Proxy Based Pre Cluster Model for Detecting Tumor and Lymph Nodes

Sneha.G[#1], Suguna.M[#2], Sharmila.D[#3]

[#1]PG Scholar
Department of Information Technology
SNS College of Technology, Coimbatore -641035
snehacce@gmail.com

[#2]Associate Professor
Department of Information Technology
SNS College of Technology, Coimbatore -641035
suguna.marappan@gmail.com

[#3]Professor & Head EIE Department,
Bannari Amman Institute of Technology,
Sathyamangalam.

ABSTRACT:

Lung cancer staging is important for the exploration of primary lung tumors and disease in regional lymph nodes and an automated system detects both types of abnormalities. Detecting both tumors and abnormal lymph nodes simultaneously from positron emission tomography – computed tomography (PET/CT) thoracic images will be supportive for clinical routine. An evidential segmentation scheme of PET/CT images is used for the detection of lung tumors. The modeling by means of evidence theory is well suited to the processing of redundant and complementary data as the PET/CT images. A cluster based approach is designed to identify all latent abnormalities, then differentiates between tumors and lymph nodes, and finally upgrade the detected tumors for false positive reduction. The method is intended to handle a wide and complex variety of abnormal patterns found in clinical datasets, consisting of different spatial contexts of tumors and abnormal lymph nodes. The proposed system is named as Trusted Pre Cluster count (TPCC) which is combined with the existing Multistage Discriminative model based on Support Vector Machines and conditional random fields, manipulating intensity, spatial and contextual features.

***Keywords*: PET/CT, lymph node, spatial, TPCC, discriminative, Mutistage Discriminative model.**

I.INTRODUCTION

Medical imaging is used to label the set of techniques that noninvasively produce images of the internal aspects of the body.Lung cancer is the most common cause of cancer related death in men and women. In particular, non-small cell lung cancer (NSCLC) is the most prevalent type of lung cancer, accounting for about 80% of all cases. Staging, which assesses the degree of spread of the cancer from its original source, is the most important factor affecting the projection and potential treatment of lung cancer.

Positron emission tomography—computed tomography (PET-CT) with F-fluoro-deoxy-glucose(FDG) tracer is now accepted as the best imaging technique for non-invasive staging. While the CT scan provides anatomical information, it has relatively low soft tissue contrast causing difficulties in separating abnormalities from the surrounding tissues. On the other hand, the PET

scan has high contrast and reveals increased metabolism in structures with promptly growing cancer cells, but their localization is limited by the low spatial resolution in PET images. The integrated PET and CT scan thus provides complementary pathological and anatomical information.

As an empirical data analysis tool, cluster analysis aims at grouping objects of a similar kind into their respective categories. There have been a large number of clustering algorithms reported in the literature. In general, clustering of unlabeled data poses three major problems: Evaluating cluster tendency to seek many clusters through the value of c, partitioning the data into c meaningful groups, validating the c clusters discovered.

The fortitude of the number of clusters 'c' is prior to clustering. Many clustering algorithms require the number of clusters c as an input parameter, so the quality of the resulting clusters is largely reliant on on the estimation of c. For some applications, users can define the number of clusters with domain knowledge. However, in many situations, the value of c is unknown and the needs are estimated from the data themselves.

The depiction of structure of an unlabeled data is in an image format and it has a long and continuous history. First the pairwise dissimilarity information about a dataset including n objects is obtained as an n X n image, where the objects are suitably reordered so that the subsequent image is enhanced to highlight the potential cluster structure in the data.

II.RELATED WORK

The visual representation of data dissimilarity includes the **"**Reordered Dissimilarity Image" (RDI). The intensity of each pixel in the RDI resembles the dissimilarity between the pair of objects addressed by the row and column of the pixel. A useful RDI highlights potential clusters as a set of "dark blocks" along the diagonal of the image, corresponding to set of objects with low dissimilarity.

A viewer can simply estimate the number of clusters c i.e., count the number of dark blocks along the diagonal of a RDI if these dark blocks retain visual clarity. However, as the boundaries between different clusters become less distinctive (e.g., significant overlap or unclear geometries), the RDI will degrade considerably, with non-distinct boundaries. Accordingly, viewers may presume different numbers of clusters from poor-quality images. This provides an objective way to estimate c from either good or, more probably, imperfect RDIs as automatically as possible.

Many Clustering algorithms require the number of cluster 'c' as an input parameter, so the quality of the resulting clusters is largely dependent on the estimation of 'c'. To overcome this problem, few existing algorithms are introduced in VAT to generate RDIs of unlabeled data, i.e., to provide secure inputs to the DBE algorithm.

Trusted Pre Cluster Count (TPCC) method combines several common thoracic images together. Then, sequential image processing operations like region segmentation, directional morphological filtering, and distance transformation are used to segment the regions of interest in the Radiographic digital image (RDI) and to convert the filtered image into a distance-transformed image. The transformed image is assigned to the diagonal axis of the RDI, which yields a one-dimensional signal, from which the potential number of clusters is extracted in the data set using sequential signal processing operations like average smoothing and peak detection.

III. SEGMENTATION AND CLASSIFICATION

A. Image Master

Image Master acts as the gateway for preview imaging. The main advantage of using this method is it allows the thoracic images to be stored in the centralized database with the sufficient details related to that scanned tiff images like image details, image taken time and date and so on.

B. Preview Imaging

The preview imaging will be very helpful to the user in order to preview a bunch of thoracic images stored in the centralized database already. It is mainly used to select the required image soon to allow it for segmentation and classification process.

C. Segmentation

Once the thoracic images are allowed for segmentation, two panels will be available for easy segmentation of the scanned images which is present in the

form of left hand side and the right hand side. In the left hand side of the panel, the selected image from preview imaging will be displayed.

The functionality embedded in the left hand side panel is to allow the user by segmenting the image in the form of placing mouse cursor point as scoring point. Once the user left click the mouse, the pointed area is calculated as x1 and y1 points and when the user release the mouse click, the released area is calculated as x2 and y2 points.

By using the calculation ratio, it initially analyzes the bounded region of the

selected portion which is present inside the (x1, y1) and (x2, y2) segments. Now the segmented image will be displayed in the right hand side panel and in the bottom of the right side panel, the visual zoom scaling will be placed for more clear visualization of

the segmented image to the user. In the segmented area, there is a button placed called comparative study, in which if the user clicks that button, the scanned segmented image in the right hand side panel will be allowed for comparison.

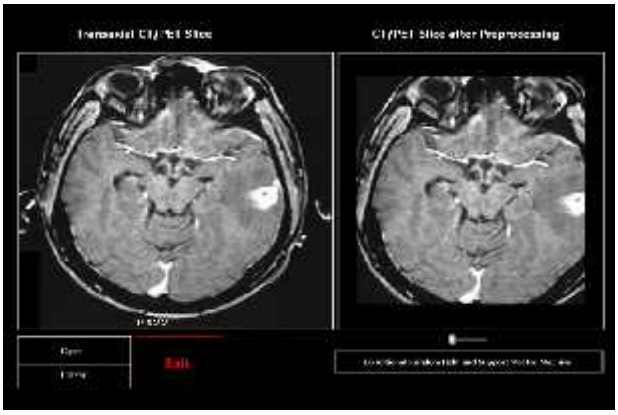

Fig 1: Transaxial CT/PET Slice after Preprocessing

D. Proxy Controller

Proxy controller is one of the main array tool which is allowed to store any particular data for time being transmission. The main advantage of using this kind of proxy controller is it will comparatively thousand times faster than a normal server, since the original server needs an external interface to store even a single data, but according to proxy, there is no need of interface since it stores all the data in a dynamic array.

IV. CONDITIONAL RANDOM FIELD FORMULATION

A CRF model integrated with SVM and a comprehensive set of features were designed to achieve an accurate discrimination between the two types of abnormalities (SVM), and minimize any misclassification by exploiting 3-D correlations (CRF). The use of CRF allowed us to incorporate the structural information in addition to the region based features, so that a 3-D ROI volume could be classified collectively. Individual regions created at the clustering step, here represented a large ROI region created by merging regions that were labeled as ROI and spatially connected in the same slice. A set of 3-D is connected (i.e., across slices) and then a 3-D ROI volume is formed, and could contain multiple ROI volumes, e.g., a primary lung tumor and several abnormal lymph nodes.

A CRF model consists of F = $<f_1, \ldots, f_k>$, a vector of "feature functions", $\theta = < \theta_1, \ldots, \theta_k>$, a vector of weights for each feature function, $O = < o_1, \ldots, o_T>$ be an observed sentence, $A = <a_1 \ldots a_T>$ be the lambent variables.

$$P(\mathbf{A} = \mathbf{y} \mid \mathbf{O}) = \frac{\exp(\boldsymbol{\theta} \cdot \mathbf{F}(\mathbf{y}, \mathbf{O}))}{\sum_{\mathbf{y}'} \exp(\boldsymbol{\theta} \cdot \mathbf{F}(\mathbf{y}', \mathbf{O}))}$$

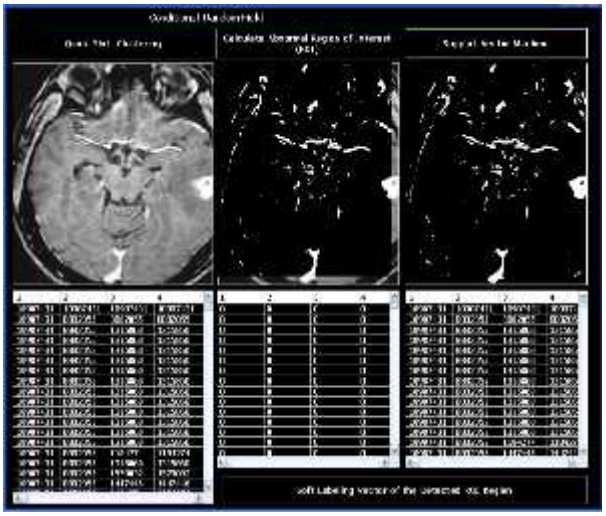

Fig 2: CRF Integrated With Support Vector Machine

Given a set of observations **o** and the correct labels **y** for each, determine the best θ:

$$\arg\max_{\theta} P(\mathbf{y} \mid \mathbf{o}, \boldsymbol{\theta})$$

The Standard Uptake Value is evaluated using the original Standard Uptake Value and the Normalized Standard Uptake Value. The soft labeling is done with the low intensity level and the high intensity level. The affected cluster count comparison is made between original and the normalized Standard Uptake Value.

Fig 3: Affected Cluster Count Comparison

Using these cluster counts the intensity level is calculated based upon the spatial and contextual features. The level of intensity is shown as low, high or equal. Here,

* - represents the intensity level as equal

0 - represents low level intensity

1 - represents high level intensity

The CRF equation is just a special form of the maximum entropy equation, it is trained exactly in the same way as determine the gradient, step in the direction of the gradient, repeat until convergence.

V. AFFECTED CLUSTER COUNT COMPARISON

The use of histogram allowed some spatial variations of the shape of the lung field, and the rigid subdivision of cells and channels helped to identify differences in shapes of lung fields. The shape of lung field was not used, since normally only a varying portion of the myocardium would display high SUVs, and would then exhibit a large variety of shapes. A binary SVM was then used to classify to either or, and the unary cost was computed from the probability estimates of the classifier.

The false positives were detected at the high uptake regions in the mediastinum, because of either reason: The region represented high-uptake myocardium or the tumor in the same image set exhibited quite low SUV and hence the high-uptake region showed similar SUV to the tumor. For case 1), the stage-3 is targeted to the detection of high uptake myocardium, for case 2), since the high-level feature worked based on the SUV contrast between ROIs and the mediastinum, if the contrast level was really low, such false positives could then remain.

VI. CONCLUSION

A fully automatic detection method for lung tumor and disease in regional lymph node from PET-CT thoracic images is engendered. Based on the low-level intensity and neighborhood features and high-level contrast-type features, abnormalities are first detected with a two-level SVM classification. The detected abnormalities are then differentiated into tumors or abnormal lymph nodes with a CRF model, based on the unary level contextual and spatial features and pairwise-level spatial features. Another CRF model is then activated to relabel the detected tumors as either true tumor or mediastinum by filtering the high-uptake myocardium areas. The detection recall and precision were measured for each stage, and the discriminative power of each feature set was also estimated. The high detection performance and capability in handling a wide variety of abnormal patterns are shown in SUV graph comparison.

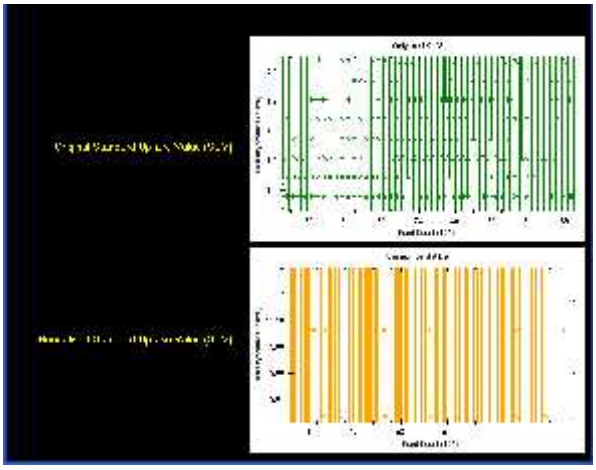

Fig 4: Graph Comparison

VII. FUTURE ENHANCEMENT

TPCC is more reliable method compared to other clustering methods. It should not be too hard to find an approximate center sample for each meaningful cluster from any well-structured RDI. Although DBE is not a clustering method, it may provide some advantageous information on object labels, especially for objects around the crowning in the projection signals. If such label information could be used, only the remaining boundary objects are clustered for reducing the amount of clustering data.

REFERENCES

[1] Yang Song*, Student Member, IEEE, WeidongCai, Member, IEEE, Jinman Kim, Member,IEEE, and David Dagan Feng, Fellow, IEEE "A Multistage Discriminative Model for Tumor and Lymph Node Detection in Thoracic Images" IEEE Transactions On Medical Imaging, Vol. 31, No. 5, May 2012 1061.

[2] R. J. Hathaway, J. C. Bezdek, and J. M. Huband, "Scalable visual assessment of cluster tendency for large data sets," Pattern Recogn., vol.39, pp. 1315–1324, 2006.

[3] "bigVAT: Visual assessment of cluster tendency for large data set," Pattern Recogn., vol. 38, no. 11, pp. 1875–1886, 2005.

[4] J. M. Huband, J. C. Bezdek, and R. J. Hathaway, "Revised visual assessment of (cluster) tendency (reVAT)," in Proc. North Amer. Fuzzy Inform. Process.Soc. (NAFIPS). Banff, Canada: IEEE Press, 2004, pp. 101–104.

[5] J. C. Bezdek and R. J. Hathaway, "VAT: A tool for visual assessment of (cluster) tendency," in Proc. Int. Joint Conf. Neural Networks, Honolulu, I, 2002, pp. 2225–2230.

[6] W. Wever, S. Stroobants, J. Coolen, and J. Verschakelen, "Integrated PET/CT in the staging of nonsmall cell lung cancer: Technical aspects and clinical integration," Eur. Respir. J., vol. 33, pp. 201–212, 2009.

[7] Y. Song, W. Cai, S. Eberl, M. Fulham, and D. Feng, "Discriminative pathological context detection in thoracic images based on multi-level inference," in MICCAI 2011, LNCS, 2011, vol. 6893, pp. 185–192.

[8] C. Cortes and V. Vapnik, "Support-vector networks," Mach. Learn., vol. 20, no. 3, pp. 273–297, 1995.

[9] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in Proc. ICML, 2001, pp. 282–289.

[10] Y. Song, W. Cai, S. Eberl, M. Fulham, and D. Feng, "Thoracic image case retrieval with spatial and contextual information," in Proc. ISBI, 2011, pp. 1885–1888.

[11] I. Jafar, H. Ying, A. Shields, and O. Muzik, "Computerized detection of lung tumors in PET/CT images," in Proc.EMBC, 2006, pp. 2320–2323