A test of a pattern recognition system for identification of spiders

M.T. Do[1], J.M. Harp[2], and K.C. Norris[3]*

[1]Department of Computer Sciences, University of Tennessee, and Oak Ridge National Laboratory, Life Sciences Division: [2]Graduate School of Biomedical Sciences, University of Tennessee / Oak Ridge National Laboratory: [3]Department of Ecology and Evolutionary Biology, 569 Dabney Hall, University of Tennessee, Knoxville, TN 37996-1610, USA

Abstract

Growing interest in biodiversity and conservation has increased the demand for accurate and consistent identification of arthropods. Unfortunately, professional taxonomists are already overburdened and underfunded and their numbers are not increasing with significant speed to meet the demand. In an effort to bridge the gap between professional taxonomists and non-specialists by making the results of taxonomic research more accessible, we present a partially automated pattern recognition system utilizing artificial neural networks (ANNs). Various artificial neural networks were trained to identify spider species using only digital images of female genitalia, from which key shape information had been extracted by wavelet transform. Three different sized networks were evaluated based on their ability to discriminate a test set of six species to either the genus or the species level. The species represented three genera of the wolf spiders (Araneae: Lycosidae). The largest network achieved the highest accuracy, identifying specimens to the correct genus 100% of the time and to the correct species an average of 81% of the time. In addition, the networks were most accurate when identifying specimens in a hierarchical system, first to genus and then to species. This test system was surprisingly accurate considering the small size of our training set.

Introduction

Physicist and Nobel laureate, Richard Feynman, once said that knowing the scientific name of an organism tells you only the name and nothing else (NOVA, 1975). Though literally true, the scientific name is the handle by which all known information regarding the species can be accessed. The scientific name is also the common 'currency' in biodiversity studies. An incorrect identification can be disastrous (see Davis, 1995; Miller & Rossman, 1995).

Given the need for making accurate identifications, we find that there are two major obstacles. The first is a general lack of funding and personnel for doing taxonomic research to properly classify and describe the vast diversity of organisms. This problem has been comprehensively documented (Systematics Agenda 2000, 1994). The second obstacle to accurate identifications is the difficulty involved in becoming proficient at recognizing arachnids and other arthropods at the species level. Species identification is a daunting task for the non-specialist and the results are often disappointing and inaccurate. The cost of acquiring proficiency is high and, for the non-specialist, the long-term benefit is low.

A partial solution to the problem, which would serve both to alleviate the time demands on taxonomists and to make specimen identification easier and more accurate for the non-specialist, is to partially automate the process. We present a computerized pattern recognition system that, though potentially useful to the systematist, is designed to make the results of taxonomic research available to workers in disciplines that require identification of collected specimens. We have chosen artificial neural networks as our

* Author for correspondence.
Fax: (423) 974 3067
E-mail: kimnorris@utk.edu

https://doi.org/10.1017/S0007485399000334 Published online by Cambridge University Press

pattern recognition tool. Neural networks are programming algorithms which simulate the structure of the brain and the processing of information therein (see Boddy *et al*. (1990) for an introduction to neural networks). Neural networks have been shown to be remarkably apt at learning. They are also capable of detecting subtle differences between similar objects. Once trained, a network can classify objects (e.g. individuals) that it has never encountered before as long as the group they belong to (e.g. genus or species) was part of the training process. It can also be trained to identify unknown objects as such. After the training process the network is time efficient, making rapid identifications while using insignificant computer time and resources.

One of the more commonly known applications of neural network technology comes from the field of forensic science. Finger-print analysis technology developed by NIST (National Institute of Standards and Technology) uses a probabilistic neural network to look for similarities in location of whorls and curves in order to determine whether an unknown print is the same as any it has been previously shown. This is a simpler procedure than training a network to identify an organism to the species level, since there is much greater variability between two individuals of the same species than two fingerprints from the same individual. Handwriting analysis provides a more reasonable comparison, as handwriting varies from signature to signature for the same person.

The possibility of using computer-aided identification systems in biodiversity studies has recently been reviewed by Edwards & Morse (1995) and Weeks & Gaston (1997). Concerned primarily with invertebrate identification, Weeks & Gaston (1997) suggest that multi-access keys, or something similar, might be useful for identification to higher taxonomic levels, but that species identification could be better attained using image analysis and/or neural networks. They review the techniques being developed and the opportunities and limitations of each.

Microbiologists, marine biologists and entomologists have been working on ways to use this technology to differentiate species of bacteria (Bungay & Bungay, 1991; Rataj & Schindler, 1991), classify human cell types (Moallemi, 1991), identify phytoplankton species (Simpson *et al.*, 1992; Boddy *et al.*, 1994; Wilkins *et al.*, 1996), and discriminate between closely related species of ichneumonid wasps (Yu *et al.*, 1992; Weeks *et al.*, 1997). Weeks *et al.* (1997) achieved an identification accuracy of 94% by using principal components analysis to distinguish five species of ichneumonid wasp using information derived from images of wing venation. A disadvantage of this system is its limitation to one sort of information input at a time (i.e. wing morphology). With the exception of the wasp identification, all of these studies made use of neural networks to discriminate visually between groups.

This study is the first attempt that we know of to use neural networks to identify macroscopic organisms. We will demonstrate that this system can classify spider individuals to genus and to species based only on digital images of the ventral view of the female epigyna. In this preliminary study, the training and testing sets are small and the neural networks are required to make identifications based only on single photomicrographs of each test individual. A human making the same identifications would have access to much larger quantities of information. More information could be incorporated into the system, but we hope to demonstrate the utility of this method by showing its performance under such minimal conditions.

Materials and methods

The data

Two species from each of three different genera of the family Lycosidae were used in the training sessions. The species were: *Arctosa rubicunda* (Keyserling), *Arctosa emertoni* Gertsch, *Pardosa groenlandica* (Thorell), *Pardosa dromaea* (Thorell), *Alopecosa aculeata* (Clerck), and *Alopecosa kochii* (Keyserling) (see Dondale & Redner, 1990). These specimens allowed us to test the network's ability to classify individuals in three ways: to genus, to species, and to species within genus (i.e. the program first classifies to genus and then to species for each specimen).

Image acquisition

Due to variability in the condition of the epigyna and the quality of the resulting images, between 14 and 21 individuals of each species were photographed. The epigyna were photographed in 70% ethanol using an Olympus SZX70 microscope equipped with a Sony CCD video camera. Images from the CCD camera were captured using a Snappy™ and recorded in Tag Image Format (TIF). The preparation of each specimen for imaging involved aligning the plate of the epigynum approximately normal to the viewing axis of the microscope through the use of forceps and cotton padding. All specimens were illuminated using a fibre optic light source.

Image pre-processing

The images were cropped to include only the epigynal boundaries. The two dimensional wavelet transform, described below, requires the input image to be a square with dimension $2^j \times 2^j$. As a consequence, the image was scaled down to a dimension of 128×128 pixels. Beyond this routine processing, no attempt was made to scale the images to account for overall body size. An example of a cropped original image is shown in fig. 1A.

Wavelet transforms

Neural networks vary in the amount of information they are designed to receive. In general, a network will have n inputs, corresponding to n numerical values. If n is too large, one of two problems may arise; the computer resources required may be excessive, or the network may lose the ability to generalize when discriminating images unseen in the training set. We found 256 inputs to be the maximum feasible in this study. Thus the input data had to be tailored to the size of the input layer. In this case the input was an image, so one option was to input the greyscale values of each pixel. To do this, the image would have to be reduced to the appropriate number of pixels. For a network with 256 inputs, our maximum, this would mean an image of 16×16 pixels. Such an image contains very little information about the shape of the epigynum, which we know is the most useful information when discriminating species. Therefore, what was needed was a way to decrease selectively the information contained in the original high-resolution image

https://doi.org/10.1017/S0007485399000334 Published online by Cambridge University Press

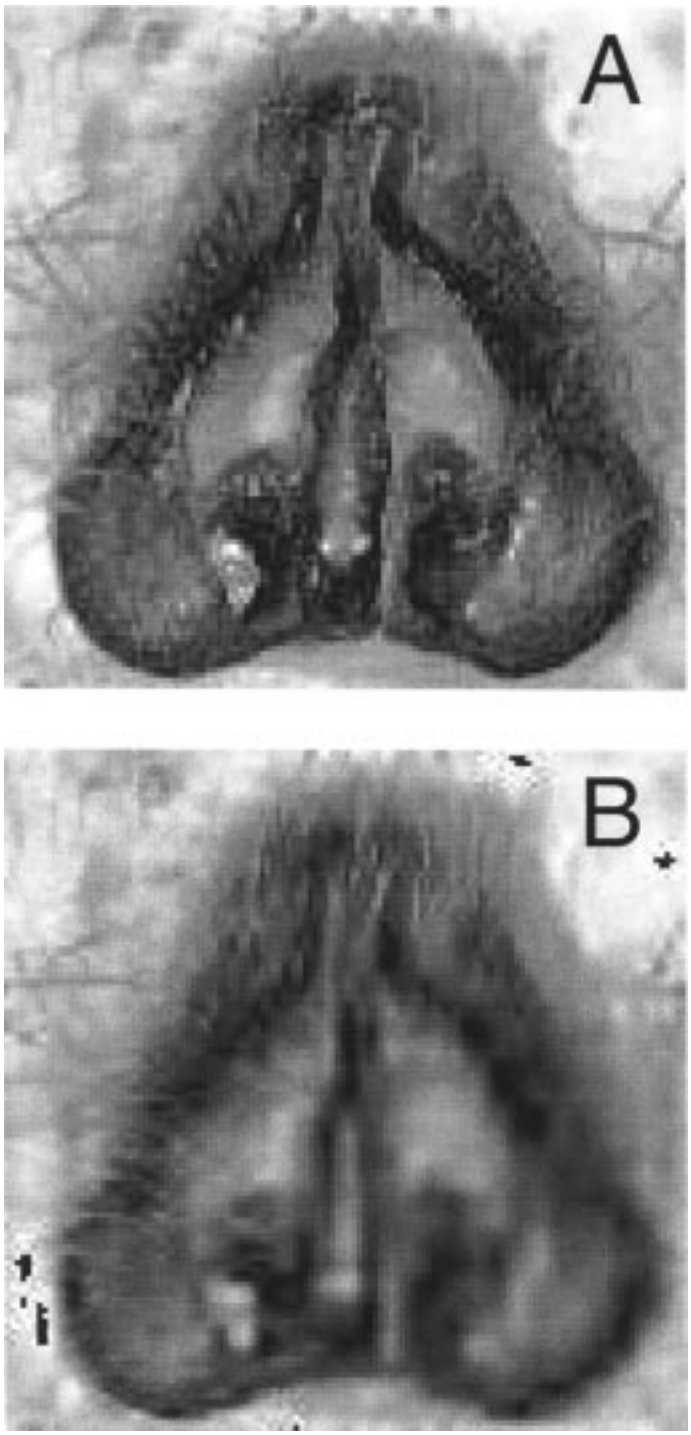

Fig. 1. (A) An epigynum as viewed on the monitor through the CCD camera. (B) An epigynum after wavelet transformation illustrating the loss of high resolution detail with the maintenance of gross shape information.

down to a maximum of 256 values, such that less useful details such as spines and hairs were eliminated while shape information (to allow species identification) was maintained. This was accomplished using a type of mathematical transform called wavelet transform (Graps, 1995).

Wavelet transforms are similar to the more commonly encountered Fourier transform. They are an iterative procedure in which an image is successively reduced to a coarser version of itself, through the removal of high frequency information contained in wavelet coefficients

https://doi.org/10.1017/S0007485399000334 Published online by Cambridge University Press

(sometimes referred to as detail coefficients). These coefficients are parameters that modify the shape of a predetermined function, called a wavelet. The particular wavelet function chosen for this work was the 'Daubechies 4', based on the success of a previous application of neural nets to static signature analysis (McCormack, 1994). At each iteration, the image is partitioned into one vector space containing the low frequency information (low resolution) and three vector spaces containing the higher frequency information. At this point, each vector space contains many wavelet coefficients. In the next iteration, the low frequency vector space is itself partitioned into one low, and three high frequency vector spaces (see fig. 2). Through this procedure, more and more of the high frequency information is removed until, finally, all that is left are the four vector spaces, each with a single coefficient. The original, high resolution image can be reconstructed by successively re-applying the high frequency data represented by the wavelet coefficients contained in the higher frequency vector spaces to the lower frequency information contained in the low frequency vector spaces. Of course, when the image is reduced to only a 4×4 matrix of wavelet coefficients, it will be considerably blurred and barely recognizable. Figure 1B shows the image from fig. 1A, after wavelet transform followed by reconstruction using only the coefficients from the later iterations. The transformed image, while containing as many pixels as the original, shows a loss of high resolution detail. The gross shape of the epigynum, however, is still evident. The input into the network actually consists of the remaining wavelet coefficients which could be used to reconstruct the image.

As mentioned above, the set of coefficients corresponding to each iteration is called a vector space. These vector spaces are numbered as follows: V_0 is the single wavelet coefficient to which the image is reduced, plus its three detail coefficients; V_1 is the four values of V_0, plus the set of 12 detail coefficients that would be applied to the four pixels of V_0 after reconstruction; V_2 is the 16 values of V_1, plus the corresponding 48 detail coefficients; and so on (see fig. 2). Thus, for our smallest ANN, the SANN (16 inputs), we use the coefficients from V_1, for the medium sized ANN, the MANN (64 inputs), we use V_2, and for our largest ANN, the LANN (256 inputs), we use V_3. The larger networks get input from detail coefficients that correspond to finer resolution, but none of our networks get the finest detail from the original image, which has 128×128 = 16383 pixels and would require V_5.

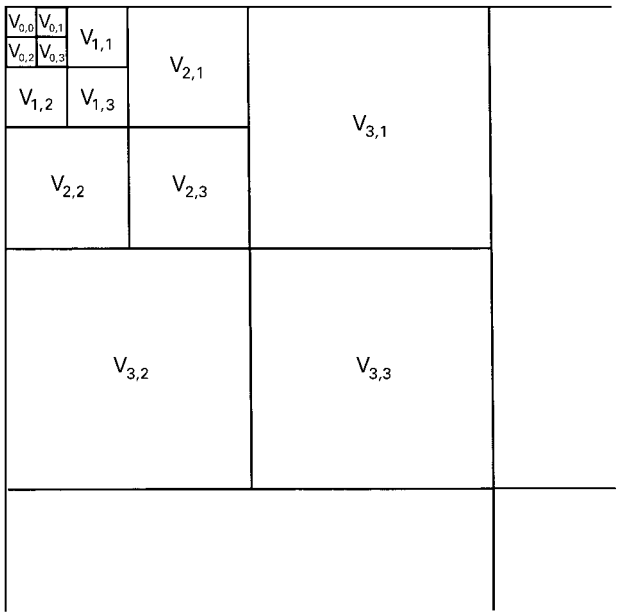

Fig. 2. A wavelet coefficient layout diagram for an image that has undergone wavelet transformation. The higher resolution information is represented by the wavelet coefficients at the higher level vector spaces (toward the right side of the figure).

Image classification

Artificial neural network structure

An artificial neural network (ANN) is a computing algorithm based on a simplistic model of the brain or, perhaps more accurately, a ganglion. The massively parallel architecture of the ANN consists of multiple layers of simple computing elements with many interconnections between the layers. The computing elements are functionally analogous to neurons. They receive signals and in turn transmit a signal which is a function of the inputs. The function by which the inputs are evaluated may be a simple logic gate but more generally involves summation of weighted input signals. A threshold function is then applied to the weighted inputs to determine the output of the neuron. A simplified ANN architecture is presented in fig. 3. This is a fully connected three layer network.

The initial architectures of our ANNs were established according to the number of input neurons and the number of classifications that the program was being designed to distinguish. Each initial ANN consisted of a layer of input neurons and a layer of output neurons, fully interconnected by random initial weights. Each input layer neuron corresponded to a wavelet coefficient, which represented the detail contained in a set of pixels. Each output neuron was assigned to a genus or species that we were attempting to identify. Separate ANNs were developed and tested with 16, 64, and 256 inputs (corresponding to different amounts of detail in the images) and three, six, or two output neurons (corresponding to identification to genus, species, or species within genus). These last ANNs, with only two output neurons, could then be used to evaluate the ability of the pattern recognition system to identify specimens first to genus, then to species within a genus – the hierarchical approach. Other ANNs were developed with six output nodes and trained using the entire set of species in the three genera. These ANNs were used to evaluate the performance of the ANN in identifying species without regard to genus classification – the nonhierarchical approach.

After acquiring and processing the images from all samples, the set of images was divided into a training set and a testing set. The composition of the training set is shown in table 1. Subsets of the entire testing set were made to test ANNs at different hierarchical levels as described above.

https://doi.org/10.1017/S0007485399000334 Published online by Cambridge University Press

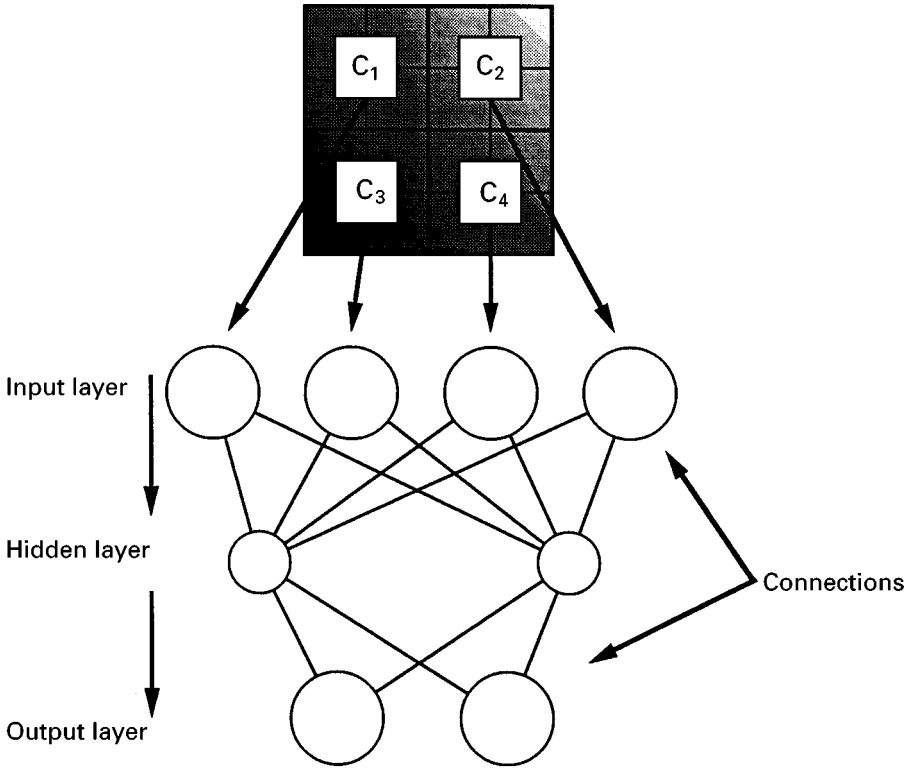

Fig. 3. A simplified three layer, fully connected artificial neural network showing the input of a hypothetical image containing four wavelet coefficients (C1 through C4), encapsulating the information of 16 pixels. The number of nodes in the input and output layers are determined by the number of inputs to the network and the number of classification options, respectively.

The training cycle

The following is a generalized description of the steps necessary to train and test any ANN, along with details specific to this study. Training involves effort in three phases: (i) assembly of the collection of object classes (in our case, spider species) by a taxonomist; (ii) construction of the training data set and processing of the data into a form capable of input into the neural network (digital images in this study); and (iii) the actual training of the neural network to recognize the object classes based on the training data set. Training continues until the desired level of accuracy is attained. It is at this point that the network is tested with previously unseen individuals to assess its ability to classify them into appropriate groups based on what it has learned from the training process (i.e. the network's ability to generalize). The idea is to 'teach' the ANN to set the output neuron, assigned to a given genus or species, to its maximum value of '1.0' when it sees a pattern indicative of that genus or species and set all other output neurons to a minimum value of '0.0'. In practice, the ANN will set the output neurons to an intermediate value depending on the certainty of its identification (e.g. an output of 0.9999 indicates certainty while 0.6 indicates less certainty). The training process introduces new neurons in a hidden layer between the input and output layers. These act as feature detectors to look for a specific pattern unique to a given genus or species. The resulting output vector is then evaluated against the target function to compute an error.

Table 1. Composition of the training sets used to train the artificial neural networks.

Training set	Genus	Species	N
	Alopecosa	*A. aculeata*	11
		A. kochii	7
Lycosidae	*Pardosa*	*P. groenlandica*	8
		P. dromaea	10
	Arctosa	*A. rubicunda*	8
		A. emertoni	9

The training sets consisted of digitized images of epigyna taken from a number, N, of individual specimens.

This error is then used to modify the weights in the connections. An entire training cycle is referred to as an epoch. Training stops when the error becomes sufficiently small (in our case 0.001). The number of epochs required to reach the target error varies considerably (tens to many thousands) depending on the ease/difficulty in discriminating between the output groups and the amount and resolution of data being used. In this work, all ANNs were trained over 300 epochs and adjustable parameters such as learning rate and momentum were held constant.

Several algorithms exist to train an ANN, each of which possesses certain strengths and weaknesses. We used Cascade correlation in conjunction with quick propagation

https://doi.org/10.1017/S0007485399000334 Published online by Cambridge University Press

(Fahlman, 1988, 1991). Cascade correlation adds hidden layer neurons one at a time throughout training. This allowed for quick learning and the ability to generate a near minimal ANN consisting of only those feature detectors that are needed. Minimizing computational time and resources is essential if this technology is to be widely accessible.

The time required for training should not be overly long. However, since the performance of the ANN is very fast after training, sufficient attention should be paid to the training process as to make the final structure as accurate as possible. At present, training the ANN requires considerable attention from the human operator. As we gain more experience with the behaviour of the algorithms when applied to spiders, the process should become standardized.

Testing

For each size of network (SANN, MANN, LANN) and for each level of identification (genus, species alone, and species within genus), inputs were taken from subsets of a testing set consisting of images which the network had not seen before. The composition of the testing set can be read from tables 2–4. Specimens from unknown species or genera were not presented to the network as part of this study.

Results

Recognition of epigyna at the genus level

Training was performed over the entire Lycosidae training set (17–18 individuals per genus). The results of testing runs on the trained ANNs are given in table 2. Both the MANN and the LANN were 100% accurate in their identifications to genus, while the SANN had an average accuracy level of 90%. Therefore, it is apparent that the system was more than adequate for distinguishing spiders to genus based solely on shape features of the female epigynum. This was true even given the small size of the training set**.**

Recognition of epigyna at the species level

The non-hierarchical approach

Only the MANN and LANN networks were used for this experiment. The ANNs were trained over the entire Lycosidae training set. The results of testing runs on this series of ANNs are given in table 3. The overall accuracy level for identifications to species for the MANN and LANN were 69% and 73%, respectively. The overall results suggest that the LANN performed better than the MANN although some peculiarities exist in the results. The ability of the LANN to identify *Alopecosa* species was considerably better than that of the MANN but the performance of the LANN actually fell below that of the MANN for *A. rubicunda.* We are unable to explain these discrepancies except to point out that the training set was quite small (only 7–11 individuals per species). It should be noted that the probability of correctly identifying, by chance, a specimen taken at random from a set containing six species is only 16.7%.

The hierarchical approach

The SANN, MANN, and LANN were trained over the appropriate subset of the Lycosidae training set. The results of testing runs for all three sets are given in table 4. The overall accuracy levels for the SANN ranged from 63% to 75%, from 69% to 100% for the MANN and from 74% to 88% for the LANN. Here again, the overall performance of the ANN improved with increasing number of included vector spaces represented in the input vectors. In this case, the performance of the LANN was less than that of the MANN for *A. kochii* and the greatest improvement with size of ANN was seen in the *Arctosa* ANNs. It is clear that the SANN was inadequate but these results suggest that there may a point at which increasing the size of the ANN will not significantly improve performance.

Table 2. Testing results for the artificial neural networks trained using the full Lycosidae training set.

Taxon	N	SANN (%)	MANN (%)	LANN (%)
Alopecosa	16	100	100	100
Pardosa	19	74	100	100
Arctosa	16	100	100	100
Overall	51	90	100	100

The ANNs possessed three output neurons corresponding to the three genera in the training set and varied in the number of vector spaces included so that the small ANN (SANN) contained 16 input neurons, the medium ANN (MANN) contained 64 and the large ANN (LANN) contained 256 input neurons. The data represent the percentage of correct responses from the ANN out of the total number, N, of unknowns presented to the trained ANN.

Table 3. Non-hierarchical testing results for the artificial neural networks (MANN and LANN, only) trained on the Lycosidae training set.

Taxon	N	MANN (%)	LANN(%)
Alopecosa aculeata	10	70	80
A. kochii	7	57	71
Pardosa groenlandica	9	44	44
P. dromaea	9	89	89
Arctosa rubicunda	9	67	56
A. emertoni	9	86	86
Overall	52	69	73

The data represent the percentage of correct responses out of the total number, N, of unknowns presented to the trained ANN.

Discussion

The performance of the trained ANNs in the testing runs demonstrated that the pattern recognition system was capable of identifying lycosids with an accuracy level of 100% to genus and an average of 81% to species. This level of accuracy was surprising given the small size of the training set used in this study (7–11 individuals per species) and the limited amount of information on which the identification system was based. We feel that the size of this training set was inadequate to assess fully the abilities of the pattern recognition system; this was designed solely as a proof-of-principle study to evaluate the feasibility of using such pattern recognition systems for species identification in spiders. Increased accuracy could be attained simply by increasing the size of the training set (Simpson *et al.*, 1992).

https://doi.org/10.1017/S0007485399000334 Published online by Cambridge University Press

Table 4. Hierarchical testing results for the three artificial neural networks using training sets containing only specimens within a single genus.

Training set	Taxon	N	SANN (%)	MANN (%)	LANN (%)
Pardosa	*P. groenlandica*	10	40	50	50
	P. dromaea	9	89	100	100
	Overall	19	63	74	74
Arctosa	*A. rubicunda*	9	56	56	78
	A. emertoni	7	86	86	86
	Overall	16	63	69	81
Alopecosa	*A. aculeata*	8	75	100	100
	A. kochii	8	75	100	75
	Overall	16	75	100	88

The data represent the percentage of correct responses out of the total number, N, of unknowns presented to the trained ANN.

We also limited the information input to shape features of the female epigynum. A fully developed system could incorporate more types of information (e.g. carapace length / width, locality, distinguishing markings, etc.) which would again increase accuracy, particularly when species have very similar genitalia. In addition, the use of higher quality cameras, microscopes, etc. during the training process might enhance precision. However, the accuracy obtained through the use of such accessible equipment as used in this study further demonstrates the potential of this system and the feasibility of its use by the general scientific community. Once trained, the neural network software could run on most personal computers.

Based on the comparative performances of the small (SANN) and large (LANN) networks, it was apparent that the SANN did not incorporate adequate data to distinguish spider species. Although it is clear that future ANNs should be at least the size of the MANN (medium sized with 64 input neurons), it is not clear whether they will necessarily always perform better with larger amounts of input neurons. The data also indicated that a hierarchical ANN was preferable to a non-hierarchical ANN. In other words, higher accuracy was attained when the network first classified the specimen to genus and then to species as opposed to going straight to species (81% vs. 73%, respectively). We would expect this to be true as long as the higher organizational group (in this case, the genus) is well defined with respect to the characters examined (in this case, the epigyna). Boddy *et al.* (1994) found that the hierarchical approach was not as good when identifying phytoplankton because the classification scheme used was not compatible with the characters fed into the network. Some species within one group resembled those in another group enough that they were always initially misclassified and therefore always misidentified. In the identification of spiders using genitalia, we speculate that although it would be beneficial to use the hierarchical approach to the level of genus, it would not be acceptable to have the network classify to the family level since these classifications are often made based on other characteristics. A separate network for each family may be the best option, since identification to family is more easily accomplished by the non-specialist. In future work, we will explore these options by testing the network on more than one family group.

In this study, we chose to use the epigynum as a test case. The essentially two dimensional nature of the ventral view of the structure served to simplify technical aspects of the work. We also used a subset of spiders which can be readily identified based on external genitalia; this will not always be the case, as some species can only be identified based on internal genitalia or other characteristics. Nonetheless, this system has the potential to read any visual input and artificial neural networks can be designed and trained specifically for any group. The next step in the development process will be to make an ANN capable of identifying species on the basis of a single view of the adult male spider's genitalia, the palpus. Future development will involve the training of an ANN using multiple views of the adult male palpus or female epigynum for those species which require this information for accurate identification. It is proposed that the ANN will avoid the need of reconstructing a three-dimensional image of these complex structures simply by learning to recognize a set of two, two-dimensional images taken at arbitrary angles. This technology is not limited to arachnology. The methods being developed here will transfer to any other taxa for which visual traits are used to distinguish genera or species (e.g. wasps (wing venation), scale insects (scales), many other insect groups (genitalia)).

We have undertaken this study as a first step toward making routine, accurate species identification of spiders accessible. We suggest that the ANN as a pattern recognition system will be useful in making the results of taxonomic research widely available by encapsulating the subjective impression of shape, and variability in shape, and making that information available to any user who may be viewing a structure for the first time. Because the network is trained on multiple individuals of each species, it has the advantage of incorporating intraspecific variation. Obviously, this system must be tested on a much larger set of species before we will know for sure how well it will perform as a tool in biodiversity studies. If the system continues to perform adequately when processing larger numbers of genera and species and / or different kinds of visual data, we suggest that these systems could become one end product of taxonomic research. Databases could incorporate artificial neural networks to make taxonomic knowledge available to anyone who could benefit from the ability to make accurate identifications of genera or species. At the very least, these systems could be used in biological monitoring programmes where the ANN could be trained with the initial species collected at the site of interest. All subsequent collections would then at least be 'standardized' to the original, regardless of changes in taxonomic nomenclature and / or personnel.

https://doi.org/10.1017/S0007485399000334 Published online by Cambridge University Press

We should emphasize that the methods used in this work cannot supplant the role of the taxonomist. Pattern recognition systems are only useful for encapsulating the results of taxonomic investigations and are not capable of independently ordering or explaining organic diversity. The ANN is a system which can simulate learning and make use of that learning. But, one should bear in mind that the software can only learn what it is presented during the training process. Also, the ANN cannot replace the revision as an end product of taxonomic investigation. What we have done here is to address the limitation of the revision as a vehicle for making the results of taxonomic investigation accessible to the end user. By freeing taxonomists from the burden of identifying collections of known species sent to them by other workers, they will have more time for the most important aspect of their work – description and revision of species and species groups.

Acknowledgements

We would like to thank Dr C.D. Dondale for generously loaning us the majority of specimens used in our study. Thanks to Dr S.E. Riechert, Dr G.J. Russell, and two anonymous reviewers for their helpful comments on the manuscript. We would also like to thank Dr G.J. Bunick and Dr P. Lambdin for giving us access to their imaging equipment. Thanks to Dr Bruce Maclennan who served as Mr Do's advisor during this project.

References

Boddy, L., Morris, C.W. & Wimpenny, J.W.T. (1990) Introduction to neural networks. *Binary* **2**, 179–185.

Boddy, L., Morris, C.W., Wilkins, M.F., Tarran, G.A. & Burkill, P.H. (1994) Neural network analysis of flow cytometric data for 40 marine phytoplankton species. *Cytometry* **15**, 283–293.

Bungay, H. & Bungay, M.L. (1991) Identifying microorganisms with a neural network. *Binary* **3**, 51–52.

Davis, G.M. (1995) Systematics and public health. *BioScience* **45**, 705–714.

Dondale, C.D. & Redner, J.H. (1990) *The wolf spiders, nurseryweb spiders and lynx spiders of Canada and Alaska. Insects and arachnids of Canada* part 17. Ottawa, Agricultural Canada.

Edwards, M. & Morse, D.R. (1995) The potential for computer-aided identification in biodiversity research. *Trends in Ecology and Evolution* **10**, 153–158.

Fahlman, S.E. (1988) An empirical study of learning speed in back propagation network. Technical Report. CMU-CS-88-162, Carnegie Mellon University, Pittsburgh, Pennsylvania.

Fahlman, S.E. & Lebiere, C. (1991) The cascade-correlation learning architecture. Technical Report. CMU-CS-90-106. Carnegie Mellon University, Pittsburgh, Pennsylvania.

Graps, A. (1995) An introduction to wavelets. *The Institute of Electrical and Electronics Engineers (IEEE) Computational Science and Engineering* **2**, 50–61.

McCormack, D.K.R. (1994) An investigation into the representation of data for the neural implementation of a handwritten static signature verification system. PhD thesis, University of Wales, College of Cardiff.

Miller, D.R. & Rossman, A.Y. (1995) Systematics, biodiversity, and agriculture. *BioScience* **45**, 680–-696.

Moallemi, C. (1991) Classifying cells for cancer diagnosis using neural networks. *IEEE Expert* **6**, 8–12.

NOVA (1975) 'Take the World from Another Point of View', episode origially broadcast 2 February 1975, PBS.

Rataj, T. & Schindler, J. (1991) Identification of bacteria by multi layer neural networks. *Binary* **3**, 159–164.

Simpson, R., Williams, R., Ellis, R. & Culverhouse, P.F. (1992) Biological pattern recognition by neural networks. *Marine Ecology Progress Series* **79**, 303–308.

Systematics Agenda 2000 (A Consortium of the American Society of Plant Taxonomists, the Society of Systematic Biologists, and the Willi Hennig Society, in cooperation with the Association of Systematics Collections) (1994) Systematics Agenda 2000: Charting the biosphere. Technical Report.

Weeks, P.J.D. & Gaston, K.J. (1997) Image analysis, neural networks, and the taxonomic impediment to biodiversity studies. *Biodiversity and Conservation* **6**, 263–274.

Weeks, P.J.D., Gauld, I.D., Gaston, K.J. & O'Neill, M.A. (1997) Automating the identification of insects: a new solution to an old problem. *Bulletin of Entomological Research* **87**, 203–211.

Wilkins, M.F., Boddy, L., Morris, C.W. & Jonker, R. (1996) A comparison of some neural and non-neural methods for identification of phytoplankton from flow cytometry data. *Cabios* **12**, 9–18.

Yu, D.S., Kokko, E.G., Barron, J.R., Schaalje, G.B. & Gowen, B.E. (1992) Identification of ichneumonid wasps using image analysis of wings. *Systematic Entomology* **17**, 389–95.

(Accepted 2 March 1999)
© CAB *International*, 1999

https://doi.org/10.1017/S0007485399000334 Published online by Cambridge University Press