

Identification of Botanical Specimens using Artificial Neural Networks

Jonathan Y. Clark

Abstract—This paper describes a method of training an artificial neural network, specifically a multilayer perceptron (MLP), to identify plants using morphological characters collected from herbarium specimens. A practical methodology is presented to enable taxonomists to use neural networks as advisory tools for identification purposes, by collating results from a population of neural networks. A comparison is made between the ability of the neural network and that of other methods for identification by means of a case study in the ornamental tree genus *Tilia* L. (Tiliaceae). In particular, a comparison is made with taxonomic keys generated by means of the DELTA system, a suite of programs commonly used by botanists for that purpose. In this study, the MLP was found to perform better than the DELTA key generator.

Index Terms—Herbarium specimens, Multilayer perceptrons, Neural network applications, Taxonomic keys, *Tilia*

I. INTRODUCTION

TAXONOMIC identification is an important issue, both for those interested in biodiversity, and for others who need to be certain which organism they are dealing with. Much of this biological identification is still carried out using a "taxonomic key", which is a classical paper-based kind of expert system. An example of this is shown in the DELTA-generated key in the Appendix. This kind of printed identification guide must usually be followed manually, although on-line computer-based methods are becoming more widespread. When using such a key, the user makes a series of choices from successive groups of contrasting statements, culminating in a name. Each statement concerns the state of at least one character or attribute. Often there are additional confirmatory characters, although the first character state mentioned is usually the most important. The success and accuracy of this identification relies heavily on the experience of the expert who compiled it, and the care of its interpretation by the user. Although interactive computer-based systems are useful, and becoming more popular, the performance of such systems is difficult to compare with that of other methods, because characters can be chosen in any order. The DELTA KEY program [1], however,

Manuscript received April 12, 2004. This work was carried out in part towards a PhD in Cybernetics, University of Reading, United Kingdom.

J. Y. Clark is now with the Department of Computing (H3), University of Surrey, Guildford, Surrey, GU2 7XH, United Kingdom (phone: +44-(0)1483-683425; fax: +44-(0)1483-686051; e-mail: j.y.clark@surrey.ac.uk).

can be used to generate a conventional printable taxonomic key, whose performance can then be directly compared with other methods. The interested reader is referred to [2] for a good account of identification methods used by biologists.

A neural network based system has the ability to learn from examples and can perform generalized recognition of previously unseen patterns. A multilayer perceptron (MLP) is the kind of artificial neural network (ANN) most suitable for identification, because it employs supervised training. Training is carried out by presenting the network with a succession of data records, constituting the training set, each one containing data from a specimen or record of known identity. The generalization ability of the network to recognize previously unseen patterns is periodically tested using an independent "validation" dataset, also containing known classes. By means of testing the network's performance against this validation set, training can be terminated before over-training occurs. A completely independent test dataset, containing the data records to be identified is then presented to the network. Information derived from this test set must not be used to optimize network parameters. For further information about ANNs, see [3] and [4].

The case study presented here is of cultivated species of the genus *Tilia*. This genus comprises about 30 species of woody trees, widely distributed in the north temperate regions, of which around 19 are cultivated in European gardens. Commonly known in England as limes, they are unrelated to the citrus tree of that name, and are otherwise referred to as lindens or basswoods. Limes are deciduous trees, usually 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 [5]. To date, there are no known computer-based identification systems relating to *Tilia*, except that of Rath [6], and earlier work by the author [7]. In Rath's work, in which 13 species of woody trees (in 12 genera) were separated by means of a neural network using leaf image data, *Tilia cordata* was the only lime.

The work presented here is the first involving both artificial neural networks and the use of the DELTA system and key generator in studies of a large number of *Tilia* species, and is derived from earlier work by the author [7]. A similar study has already been performed with respect to identification of 35 species of the genus *Lithops* (Aizoaceae) [8].

The scope of the project was restricted to species grown in

gardens in Europe, and included in the account of the genus in the European Garden Flora [5]. Of these, *Tilia neglecta* is omitted, because it is now generally included in *T. americana*. Cultivated species were chosen for this study because trees and herbarium specimens were readily available. Furthermore, cultivated *Tilia* trees present a particular difficulty with respect to identification because many species readily hybridize. A total of 19 species are thus considered here.

It is valuable at this point to discuss the selection of characters. The neural network key generated here is intended for identification of mature flowering 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 fruit characters are often of great diagnostic importance [9], it was decided to omit them from this study in order to avoid destruction of fruits on important specimens (e.g. nomenclatural types). In fact, some authors have produced separate keys to flowering and fruiting specimens [10].

A decision was made to concentrate on measurements rather than subjective descriptions of characters, since they can be more objectively evaluated by unskilled workers. Therefore, the descriptive character of leaf shape was rejected in favor of the more objective measurements of leaf length and width. (In this paper, the term leaf is used to refer to the flat part of the leaf, or blade, and excluding the leaf stalk, or petiole.) Some characters, such as presence or absence of staminodes, are usually invariable within a species, whereas many others, such as leaf length are extremely variable. Here, 3-4 separate measurements were recorded and the mean calculated. The variability of many character states within taxa, between seasons and even between branches on the same tree is extreme. Much of this variation results from different ages of leaves; therefore atypically small and immature leaves were not included here.

II. MATERIALS AND METHODS

A. Datasets

Training was carried out using three examples of each species, that is, using three different data records for each of 19 cultivated species, and each derived from a different (mostly wild-collected) herbarium specimen. This resulted in 57 training records, each containing data from a single herbarium specimen, including type material where possible and practical.

The training and validation datasets were constructed from herbarium specimens held in the herbaria of the Royal Botanic Gardens, Kew (K) and the Natural History Museum in South Kensington, London (BM). Data for 22 morphological characters were extracted from three field-collected specimens of each of the 19 species considered here. Flowering specimens were chosen because extra characters such as

'number of flowers in inflorescence' and 'presence or absence of staminodes' would also be available. Furthermore, the character of the number of flowers in the inflorescence is more reliable in flowering specimens – fruits readily drop off and are lost. It is rare for a single specimen to have both flowers and fruits, so only flowering specimens were used.

In the neural network study, the validation dataset, against which the network was tested periodically to determine when to stop training, consisted of data from one specimen of each species, the remaining two being left in the training set. Three different pairs (A, B and C) of training and validation sets were produced in which the one record to be transferred to the validation set was chosen randomly, the remaining records then being placed in the respective training set. The ANN tests outlined below were carried out using each partition pair.

The data for the test set were collected from cultivated herbarium specimens of known identity, but ones whose identity was not provided to the network. Testing was thus carried out using an independent data set derived from 30 herbarium specimens of lime trees cultivated in the Arboretum at the Royal Botanic Gardens, Kew, the specimens themselves being held in the cultivated folders of the Kew Herbarium. In a few cases some characters were not visible on the specimen and it was necessary to collect further material (from the original tree), which was pressed, dried, and mounted before examination.

The primary data were held in the DELTA format [1], [11], [12] and [13] for consistency and standardization of method with other computer-based botanical data analyses. This was to facilitate direct comparison of the neural network results with those obtained using the standard methods. The variant items methodology [13] was used. That is, the extra items were denoted by the appropriate 3-letter species acronym in the name field. Thus, data for each extra specimen or item was logically included in the concept of the species when processed by the KEY generator. A list of sources of material is available in [7]; the species acronyms are given in Table 8.

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 taxon consisted of a single line, starting with a short acronym representing the taxon name followed by the character states, with each record terminated by the class number, corresponding to the species. During conversion from DELTA format, extreme values were removed, and ranges replaced by their mean value, as when a key is generated using DELTA.

B. Neural Network

A simple feed-forward MLP with one input layer, one hidden layer, and one output layer was used for this study. One input node was designated for each character, the number of hidden nodes was variable, and one output node was assigned to represent each species to be identified. There were no connections between nodes in the same layer, and no recurrent connections. A representation of the architecture is presented in Fig. 1, although the actual number of nodes in each layer is

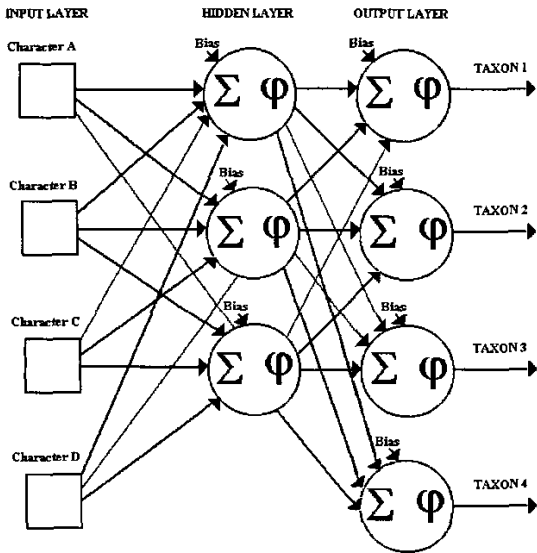


Fig. 1. MLP neural network for identification of taxa.

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. This normalization was carried out for each character independently over all training records to preclude initial character weighting. The maximum and minimum values for each character were retained for use during normalization of the validation and test data to ensure comparable scaling.

The network weights were initialized to small random values in the range ± 0.5 [3]. The presentation order of input vectors was randomized between epochs and a bias input of 1.0 was used. For further details on the parameters of the network and the training algorithms used, see [8]. The error value reported was the Squared Error Percentage (E) [14], with corrections [7], 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 that could be used in training, in this case 0.9 and 0.1 respectively. N is the number of output nodes (here equal to the number of species), 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. After determining the optimum number of hidden nodes, that was then fixed, and the learning rate varied to find

the best 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 training/validation partition sets. Thus a population of networks and their results was established. The overall neural network results were collated from the results obtained by this population.

C. DELTA Key generation

Although it would have been possible to generate an interactive computerized key using the INTKEY program [14] included with the DELTA system, it was decided to focus on the traditional key generator to enable a better comparison with the MLP technique. The interactive key would have introduced the problem of choosing the order of character selection.

The DELTA keys were all produced from the MLP training and validation data. Each species was given its own record in the DELTA ITEMS file, with each specimen of the species being treated as a variant item. In effect, this was to cause the key generator to regard all three specimens of each species to be the same species.

A completely unweighted key, shown in full in the Appendix, was generated using the DELTA KEY program, in order to force the KEY generator to choose the most useful characters itself, together with the order of their presentation. Continuous characters were divided into ranges specified in the TOKEY file; the same ranges used for character analysis tests performed earlier [7]. All other DELTA parameters (directives) were allowed to remain set to the default values and settings. This meant that no additional CONFIRMATORY CHARACTERS were sought - each statement in the key would be concerned with contrasting states of only one character.

A second DELTA key was then generated, with CHARACTER RELIABILITIES set to reflect the importance of the chosen characters, as far as is possible to determine from the traditional published key [5]. Here, the number of instances each character was used at a decision point in the key was counted. The CHARACTER RELIABILITIES were then set in the KEY file to reflect the relative number of instances. If the human-developed key did not use a character used in this study, then a value of zero was used. The character with the highest number of usage instances was given a value of ten. The other characters were given values normalized within this range.

TABLE 1 DETERMINATION OF OPTIMIZED NUMBER OF HIDDEN NODES

H nodes	64		72		80		88		96		104	
DataSet	<i>E_{val}</i>	<i>R_{val}</i>	<i>E_{val}</i>	<i>R_{val}</i>	<i>E_{val}</i>	<i>R_{val}</i>	<i>E_{val}</i>	<i>R_{val}</i>	<i>E_{val}</i>	<i>R_{val}</i>	<i>E_{val}</i>	<i>R_{val}</i>
A	3.00	73.68	3.06	68.42	3.03	68.42	2.84	68.42	2.82	73.68	2.77	73.68
B	3.04	73.68	3.03	78.95	2.75	78.95	2.62	78.95	2.63	73.68	2.95	68.42
C	2.89	78.95	3.25	78.95	3.30	57.89	2.73	73.68	2.92	78.95	2.75	73.68
Mean	2.98	75.44	3.11	75.44	3.03	68.42	2.73	73.68	2.79	75.44	2.82	71.93

TABLE 2 DETERMINATION OF OPTIMIZED LEARNING RATE

Learning rate	0.05		0.10		0.15		0.20		0.25		0.30	
DataSet	<i>E_{val}</i>	<i>R_{val}</i>	<i>E_{val}</i>	<i>R_{val}</i>	<i>E_{val}</i>	<i>R_{val}</i>	<i>E_{val}</i>	<i>R_{val}</i>	<i>E_{val}</i>	<i>R_{val}</i>	<i>E_{val}</i>	<i>R_{val}</i>
A	2.85	68.42	2.84	68.42	2.85	73.68	2.82	68.42	2.81	68.42	2.82	68.42
B	2.62	78.95	2.62	78.95	2.61	78.95	2.63	84.21	2.76	78.95	2.72	84.21
C	2.74	68.42	2.73	73.68	2.76	73.68	2.72	73.68	2.89	68.42	2.89	68.42
Mean	2.74	71.93	2.73	73.68	2.74	75.44	2.72	75.44	2.82	71.93	2.81	73.68

TABLE 3 SPECIES-BASED MISIDENTIFICATION MATRIX

Species	AME	AMU	CAR	CHI	COR	DAS	HEN	HET	INS	JAP	KIU	MAN	MAX	MIQ	MON	OLI	PLA	TOM	TUA	<i>R_{win}</i>	
TAME	100																				100
TAMU		0.0		6.7		3.3			80.0	10.0											0
TCOR		15.0			47.5	0.8			15.8								20.8				100
TDAS	3.3	3.3				6.7			33.3				23.3				30.0				0
THEN			6.7				13.3								80.0						0
THET								0.0			3.3					3.3		93.3			0
TINS	11.1			1.1	2.2	32.2			33.3	14.4							5.6				100
TKIU											100										100
TMAX				6.7									93.3								100
TMIQ														98.3		1.7					100
TMON	3.3									3.3					93.3						100
TOLI									10.0				2.2	44.4		37.8		5.6			0
TPLA						0.7											99.3				100
TTOM				1.7							22.5	7.5	1.7		35.8			30.8			0
%Conf	84.9	0.0	-	-	95.5	15.2	100	0.0	19.3	-	100	-	73.8	68.1	53.8	48.1	63.8	23.8	-		57.1

The generated keys were then parsed manually using the test data, that is, the key was used for identification in the traditional way. If any given character state was valid for more than one key statement in a couplet, then the closest match was chosen. For instance, if the choice was between a character state of 3-7 cm or 7-12 cm and the character on the specimen had a value of 6-11 cm, then the second path would be followed. If the matches were equally favorable, then the subsequent paths from both key statements were followed. If a

species name was reached without any branching of the decision path, then the identification reached was given a level of 100%. If a name was reached after only one branching of the decision path, then that species was considered to be identified to a 50% level (unless the other branch resulted in the same diagnosis). If after a 50% branch, there was a further branch, the identification reached was said to be at a 25% level. This process was continued until all branches had reached a name. Then percentage totals for all the species

reached were collated to produce percentages for each species to which the specimen was referred.

D. Assessment of performance

The R_{test} results could not be used for comparison because they were ANN-specific. Instead, in both the ANN and DELTA key trials, a misidentification matrix was produced showing the species identifications. This is a confusion matrix similar in concept to the misclassification matrix [15] and misidentification matrix [16] of Boddy *et al.*, in that it shows the percentage of identifications referred to each species by the system. All identification attempts by the network were summed to produce the results in the table.

On the bottom row, the matrix also shows the confidence of correct identification ($\%Conf$). This is identical to the confidence of correct classification used by Morgan *et al.* (1998), and is a measure of the likelihood that a given species identification is correct, given that the network has identified an unknown specimen as that taxon. It is calculated by expressing as a percentage the proportion of correct identifications with respect to the total number of identifications (including wrong identifications).

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

For each species, a winner-takes-all percentage value (R_{win}) was evaluated. This was given a value of 100% if a correct identification had the highest percentage identification, or shared between joint winners e.g. if the winning identification was equal for two species, then R_{win} was 50%. The mean R_{win} over all species is presented at the bottom of the R_{win} column in the misidentification matrix.

Statistical significance tests were then carried out to evaluate the differences between these results. The R_{win} result for each test specimen in turn was extracted from raw specimen-based misidentification matrices, resulting in an ordered set for each kind of test. This was also done in the case of the comparable misidentification matrices from the MLP studies. Although both specimen-based and species-based matrices were produced, the statistical calculations were made using the specimen-based matrices. For clarity and simplicity, however, only the species-based matrices are shown here. Thus three sets of these matrices were produced: one from the MLP results, one from the DELTA key with default parameters, and one from the DELTA key with character weightings. The significance of differences between each pair of sets was evaluated using a standard paired t-test [18].

III. RESULTS

In Table 1, the error (E_{val}) and recognition accuracy (R_{val}) produced by the network on presentation of the validation set at the point of training termination are shown. The results are given for different numbers of nodes in the single hidden layer,

varied between 64 to 104.

The number of hidden nodes which resulted in the lowest mean validation error (E_{val}) was 88; this also gave the highest mean R_{val} . Table 2 shows results produced using networks with 88 hidden nodes, with the learning rate varied between 0.05 and 0.30. Having fixed the number of hidden nodes to 88, the optimized learning rate was found to be 0.2.

A summary of the results from tests using the above network parameters with the 10 different random seeds is shown in Table 5. As described by Prechelt [14], *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.

The species-based misidentification matrix is shown in Table 3. The rows refer to the species in the test set (T...). Similarly, the columns are the species to which the test plants 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 for which there was a test specimen.

Table 6 shows the misidentification table results from the MLP neural network tests and the DELTA tests. This table compares results for the MLP neural network tests with those from the DELTA tests using no *a priori* character weighting (*TD22*), and those using character weightings (*TDC22*) suggested by the existing traditional botanical key [5].

The results compared in the statistical tests are the mean R_{win} values obtained using the winner-takes-all approach described earlier, performed on the appropriate specimen-based misidentification matrices, that provided the raw data for the species-based matrices. Since the recognition of each test specimen was tested in turn, and in the same order for each test, a paired t-test was sufficient, and it was not necessary to carry out a prior F-test. The probabilities returned by the t-test for each comparison are shown in Table 7, explained as follows:

Comparison between the MLP (*Tilia22*) performance and DELTA key with default parameters: the null hypothesis is that there is not a significant difference between the results of the two tests. A paired 2-tailed t-test, not assuming equal variance, shows that this hypothesis is refuted, there being a significant difference at the 5% level.

TABLE 4 DELTA KEY SPECIES-BASED MISIDENTIFICATION MATRIX

Species	AME	AMU	CAR	CHI	COR	DAS	HEN	HET	INS	JAP	KIU	MAN	MAX	MIQ	MON	OLI	PLA	TOM	TUA	R _{win}
TAME	100																			100
TAMU		0.0			100															0
TCOR		25.0			50.0				12.5								12.5			100
TDAS					50.0	50.0														50
THEN	100						0.0													0
THET								0.0						100						0
TINS	33.3	8.3			25.0	33.3			0.0											0
TKIU											100									100
TMAX								100					0.0							0
TMIQ														100						100
TMON		50.0			50.0										0.0					0
TOLI												66.7		33.3		0.0				0
TPLA																	100			100
TTOM			25.0									6.3		6.3		31.3		31.3		50
%Conf	42.9	0.0	-	-	18.2	60.0	0.0	0.0	0.0	-	100	-	0.0	41.7	0.0	0.0	88.9	100	-	42.9

TABLE 5 TRAINING, VALIDATION AND TEST RESULTS

Tilia22 DataSet		Total Epochs	Relevant Epochs	E _{tr}	R _{tr}	E _{val}	R _{val}	E _{test}	R _{test}
A	Mean	135.10	60.90	1.28	96.58	2.86	70.00	3.75	55.34
	StDev	85.74	30.07	0.37	3.52	0.15	4.99	0.19	5.02
	Best R _{test}	77.00	63.00	1.17	97.37	2.92	73.68	3.68	63.33
	Lowest E _{val}	182.00	63.00	1.01	97.37	2.65	78.95	3.71	56.67
B	Mean	133.70	70.70	1.09	98.42	2.73	78.42	3.63	60.33
	StDev	59.25	15.28	0.30	1.84	0.17	6.30	0.16	4.57
	Best R _{test}	133.00	70.00	1.11	100.00	2.79	78.95	3.58	66.67
	Lowest E _{val}	154.00	84.00	0.90	100.00	2.49	84.21	3.67	63.33
C	Mean	93.80	62.30	1.27	96.05	2.81	74.21	3.78	56.33
	StDev	45.63	22.00	0.50	5.58	0.19	4.61	0.17	6.37
	Best R _{test}	119.00	77.00	1.02	97.37	2.79	73.68	3.75	63.33
	Lowest E _{val}	77.00	56.00	1.13	100.00	2.56	78.95	3.76	60.00
Overall	Mean	120.87	64.63	1.21	97.02	2.80	74.21	3.72	57.33
Overall	StDev	66.30	22.86	0.39	3.95	0.17	6.24	0.18	5.63
Best R _{test}	Mean	109.67	70.00	1.10	98.25	2.83	75.44	3.67	64.44
Best R _{test}	StDev	29.14	7.00	0.08	1.52	0.07	3.04	0.09	1.93
Lowest E _{val}	Mean	137.67	67.67	1.01	99.12	2.56	80.70	3.71	60.00
Lowest E _{val}	StDev	54.37	14.57	0.12	1.52	0.08	3.04	0.04	3.33

b). Comparison between the MLP (*Tilia22*) performance and DELTA key with character weighting (*TDC22*): the null hypothesis is that there is not a significant difference between the results of the two tests. A paired 2-tailed t-test, not assuming equal variance, shows that this hypothesis is refuted, there being a significant difference at the 5% level.

c). Comparison between the DELTA key with default

parameters (*TD22*) and DELTA key with character weighting (*TDC22*): the null hypothesis is that there is not a significant difference between the results of the two tests. A paired 2-tailed t-test, not assuming equal variance, shows that this hypothesis is not refuted, there being no significant difference found at the 5% level.

IV. DISCUSSION AND CONCLUSIONS

In conclusion, the results presented here (see Table 6) demonstrate that the MLP neural network has a recognition performance consistently better than that of a key generated

TABLE 6
IDENTIFICATION PERFORMANCE

Identification Method	Mean Rwin		Confidence	
	Specimen	Species	>50%	100%
ANN MLP	66.7	57.1	8	2
DELTA TD22: unweighted	41.7	42.9	4	2
DELTA TDC22: weighted	43.9	39.3	6	4

TABLE 7
STATISTICAL TESTS

Paired T-test	TD22: unweighted	TDC22: weighted
ANN MLP	0.02	0.03
DELTA TD22 (unweighted)	N/A	0.81

TABLE 8
ACRONYMS AND SPECIES NAMES

Acronym	Species	Acronym	Species
AME	<i>americana</i>	KIU	<i>kiusiana</i>
AMU	<i>amurensis</i>	MAN	<i>mandshurica</i>
CAR	<i>caroliniana</i>	MAX	<i>maximowicziana</i>
CHI	<i>chinensis</i>	MIQ	<i>miqueliana</i>
COR	<i>cordata</i>	MON	<i>mongolica</i>
DAS	<i>dasystyla</i>	OLI	<i>oliveri</i>
HEN	<i>henryana</i>	PLA	<i>platyphyllos</i>
HET	<i>heterophylla</i>	TOM	<i>tomentosa</i>
INS	<i>insularis</i>	TUA	<i>tuan</i>
JAP	<i>japonica</i>		

using the DELTA system, when using characters obtained from herbarium specimens. Indeed, the results obtained here are better than those obtained in an earlier similar study of the genus *Lithops* [8]. This may be because the training data for the *Lithops* study were obtained from published species descriptions, whereas the *Tilia* study was based on real data obtained from three actual specimens of each species.

The neural network methodology, like any other identification system, is clearly of use in suggesting inadequacies with existing classifications. The *T. amurensis* test specimen (TAMU) was consistently identified as its close ally, *T. insularis* (INS). This is coincident with a human expert on the genus also concluding that they should both be considered *T. amurensis* [19]. In theory, after such reclassification, the system would perform even better.

Although further research is needed, it seems likely that the tree from which the *T. heterophylla* test specimen (THET) was

taken was incorrectly labeled, and is actually *T. X moltkei*, a putative hybrid between *T. americana* and *T. tomentosa* (or its 'Petiolaris' cultivar). This test specimen was identified by the neural network to be *T. tomentosa* (TOM), which it clearly was not, since the fruits differ significantly from those of that species. However, this implied that the tree might have contained some characters from *T. tomentosa*. On further investigation, that test specimen seems to be a close match to specimens labeled as being the hybrid in the Kew Herbarium, and it is also identified as the hybrid when Pigott's key [5] is used. Although the neural network could not identify that specimen as the hybrid, because a class for the hybrid was not included in training, the neural network's confusion is now understandable. The neural network tool was therefore not completely wrong, showing that such incorrect identifications should be investigated further, and can highlight problems with the specimen data. Inclusion of training data from such hybrids might be useful for neural network-based identification, where the specimens to be identified are suspected of being of hybrid origin.

It is interesting that the statistical tests reveal no statistically significant difference between the performance of the two DELTA key tests. This suggests that no significant value was added by the inclusion of expert-determined character weights. However, that conclusion might be premature, since characters in the human-derived key were ignored if they had not already been chosen for the neural network tests.

The limitations of the neural network method are largely the same as those of a human expert, namely that success depends on the quantity, validity, and accuracy of training data. It is well known that 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, and furthermore are the traditional primary source of information for the botanical taxonomist. The use of neural networks as tools for herbarium systematics therefore shows much promise in this field. In this study, the neural network results were collated from a population of trained networks. The variation in performance suggests that evolutionary computation would be useful for future further refinement of network parameters to enhance identification performance. Existing work in this field [20] and [21] shows that such evolutionary artificial neural networks (EANNs) can have an advantage over neural networks or evolutionary algorithms alone.

APPENDIX

DELTA Key to Flowering Specimens of *Tilia* species cultivated in European Gardens

1(0).	Leaf underside: stellate hairs absent	2
	Leaf underside: stellate hairs few.....	15
	Leaf underside: stellate hairs clearly present.....	16
2(1).	Leaf width 2.1 to 2.7 cm.....	KIU
	Leaf width 4.1 to 10.8 cm.....	3

3(2).	Leaf top: simple hairs absent.....	4
	Leaf top: simple hairs few.....	12
	Leaf top: simple hairs clearly present.....	PLA
4(3).	Leaf length 3.9 to 6.9 cm.....	5
	Leaf length 7.5 to 11.3 cm.....	11
5(4).	Bract petiole length 2.5 to 9.3 mm.....	AMU or COR
	Bract petiole length 10.8 to 11.5 mm.....	6
	Bract petiole length 13 to 13.8 mm.....	8
	Bract petiole length 16 to 16.5 mm.....	AMU
	Bract petiole length 19.8 to 21.8 mm.....	9
6(5).	Leaf top: small brown hairs absent.....	7
	Leaf top: small brown hairs few.....	COR
7(6).	Leaf base cordate.....	JAP
	Leaf base truncate.....	MON
	Leaf base cuneate.....	MON
8(5).	Staminodes absent.....	COR
	Staminodes clearly present.....	MON
9(5).	Style hairless.....	INS
	Style partly hairy (i.e. for at least 1/4 length, but not full length).....	10
10(9).	Bract length 4.3 to 5.9 cm.....	JAP
	Bract length 6.4 to 7.4 cm.....	DAS
11(4).	Staminodes absent.....	DAS
	Staminodes clearly present.....	AME
12(3).	Leaf underside: simple hairs absent.....	13
	Leaf underside: simple hairs few.....	14
13(12).	Bract length 4.3 to 5.9 cm.....	JAP
	Bract length 6.4 to 7.4 cm.....	COR
	Bract length 8 to 12.1 cm.....	JAP
14(12).	Bract petiole length 2.5 to 9.3 mm.....	COR
	Bract petiole length 10.8 to 11.5 mm.....	COR
	Bract petiole length 19.8 to 21.8 mm.....	INS
15(1).	Leaf top: simple hairs absent.....	AME
	Leaf top: simple hairs few.....	DAS
	Leaf top: simple hairs clearly present.....	KIU
16(1).	Axillary tufts absent.....	17
	Axillary tufts indistinct or sparse.....	22
	Axillary tufts clearly present.....	23
17(16).	Bract length free 1.9 to 5.9 cm.....	18
	Bract length free 6.5 to 6.8 cm.....	MAX
18(17).	Bract length 6.4 to 7.4 cm.....	19
	Bract length 8 to 12.1 cm.....	20
19(18).	Peduncle length 4.8 to 22.5 mm.....	TOM or OLI
	Peduncle length 25.3 to 36.3 mm.....	MIQ
20(18).	Leaf top: small brown hairs absent.....	MAN
	Leaf top: small brown hairs few.....	MAN
	Leaf top: small brown hairs clearly present.....	21
21(20).	Bract petiole length 0 to 1.5 mm.....	OLI
	Bract petiole length 2.5 to 9.3 mm.....	MIQ
22(16).	Bract width 8.3 to 17.5 mm.....	CAR
	Bract width 20 to 21.5 mm.....	HET
	Bract width 27.3 to 31.3 mm.....	MAX
23(16).	Bract length 4.3 to 5.9 cm.....	CHI
	Bract length 6.4 to 7.4 cm.....	24
	Bract length 8 to 12.1 cm.....	25
	Bract length 13.8 to 14.2 cm.....	28
24(23).	Leaf length 3.9 to 6.9 cm.....	TUA
	Leaf length 7.5 to 11.3 cm.....	HET
25(23).	Leaf margin teeth: pitch 1.7 to 4.6 mm.....	26
	Leaf margin teeth: pitch 5.1 to 5.9 mm.....	27
	Leaf margin teeth: pitch 7.4 to 8.1 mm.....	HEN
26(25).	Leaf underside: small brown hairs few.....	CAR
	Leaf underside: small brown hairs clearly present.....	MAX
27(25).	Bract length free 1.9 to 5.9 cm.....	TUA
	Bract length free 6.5 to 6.8 cm.....	HEN
28(23).	Leaf margin teeth: pitch 1.7 to 4.6 mm.....	HET
	Leaf margin teeth: pitch 5.1 to 5.9 mm.....	TUA

ACKNOWLEDGMENT

Thanks are due to Donald Pigott for many helpful discussions. Gratitude is also due to Simon Owens and Martin Cheek of the Royal Botanic Gardens, Kew for permission to study specimens in the Herbarium. Similarly, thanks are due to Nigel Taylor for permission to collect specimens from living trees in the Gardens. Grateful thanks are also due to Roy Vickery of the Natural History Museum in South Kensington, London for access to the Herbarium.

REFERENCES

- [1] M.J. Dallwitz, "A flexible computer program for generating identification keys," *Systematic Zoology*, vol 23, 1974, pp. 50-57.
- [2] R.J. Pankhurst, *Practical Taxonomic Computing*. UK, University of Cambridge Press, 1991.
- [3] J.A. Freeman and D.M. Skapura, *Neural networks: algorithms, applications, and programming techniques*, Reading, Massachusetts, USA: Addison-Wesley, 1992.
- [4] S. Haykin, *Neural networks - a comprehensive foundation*. New York, USA: Macmillan College Publishing Company, Inc., 1994.
- [5] C.D. Pigott, "Tilia", in *European Garden Flora*, S.M. Walters *et al.*, Eds., pp.205-212, Cambridge University Press, UK, 1997.
- [6] T. Rath, "Klassifikation und Identifikation gartenbaulicher Objekte mit künstlichen neuronalen Netzwerken", *Gartenbauwissenschaft*, vol. 61 (4), 1996, pp. 153-159.
- [7] J.Y. Clark, *Botanical identification and classification using artificial neural networks*, PhD Thesis, Dept. of Cybernetics, University of Reading, UK, 2000.
- [8] J.Y. Clark, "Artificial neural networks for species identification by taxonomists," *BioSystems*, vol. 72, 2003, pp. 131-147.
- [9] C.D. Pigott, Personal communication, 1999.
- [10] G.N. Jones, "Taxonomy of American Species of Linden (Tilia)", *Illinois Biological Monographs*, vol 39: Univ. Illinois Press, 1968.
- [11] M.J. Dallwitz, "A general system for coding taxonomic descriptions", *Taxon*, vol 29, 1980, pp. 41-46.
- [12] T.R. Partridge, M.J. Dallwitz and L. Watson, *A primer for the DELTA system*, 3rd edition, Canberra, Australia: CSIRO Division of Entomology, 1993.
- [13] M.J. Dallwitz, T.A. Paine, and E.J. Zurcher, *User's guide to the DELTA system - a general system for processing taxonomic descriptions*, edition 4.07, Canberra, Australia: CSIRO Division of Entomology, 1997.
- [14] L. Prechelt, "Proben1 - A set of neural network benchmark problems and benchmarking rules" Technical Report 21/94, Universität Karlsruhe, Germany, 1994.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] W.H. Press, S.A. Teukolsky, W.T. Vetterling and B.P. Flannery, *Numerical Recipes in C - The Art of Scientific Computing*. Second Edition, UK: Cambridge University Press, 1994.
- [19] C.D. Pigott, "The taxonomic status of *Tilia insularis*", *The New Plantsman*, vol. 7(3), 2000, pp. 178-183.
- [20] X. Yao, "Evolving artificial neural networks", *Proceedings of the IEEE*, vol 87(9), 1999, pp. 1423-1447
- [21] G.B. Fogel, K. Chellapilla and D.B. Fogel (2003), "Identification of coding regions in DNA sequences using artificial neural networks", in G.B. Fogel and D.W. Corne (eds.), *Evolutionary Computation in Bioinformatics*, San Francisco: Morgan Kaufmann, 2003, pp. 195-218, .