specifying a maximum margin that separates the hyperplanes from the classes [54]. Samples violating this margin are penalized.

Once again, the exact same input vectors and training and test data sets that were used in LVQ were also used for the support vector machines models (Table S1). All C-SVM algorithms were implemented with LIBSVM (Version 3.0) under MATLAB software [55].

Cluster analysis

Cluster analysis is the process of grouping data based on objects’ attributes into similar and dissimilar groups. In this research, we use clustering analysis to classify 5 sections in genus *Camellia* based on the leaf architecture data (31 attributes) and to compare the results with Chang [10]. The clustering approach used is based on the Unweighted Pair-Group Method with Arithmetic Means (UP-GMA). To address multidimensional scaling, the Gower General Similarity Coefficient is applied. The cluster analysis is conducted using MVSP software (Version 3.13n, Kovach Computing Services). The result of clustering analysis is presented in section 3.2.

Table 2. The classification results of LVQ-ANN with different number of competitive layer neuron based on LVQ1 and LVQ2 learning algorithm.

Samples	Sample numbers	Learning algorithm	Identification rate of different numbers of competitive layer neuron numbers										
			20	21	22	23	24	25	26	27	28	29	30
Sect. <i>Furfuracea</i>	5	LVQ1	80.00%	20.00%	80.00	20.00%	60.00%	20.00%	40.00%	60.00%	20.00%	40.00%	20.00%
		LVQ2	0.00%	0.00%	0.00%	80.00	0.00%	80.00%	80.00%	0.00%	0.00%	20.00%	20.00%
Sect. <i>Paracamellia</i>	8	LVQ1	37.50%	37.50%	37.50%	37.50%	37.50%	37.50%	37.50%	37.50%	37.50%	37.50%	37.50%
		LVQ2	0.00%	75.00%	75.00%	75.00%	0.00%	75.00%	75.00%	0.00%	75.00%	75.00%	75.00%
Sect. <i>Tuberculata</i>	6	LVQ1	0.00%	16.67%	16.67%	16.67%	16.67%	16.67%	16.67%	0.00%	16.67%	0.00%	16.67%
		LVQ2	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Sect. <i>Camellia</i>	16	LVQ1	62.50%	68.75%	62.50%	68.75%	68.75%	68.75%	68.75%	68.75%	62.50%	62.50%	62.50%
		LVQ2	75.00%	75.00%	87.50%	68.75%	81.25%	93.75%	81.25%	81.25%	81.25%	75.00%	75.00%
Sect. <i>Theopsis</i>	10	LVQ1	90.00%	90.00%	90.00%	90.00%	90.00%	90.00%	90.00%	90.00%	90.00%	90.00%	90.00%
		LVQ2	100%	0.00%	0.00%	0.00%	100%	0.00%	0.00%	100%	0.00%	0.00%	0.00%
Total accuracy (%)		LVQ1	57.78%	55.56%	60.00%	55.56%	60.00%	55.56%	57.78%	57.78%	53.33%	53.33%	53.33%
		LVQ2	48.89%	40.00%	44.44%	46.67%	51.11%	55.56%	51.11%	51.11%	42.22%	48.89%	42.22%

doi:10.1371/journal.pone.0029704.t002

Table 3. Comparing best classification accuracy of LVQ1, LVQ2, DAN2 and RBF-SVM.

Class	LVQ1 Training	LVQ1 Testing	LVQ2 Training	LVQ2 Testing	DAN2 Training	DAN2 Testing	SVM Training	SVM Testing
Sect. <i>Furfuracea</i>	83.33% (5/6)	80% (4/5)	33.33% (2/6)	80% (4/5)	100% (6/6)	100% (5/5)	100% (6/6)	100% (5/5)
Sect. <i>Paracamellia</i>	62.5% (5/8)	37.50% (3/8)	75% (6/8)	75.00% (6/8)	100% (8/8)	87.50% (7/8)	100% (8/8)	100% (8/8)
Sect. <i>Tuberculata</i>	16.67% (1/6)	16.67% (1/6)	0% (0/6)	0% (0/6)	83.33% (5/6)	66.70% (4/6)	83.33% (5/6)	83.33% (5/6)
Sect. <i>Camellia</i>	94.44% (17/18)	62.50% (10/16)	100% (18/18)	93.75% (15/16)	100% (18/18)	93.75% (15/16)	100% (18/18)	100% (16/16)
Sect. <i>Theopsis</i>	80.00% (8/10)	90.00% (9/10)	0% (0/10)	0% (0/10)	100% (10/10)	100% (10/10)	100% (10/10)	100% (10/10)
Total Accuracy (%)	75.00% (36/48)	60.00% (27/45)	54.17% (26/48)	55.56% (25/45)	97.92% (47/48)	91.11% (41/45)	97.92% (47/48)	97.78% (44/45)

doi:10.1371/journal.pone.0029704.t003

Results

Leaf architecture data and related morphological data of samples

(Table S2) presents leaf architecture data for each *Camellia* species. The data shows that the leaf architecture for sect. *Furfuracea*, sect. *Paracamellia*, sect. *Tuberculata*, sect. *Camellia*, and sect. *Theopsis* are different. The most pronounced difference is in leaf vertical length (Table S2, column 20). Most vary from 5 to 10 cm; however, a few are closer to 15 cm (species in sect. *Furfuracea*), or less than 5 cm (sect. *Theopsis*). But for species in sect. *Paracamellia*, sect. *Tuberculata*, and sect. *Camellia*, the leaf vertical length values are diverse and vary widely. However, genus *Camellia* becomes a more natural group since it does have a series of common traits [6]. The common leaf architecture characteristics of the five sections are: leaf blade is symmetric, angulations between secondary veins and primary veins on upper part, on middle part, and on lower part is always at acute angle (Table S2, column 11–13), veinlets are 1–2 times branched (Table S2, column 16), and areoles development is incomplete (Table S2, column 17). As shown in (Table S2, columns 6–9, 14–15), there are differences in leaf venation characteristics such as the reticulate

veins (column 6), margin shape (column 7), margin spacing (column 8), secondary veins shape (column 9), the number of secondary veins variations in angle of divergence between primary and secondary veins (columns 14), number of secondary veins (column 15).

Cluster analysis based on leaf architecture data

The dendrogram resulting from hierarchical cluster analysis grouped the 93 species into two main clusters (Figure 4). Cluster 1 (C1), included all species of sect. *Theopsis*, and most species of sect. *Paracamellia*, Cluster 2 (C2) had all species of sect. *Furfuracea*, sect. *Tuberculata*, and most species of sect. *Camellia*. On closer inspection, C1 contained two subclusters: subcluster 1a (SC1a) comprised of all special species of sect. *Theopsis*, *C. semiserrata* which belongs to sect. *Camellia*, and *C. parvimurivata* which belongs to sect. *Tuberculata*. Subcluster 1b (SC1b) comprised of all species of *Paracamellia*. C2 contained the remaining three sections and differed from previous classification [10]. For instance, subcluster 2a (SC2a) and subcluster 2b (SC2b) are mainly comprised of sect. *Camellia* and sect. *Tuberculata*, indicating significant closeness with species affinities. Therefore, we suggest that sect. *Camellia* and sect. *Tuberculata* may be merged into one section. However, branch I (I)

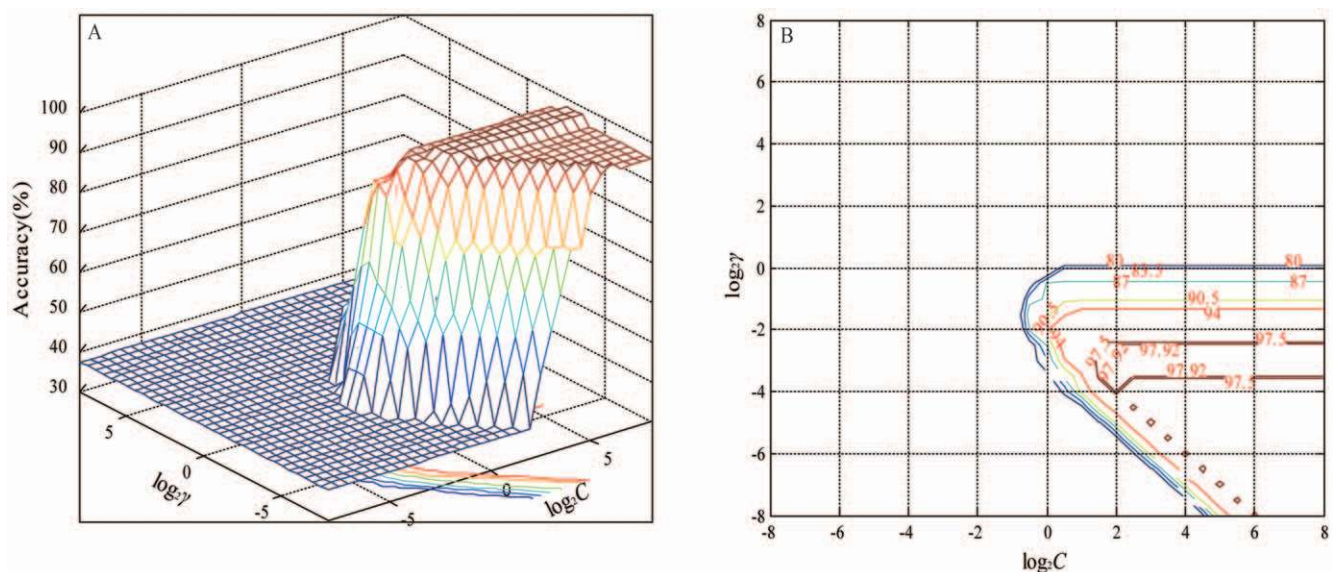


Figure 5. Classification accuracy in different kernel parameter (C) and regularization parameter (γ) by cross-validation. Three dimension diagram (A) and contour map (B). doi:10.1371/journal.pone.0029704.g005

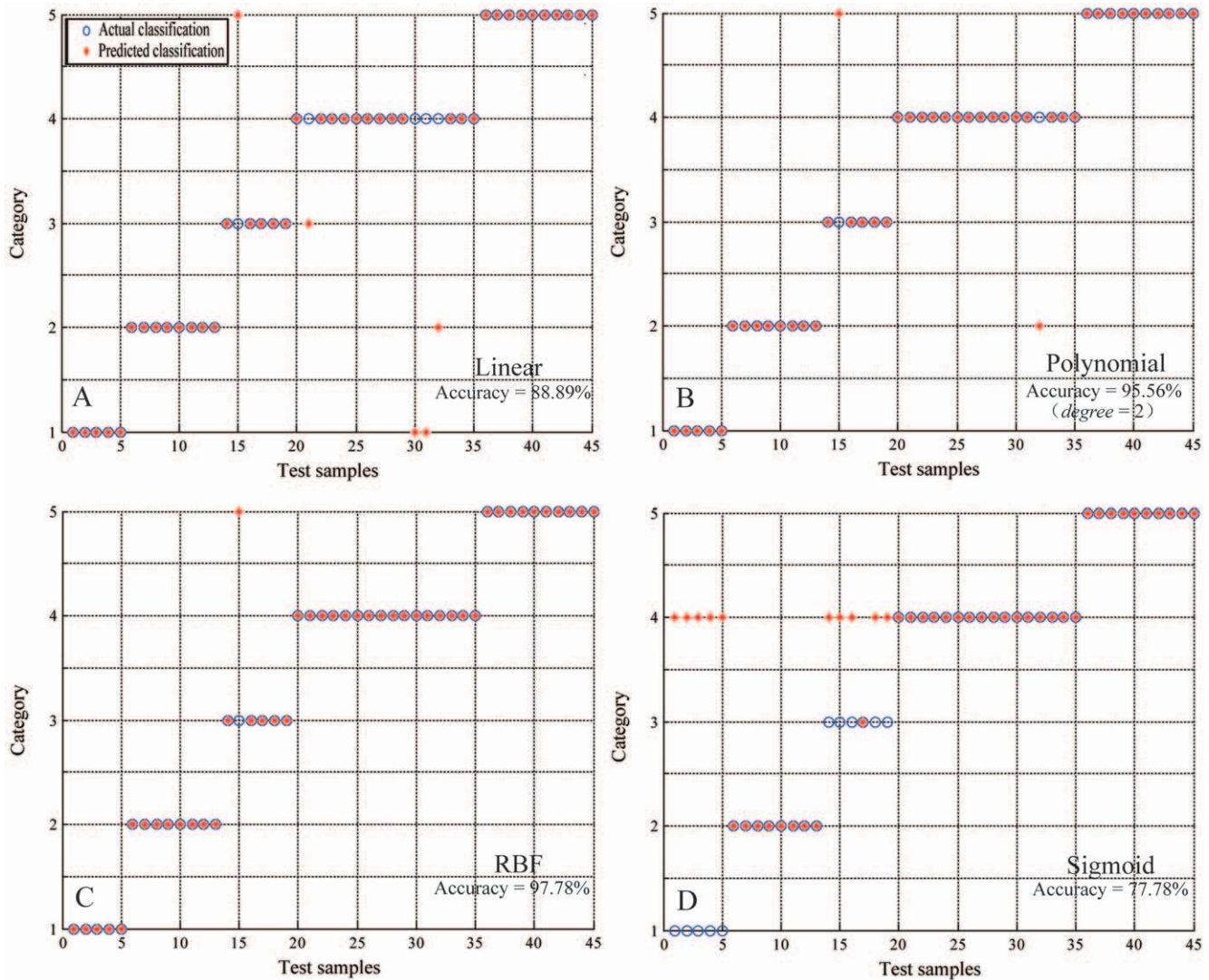


Figure 6. The classification results of linear, polynomial, RBF and sigmoid SVMs with the optimal parameters.
doi:10.1371/journal.pone.0029704.g006

and branch II (II) were clearly divided by species of sect. *Furfuracea* and some species of sect. *Camellia*. These results provide mostly the same categorization of genus *Camellia* as specified by Chang [10].

LVQ-ANN, DAN2, and SVMs classification based on leaf architecture data

(Table 2) shows the classification results for the Learning Vector Quantization neural network (LVQ-ANN) with different number of

Table 4. The classification results of the polynomial SVM with different degrees in the optimal parameters ($C = 2.828, \gamma = 0.088$).

Samples	Sample Numbers	Identification rate of polynomial classifiers in different degree							
		3	4	5	6	7	8	9	10
<i>Sect. Furfuracea</i>	5	100%	100%	100%	100%	100%	100%	100%	100%
<i>Sect. Paracamellia</i>	8	100%	87.50%	87.50%	87.50%	75.00%	75.00%	87.50%	87.50%
<i>Sect. Tuberculata</i>	6	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	66.67%	66.67%
<i>Sect. Camellia</i>	16	93.75%	93.75%	93.75%	93.75%	87.50%	81.25%	75.00%	68.75%
<i>Sect. Theopsis</i>	10	100%	100%	100%	100%	100%	100%	100%	100%
Total accuracy (%)		95.56%	93.33%	93.33%	93.33%	88.89%	86.67%	84.44%	82.22%

doi:10.1371/journal.pone.0029704.t004

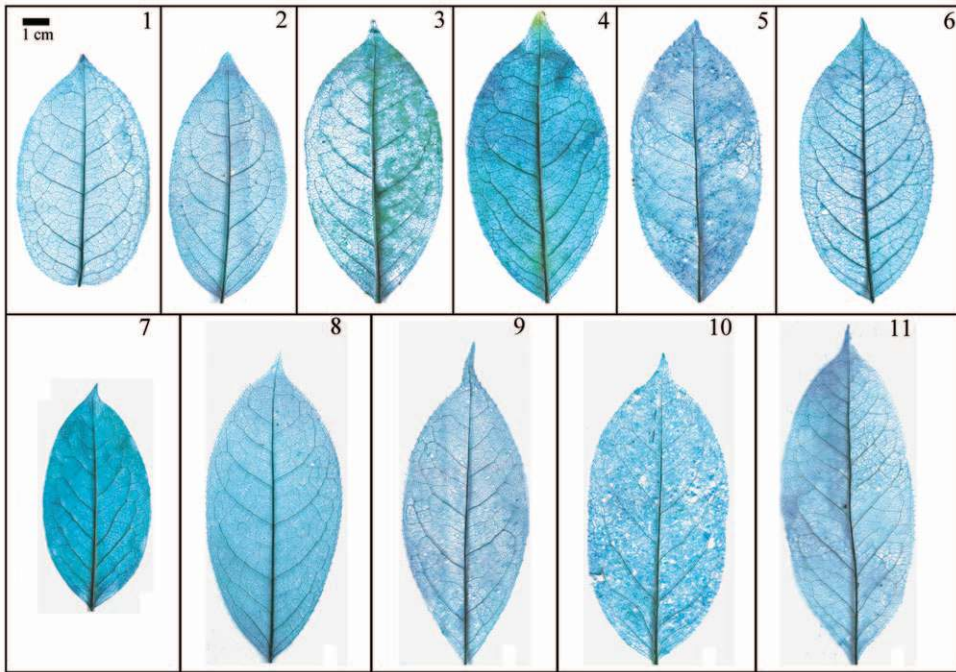


Figure 7. Leaf specimen of sect. *Furfuracea*. Numbers in figure corresponding species numbers in Table S1. Scale bar = 1 cm. doi:10.1371/journal.pone.0029704.g007

competitive layer neurons. Following suggestion of other researchers [32], and in order to achieve the best performance for the LVQ-

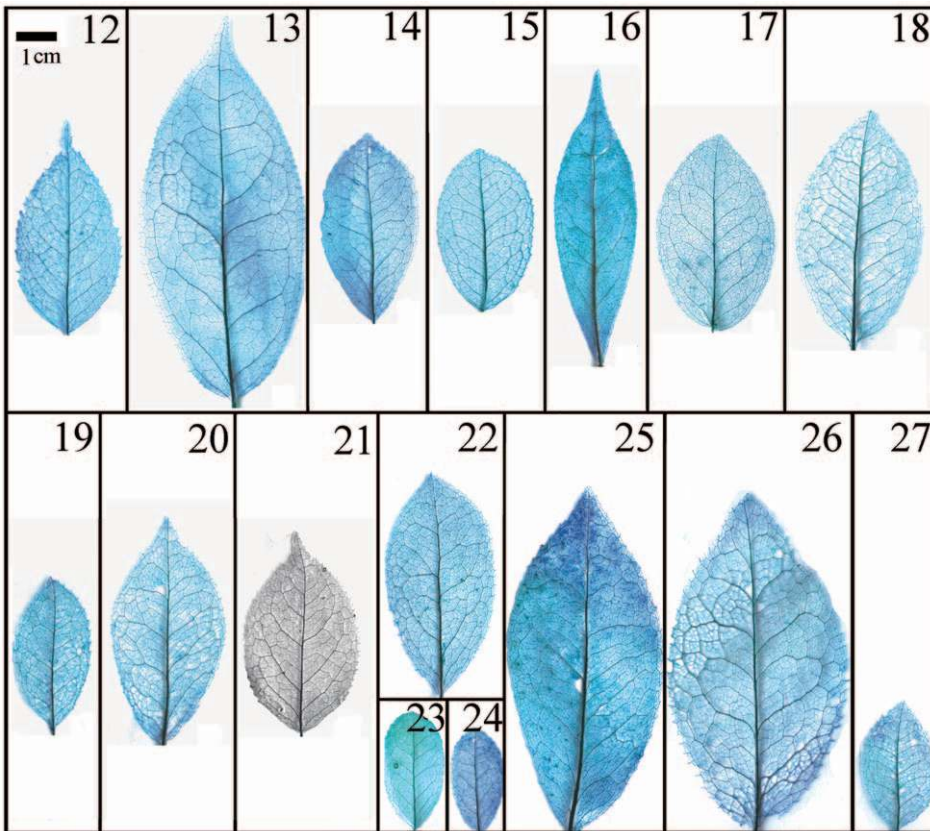


Figure 8. Leaf specimen of sect. *Paracamellia*. Numbers in figure corresponding species numbers in Table S1. Scale bar = 1 cm. doi:10.1371/journal.pone.0029704.g008

ANN algorithms, we experimented by varying the number of neurons in the competitive layer. Both the LVQ1 and LVQ2 learning algorithms reached their highest classification accuracy with the competitive layer neuron number set to 24 and 25, respectively. Comparing the two LVQ-ANN methods, revealed that LVQ1 learning algorithm produces a more accurate results, for both the training and testing data sets (75% and 60%), than LVQ2 (54.17% and 55.56%) (Table 3). Although, the classification of sect. *Theopsis* by LVQ2-ANN reached 100.00% accuracy, when the number of competitive layer neuron was 20, 24, or 27 (Table 2); overall, the classification results produced with LVQ1 learning algorithm were more stable, especially in the sect. *Theopsis* classification (accuracy of 90.00%), and the sect. *Paracamellia* (accuracy of 37.50%) (Table 3). Although LVQ-ANN does not provide acceptably accurate results for this data set, the advantage of this model is in its simplicity and the fact that the input data does not need to be normalized or orthogonalized. Thus, LVQ-ANN may be used as a simple control method for classification.

DAN2 is a dynamic neural network model that does not require model configuration or parameter optimization. DAN2's algorithm, at every iteration, solves a nonlinear minimization problem. Specifically, the nonlinear optimization strategy used in DAN2 estimates a nonlinear parameter. Like all nonlinear optimization methods for non-convex/non-concave functions, obtaining global optimization is never guaranteed. Similar to other optimization applications, choice of a good starting point can improve convergence to local optimum and beginning the search at various starting points can facilitate reaching multiple local optima. Ghiassi and Saidane [50] identify the starting point as

$F_0(X)$ and use the training data and the standard linear regression to obtain its value. In classification problems, $F(X)$ only takes binary (or integer) values, so in addition to the standard MLR; we have experimented with using a rudimentary kNN solution to obtain a good starting point. The kNN approach used is a simplified method that only considers one or two values for k ($k=1$ or $k=3$) to quickly obtain a starting point value. In this study, we use the exact same data sets used in the LVQ model to train DAN2 models. During DAN2 training, we iteratively reduce training error tolerance by specifying a SSE/MSE value. The model training stops either when it reaches this value or a predefined number of iterations. The value of this error level can be iteratively reduced to a desired level. The model uses internally defined metrics to avoid overfitting [51]. DAN2 model uses the "one-vs.-all" classification approach detailed in [53]. Results from this model are presented in (Table 3). The overall training and testing accuracy for this model is 97.92% and 91.11% respectively (sect. *Furfuracea*-100%, sect. *Paracamellia*-87.5%, sect. *Tuberculata*-66.67%, sect. *Camellia*-93.75%, and sect. *Theopsis*-100%, respectively, for the test data set). DAN2 model presents better results than LVQ models and does not require model configurations.

We next present results of using SVM for this analysis. To use the support vector machines (SVMs) model, and in order to obtain the best performance, the two SVM parameters of regularization (C) and kernel parameter (γ) are optimized using cross validation. Linear, polynomial, sigmoid, and radial basis function (RBF) kernel classifiers were tested in this study. We used the same input vectors, training and testing data sets for the LVQ models and the

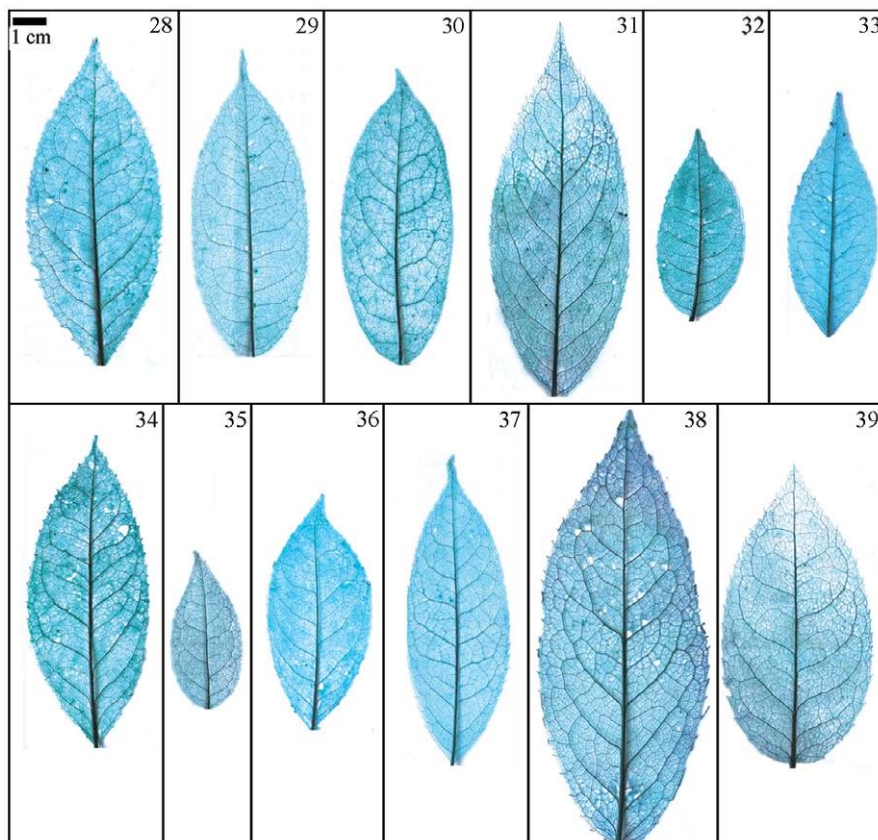


Figure 9. Leaf specimen of sect. *Tuberculata*. Numbers in figure corresponding species numbers in Table S1. Scale bar = 1 cm. doi:10.1371/journal.pone.0029704.g009

SVM models. As seen in (Figure 5), the highest accuracy of 97.92% was achieved when $C=2.828$ and $\gamma=0.088$ for the training data set. All SVM models are optimized by manipulating C and γ parameters to obtain the best training accuracy. (Figure 6) presents the classification results of all the SVM models, with optimal parameters for the test data sets. The linear kernel overall accuracy for the test data set is 88.89%. For the polynomial model, the degree parameter (d) ranges from three to ten. (Table 4) shows the classification results of polynomial SVM models with different degrees. The best results, 95.56% accuracy for the test data set, was obtained for $d=3$. The polynomial SVM classifier with polynomial degree $d=7$ had a classification rate of 88.89% for the test data set, which was similar to the linear model. (Table 4) also shows that classification accuracies do not improve for the polynomial degree larger than three. The sigmoid model performed less accurately than other SVM models. The overall accuracy for the test data for this model was only 77.78%. (Figure 6) shows that the RBF SVM classifier offers the best results with overall accuracy of 97.78% for the test data set (sect. *Furfuracea*-100%, sect. *Paracamellia*-100%, sect. *Tuberculata*-83.33%, sect. *Camellia*-100%, and sect. *Theopsis*-100%, respectively, for the test data sets).

Discussion

Taxonomical classification based on description of leaf morphology is an effective approach [13]. Leaf architecture has been the subject of several studies in taxonomy and evolutionary relationships of taxa with controversial genera [24]. The architectural properties of leaf venation patterns for systematic classification have also been studied [56,57,58]. Macrofossils studies have shown that the leaf venation patterns can be extensively utilized in identifying fossil taxa in palaeobotany [59]. The lamina morphological and venation character details of *Camellia* are also shown in the Figure 7 through 17. We found significant results using the leaf venation pattern for identifying various *Camellia* species indicating the importance of this tool for classification.

To identify and distinguish *Camellia* plants, the floras edited by botanists such as Chang [10] and Ming [6] are commonly used as a comprehensive resource [60]. Indented dichotomous keys in the literature are commonly used as the identification keys. When a new unknown species needs to be classified, we always turn to these floras, and the identification process often follows a predefined path with the observed characteristics. However, since the traditional information retrieval processes are tedious, the final

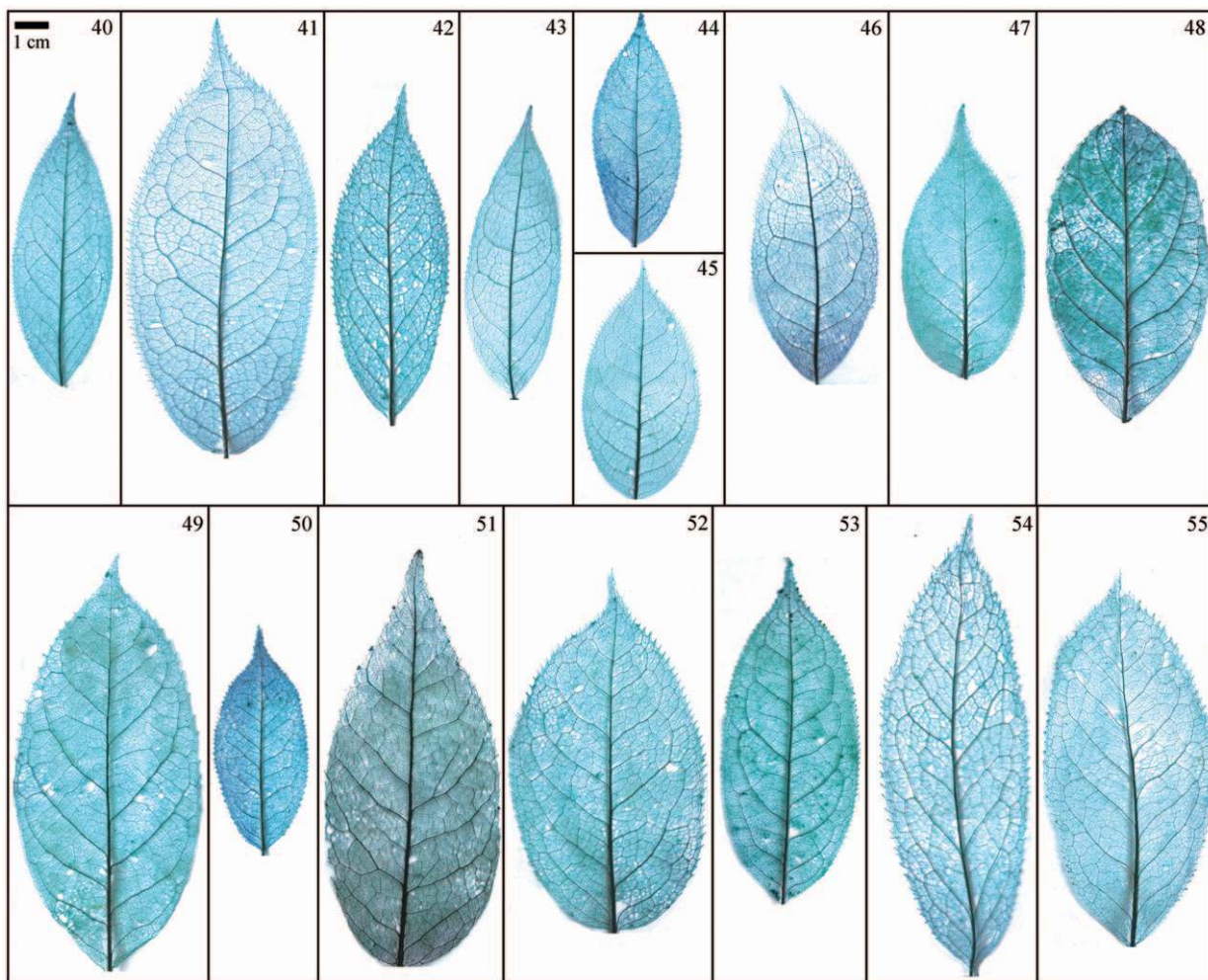


Figure 10. Leaf specimen of sect. *Camellia* (No. 40–55). Numbers in figure corresponding species numbers in Table S1. Scale bar = 1 cm.
doi:10.1371/journal.pone.0029704.g010

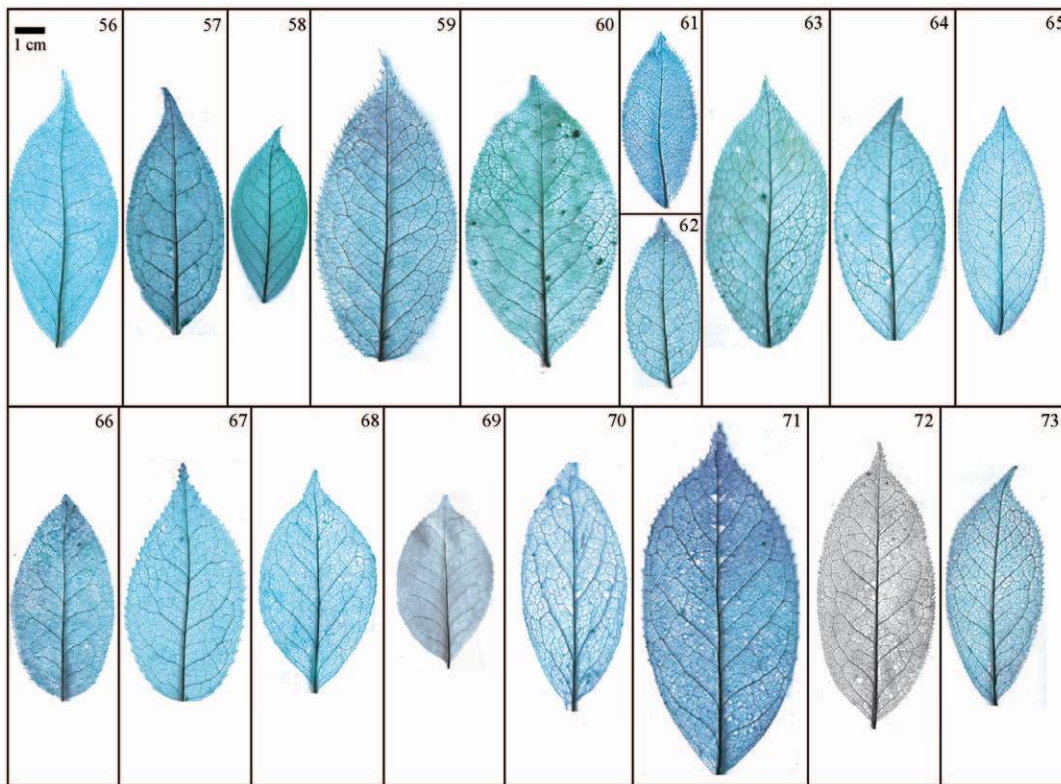


Figure 11. Leaf specimen of sect. *Camellia* (No. 56–73). Numbers in figure corresponding species numbers in Table S1. Scale bar= 1 cm.
doi:10.1371/journal.pone.0029704.g011

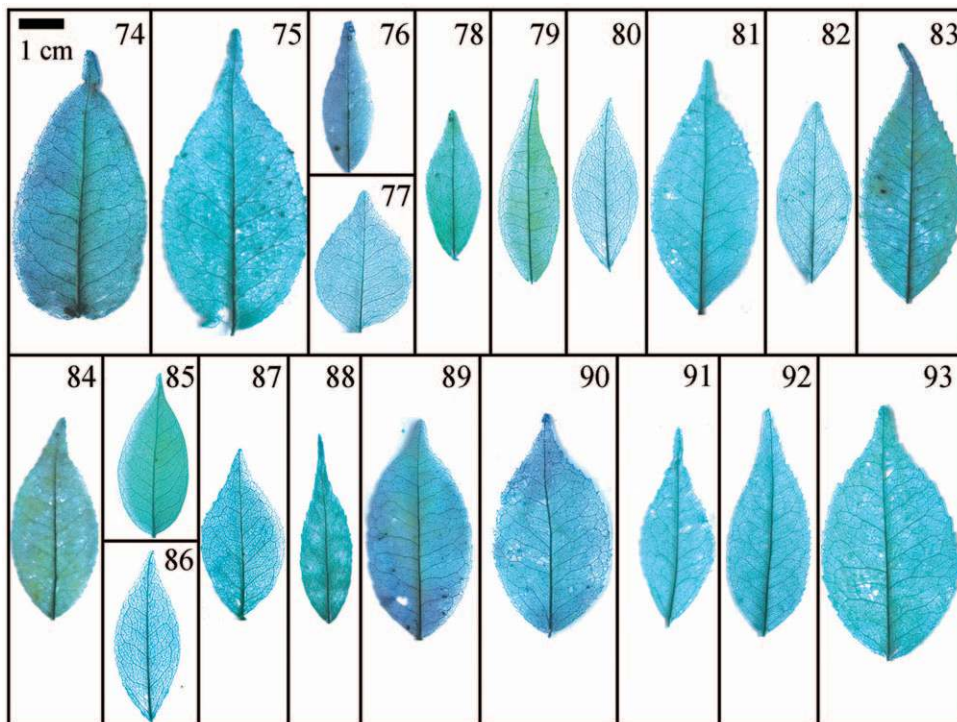


Figure 12. Leaf specimen of sect. *Theopsis*. Numbers in figure corresponding species numbers in Table S1. Scale bar= 1 cm.
doi:10.1371/journal.pone.0029704.g012

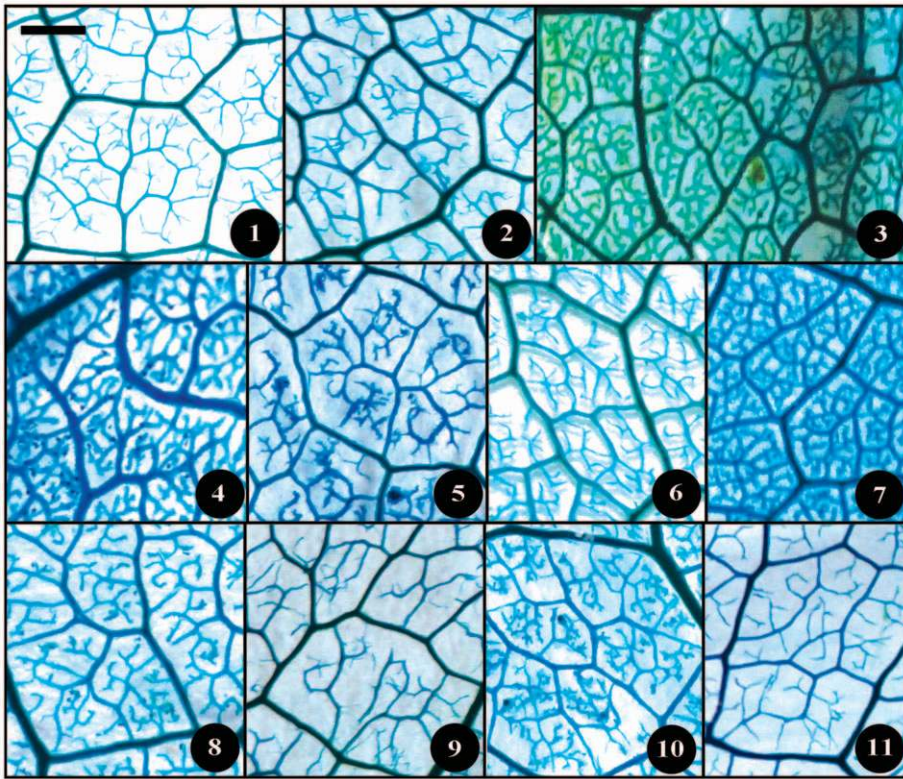


Figure 13. Detail venation characters of sect. *Furfuracea*. Numbers in figure corresponding species numbers in Table S1. Scale bar = 1 mm.
doi:10.1371/journal.pone.0029704.g013

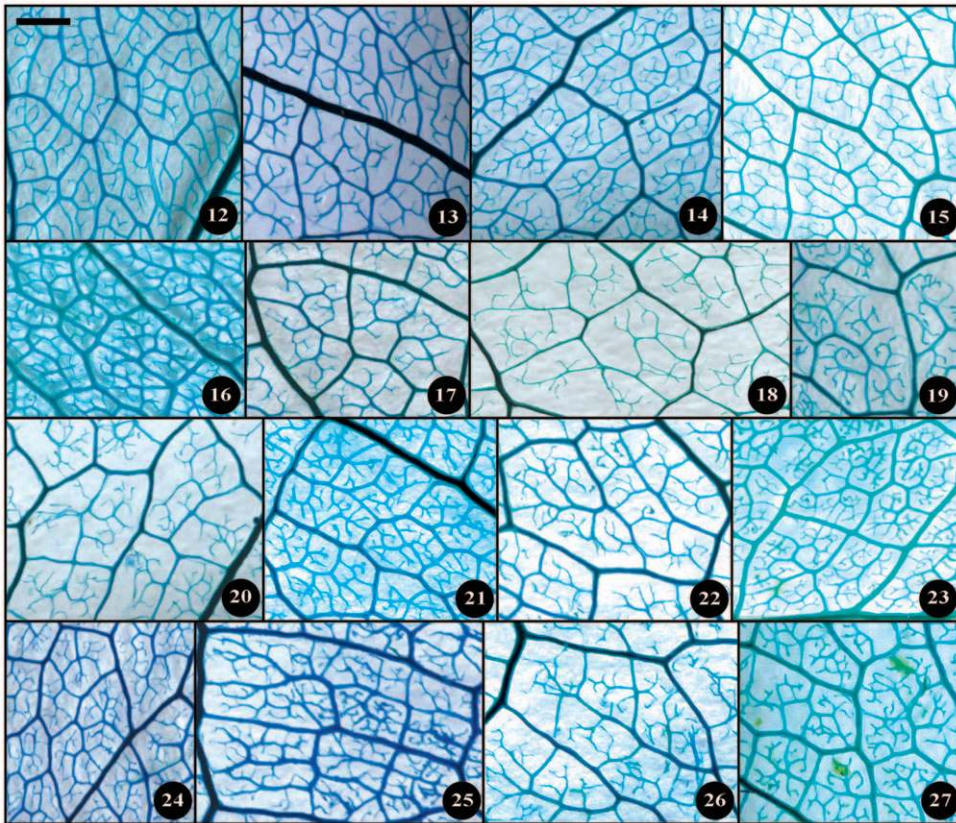


Figure 14. Detail venation characters of sect. *Paracamellia*. Numbers in figure corresponding species numbers in Table S1. Scale bar = 1 mm.
doi:10.1371/journal.pone.0029704.g014

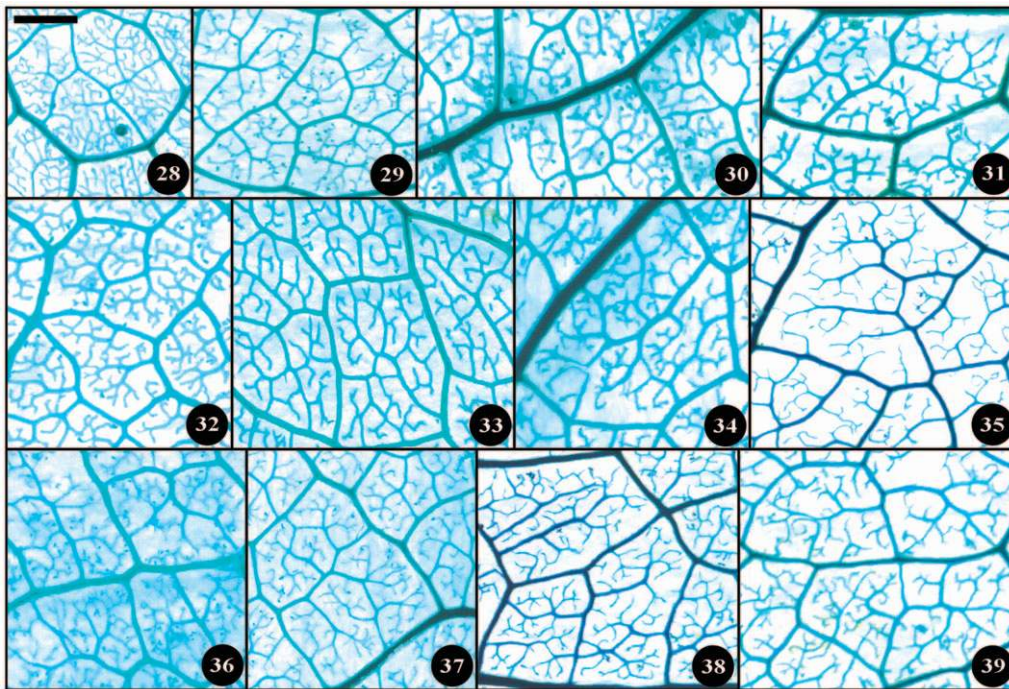


Figure 15. Detail venation characters of sect. *Tuberculata*. Numbers in figure corresponding species numbers in Table S1. Scale bar = 1 mm.
doi:10.1371/journal.pone.0029704.g015

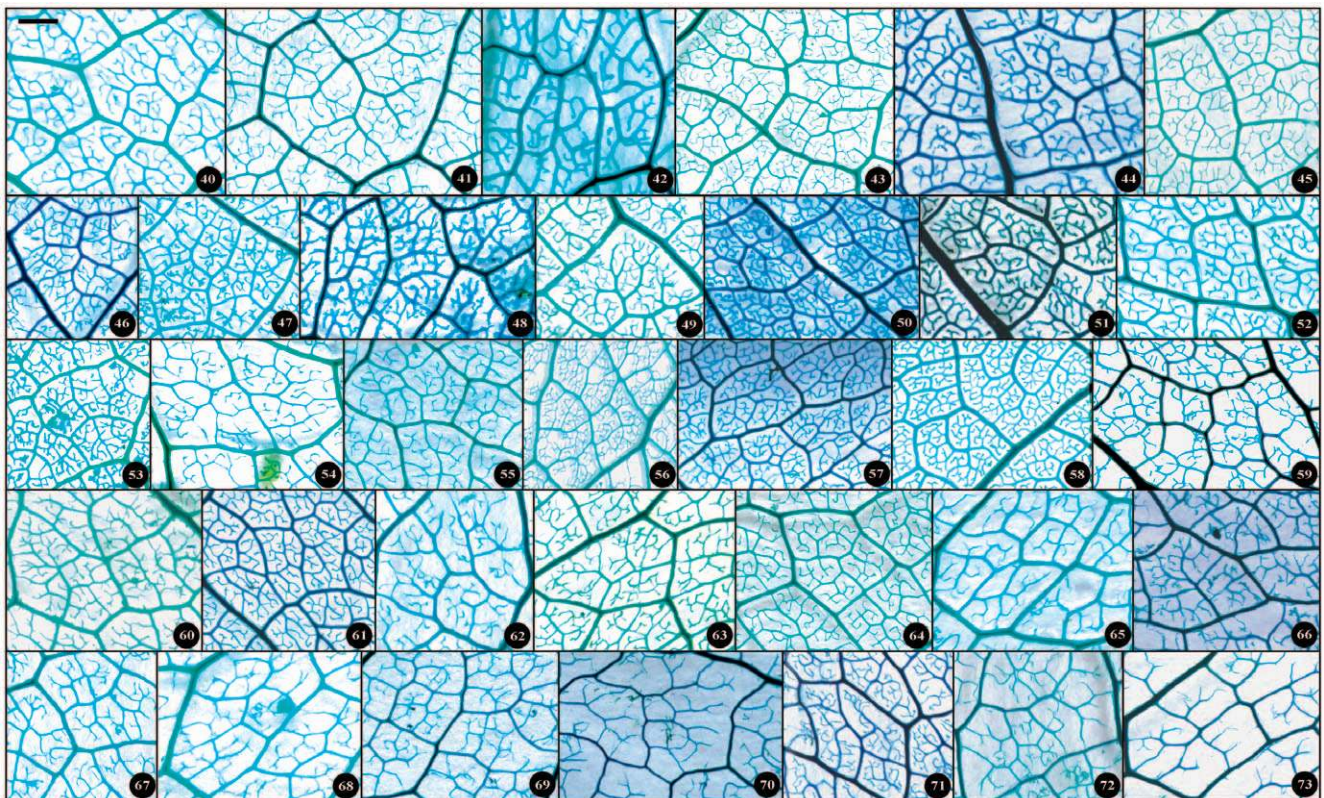


Figure 16. Detail venation characters of sect. *Camellia*. Numbers in figure corresponding species numbers in Table S1. Scale bar = 1 mm.
doi:10.1371/journal.pone.0029704.g016

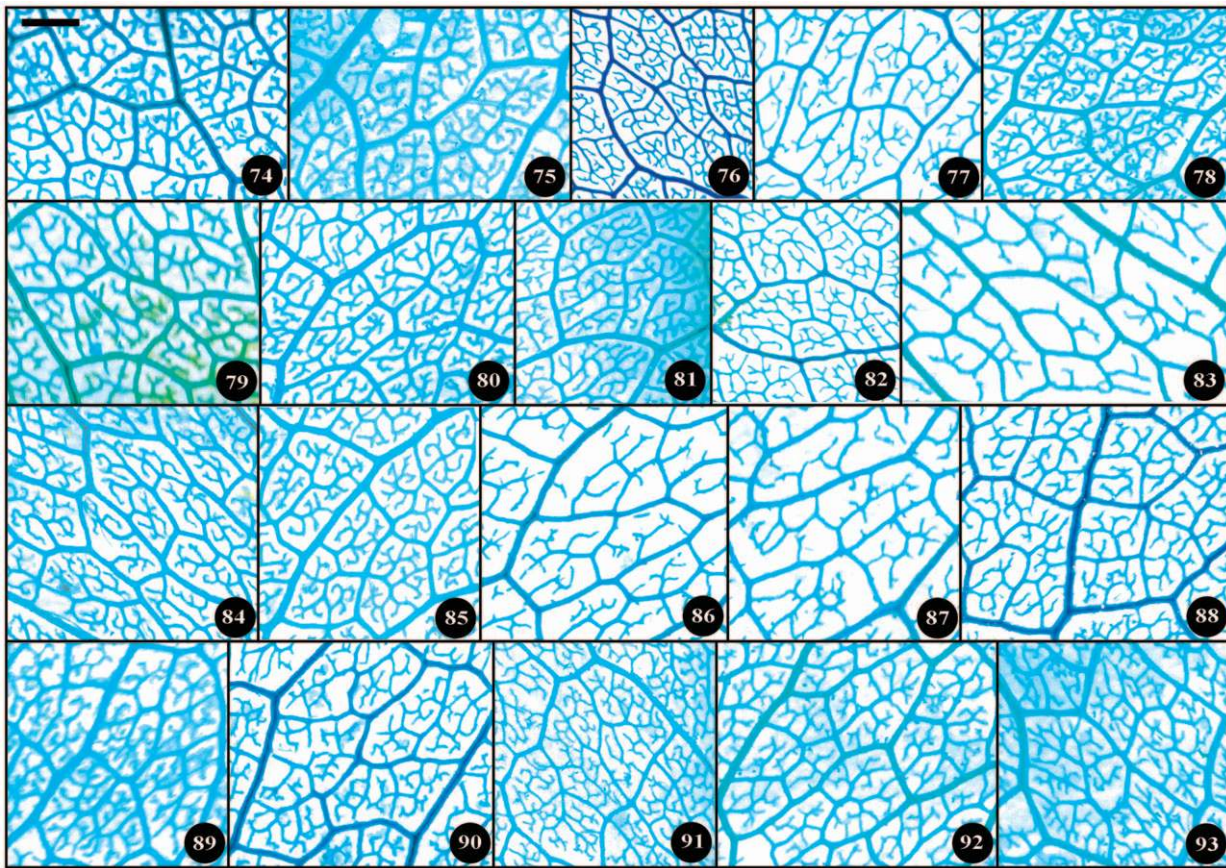


Figure 17. Detail venation characters of sect. *Theopsis*. Numbers in figure corresponding species numbers in Table S1. Scale bar = 1 mm.

doi:10.1371/journal.pone.0029704.g017

classification is often subjective. Clustering and pattern recognition techniques, especially DAN2 and SVM, used in this research, are shown to be an effective and objective classification tools that can be used to classify new species. We present results from using these tools along with the leaf architecture data for classifying 93 *Camellia* species (Table S2).

The classification of species from our dendrogram is mostly in agreement with previous research, indicating that the discrimination of these species by leaf architecture data reflects their phylogenetic relations. In this discussion, we compare and contrast our results from applying cluster analysis and pattern recognition methods using leaf architecture-based data, with existing classifications. Specifically, we compare results from the cluster analysis, and the two pattern recognition methods with the best results (DAN2, and the RBF-SVM) with those of Chang [10] and Ming [6].

Analysis of leaf characters data has been successfully employed to investigate plant taxonomy. Our study suggests that leaf architecture-based *Camellia* classification using pattern recognition techniques can be used to discriminate plants at the genus level. In this study, the results of cluster analysis using leaf architecture data mainly support Chang's [10] classification of *Camellia*. However, our results continue to strengthen the controversy about a number of species.

The separation of 93 species in the dendrogram obtained in this study using clustering analysis (Figure 4) was mostly in agreement with the taxonomy of Chang [10]. However, as illustrated in (Figure 4), *C. weiningensis* (No. 25) has similar attributes with species

belonging to sect. *Camellia*. This finding makes it reasonable to merge *C. weiningensis* into sect. *Camellia*; thus, validating Ming's [6] classification of the *C. weiningensis*. Similarly, Chang [10] classifies *C. semiserrata* (No. 49) to belong to sect. *Camellia*, and *C. parvimuricata* (No. 35) to belong to sect. *Tuberculata*. We find these two species (Nos. 35 & 49) to have closer relationship with sect. *Theopsis*. Therefore, we find it more reasonable to merge them into sect. *Theopsis*. In addition, Chang's taxonomic treatment advocates sect. *Tuberculata* and sect. *Camellia* as two independent sets. However, as depicted in (Figure 4), species of sect. *Tuberculata* are closer to sect. *Camellia*. We recommend that they should be merged into one section. For sect. *Furfuracea*, all species are grouped together, validating Chang's taxonomic treatment. We disagree with Ming's [6] suggestions that sect. *Furfuracea* should be canceled and that its species arrangements should be adjusted. Studies of Ming [6] suggest that the *C. hiemalis* (No. 18) species should be classified as a variant of *C. sasanqua* (belonging to sect. *Oleifera*), whereas our hierarchical dendrogram based on leaf architecture data shows *C. hiemalis* to be similar to the species of sect. *Paracamellia* and does not support merging of *C. hiemalis* into *C. sasanqua*. Our findings support Chang's [10] treatments of these two species. Moreover, *C. obovate* (No. 6) and *C. parafurfuracea* (No. 10) are classified as one species class by Ming [6]. Our study shows that the bases of *C. obovate* and *C. parafurfuracea* are round and both species have similar leaf architecture characteristics. Our cluster analysis reaffirms Ming's treatment of these two species so it is reasonable to consider *C. obovate* and *C. parafurfuracea* as one species. The two species *C. parvilimba* var. *brevipes* (No. 87) and *C. parvilimba* (No. 86) are very

similar, indicating a high degree of homogeneity. For these species, we agree with Chang in considering *C. parvilimba* var. *brevipes* as a variety of *C. parvilimba*. These results augment the usefulness of leaf architecture data for plant taxonomic treatments. We also note that deviation from the classification needs to be further investigated to see if a misclassification is due to the underlying algorithm's fitting of the data or Chang's [10] designation of the species.

In analyzing results from the pattern recognition techniques, we note that although LVQ-ANN did not produce very accurate results, when comparing this approach with other ANNs, LVQ has the advantage that it can classify any set of input vectors, has a fast learning algorithm [61] and is used extensively in the literatures [62,63,64,65,66,67,68].

In analyzing DAN2 results (Table 3), we note that all species of sect. *Furfuracea* and Sect. *Theopsis* conform to Chang's classifications. In sect. *Paracamellia*, DAN2 suggests the *C. winingensis* (No. 25) species to belong to sect. *Camellia*. This result is similar to the clustering algorithm's results and we disagree with Chang's classification. We suggest *C. winingensis* to belong to sect. *Camellia*, and agree with Ming's taxonomic results. Additionally, our cluster analysis show that sect. *Camellia* and sect. *Tuberculata* have significant closeness with species affinities. DAN2 results classify *C. hupehensis* (No. 36), *C. zengii* (No. 37) and *C. crassifolia* (No. 39) species to belong to sect. *Camellia*. This conclusion validates Chang's view about the close evolutionary relationship between sect. *Camellia* and sect. *Tuberculata*. Furthermore, this shows that *C. hupehensis*, *C. zengii* and *C. crassifolia* may indeed have underlying links in biological evolutionary principles with species of sect. *Camellia*. This finding emphasizes the need for further research in this branch.

In analyzing the classification results from the SVM approach (Figure 6), we note that the *C. fluviatilis* (No. 16) in sect. *Paracamellia* was incorrectly identified by all SVM classifiers. This specie was incorrectly identified as sect. *Theopsis* by linear, polynomial ($d=2$), and RBF classifiers and as sect. *Camellia* by sigmoid classifier. The results suggest *C. fluviatilis* to be similar to the species of sect. *Theopsis* or sect. *Camellia*. We also note that in the clustering analysis, (Figure 4) shows *C. fluviatilis* to have closer relation with sect. *Camellia*. Therefore, it may be more reasonable to merge it into sect. *Camellia* rather than merging it into sect. *Paracamellia* as suggested by Chang [10]. Finally, the RBF-SVM classifier offers the best conformance to Chang's classification, validating its effectiveness as a classification tool for plants.

In general, for this data set, the SVM approach shows better generalization than LVQ-ANN and DAN2. As pointed out by Pandolfi et al. [38], success of ANN methods usually depends on the quantity, validity, and accuracy of training data. However, other researchers have shown SVM to perform well for ill-posed problems with few training records [45,69,70,71]. Our results confirm this property of SVM. The RBF-SVM kernel used in this study offers the best results by conforming to Chang [10] classification. However, it should be noted that using Chang's classification as a reference is controversial and literature suggests variation from this classification. Although DAN2 displayed lower

classification accuracy in conformance to Chang's, we cannot dismiss the correctness of its results. Taxonomy is a dynamic field and existing theory does not support 100% accuracy of any classification due in part to the fact that taxonomic treatments based on different features may generate different results. Therefore, it should not be surprising to see some divergences among different tools such as those observed in using DAN2 and RBF-SVM models in comparison with *Camellia* taxonomic systems of Chang [10] and Ming [6]. Overall results from using the leaf architecture data combined with pattern recognition and discrimination methods (LVQ-ANN, DAN2, and SVM), is shown to be an effective tool for identification of genus *Camellia*.

Conclusion

In conclusion, lamina morphological and venation characters of 93 species in five sections (sect. *Furfuracea*, sect. *Paracamellia*, sect. *Tuberculata*, sect. *Camellia*, and sect. *Theopsis*) are reported. The hierarchical dendrogram based on leaf architecture data confirms the morphological classification of the five sections proposed by Chang's taxonomic treatment. LVQ-ANN, DAN2, and SVMs models based on the 31 leaf architecture attributes were constructed. In LVQ-ANN models, the best classification accuracy is 60.00% for the test data set when number of competitive layer neuron is 22 or 24 using the LVQ1 learning algorithm. The best DAN2 model offers a classification accuracy of 91.11% for the test data. In SVM models, the best classification accuracy is 97.78% using the RBF SVM classifier with $C=2.828$ and $\gamma=0.088$. The overall results indicate that leaf architecture analysis using pattern recognition tools, especially DAN2 and SVM algorithms, can be effectively used to distinguish the *Camellia* genus and other plant taxa.

Supporting Information

Appendix S1 Collection localities and vouchers of studied specimens. (DOC)

Table S1 Species studied, as classified by Chang (1998). (DOC)

Table S2 Data matrix of characteristics on leaf architecture of genus *Camellia*. (DOC)

Acknowledgements

Authors are grateful to S. S. Hong, W. Zhang, S. M. Zhang and F. Fang for assistance in species collection.

Author Contributions

Conceived and designed the experiments: HL WJ MG. Performed the experiments: HL WJ MG. Analyzed the data: WJ MG SL. Contributed reagents/materials/analysis tools: HL WJ MG SL MN. Wrote the paper: WJ MG MN.

References

- Lu HF, Pi EX, Peng QF, Wang LL, Zhang CJ (2009) A particle swarm optimization-aided fuzzy cloud classifier applied for plant numerical taxonomy based on attribute similarity. *Expert Systems with Applications* 36: 9388–9397.
- Antonios P (2005) World tea production reaches new high. *FAO Newsroom*, 14 July 2005. Food and Agriculture Organization of the United Nations, Rome, Italy. Website: <http://www.fao.org/newsroom/EN/news/2005/105404/index.html>. (Last accessed on 5 May 2011).
- Chen LH, Wang SZ, Nelson M (2005) Genetic diversities within *Camellia* species confirmed by random amplified polymorphic DNA (RAPD) markers. *Journal of the American Society for Horticultural Science* 40: 993–1147.
- Chen L, Zhou ZX, Yang YJ (2007) Genetic improvement and breeding of tea plant (*Camellia sinensis*) in China: from individual selection to hybridization and molecular breeding. *Euphytica* 154: 239–248.
- Vijayan K, Zhang WJ, Tsou CH (2009) Molecular taxonomy of *Camellia* (Theaceae) inferred from nrITS sequences. *American Journal of Botany* 96: 1348–1360.

6. Ming TL (2000) Monograph of the Genus *Camellia*. Kunming, China: Yunnan Science and Technology Press.
7. Gao JY, Parka CR, Du YQ (2005) Collected species of the genus *Camellia* — an illustrated outline. Hangzhou, Zhejiang, China: Zhejiang Science and Technology Press.
8. Sealy JR (1958) A Revision of the Genus *Camellia*. London: Royal Horticultural Society.
9. Chang HT, Bartholomew B (1984) *Camellias*. Portland, Oregon, USA: Timber Press.
10. Chang HT (1998) *Camellia*. In: , Fl. Reipubl. Popularis Sin. (eds). Flora. Science Press, Beijing, China.
11. Lu HF, Jiang B, Shen ZG, Shen JB, Peng QF, et al. (2008a) Comparative leaf anatomy, FTIR discrimination and biogeographical analysis of *Camellia* section *Tuberculata* (Theaceae) with a discussion of its taxonomic treatments. *Plant Systematics and Evolution* 274: 223–235.
12. Lu HF, Shen JB, Lin XY, Fu JL (2008b) Relevance of Fourier transform infrared spectroscopy and leaf anatomy for species classification in *Camellia* (Theaceae). *Taxon* 57: 1274–1288.
13. Pi EX, Peng QF, Lu HF, Shen JB, Du YQ, et al. (2009) Leaf morphology and anatomy of *Camellia* section *Camellia* (Theaceae). *Botanical Journal of the Linnean Society* 159: 456–476.
14. Jiang B, Peng QF, Shen ZG, Moller M, Pi EX, et al. (2010) Taxonomic treatments of *Camellia* (Theaceae) species with secretory structures based on integrated leaf characters. *Plant Systematics and Evolution* 290: 1–20.
15. Pandolfi C, Messina G, Mugnai S, Azzarello E, Masi E, et al. (2009) Discrimination and identification of morphotypes of *Banksia integrifolia* (Proteaceae) by an Artificial Neural Network (ANN), based on morphological and fractal parameters of leaves and flowers. *Taxon* 58: 925–933.
16. Mugnai S, Pandolfi C, Azzarello E, Masi E, Mancuso S (2008) *Camellia japonica* L. genotypes identified by an artificial neural network based on phyllometric and fractal parameters. *Plant systematics and evolution* 270: 95–108.
17. Yang JB, Li HT, Yang SX, Li DZ, Yang YY (2006) The application of four DNA sequences to studying molecular phylogeny of *Camellia* (Theaceae). *Acta Botanica Yunnanica* 28: 108–114. (in Chinese, with English abstract).
18. Linnaeus C (1753) *Species plantarum*. Stockholm: Salvia.
19. Meade C, Parnell J (2003) Multivariate analysis of leaf shape patterns in Asian species of the *Uvaria* group (Annonaceae). *Botanical Journal of the Linnean Society* 143: 231–242.
20. Plotze RDO, Falvo M, Pádua JG, Bernacci LC, Vieira ML, et al. (2005) Leaf shape analysis using the multiscale Minkowski fractal dimension, a new morphometric method: A study with *Passiflora* (Passifloraceae). *Canadian Journal of Botany* 83: 287–301.
21. Ye P, Weng GR (2011) Classification and recognition of plant leaf based on neural networks. *Key Engineering Materials* 464: 38–42.
22. Lin XY, Peng QF, Lv HF, Du YQ, Tang BY (2008) Leaf anatomy of *Camellia* Sect. *Oleifera* and Sect. *Paracamellia* (Theaceae) with reference to their taxonomic significance. *Journal of Systematics and Evolution* 46: 183–193.
23. Hickey LJ (1973) Classification of the architecture of dicotyledonous leaves. *American Journal of Botany* 60: 17–33.
24. Premoli AC (2008) Leaf architecture of South American *Nothofagus* (Nothofagaceae) using traditional and new methods in morphometrics. *Botanical Journal of the Linnean Society* 121: 25–40.
25. Melville R (1976) The terminology of leaf architecture. *Taxon* 25: 549–561.
26. Hickey LJ (1979) *A revised classification of the architecture of dicotyledonous leaves*. In: Metcalf CR, Chalk L, eds. *Anatomy of the dicotyledons*. Oxford: Clarendon Press. pp 25–39.
27. Rao VS, Shenoy KN, Inamdar JA (1980) Clearing and staining technique for leaf architectural studies. *Microsc. Acta* 83: 307–311.
28. Calvillo-Canadell L, Cevallos-Ferriz SRS (2002) *Bauhis moranii* gen. et sp. nov. (Cercideae, Caesalpinieae), an Oligocene plant from Tepexi de Rodríguez, Puebla, Mex., with leaf architecture similar to *Bauhinia* and *Cercis*. *Review of Palaeobotany and Palynology* 122: 171–184.
29. Jain AK, Murty MN, Flynn PJ (1999) Data Clustering: A Review. *ACM Computing Surveys (CSUR)* 31: 264–3239.
30. Foody GM, Mathur A (2006) The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM. *Remote Sensing of Environment* 103: 179–189.
31. Lu HF, Zheng H, Hu Y, Lou HQ, Kong XC (2011) Bruise detection on red bayberry (*Myrica rubra* Sieb. & Zucc.) using fractal analysis and support vector machine. *Journal of Food Engineering* 104: 149–153.
32. Roggo Y, Duponchel L, Huvette LP (2003) Comparison of supervised pattern recognition methods with McNemar's statistical test: Application to qualitative analysis of sugar beet by near-infrared spectroscopy. *Analytica Chimica Acta* 477: 187–200.
33. Chen QS, Zhao JW, Lin H (2009) Study on discrimination of Roast green tea (*Camellia sinensis* L.) according to geographical origin by FT-NIR spectroscopy and supervised pattern recognition. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 72: 845–850.
34. Bila S, Harkouss Y, Ibrahim M, Rousset J, N'Goya E, et al. (1999) An accurate wavelet neural-network-based model for electromagnetic optimization of microwave circuits. *International Journal of RF and Microwave Computer-Aided Engineering* 93: 297–306.
35. Lu HF, Zheng H, Lou HQ, Jiang LL, Chen Y, et al. (2010) Using neural networks to estimate the losses of ascorbic acid, total phenols, flavonoid, and antioxidant activity in *Asparagus* during thermal treatments. *Journal of Agricultural Food Chemistry* 58: 2995–3001.
36. Zheng H, Fang SS, Lou HQ, Chen Y, Jiang LL, et al. (2011) Neural network prediction of ascorbic acid degradation in green asparagus during thermal treatments. *Expert Systems with Applications* 38: 5591–5602.
37. Zheng H, Lu HF (2011) Use of kinetic, Weibull and PLSR models to predict the retention of ascorbic acid, total phenols and antioxidant activity during storage of pasteurized pineapple juice. *LWT-Food Science and Technology* 44: 1273–1281.
38. Pandolfi C, Mugnai S, Azzarello E, Bergamasco S, Masi E, et al. (2009) Artificial neural networks as a tool for plant identification: a case study on Vietnamese tea accessions. *Euphytica* 166: 411–421.
39. Chen QS, Guo ZM, Zhao JW (2008) Identification of green tea's (*Camellia sinensis* (L.)) quality level according to measurement of main catechins and caffeine contents by HPLC and support vector classification pattern recognition. *Journal of Pharmaceutical and Biomedical Analysis* 48: 1321–1325.
40. Dong XW, Liu YL, Yan JY, Jiang CY, Chen J, et al. (2008) Identification of SVM-based classification model, synthesis and evaluation of prenylated flavonoids as vasorelaxant agents. *Bioorganic & Medicinal Chemistry* 16: 8151–8160.
41. Zhao JW, Ouyang Q, Chen QS, Wang JH (2010) Detection of bruise on pear by hyperspectral imaging sensor with different classification algorithms. *Sensor Letters* 8: 570–576.
42. Zheng H, Lu HF, Zheng YP, Lou HQ, Chen CQ (2010) Automatic sorting of Chinese jujube (*Zizyphus jujuba* Mill. cv. 'hongxing') using chlorophyll fluorescence and support vector machine. *Journal of Food Engineering* 101: 402–408.
43. Cortes C, Vapnik V (1995) Support-vector networks. *Machine learning* 20: 273–297.
44. Vapnik V (1995) *The Nature of Statistical Learning Theory*. NY, New York, USA: Springer-Verlag.
45. Burges CJC (1998) A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery* 2: 121–167.
46. Zhang XH, Xia NH (2007) Leaf architecture of subtribe Micheliinae (Magnoliaceae) from China and its taxonomic significance. *Acta Phytotaxonomica Sinica* 45: 167–190.
47. Kohonen T, Barna G, Chrisley R (1988) Statistical pattern recognition with neural networks: benchmarking studies. *Neural Network, IEEE International Conference on* 1: 61–68.
48. Gayatri M, Kumar A, Janghu M, Kaur M, Prasad TV (2010) Implementation of epileptic EEG using recurrent neural network. *International Journal of Computer Science and Network Security* 10: 290–295.
49. Umer MF, Khayal MSH (2007) Classification of Textual Documents using Learning Vector Quantization. *Information Technology Journal* 6: 154–159.
50. Ghiassi M, Saidane H (2005) A dynamic architecture for artificial neural network. *Neurocomputing* 63: 397–413.
51. Ghiassi M, Saidane H, Zimbra DK (2005) A dynamic artificial neural network model for forecasting time series events. *International Journal of Forecasting* 21: 341–362.
52. Ghiassi M, Nangoy S (2009) A dynamic artificial neural network model for forecasting nonlinear processes. *Computers & Industrial Engineering* 57: 287–297.
53. Ghiassi M, Burnley C (2010) Measuring effectiveness of a dynamic artificial neural network algorithm for classification problems. *Expert Systems with Applications* 37: 3118–3128.
54. Chen QS, Zhao JW, Fang CH, Wang DM (2007) Feasibility study on identification of green, black and Oolong teas using near-infrared reflectance spectroscopy based on support vector machine (SVM). *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 66: 568–574.
55. Chang CC, Lin CJ (2001) LIBSVM: a library for support vector machines. Software available at: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>. (Last accessed on 15 May 2011).
56. Zetter R (1984) Morphologische Untersuchungen an *Fagus-Blättern* aus dem Neogen von Österreich. *Beiträge zur Palaontologie von Österreich* 11. Wien.
57. Kohler E (1993) *Blattneratur-Muster der Buxaceae Dumortier und Simmondsiaceae Van Tieghem*. *Feddes Repert* 104: 145–167.
58. 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.
59. Cleal CJ (1981) A new species of *Neuropteris* from the middle Westphalian of Palencia. *Estudios geológicos* 37: 77–82.
60. Pi EX, Lu HF, Jiang B, Huang J, Peng QF, et al. (2011) Precise plant classification within genus level based on simulated annealing aided cloud classifier. *Expert systems with applications* 38: 3009–3014.
61. Olmez T, Dokur Z (2003) Classification of heart sounds using an artificial neural network. *Pattern Recognition Letters* 24: 617–729.
62. Alirezaic J, Jernigan ME, Nahmias C (1997) Neural network-based segmentation of magnetic resonance images of the brain. *Nuclear Science, IEEE Transactions on* 44: 194–198.
63. Goren-Bar D, Kuflik T, Lev D, Shoval P (2001) Automating personal categorization using artificial neural networks. *Lecture Notes in Computer Science* 2109: 188–198.

64. Kusumoputro B, Budiarto H, Jatmiko W (2002) Fuzzy-neuro LVQ and its comparison with fuzzy algorithm LVQ in artificial odor discrimination system. *ISA Transactions* 41: 395–407.
65. Zhang QY, Xie CS, Zhang SP, Wang AH, Zhu BL, et al. (2005) Identification and pattern recognition analysis of Chinese liquors by doped nano ZnO gas sensor array. *Sensors and Actuators B: Chemical* 110: 370–376.
66. Berrueta LA, Alonso-Salces RM, Heberger K (2007) Supervised pattern recognition in food analysis. *Journal of Chromatography A* 1158: 196–214.
67. Liu ZY, Wu HF, Huang JF (2010) Application of neural networks to discriminate fungal infection levels in rice panicles using hyperspectral reflectance and principal components analysis. *Computers and Electronics in Agriculture* 72: 99–106.
68. Wefky A, Espinosa F, Prieto A, Garcia JJ, Barrios C (2011) Comparison of neural classifiers for vehicles gear estimation. *Applied Soft Computing* 11: 3580–3599.
69. Jack LB, Nandi AK (2002) Fault detection using support vector machines and artificial neural networks: augmented by genetic algorithms. *Mechanical Systems and Signal Processing* 16: 373–390.
70. LV GY, Cheng HZ, Zhai HB, Dong LX (2005) Fault diagnosis of power transformer based on multi-layer SVM classifier. *Electric Power Systems Research* 74: 1–7.
71. Kumar KS, Jayabarathi T, Naveen S (2011) Fault Identification and Location in Distribution Systems using Support Vector Machines. *European Journal of Scientific Research* 51: 53–60.