AUTOMATIC PET VOLUME ANALYSIS AND CLASSIFICATION BASED ON ANN AND BIC

Mhd Saeed Sharif [1], *Maysam Abbod* [1], *Benjamin Krill* [2], *Abbes Amira* [2], *and Habib Zaidi* [3,4]

[1]School of Engineering and Design, Brunel University, West London, United Kingdom
mhd.sharif@brunel.ac.uk
[2]Nanotechnology and Integrated, Bio-Engineering Centre, University of Ulster, United Kingdom
[3]Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, CH-1211 Geneva, Switzerland
[4]Geneva Neuroscience Center, Geneva University, CH-1205 Geneva, Switzerland

ABSTRACT

The increasing number of patient scans and the prevailing application of positron emission tomography (PET) in clinical oncology have led to a real need for efficient PET volume handling and the development of new volume analysis and classification approaches to aid clinicians in the diagnosis of diseases and planning of treatment. A novel automated approach for oncological PET volume classification is proposed in this paper. The proposed intelligent system deploys artificial neural networks (ANN) for classifying phantom and clinical PET volumes. Bayesian information criterion (BIC) has been used in this system to assess the optimal number of classes for each PET data set and assist the ANN block to achieve accurate automatic classification for the region of interest (ROI). ANN performance evaluation has been carried out using confusion matrix and receiver operating characteristic curve. The proposed classification methodology of phantom and clinical oncological PET data has shown promising results and can successfully classify patient lesion.

Index Terms— Positron emission tomography, Artificial neural networks, Bayesian information criterion, Medical volume analysis, Tumour.

1. INTRODUCTION

The clinical use of positron emission tomography (PET) is mainly in oncology, where ^{18}F-fluorodeoxyglucose (FDG) remains the most widely used tracer. It has already had a large valuable effect on cancer staging and treatment and its use in clinical oncology practice continues to evolve [1, 2].

Artificial neural networks (ANN) is one of the powerful artificial intelligence (AI) techniques that has the capability to learn and memorise a set of data and construct weight matrices to represent the learning patterns. The ANN is a mathematical model which emulates the activity of biological neural networks in the human brain. It consists of two or several layers, each one has many interconnected groups of neurons. As ANN has some advantages such as its non-parametric and non-linear nature, arbitrary decision boundary capabilities, easy adaptation to different types of data and input structures, and good generalisation capabilities, it has been successfully used in many applications including pattern classification, decision making, forecasting, computer-vision, and adaptive control. Many research studies have been carried out utilising ANN for different applications. A neural edge enhancer (NEE) based on a modified multilayer neural network for enhancing the desired edges clearly from noisy images was proposed in [3]. Multilayer perceptron (MLP) neural network have been used in [4] to identify breast nodule malignancy using sonographic images. Multiple classifier system using five neural networks and five sets of texture features extraction for the characterization of hepatic tissue from CT images is presented in [5]. Kohonen self-organising neural network for segmentation and a multilayer backpropagation neural network for classifying multispectral MRI images have been used in [6]. Kohonen neural network also used for image segmentation in [7]. Computer-aided diagnostic (CAD) scheme to detect nodules in lung using multiresolution massive training artificial neural network (MTANN) is presented in [8]. Other research studies have been carried out deploying ANN for medical image segmentation [9, 10].

This paper aims to develop an automatic robust PET volume classification system using ANN combined with Bayesian information criterion (BIC). BIC has been used in this system to select the optimal number of clusters for each PET data set. Two PET phantom data sets, a clinical PET volume of non-small cell lung cancer patient, and PET volumes from seven patients with laryngeal tumours have been used to validate the proposed system.

This paper is organised as follows. Section 2 presents the main system components including ANN, BIC, and performance evaluation metrics including confusion matrix (CM) and receiver operating characteristic (ROC) curve. The results and analysis are illustrated in section 3 including a full description about the phantom and clinical PET data sets, and

978-1-61284-842-6/11/$26.00 ©2011 IEEE

finally conclusions and future work are presented in section 4.

2. THE PROPOSED SYSTEM

2.1. Artificial neural network

Multilayer feedforward neural network [11] has been used in this study, as it is flexible, nonlinear models consisting of different number of neurons arranged into multiple layers. An experimental study has been run at the beginning to determine the best ANN design and the training algorithm for our application. The best ANN performance for the proposed application has been achieved using the following design: one input layer, one hidden layer, and an output layer. Hyperbolic tangent sigmoid transfer function has been used for input and hidden layers as this activation function has a nonlinear character. It is also useful in order to discriminate the complex relationships between the features. It also produces the scaled output over the -1 to +1, where -1 and +1 output values are obtained for minus and plus infinity respectively [12, 13]. On the other hand having a broader output space makes this function more efficient for the classification performance. The experimental study shows that a linear activation function is useful for the output layer. Levenberg-Marquardt backpropagation training algorithm has been used during the evaluation of neurons numbers in the hidden layer benefiting from having advantage of being fully automated, and including no user dependent parameters. In order to determine the best number for hidden neurons, a trial-and-error method is used. The results obtained after this evaluation shows that the best number of the hidden neurons which corresponds to the smallest mean squared error (MSE), and good ANN outputs is 70 hidden neurons. The maximum number of iterations used in ANN training is 1000, and each experiment has been repeated 10 times and the average is considered. Due to space constraint the details of these primary experiments and the results of different training algorithms evaluation are not included in this paper.

2.2. Bayesian information criterion

A model selection criterion known as BIC is employed in this study which uses Bayesian inference to assess the optimal number of output classes to be retained. BIC has gained notoriety as a significant approach for model selection and has been used in contexts varying from image processing and analysis [14], to biological and sociological research [15, 16]. BIC values are calculated incrementally, increasing from K = 2 to K = 8. The number of classes K is not further increased, as in this medical application, any additional separation is unnecessary based on expert consultation and comments. BIC values tend to increase indefinitely as the number of components increases. An increase in BIC value indicates improved model fit, however, this value typically stabilises on an approximate curve plateau (when plotted against the number of classes K) the beginning of which is usually taken to indicate the optimal number of output classes.

2.3. Performance metric

In the field of AI a number of performance metrics can be employed to evaluate the performance of ANN. Confusion matrix is a visualisation tool typically used in supervised and unsupervised learning approaches. Each row of the matrix represents the instances in a predicted class. One benefit of a confusion matrix is that it is easy to see if the system is confusing any two classes (the tumour and the remaining tissues in the proposed application). The other performance evaluation approach is receiver operating characteristic (ROC). This approach can be represented by plotting the fraction of true positives rate (TPR) versus the fraction of false positives rate (FPR), where the perfect point in the ROC curve is the point (0,1). The area under the ROC curve (AUC) is a reasonable performance statistic for classifier systems assuming no knowledge of the true ratio of misclassification costs, the best AUC value for a classifier is 1 [17].

2.4. System description

The proposed medical PET volume analysis system is illustrated in Fig. 1. The 3D PET volume acquired from the scanner goes through the preprocessing block, where thresholding, and median filter are utilised to remove external artefacts and enhance smoothly the quality of slices features. The optimal class number is determined by plotting BIC values against different values of K. The optimal number of classes for each slice is considered on an approximate curve plateau, at which the model begins to stabilise. This number is fed to ANN which classifies each processed slice into the corresponding number of classes. Where each voxel is classified into its corresponding class. The classification performance is evaluated then using CM and ROC. The outputs are finally selected and displayed.

3. RESULTS AND ANALYSIS

3.1. PET phantom data

3.1.1. Phantom data set 1

The first data set used in this study is obtained from NEMA IEC image quality body phantom which consists of an elliptical water filled cavity with six spherical inserts suspended by plastic rods of volumes 0.5, 1.2, 2.6, 5.6, 11.5, and 26.5 ml. The inner diameters of these spheres are: 10, 13, 17, 22, 28 and 37 mm. The scanner used for acquiring this volume has a resolution of 4.07 mm x 4.07 mm x 5 mm, with voxel volume 0.0828 ml, while the size of the obtained phantom

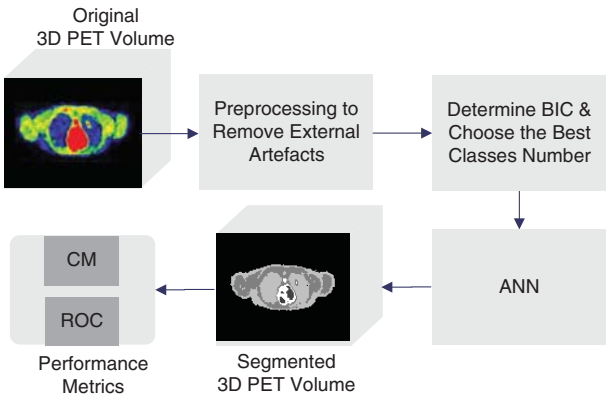

Fig. 1. Proposed system for oncological PET volume classification.

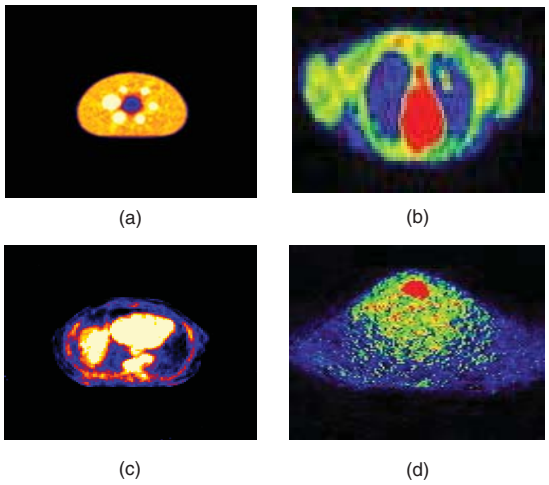

Fig. 2. Original slices: (a) phantom data set 1, (b) phantom data set 2, (c) clinical data set 1, (d) clinical data set 2 (patient 1).

DICOM (digital imaging and communications in medicine) volume is 168 x 168 x 66. This phantom was extensively used in the literature for assessment of image quality and validation of quantitative procedures [18, 19, 20, 21]. Other variants of multisphere phantoms have also been suggested [22]. The PET scanner used for acquiring the data is the Biograph 16 PET/CT scanner (Siemens Medical Solution, Erlangen, Germany) operating in 3D mode [23]. Following Fourier rebinning and model-based scatter correction, PET images were reconstructed using two-dimensional iterative normalized attenuation-weighted ordered subsets expectation maximization (NAW-OSEM). CT-based attenuation correction was used to reconstruct the PET emission data. The default parameters used were ordered OSEM iterative reconstruction with four iterations and eight subsets followed by a post-processing Gaussian filter (kernel full-with half-maximal height, 5 mm).

The total number of voxels in NEMA phantom, Fig. 2.a, is 28224 voxels, which are divided into 4 classes; class 1 has 24939 voxels, class 2 has 287 voxels, class 3 has 2847 voxels, and class 4 has 151 voxels. This number of classes matches the optimal number of classes for this data set decided using BIC plot which is 4 classes. Fig. 3 shows BIC values against the class number K, where the graph starts to stabilise on an approximate curve plateau at value of 4. The ANN outputs (Fig. 7.a) evaluated using CM which shows that 5 voxels are misclassified from class 2, and 6 voxels are misclassified from class 3 (2 voxel in class 2 and 4 voxels in class4), while class 1 and 4 are precisely classified. The fraction of samples misclassified in total is 0.001417, which indicates a good ANN performance for classifying this data set. CM for this data set is illustrated in Table 1. The other performance metric used is calculating the AUC of ROC for all classes which is almost 1 as shown in Table 2.

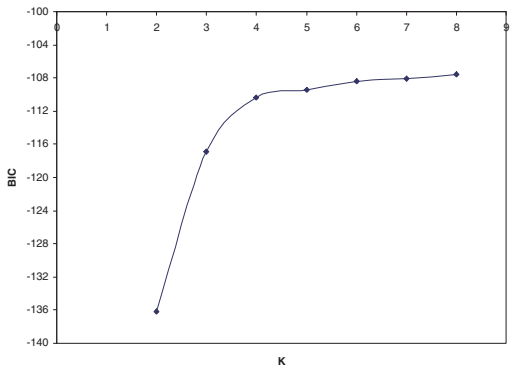

Fig. 3. Plot of BIC values for phantom data set 1, scaled by a factor of 1000.

3.1.2. Phantom data set 2

The second data set consists of Monte Carlo simulations of the Zubal antropommorphic model where two volumes were generated. The first volume contains a matrix with isotropic voxels, the size of this volume is 128 x 128 x 180. The second volume contains the same matrix of the first one but with non-isotropic voxels, it has a size of 128 x 128 x 375. The voxel size in both volumes is 5.0625 mm x 5.0625 mm x 2.4250 mm. The second data volume has 3 tumours in the lungs whose characteristics are given in Table 3.

The total number of voxels in this data set (tumour 1), Fig. 2.b, is 16384 voxels, which are divided into 5 classes; class 1 has 5327 voxels, class 2 has 7895 voxels, class 3 has 971 voxels, class 4 has 3 voxels, and class 5 has 2188 voxels. For isotropic voxels in phantom data set 2, the optimal class number obtained from BIC plot is 5 classes. The number of

Table 1. CM for phantom data set 1 classes

Class1	Class2	Class3	Class4
24939	0	0	0
5	282	0	0
0	4	2841	2
0	0	0	151

Table 2. AUC of ROC for all data sets

AUC	Class1	Class2	Class3	Class4	Class5
Data set1	1.00	0.99	0.99	0.99	-
Data set2	0.99	0.99	0.99	0.99	0.99
Clinical set1	0.88	0.86	0.94	0.79	0.66
Clinical set2	0.99	0.84	0.80	0.94	0.87

optimal class is the same for all tumours 1, 2, and 3. For volume with non isotropic voxels in phantom data set 2, the optimal class number obtained from BIC plot is also 5 classes for all the three tumours. Fig. 4 shows BIC values against the class number K for isotropic phantom data set 2, where the graph starts to stabilise on an approximate curve plateau at value of 5. The ANN outputs for tumour 1 is illustrated in Fig. 7.b. The CM for this data set shows that 1 voxel is misclassified from class 4 (the ROI), and the rest of the classes are precisely classified, Table 4. The fraction of samples misclassified in total is 4.068348E-04. The AUC of ROC for all classes is almost 1 as indicated in Table 2.

3.2. Clinical PET volume

3.2.1. Clinical data set 1

Clinical PET volumes of patients with histologically proven NSCLC (clinical Stage Ib-IIIb) who have undertaken a diagnostic whole-body PET/CT scan were used for assessment of the proposed classification technique. Patients fasted no less than 6 hours before PET/CT scanning. The standard protocol involved intravenous injection of ^{18}F-FDG followed by a physiologic saline (10 ml). The injected FDG activity was adjusted according to patient's weight using the following formula: A (Mbq) = weight 4 + 20. After 45 min uptake time, free-breathing PET and CT images were acquired. The data were reconstructed using the same procedure described for the phantom studies. The maximal tumour diameters measured from the macroscopic examination of the surgical specimen served as ground truth for comparison with the maximum diameter estimated by the proposed classification technique.

Table 3. Tumours characteristics for phantom data set 2 with two types of voxels

Tumours	Isotropic Voxels		Non Isotropic Voxels	
	Position	Size	Position	Size
1	Slice 68	2 Voxels	Slice 142	2 Voxels
2	Slice 57	3 Voxels	Slice 119	3 Voxels
3	Slice 74	2 Voxels	Slice 155	2 Voxels

Table 4. CM for phantom data set 2 classes

Class1	Class2	Class3	Class4	Class5
5327	0	0	0	0
0	7895	0	0	0
0	0	971	0	0
0	0	0	2	1
0	0	0	0	2188

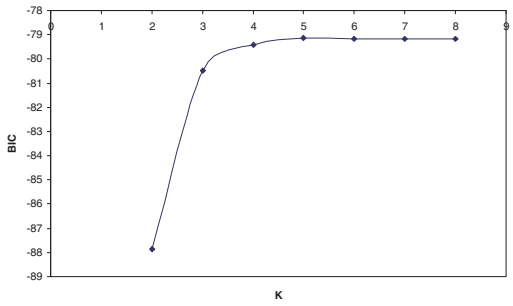

Fig. 4. Plot of BIC values for isotropic phantom data set 2, scaled by a factor of 1000.

The total number of voxels in this data set, Fig. 2.c, is 16384 voxels, which are divided into 5 classes; class 1 has 7799 voxels, class 2 has 5420 voxels, class 3 has 2518 voxels, class 4 has 504 voxels, and class 5 has 143 voxels. This number of classes matches the optimal class number obtained from BIC plot for clinical PET volume of non-small cell lung cancer patients which is 5 clusters. Fig. 5 shows BIC values against the class number K, where the graph starts to stabilise on an approximate curve plateau at value of 5. The ANN outputs which shows the segmented ROI is illustrated in Fig. 7.c. The CM for this clinical data set shows that 466 voxels are misclassified from class 1, 949 voxels are misclassified from class 2, 193 voxels from class 3, 162 voxels from class 4, and 64 voxels from class 5 are misclassified, as illustrated in Table 5. The fraction of samples misclassified in total was 0.111938. The AUC of ROC for all classes is illustrated in Table 2.

3.2.2. Clinical data set 2

The second clinical data set used in this study is PET volumes from seven patients with T3 - T4 laryngeal squamous cell carcinoma. Prior to treatment, each patient underwent an FDG-PET study. Patients were immobilized with a customized thermoplastic mask (Sinmed, Reeuwijk, the Nether-

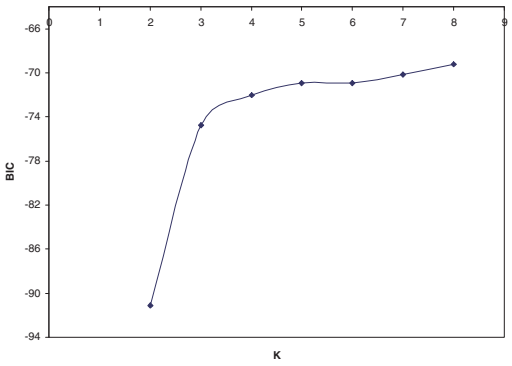

Fig. 5. Plot of BIC values for clinical PET volumes of non-small cell lung cancer patient, scaled by a factor of 1000.

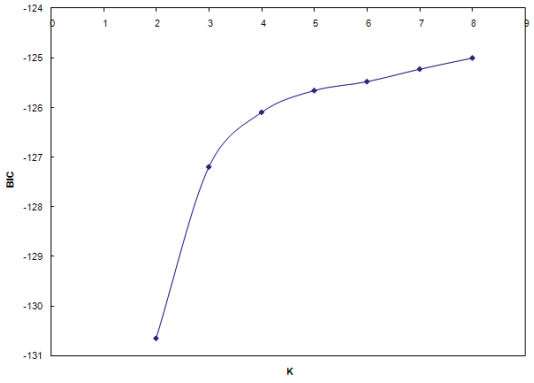

Fig. 6. Plot of BIC values for clinical PET data set 2 (patient 1), scaled by a factor of 1000.

Table 5. CM for clinical data set 1 classes

Class1	Class2	Class3	Class4	Class5
7333	466	0	0	0
857	4471	92	0	0
0	123	2325	69	1
0	0	69	342	93
0	0	2	62	79

lands) fixed to a flat table-top to prevent complex neck movements. First, a 10-min transmission scan was obtained on the Siemens Exact HR camera (CTI, Knoxville, USA). Immediately after intravenous injection of 185-370 MBq (5-10 mCi) of FDG, a 1-h dynamic 3D emission scan was performed. It consisted of eight frames with variable duration ranging from 90 to 600 s. All images were corrected for dead time, random, scatter, attenuation and decay and then reconstructed using a 3D OSEM algorithm, as used in the clinics for patients with head and neck tumours [24, 25, 26]. The size of this data set is $128 \times 128 \times 47$ for each patient.

The total number of voxels in each slice of this data set, Fig. 2.d, is 16384 voxels, which are divided into 5 classes; class 1 has 799 voxels, class 2 has 9353 voxels, class 3 has 5132 voxels, class 4 has 1012 voxels, and class 5 has 88 voxels. Fig. 6 shows BIC values against the class number K, where the graph starts to stabilise on an approximate curve plateau at value 5. The ANN outputs which shows the segmented tumour in clinical PET data set 2 (patient 1) is illustrated in Fig. 7.d, where the ROI is clearly segmented and quantified. The CM for this clinical data set shows that class 1,and 2 precisely classified, 1211 voxels are misclassified from class 3, 920 voxels from class 4, 34 voxels from class 5. The fraction of samples misclassified in total was 0.132141. The CM for this data set is shown in Table 6, while the AUC of ROC for all classes is illustrated in Table 2.

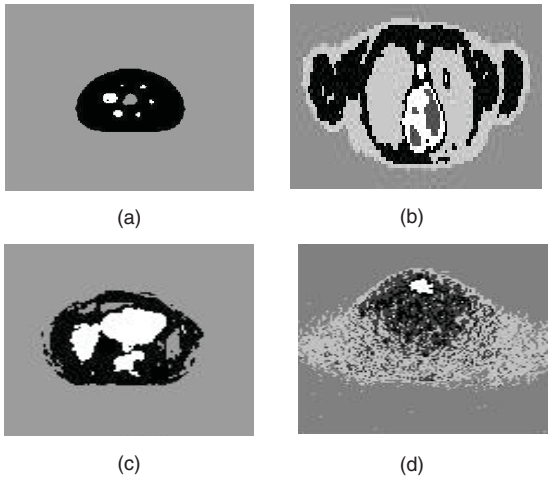

Fig. 7. Segmented slices: (a) phantom data set 1, (b) phantom data set 2, (c) clinical data set 1, (d) clinical data set 2 (patient 1).

4. CONCLUSIONS AND FUTURE WORK

An artificial intelligent approach based on ANN and BIC has been proposed for 3D oncological PET volume classification. Two PET phantom data sets, clinical PET volume of non-small cell lung cancer patient, and PET volumes from seven patients with laryngeal tumours have been used in this study to evaluate the performance of the proposed system. BIC approach has been deployed to obtain the optimal class number used by ANN to classify each slice in the processed PET volume. A detailed evaluation has been carried out on the system outputs, which has shown promising results. The performance evaluation has been carried out using confusion matrix and AUC of receiver operating characteristic curve. The ROI is precisely classified in all phantom data sets, and the application of different clinical data sets has also shown promising

Table 6. CM for clinical data set 2 classes

Class1	Class2	Class3	Class4	Class5
799	0	0	0	0
0	9353	0	0	0
0	1211	3921	0	0
0	0	920	92	0
0	10	0	24	54

results in detecting and classifying patient lesion. Ongoing research is focusing on the exploitation of other artificial intelligence and feature extraction approaches for 3D medical volume segmentation and analysis.

5. REFERENCES

[1] D. A. Mankoff, M. Muzi, and H. Zaidi, *Quantitative analysis in nuclear oncologic imaging*, Springer, 2006.

[2] S. Basu, "Selecting the optimal image segmentation strategy in the era of multitracer multimodality imaging: a critical step for imageguided radiation therapy," *Eur J Nucl Med Mol Imaging*, vol. 36, no. 2, pp. 180–181, 2009.

[3] K. Suzuki, I. Horiba, and N. Sugie, "Neural edge enhancer for supervised edge enhancement from noisy images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, pp. 1582–1596, December 2003.

[4] S. Joo, W. K. Moon, and H. C. Kim, "Computer-aided diagnosis of solid breast nodules on ultrasound with digital image processing and artificial neural network," *Proceedings of the 26th Annual International Conference of the IEEE EMBS*, 2004.

[5] S. G. Mougiakakou, I. Valavanis, K. S. Nikita, A. Nikita, and D. Kelekis, "Characterization of CT liver lesions based on texture features and a multiple neural network classification scheme," *Proceedings of the 25' Annual lnternational Conference of the IEEE EMBS*, 2003.

[6] W. E. Reddick, J. O. Glass, E. N. Cook, T. D. Elkin, and R. J. Deaton, "Automated segmentation and classification of multispectral magnetic resonance images of brain using artificial neural networks," *IEEE Transactions On Medical Imaging*, vol. 16, no. 6, pp. 911–918, December 1997.

[7] C. C. Reyes-Aldasoro and A.L. Aldeco, "Image segmentation and compression using neural networks," *Advances in Artificial Perception and Robotics CIMAT*, October 2000.

[8] K. Suzuki, H. Abe, H. MacMahon, and K. Doi, "Image-processing technique for suppressing ribs in chest radiographs by means of massive training artificial neural network (MTANN)," *IEEE Transactions on Medical Imaging*, vol. 25, no. 4, pp. 406–416, 2006.

[9] M. Hatt, C. Cheze le Rest, A. Turzo, C. Roux, and D. Visvikis, "A fuzzy locally adaptive bayesian segmentation approach for volume determination in PET," *IEEE Trans Med Imaging*, vol. 28, pp. 881–93, 2009.

[10] Y. Li and Z. Chi, "MR brain image segmentation based on self- organizing map network," *International Journal of Information Technology*, vol. 11, no. 8, 2005.

[11] G. F. Luger, *Artificial intelligence : structures and strategies for complex problem solving*, Pearson Education Inc., 2009.

[12] W. Duch and N. Jankovski, "Survey of neural transfer functions," *Neural Computing Surveys*, vol. 2, pp. 163–212, 1999.

[13] I. Kanellopoulos and G.G. Wilkinson, "Strategies and best practice for neural network image classification," *nternational Journal of Remote Sensing*, vol. 18, no. 4, pp. 711–725, 1997.

[14] C. Collet and F. Murtagh, "Multiband segmentation based on a hierarchical markov model," *Pattern Recognition*, in press.

[15] D. L. Weakliem, "A criteque of the bayesian information criterion for model selection," *Sociological Methods and Research*, vol. 27, no. 3, pp. 359–397, 1999.

[16] C. T. Volinsky and A. E. Raftery, "Bayesian information criterion for censored survival models," *Biometrics*, vol. 56, pp. 256–262, 2000.

[17] T. Fawcett, "ROC graphs: Notes and practical considerations for researchers," *Technical report, HP Laboratories, MS 1143, 1501 Page Mill Road, Palo Alto CA 94304, USA*, 2004.

[18] C. Jonsson, R. Odh, P.O. Schnell, and S.A. Larsson, "A comparison of the imaging properties of a 3- and 4-ring biograph pet scanner using a novel extended nema phantom," *IEEE Nuclear Science Symposium Conference*, vol. 4, pp. 2865 – 2867, 2007.

[19] M. D. R. Thomas, D. L. Bailey, and L. Livieratos, "A dual modality approach to quantitative quality control in emission tomography," *Physics in Medicine and Biology*, vol. 50, no. 15, pp. 187–194, 2005.

[20] H. Herzog, L. Tellmann, C. Hocke, U. Pietrzyk, M.E. Casey, and T. Kuwert, "Nema nu2-2001 guided performance evaluation of four siemens ecat pet scanners," *IEEE Transactions on Nuclear Science*, vol. 51, no. 5, pp. 2662–2669, 2004.

[21] H. Bergmann, G. Dobrozemsky, G. Minear, R. Nicoletti, and M. Samal, "An inter-laboratory comparison study of image quality of pet scanners using the nema nu 2-2001 procedure for assessment of image quality," *Physics in Medicine and Biology*, vol. 50, no. 10, pp. 2193–2207, 2005.

[22] J. M. Wilson and T. G. Turkington, "Multisphere phantom and analysis algorithm for pet image quality assessment," *Physics in Medicine and Biology*, vol. 53, no. 12, pp. 3267–3278, 2008.

[23] H. Zaidi, F. Schoenahl, and O. Ratib, "Geneva PET/CT facility: Design considerations and performance characteristics of two commercial (biograph 16/64) scanners," *Eur J Nucl Med Mol Imaging*, vol. 34, Suppl 2, S166, 2007.

[24] J. F. Daisne, M. Sibomana, A. Bol, G. Cosnard, M. Lonneux, and V. Gregoire, "Evaluation of a multimodality image (CT, MRI, and PET) coregistration procedure on phantom and head and neck cancer patients: accuracy, reproducibility and consistency," *Radiother Oncol*, vol. 69, pp. 237–245, 2003.

[25] J. F. Daisne, T. Duprez, B. Weynand, M. Lonneux, M. Hamoir, H. Reychler, and V. Gregoire, "Tumor volume in pharyngolaryngeal squamous cell carcinoma: Comparison between CT, MR imaging, and FDG PET and validation with surgical specimen," *Radiology*, vol. 233, pp. 93–100, 2004.

[26] X. Geets, J. A. Lee, A. Bol, M. Lonneux, and V. Gregoire, "A gradient-based method for segmenting FDG-PET images: methodology and validation," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 34, no. 9, pp. 1427–1438, 2007.