

Lung Graph–Model Classification with SVM and CNN for Tuberculosis Severity Assessment and Automatic CT Report Generation

Participation in the ImageCLEF 2019 Tuberculosis Task

Yashin Dicente Cid¹ 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 2019, ImageCLEF proposed a task using CT (Computed Tomography) scans of patients with tuberculosis (TB). The task was divided into two subtasks: TB severity assessment (SVR subtask) and automatic CT report generation (CTR subtask). In this work we present our participation in the task. We participated with a graph model of the lungs with a morphology–based structure that was previously validated for the detection of TB types. The graph is based on a parcellation of the lung fields into supervoxels, where each region is identified as a node of the graph. A weighted edge is defined between nodes representing adjacent regions. The associated weight is computed as the distance between 3D texture descriptors extracted from the two connected regions. This model encodes the texture distribution along the lungs, making it suitable for detecting the tissue abnormalities present in TB patients. In this work we explore two techniques to classify these graphs: (i) a lung descriptor vector based on the aggregation of graph centrality measures and (ii) a set of 2D histograms encoding the binary distribution of node features. The final classification is performed with support vector machines for the lung descriptor vector and with convolutional neural networks for the 2D histograms. The results show the strength of the technique, leading to 6th and 3rd place in the SVR and CTR subtasks, respectively.

Keywords: Lung Graph Model, Graph Kernel, CNN, 3D Texture Analysis, Computed Tomography, 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

Copyright © 2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). CLEF 2019, 9-12 September 2019, Lugano, Switzerland.

retrieval task (ImageCLEF) has been organized [2, 3], usually based on tasks specifically requested by radiologists [4] or making knowledge of public visual data [5]. The ImageCLEF 2019 [6] challenge included for the third year a task based on CT (Computed Tomography) volumes of patients with tuberculosis (TB), the ImageCLEF 2019 TB task [7]. In this task, a dataset of lung CT scans with associated meta-data was provided and two subtasks were proposed. The 2018 edition of the ImageCLEF TB task [8] already included the TB severity score prediction (SVR) subtask and this year it included the automatic CT report generation (CTR) subtask.

When tuberculosis affects the lungs, several visual patterns can be seen in a CT image, 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 requires other analyses rather than only the images [9]. We participated in the 2017 and 2018 challenge with a texture-based graph model of the lungs [10, 11] capable to offer a global descriptor of the lung texture. In both cases, the structure underlying the graph was based on a geometric parcellation of the lung fields into a fixed number of regions. This graph model offers the possibility of describing each patient with a vector of weights that has the same size for all patients, hence enabling its direct comparison. Later on, we developed a new graph model with morphology-based structure that required a graph kernel transformation in order to compare between patients. This model was tested on the same data from the ImageCLEF 2018 TB task, obtaining much higher results than the previous graph with fixed structure in the TB type classification task [12]. For the 2019 edition of the TB task we used our latest graph model with morphology-based structure and we investigated two approaches to classify the resulting data: the first is based on the graph kernel developed in [12] and the second one, proposed by Tixier *et al.* [13], is based on transforming the graphs in a set of 2D images and apply a 2D convolutional neural network (CNN).

The following section contains a brief overview of the subtasks and datasets of the ImageCLEF 2019 TB task. More detailed information on the subtasks and data can be found in the overview article of the task [7]. Section 3 first summarizes the process to build the texture-based graph model of the lungs with morphology-based structure, followed by the graph kernel and the 2D CNN approaches. The results obtained by these approaches in the two subtasks are shown in Section 4. Finally, Section 5 concludes our participation in this challenge.

2 Subtasks and Dataset

The ImageCLEF 2019 TB task proposed two subtasks: i) Severity score assessment (SVR subtask) and ii) automatic CT report generation (CTR subtask). Both subtasks share the same dataset, composed of 335 volumetric chest CT scans with associated meta-data. Moreover, automatic segmentations of the lungs were provided by the organizers, obtained with the method described

in [14]. The masks provided were used and no other lung segmentation was attempted in this work.

The challenge was divided into two phases. In the first phase, the organizers released a set of 218 CT volumes for training, along with their lung masks, ground truth labels and meta-data. In the second phase, the test set (117 CT scans) with the corresponding lung segmentations and meta-data was provided. The ground truth for the test data was never made available. In our approach we only used the CT images and the lung masks but not the meta-data. The evaluation on the test set was performed by the organizers after the scheduled deadline. The exact details of the dataset, such as number of patients per class, are included in the overview article of the TB task [7].

The SVR subtask consisted of assigning a TB severity score to each CT image. The original score varied between 1 and 5 and the organizers reduced it to two scores, HIGH and LOW, therefore transforming the original 5-class problem into a binary classification. In the case of the CTR subtask, the objective was to detect the presence/absence of 6 TB related CT-findings in each CT scan. In our approach we treated this subtask as a multi-binary problem, assuming complete independence between the findings. For the 7 binary labels to detect (TB severity + 6 CT-findings) we used the same graph model, only varying the training labels in the last classification step of our pipeline.

3 Methods

3.1 Lung Graph Model

To build the graph model of the lungs with morphology-based structure we followed the pipeline introduced in [12] and that is shown in Figure 1. The key details of each of the steps in the pipeline, including the preprocessing applied to the CT scans, follow. Unless otherwise stated, any other detail remains the same as in the original work [12].

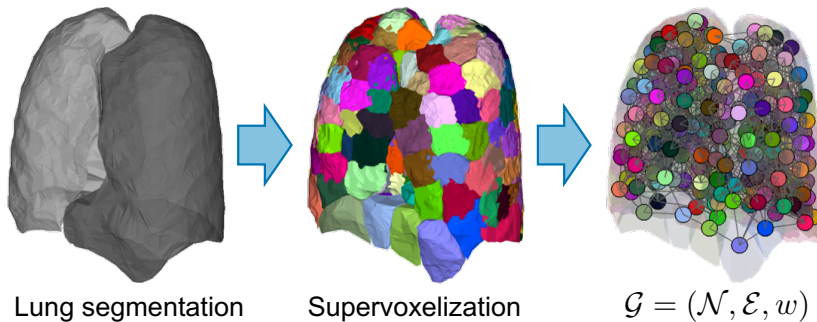


Fig. 1. Construction pipeline of the graph model consisting of 3 steps: lung parcellation, regional feature extraction and graph entity formation.

Preprocessing: For our approach we use rotation-invariant 3D texture features that require having isometric voxels. We first made the 3D images and the lung masks isometric. After analyzing the different resolutions present in the dataset, a voxel size of 1 mm was selected to capture the maximum textural information.

Supervoxelization: The parcellation of the lung fields was done using a 3D supervoxelization algorithm based on a generalization for 3D volumes of the SLIC algorithm [15, 16], that created homogeneous regions in terms of Hounsfield units (HU).

Regional 3D Texture Features: Each region in the supervoxelization was described using two state-of-the-art 3D texture descriptors: the histogram of oriented gradients based on the Fourier transform HOG (FHOG) introduced in [17] and the locally-oriented 3D Riesz-wavelet transform introduced in [18]. The former resulted in a 28-dimensional vector ($\mathbf{f}_H(v) \in \mathbb{R}^{28}$) and the latter in a 40-dimensional vector ($\mathbf{f}_R(v) \in \mathbb{R}^{40}$) for each voxel v in the region. Finally, given a region r , we extracted the mean (μ) and standard deviation (σ) of the above mentioned features inside the region, *i.e.* $\mu(\mathbf{f}_H(r))$, $\sigma(\mathbf{f}_H(r))$, $\mu(\mathbf{f}_R(r))$ and $\sigma(\mathbf{f}_R(r))$. Hence, four feature vectors were obtained for each region.

Texture-based Graph Model of the Lungs: We used a weighted undirected graph $\mathcal{G}(\mathcal{N}, \mathcal{E}, w)$ using the supervoxelization as underlying structure, *i.e.* each node $N_i \in \mathcal{N}$ corresponds to a region r_i in the supervoxelization. Then, an undirected edge $E_{i,j}$ with associated weight $w_{i,j}$ is defined between nodes N_i and N_j if regions r_i and r_j are 3D adjacent inside the lung parcellation. The weight $w_{i,j}$ is defined as the correlation distance between the regional feature vectors.

Since we used four regional feature vectors (μ_H , σ_H , μ_R and σ_R), four graphs with same edges but different edge-weights were generated for each patient: \mathcal{G}_{μ_H} , \mathcal{G}_{σ_H} , \mathcal{G}_{μ_R} and \mathcal{G}_{σ_R} .

3.2 Graph Classification

At this point, each patient is described with four graphs that share the same number of nodes and edges, but differ from any graph from another patient. Therefore, a method is required that translates these graphs into a common space where they can be compared. In this section we describe the two methods tested in 2019. One is based on extracting node centrality measures and the second one on describing each graph as a set of 2D images.

3.3 Graph Classification using Graph Kernels

In this case, we described each graph with a fixed number of values derived from graph measures. This method was developed in [12] and it is summarized

here: For each node N in a graph \mathcal{G} we computed five graph centrality measures: the weighted degree $d^w(N)$, the relative weighted degree $d^r(N) = \frac{d^w(N)}{d(N)}$, the weighted closeness $c^w(N)$, the relative weighted closeness $c^r(N) = \frac{c^w(N)}{d(N)}$ and the weighted betweenness $b^w(N)$. Considering the entire set of nodes \mathcal{N} in \mathcal{G} , each of these five measures was modeled as a distribution X : $D_{\mathcal{G}}^w = \{d^w(N)\}$, $D_{\mathcal{G}}^r = \{d^r(N)\}$, $C_{\mathcal{G}}^w = \{c^w(N)\}$, $C_{\mathcal{G}}^r = \{c^r(N)\}$ and $B_{\mathcal{G}}^w = \{b^w(N)\}$, where $N \in \mathcal{N}$. Then, we described each of these distributions using 10 equidistant percentiles from 0 to 100, with percentiles 0 ($\pi_1(X)$) and 100 ($\pi_{10}(X)$) being the minimum and maximum values in the distribution, respectively. Let $\boldsymbol{\pi}(X) = (\pi_1(X), \dots, \pi_{10}(X))$ be the vector composed of the 10 percentiles $\pi_k(X)$ of a distribution X . Then, our graph-based lung descriptor is defined as:

$$\boldsymbol{\omega}(\mathcal{G}) = (\mu_w, \boldsymbol{\pi}(D_{\mathcal{G}}^w), \boldsymbol{\pi}(D_{\mathcal{G}}^r), \boldsymbol{\pi}(C_{\mathcal{G}}^w), \boldsymbol{\pi}(C_{\mathcal{G}}^r), \boldsymbol{\pi}(B_{\mathcal{G}}^w))$$

where μ_w is the mean of the weights in the graph.

For each patient p with graph model \mathcal{G}_p , its graph-based lung descriptor $\boldsymbol{\omega}(\mathcal{G}_p)$ belongs to \mathbb{R}^{51} . From now on, $\boldsymbol{\omega}(\mathcal{G}_p)$ is referred to as $\boldsymbol{\omega}_{\mathbf{f},p}$, where \mathbf{f} corresponds to the regional feature used to build the graph \mathcal{G}_p (see Section 3.1). Since four regional features were computed in each region of the lung parcellation ($\boldsymbol{\mu}_H(r)$, $\boldsymbol{\sigma}_H(r)$, $\boldsymbol{\mu}_{\mathcal{R}}(r)$, and $\boldsymbol{\sigma}_{\mathcal{R}}(r)$), the final lung descriptor $\hat{\boldsymbol{\omega}}_p$ used in our experiments was defined as the concatenation of the four graph-based lung descriptors:

$$\hat{\boldsymbol{\omega}}_p = (\boldsymbol{\omega}_{\boldsymbol{\mu}_H,p}, \boldsymbol{\omega}_{\boldsymbol{\sigma}_H,p}, \boldsymbol{\omega}_{\boldsymbol{\mu}_{\mathcal{R}},p}, \boldsymbol{\omega}_{\boldsymbol{\sigma}_{\mathcal{R}},p}) \in \mathbb{R}^{204}.$$

Finally, these lung descriptor vectors $\hat{\boldsymbol{\omega}}_p$ were fed into a Support Vector Machine (SVM) classifier to provide the prediction scores. For each of the 7 binary problems (see Section 2), feature normalization and dimensionality reduction was applied with respect to the training data before the SVM classifier was used. The SVM parameters were found by grid-search and 10-fold cross-validation with accuracy as the performance measure to optimize. Using this approach we submitted 2 run files: *SVR_SVM.txt* to the SVR subtask and *CTR_SVM.txt* to the CTR subtask.

3.4 Graph Classification using 2D CNNs

The second graph classification method used was based on the approach proposed by Tixier *et al.* in [13]. Their original method was composed of four steps:

1. They first applied a graph node embedding called *node2vec* [19] that embeds each node in a D -dimensional space (D is a priori different for each graph).
2. Since each graph \mathcal{G}_k is described with a set of n_k vectors of dimension D_k , where n_k is the number of nodes in graph \mathcal{G}_k , they used PCA to align the dimensions of the embeddings and they selected the d first dimensions from each embedding (where $d < D_k, \forall k$).
3. Then, they extracted 2D histograms by slicing the d -dimensional PCA node embedding space. As a result, each graph is then represented as a stack of $\frac{d}{2}$ 2D histograms with size 28×28 .

4. As a final step, they considered the 2D histograms as 2D images and they trained a 2D CNN with input size $(\frac{d}{2}, 28, 28)$ to classify the initial graphs.

In our experiments we followed the same pipeline with $d = 5$ and we used the same CNN architecture³. However, since for each patient we extracted four graphs (see Section 3.1), after step 3 we concatenated their respective 5-dimensional embeddings, obtaining $4 \cdot \frac{5}{2} = 10$ 2D histograms. The network used was then set up to have input size (10,28,28). With this technique we submitted two runs, one per subtask: *SVR_GNN_node2vec_pca.csv* and *CTR_GNN_node2vec_pca.csv*.

A variant of the same method was also tested using the graph centrality measures described in Section 3.3. In this case, we replaced the node2vec and PCA steps by directly describing each node with the five abovementioned measures. The next steps were maintained, *i.e.* the concatenation of the embeddings, the creation of the 2D histograms of size 28×28 and the classification using a 2D CNN with input size (10,28,28). Two more runs, one for each subtask, were submitted using this approach: *SVR_GNN_nodeCentralFeats.csv* and *CTR_GNN_nodeCentralFeats.csv*.

Our last set of submitted runs was based on selecting the best channels before the CNN: For each binary problem we selected during training the channels where the average 2D histograms differentiated stronger between classes. This selection was done manually by visual inspection of the average 2D histograms for each problem and class and the number of selected channels varied for each binary CT-finding. Figure 2 shows the average 2D histograms obtained for each embedding (based on graph centrality measures and node2vec) for the SVR subtask. For the shown example we selected the channels {2,4,7,9} in both embeddings. With this last approach we submitted four additional runs *SVR_GNN_node2vec_pca_sc.csv*, *SVR_GNN_nodeCentralFeats_sc.csv*, *CTR_GNN_node2vec_pca_sc.csv* and *CTR_GNN_nodeCentralFeats_sc.csv*.

For all our experiments we used the same training scheme: 80–20% training and validation split with early stopping based on the validation accuracy. All the other parameters remained the same than in the original work [13].

4 Results

Tables 1 and 2 show the results obtained by our runs on the test set. The ground truth labels of the test set were never released and we just reproduce the results provided by the organizers of the ImageCLEF TB task here. In both cases we included the best run for each subtask that in both cases is from the *UIIP_BioMed* group.

5 Discussion and Conclusions

In this work we applied our latest graph model of the lungs with morphology-based structure to assess the TB severity score and to detect the presence of

³ https://github.com/Tixierae/graph_2D_CNN

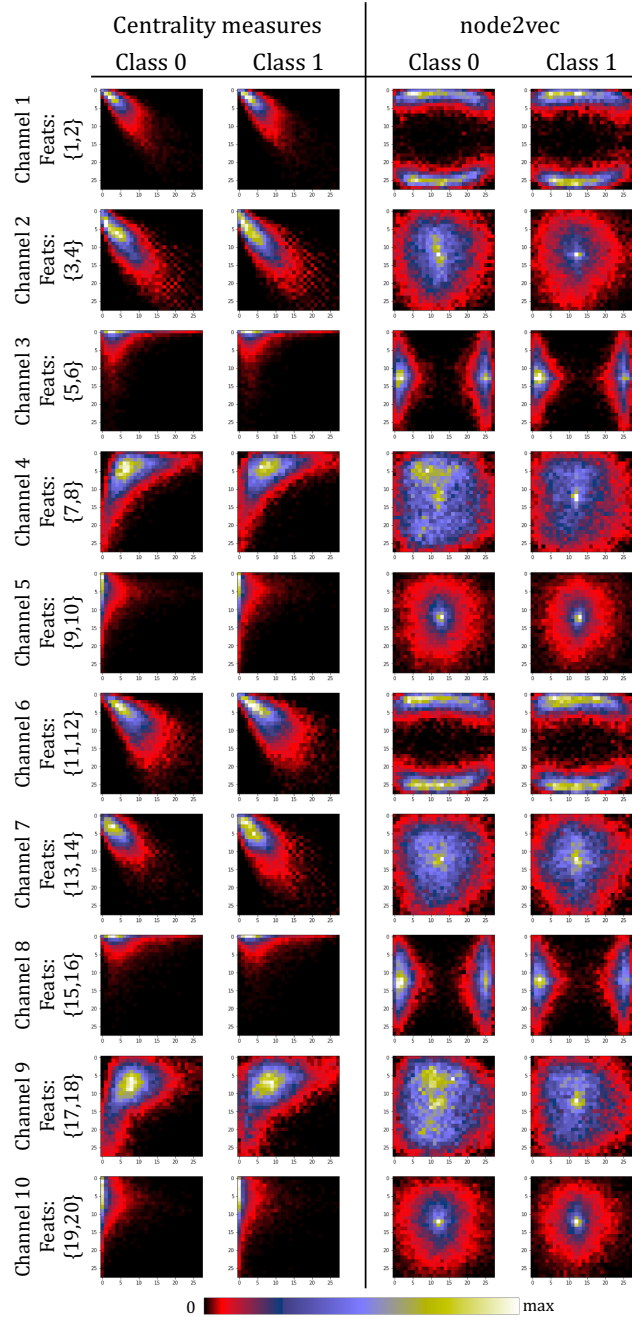


Fig. 2. Average 2D histograms over the training set for each SVR class obtained with each embedding (graph centrality measures and node2vec). In this case, class 0 refers to LOW and class 1 to HIGH TB severity.

Table 1. Results obtained in the SVR subtask for each submitted run.

Group name	Run	AUC	Accuracy	Rank
UIIP_BioMed	SRV_run1_linear.txt	0.7877	0.7179	1
MedGIFT	SVR_SVM.txt	0.7196	0.6410	9
MedGIFT	SVR_GNN_nodeCentralFeats_sc.csv	0.6457	0.6239	26
MedGIFT	SVR_GNN_nodeCentralFeats.csv	0.5496	0.4701	42
MedGIFT	SVR_GNN_node2vec_pca.csv	0.4933	0.4615	47
MedGIFT	SVR_GNN_node2vec_pca_sc.csv	0.4076	0.4274	53

Table 2. Results obtained in the CTR subtask for each submitted run.

Group Name	Run	Mean AUC	Min AUC	Rank
UIIP_BioMed	CTR_run3_pleurisy_as_SegmDiff.txt	0.7968	0.6860	1
MedGIFT	CTR_SVM.txt	0.6795	0.5626	5
MedGIFT	CTR_GNN_nodeCentralFeats_sc.csv	0.5381	0.4299	27
MedGIFT	CTR_GNN_node2vec_pca_sc.csv	0.5261	0.4435	29
MedGIFT	CTR_GNN_nodeCentralFeats.csv	0.5104	0.4140	31
MedGIFT	CTR_GNN_node2vec_pca.csv	0.5016	0.2546	33

six TB-related abnormalities on lung CT scans. Once the graph was built, two approaches were performed to classify the lung graphs. It is worth to remember that their comparison is not straight forward since the graphs contain a different number of nodes and edges. Among the two classification techniques applied, the CNN with the 2D node embeddings performed much worse than the SVM classifier using the lung descriptor vectors (see Sections 3.3 and 3.4). However, among the two node embeddings attempted there is no clear winner. Analyzing the results obtained after reducing the number of channels (runs with *_sc* suffix), there is improvement in both subtasks only when using the embedding based on the graph centrality measures. In any case, the results obtained with the CNN-based classification are really low for both subtasks.

Considering our best approach, we ranked 6th and 3rd at the group level, in the SVR and CTR subtasks respectively, with results clearly above random. This supports the suitability of modeling the lungs as a graph for the proposed subtasks. However, the results show that there is still room for improvement, particularly in the CTR subtask. In the case of the SVR subtask, better results could be obtained by using the provided meta-data or even the predicted CT-findings. For the CTR subtask, we believe that using a classifier that could keep the relations between the presence/absence of the diverse CT-findings will boost our results.

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., Péteri, R., Dicente Cid, Y., Liauchuk, V., Kovalev, V., Klimuk, D., Tarasau, A., Abacha, A.B., Hasan, S.A., Datla, V., Liu, J., Demner-Fushman, D., Dang-Nguyen, D.T., Piras, L., Riegler, M., Tran, M.T., Lux, M., Gurrin, C., Pelka, O., Friedrich, C.M., de Herrera, A.G.S., Garcia, N., Kavallieratou, E., del Blanco, C.R., Rodríguez, C.C., Vasilopoulos, N., Karampidis, K., Chamberlain, J., Clark, A., Campello, A.: ImageCLEF 2019: Multimedia retrieval in medicine, lifelogging, security and nature. In: Experimental IR Meets Multilinguality, Multimodality, and Interaction. Volume 2380 of Proceedings of the 10th International Conference of the CLEF Association (CLEF 2019), Lugano, Switzerland, LNCS Lecture Notes in Computer Science, Springer (September 9-12 2019)
7. Dicente Cid, Y., Liauchuk, V., Klimuk, D., Tarasau, A., Kovalev, V., Müller, H.: Overview of ImageCLEFtuberculosis 2019 - automatic CT-based report generation and tuberculosis severity assessment. In: CLEF2019 Working Notes. Volume 2380 of CEUR Workshop Proceedings., Lugano, Switzerland, CEUR-WS.org <<http://ceur-ws.org/Vol-2380>> (September 9-12 2019)
8. 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)
9. Jeong, Y.J., Lee, K.S.: Pulmonary tuberculosis: up-to-date imaging and management. *American Journal of Roentgenology* **191**(3) (2008) 834–844
10. 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)
11. Dicente Cid, Y., Müller, H.: Texture-based graph model of the lungs for drug resistance detection, tuberculosis type classification, and severity scoring: Participation in ImageCLEF 2018 tuberculosis task. In: CLEF2018 Working Notes. CEUR

Workshop Proceedings, Avignon, France, CEUR-WS.org <<http://ceur-ws.org>> (September 10-14 2018)

12. Dicente Cid, Y., Jimenez-del-Toro, O., Poletti, P.A., Müller, H.: A graph model of the lungs with morphology-based structure for tuberculosis type classification. In Chung, A.C.S., Gee, J.C., Yushkevich, P., Siqu, B., eds.: The 26th International Conference on Information Processing in Medical Imaging (IPMI). (2019)
13. Tixier, A.J.P., Nikolentzos, G., Meladianos, P., Vazirgiannis, M.: Graph classification with 2d convolutional neural networks. arXiv preprint arXiv:1708.02218 (2017)
14. 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
15. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**(11) (November 2012) 2274–2282
16. Schabdach, J., Wells, W., Cho, M., Batmanghelich, K.N.: A likelihood-free approach for characterizing heterogeneous diseases in large-scale studies. In: International Conference on Information Processing in Medical Imaging, Springer (2017) 170–183
17. 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
18. 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
19. Grover, A., Leskovec, J.: node2vec: Scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining