

# Classification of different cancerous animal tissues on the basis of their $^1\text{H}$ NMR spectra data using different types of artificial neural networks

A.R. Mehridehnavi\*

*Department of Medical Physics and Medical Engineering, School of Medicine, Isfahan University of Medical Sciences, Isfahan, I.R.Iran.*

---

## Abstract

The main objective in classification of the NMR spectra of cancerous and healthy tissue, with high number of features is the prerequisites of the minimum number of samples. Therefore the use of conventional classifier on this type of the data is not recommended. In the current work, different structures of the artificial neural networks (ANN) were tried on classification of different cancerous and healthy tissues. The use of nonlinear classifiers such as ANN could be a proper alternative in NMR spectra classification. The data consists of five type of cancerous and three types of NMR spectra of healthy animal tissues. In the case of multi layer perceptron (MLP), two-layer network with 11 hidden nodes gave the best solution on the current data. In addition MLP + single layer perceptron (SLP) were tried on the current data but it did not make any improvement. Finally the SLP network with log likelihood cost function was tried and in addition to giving the best classification it had a fast convergence time and gave a unique solution on the data independent of the initial seed value off the training. The SLP classifier is better than any other classifiers for the current data.

**Keywords:** NMR spectra; artificial neural networks; Cancer; Single layer perceptron

---

## INTRODUCTION

In cancer, the search for tumor markers started over 30 years ago (1). Clinically useful tumor markers, now exist for a few malignancies, and used for screening or post-treatment management of the disease. To be useful, any such tests (including classification methods) must be highly reliable, sensitive and specific, and must provide information in complement with currently accepted clinical procedures.

The NMR spectra of biological materials are very rich in information which is carried in the overall pattern of metabolite and tissue resonance. However, the information contained in the  $^1\text{H}$  NMR of tissue and bio-fluid metabolite patterns is often extremely complex. Subtle

biochemical alterations may be lost even in extensive conventional quantitative and statistical analysis of the spectral data (2).

A common objective of almost all conventional classifiers, for discrimination between two or more groups, is to have a non-singular covariance matrix. In problems such as the classification of NMR spectra where each individual has many variables compared to the number of patterns, there should be a pre-processing stage of variable reduction, using, for instance, principal component analysis or partial least squares (3). The applicability of conventional classifiers methods on NMR spectra requires that the number of independent objects in the training set (i.e. spectral samples) to be larger by at least a factor of three more than the number of

---

\*Corresponding author: Dr. A.R. Mehridehnavi  
Tel. 0098 311 7922485, Fax. 0098 311 6688597  
Email: mehri@med.mui.ac.ir

variables. Unfortunately, current data could not fulfil this requirement. In artificial neural networks (ANN) classifiers the non-singularity of the covariance matrix is not important and consequently there is no need for prior transformation of the data (4).

There are different methods of classification which are used in biomedical and biochemistry classifications i.e. linear discriminants, k-nearest neighbour, Bayesian methods, ANN classifiers and cluster analysis.

As magnetic resonance (MR) Imaging is growing, the use of NMR spectroscopy in diagnostic of cancerous tissue becomes an important issue. There is a major interest to distinguish between different types of cancer on the basis of FID (Free Induction Decay) spectra (5).

In this study, it was tried to compare different classification methods on NMR spectra data of cancerous tissues. This work is in line with the work of Howells et al. (6), Branston et al. (7), and Lisboa et al (8,9) on the same data with 180 frequency components. Howells et al. (6) used principal component analysis to reduce the dimension of the spectral data. The authors showed that the best classification could be achieved, using the first 13 principal component scores following by cluster analysis method. Their success was about 64%. Branston et al. (7) improved previous work (6) by using a multi-layer perceptron. Again the same first 13 principal components were used to represent the data. The multi-layer perceptron showed a better performance than clustering using the dendritic tree. Their success was about 77%.

The objective of this study was to review the classification of NMR spectra of tumor extracts with different classifiers. ANNs are finding applications in different fields of science and engineering for classification and pattern recognition, and in medicine for medical diagnostics, where

they complement human expertise (10,11). They are also used in clustering of proteins and DNA sequences, in the prediction of properties of molecules such as reactivity, in the deduction of the composition of chemical mixtures, and in the classification of NMR spectra (12).

## MATERIALS AND METHODS

### *Tumor processing and data extraction*

Five of the tumors were from established cell line grown subcutaneously in the flanks of a variety of rat strains. Morris hepatoma lines H7777 (poorly differentiated, fast growing) and H9618A (well differentiated and slower growing) were grown in male or female buffalo rats; a pituitary tumor (GH3 prolactinoma) in Wistar-Furth rats; Walker 256 carcinosarcoma in female wistar rats; LBDS<sub>1</sub> fibrosarcoma in male BD9 rats. Chemically-induced primary mammary tumors were developed in female Ludwin-Olac rats following injection of methyl-nitrosourea. Normal tissues (liver, spleen and kidney) were obtained from different strains of healthy rats. Tissue and tumors were frozen in liquid nitrogen, extracted using 6% perchloric acid, neutralised with potassium hydroxide and then freeze-dried. This procedure results in minimal breakdown of low molecular weight metabolites. Macromolecules such as proteins and DNA and lipophilic components will be precipitated out during the extraction and will not contribute to the NMR signals (6).

<sup>1</sup>H NMR spectra were obtained at 25 °C on a Bruker AM-400 spectrometer after dissolution of 70 mg of freeze dried material in 0.5 ml D<sub>2</sub>O and readjustment to pH 6.80 ± 0.05. The signal at zero ppm was from a sodium 3'-(trimethylsilyl)-1-propane sulphonate (TPS or TSP) which was added for chemical shift and quantification referencing. Data acquisition involved a pulse repetition time of 10 s, a

spectral width of 8 kHz and collection of 16 k data points (6).

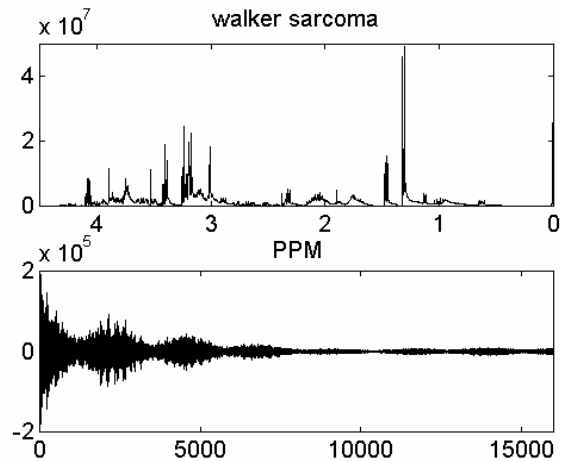
NMR signals were acquired in the time domain and converted to frequency domain using Fourier Transform. Note that the frequency axis of the NMR spectrum is also known as the chemical shift axis since it is the chemical environment of each  $^1\text{H}$  nucleus which influences the resonance frequency of its NMR signal. By convention, this chemical shift axis is scaled in ppm to account for the operating frequency of the spectrometer.

The peak from dissolved deuterated water at 4.7 ppm had been removed selectively from the spectrum. The baseline distortion is minimised by applying a linear baseline correction in the range of 4.5 to 0.0 ppm, over which signal heights were measured. It should be noted that the concentration of the metabolites in the samples are proportional to the area under each NMR signal. However, the presence of large numbers of overlapping signals makes it difficult to measure the areas under many of the NMR peaks accurately and so the data is reduced by measuring the maximum peak height within each narrow (0.025 ppm) window in the spectrum. 180 peak heights were extracted and normalised with respect to the TSP reference peak and therefore reflect the pattern of metabolite concentration in each sample (6).

#### **Data organization**

The data included NMR spectra of five types of cancer tissue and three NMR spectra of the healthy animal tissue. That is, 12 fibrosarcoma, 5 pituitary tumors, 8 Walker sarcoma, 12 hepatoma H7777, 11 hepatoma H9618A, 12 healthy spleen tissues, 11 healthy kidney tissues, and 12 healthy liver tissues, totally accounts for 83. A typical spectrum example of MR and Spin Echo signal of Walker sarcoma are given in Fig. 1. The data were divided into five independent subsets. Five cross-

validation subsets of the data were used for the training and the test. In each trail, four subsets were used as training data and the left set as test data.



**Fig. 1.** Magnetic Resonance and Spin Echo signal of Walker sarcoma tissue and its spectra.

#### **Data classification**

The NMR data samples were classified with the use of the ANN and a few types of the conventional classifiers. In this work, an ANN programme written in C language was developed and used. For statistical trails Matlab statistical tool box was used. ANNs: In this work three type of the ANN including MLP (multi layer perceptron), MLP+SLP, and SLP, (single layer perceptron), were tried.

The procedure outlined in 1 to 5 was repeated with different numbers of hidden nodes and for all five cross-validation permutations. In this work, training was stopped at the point of lowest generalization error for a given training/test data combination. The exception to this was when the two errors reduced asymptotically together, indicating that the particular test set was probably always interpolating the training data. In this case, training was stopped when the training and test error flattened out.

Output nodes required the selection of

an appropriate number of hidden nodes. Tests were carried out with a range of hidden units, namely 4, 6, 8, 10, 11, 13 and 15 to find the optimum number for the best classification. The networks were trained by back-propagation using least mean square error minimisation. The test squares of the classification error (network response - target response), were summed over all the test samples in each group covering the complete data set, i.e. all 83 patterns, averaged and square-rooted, to produce the RMS (root mean square) figures. A pattern was considered correctly classified if the output unit with a target value of 1.0 gave 0.6 or higher and output units with a target value of 0.0 gave less than or equal to 0.4.

## RESULTS

Different ANN with different number of hidden nodes tried on current data results are given in Table 1. In the next stage with respect to results and over generalization problem 11 nodes for hidden was chosen.

Combined network (MLP with direct connections): An alternative neural network structure is a MLP network with additional direct input to output connections with direct weights. The direct input/output connections comprise a SLP which operates in parallel with a MLP.

The network was simulated with two combinations of cost function, the first using the least mean square (LMS) as a cost function in both parts of the network, and in the second trial the LMS error was used for error propagation in the MLP part and LL (log likelihood) error cost function was used for error propagation in direct part (SLP) of the network.

The network was simulated with direct connection and a MLP with 11 hidden nodes as chosen in the previous section. The LMS cost function was used in both parts of the network. The number of correctly classified samples in the test sets

listed in row 1 of Table 2.

Again the previous network structure was used but, the LL error was used as the cost function in the direct part of the network. This network converged very fast in comparison with the previous one which used the LMS cost function in the direct part. The number of correctly classified samples in the test sets listed in row 2 of Table 2.

Single layer network using the LL cost function: A single layer network using the LL error cost function and the generalised delta rule simulated. The network was trained and tested with all five data groups. Classification test results on the five cross validation tests are listed in row 3 of Table 2. Nearest class mean classifier: In this part the mean value of all eight classes of training sets in each data group was computed separately. The euclidean distances of each pattern including test and training patterns were calculated with respect to all eight class mean values. A pattern was assigned to the class from which it had the smallest distance (6.9). The results of train and test for this method are listed in rows 4 and 5 of Table 2.

The first row shows the number of correct classifications over the number of training patterns in the group and the second row shows the same for test subsets.

The k-Nearest Neighbour (k-NN) was used for the classification of the current data. There were different suggestions for the choice of  $k$ , such as  $k_n$  should be proportional to the square root of  $n$ , the number of samples, or it should be a function of the data dimensionality  $m$ . Unfortunately this was not the case in this application, the number of training samples were few compared with the number of features in data space, therefore the  $k$  could not be certain. Different values for  $k$  were tried from 1 to 6 for all five groups of data. The best classification occurred with small  $k$  such as  $k=1, 2$ , and 3.

**Table 1.** Classifiers results summary

Classifier	Training error	Training STD error	Test error	Test STD error	Correct train (%)	Correct test (%)
MLP (180-4-8)	0.0252	0.0023	0.0672	0.0317	100	94
MLP (180-6-8)	0.0196	0.0007	0.0632	0.0339	100	94
MLP (180-8-8)	0.0181	0.0007	0.0613	0.0285	100	94
MLP (180-10-8)	0.0169	0.0007	0.0612	0.0273	100	94
MLP (180-11-8)	0.0160	0.0006	0.0561	0.0253	100	96.4
MLP (180-13-8)	0.0157	0.0006	0.0628	0.0226	100	96.4
MLP (180-15-8)	0.0150	0.0006	0.0620	0.0283	100	96.4
MLP+SLP (LMS & LMS)	0.0249	0.0030	0.0778	0.0302	100	90.4
MLP+SLP (LMS & LL)	0.0050	0.0011	0.0560	0.0299	100	95.2
SLP (LL)	0.0051	0.0022	0.0583	0.0226	100	96.4
Minimum distance	-	-	-	-	78.3	66.3
1-NN	-	-	-	-	100	85.5
2-NN	-	-	-	-	93	83.1
3-NN	-	-	-	-	91	79.5

**Table 2.** Number of correctly classified in ANN's and nearest class mean classifier.

Group	G.1	G.2	G.3	G.4	G.5
MLP (180-11-8) + SLP (LMS+LMS)	14/15	15/15	13/15	17/19	16/19
MLP (180-11-8) + SLP (LMS+LL)	14/15	15/15	13/15	19/19	18/19
SLP (LL)	14/15	15/15	15/15	18/19	18/19
Minimum Distance (train)	54/68	52/68	50/68	50/64	54/64
Minimum Distance (test)	9/15	9/15	11/15	13/19	13/19

## DISCUSSION

The best number of hidden nodes was 11, which could classify current data with the smallest average RMS error and smallest standard deviation of the test error. Also it gave the maximum number of correct classifications as shown in Table 1. For further stages the network with 11 hidden nodes was chosen.

The simulation results of combined network, MLP and direct connection using LMS in both part did not show any improvement. The number of correctly classified samples in the test set was worse. The convergence behaviour of this structure was similar to the MLP.

In the combined network with LL error in the direct part the training and test RMS errors were smaller and converged faster than before. The training and test RMS

errors of all groups had similar convergences as before. The network with the LL error cost function at direct part behaved differently from the network with the LMS error cost function at direct part. The network with the LL error cost function was more stable than the network with the LMS error cost function and behaved independent from the random seed.

In the single layer network the training error went down to very small values in comparison with the network with hidden layers. In this type of network we can arrive at the overtraining point very soon. The test error declined to the smallest possible value, following the training error, and by further training the test error was increased while the training error was declining.

From the classification results in Table 1

and works result of Howells et al. (6) and Branston et al (7), the method which could classify more efficiently than the others, is the ANN. It can be seen that for linear data it is better to use a single layer network with the log likelihood cost function as it converges faster than the other types of network such as the MLP using least mean squares. Since data are not linearly separable it is better to use the MLP with the least mean square error. It seems that the use of MLP network for classification of the data with large dimensions in comparison with the data samples is not useful and is time consuming. It is recommended to use single layer perceptron ANN for classification of these types of the data.

### ACKNOWLEDGMENT

I should thank Maxwell RJ and his colleague letting us to use their data for this study.

### REFERENCES

1. Chmurny GN, Hilton BD, Halverson D, McGregor GN, Klose J, Issaq HJ, et al. An NMR blood test for cancer: a critical assessment. *NMR Biomed.* 1988;1:136-50.
2. Gartland KP, Sanins SM, Nicholson JK, Sweatman BC, Beddell CR, Lindon JC. Pattern recognition analysis of high resolution  $^1\text{H}$  NMR spectra of urine. A nonlinear mapping approach to the classification of toxicological data. *NMR Biomed.* 1990;3:166-172.
3. Krzanowski WJ, Jonathan P, McCarthy WV, Thomas MR. Discriminant analysis with singular covariance matrices: Methods and applications to spectroscopic data. *N Appl Statist.* 1995;44:101-115.
4. Kowalski BR, Wold S. Pattern recognition in chemistry. in *Handbook of Statistics Vol.3* By P.R.Krishnaiah and L.N. Kanal. North Holland, Amsterdam, 1982. p. 673-697.
5. Shah N, Sattar A, Benanti M, Hollander S, Cheuck L. Spectroscopy as an imaging tool for cancer. *J Am Osteopathic Assoc.* 2006;106:23-27.
6. Howells SL, Maxwell RJ, Griffiths JR. Classification of tumor  $^1\text{H}$  NMR spectra by pattern recognition. *NMR Biomed.* 1992;5:59-64.
7. Branston NM, Maxwell RJ, Howells SL. Generalization performance using back-propagation algorithms applied to the patterns derived from tumor  $^1\text{H}$  NMR spectra. *J Microcomputer Appl.* 1993;16:113-123.
8. Lisboa PJG, Kirby SPI, Vellido A, Lee YYB, El-Deredy W. Assessment of statistical and neural networks methods in NMR spectral classification and metabolite selection. *NMR Biomed.* 1998;11:225-234.
9. Lisboa PJG, Branston NM, El-Deredy W, Vellido A. Tissue characterization with NMR spectroscopy. Current state and future prospects for the application of neural networks analysis. *Proceedings of the IEEE/INNS International Joint Conference on Neural Networks;* 1997 1385-1390; Houston, USA.
10. Ringner M, Saal LH, Ladanyi M, Westermann F, Berthold F, Schwab M. et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med.* 2001;7:673-679.
11. Ziaei L, Mehridehnavi AR, Salehi M. Application of artificial neural networks in cancer classification and diagnosis prediction of a subtype of lymphoma based on gene expression profile. *J Res Med Sci.* 2006;11:13-17.
12. Burns JA, Whitesides GM. *Chemical Reviews: Feed forward neural networks in chemistry: Mathematical systems for classification and pattern recognition.* *Am Chem Soc.* 1993;93:2583-2601.