

Texture-based Graph Model of the Lungs for Drug Resistance Detection, Tuberculosis Type Classification, and Severity Scoring: Participation in the ImageCLEF 2018 Tuberculosis Task

Yashin Dicente Cid^{1,2} and Henning Müller^{1,2}

¹ University of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland;

² University of Geneva, Switzerland

yashin.dicente@hevs.ch

Abstract. In 2018, ImageCLEF proposed a task using CT (Computed Tomography) scans of patients with tuberculosis (TB). The task was divided into three subtasks: multi-drug resistance detection, TB type classification, and severity scoring. In this work we present a graph model of the lungs capable of characterizing TB patients with different lung problems. The graph contains a fixed number of nodes with weighted edges based on dissimilarity measures between texture descriptors computed in the nodes. This model encodes the texture distribution along the lungs, making it suitable for describing patients with different TB types. The results show the strength of the technique, leading to high results in the three subtasks.

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], usually based on tasks specifically requested by radiologists [4] or making knowledge of visual data accessible [5]. The ImageCLEF 2018 [6] challenge included a task based on CT (Computed Tomography) volumes of patients with tuberculosis (TB), the ImageCLEF 2018 TB task [7]. In this task, a dataset of lung CT scans was provided and 3 subtasks were proposed. The 2017 edition of the ImageCLEF TB task [8] already included 2 of the 3 subtasks: multi-drug resistance detection and TB type classification. However, the dataset was smaller than in 2018. We participated in the 2017 challenge with a texture-based graph model of the lungs [9]. For the 2018 edition we applied the lessons learned in 2017 and we participated in the three subtasks with a refined approach. The new third subtask targeted the prediction

of a general severity score of the disease. Health professionals face this task by mainly visual inspection of the CT volumes. However, they base their final score on other clinical data as well as on the image.

When tuberculosis affects the lungs, several visual patterns can be seen in a CT image. These patterns are usually characteristic of the underlying TB type. Moreover, their spread into parts of the lung is a good indicator of the severity of the diseases. However, the final diagnosis usually required other analyses than only the images [10]. Our approach is able to obtain a global texture-based description of the lungs. It consists of creating a graph model of the lungs where the nodes represent lung regions and the edges encode relations between the texture inside the regions.

The following section contains a brief overview of the subtasks and datasets of the ImageCLEF 2018 TB task. More detailed information on the task can be found in the overview article [7]. Section 3 explains the process of building the texture-based graph model of the lungs and all the variations tested for this task in detail. The results obtained by this approach in the three subtasks are shown in Section 4. Finally, Section 5 concludes our participation in this challenge.

2 Subtasks and Datasets

The ImageCLEF 2018 TB task proposed three subtasks: i) Multi-drug resistance (MDR) detection, ii) Tuberculosis type (TBT) classification, iii) and severity scoring (SVR). For the three subtasks, volumetric chest CT images and automatic segmentations of the lungs were provided by the organization. No other lung segmentation was attempted in this work and the masks provided were used. These masks were obtained with the method described in [11].

The challenge was divided into two phases. In the first phase, the organizers released for each subtask a set of patient CT volumes as training set along with their lung masks, groundtruth labels and meta-data including age and gender of the patients. In the second phase, the test set with the corresponding lung segmentations and meta-data were provided. The groundtruth for the test data was never released. In our approach we only used the CT images and no meta-data. The evaluation on the test set was performed by the organizers after the scheduled deadline for all runs that were submitted in time. The number of CT volumes for each subtask are specified in Tables 1, 2, and 3.

Table 1. Dataset of the MDR detection subtask. DS means drug-sensible.

Patient set	Train	Test
DS	134	99
MDR	125	137
Total patients	259	236

Table 2. Dataset of the TBT classification subtask.

Patient set	Num. Patients (CTs)			
	Train		Test	
Type 1	228	(376)	89	(179)
Type 2	210	(273)	80	(115)
Type 3	100	(154)	60	(86)
Type 4	79	(106)	50	(71)
Type 5	60	(99)	38	(57)
Total patients (CTs)	677	(1,008)	317	(505)

Table 3. Dataset of the SVR subtask.

Patient set	Train	Test
Low severity	90	62
High severity	80	47
Total patients	170	109

3 Methods

This section details the process of building our graph-based model, the extraction of a descriptor vector from the CT images, and the classification algorithm applied. The same technique was applied for describing the patients of the three subtasks. Our method consists of creating a graph model of the lungs, with nodes based on a geometrical atlas and weighted edges encoding dissimilarities between 3D texture descriptors of each atlas region. Figure 1 shows the pipeline to obtain the graph model.

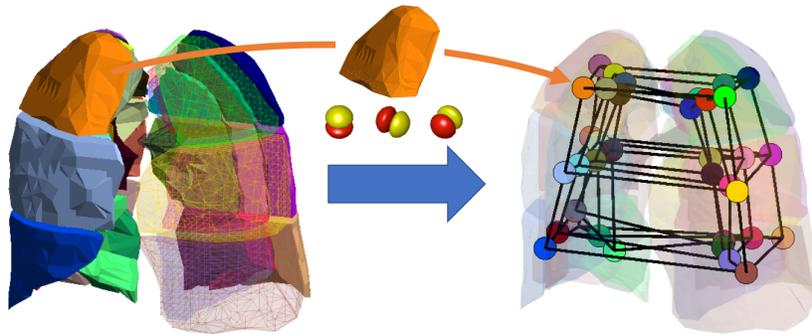


Fig. 1. Construction pipeline of the graph model consisting of 3 steps: lung mask division, regional feature extraction and graph entity formation.

3.1 Preprocessing

Our approach is based on 3D texture features that require having isometric voxels. We first made the 3D images and the lung masks isometric. After analyzing the multiple resolutions and the inter-slice distances in the dataset, we opted for a voxel size of 1 mm to capture a maximum of the available information.

3.2 Atlas of the Lung

To build a graph with a fixed structure over the lung anatomy we opted for a geometric division of the lung with a fixed number of regions. We chose the atlas developed by Depeursinge et al. in [12]. This atlas is only based on the mask of the lungs and provides 36 geometric regions dividing the lungs as shown in Figure 1.

3.3 3D Texture Features

Two state-of-the-art 3D texture features were selected to describe the texture inside the lungs. The first method is a histogram of gradients based on the Fourier transform HOG (FHOG) introduced in [13]. We used 28 3D directions 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 [14] and based on [15]. The parameters that obtained the best results in the above mentioned article were used in our approach. These are: 3rd-order Riesz transform, 4 scales and 1st-order alignment. This configuration resulted in 40-dimensional feature vectors for each image voxel. The feature vector for a single voxel was then reduced to be 10-dimensional containing the energy of each filter along the 4 scales ($\mathbf{f}_R \in \mathbb{R}^{10}$).

Feature Vector of a Region: Several feature vectors can be extracted from a region r using the above mentioned texture descriptors \mathbf{f}_H and \mathbf{f}_R . Given a region r_i , we extracted the mean (μ_i) and standard deviation (σ_i) of the features inside the region, i.e.: $\mu_i(\mathbf{f}_H)$, $\sigma_i(\mathbf{f}_H)$, $\mu_i(\mathbf{f}_R)$, and $\sigma_i(\mathbf{f}_R)$. Hence, four feature vectors were obtained for each atlas region.

3.4 Texture-based Graph Model of the Lungs

We used a weighted undirected graph $\mathcal{G}(\mathcal{N}, \mathcal{E})$ with 36 nodes and 84 edges to model the lung. Each node $N_i \in \mathcal{N}$ corresponds to a region r_i in the geometric atlas. The edges were based on the region adjacency defined by the atlas. In this case, there is an edge $E_{i,j}$ between nodes N_i and N_j if regions r_i and r_j are neighbors in the atlas, i.e., if they are 3D adjacent. Moreover, it has 18 additional edges connecting each pair of nodes representing opposite regions inside the atlas with respect to the left/right division of the lungs. This configuration resulted in 84 edges and is shown in Figure 1.

The weight $w_{i,j}$ of an edge $E_{i,j}$ was defined using two dissimilarity measures between the regional features: the Euclidean and the correlation distances. Depending on the measure selected, a different graph model was obtained.

3.5 Graph-based Patient Descriptor

The feature descriptor \mathbf{w}_p of a patient p was defined as the ordered list of weights $w_{i,j} \in G_p$. Since $w_{i,j} = w_{j,i}$, this descriptor was 84-dimensional for all patients. These vectors were normalized using a Z-score normalization based on the patients of the training set. The elements composing this vector can not be seen independently since they encode the structure of the graph. In this case, the normalization was applied to all the components simultaneously, i.e., the mean and variance needed for the Z-score normalization were computed over all edge weights from the training patients.

Patient Descriptor Concatenation: For a given distance between regional features (*corr* or *eucl*), four normalized patient descriptor vectors $\bar{\mathbf{w}}_p$ were obtained. These were: $\bar{\mathbf{w}}_{\mu_H}$, $\bar{\mathbf{w}}_{\sigma_H}$, $\bar{\mathbf{w}}_{\mu_R}$, and $\bar{\mathbf{w}}_{\sigma_R}$. We ran a different experiment for each of these patient descriptors. Moreover, we also tested five concatenations of these descriptors in order to better describe each patient, resulting in nine experiments. The tested concatenations were defined as:

- Mean and std of FHOG: $\hat{\mathbf{w}} = (\bar{\mathbf{w}}_{\mu_H} || \bar{\mathbf{w}}_{\sigma_H})$.
- Mean and std of Riesz: $\hat{\mathbf{w}} = (\bar{\mathbf{w}}_{\mu_R} || \bar{\mathbf{w}}_{\sigma_R})$.
- Mean of FHOG and Riesz: $\hat{\mathbf{w}} = (\bar{\mathbf{w}}_{\mu_H} || \bar{\mathbf{w}}_{\mu_R})$.
- Std of FHOG and Riesz: $\hat{\mathbf{w}} = (\bar{\mathbf{w}}_{\sigma_H} || \bar{\mathbf{w}}_{\sigma_R})$.
- Mean and std of FHOG and Riesz: $\hat{\mathbf{w}} = (\bar{\mathbf{w}}_{\mu_H} || \bar{\mathbf{w}}_{\sigma_H} || \bar{\mathbf{w}}_{\mu_R} || \bar{\mathbf{w}}_{\sigma_R})$.

Feature Space Reduction: For some of the concatenations of patient descriptors, the feature space dimension was significantly larger than the number of patients in a single class. To avoid the known problems of using such large feature spaces, we performed feature space reduction by selecting the dimensions with higher correlation with respect to the groundtruth labels. This technique reduced the size of the feature space by two approximately. The final experiments were performed both using this feature space reduction and with no reduction.

Tested runs: We performed 36 experiments (or runs) in total per subtask, generated by the different options explained. Table 4 summarizes all possible options for each step.

3.6 Classification

Multi-class support vector machine (SVM) classifiers with RBF kernel were used for each run in the three subtasks, particularly, 2-class for the MDR subtask, and

Table 4. Possible configurations for each algorithm step. With these options we run 36 experiments (or runs).

Algorithm property	Options
Texture descriptors	FHOG, Riesz, FHOG & Riesz
Regional features	mean, std, mean & std
Distance measures	corr, euc
Feature space reductions	none, mostCorr

5-class for the TBT and SVR subtasks. Grid search over the RBF parameters cost C and gamma γ was applied. Since the data were normalized, both C and γ moved in $\{2^{-10}, 2^{-9}, \dots, 2^{10}\}$. 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.

MDR subtask: The submission run for the this subtask had to contain the probability for a patient of being multi-drug resistant. In this case, we used the probabilities from the SVM algorithm.

TBT subtask: For this subtask, the dataset contained more than one CT scan per patient in most of the cases. However, the classification had to be performed at the patient level. In this case, we averaged the probabilities provided by the SVM algorithm over all the images of the same patient. The overall most probable class was assigned to the patient.

SVR subtask: In this subtask the submission file had to contain the severity score (1 to 5) and the probability of belonging to the *high* severity class for each patient. We considered the severity score as a class in the SVM algorithm, assigning the most probable class as the predicted severity score. According to the organizers, the *high* class corresponded to a severity score in [1, 3]. In this case, for each patient, we obtained the probability of class *high* by adding the SVM probabilities of belonging to classes 1 to 3.

4 Results

A total of 10 runs could be submitted in each ImageCLEF 2018 TB subtask. We chose them based on the performance obtained in the training set. For this, we used the same indicators as the ones used by the organizers. These were: Accuracy (Acc) and Area Under the Curve (AUC) for the MDR subtask, Acc and unweighted Cohen’s Kappa (Kappa) for the TBT subtask, and AUC and Root Mean Square Error (RMSE) for the SVR task. Additionally, we base our selection of the best runs on the cross-validation accuracy extracted from the SVM classifier during the training phase. Tables 5, 6, and 7 show the results obtained by each submitted run along with the best result obtained by other

participants with respect to each performance measure. All results were provided by the organizers of the task. Our group participated as the *MedGIFT* group. In the case of the MDR subtask, only 9 runs were evaluated since one submission failed.

Table 5. MDR subtask results. We participated with 9 runs as the *MedGIFT* group.

Group	Configuration	Rank		Rank	
		AUC	AUC	Acc	Acc
VISTA@UEvora	Not applicable	0.6178	1	0.5593	8
San Diego VA HCS/UCSD	Not applicable	0.6114	2	0.6144	1
MedGIFT	Riesz / std / corr / mostCorr	0.5237	22	0.5593	12
MedGIFT	FHOG / std / euc / mostCorr	0.5205	23	0.5932	2
MedGIFT	FHOG & Riesz / std / corr / none	0.5095	25	0.4873	33
MedGIFT	FHOG & Riesz / std / euc / mostCorr	0.5039	27	0.5424	19
MedGIFT	FHOG / mean / corr / mostCorr	0.4941	30	0.5551	14
MedGIFT	Riesz / mean & std / corr / mostCorr	0.4855	31	0.5212	22
MedGIFT	Riesz / mean / euc / mostCorr	0.4824	33	0.5297	21
MedGIFT	FHOG / mean and std / euc / mostCorr	0.4693	36	0.5720	6
MedGIFT	FHOG & Riesz / mean & std / corr / mostCorr	0.4568	38	0.5085	27

Table 6. TBT subtask results. We participated with 10 runs as the *MedGIFT* group.

Group	Configuration	Rank		Rank	
		Kappa	Kappa	Acc	Acc
UIIP_BioMed	Not applicable	0.2312	1	0.4227	1
MedGIFT	FHOG & Riesz / std / euc / mostCorr	0.1706	3	0.3849	2
MedGIFT	Riesz / mean & std / euc / mostCorr	0.1674	4	0.3849	3
MedGIFT	FHOG & Riesz / mean & std / corr / mostCorr	0.1531	8	0.3691	7
MedGIFT	FHOG & Riesz / mean / euc / none	0.1517	9	0.3628	8
MedGIFT	Riesz / std / euc / mostCorr	0.1494	10	0.3722	6
MedGIFT	FHOG & Riesz / mean & std / corr / none	0.1356	13	0.3628	9
MedGIFT	FHOG / mean & std / euc / mostCorr	0.0949	19	0.3344	14
MedGIFT	FHOG / std / corr / none	0.0855	21	0.3218	18
MedGIFT	FHOG & Riesz / std / corr / mostCorr	0.0787	23	0.3281	17
MedGIFT	FHOG / std / corr / mostCorr	0.0589	26	0.3060	21

Considering the accuracy, the proposed method ranked 2nd in the MDR and TBT subtasks. However, considering the AUC in the MDR subtask, the best result achieved ranked 22nd. In the case of the TBT subtask, our best run with respect to the accuracy is also the best with respect to the Kappa measure, obtaining the 3rd place. For the SVR subtask, our method ranked 1st according to the AUC and 2nd considering the RMSE. Overall, the results are good when compared with other participants. However, they are still far from perfect in the three subtasks because for real clinical applications a higher accuracy seems necessary. Moreover, they reflect that the optimization of the SVM parameters was based on the cross-validation accuracy.

Table 7. SVR subtask results. We participated with 10 runs as the *MedGIFT* group.

Group	Configuration	Rank		Rank	
		RMSE	RMSE	AUC	AUC
UIIP_BioMed	Not applicable	0.7840	1	0.7025	6
MedGIFT	FHOG / std / euc / mostCorr	0.8513	2	0.7162	5
MedGIFT	FHOG & Riesz / mean & std / euc / none	0.8883	4	0.6733	10
MedGIFT	FHOG & Riesz / mean & std / corr / mostCorr	0.8934	5	0.7708	1
MedGIFT	FHOG / mean / euc / mostCorr	0.8985	6	0.7443	3
MedGIFT	FHOG / mean / corr / mostCorr	0.9237	7	0.6450	18
MedGIFT	FHOG / mean & std / euc / none	0.9433	8	0.7268	4
MedGIFT	FHOG / mean & std / corr / mostCorr	0.9433	9	0.7608	2
MedGIFT	Riesz / mean & std / corr / none	0.9626	11	0.5535	34
MedGIFT	FHOG & Riesz / mean / euc / none	0.9954	15	0.6644	12
MedGIFT	Riesz / std / corr / mostCorr	1.0492	20	0.5841	29

5 Conclusions

This work presents an updated version of our previously developed graph model of the lung based on regional 3D texture features for describing lungs affected by tuberculosis. The participation in the ImageCLEF 2018 TB task allows for an objective comparison between methods since the evaluation was performed by the organizers. The results proved the suitability of our approach for detecting, classifying and scoring patients with TB. However, the results show that there is still room for improvement, particularly in the MDR subtask where results were relatively close to random for all participants. We believe that the results could have been better if the optimization of the classifier parameters would have been done based on other performance measures rather than only on the SVM cross-validation accuracy.

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. Markonis, D., Holzer, M., Dungs, S., Vargas, A., Langs, G., Kriewel, S., Müller, H.: A survey on visual information search behavior and requirements of radiologists. *Methods of Information in Medicine* **51**(6) (2012) 539–548
 5. Müller, H., Kalpathy-Cramer, J., Demner-Fushman, D., Antani, S.: Creating a classification of image types in the medical literature for visual categorization. In: *SPIE Medical Imaging*. (2012)
 6. Ionescu, B., Müller, H., Villegas, M., de Herrera, A.G.S., Eickhoff, C., Andrearczyk, V., Cid, Y.D., Liauchuk, V., Kovalev, V., Hasan, S.A., Ling, Y., Farri, O., Liu, J., Lungren, M., Dang-Nguyen, D.T., Piras, L., Riegler, M., Zhou, L., Lux, M., Gurrin, C.: Overview of ImageCLEF 2018: Challenges, datasets and evaluation. In: *Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Ninth International Conference of the CLEF Association (CLEF 2018)*, Avignon, France, LNCS Lecture Notes in Computer Science, Springer (September 10-14 2018)
 7. Dicente Cid, Y., , Liauchuk, V., Kovalev, V., , Müller, H.: Overview of ImageCLEFtuberculosis 2018 - detecting multi-drug resistance, classifying tuberculosis type, and assessing severity score. In: *CLEF2018 Working Notes. CEUR Workshop Proceedings, Avignon, France, CEUR-WS.org* <<http://ceur-ws.org>> (September 10-14 2018)
 8. 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)
 9. Dicente Cid, Y., Batmanghelich, K., Müller, H.: Textured graph-model of the lungs for tuberculosis type classification and drug resistance prediction: participation in ImageCLEF 2017. In: *CLEF2017 Working Notes. CEUR Workshop Proceedings, Dublin, Ireland, CEUR-WS.org* <<http://ceur-ws.org>> (September 11-14 2017)
 10. Jeong, Y.J., Lee, K.S.: Pulmonary tuberculosis: up-to-date imaging and management. *American Journal of Roentgenology* **191**(3) (2008) 834–844
 11. 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) 31–35
 12. 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
 13. 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
 14. 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
 15. Depeursinge, A., Foncubierta-Rodríguez, A., Van De Ville, D., Müller, H.: Rotation-covariant texture learning using steerable Riesz wavelets. *IEEE Transactions on Image Processing* **23**(2) (February 2014) 898–908