Neural network-based approach for the classification of wireless-capsule endoscopic images.

Vassilis Kodogiannis
Maria Boulougoura
Harrow School of Computer Science, University of Westminster

Neural network-based approach for the classification of wireless-capsule endoscopic images

V.S. Kodogiannis, M. Boulougoura
Mechatronics Group, School of Computer Science
Univ. of Westminster
London, United Kingdom
E-mail: kodogiv@wmin.ac.uk

Abstract – The importance of computer-assisted diagnosis in endoscopy is to assist the physician in detecting the status of tissues by characterising the features from the endoscopic image. In this paper schemes have been developed to extract new texture features from the texture spectra in the chromatic and achromatic domains for a selected region of interest from each colour component histogram of images acquired by the new M2A Swallowable Imaging Capsule. The concept of fusion of multiple classifiers dedicated to specific feature parameters and the implementation of an advanced intelligent scheme have been also adopted in this study. The high detection accuracy of the proposed systems provides thus an indication that such intelligent schemes could be used as a supplementary diagnostic tool in capsule endoscopy.

I. INTRODUCTION

In medical practice, endoscopic diagnosis and other minimally invasive imaging procedures, such as computed tomography, ultrasonography, con-focal microscopy, computed radiography, or magnetic resonance imaging, are now permitting visualisation of previously inaccessible regions of the body. Their objective is to increase the expert's ability in identifying malignant regions and decrease the need for intervention while maintaining the ability for accurate diagnosis. For more than 10 years, flexible video-endoscopes have a widespread use in medicine and guide a variety of diagnostic and therapeutic procedures including colonoscopy, gastroenterology and laparoscopy [1]. A miniaturised CCD-imager is integrated on the distal side on such endoscopes to acquire intra-corporeal images in video quality ("chip on a stick"). This electronic imager is substituting the fibre-optic bundle of conventional large-diameter flexible endoscopes. Conventional diagnosis of endoscopic images employs visual interpretation of an expert physician. Since the beginning of computer technology, it becomes necessary for visual systems to "understand a scene", that is making its own properties to be outstanding, by enclosing them in a general description of an analysed environment.

Computer-assisted image analysis can extract the representative features of the images together with quantitative measurements and thus can ease the task of objective interpretations by a physician expert in endoscopy. A system capable to classify image regions to normal or abnormal will act as a second - more detailed - "eye" by processing the endoscopic video. Its exceptional value and contribution in supporting the medical diagnosis procedure is high. Endoscopic images possess rich information [2], which facilitates the abnormality detection by multiple techniques. However, from the literature survey, it has been found that only a few techniques for endoscopic image analysis have been reported and they are still undergoing testing. In addition, most of the techniques were developed on the basis of features in a single domain: chromatic domain or spatial domain. Applying these techniques individually for detecting the disease patterns based on possible incomplete and partial information may lead to inaccurate diagnosis. For example, regions affected with bleeding and inflammation may have different colour and texture characteristics. Parameters in the spatial domain related with lumen can be used to suggest the cues for abnormality. For instance, small area of lumen implies the narrowing of the lumen which is often one of the symptoms for lump formation and not the presence of possible bleeding. Therefore, maximising the use of all available image analysis techniques for diagnosing from multiple feature domains is particularly important to improve the tasks of classification of endoscopic images.

Krishnan, *et al.*[3] have been using endoscopic images to define features of the normal and the abnormal colon. New approaches for the characterisation of colon based on a set of quantitative parameters, extracted by the fuzzy processing of colon images, have been used for assisting the colonoscopist in the assessment of the status of patients and were used as inputs to a rule-based decision strategy to find out whether the colon's lumen belongs to either an abnormal or normal category. The quantitative characteristics of the colon are: mean and standard deviation of RGB, perimeter, enclosed boundary area, form factor, and center of mass. The analysis of the extracted quantitative parameters was performed using

three different neural networks selected for classification of the colon. The three networks include a two-layer perceptron trained with the delta rule, a multilayer perceptron with back-propagation (BP) learning and a self-organizing network. A comparative study of the three methods was also performed and it was observed that the self-organizing network is more appropriate for the classification of colon status. Endoscopic images contain rich information of texture. Therefore, the additional texture information can provide better results for the image analysis than approaches using merely intensity information. Such information has been used in CoLD (colorectal lesions detector) an innovative detection system to support colorectal cancer diagnosis and detection of pre-cancerous polyps, by processing endoscopy images or video frame sequences acquired during colonoscopy [4]. It utilised second-order statistical features that were calculated on the wavelet transformation of each image to discriminate amongst regions of normal or abnormal tissue. A neural network based on the classic BP learning algorithm performed the classification of the features. CoLD integrated the feature extraction and classification algorithms under a graphical user interface, which allowed both novice and expert users to utilise effectively all system's functions. The detection accuracy of the proposed system has been estimated to be more than 95%. Recently, a new scheme for on-line training neural network-based colonoscopic diagnosis has been introduced [5]. A memory-based adaptation of the learning rate for the on-line back-propagation is proposed and used to seed an on-line evolution process that applies a differential evolution (DE) strategy to (re-) adapt the neural network to modified environmental conditions. The specific approach looks at on-line training from the perspective of tracking the changing location of an approximate solution of a pattern-based, and thus, dynamically changing, error-function. Results in interpreting colonoscopy images and frames of video sequences were promising and suggest that networks trained with this strategy detect malignant regions of interest with accuracy.

Intra-operative endoscopy, although used with great success, is more invasive and associated with a higher rate of complications. Though the gastrointestinal (GI) endoscopic procedure has been widely used, doctors must be skilful and experienced to reach deep sites such as the duodenum and small intestine. The cleaning and sterilisation of these devices is still a problem leading to the desire for disposable instruments. In GI tract, great skill and concentration are required for navigating the endoscope because of its flexible structure. Discomfort to the patient and the time required for diagnosis heavily depend on the technical skill of the physician and there is always a possibility of the tip of the endoscope injuring the walls. Standard endoscopic examinations evaluate only short segments of the proximal and distal small bowel and barium follow-through has a low sensitivity and specificity of only 10% for detecting pathologies. Hence, endoscopic examination of the entire small bowel has always been a diagnostic challenge. Limitations of the diagnostic techniques in detection of the lesions located in the small bowel are mainly due to the length of the small intestine, overlying loops and intra-peritoneal location. This caused also the desire for autonomous instruments without the bundles of optical fibres and tubes, which are more than the size of the instrument itself, the reason for the objections of the patients. The use of highly integrated microcircuit in bioelectric data acquisition systems promises new insights into the origin of a large variety of health problems by providing lightweight, low-power, low-cost medical measurement devices.
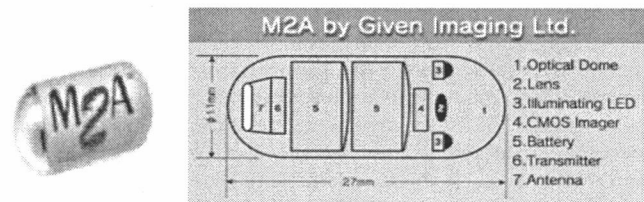

Fig. 1: Given Imaging Capsule

At present, there is only one type of microcapsule which has been introduced recently to improve the health outcome. It might sound like the sci-fi story Fantastic Voyage, but the capsule can efficiently diagnose a range of GI disorders, suspected Crohn's cases, celiac disease, unexplained bleeding, and other small-bowel disorders. This first swallowable video-capsule for the gastroenterological diagnosis has been presented by Given Imaging, a company from Israel, and its schematic diagram is illustrated in Fig. 1 [6]. The system consists of a small swallowable capsule containing a battery, a camera on a chip, a light source, and a transmitter. The camera-capsule has a one centimetre section and a length of three centimetres so it can be swallowed with some effort. In 24 hours, the capsule is crossing the patient's alimentary canal.

For the purpose of this research work, endoscopic images have been obtained using this innovative endoscopic device. They have spatial resolution of 171x151 pixels, a brightness resolution of 256 levels per colour plane (8bits), and consisted of three colour planes (red, green and blue) for a total of 24 bits per pixel. The proposed methodology in this paper is considered in two phases. The first implements the extraction of image features while in the second phase an advanced neural network is implemented / employed to perform the diagnostic task. Texture analysis is one of the most important features used in image processing and pattern recognition. It can give information about the arrangement and spatial properties of fundamental image elements. Many methods have been proposed to extract texture features, e.g. the co-occurrence matrix, and the texture spectrum in the achromatic component of the image. The definition and extraction of

quantitative parameters from endoscopic images based on texture information in the chromatic and achromatic domain is being proposed. This information is initially represented by a set of descriptive statistical features calculated on the histogram of the original image. Additionally, in this study an alternative approach of obtaining those quantitative parameters from the texture spectra is proposed both in the chromatic and achromatic domains of the image. The definition of texture spectrum employs the determination of the texture unit (TU) and texture unit number (N_{TU}) values. Texture units characterise the local texture information for a given pixel and its neighbourhood, and the statistics of the entire texture unit over the whole image reveal the global texture aspects. For the diagnostic part, the concept of multiple-classifier scheme has been adopted, where the fusion of the individual outputs was realised using fuzzy integral. An intelligent classifier-scheme based on the methodology of Extended Normalised Radial Basis Function (ENRBF) neural networks has been also implemented.

II. IMAGE FEATURES EXTRACTION

A major component in analysing images involves data reduction which is accomplished by intelligently modifying the image from the lowest level of pixel data into higher level representations.
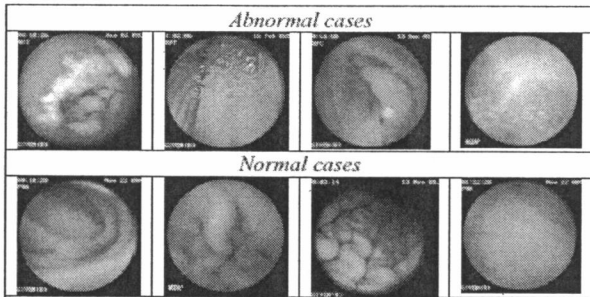

Fig. 2: Selected endoscopic images of normal and abnormal cases.

Texture is broadly defined as the rate and direction of change of the chromatic properties of the image, and could be subjectively described as fine, coarse, smooth, random, rippled, and irregular, etc. For this reason, we focused our attention on nine statistical measures (standard deviation, variance, skew, kurtosis, entropy, energy, inverse difference moment, contrast, and covariance) [7]. All texture descriptors are estimated for all planes in both RGB {R (Red), G (Green), B (Blue)} and HSV {H (Hue), S (Saturation), V (Intensity)} spaces, creating a feature vector for each descriptor $D_i=(R_i,G_i,B_i,H_i,S_i,V_i)$. Thus, a total of 54 features (9 statistical measures x 6 image planes) are then estimated. For our experiments, we have used 70 endoscopic images related to abnormal cases and 70 images related to normal ones. Fig. 2 shows samples of selected images

acquired using the M2A capsule of normal and abnormal cases. Generally, the statistical measures are estimated on histograms of the original image (1st order statistics) [8]. However, the histogram of the original image carries no information regarding relative position of the pixels in the texture. Obviously this can fail to distinguish between textures with similar distributions of grey levels. We therefore have to implement methods which recognise characteristic relative positions of pixels of given intensity levels. An additional scheme is proposed in this study to extract new texture features from the texture spectra in the chromatic and achromatic domains, for a selected region of interest from each colour component histogram of the endoscopic images.

A. N_{TU} Transformation

The definition of texture spectrum employs the determination of the texture unit (TU) and texture unit number (N_{TU}) values. Texture units characterise the local texture information for a given pixel and its neighbourhood, and the statistics of all the texture units over the whole image reveal the global texture aspects. Given a neighbourhood of $\delta \times \delta$ pixels, which are denoted by a set containing $\delta \times \delta$ elements $P=\{P_0, P_1,...., P_{(\delta \times \delta)-1}\}$, where P_0 represents the chromatic or achromatic (i.e. intensity) value of the central pixel and $P_i \{i=1,2,...,(\delta \times \delta)-1\}$ is the chromatic or achromatic value of the neighbouring pixel i, the $TU=\{E_0, E_1,...., E_{(\delta \times \delta)-1}\}$, where $E_i \{i=1,2,...,(\delta \times \delta)-1\}$ is determined as follows:

$$E_i = \begin{cases} 0, & if & P_i < P_0 \\ 1, & if & P_i = P_0 \\ 2, & if & P_i > P_0 \end{cases} \quad (1)$$

The element E_i occupies the same position as the ith pixel. Each element of the TU has one of three possible values; therefore the combination of all the eight elements results in 6561 possible TU's in total. The texture unit number (N_{TU}) is the label of the texture unit and is defined using the following equation:

$$N_{TU} = \sum_{i=1}^{(\delta \times \delta)-1} E_i \times \delta^{i-1} \quad (2)$$

Where, in our case, $\delta = 3$.

The texture spectrum histogram ($Hist(i)$) is obtained as the frequency distribution of all the texture units, with the abscissa showing the N_{TU} and the ordinate representing its occurrence frequency. The texture spectra of various image components {I (Intensity), R (Red), G (Green), B (Blue), H (Hue), S (Saturation)} are obtained from their texture unit numbers. The statistical features are then estimated on the histograms of the N_{TU} transformations of the chromatic and achromatic planes of the image (R,G,B,H,S,V).

III. FEATURES EVALUATION

Recently, the concept of combining multiple classifiers has been actively exploited for developing highly reliable "diagnostic" systems [9]. One of the key issues of this approach is how to combine the results of the various systems to give the best estimate of the optimal result. A straightforward approach is to decompose the problem into manageable ones for several different sub-systems and combine them via a gating network. The presumption is that each classifier/sub-system is "an expert" in some local area of the feature space. The sub-systems are local in the sense that the weights in one "expert" are decoupled from the weights in other sub-networks. In this study, six subsystems have been developed, and each of them was associated with the six planes specified in the feature extraction process (*i.e.* R, G, B, H, S, & V). Each subsystem was modelled with an appropriate intelligent learning scheme. In our case, the Extended Normalised Radial Basis Function Network (ENRBF) has been proposed. Such scheme provides a degree of certainty for each classification based on the statistics for each plane. The outputs of each of these networks must then be combined to produce a total output for the system as a whole as can be seen in Fig. 3.
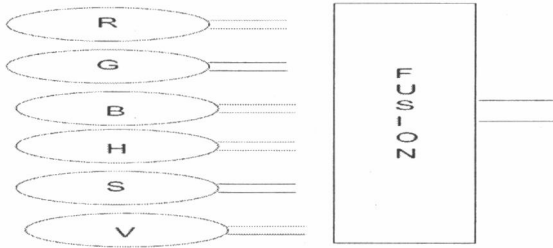

Fig. 3: Proposed fusion scheme.

While a usual scheme chooses one best subsystem from amongst the set of candidate subsystems based on a winner-takes-all strategy, the current proposed approach runs all multiple subsystems with an appropriate collective decision strategy. The aim in this study is to incorporate information from each plane/space so that decisions are based on the whole input space. The adopted in this paper methodology was to use the fuzzy integral concept. Fuzzy integral (FI) is a promising method that incorporates information from each space/plane so that decisions are based on the whole input space in the case of multiple classifier schemes. FI combines evidence of a classification with the systems expectation of the importance of that evidence. By treating the classification results a series of disjointed subsets of the input space Sugeno defined the g_λ-fuzzy measure [11].

$$g(A \cup B) = g(A) + g(B) + \lambda g(A) g(B);$$

$$\lambda \in (-1, \infty) \tag{3}$$

Where the λ measure can be given by solving the following non-linear equation.

$$\lambda + 1 = \prod_{i=1}^{K} (1 + \lambda g^i) \quad \lambda > -1 \tag{4}$$

The $g^i, i \in \{1, ..., K\}$ values are fuzzy densities relating to the reliability of each of the K feature networks and satisfy the conditions of fuzzy sets laid out by Sugeno.

A. Extended Normalised RBF Network

The classification scheme utilised here is an artificial neural network known as an Extended Normalised Radial Basis Function Network (ENRBF) [10], which utilises a series of linear models instead of the linear combiner in an RBF network. The scheme is illustrated in Fig. 4. We propose a supervised training method for this scheme that is fully supervised and self organising in terms of structure. The method incorporates training techniques from Bayesian Ying-Yang (BYY) training which treats the problem of optimisation as one of maximising the entropy between the original non-parametric data distribution based on Kernel estimates or user specified values and the parametric distributions represented by the network. This is achieved through the derivation of a series of Expectation Maximisation (EM) update equations using a series of entropy functions as the Q function or log-likelihood function.
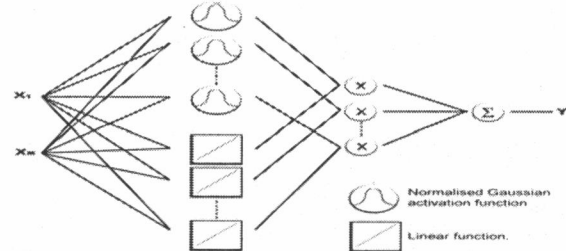

Fig. 4: ENRBF scheme.

The ENRBF network can be represented by the following equations.

$$E(z \mid x, \Theta) = \frac{\sum_{j=1}^{K} (W_j^T x + c_j) p(x \mid j, \theta_j)}{\sum_{j=1}^{K} p(x \mid j, \theta_j)} \tag{5}$$

Where z is the output of the network $z \in Z$, x is an input vector $x \in X$, $\Theta = [W, c, \theta]$ are the network parameters and $\theta = [m, \Sigma]$ are the parameters of the Gaussian activation functions given by:

$$p\left(x \mid j, \theta_j\right) = \exp\left\{-\tfrac{1}{2}\left(x - m_j\right) \Sigma_j^{-1}\left(x - m_j\right)\right\} \quad (6)$$

The BYY method attempts to maximise the degree of agreement between the expected value of z from the network and the true value of z from the training data. It is guaranteed to lead to a local optimum and unlike the original EM algorithm for learning the parameters of Gaussian functions this method encourages coordination between the input and output domains. Like the EM algorithm this method is also very fast in terms of the number of iterations needed for the parameters to converge. However, as BYY is an EM based technique it is still susceptible to locally maximal values. The Split and Merge EM (SMEM) concept for Gaussian Mixture Models (GMM) proposed initially by Ueda, has been applied to the ENRBF scheme. The original SMEM algorithm is able to move neurons from over populated areas of the problem domain to underrepresented areas by merging the over populated neurons and splitting the under-populated. The use of Eigenvectors to split along the axis of maximum divergence instead of randomly as in original SMEM has been proposed recently. The SMEM algorithm suffers from the fact that before terminating all possible combinations of Split and Merge operations must be examined. Although many options can be discounted, the training time still increases exponentially with network size and again suffers from problems inherent with k-means and basic EM in that it is essentially unsupervised. In this work we incorporate the supervised nature of BYY training with improved statistical criteria for determining the neurons which poorly fit their local areas of the problem domain.

IV. RESULTS

The proposed approach was evaluated using 140 clinically obtained endoscopic M2A images. For the present analysis, two decision-classes are considered: abnormal and normal. Seventy images (35 abnormal and 35 normal) were used for the training and the remaining ones (35 abnormal and 35 normal) were used for testing. The extraction of quantitative parameters from these endoscopic images is based on texture information. Initially, this information is represented by a set of descriptive statistical features calculated on the histogram of the original image. ENRBF network is incorporated into a multiple classifier scheme, where the structure of each individual (for R, G, B, H, S, & V planes) classifier is composed of 9 input nodes (i.e. nine statistical features) and 2 output nodes. In a second stage, the nine statistical measures for each individual image component are then calculated though the related texture spectra after applying the (N_{TU}) transformation.

A. Performance of Histograms-based Features

The multiple-classifier scheme using the ENRBF network has been trained on the six feature spaces. The network trained on the R feature space and it then achieved an accuracy of 94.28% on the testing data incorrectly classifying 2 of the normal images as abnormal and 2 abnormal as normal ones.

ENRBF Accuracy (70 testing patterns)		
Modules	Histogram-based	N_{TU}-based
R	94.28% (4 mistakes)	92.85% (5 mistakes)
G	92.85% (5 mistakes)	97.14% (2 mistakes)
B	94.28% (4 mistakes)	95.71% (3 mistakes)
H	91.43% (6 mistakes)	94.28% (4 mistakes)
S	88.57% (8 mistakes)	91.43% (6 mistakes)
V	94.28% (4 mistakes)	97.14% (2 mistakes)
Overall	94.28% (4 mistakes)	95.71% (3 mistakes)

Table 1: Performances of ENRBF

The network trained on the G feature space misclassified 2 normal images as abnormal but not the same ones as the R space. The remaining 3 images were misclassified as normal ones. The B feature space achieved an accuracy of 94.28% on the testing data with 4 misclassifications, i.e. 3 abnormal as normal ones and the remaining one image as abnormal ones.
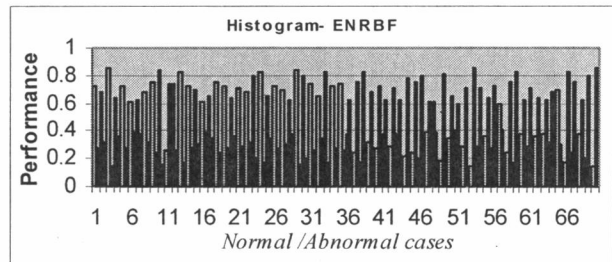

Fig. 5: Histogram-based Performance

The network trained on the H feature space achieved 91.43% accuracy on the testing data. The network trained on the S feature space achieved an accuracy of only 88.57% on the testing data. Finally, the network for the V feature space misclassified 2 normal cases as abnormal and 2 abnormal as normal ones, giving it an accuracy of 94.28% on the testing data. The fuzzy integral (FI) concept has been used here to combine the results from each sub-network and the overall system misclassified 1 normal cases as abnormal and 3 abnormal as normal ones, giving the system an overall accuracy of 94.28%. These results are illustrated in Fig. 5, while Table 1 presents the performance of individual components. It can be shown that in general the confidence levels for each correct classification is above 0.6.

B. Performance of N_{TU}-based Features

In the N_{TU}-based extraction process, the texture spectrum of the six components (R, G, B, H, S, V) have been obtained from the texture unit numbers, and the same nine statistical

measures have been used in order to extract new features from each textures spectrum. In a similar way, a multi classifier consisting of ENRBF networks with 9 input nodes and 2 output nodes was trained on each of the six feature spaces. The N_{TU} transformation of the original histogram has produced a slight but unambiguous improvement in the diagnostic performance of the multi-classifier scheme. Table 1 illustrates the performances of the network in the individual components.
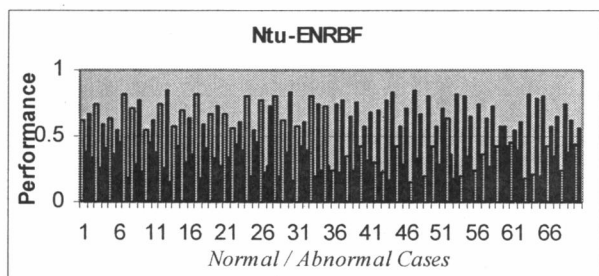

Fig. 6: N_{TU}-based Performance

The ENRBF network trained on the R feature space and it then achieved an accuracy of 92.85% on the testing data incorrectly classifying 3 of the normal images as abnormal and 2 abnormal as normal ones. The network trained on the G feature space misclassified 2 normal images as abnormal but not the same ones as the R space. The B feature space achieved an accuracy of 95.71% on the testing data with 3 misclassifications, i.e. 2 abnormal as normal ones and the remaining one image as abnormal one. The network trained on the H feature space achieved 94.28% accuracy on the testing data. The network trained on the S feature space achieved an accuracy of only 91.43% on the testing data. Finally, the network for the V feature space misclassified 1 normal case as abnormal and 1 abnormal as normal one, giving it an accuracy of 97.14% on the testing data. The fuzzy integral (FI) concept has been used here to combine the results from each sub-network and the overall system provided an accuracy of 95.71%. More specifically, 1 normal case as abnormal and 2 abnormal as normal ones provide us a good indication of a "healthy" diagnostic performance. However the level of confidence in this case was slight less than the previous case (i.e. the histogram), that is 0.54 as shown in Fig. 6.

V. CONCLUSIONS

An approach on extracting texture features from endoscopic images using the M2A Given Imaging capsule has been developed. Statistical features based on texture are important features, and were able to distinguish the normal and abnormal status in the selected clinical cases. The multiple classifier approach used in this study with the inclusion of an advanced

neural network algorithm provided encouraging results. Two approaches on extracting statistical features from endoscopic images using the M2A Given Imaging capsule have been developed. In addition to the histogram-based texture spectrum, a new approach of obtaining those quantitative parameters from the texture spectra is proposed both in the chromatic and achromatic domains of the image by calculating the texture unit numbers (N_{TU}) over the histogram spectrum. Future studies will be focused on further development of this "diagnostic" system by incorporating additional features, investigation of algorithms for reduction of input dimensionality as well as the testing of this approach to the IVP-endoscopic capsule which is under development through a European research project.

REFERENCES

[1] Y. Haga, M. Esashi, "Biomedical Microsystems for Minimally Invasive Diagnosis and Treatment", *Proceedings of IEEE*, Vol. 92, pp. 98-114, 2004

[2] K. Nagasako, T. Fujimori, Y. Hoshihara, M. Tabuchi, *Atlas of Gastroenterologic Endoscopy by High-resolution Video-Endoscopy*, IGAKU-SHOIN Ltd., Tokyo, 1998.

[3] S. Krishnan, P. Wang, C. Kugean, M. Tjoa, "Classification of endoscopic images based on texture and neural network", *Proc. 23rd Annual IEEE Int. Conf. in Engineering in Medicine and Biology*, Vol. 4, pp. 3691-3695, 2001.

[4] D.E. Maroulis, D.K. Iakovidis, S.A. Karkanis, D.A. Karras, "CoLD: a versatile detection system for colorectal lesions endoscopy video-frames", *Computer Methods and Programs in Biomedicine*, Vol. 70, pp. 151–166, 2003.

[5] G. Magoulas, V. Plagianakos, M. Vrahatis, "Neural network-based colonoscopic diagnosis using on-line learning and differential evolution", *Applied Soft Computing*, Vol. 4, pp. 369–379, 2004

[6] G. Idden, G. Meran, A. Glukhovsky and P. Swain, "Wireless capsule endoscopy", *Nature*, pp. 405-417, 2000

[7] R.M. Haralick, "Statistical and structural approaches to texture", *IEEE Proc.*, Vol. 67, pp. 786- 804, 1979

[8] M. Boulougoura, E. Wadge, V.S. Kodogiannis, H.S. Chowdrey, "Intelligent systems for computer-assisted clinical endoscopic image analysis", *2nd IASTED Int. Conf. on BIOMEDICAL ENGINEERING, Innsbruck, Austria*, pp. 405-408, 2004.

[9] E. Wadge, V.S. Kodogiannis, "Intelligent diagnosis of UTI in vivo using gas sensor arrays", *Int. Conf. on Neural Networks and Expert Systems in Medicine and HealthCare, NNESMED 2003, Sheffield UK*, pp. 93-98, 2003.

[10] E. Wadge, V. Kodogiannis, D. Tomtsis, "Neuro-Fuzzy Ellipsoid Basis Function multiple classifier for diagnosis of urinary Tract Infections", *Proc. ICCMSE 2003, Greece*, pp. 673-677, 2003

[11] L.I. Kuncheva, *Fuzzy Classifier Design*, Physica-Verlag, 2000