Textured Graph-model of the Lungs for Tuberculosis Type Classification and Drug Resistance Prediction: Participation in ImageCLEF 2017

Yashin Dicente Cid[1,2], Kayhan Batmanghelich[3], and Henning Müller[1,2]

[1] University of Applied Sciences Western Switzerland (HES–SO), Sierre, Switzerland;
[2] University of Geneva, Switzerland;
[3] University of Pittsburgh, USA
yashin.dicente@hevs.ch

Abstract. In 2017, the ImageCLEF benchmark proposed a task based on CT (Computed Tomography) images of patients with tuberculosis (TB). This task was divided into two subtasks: multi-drug resistance prediction, and TB type detection. In this work we present a graph-model of the lungs capable of characterizing TB patients with different lung problems. This graph contains a fixed number of nodes with weighted edges based on distance measures between texture descriptors computed on the nodes. This model attempts to encode the texture distribution along the lungs, making it suitable for describing patients with different tuberculosis types. The results show the strength of the technique, leading to best results in the competition for multi-drug resistance (AUC = 0.5825) and good results in the tuberculosis type detection (Cohen's Kappa coef. = 0.1623), with many of the good runs being fairly close.

Keywords: lungs graph-model, 3D texture analysis, tuberculosis

1 Introduction

ImageCLEF (the image retrieval and analysis evaluation campaign of the Cross-Language Evaluation Forum, CLEF) has organized challenges on image classification and retrieval since 2003 [1]. Since 2004, a medical image analysis and retrieval task has been organized [2, 3].

The ImageCLEF 2017 [4] challenge included a task based on tuberculosis CT (Computed Tomography) volumes, the ImageCLEF 2017 tuberculosis task [5]. In this task, a dataset of lung CT scans was provided and 2 subtasks were proposed. When tuberculosis affects the lungs, several visual patterns can be seen in a CT image. However, the final diagnosis usually required other analyses than only the images [6]. A preliminary visualization of the CT volumes available in the ImageCLEF task showed lighter regions forming patterns that could be characterized with an holistic description of the lung. In [7], a graph-model of the lungs capable to differentiate between pulmonary hypertension and pulmonary

embolism patients is presented. Both diseases present somewhat similar visual defects in lung CT scans, with different shapes and distributions. The graph was based on dividing the lung into several regions and use these as nodes of a graph. The edges of the graph encoded the difference between HU distributions in the lung regions. However, Dual Energy CT (DECT) scans were used in the study and the HU distribution can be described with more detail than in a standard single energy CT. Preliminary results showed that only one energy level did not contain enough information about the HU distribution to differentiate between the pulmonary hypertension and embolism patients.

Following the same approach, more complex features are used in this work. The descriptors based on the HU distribution were replaced by 3D texture features. Moreover, a deep analysis of the edges of the graph are helpful to describe the tuberculosis patterns. Our hypothesis is that a holistic analysis of the relations between regional texture features is able to encode subtle differences between patients with different tuberculosis type and drug resistance.

The following section contains a brief overview of the subtasks and dataset of the ImageCLEF 2017 tuberculosis task. More detailed information on the task can be found in the overview article [5]. Section 3 explains the process of building the textural graph-model of the lungs and all the variations tested for this task in detail. The results obtained by this approach in both subtasks are shown in Section 4. Finally, Section 5 concludes our participation in this challenge.

2 Subtasks and Datasets

The ImageCLEF 2017 TB task proposed two subtasks: i) Multi-drug resistance (MDR) prediction, and ii) Tuberculosis type (TBT) detection. The MDR task is a 2-class problem and the TBT task contains 5 classes. For both subtasks volumetric chest CT images with different voxel sizes and automatic segmentations of the lungs were provided. No other lung segmentation was attempted in this work and the masks provided were used. These masks were obtained with the method described in [8].

The competition was divided into two phases. In the first phase, the organizers released for each subtask a set of patient CT volumes as training set with their lung masks and ground truth labels. In the second phase, the test set with its lung segmentations but no labels were provided. The evaluation on the test data was performed by the organizers after the scheduled deadline for all runs that were submitted in time. The number of CT volumes for each task and set are specified in Tables 1 and 2.

3 Methods

This section details the process for the feature vector extraction from the CT images provided by the ImageCLEF TB task. The same technique was applied for describing the patients of both subtasks. This technique consists of creating graph-models of the lungs, with nodes based on a geometrical atlas and

Table 1. Dataset for the multi-drug resistance task. DS means drug-sensible.

Patient set	Train	Test
DS	134	101
MDR	96	113
Total patients	230	214

Table 2. Dataset for the tuberculosis type detection.

Patient set	Train	Test
Type 1	140	80
Type 2	120	70
Type 3	100	60
Type 4	80	50
Type 5	60	40
Total patients	500	300

weighted edges encoding dissimilarities between 3D texture descriptors of each atlas region.

3.1 Isometric Volumes

The approach is based on 3D texture features. These features require having isometric voxels. The first step of our approach is to make the 3D images and masks provided by the organizers isometric. After analyzing the multiple resolutions found in the dataset and inter-slice distances, we opted for a voxel size of 1 mm to capture a maximum of information.

3.2 Atlas of the Lung

To build a graph with a fixed structure over the lung physiology we first need a localization system. In this case we chose the atlas developed by Depeursinge et al. in [9]. This atlas is only based on the mask of the lungs and provide 36 geometric regions dividing the lungs as shown in Figure 1. It is not based on the lung lobes.

3.3 3D Texture Features

Two state-of-the-art 3D texture features were selected to describe the texture inside the lung. The first method is a histogram of gradients based on the Fourier transform HOG (FHOG) introduced in [10]. 28 3D directions are used for the histogram obtaining a 28-dimensional feature vector per image voxel ($\mathbf{f}_H \in \mathbb{R}^{28}$).

The second approach is the locally-oriented 3D Riesz-wavelet transform introduced by Dicente et al. in [11]. The parameters that obtained the best results in this article were used in the approach. These are: 3rd-order Riesz transform, 4 scales and 1st-order alignment. This configuration results in 40-dimensional

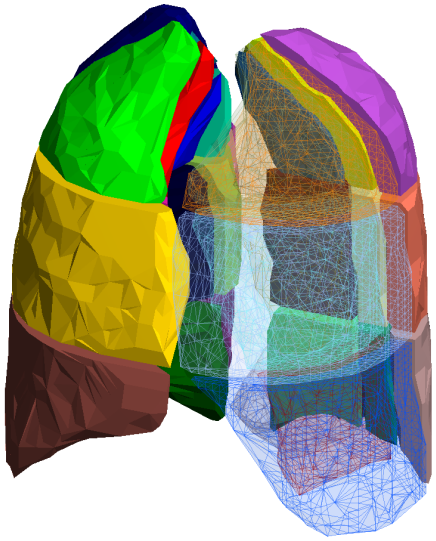


Fig. 1. Visualization of the 36-region atlas of the lungs.

feature vectors for each image voxel. The feature vector for a single voxel is then 10-dimensional containing the energy of each filter along the 4 scales ($\mathbf{f}_{\mathcal{R}} \in \mathbb{R}^{10}$).

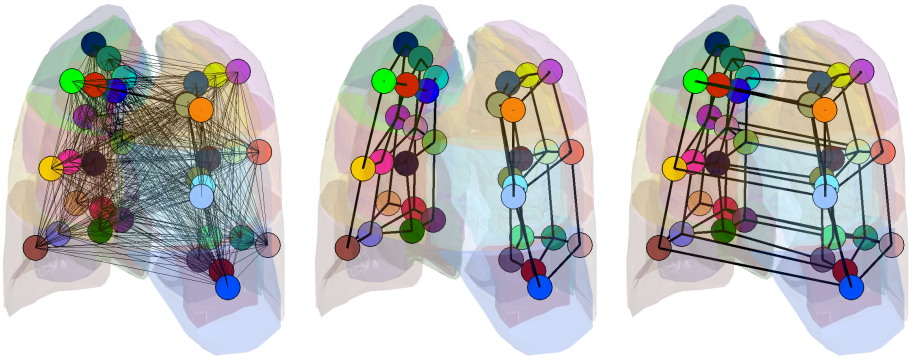
3.4 Textural Graph-Model of the Lungs

Nodes Using as a base the 36-region atlas, the texture information from the 3D textural features was embedded in a 36-node graph. With this fixed number of nodes, we defined several graphs varying the number of edges and their weights. The following notation is used: given a patient $p \in P$ and its 36-region atlas A_p of the lungs with regions $\{r_1, \ldots, r_{36}\} \in A_p$ we let G_p be the 36-node graph of the patient p with nodes $\{N_1, \ldots, N_{36}\}$. Each node N_a represents one region $r_a \in A_p$.

Edges Three graphs were defined.

- *Graph_Full*: This is the fully connected 36-node graph. For every pair of nodes N_a and N_b with $a \neq b$ there exists an undirected edge $E_{a,b}$. The total number of edges in this case is 630 ($\frac{36 \cdot 35}{2}$).
- *Graph_66*: Based on the region adjacency defined by the atlas A_p, in this case, there exists edge $E_{a,b}$ between nodes N_a and N_b if regions r_a and r_b are 3D adjacent in the atlas. This graph contains 66 edges in total.
- *Graph_84*: The last graph has the same 66 edges as *Graph_66*. Moreover, it has 18 additional edges connecting each pair of nodes representing opposite regions inside the atlas.

The three graphs are shown in Figure 2.



Graph_Full *Graph_66* *Graph_84*

Fig. 2. Comparison of the three graphs used in our approach.

Weights The graphs were always undirected weighted graphs. The weight $w_{a,b}$ of an edge $E_{a,b}$ was defined in several ways based on the relations between the features of the corresponding nodes N_a and N_b. Four measures were used to compute the weights. Considering $\mathbf{f}_a$ and $\mathbf{f}_b$ the feature vectors of regions r_a and r_b respectively, the measures used are:

- *Correlation (corr)*: $w_{a,b} = corr(\mathbf{f}_a, \mathbf{f}_b)$
- *Cosine similarity (cos)*: $w_{a,b} = cos(\mathbf{f}_a, \mathbf{f}_b)$
- *Euclidean distance (euc)*: $w_{a,b} = ||\mathbf{f}_a - \mathbf{f}_b||_2$
- *Norm of the sum (sumNorm)*: $w_{a,b} = ||\mathbf{f}_a + \mathbf{f}_b||_2$

Feature Vector of a Region Several feature vectors can be extracted from a region r. Section 3.3 introduced the features extracted for a single voxel, $\mathbf{f}_H$ and $\mathbf{f}_\mathcal{R}$. Given a region r_a, we extracted the mean (μ_a) and standard deviation (σ_a) of the features inside the region, i.e.: $\mu_a(\mathbf{f}_H)$, $\sigma_a(\mathbf{f}_H)$, $\mu_a(\mathbf{f}_\mathcal{R})$, and $\sigma_a(\mathbf{f}_\mathcal{R})$.

Feature Vector of a Patient Finally, the feature vector $\boldsymbol{w}_p$ of a patient p, is defined as the ordered concatenation of the weights $w_{a,b} \in G_p$. Depending on the graph used, this feature vector can be 630-, 66-, or 84-dimensional.

Feature Normalization The last step before using the feature vectors $\boldsymbol{w}_p$ with the classifier is to normalize them along the set of training patients P_{TRN}. Each component $w_{p,i}$ of a vector $\boldsymbol{w}_p$ corresponds to the weight of a different edge in the graph. Then, the components of a feature vector can not be seen independently and the normalization is required to keep these relations. Thus, the normalization was done for all components simultaneously. Several normalizations were tested: The first is linearly resizing the min and max values among all components between 0 and 1, i.e.:

$$[\min_{p \in P_{TRN}} \{\boldsymbol{w}_p\}, \max_{p \in P_{TRN}} \{\boldsymbol{w}_p\}] \rightarrow [0, 1]$$

. The second normalization aims to remove outliers. It considers the feature vector components as elements of a Gaussian distribution, and it centers the data around 0 with a standard deviation of 1, i.e.:

$$[\mu_{p\in P_{TRN}}(\boldsymbol{w}_p) - \sigma_{p\in P_{TRN}}(\boldsymbol{w}_p), \mu_{p\in P_{TRN}}(\boldsymbol{w}_p) + \sigma_{p\in P_{TRN}}(\boldsymbol{w}_p)] \rightarrow [-1, 1]$$

. These normalizations were also applied to the set of test patients P_{TST} using the limits computed on the training set.

Feature Concatenation Fixing a graph structure (*Graph_Full*, *Graph_66*, or *Graph_84*) and a measure between features (*corr*, *cos*, *euc*, or *sumNorm*), several sets W_d of feature vectors $\boldsymbol{w}_{d,p}$ were defined when varying the features used to encode the 3D texture in the regions. The possible features are $\mu_a(\mathbf{f}_H)$, $\sigma_a(\mathbf{f}_H)$, $\mu_a(\mathbf{f}_\mathcal{R})$, and $\sigma_a(\mathbf{f}_\mathcal{R})$. After normalizing each set of vectors W_d, concatenations of these descriptors were tested in order to better describe each patient. These were based on the underlying features. The combinations tested were:

- *FHOG* using the mean and the std ($\mu(\mathbf{f}_H)$ and $\sigma(\mathbf{f}_H)$),
- *Riesz* using the mean and the std ($\mu(\mathbf{f}_\mathcal{R})$ and $\sigma(\mathbf{f}_\mathcal{R})$),
- and *FHOG* and *Riesz* using the mean, the std, and both ($\mu(\mathbf{f}_H)$ and $\mu(\mathbf{f}_\mathcal{R})$, $\sigma(\mathbf{f}_H)$ and $\sigma(\mathbf{f}_\mathcal{R})$, and $\mu(\mathbf{f}_H)$ and $\mu(\mathbf{f}_\mathcal{R})$ and $\sigma(\mathbf{f}_H)$ and $\sigma(\mathbf{f}_\mathcal{R})$),

Feature Space Reduction When using the graph *Graph_Full* and feature concatenations, the feature space dimension was much larger than the number of patients. To avoid the known problems when using such large feature spaces, we tested two feature space reduction techniques. Both were applied in the training phase. In the first case, we considered the ground truth labels to be an step function, with values $\{0, 1\}$ for the MDR subtask, and $\{1, \ldots, 5\}$ for the TB task. Then, we computed the correlation between each feature dimension and these step functions, and selected the feature dimensions that correlated best with them. The threshold was set up as the mean absolute correlation of the feature dimensions with the labels. The second technique is based on the standard deviation of each feature. Only the components with a standard deviation higher than the mean of the standard deviations were selected. Both techniques reduced the size of the feature space by 2 approximately.

3.5 Classification

Multi-class support vector machine (SVM) classifiers with RBF kernel were used for each run of both subtasks, particularly, 2-class for the MDR task and 5-class for the TBT task. Grid search over the RBF parameters cost C and gamma γ was applied. Since the data were normalized, both C and γ moved between $[2^{-10}, 2^{-9}, \ldots, 2^{10}]$. The best C and γ combination for a run was set as the one with highest cross-validation accuracy in the training set of each subtask.

3.6 Tested Runs

The procedure explained in Section 3.4 resulted in 648 runs per subtask. Table 3 summarizes all possible options for each step using the same name coding as in the results tables.

Table 3. Possible configurations for each step. With this option there are $3 \times 4 \times 3 \times 3 \times 2 \times 3 = 648$ combinations.

Property	Options
Graph	Graph_Full, Graph_66, Graph_84
E. weight	corr, cos, euc, sumNorm
Features	FHOG, Riesz, FHOG and Riesz
F. measure	mean, std, mean and std
F. norm.	[0,1], Gauss[-1,1]
F. reduct.	none, mostCorr, mostStd

Submitted Runs A total of 10 runs could be submitted in the ImageCLEF 2017 TB task. 5 runs were submitted for each task. The 5 runs selected for each task were either one of the 5 best runs with respect to the cross-validation accuracy on the training set (Acc_{TRN}), or a late fusion of a few of them. The late fusion was executed using the probabilities that the SVM classifier returned and the mean probability of belonging to each class. The same procedure was applied in both tasks. Tables 4 and 5 show the identifier and run setup of the 5 best runs for each task respectively.

Table 4. Best runs for the MDR task.

Run Id.	Graph	Features	F. measure	E. weight	F. norm.	F. reduct.	Acc_{TRN}
MDR_Top1	Graph_84	FHOG and Riesz	mean and std	corr	Gauss[-1,1]	mostCorr	0.6900
MDR_Top2	Graph_66	FHOG and Riesz	std	cos	[0,1]	mostCorr	0.6856
MDR_Top3	Graph_84	FHOG	mean	corr	[0,1]	none	0.6812
MDR_Top4	Graph_66	FHOG and Riesz	mean and std	corr	[0,1]	mostCorr	0.6725
MDR_Top5	Graph_66	FHOG	mean	corr	Gauss[-1,1]	mostCorr	0.6725

Table 5. Best runs for the TBT task.

Run Id.	Graph	Features	F. measure	E. weight	F. norm.	F. reduct.	Acc_{TRN}
TBT_Top1	Graph_66	FHOG and Riesz	mean and std	sumNorm	Gauss[-1,1]	none	0.5276
TBT_Top2	Graph_84	FHOG and Riesz	mean and std	sumNorm	Gauss[-1,1]	none	0.5174
TBT_Top3	Graph_66	FHOG and Riesz	mean and std	sumNorm	[0,1]	none	0.5112
TBT_Top4	Graph_66	FHOG and Riesz	mean and std	sumNorm	Gauss[-1,1]	mostCorr	0.5112
TBT_Top5	Graph_84	FHOG and Riesz	mean and std	sumNorm	[0,1]	none	0.5092

4 Results

This section shows the results obtained on the training set by the submitted runs (Acc_{TRN}), and the final performance in the competition (Acc_{TST}). The final ranking was based on the Area Under the ROC Curve (AUC) for the MDR task, and on the unweighted Cohen's Kappa coefficient (Kappa) for the TBT task. Table 6 shows the results for the MDR subtask ordered by the ranking provided by the task organizers. The run identifiers *MDR_TopBest3* and *MDR_TopBest5* were obtained by doing late fusion with the 3 and 5 best runs respectively (see Section 3.6). The results for the TBT task are shown in Table 7. Again, the run identifiers *TBT_TopBest3* and *TBT_TopBest5* correspond to the late fusion of the 3 and 5 best runs respectively. The best run of the competition is shown in the same table.

Table 6. Results of Task 1 — Multi-drug resistance detection.

Run Id.	Run Filename	AUC	Acc_{TST}	Acc_{TRN}	#Rank
MDR_Top1	MDR_Top1_correct.csv	0.5825	0.5164	0.6900	1
MDR_TopBest3	MDR_submitted_topBest3_correct.csv	0.5727	0.4648	–	2
MDR_TopBest5	MDR_submitted_topBest5_correct.csv	0.5624	0.4836	–	3
MDR_Top2	MDR_Top2_correct.csv	0.5337	0.4883	0.6856	10
MDR_Top3	MDR_Top3_correct.csv	0.5112	0.4413	0.6725	17

Table 7. Results of Task 2 — Tuberculosis type classification.

Run Id.	Run Filename	Kappa	Acc_{TST}	Acc_{TRN}	#Rank
–	TBT_resnet_full.txt	0.2438	0.4033	–	1
TBT_Top1	TBT_Top1_correct.csv	0.1623	0.3600	0.5276	10
TBT_TopBest3	TBT_submitted_topBest3_correct.csv	0.1548	0.3500	–	12
TBT_TopBest5	TBT_submitted_topBest5_correct.csv	0.1410	0.3367	–	15
TBT_Top4	TBT_Top4_correct.csv	0.1352	0.3300	0.5112	16
TBT_Top2	TBT_Top2_correct.csv	0.1235	0.3200	0.5174	17

5 Conclusions

This work presents a new graph-model of the lung based on regional 3D texture features for describing lungs affected by tuberculosis. The participation in the ImageCLEF 2017 tuberculosis 2017 allows for an objective comparison between methods since the ground truth for the test set was never released. For the MDR task, our method participated with 5 runs and obtained the 1st, 2nd and 3rd place in the challenge. Also in the case of the TBT subtask 5 runs were submitted but the best rank obtained was 10. The results underline the difficulty of both

tasks and the suitability of our approach for describing TB patients. However, the results also suggest some overfitting by our method when comparing the accuracies obtained for the training and test sets.

Acknowledgements

This work was partly supported by the Swiss National Science Foundation in the project PH4D (320030–146804).

References

1. Müller, H., Clough, P., Deselaers, T., Caputo, B., eds.: ImageCLEF – Experimental Evaluation in Visual Information Retrieval. Volume 32 of The Springer International Series On Information Retrieval. Springer, Berlin Heidelberg (2010)
2. Kalpathy-Cramer, J., García Seco de Herrera, A., Demner-Fushman, D., Antani, S., Bedrick, S., Müller, H.: Evaluating performance of biomedical image retrieval systems: Overview of the medical image retrieval task at ImageCLEF 2004–2014. Computerized Medical Imaging and Graphics **39**(0) (2015) 55 – 61
3. Villegas, M., Müller, H., Gilbert, A., Piras, L., Wang, J., Mikolajczyk, K., García Seco de Herrera, A., Bromuri, S., Amin, M.A., Kazi Mohammed, M., Acar, B., Uskudarli, S., Marvasti, N.B., Aldana, J.F., Roldán García, M.d.M.: General overview of ImageCLEF at the CLEF 2015 labs. In: Working Notes of CLEF 2015. Lecture Notes in Computer Science. Springer International Publishing (2015)
4. Ionescu, B., Müller, H., Villegas, M., Arenas, H., Boato, G., Dang-Nguyen, D.T., Dicente Cid, Y., Eickhoff, C., Garcia Seco de Herrera, A., Gurrin, C., Islam, B., Kovalev, V., Liauchuk, V., Mothe, J., Piras, L., Riegler, M., Schwall, I.: Overview of ImageCLEF 2017: Information extraction from images. In: Experimental IR Meets Multilinguality, Multimodality, and Interaction 8th International Conference of the CLEF Association, CLEF 2017. Volume 10456 of Lecture Notes in Computer Science., Dublin, Ireland, Springer (September 11-14 2017)
5. Dicente Cid, Y., Kalinovsky, A., Liauchuk, V., Kovalev, V., , Müller, H.: Overview of ImageCLEFtuberculosis 2017 - predicting tuberculosis type and drug resistances. In: CLEF 2017 Labs Working Notes. CEUR Workshop Proceedings, Dublin, Ireland, CEUR-WS.org <http://ceur-ws.org> (September 11-14 2017)
6. Jeong, Y.J., Lee, K.S.: Pulmonary tuberculosis: up-to-date imaging and management. American Journal of Roentgenology **191**(3) (2008) 834–844
7. Dicente Cid, Y., Müller, H., Platon, A., Janssens, J.P., Lador, F., Poletti, P.A., Depeursinge, A.: A lung graph–model for pulmonary hypertension and pulmonary embolism detection on DECT images. In: MICCAI Workshop on Medical Computer Vision: Algorithms for Big Data, MCV 2016. (2016)
8. Dicente Cid, Y., Jimenez-del-Toro, O., Depeursinge, A., Müller, H.: Efficient and fully automatic segmentation of the lungs in CT volumes. In Orcun Goksel, Jimenez-del-Toro, O., Foncubierta-Rodriguez, A., Müller, H., eds.: Proceedings of the VISCERAL Challenge at ISBI. Number 1390 in CEUR Workshop Proceedings (Apr 2015)
9. Depeursinge, A., Zrimec, T., Busayarat, S., Müller, H.: 3D lung image retrieval using localized features. In: Medical Imaging 2011: Computer–Aided Diagnosis. Volume 7963., SPIE (2011) 79632E

10. Liu, K., Skibbe, H., Schmidt, T., Blein, T., Palme, K., Brox, T., Ronneberger, O.: Rotation-invariant hog descriptors using fourier analysis in polar and spherical coordinates. International Journal of Computer Vision **106**(3) (2014) 342–364
11. Dicente Cid, Y., Müller, H., Platon, A., Poletti, P.A., Depeursinge, A.: 3–D solid texture classification using locally–oriented wavelet transforms. IEEE Transactions on Image Processing **26**(4) (April 2017) 1899–1910