Neural network segmentation of magnetic resonance spin echo images of the brain

S. Cagnoni*, G. Coppini†, M. Rucci‡, D. Caramella** and G. Valli*

*Department of Electronic Engineering, University of Florence; †CNR Institute of Clinical Physiology, Pisa; ‡Scuola Superiore S. Anna, Pisa; **Department of Radiology, University of Pisa, Italy

Received September 1992, accepted December 1992

ABSTRACT

This paper describes a neural network system to segment magnetic resonance (MR) spin echo images of the brain. Our approach relies on the analysis of MR signal decay and on anatomical knowledge; the system processes two early echoes of a standard multislice sequence. Three main subsystems can be distinguished. The first implements a model of MR signal decay; it synthesizes a four-echo multiecho sequence, in order to add images characterized by long echo-times to the input sequence. The second subsystem exploits a priori *anatomical knowledge by producing an image, in which pixels belonging to brain parenchyma are highlighted. Such anatomical information allows the following submodule to distinguish biologically different tissues with similar water content, and hence similar appearance, which might produce misclassifications. The grey levels of the reconstructed sequence and the output of the second module are processed by the third subsystem, which performs the segmentation of the sequence. Each pixel is assigned to one of five different tissue classes that can be revealed with brain MR spin echo imaging. With a suitable encoding, a five-level segmented image can then be produced. The system is based on feed-forward networks trained with the back-propagation algorithm; experiments to assess its performance have been carried out on both simulated and clinical images.*

Keywords: Magnetic resonance imaging, image segmentation, artificial neural networks, spin echo sequences, MR image synthesis

INTRODUCTION

The automatic segmentation of anatomical structures is one of the basic processes in most applications of computer vision to medical imagery[1]; segmentation has immediate applications, such as the estimation of dimensional and morphological parameters of biological structures. Moreover, the extraction of the sections of a given organ from a series of tomograms allows 3D information to be explicitly recovered[2].

We have developed a neural-network based system which embodies different kinds of knowledge required to segment magnetic resonance (MR) images of the brain. Independently of its being performed by a human observer or by an artificial system, the identification of different organs (or, more generally, of regions of interest) requires that at least two kinds of information be processed. The first derives from the physical phenomena involved in the image-generation process which produces a grey-level representation of the imaged object. The second derives from specific anatomical knowledge about the district that is being analysed.

As regards the generation of MR images, an MR signal depends on both the biochemical characteristics of the tissues and the acquisition parameters. In particular, a strong dependence on the T_2 relaxation time (which conspicuously differentiates liquid from parenchymal tissues) can be achieved with spin echo sequences. Therefore, by analysing such sequences, it is possible to evaluate T_2 and classify the imaged tissues accordingly.

The most straightforward procedure for estimating T_2 is least-square fitting performed on raw signals obtained at different echo times (TE). Such a method has been used both to distinguish healthy from pathological tissues and to generate synthetic spin echo sequences with any desired TE[3,4]. However, the method requires that: (i) the fit be performed on a large number of acquisitions with different TEs; (ii) such acquisitions be performed in a sufficiently long time-span[5]. Usually, neither of these conditions is satisfied in clinical settings, where acquisition time must be kept as short as possible. Furthermore, most hospitals do not usually archive raw signals, but images that have been preprocessed by expanding their dynamic range, thus altering the exponential decay of the raw signal. Hence, although sequences acquired with similar parameters from the same anatomical district are similarly processed, a correct estimation of signal decay is not easily obtained with conventional statistical methods.

Therefore, the development of more flexible

Correspondence and reprint requests to: Dr Stefano Cagnoni, Department of Electronic Engineering, University of Florence, Via di S. Marta, 3, I50139 Firenze, Italy

0141–5425/93/05355–08

models of the MR signal is needed and artificial neural networks (ANNs) could be suitable for this purpose. By means of a proper training procedure, ANNs are able to emulate a wide class of multivariate functions. In addition, good noise-rejection properties are typical of ANN-based systems. This enables reliable extraction of underlying prototypes from a set of examples[6,7,8].

As regards the use of *a priori* anatomical knowledge, the following should be considered. Although the decay of MR spin echo signals carries relevant information about tissue properties, in general its analysis does not ensure safe recognition of biological structures. As an example, the acquisition parameters are usually set in order to enhance the contrast between well-hydrated and parenchymal tissues. Thus, grey levels corresponding to tissues with comparable water content look similar, even if the tissues are biologically different. This makes tissue discrimination and classification difficult.

In our first attempt to classify brain tissues using an ANN model, subcutaneous fat was often misclassified as white matter. A similar result was reported by Piraino and co-workers[9]. Such a problem can be overcome if a primary segmentation that extracts the brain parenchyma is performed. Indeed, the three main brain tissues (white matter, grey matter and cerebrospinal fluid) are well differentiated and can only be found in the cerebral region. However, representation of the required anatomical knowledge about the shape and position of the imaged organs is hampered by the typical complexity and variability of the biological structures. The capability of ANN systems to represent and use prototypical knowledge implicitly contained in the training set has allowed us to solve this problem.

In the following sections, we describe the structure of our system and its implementation, and report the experimental results obtained on both simulated and clinical images.

SYSTEM ARCHITECTURE

As schematized in *Figure 1*, we outlined the following main computational processes: (i) analysis of the information carried by signal decay; (ii) identification of the brain parenchyma; (iii) pixel classification into predefined tissue groups. Consequently, three main subsystems (modules) were designed and implemented.

The first module, which we have called sequence reconstructor (SR), derives a four-image multi-echo sequence from a standard two-echo multislice sequence by synthesizing two long-TE images. In this way it enhances differences in MR signal decay among the various tissue components, besides improving the signal-to-noise ratio of the sequence.

The second module (brain detector [BD]) recognizes the brain parenchyma and generates a primary segmentation in which the brain region is outlined. *A priori* knowledge of the shape of the brain in the slices considered is stored in the weights of a properly trained neural network.

The third module, which we have called classification module (CM), processes the outputs of the other modules and classifies the pixels as belonging to five different classes, representing the main tissues constituting the head.

In all modules we have utilized multilayer feed-forward networks, made up of processing elements with sigmoidal activation functions. Network training was performed using the back-propagation algorithm. This neural paradigm has been widely applied to different fields, thanks to its generally non-critical behaviour. In fact, the absence of feedback connec-

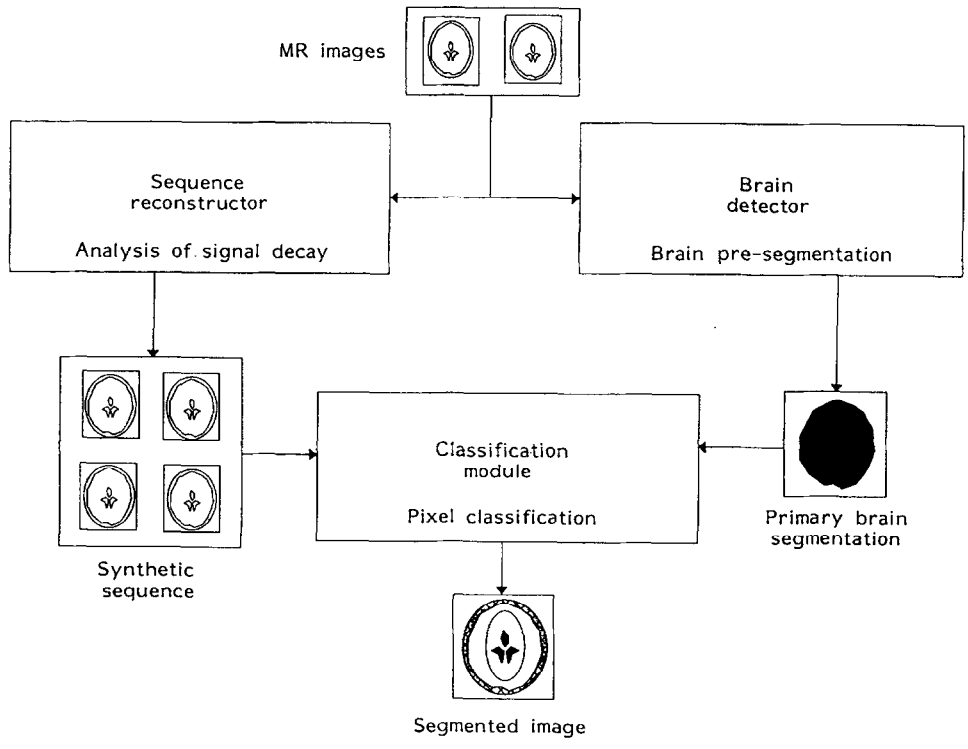

Figure 1 Structure of the neural network system

tions ensures network stability. In addition, only moderate computing resources are needed. On the other hand, its main drawbacks are usually slow convergence and the chance of getting trapped into local minima of the error function[6,7,8].

The sequence reconstructor: a neural-network model of MR spin echo signal

MR spin echo signals are characterized by an exponentially decaying behaviour, depending on both the acquisition parameters' echo time TE and repetition time TR, which can be set by the operator, and on the molecular characteristics of the sample (proton density $N(H)$ relaxation times T_1 and T_2).

For spin echo sequences that are commonly used in clinical onsets, the dependence of pixel intensity I on tissue properties and sequence parameters is described by the following equation:

$$I \cong c\,N(H)(1 - e^{-(TR - TE/2T_1)} + e^{-(TR/T_1)})\, e^{-(TE/T_2)} \quad (1)$$

where c is a constant.

When typical parameters are considered ($TE <$ 200 ms, $TR > 1000$ ms), (1) can be approximated as follows:

$$I \cong k(N(H), T_1, TR)\, e^{-(TE/T_2)} \quad (1b)$$

which shows that such images are strongly T_2-dependent. As a consequence, the intensity of pixels belonging to parenchymal tissues (short T_2) is higher in images acquired with short TEs, while well-hydrated tissues and liquids (long T_2) are brighter in long-TE images[10].

The least-squares methods to estimate T_2 and $N(H)$ are based on the linearization of the exponential model of equation (1b). This leads to linear equations in which the logarithm of proton density $N(H)$ and the reciprocal of relaxation time T_2 are the unknown quantities[4]. The coefficients of such equations are functions of TE and the corresponding values of I. The major drawback of such a least-squares estimate of T_2 is a high sensitivity to noise, whose effects can be limited with the acquisition of a large number of images (with different TEs) during the longest possible time-span. In this situation, the fit proves useful and accurate when T_2 or proton density maps are to be reconstructed. However, when used to synthesize images with intermediate TEs, it adds little information to the original sequence. If the input images have been preprocessed, the aforementioned equations are, in general, no longer adequate to model the signal, and another approach is needed.

The neural-network module that we have developed to perform such a task (sequence reconstructor) consists of a four-layer network with $18+6+6+3$ processing elements (see *Figure 2*). The 18 inputs are the grey levels of two corresponding 3×3 windows taken from two short-TE images of the same slice. The three ouputs represent the estimated intensity values of the central pixel of such windows in three images with the same TR and prefixed TEs. Therefore, by successively processing pairs of 3×3 windows centred on each pixel of the input images, the network synthesizes long-TE images, thus building a longer multiecho sequence.

The training set was derived from multiecho sequences, where $TE = 50$, 100, 150, 200 ms and $TR = 2000$ ms. The input patterns are the pixel intensities of the first two images. The teaching inputs are the intensity values of images acquired with $TE =$ 100, 150, 200 ms and $TR = 2000$ ms.

The training set includes 4500 examples, randomly selected from a set of three spin echo multiecho sequences. Pixels belong to square regions of interest that include only significant parts of the images. Background pixels were not included in the training

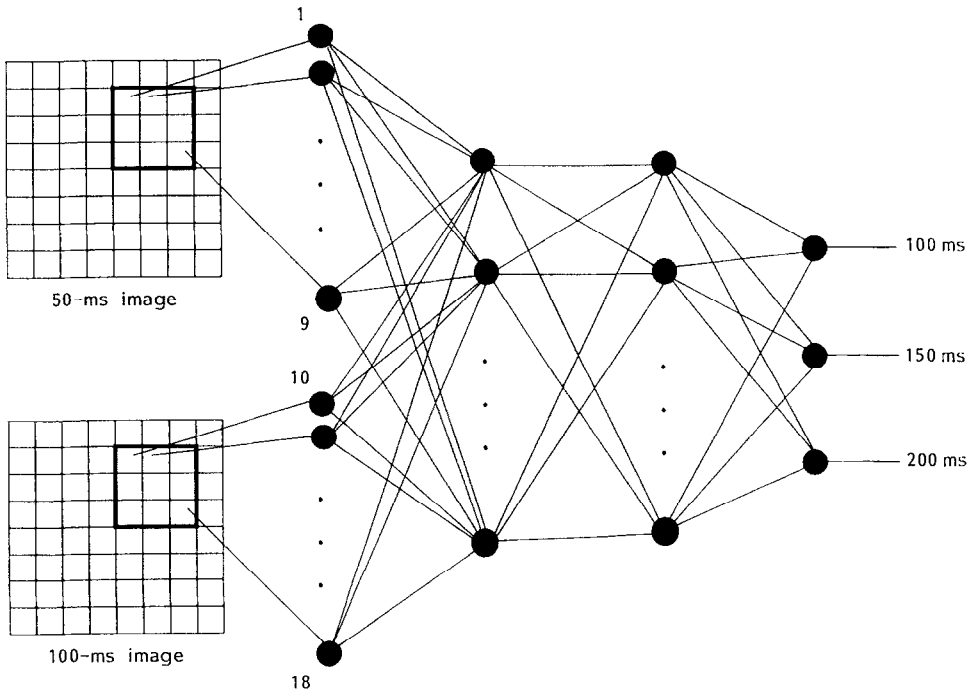

Figure 2 Neural network which makes up the sequence reconstructor subsystem

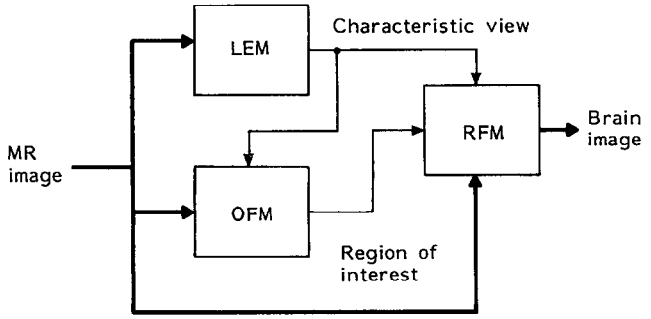

Figure 3 Structure of the brain detector subsystem

sets, as this could result in the network producing null outputs in the presence of low-signal significant inputs[11].

The brain detector

As was explained in the introduction, the brain detector identifies the brain by using specific knowledge about its prototypical shape. The BD is a specific implementation of a previously developed neural-network architecture for tomogram segmentation[12]. In this section we outline its basic structure.

As shown in *Figure 3*, this module includes three ANN submodules: a level evaluator module (LEM), an organ focuser module (OFM), and a region finder module (RFM). LEM and OFM implement an attention-focusing mechanism. Such a mechanism has allowed the performances of RFM to be optimized by making its response invariant with respect to translations and scaling of the observed object. Attention focusing is performed by processing low-resolution images (typically 16×16 or 32×32 matrices) obtained by decimation of the original images after performing a Gaussian smoothing. In this way, the focusing process is mainly driven by large-scale features, rather than by fine details (which, being extremely variable, can be misleading).

As concerns LEM operation, we have chosen six characteristic views which form a minimal set of prototypical images that differ from one another in number and/or position of the imaged organs. The task of LEM is to estimate the characteristic view to which the analysed slice is most similar. In the first phase of its operation, LEM estimates the position of centroid C of the head slice by a network with 16×16 input units, $32 + 32$ hidden units and 32 output units (16 units for each coordinate of C). Afterwards, C is taken as the origin of a polar reference system $[r,\theta]$ and a logarithmic–polar transformation is performed by relocating the image pixels from (r,θ) to $(\log(r+1),\theta)$. The transformed image is then processed by a network with 16×16 input units, $32 + 32$ hidden units, and 6 output units. Each of the output neurons corresponds to one of the characteristic views. It is worth noting that logarithmic processing forces the network to operate chiefly on the basis of the imaged anatomical structures rather than their dimensions.

Following LEM processing, OFM locates the centre of the imaged brain structures and extracts a rectangular region of interest (ROI) around it. The centre is computed by a network which processes: (i) a 16×16 image; (ii) the code of the characteristic view as estimated by LEM. The network structure also includes $32 + 32$ hidden neurons and 32 output neurons.

Once attention focusing has been performed, RFM identifies the ROI pixels that belong to the brain parenchyma. RFM is a five-layer network with two different processing paths: in the first path the grey levels of the pixels enclosed within a 9×9 moving mask are processed, while the position of the moving mask is the input to the second path. These paths converge into the fourth layer of the network, which includes three neurons. The network output is a number belonging to [0, 1], which represents the 'brain-membership' of the pixel at the current mask position.

All the networks of BD were trained with image sets taken from three patients. A total number of 36 images was considered. Each of them was acquired using a spin echo sequence (to form axial slices) with *TE* and *TR* set equal to 100 and 2000 ms, respectively.

The classification module

The sequence reconstructor enhances the signal-decay information, while BD utilizes *a priori* anatomi-

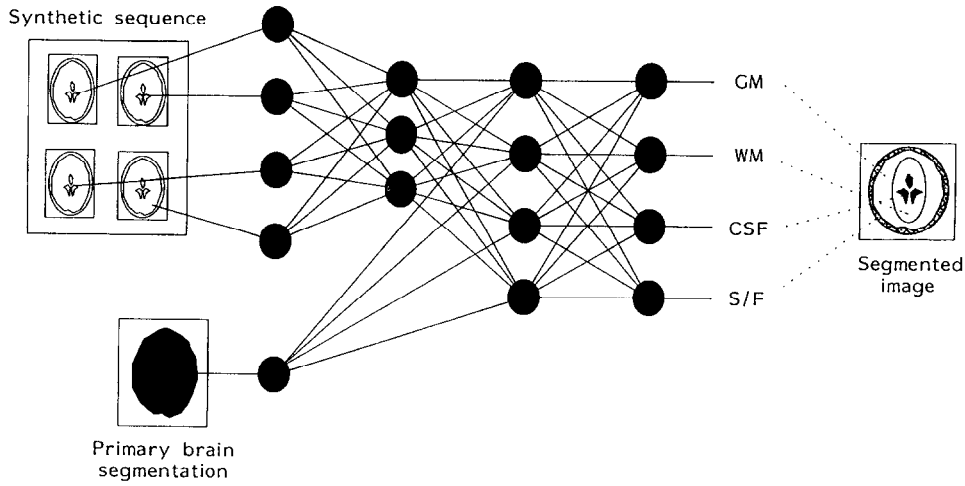

Figure 4 Classification module network

cal knowledge. The classification module processes the SR and BD outputs to produce the final segmentation of the image. The CM consists of a neural network with $5+3+4+4$ units (see *Figure 4*).

The network inputs are: (i) the grey levels of four corresponding pixels from the SR-generated synthetic sequence; (ii) the value of the considered pixel in the primary segmentation generated by the BD.

Each output is representative of a class of tissues: grey matter (GM), white matter (WM), cerebrospinal fluid (CSF) and other hydrated tissues, skin or fat (S/F). The input pixel is assigned to the class represented by the output unit with the highest activation, provided such an activation is greater than a threshold of 0.5. If no outputs are above 0.5, the pixel is classified as background (BG).

As regards network topology, inputs reach two different layers: grey levels are fed into the first layer and anatomical information into the second layer. In this way the latter input has approximately the same importance as the grey-level information coming from the synthetic sequence.

A total of 2500 examples were taken from 25 images coming from three spin echo multislice sequences ($TE = 50$, 100 ms, $TR = 2000$ ms) and from the corresponding segmented image, generated by BD. These examples were classified by an expert radiologist. Half of them (50 examples per image) were used to build the training set. The other 1250 were used to test the system. The values of the teaching inputs were assigned as follows: (i) the output corresponding to the tissue class of the input was given a value of 1; (ii) if the input was classified as background, all outputs were set to 0; (iii) the outputs corresponding to tissues that were different from the correct one, were given values of 0.3 or 0.5 (*Table 1*). With these values we observed a smooth and sufficiently fast decrease of the error function during the training phase. The choice of 0.5 as the threshold value for classification actually derives from this training strategy.

SYSTEM IMPLEMENTATION AND RESULTS

The networks were implemented and trained with a software simulation program written in C language[13]. As regards the SR, the learning phase was fast, with a smooth decrease of the error function. In fact, we observed that learning sessions with different initial conditions and learning rates converge to approximately the same weight set.

Further evidence of this fact is given by the relatively short time needed to reach satisfactory results. In fact, only about 300 iterations are necessary to reach a mean squared error of 1.34×10^{-2} per pattern, and usually, after about 50 iterations, the value of the error is already lower than 1.55×10^{-2}.

Table 1 Teaching input values used to train BS*

	GM	WM	CSF	S/F
Grey matter	1	0.5	0.3	0.3
White matter	0.5	1	0.3	0.3
CSF	0.3	0.3	1	0.3
Skin or fat	0.3	0.3	0.3	1
Background	0	0	0	0

*Each row is the teaching pattern for a given tissue class

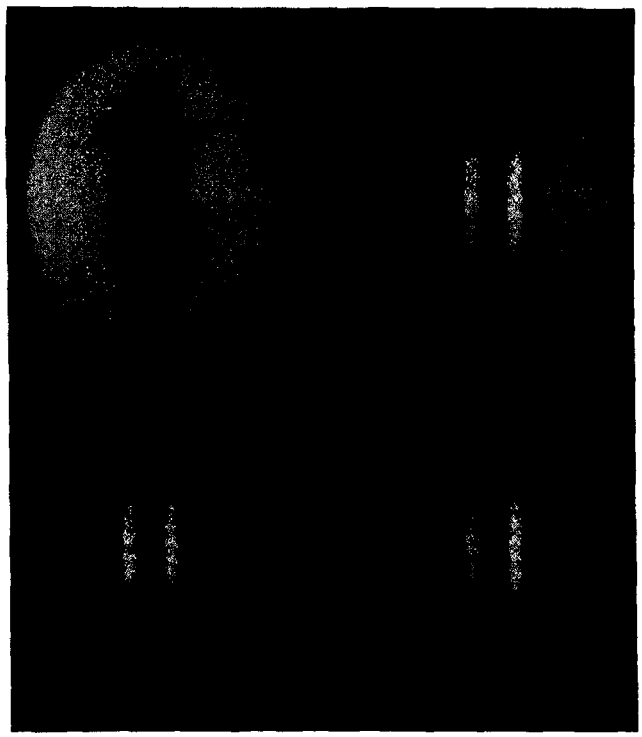

Figure 5 Phantom used for the tests

The module that has imposed the heaviest computational load is the BD. In particular, training of the RFM network has required about 5000 iterations to converge to a mean squared error of 4.5×10^{-3} per pattern.

During the training phase of the CM, we observed that lower learning rates yield better performances, although this causes the convergence to be slower. This has led us to choose a value of 0.15 for the learning rate. As for the SR, no significant influence of the initial values of the weight set was observed. The SR reached an average squared error per pattern of 3.8×10^{-2} in 10 000 iterations.

To quantify the system performances we used real MR images and a software-generated phantom (*Figure 5*). The phantom is made up of seven overlapping ellipses. Moving from the outside inwards, they represent skin or fat (S/F), grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF), respectively. Our phantom was 'imaged' with a simulated spin echo multiecho sequence with four echoes ($TE = 50$, 100, 150, 200 ms, $TR = 2000$ ms). The grey levels for each region were calculated by averaging the grey levels that characterize each tissue on the whole training set. A null value was given to the background.

Uniform noise, ranging from $-L$ to L grey levels ($L = 5$, 10, 15, 20), with a dynamic range of 256 grey levels, was then added to the first image of the sequence. In the other images of the sequence, higher noise levels were added. In this way, we tried to simulate usual archiving procedures, in which signal decay is compensated for with proper rescaling. Thus images corresponding to 100–200 ms echoes were expanded by factors ranging from 1.39 to 1.56.

The SR was tested to assess its noise sensitivity.

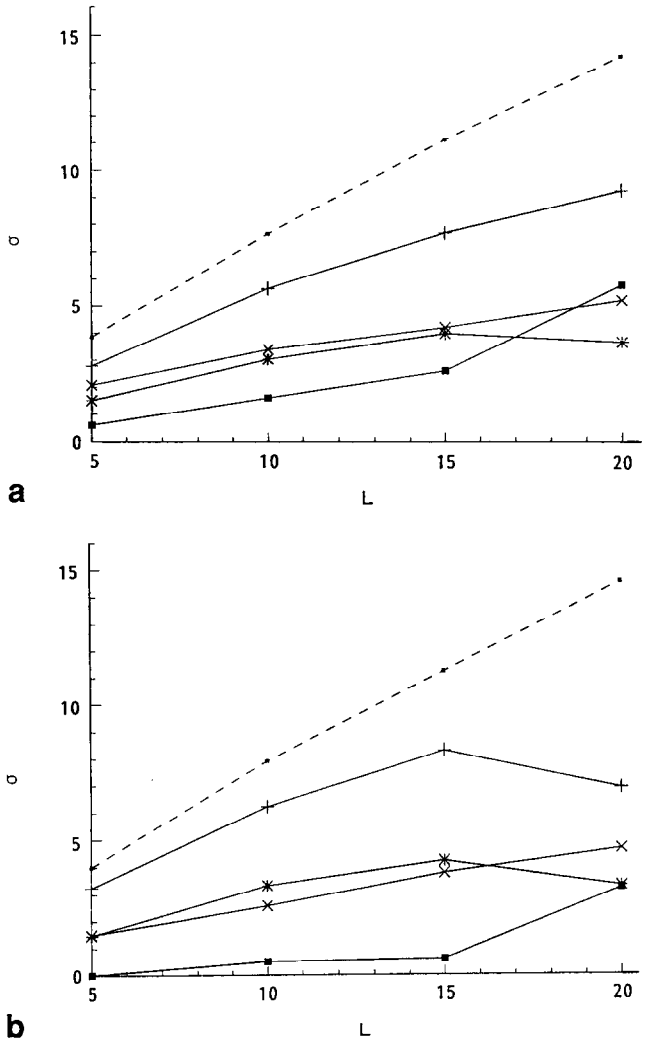

Figure 6 Comparison between the noise levels assessed in the raw images and in different regions of the reconstructed images of the phantom with **a,** $TE = 150$ ms and **b,** $TE = 200$ ms; ■ raw sequence; + grey matter; * white matter; −■− cerebrospinal fluid; × skin or fat

Quantitative tests were performed by comparing the noise levels of a reconstructed sequence of the phantom with those of the 'raw' sequence (the software-generated sequence with added uniform noise). Noise levels were evaluated as the standard deviations σ of signal in constant-level regions of the phantom. The results reported in *Figure 6* demonstrate that the SR has good noise-rejection properties. Comparable results and other data have been reported[10], which were obtained using the same network architecture to synthesize multiecho sequences with $TE = 120$, 150, 180 ms and $TR = 2000$ ms. The SR was also tested with real images. In *Figure 7*, a real multiecho coronal sequence (top) is compared with the corresponding one, synthesized by the SR (bottom).

According to these results, the synthetic images generated by the network are faithful, low-noise emulations of the real ones from the late echoes of a spin echo sequence. Furthermore, since comparable results have been obtained on both coronal and axial sequences, SR performance does not seem to depend on the orientation of the slicing plane. The synthesis of the sequence is quite fast. On an HP-Apollo 710 workstation, about 11 s are needed to generate a sequence, with a 256×256 matrix.

Extensive results concerning the BD were obtained from a set of three brain images[12]. In particular, an average true-positive rate of about 95% and a true-negative rate of 99% were observed in the identification of brain pixels.

A first series of tests on the CM have been performed using the phantom. For each noise level, the 'raw' noisy sequences were processed first. In order to evaluate the performances of the whole system, the same measurements were repeated using the SR to generate the longest-TE images of the input sequence.

The results of such tests are compared in *Figure 8* and *Table 2*. As can be seen, the CM is insensitive to

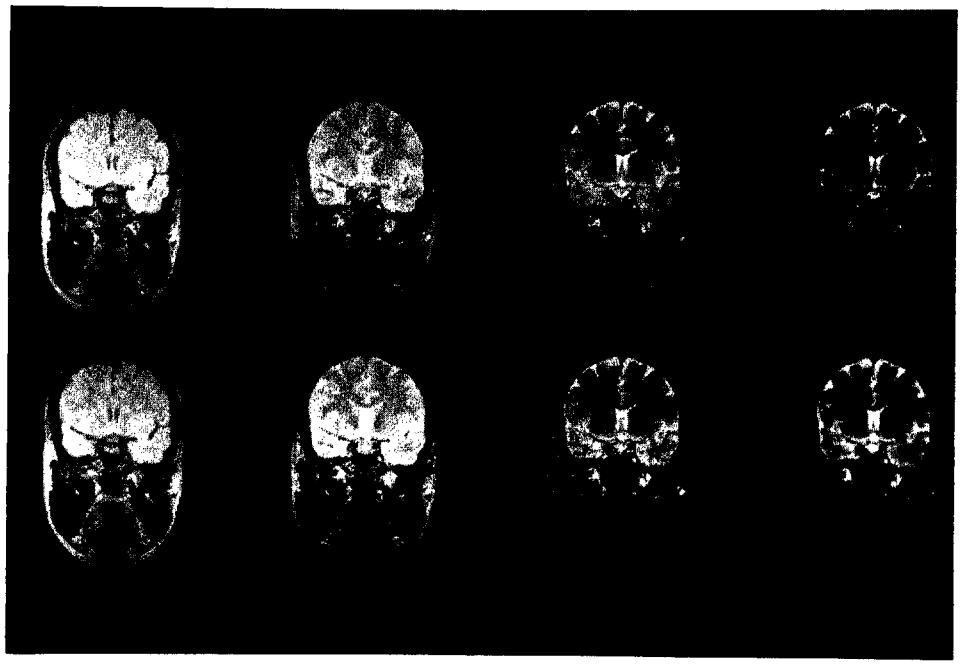

Figure 7 Comparison between a real multiecho sequence with $TE = 50$, 100, 150, 200 ms, $TR = 2000$ ms (top) and the corresponding one reconstructed by the SR (bottom)

Figure 8 Segmentations of the phantom sequence obtained with and without using the SR. Top row: the raw images of the phantom with increasing added noise ($L = 5$, 10, 15, 20); mid row: the corresponding segmentations obtained without SR processing; bottom row: the corresponding segmentations obtained after SR processing

Table 2 Number of misclassifications for different L values and total number of pixels for each tissue class in phantom sequences

	Raw sequence				Reconstructed sequence				Total no. of pixels
L	5	10	15	20	5	10	15	20	
Grey matter	0	0	0	0	0	0	0	0	18382
White matter	0	7	493	3892	0	0	2	2464	7980
CSF	0	0	0	14	17	28	41	289	1844
Skin or fat	0	0	0	0	0	0	0	0	6436
Background	0	0	0	0	0	0	0	0	30894

noise levels computed with $L = 5$, 10. The few misclassifications that can be observed (at $L = 5$, 10) on the sequence reconstructed by the SR are not caused by noise, but by a spatial smoothing introduced in the reconstruction process. In fact, all such misclassifications can be observed only on the boundaries between the regions. The smoothing effect probably derives from the architecture of the first module, which operates on a square input window. Nevertheless, this is a low price to be paid to achieve a good noise rejection. In fact, in the tests performed with higher noise levels ($L = 15$, 20), better results were achieved on the reconstructed sequence. The greater difference between the performance achieved with

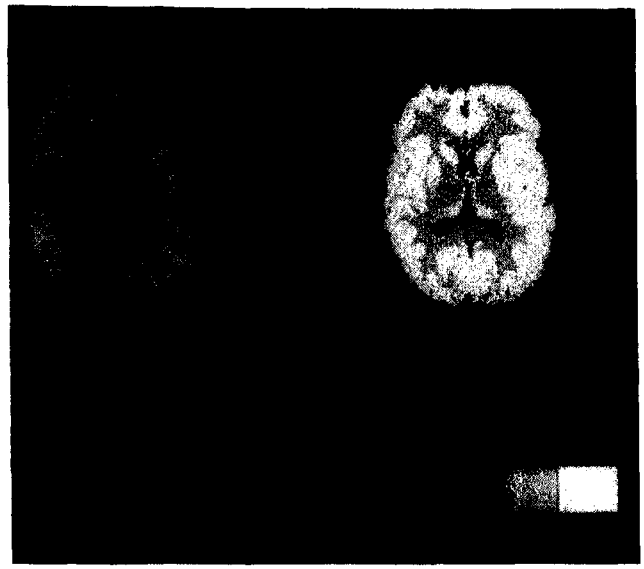

Figure 9 An axial tomogram (left) and the corresponding segmentation (right)

and without the reconstructed sequence can be observed on the white matter. The misclassified pixels in the reconstructed sequence were about 95% less ($L = 15$) and 40% less ($L = 20$) than in the original sequence. It should also be noted that the noise levels corresponding to $L = 20$ are well above the average noise level typically observed in real MR experiments.

To test classification accuracy, the system was separately tested with the training and the test set patterns. *Table 3* shows the so-called confusion matrices calculated on both sets[14]. In such matrices, the diagonal element C_{ii} is the percentage of pixels belonging to class i that have been correctly classified; the off-diagonal element C_{ij} is the percentage of pixels belonging to class i that have been misclassified as belonging to class j. The system achieved an average accuracy of about 92.2% on the training set, with its performance ranging from 89.9% (CSF) to 98.2% (S/F). On the test set, an overall accuracy of 94.5% (worst 84.2% [CSF], best 99% [S/F]) was achieved. *Figure 9* shows an axial tomogram on the left, and the corresponding segmented image on the right.

CONCLUSIONS

Artificial neural networks have proved adequate for integrating the use of *a priori* anatomical knowledge

Table 3 Confusion matrices and overall performances of CM calculated on the training and on the test set

	Training set					Test set				
	GM	WM	CSF	S/F	BG	GM	WM	CSF	S/F	BG
Grey matter	91.4	7.6	0.8	0.2	0	92.9	6.0	1.1	0	0
White matter	7.7	92.2	0.2	0	0	6.9	92.0	0.4	0	0.7
CSF	5.9	4.2	89.9	0	0	7.9	1.3	84.2	5.3	1.3
Skin or fat	0	0	0.9	98.2	0.9	0.5	0	0	99.0	0.5
Background	1.3	0	0.6	0.6	97.5	0	0	0	0.4	98.6
Global accuracy (%)			92.96					94.56		

with low-level processing capabilities, as is usually needed to segment medical images.

As shown by the experimental results, the system is characterized by accuracy and noise immunity. Moreover, good spatial coherence is typical of the segmentations produced by our system, even though such segmentations rely mainly on a pixel-based classification.

Although we have trained the system with examples taken only from healthy brain tissues, the same architecture could be used for detecting and classifying pathological tissues (not necessarily in the brain) as well. In this case, 3D shape and dimensional information about pathological entities, such as tumours, can be recovered from a stack of segmented slices. We are planning to use the system to detect demyelinating plaques caused by multiple sclerosis and to study their time evolution.

ACKNOWLEDGEMENTS

This work was partially supported from MURST (Ministero per l'Università e la Ricerca Scientifica e Tecnologica), and CNR (Consiglio Nazionale delle Ricerche).

REFERENCES

1. Ballard DH, Brown CM. *Computer Vision*. Englewood Cliffs: Prentice-Hall, 1982.
2. Udupa JK. Display of 3-D information in discrete 3-D scenes produced by computerized tomography. *Proc IEEE* 1983; **71:** 420–31.
3. Riederer SJ, Suddarth SA, Bobman SA, Lee JN, Wang HZ, MacFall JR. Automated MR image synthesis: feasibility studies. *Radiology* 1984; **153:** 203–6.
4. MacFall JR, Riederer SJ, Wang HZ. An analysis of noise propagation in computed T2, pseudodensity, and synthetic spin-echo images. *Med Phys* 1986; **13** (3): 285–92.
5. Darwin RH, Drayer BP, Riederer SJ, Wang HZ, MacFall JR. T2 estimates in healthy and diseased brain tissue: a comparison using various MR pulse sequences. *Radiology* 1986; **160:** 375–81.
6. Wassermann PD. *Neural Computing. Theory and Practice*. New York: Van Nostrand Reinhold, 1989.
7. Maren A, Harston C, Pap R. *Handbook of Neural Computing Applications*. San Diego: Academic Press, 1990.
8. Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. In: Rumelhart DE, McClelland JL, eds. *Parallel Distributed Processing. Explorations in the Microstructure of Cognition*, Vol 1. Cambridge, MA: MIT Press, 1985; 318.
9. Piraino D, Sundar S, Richmond B, Schils J, Thome J. Segmentation of magnetic resonance images using a backpropagation neural network. *Proc Ann Int Conf IEEE EMBS* 1991, **13**(3): 1466–7.
10. Brant-Zawandzki M, Norman D. *Magnetic Resonance Imaging of the Central Nervous System*. New York: Raven Press, 1987.
11. Cagnoni S, Caramella D, De Dominicis R, Valli G. Neural network modelling of spin echo multiecho sequences. *J Dig Imag* 1992; **5** (2): 89–94.
12. Coppini G, Poli R, Rucci M, Valli G. A neural network architecture for understanding discrete 3D scenes in medical imaging. *Comp Biomed Res* 1992; **25:** 569–85.
13. Poli R, Cagnoni S, Coppini G, Livi R, Valli G. A neural network expert system for diagnosing and treating hypertension. *Computer* 1991; **24:** 64–71.
14. Özkan M, Sprenkels HG, Dawant BM. Multi-spectral magnetic resonance image segmentation using neural networks. *Proc Int Joint Conf on Neural Networks* 1990; **I:** 429–30.