BatmanLab in the ImageCLEF Tuberculosis Task 2017

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

[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 this work we present our participation in the ImageCLEF 2017 tuberculosis task. The task consists of detecting five tuberculosis (TB) types and predicting drug resistance from lung CT (Computed Tomography) volumes. Our approach is based on a previously developed non-parametric method. Tested on CT images of Chronic Obstructive Pulmonary Disease (COPD) patients, it consists of describing each subject as a collection of local feature descriptors embedded in a dissimilarity space. The set of local features was extended for this work adding new 3D texture descriptors. The results shows that our approach is able to characterize several TB types, achieving a Cohen's Kappa coefficient of 0.1533, but does not suit for predicting drug resistance were it only achieved an AUC of 0.5241.

Keywords: non-parametric dissimilarity model, heterogeneous diseases, 3D texture features, tuberculosis

1 Introduction

Tuberculosis (TB) is an infectious disease considered an epidemic by the World Health Organization [1]. Usually, the main organ affected is the lung. Early detection and classification is essential for proper treatment. The diverse TB types and resistance to drugs challenge the appropriate treatment. With this scope, the ImageCLEF[4] initiative proposed a tuberculosis task in 2017 [2]. The aim of the TB task is to detect multi-drug resistance cases early and to identify the different TB types from CT (Computed Tomography) volumes. More details about the the other ImageCLEF 2017 tasks can be found in [3]. ImageCLEF is an evaluation campaign on medical image analysis and retrieval that has been organizing medical tasks since 2004 [4–6].

The Batman lab decided to participate in this task with an approach previously used for Chronic Obstructive Pulmonary Disease (COPD) cases [7]. A first visualization of the training set offered by ImageCLEF suggested different alterations in the lung parenchyma but with non-specific location, similar to

[4] http://www.imageclef.org/ as of 31.05.2017

what happens in COPD patients. The approach consists of extracting regional texture features on quasi-homogeneous regions of the lungs, considering each patient as a bag of words (BoW). However, in contrast to a traditional BoW approach, we structured the words building an underlying distribution between the words of the different patients in a graph embedding. The procedure for building this distribution is not based on any priors and makes it suitable for cases when no pattern is known. It is supposed to differ along patients. Finally, the classification is performed on the dissimilarity space of these distributions.

In the next section we present the details of the ImageCLEF TB datasets, followed by the construction of the dissimilarity space, the classification methods used, and the selection of the runs submitted to the challenge. Section 3 summarizes the results obtained in the task and finally Section 4 exposes the conclusions of our participation.

2 Methods

2.1 ImageCLEF TB Datasets

The ImageCLEF TB task 2017 was divided into two subtasks, both based on lung CT images of patients with tuberculosis. The first task consisted of predicting multi-drug resistant (MDR) patients versus non-drug-resistant cases (DS, drug-sensitive). The dataset was divided into two classes (MDR and DS) with approximately 200 patients in each group. Table 1 contains the exact number of subjects for both training and test sets. The second subtask consisted of a multi-class classification problem. It contained patients with five tuberculosis types. No information about the relation of the classes was suggested by the ImageCLEF organizers, so they were considered independent in this approach. The detailed numbers for this second dataset are specified in Table 2. Moreover, the ImageCLEF organizers also provided automatic lung segmentations extracted with the method introduced in [8]. Our approach used this segmentations to restrict the region of interest to the lung fields.

Table 1. Number of patients per class in the multi-drug resistance dataset.

Class	Train	Test
DS	134	101
MDR	96	113

Table 2. Number of patients per class in the tuberculosis type dataset.

Class	Train	Test
Type 1	140	80
Type 2	120	70
Type 3	100	60
Type 4	80	50
Type 5	60	40

The structure of the challenge was the same for both subtasks. The training CT images and their labels were released 1.5 months before the test set. The test labels were never release but the submitted runs were evaluated. The participants

could submit up to 10 runs containing the predicted labels and the organizers computed the performance measures. The results were made available on the ImageCLEF 2017 TB task web page [5] at the end of the challenge.

2.2 Non-Parametric Dissimilarity Matrix

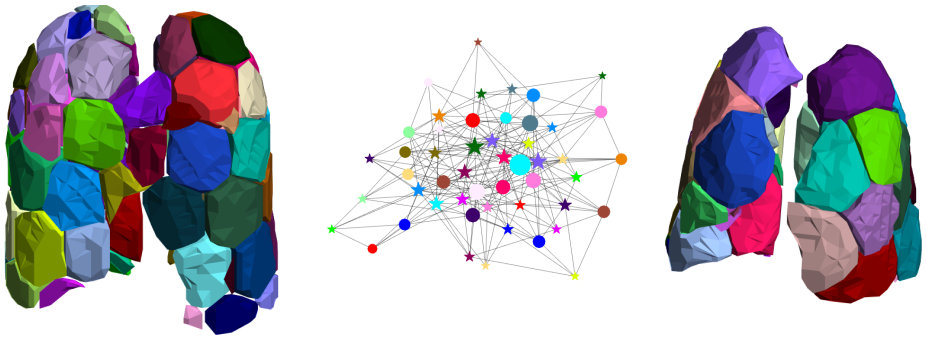

Fig. 1. Left and right corresponds to 3D visualizations of the supervoxelization results for two patients of the dataset. Each color represents a different supervoxel label. The number of supervoxels may vary for each patient as shown in this example. The center of the figure illustrates an example of the KNN graph in the feature space. Each node shape corresponds to a different patient (circle and star in this case). The color of each node is equal to the color of the respective supervoxel label. The size of the node is associated with its popularity (node degree).

The pipeline introduced for COPD detection in [7] was followed. This approach is referred to as the *COPD approach*. The lung area was first divided into homogeneous regions using a supervoxel algorithm [9]. Figure 1 shows a 3D visualization of these regions for two patients of the dataset. Each patient is represented by a set of features extracted from these regions. A total of 4 features were used for this approach. The features are: a 32-bins intensity histogram based on [10] (*Hist*); Haralick features from the Gray-Level Co-occurrence Matrix (GLCM) following [11] (*Haril*); a rotation-invariant histogram of gradients based on the Fourier transform introduced in [12] (*sHOG*); and features based on the locally-oriented Riesz-wavelet transform presented in [13] (*Riesz*). The latter descriptor was not used in the COPD approach. For this work we chose the 1st-order alignment method with 3rd-order Riesz filters and 4 scales.

The next step in the pipeline is to see each patient as a bag of features with unknown density. Then, the Kullback-Leibler (KL) divergence was used to compute the dissimilarities between the densities because presented better results in the COPD case. To compute this measure, the densities were defined with a k-Nearest Neighbor Graph (see Figure 1). The details of this procedure

[5] `http://imageclef.org/2017/tuberculosis` as of 31.05.2017

can be found in the COPD approach article. Since this dissimilarity measure is not a distance, we applied the same technique as in the previous work and we define the similarity kernel between subjects by exponentiating the symmetric KL divergence and projecting the resulting matrix onto a positive semi-definite cone. The projection was done by setting all negative eigenvalues to zero.

2.3 Classification

Once the dissimilarity matrix was generated, several classifiers were tested in the dissimilarity space. The classifiers are: Random Forests (RF), Logistic Regression (LR), Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and the Gradient Boosting Classifier (GBC).

Given that the sample size is small and feature space has a large dimensionality, it is conceivable that the separating hyperplane is sensitive based on the patients in the training set. In other words, the separating hyperplane slightly changes for each fold in the cross-validation procedure. Therefore, patients close to the boundary of the classes may change labels based on the fold but those further away do not. To account for this phenomenon, we resemble the training set and average the probability of the labels (bootstrap).

2.4 Run Selection

Five runs per subtask were finally submitted to the ImageCLEF TB challenge. The runs were selected among several combinations of regional features and classifiers based on the best accuracy obtained in the training sets. Moreover, late fusion was attempted. The best runs and their configuration are shown in Table 3.

Table 3. Run configuration with best results per subtask. The number in the run name corresponds to the accuracy obtained in the training phase.

Subtask	Run name	Regional features	Classifier
MDR	MDR_SuperVx_Hist_FHOG_rf_0.648419	Hist and sHOG	RF
MDR	MDR_SuperVx_FHOG_rf_0.637994	sHOG	RF
MDR	MDR_SuperVx_Reisz_knn_0.624984	Riesz	KNN
MDR	MDR_SuperVx_Hist_Reisz_knn_0.605056	Hist and Riesz	KNN
MDR	MDR_SuperVx_Hist_FHOG_gbc_0.603953	Hist and sHOG	GBC
TBT	TBT_SuperVx_Hist_FHOG_lr_0.414000	Hist and sHOG	LR
TBT	TBT_SuperVx_Hist_FHOG_Reisz_lr_0.426000	Hist, sHOG and Riesz	LR
TBT	TBT_SuperVx_Hist_Reisz_lr_0.426000	Hist and Riesz	LR

The submitted run files were a subset of the best runs and a few fusion approaches of several runs. For the MDR subtask these were:

- *MDR_SuperVx_Hist_FHOG_rf_0.648419.csv*.
- *MDR_SuperVx_FHOG_rf_0.637994.csv*.

- *MDR_submitted_top4_0.656522.csv* (late fusion method). Class probability averaged over the following runs: MDR_SuperVx_Hist_FHOG_rf_0.648419, MDR_SuperVx_FHOG_rf_0.637994, MDR_SuperVx_Reisz_knn_0.624984, and MDR_SuperVx_Hist_Reisz_knn_0.605056.
- *MDR_submitted_top5.csv*: Late fusion method using the top 5 runs, i.e. the same four runs than in MDR_submitted_top4_0.656522.csv plus the run MDR_SuperVx_Hist_FHOG_gbc_0.603953.
- *MDR-submitted_top1.csv*: Same late fusion technique but only using the best run MDR_SuperVx_Hist_FHOG_rf_0.648419. It was used to test the late fusion procedure.

In the case of the TBT subtask, the submitted runs were:

- *TBT_SuperVx_Hist_FHOG_lr_0.414000.csv*.
- *TBT_SuperVx_Hist_FHOG_Reisz_lr_0.426000.csv*
- *TBT_submitted_bootstrap.csv* (bootstrap method).
- *TBT_submitted_top2_0.430000.csv* (late fusion method): Class probability averaged over the following runs: TBT_SuperVx_Hist_FHOG_lr_0.414000, and TBT_SuperVx_Hist_FHOG_Reisz_lr_0.426000.
- *TBT_submitted_top3_0.490000.csv*: Late fusion method like the previous one adding the run TBT_SuperVx_Hist_Reisz_lr_0.426000.csv

3 Results

Tables 4 and 5 provide the results obtained by our method and the best run in each of the subtasks. The selection of the runs was based on the cross-validation accuracy obtained in the training phase. However, the final ranking was established by the Area Under the Curve (AUC) in the MDR subtask, and by the unweighted Cohen's Kappa coefficient (Kappa).

Table 4. Results for our submitted runs in the MDR detection task. The best run of the task is given as reference.

Group Name	Run	Run Type	AUC	ACC	Rank
MedGIFT	MDR_Top1_correct.csv	Automatic	0.5825	0.5164	1
BatmanLab	MDR_submitted_top5.csv	Automatic	0.5241	0.5164	13
BatmanLab	MDR_submitted_top4_0.656522.csv	Automatic	0.5130	0.5024	16
BatmanLab	MDR-submitted_top1.csv	Automatic	0.4899	0.4789	24
BatmanLab	MDR_SuperVx_Hist_FHOG_rf_0.648419.csv	Automatic	0.4899	0.4789	25
BatmanLab	MDR_SuperVx_FHOG_rf_0.637994.csv	Automatic	0.4601	0.4554	27

4 Conclusions

This work presents our method for the ImageCLEF 2017 tuberculosis task. The same approach was applied to the two subtasks obtaining quite different results.

Table 5. Results for our submitted runs in the TB type classification. The best run in the competition is given as reference.

Group Name	Run	Run Type	Kappa	ACC	Rank
SGEast	TBT_resnet_full.txt	Not applicable	0.2438	0.4033	1
BatmanLab	TBT_SuperVx_Hist_FHOG_lr_0.414000.csv	Automatic	0.1533	0.3433	13
BatmanLab	TBT_submitted_bootstrap.csv	Automatic	0.1057	0.3033	18
BatmanLab	TBT_submitted_top3_0.490000.csv	Automatic	0.1057	0.3033	19
BatmanLab	TBT_SuperVx_Hist_FHOG_Reisz_lr_0.426000.csv	Automatic	0.0478	0.2567	20
BatmanLab	TBT_submitted_top2_0.430000.csv	Automatic	0.0437	0.2533	21

For the MDR subtask our method achieved results slightly better than random (both in AUC and accuracy). Although they do not differ much from the best results, they are significantly worse than the results obtained with the COPD dataset. This suggests that our approach was not perfectly suitable for this task. In the case of the TB type classification task, the results are better, being farther from the random performance. The results are not as good as other methods of participants. It can be concluded that the different TB types present differentiable visual patterns. The drug resistance patterns seem hard to identify by structural defects. The fact that we could re-use a framework tested in a different disease obtaining better results than random encourages us to follow this line for characterizing other heterogeneous diseases.

Acknowledgements

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

References

1. World Health Organization, et al.: Global tuberculosis report 2016. (2016)
2. 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)
3. 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)
4. 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

5. 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)
6. 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)
7. 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 (accepted) 30–42
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. Holzer, M., Donner, R.: Over-segmentation of 3d medical image volumes based on monogenic cues. Proceedings of the CVWW **14** (2014)
10. Sorensen, L., Nielsen, M., Lo, P., Ashraf, H., Pedersen, J.H., De Bruijne, M.: Texture-based analysis of copd: a data-driven approach. IEEE transactions on medical imaging **31**(1) (2012) 70–78
11. Vogl, W.D., Prosch, H., Müller-Mang, C., Schmidt-Erfurth, U., Langs, G.: Longitudinal alignment of disease progression in fibrosing interstitial lung disease. In: MICCAI (2). (2014) 97–104
12. 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
13. 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