Texture Analysis from 3D Model and Individual Slice Extraction for Tuberculosis Task

Notebook for ImageCLEF 2018

Md Sajib Ahmed, Md Obaidullah Sk, Mohan Jayatilake,
Teresa Gonçalves, and Luís Rato

Department of Informatics, University of Évora, Portugal
{jack6148,sk.obaidullah,jayatiml}@gmail.com,
{tcg,lmr}@uevora.pt

Abstract. Tuberculosis (TB) is a dreaded bacterial infection that affects human lungs. It has been known to mankind since ancient ages. ImageCLEF 2018 Tuberculosis task proposes three challenging subtasks based on Computed Tomography (CT) scan images of patients' lungs: multi-drug resistance (MDR) detection, tuberculosis type (TBT) classification and severity scoring (SVR). In this work, two different approaches are presented: 3D Modeling and Slice Extraction. Several feature descriptors were calculated (mean and higher order moments, fractal dimension and texture analysis based measures) from CT scans and different classifiers were tested. The 3D Modeling approach uses six features (Mean, Skewness, Kurtosis, Fractal Dimension, Homogeneity, and Energy) and Slice Extraction approach calculates a vector of 96 features (based on Mean, Correlation, Contrast, Homogeneity, Energy, and Entropy). In accordance with the ranking given by the organizers, systems submitted were ranked 1^{st} for multi-drug resistance detection, 5^{th} for tuberculosis type classification and 3^{rd} tuberculosis severity scoring.

Keywords: Tuberculosis, Computed Tomography, Classification, Texture Analysis, Fractal Dimension, Machine Learning

1 Introduction

Tuberculosis (TB) is an infection caused by a bacteria named Mycobacterium tuberculosis. This bacteria generally attacks the lungs but sometimes it can damage other parts of the body. TB spreads through the air when an infected person coughs, sneezes or talks. The World Health Organization (WHO) report states that TB is the ninth leading cause of death worldwide [9].

The biggest problem a TB patient can face is that the organisms become resistant to two or more of the standard drugs. In contrast to drug sensitive (DS) tuberculosis, its multi-drug resistant (MDR) form is much more difficult and costly to recover from. Thus, early identification of the drug resistance (DR) status is of great importance for an effective treatment. The most frequently used

methods for DR detection are either costly or take a long time (up to several months), hence, there is an urgent need for fast, precise and cheap techniques. One of the possible techniques is to analyze CT scan images of patient's lungs to get insight about the specificity of the disease, including if it's a multi-drug resistance (MDR) form, what's the tuberculosis type (TBT) and score its severity (SVR).

ImageCLEF (Image Retrieval and Analysis Evaluation Campaign of the Cross Language-Evaluation Forum) organizes challenges since 2003 [7] and medical image analysis and retrieval tasks since 2004 [5]. This work presents an approach to tackle ImageCLEF 2018 Tuberculosis task [1], which is on its second edition.

The rest of the paper is organized as follows: Section 2 describes the Theory and Approaches taken and Section 3 introduces the Experiments and Submitted Runs. Finally, Section 4 concludes the paper.

2 Methodology

To tackle ImageCLEF 2018 tuberculosis task [1], two different approaches were tested: one based on 3D modeling and another based on slice extraction information. Next subsections present the underlying theory and techniques and parameters of each proposed approach.

2.1 Theory

This work is based on the mean and higher order moments (skewness and kurtosis) and texture analysis based features to classify tuberculosis images. This subsection introduces these measures and the used parameters to obtain the input features.

Mean and Higher Order Moments. The skewness (3^{rd} standardized moment) and kurtosis (4^{th} standardized moment) are measure descriptors of the shape of a probability distribution. Skewness is a measure of the asymmetry of the probability distribution. Kurtosis is related to the tails of the distribution; infrequent extreme deviations (or outliers) result in higher values of kurtosis.

Fractal Dimension. According to Peleg *et. al.* [8], the image within the ROI can be treated as a hilly terrain with its height being proportional to the gray level of the image. The fractal dimension of the surface can then be estimated.

Texture Features. Gray level co-occurrence matrix (GLCM), also known as gray-level spatial dependence matrix, characterize the texture of an image by calculating how often pairs of the pixel with specific values and in a specified spatial relationship occur in an image. From this matrix, several statistical measures can then be extracted [4][6]. Contrast, correlation, homogeneity, energy,

and entropy were used in this work. Contrast returns a value after measuring the intensity contrast between a pixel and its neighbors over the entire image; it is also known as variance or inertia. Correlation returns a value after measuring how correlated a pixel is to its neighbors over the entire image; it ranges between -1 and 1: 1 or -1 for a perfectly positively or negatively correlated image and NaN for a constant image. Homogeneity measures the similarity of pixels in the matrix; a diagonal GLMC gives homogeneity of 1 and the value becomes large if local textures only have minimal changes. Energy is a measure of the extent of pixel pair repetitions; when pixels are very similar, the energy value will be large. Entropy is a statistical measure of randomness of an image.

2.2 Preprocessing

All CT images were in NIFTI (Neuroimaging Informatics Technology Initiative); the masks provided [2] was also in NIFTI format. First slides are extracted and, for each slice, an ROI is extracted based on the given mask. Figures 1 and 2 show a slice of Tuberculosis CT scan image and the corresponding mask, respectively.

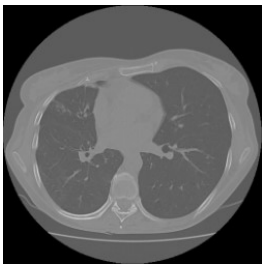

Fig. 1. A slice of Tuberculosis CT scan

Fig. 2. Provided Mask image of the lung

2.3 3D Modeling Approach

The block diagram of the 3D Modeling Approach is presented in Figure 3. In this approach, 2D data is converted to 3D data and a threshold value of 100000 is defined to extract only bronchi within the lung resulting in a 3D model of the bronchi. With this model, the statistical features of mean and higher order moments (skewness and kurtosis), fractal dimension, homogeneity and energy are computed and several ML classifiers are applied and their performance is analyzed.

The whole 3D bronchi ROI computed mean is a weighted average of slice values (using all pixel values within the multi-slice ROIs encompassing the

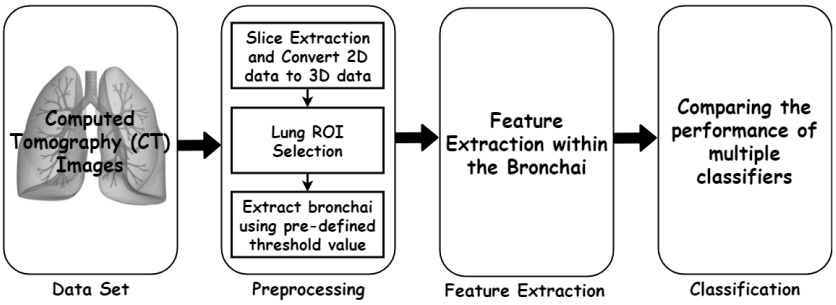

Fig. 3. Overview of the 3D modeling approach

bronchi).Skewness, kurtosis and fractal dimension were calculated within the 3D bronchi ROI pixel value distribution.

Energy and homogeneity were calculated with the help of the GLCM obtained for each 2D slice with the direction of 0° and offset equal to 1 (pixel to the right). Then the average was calculated over all slices.

2.4 Slice Extraction Approach

In this approach, a focus is taken on individual slices. Observing each slice pattern, a threshold value of 15000 is chosen to ensure that no slices with meaningful information would be missed. Here, meaningful information corresponds to some dots being present in the ROI. Then, texture analysis features were extracted for each selected ROI and an average over slice-wise values was calculated. Finally, the computed feature vector was used as input to several ML classifiers and their performance was analyzed. The block diagram is presented in Figure 4.

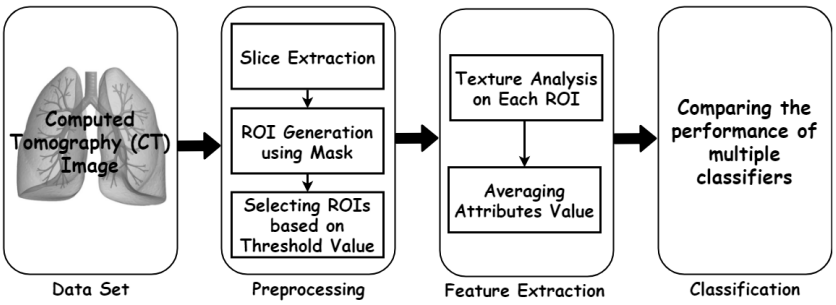

Fig. 4. Overview of the slice extraction approach

For each ROI, 16 offsets (4-pixel distances: 1 to 4, and 4 directions: 0, 45, 90 and 135 degrees) are considered for GLCM generation. Then, for each of the 16

co-occurrence matrices, 6 features were computed (mean, contrast, correlation, homogeneity, energy and entropy) resulting in a $16^*6 = 96$ dimensional feature vector. Finally, averaging feature values for all slices was done to generate the final feature vector.

2.5 Classifiers

For each of the previous approaches, several machine learning algorithms were used to build classification models. They were: Linear Discriminant Analysis (LDA), Logistic Regression (L), Multilayer Perceptron (MLP), Simple Logistic (SL), Sequential Minimal Optimization (SMO), Logistic Model Trees (LMT), Random Forest (RF) and Random Tree (RT). A simple voting scheme (Vote) was also tested.

3 Experiments and Submitted Runs

As already mentioned, this year challenge has 3 different subtasks: multi-drug resistant (MDR) detection, tuberculosis type (TBT) classification and severity scoring (SVR). This section describes the datasets, introduces the evaluation measures and system configuration and presents the results of the submitted runs for each subtask.

3.1 Datasets Description

Table 1 presents the number of subjects for the training and test sets, for the MDR sub-task ((multi-drug resistant vs. drug-sensitive patients).

Table 1. MDR dataset: mumber of patients per class

Patients	Training set	Test set
Drug sensitive	134	101
Multi-drug resistant	125	113
Total	259	214

The second sub-task is a multi-class classification problem with five tuberculosis types: infiltrative, focal, tuberculoma, miliary, and fibro-cavernous. No information about the relation between these classes is given. Table 2 presents the number of patients and CT scans for the training and test sets; values in parenthesis are the number of CT scans which means that some patients have more than one CT scan.

The third subtask aims at scoring TB severity (value between 1 to 5); the severity level, given as 'low' or 'high') is also available. Table 3 presents the details for training and test sets.

Table 2. TBT dataset: number of patients per class

Patients	Training set	Test set
Infiltrative	228 (376)	89 (176)
Focal	210 (273)	80 (115)
Tuberculoma	100 (154)	60 (86)
Miliary	79 (106)	50 (71)
Fibro-cavernous	60 (99)	38 (57)
Total	677 (1008)	317 (505)

Table 3. SVR dataset: number of patients per class

Patients	Training set	Test set
Low severity	90	62
High severity	80	47
Total	170	109

3.2 Evaluation Metrics and System Configuration

For the MDR subtask, the performance of the system was measured using the Area Under the Curve (AUC) of the Receiver Operator Characteristic (ROC); the ROC curve is created by plotting the true positive rate against the false positive rate. For the TBT subtask, Cohen's Kappa coefficient (Kappa) was used to measure the system performance; Kappa statistic measures inter-rater agreement for categorical items. For the SVR subtask, the performance is measured through the Root Means Square Error (RMSE); RMSE represents the sample standard deviation of the differences between observed and predicted values.

To evaluate the models generated by the ML algorithms, a stratified K-fold cross-validation approach was used (K=5). Regarding resources, all experiments were carried out using MATLAB 2017b software and Weka 3.8.1 toolkit [3] in a system with 3.5 GHz CPU, 8 GB RAM.

3.3 Top-5 Submitted Runs for each Subtask

From the total 30 runs allowed for submission (10 runs for each subtask), 23 runs were uploaded. The runs were selected based on best AUC, Kappa coefficient and RMSE measures (obtained with the cross-validation procedure over the training dataset), with the machine learning algorithms fine-tuned experimentally.

Tables 4, 5 and 6 show the configuration of the best 5 runs for MDR, TBT and SVR subtasks, respectively.

3.4 Results on Test Set

Tables 7, 8 and 9 show the performance values and the rank for the best 3 submitted runs for MDR, TB type, and SVR subtasks, respectively.

Table 4. MDR subtask: top five submitted runs.

Method	**Classifier**	**Run Name**
3D modeling	Sl	MDR-Run-06-Mohan
3D modeling	Vote (LDA, SMO)	MDR-Run-08-Mohan
Slice extraction	SL	MDR-Run-09-Sk
3D modeling + Slice extraction	Vote (LDA, SL)	MDR-Run-10-Mix
Slice extraction	LDA	MDR-Run-07-Sk

Table 5. TBT subtask: top five submitted runs.

Method	**Classifier**	**Run Name**
3D modeling	RF	TBT-Run-02-Mohan
3D modeling	RF	TBT-Run-05-Mohan
3D modeling	RF	TBT-Run-03-Mohan
3D modeling + Slice extraction	RF	TBT-Run-06-Mix
3D modeling	Vote (RF, LMT)	TBT-Run-04-Mohan

Table 6. SVR subtask: top five submitted runs.

Method	**Classifier**	**Run Name**
3D modeling	MLP	SVR-Run-07-Mohan
3D modeling	MLP	SVR-Run-03-Mohan
3D modeling	Vote(MLP,Sl)	SVR-Run-06-Mohan
3D modeling	RF	SVR-Run-02-Mix
3D modeling	RF	SVR-Run-05-Mohan

Table 7. MDR subtask: AUC and accuracy (ACC) of the test set.

Run	AUC	ACC	Rank
MDR-Run-06-Mohan	0.6178	0.5593	1
MDR-Run-08-Mohan	0.6065	0.5424	3
MDR-Run-09-Sk	0.5921	0.5763	4

Table 8. TBT subtask: Kappa and accuracy (ACC) of the test set.

Run	Kappa	ACC	Rank
TBT-Run-02-Mohan	0.1664	0.3785	5
TBT-Run-05-Mohan	0.1621	0.3754	7
TBT-Run-03-Mohan	0.1335	0.3502	14

Table 9. SVR subtask: RMSE and accuracy (ACC) over the test set.

Run	RMSE	ACC	Rank
SVR-Run-07-Mohan	0.8883	0.6239	3
SVR-Run-03-Mohan	1.0091	0.6371	17
SVR-Run-06-Mohan	1.0536	0.6356	21

As can be seen, very competitive results were achieved with six setups reaching the top 10 ranking. For MDR, TBT and SVR subtasks, a 1^{st}, 5^{th} and 3^{rd} positions were obtained, respectively.

The best results were obtained with the 3D modeling approach. For the MDR subtask it uses the Simple Logistic classifier algorithm; for the TBT subtask, it uses the Random Forest algorithm (with the following parameters: `numFeatures=20`, `numIterations=1500` and `seed=20`); for the SVR subtask, it uses the multi-layer perception algorithm (MLP) (with parameter `trainingTime=100`).

4 Conclusion

This work presents an approach to building CT scans image classifiers for 3 different subtasks: multi-drug resistance (MDR) detection, tuberculosis type (TBT) classification and severity scoring (SVR). Two different approaches were tested: one based on 3D modeling and another based on slice extraction. Different measures were extracted, namely, the Mean, Skewness, Kurtosis, Fractal Dimension, and texture analysis ones extracted from the GLCM (Mean, Correlation, Contrast, Homogeneity, Energy, and Entropy). Their individual and combined performances were tested using different machine learning classifiers.

Though in terms of accuracy both approaches were very competitive, using the TB task performance measures (AUC for MDR, Kappa for TBT and RMSE for SVR subtasks), 3D modeling features are more promising.

As future, we intend to use the patient clinical information to improve the overall performance of three tasks.

5 Acknowledgment

The authors thank the LEADER and gLINK Erasmus Mundus projects for supporting this research.

References

1. 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, CEUR-WS.org <http://ceur-ws.org>, Avignon, France (September 10-14 2018)
2. Dicente Cid, Y., Jiménez del Toro, O.A., Depeursinge, A., Müller, H.: Efficient and fully automatic segmentation of the lungs in ct volumes. In: Goksel, O., Jiménez del Toro, O.A., Foncubierta-Rodríguez, A., Müller, H. (eds.) Proceedings of the VISCERAL Anatomy Grand Challenge at the 2015 IEEE ISBI. pp. 31–35. CEUR Workshop Proceedings, CEUR-WS (May 2015)
3. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: an update. SIGKDD Explorations 11(1), 10–18 (2009)

4. Haralick, R.M., Shanmugam, K., et al.: Textural features for image classification. IEEE Transactions on systems, man, and cybernetics SMC-3(6), 610–621 (1973)
5. Kalpathy-Cramer, J., de Herrera, A.G.S., Demner-Fushman, D., Antani, S., Bedrick, S., Müller, H.: Evaluating performance of biomedical image retrieval systems—an overview of the medical image retrieval task at imageclef 2004–2013. Computerized Medical Imaging and Graphics 39, 55–61 (2015)
6. Materka, A., Strzelecki, M., et al.: Texture analysis methods–a review. Technical university of lodz, institute of electronics, COST B11 report, Brussels pp. 9–11 (1998)
7. Müller, H., Clough, P., Deselaers, T., Caputo, B., CLEF, I.: Experimental evaluation in visual information retrieval. The Information Retrieval Series 32, 1–554 (2010)
8. Peleg, S., Naor, J., Hartley, R., Avnir, D.: Multiple resolution texture analysis and classification. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-6(4), 518–523 (1984)
9. World Health Organization: Global Tuberculosis Report 2017 (2017), `http://www.who.int/tb/publications/global_report/en/`