Classification of Confocal Endomicroscopy Patterns for Diagnosis of Lung Cancer

*************************************** ⋆

No Institute Given

Abstract. Confocal Laser Endomicroscopy (CLE) is an emerging imaging technique that allows the in-vivo acquisition of cell patterns of potentially malignant lesions. Such patterns could discriminate between inflammatory and neoplastic lesions and, thus, serve as a first in-vivo biopsy to discard cases that do not actually require a cell biopsy.

The goal of this work is to explore whether CLE images obtained during videobronchoscopy contain enough visual information to discriminate between benign and malign peripheral lesions for lung cancer diagnosis. To do so, we have performed a pilot comparative study with 12 patients (6 adenocarcinoma and 6 benign-inflammatory) using 2 different methods for CLE pattern analysis: visual analysis by 3 experts and a novel methodology that uses graph methods to find patterns in pre-trained feature spaces. Our preliminary results indicate that although visual analysis can only achieve a 60.2% of accuracy, the accuracy of the proposed unsupervised image pattern classification raises to 84.6%.

We conclude that CLE images visual information allow in-vivo detection of neoplastic lesions and graph structural analysis applied to deeplearning feature spaces can achieve competitive results.

1 Introduction

Incidence of Solitary Pulmonary Nodule (SPN) is increasing, but its diagnose can be challenging, especially when it is located in the periphery. Bronchoscopy is a minimally invasive procedure used in the diagnosis of central airways and lung tissue pathologies. Nowadays, new bronchoscopic procedures such as endobronchial ultrasound, ultrathin bronchoscopy guided by fluoroscopy or electromagnetic navigation, where a combination of endoscopic and radiological images are used, have increased bronchoscopy diagnostic yield.

Confocal laser endomicroscopy (CLE) is a new technique that can be combined with bronchoscopy to provide images of cells. CLE is a probe containing optic fibres that is introduced through the channel within the bronchoscope. Once CLE is in touch with the airways wall (proximal or distal airways) it will show microscopic images of lung tissue [1]. Up to now, CLE has been mainly used in gastrointestinal endoscopy to detect changes in mucosa and cellular patterns for in-vivo classification of neoplastic and inflammatory lesions in colon cancer

---
⋆

diagnosis [1–3]. Only in recent studies, CLE using fluorescence has been applied to the visualization of lung structures (alveolar ducts and bronchial microstructures) [4, 5]. In particular there are 2 works describing the visual characteristics of neoplastic cellular patterns observed in CLE.

In [6], the authors analyzed the patterns observed in CLE images from 12 patients with nodules in main airways. In that study, the visual patterns for 3 different types of cancerous lesions were described and compared to the visual appearance of normal mucosa. The main conclusion of that work is that the distribution and size of cell nuclei is more heterogeneous and sparse for malign nodules. The other study reports about cellular patterns of CLE with methylene blue dye taken from peripheral lesions of 3 patients [7]. In that work the authors also conclude that the technique might allow in-vivo assessment of solitary peripheral nodules during the bronchoscopic exploration. Although clinical studies indicate that CLE images of the lung could provide enough visual information to differentiate between malign and benign cellular patterns, as far as we known, there are no studies assessing CLE visual patterns in an quantitative systematic way. The goal of this work is to characterize cellular patterns in CLE images of peripheral lung nodules and assess their potential to discriminate between malign and benign lesions.

A main challenge in the development of image processing and classification methods for emerging medical imaging technologies is the limited amount of data available to design, train and test the decision support system. Such a condition has a negative impact in state-of-art classification methods that demand huge amounts of annotated data to achieve accurate and reliable results, such as Support Vector Machines (SVM) [8] and more recently deep learning algorithms [9].

Alternatives to supervised classification methods are either retrieval systems or unsupervised clustering algorithms. Both approaches rely on a similarity measure defined in the image feature space to either retrieve annotated cases similar to an unknown input image or group images presenting similar visual features. In fact, the only work addressing assessment of cellular patterns in CLE proposes a content-based image retrieval system based on Bag of Words and k-nearest neighbour (kNN) clustering[2]. Although, the use of such software restricts to classification of CLE patterns in gastrointestinal endoscopic procedures its 89.6% of accuracy proves the potential of unsupervised methods to achieve clinically meaningful conclusions in case of a low number of samples.

We present a novel approach to find patterns in visual feature spaces by exploring the space topological structure. In order to extract topology, proximity relationships among a set of discrete samples is codified using a graph representation. The open sets defining the topology correspond to subgraphs fulfilling some connectivity criterion. To compute such subgraphs, we use community detection methods from social network analysis. Finally, the profile computed using the annotated samples, belonging to each community that a CLE image belongs to, defines the diagnose of non-annotated cases. The preliminary results obtained for

171 images with malign and benign patterns achieve a 84.6% of overall accuracy and a 81.1% of recall in cancer detection.

2 Classification of Confocal Patterns

Our approach consists in constructing a graph representation of CLE images based on visual features and try to discover potentially overlapping groups of images that share common properties. We refer to such groups as communities borrowing the terminology from the social network community. The rational of this approach is the following. Given a network of images (users) connected by means of visual appearance, we apply community detection algorithms to infer group of images that share similar properties. The annotations (clinical data in our case) made on the community images allow the definition of a community profile that can be assigned to images in the community without annotations. Since a CLE image can belong to more than one community we have to merge the community profiles it belongs to. In this section we describe all the main steps of our method: construction of the graph codifying feature space connectivity, graph community detection representing space topology and definition of profile (diagnosis in our case) for non-annotated images (see Fig. 1).
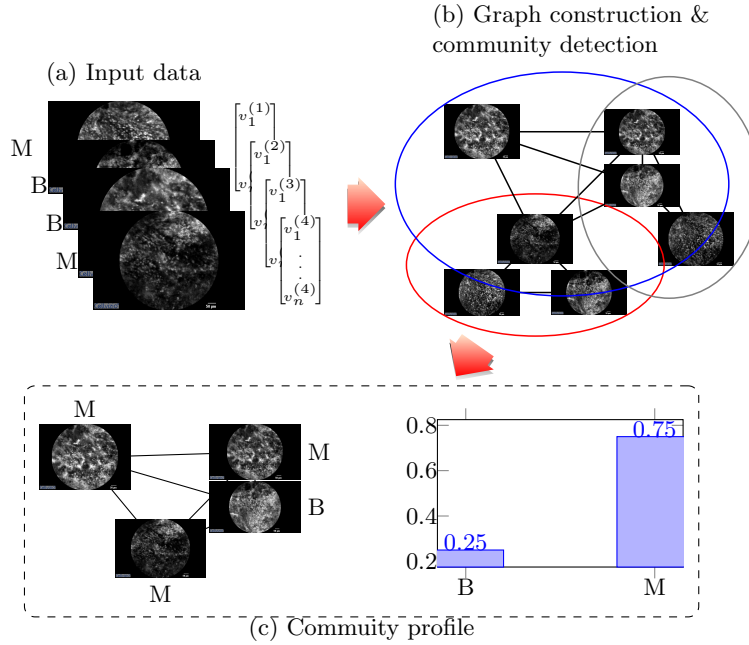

Fig. 1. Construction of CLE image clusters: (a) original images and feature vectors, (b) graph construction and community detection and (c) community profile.

The topology of a space is given by a collection of open subsets that cover the whole space [10]. Such collection defines, for each point of the space, a set of neighbouring points given by all points included in the subsets the point belongs to. Conversely, the set of neighbours of each point defines a topology in the space. Although unsupervised clustering is also based on neighbourhoods, a main difference with a topological approach is the criterion used to compute the open sets. In most unsupervised clustering such neighbourhoods are defined using a metric on the Euclidean space given by feature vectors. In a metric space, the collection of open sets are "balls" (points at equal distance to a given point). Since all open sets have equal dimension and shape, the resulting topology is very rigid in the sense that might have limited capability to describe clusters with complex structures of different dimensions.

The first step to obtain a topology is to represent the proximity among feature points. In the discrete domain, the spatial proximity of a set of points (images), sampled in a feature space, can be represented using an undirected graph, being the nodes the feature points representing images and being the edges defined according to similarity measure, or *friendship*, among neighbouring images. In the context of social networks, most applications (like Facebook or Tinder) grant connection between two users in case that they have mutually accepted a friendship request. An equivalent criterion on a image network is to use the mutual-kNN to define the connectivity of the graph [11]. The mutual-kNN criterion connects two images, I_i and I_j, if their feature points, $x_i \in \mathbb{R}^n$ and $x_j \in \mathbb{R}^n$, both of them are among the k-th nearest neighbours of the other. That is, two images are connected if x_i is ranked on the top k nearest neighbours of x_j and reciprocally. We use the Euclidean distance in $\mathbb{R}^n$ to compute nearest neighbours.

In order to obtain the groups of neighbouring points that define the topology, we will use methods for community detection in social networks. Such methods group users (images in our case) according to the adjacency of the graph representing the network. A main difference with other clustering techniques (like k-means) is that social community detection methods allow group overlapping. So that nodes (users) can belong to more than one community, characterized by each community metadata.

To detect CLE image communities we use the intrinsic longitudinal community detection (iLCD) algorithm, which analyses social network communities along time [12]. In that work, the authors define the smallest community that they want to detect (a clique of order 3 or 4) and a set of rules that detect new communities, merge them or even destroy them. A clique of order c is a fully connected subgraph with c nodes. A subgraph is said to be fully connected if each node is connected to all the other nodes belonging to the clique.

A clique represents a simplex and it is a continuous mapping (homeomorphism) into a topological space of a polyhedron defining a reference in Euclidean space [10]. A simplex defines an open set and thus the collection of all simplex define a topology. It follows that iLCD communities are giving the open sets of

a topology defined by simplicial complexes, so we can describe richer geometries of our feature space.

The last step of our method is to take advantage of the above geometric representation to compute CLE image profiles for clinical diagnosis purposes. The profile of a non-annotated image is given by the fusion of the community profiles, in which it belongs to. Given that in this work we face a 2-class problem, we use the argmax criterion on the diagnosed: malign or benign cases, and therefore assign the most frequent label to the new image.

3 Experiments

3.1 Experimental Set-up

A total of 12 patients (cases) from Hospital Universitari de Bellvitge with SPN or lung mass referred for bronchoscopy were included in the study. In the end, 6 patients had adenocarcinoma and 6 patients showed benign inflammatory changes. CLE images were obtained using a flexible bronchoscope BF-160, Olympus and a Alveloflex-Cellvizio (Manua Kea Technologies) 660nm miniprobe for acquisition of CLE images. Virtual bronchoscopy using Lung Point was used to plan the path to the lesions. Methylene blue dye was used to enhance cellular pattern.

For each case, 2-6 video sequences lasting between 1-2.5 minutes each were acquired and processed with Cellvizio Viewer to enhance the blue channel. Video sequences were visually explored to select the between 10 and 15 images presenting a clear cellular pattern. In total, we collected 162 images, 78 from adenocarcinoma and 79 from inflammation. Fig. 2 shows 3 images from each group. Cells nuclei correspond to brilliant spots in images. We note that for benign-inflammatory images, their distribution and size is homogeneous. Meanwhile malignant patterns are characterized by a clustering of cells that results in an overall heterogeneous appearance of images which have areas of low fluorescence.

Although any feature extraction technique can be applied, we have chosen a visual descriptor coming from deep learning architectures since they are proving to outperform handcraft descriptors in many computer vision applications. We used the VGG Convolutional Neural Network (CNN), which was trained on the ImageNet 1,000 object categories [13]. This CNN computes a global image descriptor obtained as the concatenation of the activation values of the last CNN convolution layer, and just before the fully connected layers which classify the 1,000 objects. This layer gives us a 4096-dimensional real-valued vector that we reduce using principal component analysis to 100 dimensions, which roughly represent the 90% of the energy the VGG vector.

We have compared our method to k-means and to kNN to assess the performance of the propose method in different aspects. The accuracy of k-means will indicate the discriminative power of VGG features, while comparison to kNN will show the differences between a description of clusters based on topology like ours and a standard metric approach. We assign to each cluster the most frequent label of the images belonging to it. We compare the value of such label to

each image diagnosis to compute recall each class (malign, benign) and overall classification accuracy.

Concerning algorithm parameter settings, we applied all methods using different parameter configurations and we selected the best performer for comparisons. The numbers reported in Section 3.2 were obtained with graphs constructed using $k = 10$, $k = 18$ clusters for k-means and $k = 20$ for kNN.

Finally, images were shown to 3 expert observers who ignore the histopathological diagnosis for a blind visual labelling. Recall for each class and overall accuracy were computed for each expert. Aside, 2 of the observers were asked to provide a diagnosis for each patient based on the visual inspection of the whole set of images to assess accuracy in their final diagnosis.
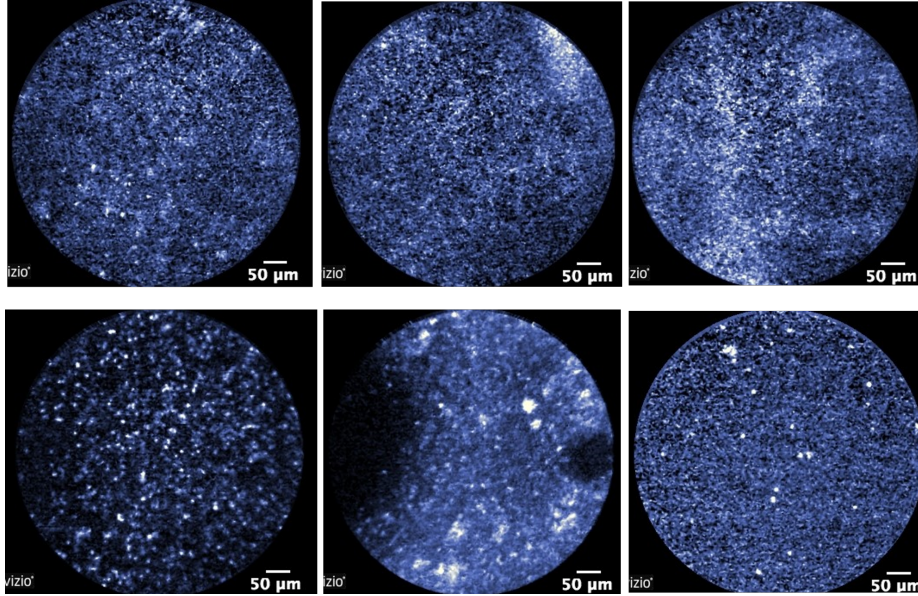

Fig. 2. Examples of CLE images for benign-inflammatory (top images) and malign-adenocarcinoma (bottom images) cases.

3.2 Results

The overall average accuracy in classification achieved by visual inspection of images was only 60.2% and presented a large variability among experts and cases. For malign images, recall was 72.8% in average with values per each observer equal to 84.2% for observer 1, 63.2% for observer 2 and 71.1% for observer 3. Benign images had even more variability with recall equal to 36.1% for observer 1, 51.2% for observer 2 and 60.5% for observer 3 and an average recall 49.2%. Results for visual classification are summarized for each case in Table 1 with benign cases indicated with (B) and malign ones with (M). Concerning accuracy

	Case 1 (M)	Case 2 (B)	Case 3 (M)	Case 4 (B)	Case 5 (M)	Case 6 (B)	Case 7 (B)	Case 8 (M)	Case 9 (M)	Case 10 (M)	Case 11 (B)	Case 12 (B)
Obs. 1	7/7	2/15	9/15	5/15	13/13	6/14	4/12	12/12	14/14	9/15	7/15	7/15
Obs. 2	6/7	11/15	4/15	7/15	13/13	7/14	3/12	10/12	13/14	2/15	2/15	14/15
Obs. 3	5/7	12/15	5/15	6/15	11/13	9/14	4/12	10/12	14/14	9/15	9/15	12/15
AA	85.7%	55.5%	40.0%	40.0%	94.8%	52.3%	30.5%	88.8%	97.6%	44.4%	40.0%	73.3%

Table 1. Analysis of CLE images done by 3 observers.

in final diagnosis, both experts issued the right diagnosis in 7/12 cases (66.6%), made a bad decision in 3/12 and did not agree in 2/12.

The overall accuracy for the proposed method (labelled GraphCom from now on) is 84.6%, 81.3% for kNN and 82.4% for k-means. Recall for malign images is 88.1% for GraphCom, 84.5% for kNN and 78.6% for k-means, while for benign cases recall is 81.2% for GraphCom, 78.2% for kNN and 86.2% for k-means. Results for each case reported in Table 2 show a lower variability than visual assessment. It is worth noticing that in case that final diagnosis was given by the most frequent label in each case, GraphCom would issue the right diagnosis in 12/12 cases (100%), kNN in 11/12 (91.6%) and k-means in 10/12 cases (83.3%). For both, kNN and k-means the wrong diagnosis were for malignant cases, which implies cancer detection rate equal to 83.3% for kNN and 66.6% for k-means.

	Case 1 (M)	Case 2 (B)	Case 3 (M)	Case 4 (B)	Case 5 (M)	Case 6 (B)	Case 7 (B)	Case 8 (M)	Case 9 (M)	Case 10 (M)	Case 11 (B)	Case 12 (B)
GraphCom	7/7	14/15	9/15	14/15	11/13	13/14	12/12	7/12	11/14	12/15	15/15	10/15
kNN	7/7	13/15	10/15	12/15	11/13	14/14	12/12	7/12	7/14	12/15	15/15	10/15
k-means	7/7	12/15	14/15	12/15	10/13	14/14	12/12	5/12	7/14	14/15	15/15	11/15

Table 2. Analysis of CLE images done by the 3 unsupervised methods.

4 Conclusions

The goal of this work was to determine whether CLE with methylene blue could potentially be useful to assess peripheral SPN. To do so, a pilot study with 12 patients (6 adenocarcinoma, 6 benign) and 171 images was conducted. Two main conclusions can be derived from our pilot study.

A first disappointing conclusion is that visual analysis by several individuals of CLE images looking for differences between neoplastic and inflammatory diseases showed an important inter and intra observer bias. In addition, identifying differences between malignancy and non-malignancy by visual inspection can be challenging given the similarities between both confocal patterns. Looking at our results, visual inspection is not enough to give a diagnosis of malignancy versus benign pathologies base on CLE images, with only 60% and 66% average accuracy between 2 and 3 observers respectively.

The good news is that image analysis can obtain enough information from CLE to discriminate between inflammatory and cancerous patterns using non-supervised techniques. The discriminative power of VGG features (given by k-

means clustering) is quite high with a 82.4% of overall accuracy. However recall for malign cases drops to 78.6% and accuracy in final diagnosis is 83.3% overall but only a 66.6 in case of adenocarcinoma. In this paper, we have proposed to use a graph representation of the feature space combined with community detection algorithms. Our method provides a more flexible description of the feature space topology that increases accuracy in malign patterns to 88.1% and reaches a 100% in final diagnosis.

In summary, our preliminary results indicate that CLE images contain enough visual information for in-vivo detection of neoplastic lesions that could be extracted using graph structural analysis. The promising results obtained in this pilot study, encourage further research on community detection and graph structural methods as tools to design diagnosis support systems able to reach clinically meaningful conclusions with scarce annotated data.

References

1. J. M. Jabbour, M. A. Saldua, J. N. Bixler, and K. C. Maitland, "Confocal endomicroscopy: Instrumentation and medical applications," *Ann Biomed Eng.*, vol. 40(2), pp. 378–97, 2012.
2. B. André, T. Vercauteren, A. M. Buchner, and et al, "Software for automated classification of probe-based confocal laser endomicroscopy videos of colorectal polyps," *World J Gastroenterology*, vol. 18(39), pp. 5560–5569, 2012.
3. S. Sanduleanu, A. Driessen, E. Gomez-Garcia, W. Hameeteman, A. de Brune, and A. Masclee, "In vivo diagnosis and classification of colorectal neoplasia by chromoendoscopy guided confocal laser endomicroscopy," *Clin Gastroenterol Hepatol*, vol. 8, pp. 371–8, 2010.
4. L. Thiberville, M. Salaun, S. Lachkar, S. Dominique, S. Moreno-Swirc, C. Vever-Bizet, and G. Bourg-Heckly, "Human in vivo fluorescence microimaging of the alveolar ducts and sacs during bronchoscopy," *Eur Respir J*, vol. 33, pp. 974–85, 2009.
5. F. Fuchs, S. Zirlik, and K. Hildner, "Fluorescein-aided confocal laser endomicroscopy of the lung," *Respiration*, vol. 81, pp. 32–8, 2010.
6. F. Fuchs, S. Zirlik, K. Hildner, J. Schubert, M. Vieth, and M. Neurath, "Confocal laser endomicroscopy for diagnosing lung cancer in vivo," *Eur Respir J*, vol. 41, pp. 1401–8, 2013.
7. T. Hassan, N. Piton, S. Lachkar, M. Salan, and L. Thiberville, "A novel method for in vivo imaging of solitary lung nodules using navigational bronchoscopy and confocal laser microendoscopy," *Lung*, vol. 193(5), pp. 773–8, 2015.
8. C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Min. Knowl. Discov.*, vol. 2, no. 2, pp. 121–167, 1998.
9. Y. Bengio, "Learning deep architectures for ai," *Found. Trends Mach. Learn.*, vol. 2, no. 1, pp. 1–127, 2009.
10. M. J. Greenberg and J. R. Harper, *Algebraic Topology: A First Course.* Addison-Wesley, 1981.
11. Q. Luo, S. Zhang, T. Huang, W. Gao, and Q. Tian, "Superimage: Packing semantic-relevant images for indexing and retrieval," in *ICMR*, 2014.
12. R. Cazabet, F. Amblard, and C. Hanachi, "Detection of overlapping communities in dynamical social networks," in *ICSC*, pp. 309–314, 2010.

13. K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," in *BMVC*, 2014.