

Neuronal networks and textural descriptors for automated tissue classification in endoscopy

GEORGE D. MAGOULAS

School of Computer Science and Information Systems, Birkbeck College, University of London, London WC1E 7HX, UK

Received September 7, 2005; Accepted November 14, 2005

Abstract. This study examines the potential of neuronal networks and textural feature extraction for recognising suspicious regions in endoscopy under variable perceptual conditions and systematic or random noise in the data. Second-order statistics and discrete wavelet transform-based methodologies are examined in terms of their discrimination abilities, and several neuronal network learning algorithms are compared in terms of success. The results provide numerical evidence that neuronal networks are capable of classifying offline and online tissue samples extracted from standard images and VHS videotape recordings of colonoscopy procedures with satisfactory success rates. This type of technology could prove to be useful for developing intelligent adaptive systems that will assist medical experts in real-time to automate minimally invasive diagnostic procedures.

Introduction

Computer-based models of neuronal networks (NNs) are increasingly being used in medicine, especially in the development of sophisticated intelligent systems for medical image interpretation (1). The aim is to assist experts in identifying malignant regions using minimally invasive imaging procedures, which will hopefully increase the ability of the expert to detect cancer regions, and decrease the need for intervention and maintain the ability for accurate diagnosis. Furthermore, it may become possible to examine larger areas, studying living tissue *in vivo*, possibly at a distance, and thus minimise the shortcomings of biopsies, such as a limited number of tissue samples, delay in diagnosis, and patient discomfort. In medical practice, endoscopic approaches and other minimally invasive techniques are now permitting the visualisation of previously inaccessible regions of the body.

Computer-based neuronal methodologies present some attractive qualities, such as 'learning from data,' generalisation,

and handling uncertainty and ambiguity in distorted or noisy images. NN-based methods are, in principle, capable of assisting human experts in diagnostic endoscopy. However, generating neuronal models on the basis of some data sets encounters several difficulties because data are often incomplete (missing parameter values), incorrect (systematic or random noise in the data), and/or inexact (inappropriate selection of parameters for the given task).

In computer-assisted endoscopy, given a medical image, the 'true' features associated with the physical properties of the tissue are not exactly known in advance. Usually, one or more feature extraction models are used to provide values for the parameter of each feature (2). The findings are then used to infer the correct interpretation. In this same task of interpretation, on the basis of local changes on the properties of the tissue under examination, the performance of human perception is considered outstanding. Furthermore, medical experts have the ability to add or remove components from an image to give meaning to what they see, and can also adapt to changes to the extent that even a distorted image can be recognised. The situation is far more challenging with computerised systems, especially when several objects of different kinds, related by a set of spatial-temporal relations, are present in the observed scene.

In this study, we focus on interpreting colonoscopic images and detecting tumours using NNs-based tissue classification. Detecting malignant regions in these video sequences encounters several problems due to the time varying nature of the process, and variable perceptual conditions (shading, shadows, lighting conditions, reflections, etc.).

Materials and methods

Computer-based classification of tissue using endoscopy images is usually quite sophisticated and involves multiple levels of processing. Fig. 1 provides an overview of the approach used in this study. We provide an overview of the two most important stages, namely feature extraction and neuronal network-based classification.

Feature extraction using statistical features of the texture. The purpose of this stage is to discriminate between normal and suspicious tissue regions on the basis of their local surface properties, as expressed by textural features. To this end, an endoscopic image was separated into 16x16 pixel windows and two feature extraction techniques were investigated: co-occurrence matrices and 2-dimensional discrete wavelet

Correspondence to: Dr George Magoulas, School of Computer Science and Information Systems, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK
E-mail: gmagoulas@dcs.bbk.ac.uk

Key words: neuronal network, endoscopy, automated tissue classification, statistical descriptor

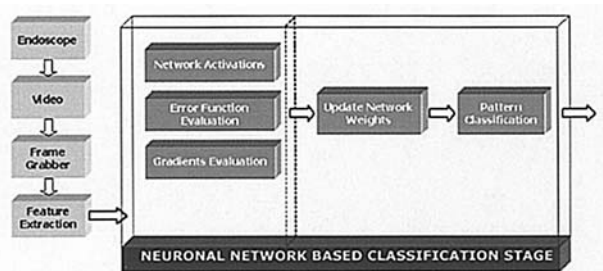


Figure 1. Main stages of the experimental approach.

transform (DWT). First, a one-level wavelet decomposition of the image regions was performed (3). This resulted in four wavelet channels: one approximate and three detail wavelet channels with frequency indexes 2, 3, and 4. Feature extraction was then conducted using the information from the co-occurrence matrices, where each element of a matrix represents the probability of going from a pixel with grey level (i) to another pixel with grey level (j) under a predefined distance and angle. We considered four angles, namely 0° , 45° , 90° , and 135° , as well as a predefined distance of one pixel. From these matrices, a set of statistical measures, or feature vectors, were computed for building different texture models. These were based on calculating the energy-angular second moment, correlation, inverse difference moment, and entropy (detailed in ref. 2) of the detail channels to create a more reliable framework for the generation of the textural descriptors (4). Hence, a set of four features was obtained for each window by calculating the four statistical measures. In this way, a feature vector containing 16 features that uniquely characterises each image window of the selected wavelet channel in the wavelet domain was formed. This procedure, as applied to the three detail channels, resulted in $3 \times 16 = 48$ relevant measures, which formed the input pattern of the neural classifier (4).

Neuronal network learning and adaptation for classification. Models of neuronal networks are inspired by information processing and computation in the human brain (5). Nodes, or artificial neurons, in these models are usually considered simplified models of biological neurons. Artificial neurons can be connected, forming multilayered neuronal structures called multilayer perceptrons (MLPs).

Learning in NNs is usually achieved by minimising a measure related to network's error, as defined by the difference between the actual network output and one that is user defined (5). This approach is popular in training and includes training algorithms that can be divided in two categories: offline, also called batch; and online learning, also called stochastic (5). Offline learning is considered the classical machine learning approach in which a set of examples is used for learning a good approximating function, i.e. train the network before it is used in the real application, whilst online learning is related to updating network parameters after the presentation of each training example, which may be sampled with or without repetition. Online learning appears to be more robust than batch learning as errors, omissions or redundant data can be corrected or ejected during the training phase.

Training data can often be generated easily and in great quantities online during the endoscopy procedure, but are

usually scarce and precious before the procedure. However, the training set may not be able to represent all possible variations of the environment in which the neuronal network will operate because endoscopy is a time varying process influenced by a number of factors, such as diffused light conditions, changes in the perceptual direction of the physician, and variations in the environment. Thus, on many occasions, the NN is not able to track the statistical variations of the environment in which it operates, as the statistical characteristics of the concepts/signals generated by the environment change with time. To alleviate this situation, the NN must adapt its parameters, and consequently its behaviour, to the temporal structure of the input (concepts/signals) in a properly controlled fashion. In computer-assisted endoscopy, this can be potentially useful to adapt to: i) changing environments, i.e. the computer system will apply knowledge/ intelligence to manipulate changes in the environment, such as concepts drift; ii) a similar setting, i.e. the system will be capable of being transferred or adjust to altered circumstances, such as new environmental conditions; and iii) a new/unknown situation, i.e. the system evolves (builds new knowledge or improves the existing one) and autonomously finds an approach to solve the problem given incomplete knowledge and a set of examples.

Although it is difficult to define precise boundaries between the different forms of adaptation that can be encountered in endoscopy applications of the future, it is hopefully clear that adaptation aims to handle the uncertainty that may arise from stochastic elements in either the environment (noise/concept drift) or evaluation process (observation noise). Thus, when the appropriateness of a solution changes, gradually or abruptly, as conditions change within the environment, an automated system can learn to adapt to variations of the environment and exhibit the ability to continually find near-optimal classifications.

The performance of a neuronal system can be enhanced by combining several individual networks trained on the same task in a structure called neural ensemble. In our approach, we first generated individual ensemble members and then created an appropriate combination to produce the ensemble output. Our ensemble exploited the notion of diversity, i.e. individual members of the ensemble made errors on different input patterns (6).

In the next section, we present experimental results using several offline and online algorithms. Our experiments explore key issues in designing a neuronal system for computer-assisted diagnosis: i) the effect of the network architecture on the effectiveness of the classification scheme; and ii) the performance success of the various training approaches.

Results and Discussion

We present results from experiments with still endoscopy images and colonoscopy video sequences provided by collaborators in the Section for Minimally Invasive Surgery of the University of Tübingen. They provide examples of different types of abnormalities recorded under various perceptual conditions and were considered appropriate for our experiments to explore how the neuronal approach will deal with changing environments. In all cases, texture features were extracted from randomly selected 16×16 pixel image windows with 64 grey levels.

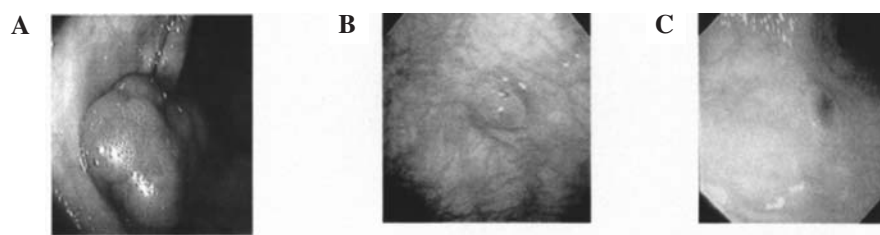


Figure 2. Sample of endoscopy images used in the experiments showing different abnormalities.

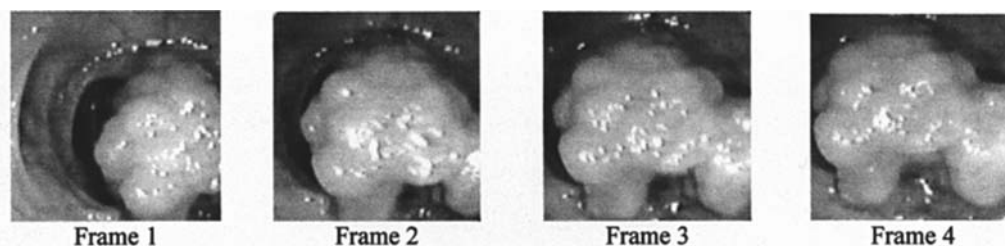


Figure 3. Four frames of a video sequence showing a polypoid tumor of the colon.

To investigate the effect of the NN architecture on the effectiveness of the tissue classification scheme, we conducted a set of experiments using MLPs with 5, 10, 15, 20, 25 and 30 hidden neurons (2). MLPs were trained with the batch adaptive learning rate backpropagation (BP) algorithm (7), using 100 different initial weight sets and samples from image A (Fig. 2). A set of 200 wavelet transform-based feature patterns were used for training and 400 for testing. Networks with 15 hidden neurons exhibited the best average tissue classification success (96.5%). With regards to the total computational cost as measured by the number of error function and gradient evaluations, MLPs with 10 to 20 hidden neurons exhibited the lowest average training cost, while the worst case performance was observed with 5 hidden neurons (765 evaluations in total).

In the second set of experiments, 1000 MLPs with varying number of hidden neurons (from 8 to 21) were trained using two batch learning algorithms, the adaptive backpropagation (ABP), (8), and the Levenberg-Marquardt (LM) (9). The co-occurrence matrices for the textural description of tissue samples was applied to the two images, and 10 normal/10 abnormal samples were randomly extracted from each image (Fig. 2B and C). Malignant regions in these images belong to two different types: image B is a low grade cancer, while image C is a moderately differentiated carcinoma (10). The performance of the ABP and LM trained MLPs was tested on a set of 40 normal and 40 malignant texture samples randomly selected from the two images and different from the training set. Depending on the NN architecture, only a small sample (7%-15%) of the 1000 MLPs trained with each algorithm exhibited a classification success of $\geq 90\%$ that may be considered 'acceptable.' In general, LM-trained MLPs outperformed ABP-trained ones in classification success; LM exhibited the highest average percentage (96.75%) using NNs with 11 hidden nodes.

In the third set of experiments, 1000 MLPs with 11 hidden neurons were trained on a set of 300 samples extracted using co-occurrence matrices from frame 3 of the video sequence (Fig. 3). Training was applied offline using the Rprop algorithm (11), scaled conjugate gradient (SCG) (12)

and LM algorithm. Although medical experts may be able to detect suspicious regions with great accuracy, this is not the case for the automatic system when tested with 3969 samples from the same frame. Only 2 MLPs of the 1000 trained with the Rprop algorithm achieved recognition success of 90%-100%. For the SCG algorithm, the corresponding number is 3 of 1000, while this number is slightly higher for the LM, with 6 of the 1000 MLPs exhibiting success between 90%-100%. The best result for each training method is 92% for the Rprop, 92.4% for the LM and 92.6% for the SCG. Rprop, despite its simplicity, was also the fastest, requiring only an average of 0.644 sec.

Results of MLPs with the best classification performance for each of the four frames of the video sequence are summarised in Table I. Rprop exhibits the best overall performance. Note that the results of Table I were achieved by offline training of a special MLP architecture with 11 hidden neurons for each frame using 300 patterns randomly chosen from that frame and testing it on the same frame. The last row of Table I shows the result obtained by a diverse ensemble that employed the Rprop algorithm for the individual NNs.

Finally, the last experiment concerned online learning (13,14). The average classification success is shown in Table II. Results of the Rprop, BP and ABP algorithms and neural ensemble were obtained using a predefined training set of 1200 patterns (300 randomly chosen patterns from each frame) and a test set of 16000 patterns (the test set constitutes the whole image region in each of the four frames). MLP training was conducted offline, as these methods are only suitable for batch learning (requiring information from the whole training set to be available beforehand). Results of the online descent (13) and online differential evolution (14) were obtained under more realistic conditions, as the weights are updated after each pattern presentation. Although the first impression might be that the large training set affected the classification success of Rprop when compared to the best single frame case (Table I), this is not totally true because the average performance of the method was significantly improved from 78% to 85% (Table II). It was possibly caused by the

Table I. Best classification success for the three first order batch learning methods and the neural ensemble.

Method	Frame 1	Frame 2	Frame 3	Frame 4
Rprop	92%	91%	92%	93%
BP	78%	74%	77%	78%
ABP	81%	85%	83%	81%
Ensemble-based	92%	93%	97%	95%

Table II. Average classification success for methods in the four frames.

Method	Average
Rprop	85%
BP	78%
ABP	79%
Ensemble-based	93%
Online BP	83%
Online evolution	90%

diversity among the patterns of the four frames included in the large training set. Thus, the larger training set allowed us to better model the changes in the environment, at least offline.

The average classification success of online evolution is close to the best results obtained by batch training methods reported above. Networks trained with online evolution are able to perform satisfactory in changing conditions, as data from different frames are presented to the same network.

In conclusion, the development and use of neuronal network-based systems for automated tissue classification in endoscopy requires following a series of stages that constitute an information life cycle through which medical information is acquired, processed, stored, and used.

NNs have a relative advantage for knowledge acquisition compared to other methods because they can handle incomplete information and noise embedded in the training data. Examples are critical for accurate diagnosis and appropriate interpretation of the findings when neuronal networks are used. This of course requires clinicians to be able to express competence and knowledge in a form that can be represented and processed by a NN. The approach proposed here was based on identifying properties of the tissue from images using statistical descriptors and can be potentially useful in cases where there is no previous experience in the interpretation of images, such as when new imaging technologies are used or when no histological confirmation is available.

With regards to processing and storing medical information, the role of the medical expert is important as he/she is responsible for the quality of data, e.g. avoid inconsistencies and cover all possible cases. Algorithms have been proposed in this study to automatically train the NNs using available data. In particular, we explored offline and online learning methods. The best results were obtained by a new approach that can be considered a hybrid scheme combining a variant of the stochastic gradient descent with a differential evolution strategy to (re)adapt the NN to modified environmental conditions. This hybrid scheme was compared to standard

methods that have traditionally been used for training NNs offline. Preliminary results suggest that NNs trained with this algorithm detect malignant regions of interest with high accuracy. Nevertheless, in a real-world setting, the medical expert must examine and interpret the NN classifications and perhaps recommend training the network again to obtain a different, more suitable model of the domain.

Finally, the way information is processed and stored in a NN may make clinicians reluctant to use these methods no matter how superior its performance. There is an obvious need to enhance the usability of these approaches by providing explanation capabilities and interfaces that will facilitate the communication of knowledge between system and clinician.

Our work in this area is ongoing, as extensive testing with long video sequences and under various conditions is necessary to investigate the properties, robustness, and effect of the parameters involved in this automated approach, and thus assess the potential benefits of integrating it into clinical practice.

Acknowledgements

Thanks to V. Plagianakos, M.N. Vrahatis (University of Patras), D. Iakovidis (University of Athens), S. Karkanis (Technological Educational Institute of Lamia, Greece), and M.O. Schurr (University of Tübingen) for insight and advice that benefited this study.

References

1. Karkanis S, Magoulas GD and Theofanous N: Image recognition and neuronal networks: intelligent systems for the improvement of imaging information. *Minim Invasive Ther Allied Tech* 9: 225-230, 2000.
2. Leondes CT: Image Processing and Pattern Recognition. *Neural Network Systems Techniques and Applications Series 5*: Academic Press, 1998.
3. Unser M: Texture classification and segmentation using wavelet frames. *IEEE Trans Image Process* 4: 1549-1560, 1995.
4. Magoulas GD, Karkanis SA, Karras DA and Vrahatis MN: Evaluation of texture-based schemes in neural classifiers training. *WSEAS Trans Comp* 3: 1729-1735, 2004.
5. Looney CG: Pattern recognition using neural networks. Oxford University Press, Oxford, 1997.
6. Sharkey AJC. and Sharkey NE: Combining diverse neural nets. *Knowledge Eng Rev* 12: 231-247, 1997.
7. Magoulas GD, Vrahatis MN and Androulakis GS: Effective backpropagation training with variable stepsize. *Neural Networks* 10: 69-82, 1997.
8. Magoulas GD, Vrahatis MN and Androulakis GS: Improving the convergence of the backpropagation algorithm using learning rate adaptation methods. *Neural Comput* 11: 1769-1796, 1999.
9. Hagan M and Menhaj M: Training feedforward networks with the Marquardt algorithm. *IEEE Trans Neural Networks* 5: 989-993, 1994.
10. Kudo SE, Kashida H, Tamura T, *et al*: Colonoscopic diagnosis and management of nonpolypoid early colorectal cancer. *World J Surgery* 24: 1081-1090, 2000.
11. Riedmiller M and Braun H: A direct adaptive method for faster back-propagation learning: the Rprop algorithm. In: *Proceedings of the IEEE International Conference on Neural Networks*, San Francisco, pp586-591, 1993.
12. Möller M: A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks* 6: 525-533, 1993.
13. Magoulas GD, Plagianakos VP and Vrahatis MN: Adaptive stepsize algorithms for online training of neural networks. *Nonlinear Anal Theory Methods Appl* 47: 3425-3430, 2001.
14. Magoulas GD, Plagianakos VP, and Vrahatis MN: Neural network-based colonoscopic diagnosis using online learning and differential evolution. *Appl Soft Comput* 4: 369-379, 2004.