ULTRASONIC TISSUE CHARACTERISATION USING NEURAL NETWORKS

Th E Schouten[1], M klein Gebbinck[1,2], J M Thijssen[2], J T M Verhoeven[2]

Katholieke Universiteit Nijmegen, Department of Informatics[1] and Department of Ophthalmology[2], Nijmegen, The Netherlands

ABSTRACT

Ultrasound Imaging is an important diagnostic tool in medical practice and research. It can be used to scan soft tissues, to characterise and to classify these according to possible diseases. In this paper diffuse liver diseases are studied, the available database consists of healthy livers and four kinds of diseases. From the echographic measurements five parameters are calculated for a tissue characterisation. To obtain a sufficiently large training set artificial data is generated using an optimal kernel estimate of the probability density function of the original data. Tissue characterisation is then performed using different kinds of neural networks: feed forward networks with error back propagation, self-organising feature maps and the ARTMAP network. The obtained results are given, compared and discussed. The results are also compared with a classification based on discriminant analysis.

1 INTRODUCTION

Several acoustic and texture parameters (1) can be extracted from echograms of a tissue on the basis of which characterisation of the tissue takes place. A classifier is then applied to classify the tissue into one of a set of predefined classes. Traditionally a statistical approach is chosen for the classifier, in recent years however, artificial neural networks have shown to be a promising alternative (2,3). More detailed information on the work described in this paper can be found in a master thesis (6).

In the next section the available data set is discussed which consists of healthy livers and four kinds of diffuse liver diseases. From the measured echograms five parameters are calculated to base the classification on. The data set is rather small and can hardly be split in a training and validation set. Therefore, a training set was generated using an optimal kernel estimate of the probability density function of the original data set. A quadratic discriminant function was determined on the basis of the training set and its performance on the validation set is presented

In the following sections the used neural networks and the methods to obtain their optimal architecture and parameter settings are discussed. To avoid over-learning the generalisation quality of the networks was measured during learning using their classification performance on a test set consisting of generated data and the original data. Finally conclusions about the used methods are drawn.

2 DATA SET AND DISCRIMINANT ANALYSIS

The data set consists of 4 diffuse liver disease classes to be discriminated against a normal population (NO, 129 cases). The disease classes are: women with primary biliary cirrhosis (PBC, 10 cases), acute hepatitis (AH, 13 cases), alcoholic hepatitis/cirrhosis (AHC, 10 cases) and other types of hepatitis/cirrhosis (HC, 27 cases). The disease was proven by a biopsy except for acute hepatitis which was assessed by clinical diagnosis.

Five parameters were selected to perform this study: the acoustic parameters β, α and S and the texture parameters μ and SNR. The slope of the attenuation coefficient determined with the multi-narrowband method provided α, forcing the slope to have zero intercept at zero frequency gave the value of β. S is the slope of the backscattering spectrum, after correction for the frequency dependent attenuation, between 1.2 and 3.2 MHz. The estimated mean grey level gives the value of μ, while SNR is that mean divided by the standard deviation of the grey level histogram. All parameters were corrected for age effects (4). For the analysis the parameters are scaled separately, projecting the mean minus 2 times the standard deviation to 0.1 and the mean plus 2 times the standard deviation to 0.9.

As the data set is small, it can hardly be divided into a training and a validation set. Therefore a training set of 500 cases per class was generated, using an optimal kernel estimate of the probability density function (pdf) of the original data set(5). As all parameters are found to be distributed according to a normal density function (6), a Bartlett kernel with smoothing parameter $h_n = 1.8218\ \sigma\ n^{-1/5}$ was used for the pdf estimation. The correlations between the parameters of the generated data are about equal to those of the original data (see tables 1a and 1b) and the mean and standard deviations differ by less than 1% respectively 5%.

TABLE 1a Correlations for class AHC of the original data

	$\rho_{\mu x}$	$\rho_{\beta x}$	$\rho_{SNR x}$	$\rho_{S x}$	$\rho_{\alpha x}$
μ	1.000	0.532	0.015	-0.466	0.468
β	0.532	1.000	0.200	-0.155	0.271
SNR	0.015	0.200	1.000	-0.622	-0.360
S	-0.466	-0.155	-0.622	1.000	0.153
α	0.468	0.271	-0.360	0.153	1.000

TABLE 1b Correlations for class AHC of the generated data

	$\rho\mu x$	$\rho\beta x$	$\rho SNRx$	ρSx	$\rho\alpha x$
μ	1.000	0.401	0.086	-0.371	0.336
β	0.401	1.000	0.168	-0.124	0.209
SNR	0.086	0.168	1.000	-0.526	-0.282
S	-0.371	-0.124	-0.526	1.000	0.192
α	0.336	0.209	-0.282	0.192	1.000

A test data set, consisting of 100 generated cases and the original cases per class, was used to determine the generalisation quality of neural networks during training.

Using the generated training data, discriminant analysis was performed using the individual within-group covariant matrices, resulting in a quadratic discriminant function (7). The classification on the validation set (the original data) resulted in a confusion matrix that is shown in table 2. There at row i and column j the percentage of cases belonging to class i and classified as class j is given.

TABLE 2 Confusion matrix (in %) for the validation data using discriminant analysis

	NO	PBC	HC	AH	AHC
NO	85.3	3.1	3.1	0.0	8.5
PBC	0.0	90.0	0.0	0.0	10.0
HC	29.6	14.8	33.3	7.4	14.8
AH	23.1	7.7	23.1	30.8	15.4
AHC	10.0	10.0	10.0	10.0	60.0

A Jack-knife method of discriminant analysis on the original data was performed. Each pattern was classified in turn using a discriminant function determined using all the other patterns. This resulted in a correct classification percentage for NO, PBC, HC, AH and AHC of 86.8%, 40.0%, 10.3%, 30.8% and 0.0% respectively. A clear improvement is thus seen using the generated data as a training set.

3 BACK-PROPAGATION NETWORK

A back-propagation (8) network implementation with learning rate and momentum as parameters was used. For each target class an output neuron was used and after each presented pattern the weights were updated (so called off-line learning). For each epoch (a presentation of all the training patterns) the order of presenting the patterns was selected randomly. As additional learning methods adding noise to the learning patterns and weight decay by adding a penalty term to the error function were implemented.

Many test runs were done to determine the best neural network architecture and parameter settings. After each training epoch the fraction of wrongly classified patterns per class was determined for the test set. The root mean square of these fractions (indicated by ICF) is an error function indicating the generalisation quality of the network. This function decreases with the number of learning epochs, at first rapidly and then slower, but then it levels off and often increases again, while the error function on which the network is trained still decreases. This is often called over-learning: the network specialises too much on correctly classifying the training patterns and looses generalisation capability. The network after the epoch giving the lowest ICF is taken as the best obtainable with the given architecture and parameter settings.

After some experimentation a learning rate of 0.01 and a momentum factor of 0.8 was taken to check several three layer (5-5-5, 5-10-5, 5-20-5 and 5-40-5) and four layer (5-10-5-5, 5-10-10-5, 5-20-10-5 and 5-20-20-5) architectures. The 5-40-5 network gave the lowest (0.530) ICF value and was selected for further testing. The learning rate was varied (0.0005, 0.01 and 0.2) with a value of 0.01 giving the best result. The momentum factor was then varied (0.0, 0.40, 0.80 and 0.95) with 0.80 giving the best result. Adding random noise (2%, 5% and 10%) to the training patterns and using weight decay (factors 0.001 and 0.01) did not decrease the ICF.

TABLE 3 Confusion matrix (in %) for the validation data using back-propagation

	NO	PBC	HC	AH	AHC
NO	81.4	2.3	3.1	5.4	7.8
PBC	0.0	100.0	0.0	0.0	0.0
HC	14.8	0.0	85.2	0.0	0.0
AH	7.7	0.0	0.0	92.3	0.0
AHC	0.0	0.0	0.0	0.0	100.0

With the so determined best network the validation set was classified and the obtained confusion matrix is shown in table 3.

4 KOHONEN NETWORK

A Kohonen map (9) type of network was implemented with learning rate and neighbourhood radius parameters decreasing linearly from their initial values to zero. The initial weights were randomly set and were then updated after each presented pattern. After some experimentation the number of training epochs was fixed at 100 and the initial neighbourhood radius was set to half the width of the map.

With a learning rate of 0.005 square maps ranging from 3x3 to 10x10 were trained. The 7x7 map provided the lowest ICF of 0.735 that occurred after 62 epochs. Several initial learning rates were then tried and a value of 0.005 gave the best result. The conscience learning variant as described in (8) was then tried improving the ICF a little to 0.733. With the so determined best network the validation set was classified and the obtained confusion matrix is shown in table 4.

TABLE 4 Confusion matrix (in %) for the validation data using kohonen

	NO	PBC	HC	AH	AHC
NO	67.4	9.3	7.8	3.9	11.6
PBC	0.0	80.0	0.0	20.0	0.0
HC	22.2	14.8	55.5	7.4	0.0
AH	15.4	15.4	7.7	61.5	0.0
AHC	0.0	10.0	10.0	0.0	80.0

5 ARTMAP

ARTMAP (11) is an interesting network to consider because it has a fast training method and usually gives a good performance. It also has a low number of parameters, so a good network is easier to find. The used implementation, see (10) for more information, uses FuzzyArt/ART2a (12,13) as its basics' networks. The best network was obtained with the base-vigilance parameter set to 0.0, the alpha parameter set to 0.1 and the beta parameter set to 0.85. With the so determined best network the validation set was classified and the obtained confusion matrix is shown in table 5.

TABLE 5 Confusion matrix (in %) for the validation data using ARTMAP

	NO	PBC	HC	AH	AHC
NO	78.3	1.6	6.2	8.5	5.4
PBC	0.0	100.0	0.0	0.0	0.0
HC	22.2	3.7	66.7	7.4	0.0
AH	7.7	7.7	0.0	69.2	15.4
AHC	0.0	0.0	0.0	0.0	100.0

6 CONCLUSION

In table 6 the obtained ICF's (the root mean square of the fractions of wrongly classified patterns per class) on the validation data are given for the discriminant analysis and the neural networks trained on the generated training data set. Also the ICF for the Jack-knife analysis of the validation data is given. The indicated times are only to be taken as an indication of the training time of the final selected networks on a workstation (SUN Sparcstation1) as no controlled timing measurements have been done.

TABLE 6 Obtained ICF for validation data and needed training time

	JK	Disc	BP	Koho	ART
ICF	0.73	0.47	0.11	0.32	0.22
time		1 m	120 h	1 h	10 m

Compared to discriminant analysis the neural networks clearly improve the classification of the patients' data at the expense of a large human and computational effort. Note however, that the time needed for classification of a pattern using a trained neural network is small for all the studied neural networks.

From the performance improvement of the discriminant analysis trained on the generated data as compared to the Jack-knife analysis and from the way of generating the training set (optimal estimate of the pdf and independent selection of the parameters), it can be hoped that the obtained good classification also extends to data obtained from new patients.

The introduced ICF determination during learning provides a good way of detecting over-learning of the neural networks and finding the best network for the classification task.

REFERENCES

(1)Thijssen, J.M., Oosterveld, M.J., Hartman, P.C. and Rosenbush, G.J.E., 1993,Ultrasound Med. Biol.,19, in press
(2)Whittaker, A.D., Park, B.S., McCauley, J.D. and Huang, Y. , 1991, Proceedings of the 1991 Symposium on Automated Agriculture for the 21ste Century, pp 116-125
(3)Ostrem, J.S., Valdes, A.D. and Edmonds, P.D., 1991, Ultrasonic Imaging 13, 298-299
(4)Hartman, P.C., Oosterveld, B.J., Thijssen, J.M. and Rosenbusch, G.E.J.,1991 Ultrasound Med. Biol. 17, 857-867
(5)Holström, L. and Koistinen, P., 1992, IEEE Trans. Neural Networks 3, 24-38.
(6)Gebbinck, M. klein, 1992, Ultrasonic Tissue Characterization Using Neural Networks, M.sc. Thesis, University of Nijmegen
(7)SAS Institute Inc., 1989, SAS/STAT User's Guide Version 6, Volume 1
(8)Rumelhart, D.E., Hinton,G., Williams, R., 1986, Nature 323, 533-536
(9)Kohonen, T., 1984, Self-organization and associative memory (Springer-Verlag)
(10)Willems, P., 1993, Parallel Implementation and Evaluation of the ART Neural Networks, M.sc. Thesis, University of Nijmegen.
(11)Carpenter, G.A., Grossberg, S., Reynolds, J.H., 1991, Neural Networks 4 , 565-588
(12)Carpenter, G.A., Grossberg, S., Rosen. D.B., 1991, Neural Networks 4 , 759-771
(13)Carpenter, G.A., Grossberg, S., Rosen. D.B., 1991, Neural Networks 4 , 493-504