A FRAMEWORK FOR AUTOMATED TUMOR DETECTION IN THORACIC FDG PET IMAGES USING TEXTURE-BASED FEATURES

G.V. Saradhi[1], *G. Gopalakrishnan*[2], *A.S. Roy*[2], *R. Mullick*[2], *R. Manjeshwar*[3], *K. Thielemans*[4], *U. Patil*[5]

[1] Computing and Decision Sciences Lab, GE Global Research, Bangalore, India
[2] Imaging Technologies, GE Global Research, Bangalore, India
[3] Functional Imaging Lab, GE Global Research, Niskayuna, USA
[4] Hammersmith Imanet, London, UK
[5] Department of Radiology, Manipal Hospital, Bangalore, India

ABSTRACT

This paper proposes a novel framework for tumor detection in Positron Emission Tomography (PET) images. A set of 8 second-order texture features obtained from the gray level co-occurrence matrix (GLCM) across 26 offsets, together with uptake value was used to construct a feature vector at each voxel in the data. Volume of Interest (VOI) samples from 42 images (7 patients with 6 gates each), marked by a radiologist, representing 5 distinct anatomy types and pathology were used to train a logit boost classifier. A ten-fold cross-validation showed a true positive rate of 96%and a false positive rate of 8% for tumor classification. The test dataset consisted of $50 \times 50 \times 40$ representative VOIs from gated PET images of 3 patients. The classifier was run on the test data, followed by an SUV-based thresholding and elimination of noise using connected component analysis. The method detected $10/12$ (83%) tumors while detecting an average of 20 false positive structures.

Index Terms— Positron Emission Tomography (PET), tumor classification, gray-level co-occurrence matrix, texture, logit boost

1. INTRODUCTION

Tumor detection in PET images is useful in clinical applications such as staging and therapy planning, and medical image analysis techniques like constraint-based non-rigid registration [1]. There are multiple automatic tumor detection techniques that have been applied successfully to the CT, Ultrasound, MR and PET modalities. Detection of tumors on PET images is a challenging problem, due to its limited spatial resolution and low signal-to-noise ratio. The problem of tumor detection on PET images has been addressed in literature in various contexts. While some attempt a full-volume segmentation of a PET image, others address tumor ROI delineation. They also differ from each other in terms of the datasets on which the techniques are proposed. Kanakatte et al. [2] present a pilot study on automatic lung-tumor segmentation using standard uptake values (SUVs) making use of the high metabolization in the tumor. They demonstrate their results on 44 slices containing tumor and heart. A three stage approach consisting of preprocessing, segmentation and asymmetry detection was proposed by Chen et al. [3]. The authors use dynamic PET frames and compress them to four slices using principal component analysis (PCA). Further they use a graph-theoretic energy-minimization approach for segmentation, followed by an asymmetric feature detection to isolate pathological lesions in neuro images. Results are demonstrated on simulated and clinical datasets. Wong et al. [4] have used cluster analysis to replace manual ROI delineation in dynamic PET images. The authors use shape and magnitude of tissue Time Activity Curves (TACs) to classify them into a smaller number of distinct characteristic classes that are mutually exclusive so that the tissue TACs within a cluster are similar to one another but are dissimilar to those drawn from other clusters. They validate their technique on a simulated phantom data and assess its performance on a real dynamic PET data. Huang et al. [5] build statistical models on mean positron emission rate, raw sinogram data and the reconstructed image, and use them to derive the test criteria for maximum likelihood ratio test and a composite hypothesis test. They demonstrate detection of lesions (with a probability of 0.9) of size 15mm with lesion-to-background contrast $1.1 : 1$. Montgomery et al. [6] demonstrate automatic PET volume segmentation using Markov Random Field Models initialized by marginal segmentation, to characterize the spatial relationship between neighboring voxels. A spherical mean shift-based segmentation technique using a user defined seed point is proposed by [7] et al. The relationship between source-to-background ratio and the iso-activity level is used for segmentation of PET volumes by Daisne et al [8]. In a recent work, Naqa et al. [9] explore feature-based approach for predicting cancer treatment outcomes. They use intensity-volume histogram metrics and shape and texture features extracted from PET images, to predict patient

response to treatment.

While there has been a significant amount of work on PET volume segmentation, tumor ROI delineation and lesion segmentation on PET neuro-images, focus on tumor detection on PET images from thoracic region is limited. Previous work [9] has suggested that texture information present in PET images is a good encapsulator of the differences in visual perception of broadly iso-intense regions of high activity such as the heart and tumors. They also show a correlation between texture-based features and treatment outcome.

We present in this work, an automated method for PET tumor detection by developing a meta-classifier with gray level co-occurence matrix (GLCM) based second-order texture features as input, followed by SUV thresholding and connected component analysis. This framework (Figure 1) is a first step towards fully automated tumor detection and it would assist a clinician to identify the potential tumor regions, saving time to browse through all the slices of the volume. It will also be useful in other applications where a few number of false positives per case is acceptable, as in constraint-based non-rigid registration [1]. In section 2, the datasets used for our experiments and a brief theoretical background of various components of our framework is presented. A discussion on the experiments performed, a brief analysis on the classifiers we experimented on, and the results obtained therefrom is presented in section 3, followed by conclusions and future work.

Fig. 1. Overall framework for tumor detection

2. MATERIALS AND METHODS

Our database consists of 7 patient cases (covering the thoracic region) obtained from a GE DSTE PET/CT scanner. Each patient data was percentage-time binned into 6 gates (gated for respiration) using trace information from an optical tracking device. A clinical expert selected VOI samples in each of the 42 images: left lung, right lung, liver, tumor, heart and background. The clinician-marked samples included a total of 101 tumor regions with the size distribution of the tumors ranging between 0.5 and 6.0 cc (median size: 2.0 cc).

For each pixel location in a 3D image, there are 26 neighbors which are one-pixel away. In a given region of the image and for a given offset, a GLCM is obtained as a table with n rows and n columns, where n is the number of gray-levels for which the GLCM is calculated. In our analysis, we divided the gray-levels into 256 bins of equal size ($n = 256$). A cell $(i, j), 1 \leq i, j \leq n$, gives the number of times the gray-level j occurs as a neighbor (for a given offset) to the gray-level i. The texture calculations require a symmetrical matrix. A GLCM is made symmetric by counting each pixel pair twice or alternatively adding the transpose of the GLCM to the GLCM matrix. A normalized GLCM is then obtained by dividing each cell in the symmetric GLCM by the total number of possible co-occurrences. A normalized symmetric GLCM G of an image region R, for a given offset (further referred to as GLCM), is thus a matrix in which each cell $G(i, j)$ gives the probability of occurrence of the gray-levels i and j as neighbors (defined for an offset).

To extract a feature vector at a given pixel location, a VOI of size $5 \times 5 \times 3$ is extracted centered at that location. GLCMs are calculated for each of the 26 offsets at the location. Let G be a normalized GLCM. We define μ as the weighted pixel average defined as

$$\mu = \sum_{i,j} i.G(i, j) = \sum_{i,j} j.G(i, j)$$

due to matrix symmetry. σ^2, the weighted pixel variance is defined as

$$\sigma^2 = \sum_{i,j} (i - \mu)^2 G(i, j) = \sum_{i,j} (j - \mu)^2 G(i, j)$$

From each GLCM, the features below are extracted using the Insight Toolkit [10].

Energy: $f_1 = \sum_{i,j} G(i, j)^2$

Entropy: $f_2 = -\sum_{i,j} G(i, j) log_2 G(i, j)$

Correlation: $f_3 = \sum_{i,j} \frac{(i-\mu)(j-\mu)G(i,j)}{\sigma^2}$

Difference Moment: $f_4 = \sum_{i,j} \frac{1}{1+(i-j)^2} G(i, j)$

Inertia: $f_5 = \sum_{i,j} (i - j)^2 G(i, j)$

Cluster Shade: $f_6 = \sum_{i,j} ((i - \mu) + (j - \mu))^3 G(i, j)$

Cluster Prominence: $f_7 = \sum_{i,j} ((i-\mu)+(j-\mu))^4 G(i, j)$

Haralick's Correlation: $f_8 = \sum_{i,j} \frac{(i,j)G(i,j)-\mu_t^2}{\sigma_t^2}$, where μ_t and σ_t are the mean and standard deviation of the row (or column, due to symmetry) sums.

We use a logit-boost classifier proposed by Friedman et al. [11], where the authors describe how the performance of many classification algorithms can be significantly improved by sequentially applying them to reweighted versions of the input data, and taking a weighted majority vote of the sequence of classifiers, thereby produced. They show that boosting can be viewed as an approximation to additive modeling on the logistic scale using maximum Bernoulli likelihood as a criterion. This algorithm fits additive logistic

regression models by stage-wise optimization of the Bernoulli log-likelihood. We use a java-based implementation of the algorithm [12], for our experiments. This classifier is used for the purpose of discriminating various classes using GLCM feature vectors. We note that we use more classes for tumor detection refine the decision boundaries of the tumor class and thus reduce the false positive rate. Though the classifier is used to classify a given voxel into any of these 6 classes, we would ultimately perform a binary classification by treating the voxels labeled as lung (left and right), heart and liver as non-tumor voxels.

3. EXPERIMENTS AND RESULTS

VOIs of size $5 \times 5 \times 3$ voxels were selected for each anatomy class from various gates of 6 patients. In all, 36 heart regions, 35 regions for each of the left-lung and right-lung, 36 liver regions, 18 background regions and 101 tumor regions were selected as training samples. The aforementioned 209 features (8×26 texture based + 1 intensity feature) were calculated using the Insight Toolkit [10], for each of these training samples. The tumor classification system was trained based on this training set using a logit-boost classifier.

We explored the performance of a few classifiers like [13] like Naive Bayes, Bayesian Networks, J48 decision tree and logistic regression and logit boost. Our initial results interms of ROC area for the threshold curve (which measures the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one) showed that logit boost performs better than the others, as shown in Figure 2.

A cross-validation experiment was run to optimize the parameters in order to achieve the best possible precision and recall values for tumor. Decision stump, which is a one-level binary decision tree [13], was used as a base classifier. It was run for 10 iterations with 100 runs for internal cross-validation. The shrinkage parameter was set to 0.25 to avoid over-fitting. At these parameter settings, the confusion matrix for a 10-fold cross-validation experiment is as shown in Table-1. Out of 101 tumors present in the dataset, 97 were were correctly classified, and among the rest, 3 of them were misclassified as liver and 1 as heart. A true positive rate of 96%, false positive rate of 8%, precision of 88% and a recall of 96% for the tumor class, was observed.

For the purpose of testing, a 3-D region of size $50 \times 50 \times 40$ voxels centered at and covering the essential thoracic extent was extracted from a given test image. The GLCM features were extracted for each voxel in the test image, thus obtaining 100000 feature vectors. These feature vectors were run through the classifier and a binary mask was generated with voxels classified as tumor being set to 1 and the remaining voxels to 0. This binary image mask was overlayed on the original image (Figure 3).

We observe that the classifier successfully eliminates the interiors of heart and liver regions which are potential false positives for tumor detection due to their high uptake values. We note that borders of these regions still get classified as tumors, as shown in Figure 3; however, border voxels were found to have significantly lower uptake than the tumor voxels. This fact was exploited so as to identify compact regions of high uptake from the image post classification, by means of an SUV-based intensity threshold.

We motivate the selection of an intensity threshold by reference to the corresponding SUV: a conservative value of SUV = 1.2, based on an ROC-analysis on the datasets we have (acquired using same protocol and from same site), was used to generate this intensity value for each PET image. A connected component analysis was performed on the resulting thresholded image and regions of size lesser than 0.98 cc (10 voxels) were eliminated as noise. As a result, compact structures with high uptake value remain as candidate 3-D structures for tumor locations. The resulting image for one of our cases is shown Figure 4. We detected $10/12$ (83%) tumors while detecting an average of 20 false positive structures.

4. CONCLUSION AND FUTURE WORK

We presented a framework for an automatic detection of tumors in PET images. Among the tumors that were missed by our method were tumors with a very low uptake. The clinical expert was able to classify them correctly based on information from other gates. An extension of this method which incorporates information from multiple gates for each voxel may alleviate this effect. The use of the extracted texture features to deduce a tumor confidence score in regions where a physician suspects a tumor, is an area for further exploration. This will be of use in images with small tumors and also in tumors with low uptake. We would like to explore further on pre-processing steps for image enhancement such as the one used in [2], increase the size of our training set to ensure that over-fitting of data is avoided to the extent possible and work on post-processing techniques to decrease the number of false positives per case.

5. ACKNOWLEDGEMENTS

We wish to thank Valentino Bettinardi, Maria Carla Gilardi, Ferruccio Fazio from Hospitale San Raffaele, Milan for providing the clinical data, Jim Miller from GE Global Research for his valuable inputs, and Manisha Roy Choudhury from Nanyung Technological University, Singapore for her efforts on initial data analysis.

6. REFERENCES

[1] J. V. Miller, G. Gopalakrishnan, M. Datar, P. R. S. Mendonça, and R. Mullick, "Deformable registration with spatially varying degrees of freedom

constraints," in *Proceedings of the 2008 IEEE International Symposium on Biomedical Imaging*, 2008, pp. 1163–1166.

[2] A. Kanakatte, J. Gubbi, N. Mani, T. Kron, and D. Binns, "A pilot study of automatic lung tumor segmentation from positron emission tomography images using standard uptake values," in *Proceedings of the 2007 IEEE Symposium on Computational Intelligence in Image and Signal Processing*, 2007, pp. 363–368.

[3] Z. Chen, D. Feng, and W. Cai, "Pathological lesion detection in 3d dynamic pet images using asymmetry," in *Proceedings of Image Analysis and Processing*, 2003, pp. 295– 300.

[4] K-P. Wong, D. Feng, S. R. Meikle, and M. J. Fulham, "Segmentation of dynamic pet images using cluster analysis," in *IEEE transactions of nuclear medicine*, 2002, vol. 49, pp. 200– 207.

[5] C. C. Huang, X. Yu, J. Bading, and P. S. Conti, "Computer-aided lesion detection with statistical model-based features in pet images," in *IEEE transactions of nuclear medicine*, 1997, vol. 44, pp. 2509–2521.

[6] D. W. G. Montgomery, A. Amira, and H. Zaidia, "Fully automated segmentation of oncological pet volumes using a combined multiscale and statistical model," in *American Association of Physicists in Medicine.*, 2007, vol. 34, pp. 722– 736.

[7] T. B. Sebastian, R. M. Manjeshwar, T. J. Akhurst, and J V. Miller, "Objective pet lesion segmentation using a spherical mean shift algorithm," in *Medical Image Computing and Computer-Assisted Intervention*, 2006, vol. 4191, pp. 782–789.

[8] J-F. Daisne, M. Sibomana, A. Bol, T. Doumont, M. Lonneux, and V.Gregoire, "Tri-dimensional automatic segmentation of pet volumes based on measured source-to-background ratios: influence of reconstruction algorithms," in *Journal of the European Society for Therapeutic Radiology and Oncology.*, 2003, vol. 69, pp. 247–250.

[9] I. El. Naqa et al., "Exploring feature-based approaches in pet images for predicting cancer treatment outcomes," in *Pattern Recogntion*, 2008.

[10] L. Ibanez, W. Schroeder, L. Ng, and J. Cates, *The ITK Software Guide*, http://www.itk.org/ItkSoftwareGuide.pdf, second edition, 2005.

[11] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting," in *Stanford University*, 1998.

Heart	Liver	L-lung	R-lung	BG	Tumor	classified as
29	4	0	1	0	2	Heart
9	19	0	0	0	8	Liver
0	0	26	7	1	1	L-lung
1	0	18	15	0	2	R-lung
0	0	0	0	18	0	BG
1	3	0	0	0	97	tumor

Table 1. Results from a 10-fold cross-validation

[12] Ian H., *Data Mining: Practical machine learning tools and techniques, 2nd Edition*, 2005.

[13] Ethem Alpaydin, *Introduction to Machine Learning*, 2004.

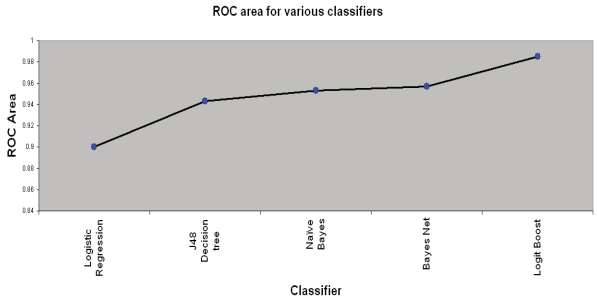

Fig. 2. ROC Area for various classifiers

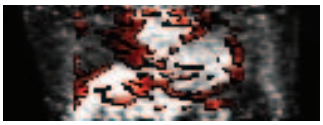

Fig. 3. Coronal view of PET image post GLCM classification; classified voxels are shown in red

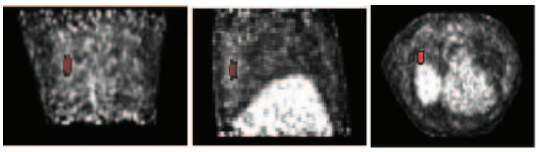

Fig. 4. Overlay of tumor (in red) detected by the method