Automatic categorisation of five species of *Cymatocylis* (Protozoa, Tintinnida) by artificial neural network

P. F. Culverhouse[1], R. Ellis[2], R. G. Simpson[1], R. Williams[3], R. W. Pierce[4], J. T. Turner[5]

[1]School of Electronic, Communication and Electrical Engineering, University of Plymouth, Plymouth PL4 8AA, United Kingdom

[2]Department of Psychology, University of Plymouth, Plymouth PL4 8AA, United Kingdom

[3]Plymouth Marine Laboratory, Natural Environment Research Council, The Hoe, Plymouth PL1 3DH, United Kingdom

[4]Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island 02882, USA

[5]Biology Department and Center for Marine Science and Technology, University of Massachusetts, Dartmouth, Massachusetts 02747, USA

ABSTRACT: Photomicrographs of 5 species of *Cymatocylis* were digitised, binarised and edited by hand to remove large debris contaminating the images. An artificial neural network (back-propagation of error) was trained to categorise 201 of these specimens after pre-processing the data by Fourier transformation. Of the 299 trials which were carried out, 28% demonstrated better than 70% correct categorisation of the data used in the training sets. The best performing network learned to differentiate the training data set with an error rate of 11%. The same network gave an error rate of 18% when presented with previously unseen data. The results of training back-propagation of error networks are presented and the performance and limitations are discussed and compared with more classical morphometric and clustering techniques for the taxonomic separation of marine plankton. This automatic technique demonstrates the potential of neural network pattern classifiers for addressing the difficult taxonomic task of congeneric classification and also has wider implications for the automatic identification of field samples of marine organisms.

KEY WORDS: Tintinnid · Taxonomy · Neural network

INTRODUCTION

Following the work of Simpson et al. (1991, 1992, 1993) and Williams et al. (in press) on biological pattern recognition of phytoplankton by neural networks, we have attempted to expand this work into microzooplankton. There is a growing requirement within ecological research for automatic techniques for taxonomic classification. The use of neural networks, in biological science, is seen to be a 'new concept' which could potentially realise this goal (Culverhouse in press). It must be recognised that artificial neural networks are not just simple techniques which can be readily applied to the problem. It is not a 'silver bullet'

technique and a great deal of work has to be done, not only to assess the efficacy of the technique but to explore its general applicability within marine biological science. There are over 20 types of neural networks, some of which are more applicable to this problem than others.

One of the main objectives of this work is to be able to take microscopic imagery of plankton from 'field' samples, directly or through a video system, into an associated computer to run through an artificial neural network classifier to obtain accurate taxonomic identifications. Reported work to date has been on prepared outline imagery of biological specimens (Simpson et al. 1993) or from flow cytometry of cultured phytoplank-

ton species (Boddy et al. in press). The problem which has to be addressed now is the 'real time' identification of net, bottle or direct *in situ* video-sampled plankton which requires separation of multiple, overlapping and detritus-contaminated imagery, highly diverse species and different orientations of the biological specimens to the field of view. It will be many years before adequate solutions are found to these problems. In the meantime the neural network application shows great promise. Certain advances, such as using digitised images of naturally occuring plankton in place of camera lucida drawings, using binarised images which are only partly cleaned of debris prior to processing and finally increasing the categories to be identified from 2 to 5 by the neural network, are presented. The task was made deliberately more difficult by using congeneric species of the tintinnid genus *Cymatocylis*, which taxonomists already experience difficulty in separating.

The classical and new multivariate statistical approaches used to separate the species of this genus are given in Williams et al. (1994, this volume), but here we are developing and assessing a new technique to be used eventually for identification of individual species in plankton samples.

METHOD

Image pre-processing. Photomicrographs of 5 species of the genera *Cymatocylis* (*C. calyciformis*, *C. drygalskii*, *C. vanhöffeni*, *C. convallaria* and *C. parva*) were obtained from microzooplankton samples taken off South Georgia (Williams et al. 1994).

Photomicrographs of 201 examples of the 5 species were digitised, binarised and Fourier transformed. The images were digitised using a monochrome Panasonic wv-cd50 camera (resolution 768 by 576 pixels, 1:1 aspect ratio) with 18 mm focal length autoiris lens and stored as digital images on a Sun 3/160 computer using a 16 MHz bandwidth framestore. Each image was clipped from 768 by 576 pixels to 512 by 512 pixels for editing and subsequently manually thresholded to provide binarised images. Large debris and other occlusions were also removed by manually painting each image (using a computer editing program). Small debris (less than approximately 50 pixels in area) were left in images where they occurred. Images were then reduced further in resolution by pixel averaging to a final size of 256 by 256 pixels to provide suitable-sized images for the 2-dimensional Fast Fourier Transform processing (2D-FFT). The resulting Fourier power spectra were converted to normalised, differenced 1-dimensional histograms as described in Simpson et al. (1991). The

resulting data constituted the data set for the neural network learning trials.

The Fourier processing employed here provides information that is invariant to ambient lighting conditions, specimen rotation and specimen translation. It does not, however, provide invariance to specimen scaling or specimen morphological variation.

Fourier transforms of the original grey-level images (as against the binarised images) were not used directly in this study. This was due to the large amount of debris cluttering some of the images. Although it has already been noted that small debris was left in the edited images, it proved impossible to remove the larger debris from the grey-level images without introducing additional features to the images as a result of the editing operation. The choice to use cleaned binary or raw grey-level images in this study was resolved by attempting to train neural networks directly on the Fourier components derived from the raw images. The networks failed to perform much better than chance, a fact assigned to the additional degrees of freedom obtained by this pre-processing method and the small size of the data sets (unpubl. results).

Data validation. Prior to the experimental work each specimen's taxonomic label was validated by an independent group of 6 experts. Procedurally the 201 photomicrographs were replicated 3 times and presented to each expert in random sequence, thus requiring 603 image categorisations from each expert. Consistency within an expert and between experts was evaluated. Experts were found to be >91% consistent with themselves and with each other. Given this high degree of consistency, the results from all the experts were used to select the validated data set for neural network training and testing. Of the 201 images, 18 were not categorised with 100% consistency across the panel of experts. These were removed from the data set of images used to train and test the neural networks. This validated data set was used in all experiments reported in this paper.

Neural network configuration. Back propagation of error networks (BPN) were generated, trained and assessed with 'unseen' data as reported in Simpson et al. (1991). Network weights were randomly set within the numerical bounds of ±0.1. Fourier histograms were presented in random sequence to each network for 1000 training iterations. Data were split into 2 parts: (1) data used for training and (2) data used for testing network generalisation ability. Trials were conducted with randomised training set size and randomised generalisation set sizes within the bounds of the total data set.

The network topology adopted for this research was a 15-3-5 net, a between-layers fully connected architecture (as previously reported). The frequencies were

chosen to reflect the minimum set of harmonic series in which the salient features could still be discerned by experts.

Individual species were not represented equally in the data set (due to availability of specimens) and comprised (A) 40 *Cymatocylis calyciformis*, (B) 24 *C. drygalskii*, (C) 38 *C. vanhöffeni*, (D) 43 *C. convallaria*, and (E) 40 *C. parva*. In an attempt to correct for the small data set size, stochastic noise was added to each of the Fourier histograms during training (Györgyi 1990). *C. drygalskii* were underrepresented in the data set and in an attempt to compensate, 18 images, drawn at random from the population, were duplicated into the data set which, in conjunction with the additive noise, balanced the training population size for *C. drygalskii*. This data set totalled 201 images.

RESULTS AND DISCUSSION

Of the 100 trials using the above network and data pre-processing configurations, 84 succeeded in learning the training set data and performing the 5-way discrimination with less than a 30% root-mean-square error. The performance of the 3 top performing networks is summarised in Table 1. The training sets for all 3 networks constituted less than half the data set of 201 images. Network trial number 219 demonstrates a mean training error of 11% and a mean test error (on unseen data) of 23%; this performance was achieved with a training set size of 103 images and a test set of 98 images. Network trial number 202 gave similar performance measures of 15% mean training error and 33% mean test error over 108 training images and 93 test images. It should be re-stated that a network is trained on the training data and tested on unseen data (the test data) to assess its ability to generalise from the training data.

The test data performance may be represented as tables of confusion, where the categorisations carried out by each network are compared to the desired categorisations (as defined by the validations of the expert panel of taxonomists and by the morphological discriminant analyses). Confusion matrices for networks 219 and 202 are presented in Table 2.

Table 1. Performance measures for the (3) best performing network trials

Network	Mean training error (%)	Mean test error (%)	Actual test error (error/test size)
202	14.5	32.5	30/90
219	10.5	23.0	23/98
224	16.1	40.0	39/98

Confusion matrices were constructed using the entire data set aggregating the training and test data sets, since the data sets were relatively small, with a maximum individual species specimen count of 43 for *Cymatocylis convallaria*. Although this hides the classification results for the 2 classes of data, it does allow comparison with both the expert panel and the morphological discriminant analyses. Both networks show good overall performance for 4 of the 5 species in the data set. However the overall classification scores are pulled down by the rather poor discrimination for *C. vanhöffeni* (C in Table 2) with regard to *C. drygalskii* (B) and to *C. calyciformis* (A). Two species, *C. convallaria* (D) and *C. parva* (E), separate well in both networks, with performances similar to that of the expert panel. The remaining 2 species, *C. calyciformis* (A) and *C. drygalskii* (B), are categorised with less than 12% error (network 219) and 17% error (network 202) respectively.

Extending the comparisons between neural network and experts to the untested data that lay outside the consensus data set (unclassified data) of the expert panel, it can be observed in Table 3 that the categories found difficult by experts were also found difficult to classify by the neural networks as well as by the morphological discriminant analysis, which showed that species C (*Cymatocylis vanhöffeni*) and species B (*C.drygalskii*) caused some problem.

More specifically, 3 of the 18 unclassified (by experts) specimens were also unclassed by the discriminant analysis. Nine of the remaining 15 speci-

Table 2. Confusion matrices, showing discrepencies between network categorisations and the desired categorisations. A: *Cymatocylis calyciformis*; B: *C. drygalskii*; C: *C. vanhöffeni*; D: *C. convallaria*; E: *C. parva*

Network 202	A	B	C	D	E
Size:	40	24	38	43	40
A	33	1	8	1	0
B	3	20	9	0	0
C	1	3	20	0	0
D	3	0	0	42	0
E	0	0	1	0	40
%	83	83	53	98	100

Network 219	A	B	C	D	E
Size:	40	24	38	43	40
A	37	0	5	1	0
B	0	21	10	0	0
C	1	3	23	0	0
D	2	0	0	42	0
E	0	0	0	0	40
%	93	88	61	98	100

Table 3. Expert unclassified data shown against unclassed/inconsistent data in networks and discriminant analysis. A: *Cymatocylis calyciformis*; B: *C. drygalskii*; C: *C. vanhöffeni*; D: *C. convallaria*; E: *C. parva*

	A	B	C	D	E
Expert panel	2	3	7	5	1
Network 202	7	4	18	1	0
Network 219	3	3	15	1	0
Discriminant	0	4	4	2	0

Table 4. Confusion resolution for data unclassified by discriminant analysis. Networks 202 and 219. A: *Cymatocylis calyciformis*; B: *C. drygalskii*; D: *C. convallaria*

Specimen	Initial classification	Discriminant	Network 202	Network 209
60	D	Unclassed	A	A
32	B	Unclassed	A	A
29	B	Unclassed	D	D

Table 5. Internal body contents size of *Cymatocylis* spp. as a percentage of body area (from thresholded images), A: *C. calyciformis*; B: *C. drygalskii*; C: *C. vanhöffeni*; D: *C. convallaria*; E: *C. parva*

	A	B	C	D	E
< 10 %	28	7	19	16	25
< 50 %	3	14	17	22	9
> 50 %	10	6	7	9	5
Total	41	27	43	47	39

mens were correctly classed by network 202 by assuming the correct class to be that defined by the morphological analysis, and 8 of the 15 were correctly classed by network 219.

There were 3 specimens that were unclassified using discriminant analysis to 95 % confidence and also were not in the set of expert consensus: specimen numbers 60 (*Cymatocylis convallaria*), 32 and 29 (both *C. drygalskii*). These were also mis-classified by the best performing networks (BPN) as shown in Table 4, but interestingly both network 202 and network 219 mis-classed these specimens in a uniform manner, ascribing specimen 60 to *C. calyciformis* instead of *C. convallaria*, and specimens 32 and 29 to *C. calyciformis* and *C. convallaria* instead of *C. drygalskii* as initially identified by an 'expert' tintinnid ecologist (Pierce & Turner 1993).

The present set of training trials were carried out on only a small number of images per species, and the variation in image data due to specimen gut density was thought to be responsible for the mis-categorisations. Table 5 summarises the internal structure of each specimen in the data set as a percentage of body area. The differences in body contents may be observed in the selected images presented in Fig. 1, where the *Cymatocylis calyciformis* and *C. convallaria* images both show bright areas of internal structure in the body, whereas the *C. vanhöffeni* and *C. parva* images show none.

There were specimens with substantial amounts of body content for all species in the data set. However, performing a linear correlation between network performance and image body occlusions (described in Table 5) showed there was no correlation between size of the body contents and neural network generalisation performance (Pearson's product moment correlation of less than 0.2 for all categories in Table 3 against performance for network 219 and network 276, with 3 degrees of freedom) this result demonstrates an additional resilience to noise within the data set by the neural network paradigm.

Both of the BPNs (202 and 219) show a high degree of similarity in their results. Both networks fail to attain a minimum of 83 % correct categorisation on only 1 of the 5 species in the study, *Cymatocylis vanhöffeni*. Both find the remaining 4 species relatively easy to dis-
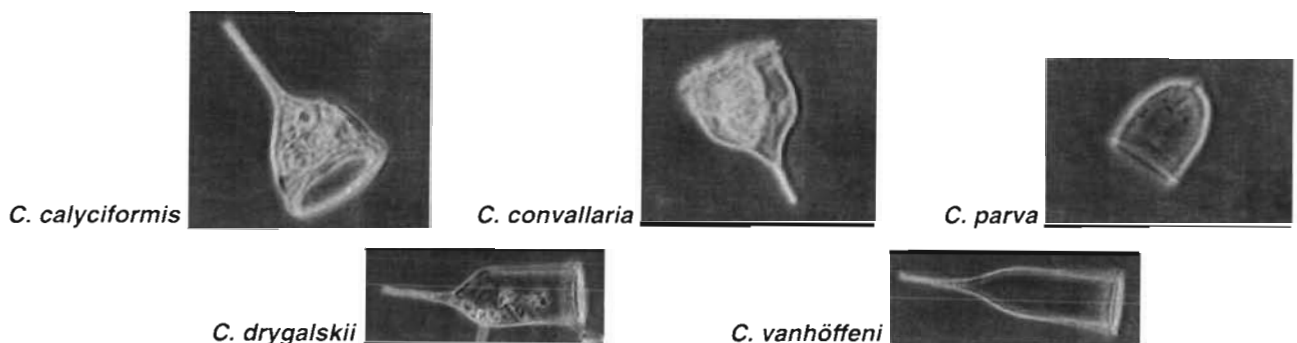

Fig. 1. *Cymatocylis* spp. used in the trials

tinguish, but with some confusion between *C. vanhöffeni* (C) with regard to *C. drygalskii* (B) and to *C. calyciformis* (A). The relationship between network learning and generalisation ability and obvious morphological differences present in the specimens is difficult to discover, due to the 2-dimensional Fourier pre-processing and the subsequent collapse to a 1-dimensional histogram as a final data reduction.

Overlays of the Fourier components of each specimen for each species in the neural network data set are shown in Fig. 2. Treating these as signatures, it may be observed that *Cymatocylis parva, C. convallaria* and *C. calyciformis* all show strong 'average' values that stand out in a visual inspection of the graphs. *C. vanhöffeni* and *C. drygalskii* have less clear signatures. This difficulty is perhaps indicative of the problem facing the BPN when being trained with these data. The variance for each species' Fourier histograms correlates well with the difficulty experienced by the BPNs.

This is strongly suggestive of the problem facing the BPNs that the performance of the neural networks is limited by the paucity of information given by the current pre-processing method. This may be because features evident in the original data (spatial representations of the specimens) may well be hidden or obscured by the Fourier analysis. The collapse of 2-dimensional information into a single histogram has great value as a data reduction technique and in reducing neural network complexity, although it can have disadvantages.

We are currently exploring other pre-processing techniques suitable for neural network utilisation, as well as developing extensions to the existing Fourier technique. The measure of utility of an artificial neural network should be the degree of correlation between the taxonomic description of a specimen (obtained through a consensus of experts) and the categorisation developed by a neural network when trained on other images of the same species. Expert opinion is normally based on a number of morphological and ecological parameters. Williams et al. (1994) demonstrate this for the 5 species in question here.

A direct comparison between these morphometric and ecological tests and BPN categorisations provides a way of objectively observing the categorisations obtained by experts and neural networks. Figs. 3 & 4 show such an objective morphometric function for the 5 tintinnid species. Overlaid on Fig. 4 is the set of specimens that the experts failed to form a consensus on; Fig. 3 is overlaid with the set of specimens that the artificial neural network (trial 219) failed to co-categorise with expert opinion.

It can be seen that in all 5 species, experts found specimens that are either between cluster boundaries (image numbers 32, 47 and 60 for example) or within
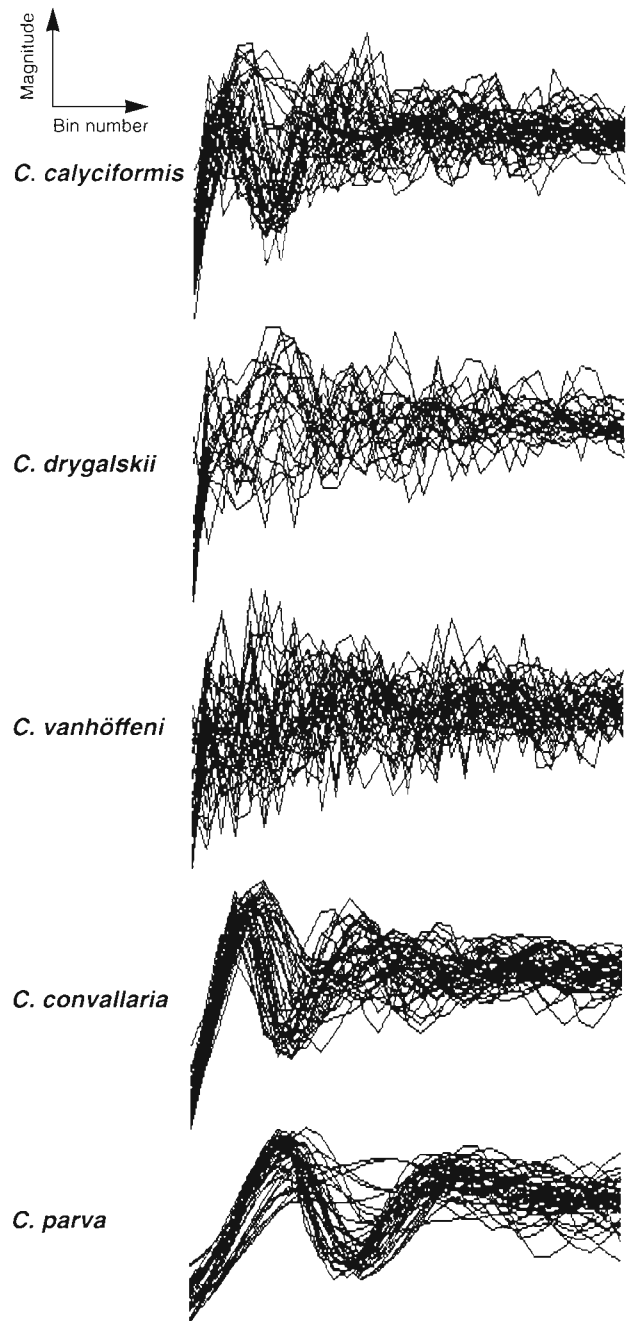

Fig. 2. Fast Fourier Transform 16 bin histograms for the entire data set (overlaid on each other)

cluster pools (image numbers 167, 70 and 126 for example). Since the experts did not adhere to the discriminant analysis provided in these figures it must be assumed that additional parameters are being employed by experts in the species recognition task.

The artificial neural network, on the other hand, used a completely different source of parametric information, the 2D-FFT data. Yet similar clusterings to the experts' have been obtained. Comparing network and
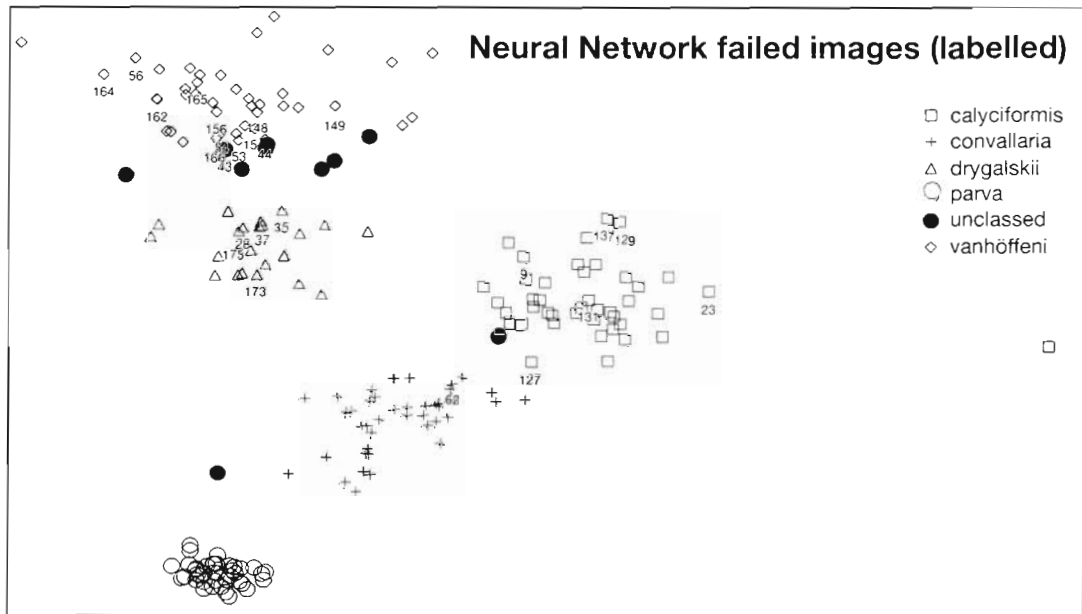
Fig. 3. Discriminant analysis plot overlaid with network 219 misclassifications (numbered data points; centre of numeral string is directly below each trial 219 misclassification) of *Cymatocylis* spp.
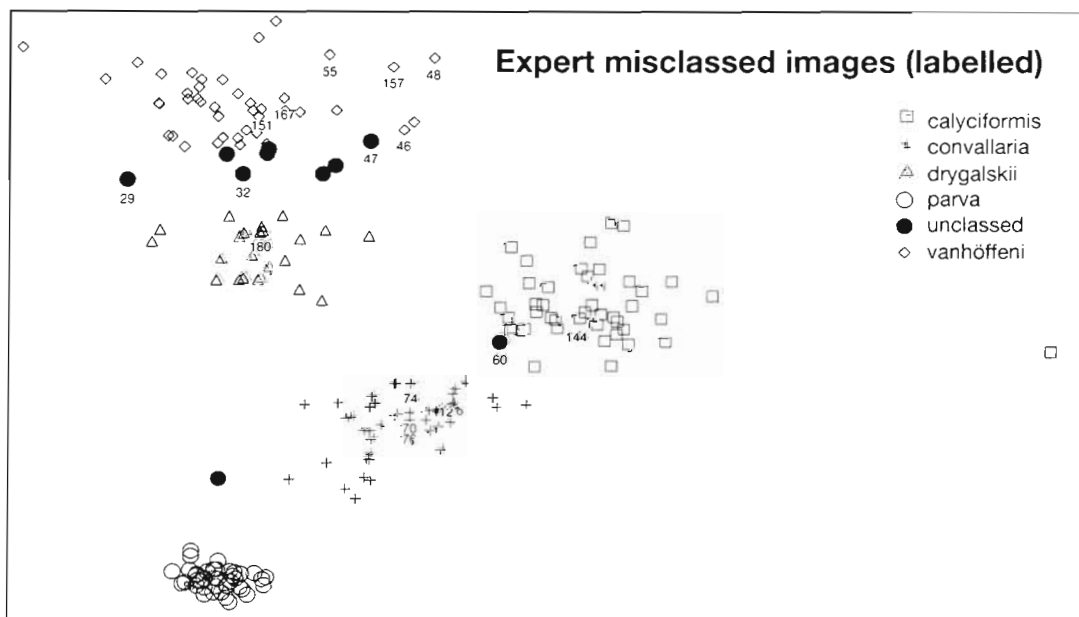

Fig. 4. Discriminant analysis plot with expert misclassifications (numbered data points; centre of numeral string is directly below each trial 219 misclassification) of *Cymatocylis* spp.

expert difficulty on a per species basis, both network 219 and network 276 failed to correctly categorise 34 of the 183 images that every expert concurred on. Of the 18 images that experts failed to agree on, 8 caused problems for the networks. A good agreement existed between the networks and the experts, in that the specimens most difficult to identify came from the same species on both counts. Table 6 summarises the

difficulties. Plotting these problem specimens on the same graph (Fig. 3) a similar set of ambiguities may be seen, with examples of inter-pool miscategorisations (43, 53, 127) and intra-pool miscategorisations (131, 62, 175). The networks demonstrate failure rates of less than 22 % for 4 of the 5 species, with *Cymatocylis van-höffeni* being the worst performer with a 49 % failure rate. Although no complete explanation can be given
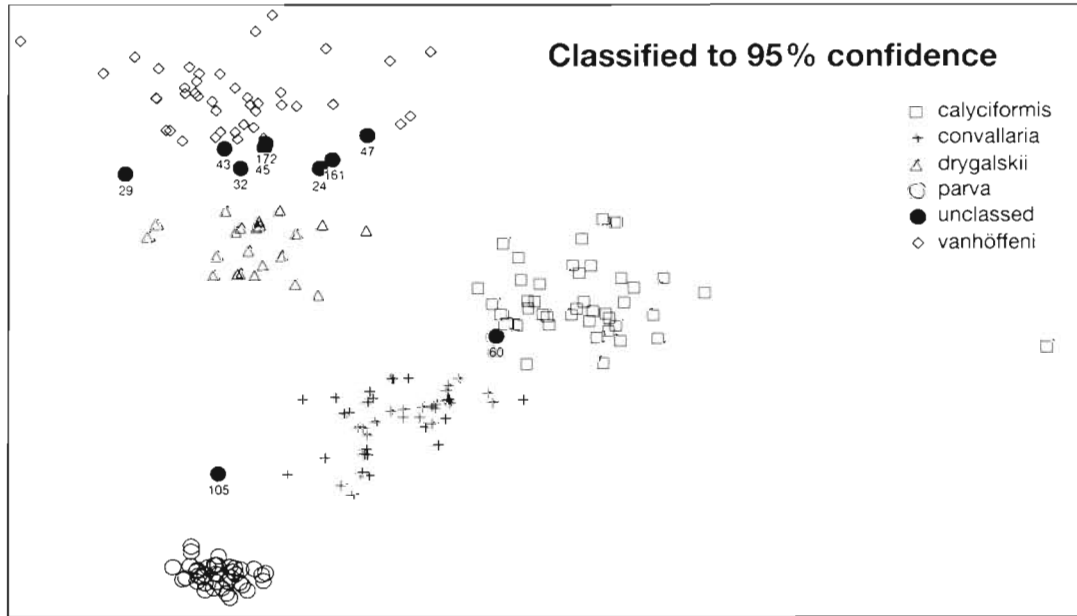
Fig. 5. Discriminant analysis plot, showing specimens of *Cymatocylis* spp. classified, and unclassified, at the 95% confidence level

for this poor performance, there is an agreement with the morphometric discriminant analysis (depicted in Fig. 5) and with expert opinion. Seven of the 10 images listed as 'mis-classed' by the discriminant analysis were classified correctly (as judged by the panel of experts) by both networks. Interestingly, all 3 tests (expert, network and morphology based) show that most confusions occur between *C. vanhöffeni* and *C. drygalskii*, shown in Figs. 3, 4 & 5 respectively. Two additional boundary problems occur with the statistical tests with specimen number 105 (between *C. parva* and *C. convallaria*) and specimen number 60 (between *C. calyciformis* and *C. convallaria*).

Given the small populations of specimens under scrutiny here, it is difficult to understand how effective a judge of species characteristics the BPN networks are. A review of the problems experienced by the expert panel in classifying the data set demonstrates a

good correlation with networks 219's behaviour (Pearson's product moment correlation 0.86, with 3 degrees of freedom).

In summary several proposals can be made: (1) that the low numbers of *Cymatocylis* images in the present trials were insufficient for good generalisations; an alternative view, (2) that the pre-processing techniques employed in this work failed to provide sufficient information specific to each species to enable a neural network to cluster the specimens correctly; and (3) that morphological variation between these species makes the categorisation task impossible.

The latter proposal may be quickly discounted, however, as it has already been shown that the categorisations are solvable using 2 morphological features common to all the *Cymatocylis* spp. in question. The 2 remaining proposals can be additive in their effect on neural network behaviour, but must be left for further work to disambiguate.

CONCLUDING REMARKS

From our previous work in attempting to apply this artificial neural network technique to taxonomic classification of planktonic images, it was predicted that over 100 images of each species type would be required for adequate generalisation of a neural network, to allow it to effectively discriminate between categories. Here we have taken 5 morphologically similar species and, deliberately using a smaller data set

Table 6. Comparison of performance. A: *Cymatocylis calyciformis*; B: *C. drygalskii*; C: *C. vanhöffeni*; D: *C. convallaria*; E: *C. parva*

	A	B	C	D	E
Training set size	40	24	38	43	40
Data set size	42	27	45	47	41
Network 202, % error	17	22	49	6	0
Network 219, % error	7	19	42	6	0
Experts, % error	5	11	16	11	2
Discriminant, % error	0	15	9	4	0

than suggested above, have attempted to automatically categorise photomicrographs of specimens. The attempt met with a qualified success.

We have demonstrated a neural network system that is capable of learning descriptions of 5 tintinnid species following training using 24 to 43 examples of these species. These results show that large data sets are not necessarily required for separation of species, although an obvious requirement of any data set is that it is a representative sample of the morphological variability of the species.

A direct comparison with Williams et al. (1994), who have applied multivariate statistical analysis to the Fourier transformed data set used in this neural network study, indicates that the data are clearly discriminable to 95 % confidence. This sets the goal for neural network performance to attain, since the multivariate analysis demonstrates the potential for discrimination using the Fourier processed data set. However, this must be set against the unequal task required of the neural network. Supervised learning of a back propagation network requires a pool of training data, and assessment of the success of training can only be measured using a previously 'unseen' (by the network) test data set, good performance at both tasks being indicative of the ease of fitting a model to the training data set and of the network's generalisation abilities from that model. The degree of success provides a measure of the applicability of the automatically generated model to the wider case of morphology of tintinnid species. Using the same terminology, the multivariate statistical analysis also constructs a 'model', but of the entire data pool; it makes no comment on the wider implications of clustering behaviour on a larger data set, given the construction of a model.

An additional burden on the neural network, evident from the morphological study of *Cymatocylis* by Williams et al. 1994), is the wide variance of morphology across the 201 specimens that constitute the data set. This variance inevitably makes the sub-sampling of the data set, necessary for the creation of training and test data sets, sensitive to sample size and results in pools of training and test data which can be unrepresentative of the whole population. If an unrepresentative data set is used to train a neural network then the ability of such a network to generalise to other morphotypes not seen in the training set will be impaired. Increasing the set size will improve the situation provided that the enlarged sample becomes more representative of the real population as more specimens are introduced to the set. This will have a direct bearing on the utility of cultured populations of species in the evaluation of new technologies and techniques for taxonomic classification, where morphological variation within cultures will be minimal.

This article was submitted to the editor

We have applied exactly the same pre-processing, neural network algorithms and experimental protocols (with the addition of using photomicrographs rather than hand drawn outlines from camera lucida images) as those used in our previous work. We are not aware of any robust discriminant techniques in existance that can be applied to morphologically diverse species, as this and our previous studies have done, and yet still obtain useful taxonomic classifications automatically.

This research extends our previous work in demonstrating the ability of artificial neural network classifiers to handle debris filled images, similar to those used by planktologists studying field samples under the microscope. To operate in true field conditions, a network must be capable of functioning correctly even when multiple specimens and highly cluttered images are presented. Although the work presented here is a step towards this goal there are still a number of profound problems to be solved.

LITERATURE CITED

Boddy, L., Morris, C. W., Wilkins, M. F., Tarran, G. A., Burkill, P. H. (in press). Neural network analysis of flow cytometric data for forty phytoplankton species common in the North Sea. Cytometry

Culverhouse, P. F. (in press). Automatic methods for the detection of harmful algae. In: Proc. 6th int. Conf. Toxic Marine Phytoplankton. Nantes, 18–22 Oct 1993

Györgyi, G. (1990). Inference of a rule by a neural network with thermal noise. Phys. Rev. Lett. 64: 2957–2960

Pierce, R. W., Turner, J. T. (1993). Global biogeography of marine tintinnids. Mar. Ecol. Prog. Ser. 94: 11–26

Simpson, R., Culverhouse, P. F., Williams, R., Ellis, R. (1991). Classification of Dinophyceae by artificial neural networks. IEEE Conf. on Neural Networks and Ocean Engineering, Washington DC, USA, August 1991. IEEE, Piscataway, NJ, p. 223–229

Simpson, R., Culverhouse, P. F., Williams, R., Ellis, R. (1993). Classification of Dinophyceae by artificial neural networks. In: Smayda, T J., Shimizu, Y (eds.) Toxic phytoplankton blooms in the sea. Elsevier Science, New York, p. 183–190

Simpson, R., Williams, R., Ellis, R., Culverhouse, P. F. (1992). Biological pattern recognition by neural networks. Mar Ecol. Prog. Ser 79: 303–308

Williams, R., McCall, H., Pierce, R. W., Turner, J. T (1994). Speciation of the tintinnid genus *Cymatocylis* by morphometric analysis of the loricae. Mar Ecol. Prog. Ser. 107: 263–272

Williams, R., Culverhouse, P. F., Simpson, R., Ellis, R., Lindley, J. A., McCall, H., Requera, B., Bravo, I., Parasini, T (in press). Identification of species of *Ceratium* and *Dinophysis* by artificial neural networks. In: Proc. 6th Int. Conf. Toxic Marine Phytoplankton. Nantes, 18–22 Oct 1993