Automated Plant Identification Using Artificial Neural Networks

Jonathan Y. Clark, David P. A. Corney and H. Lilian Tang
Department of Computing, University of Surrey
Guildford, Surrey, UK
j.y.clark@surrey.ac.uk

***Abstract*— This paper describes a method of training an artificial neural network, specifically a multilayer perceptron (MLP), to act as a tool to help identify plants using morphological characters collected automatically from images of botanical herbarium specimens. A methodology is presented here to provide a practical way for taxonomists to use neural networks as automated identification tools, by collating results from a population of neural networks. A case study is provided using data extracted from specimens of the genus *Tilia* in the Herbarium of the Royal Botanic Gardens, Kew, UK. A classification accuracy of 44% was achieved on this challenging multiclass problem.**

Keywords - Herbarium specimens; Multilayer perceptrons; Neural network applications; Plant identification; *Tilia*

I. INTRODUCTION

The identification of plants is an important matter, both for those who need to be certain which organism they are dealing with, and others who are interested in establishing levels of biodiversity. Although biological identification is still often carried out using a "taxonomic key" – a traditional paper-based kind of expert system, there is an increasing trend towards using computer-aided identification systems [1]. Printed identification guides are followed manually, the user making a series of choices from groups of contrasting statements, eventually ending in a name. Although each of these statements concerns the state of at least one character or attribute, there are often additional characters, used for confirmation. The success and accuracy of identification using this methodology relies greatly on the experience of the author of the key, and how carefully it is interpreted by the user.

Artificial neural networks are computer programs that have the ability to learn from examples and can thus also perform recognition of previously unseen patterns. A multilayer perceptron (MLP) is a supervised artificial neural network (ANN) and is therefore suitable for identification, because it employs training using data for which the classes are already known. Training is carried out by presenting a succession of data records (the training set) to the network, each record containing data from a specimen or record of known identity. The resultant ability of the network to recognize previously unseen patterns is periodically tested using an independent "validation" dataset, also containing data records of known classes. The performance of the network is periodically tested against this validation set, so that training can be terminated before over-training occurs. A completely independent test dataset that contains the data records to be identified is then presented to the network. Information derived from this test set should not be used to optimize the parameters of the network, and instead it is possible to use the performance of the validation set. For further information about ANNs, see [2] and [3].

The case study presented here is of 4 species of the genus *Tilia*. This genus comprises about 23 species of woody trees, widely distributed in the north temperate regions, of which many are cultivated in gardens. They are often known as lime trees, lindens or basswoods and they are not related to the citrus tree of similar name. They are deciduous trees, often with heart-shaped, pointed leaves.

Classical printed taxonomic keys have already been used for the identification of species in the genus *Tilia*. An example of a recent key to *Tilia* species is that by Pigott [4]. To date, there are no known computer-based identification systems relating to *Tilia*, except that by Rath [5], and also earlier work by one of the authors [6][7][8][9]. Rath used a neural network to separate 13 species of woody trees (in 12 different genera) using leaf image data. Only one species of lime, *Tilia cordata*, was included. The identification task in this paper is made especially challenging as we are considering species within a single genus, which are similar by definition. Distinguishing between genera (or other higher level taxa) is typically a simpler task.

Details of other botanical species identification methods are given in [10]. The review presented there includes leaf and flower shape analysis, leaf texture analysis and vein analysis, as well as various species identification tasks. The work presented here is the first involving both artificial neural networks and the use of data automatically extracted from images of botanical herbarium specimens.

One closely related study uses a MLP to discriminate between species and varieties of the genus *Banksia* [11]. This used software to automatically extract characters from leaves, such as area and roundness, and then identify the accessions (such as providing a species name). The characters are derived from scanned images of single, undamaged leaves as opposed to herbarium images, making the character extraction more straightforward.

This work was supported by The Leverhulme Trust under Grant F/00242/H (MORPHIDAS).

TABLE 1 LIST OF CHARACTERS USED

Character	*Description*
Total number of teeth	Total count, excluding the tip
Total area of teeth	Total area in mm^2
Mean angle	Angle at tip of tooth, averaged over all teeth
Tooth frequency	number of teeth / inner length
Total outer edge length	Total length of outer edges of all teeth
Teeth edge ratio	Average ratio of lengths of two outer edges of each tooth
Total length of teeth bases	Sum of lengths where tooth meets blade
Length of blade	From insertion point to tip
Width of blade	Width at widest point, perpendicular to length-axis
Relative position of widest point	Distance along main axis: 0 = at insertion point, 1= at tip
Width-25	Blade width, 25% from insertion point to tip
Width-50	Blade width, 50% from insertion point to tip
Width-75	Blade width, 75% from insertion point to tip
Total perimeter of blade	Including teeth
Total area of blade	Including teeth
Perimeter of blade ignoring teeth	Perimeter after teeth are removed
Shape factor	$Perimeter^2$ / blade area
Compactness	4π * blade area / $perimeter^2$
Perimeter Ratio	Total perimeter / inner perimeter
Tooth Area Blade Ratio	Total tooth area / blade area excluding teeth
Tooth Number Blade Ratio	Number of teeth / total blade perimeter
Average tooth area	Total area of teeth / number of teeth

The scope of this paper is restricted to 4 species of *Tilia* commonly grown in gardens in Europe. In an earlier paper [8] 19 species were considered. However, in that study character states were extracted manually by observation and measurement in the traditional way. In the current study, information is extracted automatically from images of specimens, with only minimal manual effort, and it is therefore necessary to consider a large number of specimens. These 4 species were the only *Tilia* specimens in the Kew herbarium that were present in sufficient numbers.

A. Selection of characters

The neural network system considered here is intended for identification of mature specimens, taken from the crown of the tree, since the morphology of the leaves often varies considerably on different parts of the tree. Leaves sprouting from the base of the trunk, called 'sprout leaves', cannot usually be identified, as they are often completely different from the normal leaves. Although such sprout leaves have not been explicitly excluded from the datasets, the fact that the images were of clearly labeled specimens in the species folders means that that they are mostly mature specimens from the crown.

In previous work, we have described algorithms and software that can locate leaves within digital images of whole herbarium specimens [12] and can then automatically extract characters from these leaves [12] [13]. The work combined image processing and machine learning methods to locate and characterize the boundary of leaves in complex images, even where the leaves were overlapping or damaged – a rather difficult task.

Table 1 lists the 22 features that we have used in this current work. Many of these reproduce features described by [14] and [15]. One key difference is that their work is based on manual measurements of leaves, whereas the data we are describing here was generated automatically from images. Inevitably, the use of automated systems risks increasing the noise in the data, but it also allows far larger data sets of be created than would otherwise be possible. Using this software, we have extracted these features from over 1600 leaves found on over 1000 specimens of wild-collected plants. In this study, we limit the data to characters derived from 129 leaves of each of four species making a total of 516 records.

II. MATERIALS AND METHODS

A. Datasets

Training of the neural network was carried out using 129 examples of each species, each derived from a different wild-collected herbarium specimen. This resulted in 516 training records, each containing data from a single leaf found on a botanical herbarium specimen. This included type material where possible and practical.

The training, validation and test datasets were constructed from images of specimens held in the Herbarium of the Royal Botanic Gardens, Kew (K). Data for 22 morphological characters, conventional and novel, were extracted automatically from images of herbarium specimens of each of the 4 species considered here (see Table 2). For the purposes of this study, data derived from all specimens in folders labeled with these species names were included.

TABLE 2 ACRONYMS AND SPECIES NAMES

Acronym	Species
AME	americana
COR	cordata
PLA	platyphyllos
TOM	tomentosa

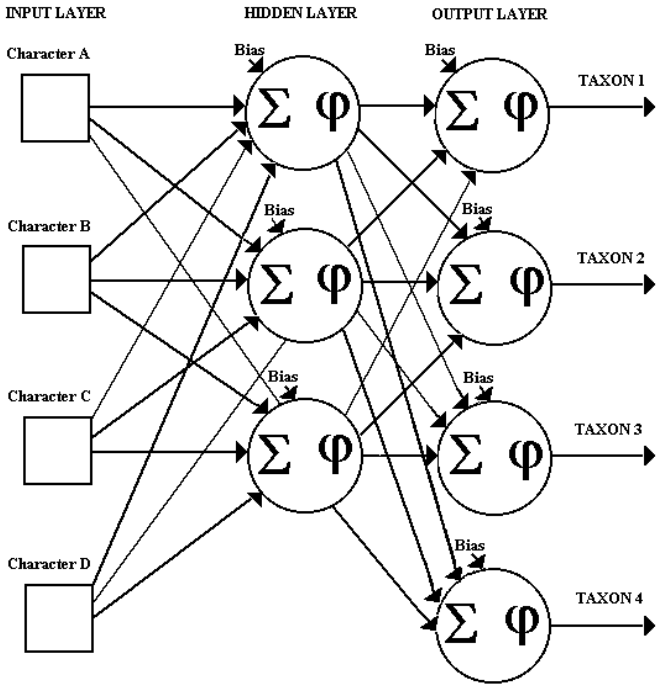

Figure 1. MLP neural network for species identification. Here each taxon is represented by a different species of *Tilia.*

Different numbers of specimens were present for each species, due to the arbitrary historical development of the herbarium. Furthermore, the software extracted different numbers of leaves from each specimen. Training ANNs with substantially unbalanced classes is notoriously difficult [16]; to simplify matters, we used random undersampling to discard random examples from the over-represented classes until all classes had the same number of examples, namely 129.

For each network, the data were divided into training, validation and test sets in a ratio of 70:20:10 respectively. Neural networks are known to be prone to overfitting, and hence not performing well on previously-unseen data. In order to reduce the risk of overfitting, a validation dataset was used to test each network periodically during training. When the validation set error began to rise, training was terminated. Only at that point the network was tested with the test data set in order to calculate the final accuracy.

Three different partition sets (A, B, and C) were created to mitigate against the chance characteristics of any particular random partition. Each partition was created for stratified cross-validation, meaning that each target class (i.e. species) was represented the same number of times in the training set, the same number of times in the validation set and the same number of times in the test set. For each partition, every record appeared in exactly one of the training, validation or test sets. For each partition, the accuracy of the neural network's predictions gives an unbiased estimate of the system's likely generalization performance in response to novel data.

These data were converted to a standard ASCII tabulated numeric format suitable for input to the neural network. In this case, each record for each leaf consisted of a single line, starting with a short acronym representing the species name followed by the character states and values, with each record terminated by the class number, corresponding to the species. Before using the training sets, the number of records was multiplied by a factor of 10, and 0.25dB of noise was added to the additional records to facilitate training, as preliminary studies showed that noise-injection would improve the final classification accuracy.

B. Neural Network

In this study, a simple feed-forward MLP with one input layer, one hidden layer, and one output layer was used. One input node corresponded to each character and one output node was assigned to represent each of the species considered (here each taxon to be identified is a species). Thus there were 22 input nodes and 4 output nodes. The number of hidden nodes was varied in order to optimize the network performance. There were no connections between nodes in the same layer, and no recurrent connections. A representation of the network architecture is shown in Fig. 1, though the number of nodes in each layer is different from that shown. The input vectors were normalized in the range ±0.9 to reduce the training time required for the inputs to the hidden nodes to reach the domain of the sigmoid activation function. Normalization was carried out independently for each character over all training records to help prevent initial weighting of characters. The minimum and maximum values for each character were kept for use during equivalent normalization of the validation and test data sets to make sure that scaling was comparable.

The weights in the network were initialized to small random values in the range ±0.5 [2]. The presentation order of input vectors was randomized between epochs. A bias input of 1.0 was used. For further details on the parameters of the network and the training algorithms used, see [7].

The error value reported was the Squared Error Percentage (E) [17], with corrections [6], given by

$$E = 100\frac{1}{NP(o_{max} - o_{min})^2}\sum_{p=1}^{P}\sum_{i=1}^{N}(o_{pi} - t_{pi})^2 \quad (1)$$

where o_{max} and o_{min} are the maximum and minimum of the output values used in the network training, here 0.9 and 0.1 respectively. N is the number of output nodes (equal to the number of species, in this case 4), and P is the number of records (patterns, or examples) in the data set under consideration. o_{pi} is the actual output at output node i when input pattern p is presented. t_{pi} is the target (desired) output at output node i when pattern p is presented.

Training was initially carried out with a constant learning rate of 0.1, and a single fixed random seed, varying only the number of nodes in the single hidden layer. Momentum was not applied. An optimized number of hidden nodes was determined by performing a number of trials using different numbers of hidden nodes. The configuration that gave the

lowest error on the validation set was considered to have an optimum number of hidden nodes. That was then fixed, and then a number of runs were carried out in which the learning rate was varied similarly to find an optimized value.

After the network parameters were set to these values, the tests were run again using the same set of 10 different random seeds on each of the three training/validation partition sets (A, B, and C). In this way a population of 30 networks and their performance results was established. The overall neural network results were collated from the results obtained from this population.

C. Assessment of performance

In the neural network trials, a misidentification matrix was produced showing species identifications. This is a confusion matrix that is similar to the misclassification matrix [18] and misidentification matrix [19] of Boddy et al. It shows the percentage of identifications referred to each species by the system. All identification attempts by the network using the test sets with the 30 different random seed trials were summed to produce the results in the table.

The matrix also shows the confidence of correct identification (%Conf). This is identical to the confidence of correct classification [20], and is a measure of the likelihood that a given species identification is correct, given that the network has identified an specimen as being that particular species. It is expressed as a percentage by calculating the proportion of correct identifications with respect to the total number of identifications.

$$\%Conf = \frac{correct}{correct + incorrect} \times 100 \quad (2)$$

III. RESULTS

The error (E_{val}) and recognition accuracy (R_{val}) produced by the network on presentation of the validation set at the point of the termination of training are shown in Table 3. The results are given for different numbers of nodes in the single hidden layer, varied between 24 and 96.

The number of hidden nodes which resulted in the lowest mean validation error (E_{val}) was found to be 56. Table 4 shows results produced using networks with 56 hidden nodes, with the learning rate varied between 0.002 and 0.050. Similarly, having fixed the number of hidden nodes to 56, the optimized learning rate was determined to be 0.005. A summary of the results from tests using the above network parameters with the 10 different random seeds is shown here in Table 5. As described by Prechelt [17], *Total Epochs* is the total number of iterations through the training set when training is actually terminated. *Relevant Epochs* is the number of training epochs at the point of minimum validation error. Also at the point of minimum validation error, E_{trn} and R_{trn} are the error and recognition accuracy respectively on the training set; E_{val} and R_{val} are the error and the recognition accuracy using the validation set. E_{test} and R_{test} are the error and recognition accuracy resulting from presentation of the test set to the trained network saved from the point of minimum validation error. The sample standard deviation (*StDev*) is also provided for all the results. Within each of the three training/validation/test partition datasets (A, B, and C), and also over the entire network population, the results for the network that produced the highest accuracy of recognition of the test set (Best R_{test}) and smallest error with the validation set (Lowest E_{val}) are also presented in Table 5. Broadly similar results were obtained using a standard linear discriminant analysis classifier (results not shown). This similarity suggests that the overall quality of the results is not specific to the MLP classifier, but reflects the challenging nature of the data set available.

The misidentification matrix is shown in Table 6. The rows refer to the 4 species to be identified in the test set. Similarly, the columns are the species to which the test species are referred by the neural network. Percentages are shown of the total samples of the row test species that are identified as belonging to the corresponding column species. Ideal (correct) identifications are shown in bold. The confidence of correct identification is given for each species.

IV. DISCUSSION AND CONCLUSIONS

In conclusion, the results presented here (see Table 5) demonstrate that the MLP neural network has a fair recognition performance when using characters obtained from herbarium specimens. Results obtained here are not as good as in earlier similar studies of the genus *Tilia* [8][9]. However, in the earlier studies, morphological data were extracted *manually* from actual specimens, and that data included details of flowers and other structures. In contrast, the work presented in this paper represents an innovative study because the information was extracted *automatically* from images of specimens, and furthermore, only characters of leaf shape and morphology have been used, and here we are performing identification from individual leaves that are still attached to the specimens.

The misidentification matrix (Table 6) shows some interesting results – the two species most easily identified by the system are *Tilia cordata* and *Tilia americana*. *Tilia platyphyllos* is often misidentified and so is *Tilia tomentosa*. The main problem here would seem to be the amount of available data in the form of specimens, and if more species are to be identified effectively, then either more data is needed, or the degree of automation is likely to need to be reduced in order to improve recognition performance. Addition of geographic information is also liable to be helpful here [9].

However, there is much potential here since neural networks train best and learn to generalize best when presented with data rich in variation. Herbarium specimens are a good source of such data, are also a traditional primary information source for botanical taxonomists. Automating such identification, or at least providing a means to speed up the identification of herbarium specimens is clearly of value. Such automation helps generate very large data sets cheaply and quickly. However, they are less accurate than manual methods, and hence the data contain more noise. The use of neural networks is therefore particularly relevant here, since

they respond well to large amounts of data, and their performance can even be enhanced by the inclusion of noise. Identification of species within a single genus is a realistic and common problem that botanists face, and much more challenging to automate than separation of widely different genera. It is clear, however, that in order for good performance to be achieved that large amounts of data are needed, in order to extract sufficient information from such noisy data. It is also likely that an increase in performance could be increased by incorporating additional methods of dimension reduction such as principal component analysis (PCA).

TABLE 3 DETERMINATION OF OPTIMIZED NUMBER OF HIDDEN NODES

H nodes	24		32		40		48		56	
DataSet	E_{val}	R_{val}	E_{val}	R_{val}	E_{val}	R_{val}	E_{val}	R_{val}	E_{val}	R_{val}
A	15.700	53.77	15.984	51.89	16.320	50.94	16.061	48.11	15.698	47.17
B	17.587	40.57	17.786	37.74	17.811	36.79	17.568	39.62	17.442	40.57
C	17.373	43.40	17.258	44.34	17.274	46.23	17.501	44.34	17.374	49.06
Mean	**16.887**	**45.913**	**17.009**	**44.657**	**17.135**	**44.653**	**17.043**	**44.023**	**16.838**	**45.600**

H nodes	64		72		80		88		96	
DataSet	E_{val}	R_{val}	E_{val}	R_{val}	E_{val}	R_{val}	E_{val}	R_{val}	E_{val}	R_{val}
A	16.246	46.23	16.063	49.06	15.975	48.11	15.985	50.94	16.093	47.17
B	17.695	36.79	17.817	35.85	17.736	36.79	17.689	38.68	17.637	41.51
C	17.426	46.23	17.415	41.51	17.525	41.51	17.324	49.06	17.288	48.11
Mean	**17.122**	**43.08**	**17.098**	**42.14**	**17.079**	**42.14**	**17.999**	**46.23**	**17.006**	**45.60**

TABLE 4 DETERMINATION OF OPTIMIZED LEARNING RATE

Learning rate	0.002		0.005		0.010		0.015		0.020	
DataSet	E_{val}	R_{val}	E_{val}	R_{val}	E_{val}	R_{val}	E_{val}	R_{val}	E_{val}	R_{val}
A	15.689	50.94	15.681	49.06	15.698	47.17	15.633	52.83	15.673	50.94
B	17.579	46.23	17.338	42.45	17.442	40.57	17.511	41.51	17.535	38.68
C	17.664	43.40	17.371	47.17	17.374	49.06	17.430	43.40	17.455	43.40
Mean	**16.977**	**46.857**	**16.797**	**46.227**	**16.838**	**45.600**	**16.858**	**45.913**	**16.888**	**44.340**

Learning rate	0.030		0.035		0.040		0.050	
DataSet	E_{val}	R_{val}	E_{val}	R_{val}	E_{val}	R_{val}	E_{val}	R_{val}
A	15.672	52.83	15.639	50.94	15.629	49.06	15.698	50.94
B	17.710	39.62	17.734	39.62	17.773	40.57	17.831	40.57
C	17.581	42.45	17.581	47.17	17.600	46.23	17.706	44.34
Mean	**16.988**	**44.97**	**16.985**	**45.91**	**17.001**	**45.29**	**17.078**	**45.28**

TABLE 5 TRAINING, VALIDATION AND TEST RESULTS

Tilia DataSet		Total Epochs	Relevant Epochs	E_{trn}	R_{trn}	E_{val}	R_{val}	E_{test}	R_{test}
A	Mean	951.00	535.00	15.54	50.23	15.97	49.99	16.91	44.71
	StDev	65.23	104.16	0.31	1.78	0.20	2.31	0.30	3.90
	Best R_{test}	1000	510	15.315	51.96	15.887	50.84	17.127	50.98
	Lowest E_{val}	890	440	15.672	49.24	15.681	49.06	16.803	45.1
B	Mean	966.00	413.00	15.81	48.88	17.64	38.21	16.59	40.59
	StDev	94.30	159.86	0.47	2.02	0.15	2.72	0.17	1.86
	Best R_{test}	960	470	15.828	48.42	17.751	37.74	16.471	43.14
	Lowest E_{val}	1000	750	14.636	54.31	17.301	43.4	16.672	43.14
C	Mean	844.00	117.00	16.86	43.42	17.33	47.36	16.44	46.47
	StDev	193.52	11.60	0.05	0.32	0.06	1.88	0.26	3.21
	Best R_{test}	1000	120	16.86	43.73	17.393	49.06	16.645	52.94
	Lowest E_{val}	450	110	16.864	43.46	17.252	48.11	16.535	47.06
Overall	Mean	**920.33**	**355.00**	**16.07**	**47.51**	**16.98**	**45.19**	**16.65**	**43.92**
Overall	StDev	**136.95**	**207.86**	**0.66**	**3.35**	**0.75**	**5.60**	**0.31**	**3.91**
Best R_{test}	Mean	**986.67**	**366.67**	**16.00**	**48.04**	**17.01**	**45.88**	**16.75**	**49.02**
Best R_{test}	StDev	**23.09**	**214.55**	**0.79**	**4.13**	**0.99**	**7.11**	**0.34**	**5.19**
Lowest E_{val}	Mean	**780.00**	**433.33**	**15.72**	**49.00**	**16.74**	**46.86**	**16.67**	**45.10**
Lowest E_{val}	StDev	**291.03**	**320.05**	**1.11**	**5.43**	**0.92**	**3.03**	**0.13**	**1.96**

TABLE 6 MISIDENTIFICATION MATRIX

	COR	PLA	AME	TOM
COR	**292**	13	35	49
PLA	162	**67**	73	87
AME	44	49	**223**	54
TOM	95	30	165	**90**
%Conf	49.24	42.14	44.96	32.14

ACKNOWLEDGMENT

Gratitude is due to the Royal Botanic Gardens, Kew for permission to study specimens in the Herbarium, and to Donald Pigott for helpful comments regarding *Tilia*. Thanks also go to Y. Hu and J. Jin for help with extraction algorithms.

REFERENCES

[1] R.J. Pankhurst, *Practical Taxonomic Computing*. UK, University of Cambridge Press, 1991.

[2] J.A. Freeman and D.M. Skapura, *Neural networks: algorithms, applications, and programming techniques*, Reading, Massachusetts, USA: Addison-Wesley, 1992.

[3] S. Haykin, *Neural networks - a comprehensive foundation*. New York, USA: Macmillan College Publishing Company, Inc.,1994.

[4] C.D. Pigott, "*Tilia*", in *European Garden Flora Volume 5*, J. Cullen *et al*. (Eds.), pp.205-212, Cambridge University Press, UK, 1997.

[5] T. Rath, "Klassifikation und identifikation gartenbaulicher objekte mit könstlichen neuronalen netzwerken", *Gartenbauwissenschaft*, vol. 61 (4), 1996, pp. 153-159.

[6] J.Y. Clark, *Botanical identification and classification using artificial neural networks*, PhD Thesis, Dept. of Cybernetics, University of Reading, UK, 2000.

[7] J.Y. Clark, "Artificial neural networks for species identification by taxonomists," *BioSystems*, vol. 72, 2003, pp. 131-147.

[8] J.Y. Clark, "Identification of botanical specimens using artificial neural networks," presented at the *IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, San Diego, USA - October 2004, pp. 87-94.

[9] J.Y. Clark, "Plant identification from characters and measurements using artificial neural networks" In MacLeod, N (Ed.) *Automated Taxon Identification in Systematics* (Systematics Association Special Volume), CRC Press, 2007, pp. 207-224.[Presented at a Symposium on Algorithmic Approaches to the Identification Problem in Systematics, Natural History Museum, South Kensington, 2005.]

[10] J. S. Cope, D. P. A. Corney, J. Y. Clark, P. Remagnino and P. Wilkin, "Plant species identification using digital morphometrics: A review", *Expert Systems with Applications*, 39(8), 2012, pp. 7562–7573

[11] G. Messina, C. Pandolfi, S. Mugnai, E. Azzarello, K. Dixon and S. Mancuso (2009) "Phyllometric parameters and artificial neural networks for the identification of *Banksia* accessions", *Australian Systematic Botany*, 22(1), 2009, pp. 31–38

[12] D.P.A. Corney, J.Y. Clark, H.L. Tang and P. Wilkin, "Automatic extraction of leaf characters from herbarium specimens" *Taxon,* 61(1), 2012, pp. 231-244.

[13] D.P.A. Corney, H.L. Tang, J.Y. Clark, Y. Hu, J. Jin, "Automating digital leaf measurement: the tooth, the whole tooth, and nothing but the tooth," *PLoS One*, submitted for publication.

[14] P.M. Huff, P. Wilf and E.J. Azumah, "Digital future for paleoclimate estimation from fossil leaves? Preliminary results." *Palaios* 18, 2003, pp. 266-274.

[15] D.J. Peppe, D.L. Royer, B. Cariglino, S.Y. Oliver, S. Newman *et al*., "Sensitivity of leaf size and shape to climate: global patterns and paleoclimatic applications." *New Phytologist* 190, 2011, pp. 724-739.

[16] H. He and E.A Garcia, "Learning from Imbalanced Data", *IEEE Transactions on Knowledge and Data Engineering*, 21 (9), 2009, pp. 1263-1284

[17] L. Prechelt, "Proben1 – A set of neural network benchmark problems and benchmarking rules" Technical Report 21/94, Universität Karlrühe, Germany, 1994.

[18] L. Boddy, C.W. Morris, and A. Morgan, "Development of artificial neural networks for identification", in *Information technology, plant pathology & biodiversity*, Bridge, P., Jeffries, P., Morse, D.R., Scott, P.R., Eds., Wallingford, UK: CAB International, 1998, pp. 221-231.

[19] L. Boddy, C.W. Morris, M.F. Wilkins, L. Al-Haddad, G.A. Tarran, R.R. Jonker and P.H. Burkill, "Identification of 72 phytoplankton species by radial basis function neural network analysis of flow cytometric data", *Marine Ecology Progress Series*, vol. 195, 2000, pp. 47-59.

[20] A. Morgan, L. Boddy, J.E.M. Mordue and C.W. Morris, "Evaluation of artificial neural networks for fungal identification, employing morphometric data from spores of *Pestalotiopsis* species" *Mycological Research*, vol. 102 (8), 1998, pp. 975-984.