Texture Analysis from 3D Model and Individual Slice Extraction for Tuberculosis MDR Detection, Type Classification and Severity Scoring

Notebook for ImageCLEF 2018

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

Abstract. Tuberculosis (TB) is a dreaded bacterial infection that affects human lungs. It has been known to mankind since ancient ages. Tuberculosis ImageCLEF 2018 proposes a set of tasks based on Computed Tomography (CT) scan images of patients' lungs. They are: multi-drug resistance (MDR) detection, tuberculosis type (TBT) classification and severity scoring (SVR). In this work, two different methods are presented to solve these problems. Texture analysis based methods (3D Modeling and Slice extraction approach) were used to generate feature values from CT scans and different classifiers were tested. 3D Modeling approach calculates seven statistical features of Mean, Skewness, Kurtosis, Homogeneity, Energy, Entropy and Fractal Dimension. And Slice extraction approach calculates 96 dimensional feature vector based on Contrast, Correlation, Energy, Homogeneity, Entropy and Mean. In accordance with the ranking given by the organizers, this approach was 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 which 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. From World Health Organization (WHO) report, tuberculosis is one of the top ten courses of death worldwide [12].

The greatest problem that can happen to a TB patient 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 requirement for fast, precise and cheap techniques. One possible technique is the automatic analysis of CT scan images of patient's lungs to help characterize tuberculosis.

ImageCLEF (the Image Retrieval and Analysis Evaluation Campaign of the Cross Language-Evaluation Forum) has organized challenges on image classification and retrieval since 2003 [11]. The 2018 edition [8] proposes 3 main tasks (and a pilot task), one being related to the analysis of tuberculosis from lung CT images [4]. This task includes three independent subtasks: multi-drug resistance (MDR) tuberculosis detection, tuberculosis type (TBT) classification and severity scoring (SVR). This work presents the University of Évora approach to tackle these subtasks.

The rest of the paper is organized as follows: Section 2 describes the Methodology (theory of methods, explanation of 3D modeling approach and slice extraction approach) and Section 3 introduces the Experiments and Submitted Runs (dataset description, evaluation metrics, system configuration, top-5 submitted runs for each subtask and results on test dataset). Finally, Section 4 concludes the paper.

2 Methodology

To tackle ImageCLEF 2018 tuberculosis task [4], 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 of the proposed approach.

2.1 Theory

The present work is based on mean, higher order moments (skewness and kurtosis) and texture analysis to classify tuberculosis images. We have proposed two basic models for the present ImageCLEF 2018 [8] competition. In this section, we present the theoretical basis for the estimation of mean and higher order moments, fractal dimension and GLCM-based texture analysis (energy and homogeneity) techniques that used in both models.

Mean and Higher Order Moments. The mean (1^{st} moment) , skewness (3^{rd} moment) and kurtosis (4^{th} moment) are measures of asymmetry and normalized form of the fourth central moment of the whole bronchi ROI respectively. The mean; found by adding the pixel values and dividing by the number of pixels. If the skewness is negative, the pixel values are spread out more to the left of the mean than to the right; if skewness is positive, the pixel values are spread out

more to the right. Kurtosis indicates the degree of peakedness of the values; it is based on the size of the tail of the pixel value distribution.

The skewness and kurtosis of the pixel value within the whole bronchi can be determined using Equation (1). Here, p_i and i represent the signal intensity in i^{th} pixel and the number of pixels within the bronchi the regions of interest (ROI) respectively. Further, p and $f(p_i)$ characterize the mean of the pixel values within the bronchi and the probability of the pixel value falling within a specific value given by the range of this variable's density, respectively.

$$n^{th}moment = \sum_{i} (p_i - p)^n f(p_i)$$
(1)

Fractal Dimension. According to Peleg *et. al.* [13], the image within the whole ROI is treated as a hilly terrain and its height is proportional to the gray level of the image. Points with distance ε from the surface on both sides create a blanket, whose thickness is 2ε . The area of the blanket can be estimated as [2,5]:

$$A(\varepsilon) = F\varepsilon^{2-D} \tag{2}$$

where F is a constant and D is the fractal dimension of the surface. When the $log(A(\varepsilon))$ is plotted against $log(\varepsilon)$ a straight line is obtained with a slope equal to 2-D, which gives an estimation of the fractal dimension (Equation(3)).

$$log(A(\varepsilon)) = log(F) + (2 - D)log(\varepsilon)$$
(3)

Gray Level Co-occurrence Matrix. Energy, homogeneity, contrast, correlation and entropy are statistical texture features. These features are based on the normalized gray-level co-occurrence matrix (GLCM); for a 2D bronchi map f, the co-occurrence matrix $M_{f,\delta}(k,l)$ represents the joint probability occurrence of pixel pairs (with a defined spatial relationship) having gray level values k and l, for a given spatial offset $\delta = (\delta x, \delta y)$ between the pair [7,14]. $M_{f,\delta}(k,l)$ is defined by Equation (4)

$$M_{f,\delta}(k,l) = \begin{cases} \sum_{x=1^n} \sum_{y=1^n} 1, & \text{if } f(x,y) = k \text{ and } f(x+\delta x, y+\delta y) = l \\ \sum_{x=1^n} \sum_{y=1^n} 0, & \text{otherwise} \end{cases}$$
(4)

The co-occurrence matrix $M_{f,\delta}(k,l)$ has dimension $n \times n$, where *n* is the number of gray levels in *f*. The GLCM accounts for the spatial inter-dependency or co-occurrence of two pixels at specific relative positions. Co-occurrence matrices are calculated for the directions of 0°, 45°, 90°, and 135° and the average matrix over all offsets can be used [1,10]. In this study, the 2D formulation with 8-connexity (computed with 8 different offsets) was used. On the basis of this matrix, the second-order statistical features of energy and homogeneity [7] for

each image slice were derived. Then, the average of energy and homogeneity over all slices was estimated.

Energy is defined as the measure of the extent of pixel pair repetitions in the matrix M and can be estimated using Equation (5). When pixels are very similar, the energy value will be large.

$$Energy = \sum_{k=1^n} \sum_{l=1^n} M_{f,\delta}(k,l)^2$$
(5)

Homogeneity is the statistical measure of the similarity of pixels in the matrix M and can be estimated using Equation (6). A diagonal gray level co-occurrence matrix gives homogeneity of 1 and the value becomes large if local textures only have minimal changes.

$$Homogeneity = \sum_{k=1^n} \sum_{l=1^n} \frac{M_{f,\delta}(k,l)}{1+|k-l|}$$
(6)

Contrast returns a value after measuring the intensity contrast between a pixel and its neighbors over the entire image. Contrast is also known as variance or inertia and is given by Equation (7).

$$Contrast = \sum_{k=1^{n}} \sum_{l=1^{n}} |k-l|^2 M_{f,\delta}(k,l)$$
(7)

Correlation returns a value after measuring how correlated a pixel is to its neighbors over the entire image. Correlation ranges between -1 and 1; it's 1 or -1 for a perfectly positively or negatively correlated image and is NaN for a constant image. It can be calculated through the Equation (8)

$$Correlation = \sum_{k=1^n} \sum_{l=1^n} \frac{(k-\mu k)(l-\mu l)M_{f,\delta}(k,l)}{\sigma_k \sigma_l}$$
(8)

Entropy returns a scalar value representing the entropy of a grayscale image. Entropy is a statistical measure of randomness that can be used to characterize the texture of an image and is defined by Equation (9).

$$Entropy = \sum_{k=1^n} \sum_{l=1^n} -ln(M_{f,\delta}(k,l)).M_{f,\delta}(k,l)$$
(9)

2.2 3D Modeling Approach

The first method is based on 3D modeling and further extraction of texture patterns. The block diagram is presented in Figure 1. First, the input data set is pre-processed: slices are extracted and 2D data is converted to 3D data by selecting lung ROIs (using masks [3] given). Then, bronchi are extracted using a pre-defined threshold value. The texture analysis is performed within the bronchi. Finally, a feature vector is computed, followed by the application of multiple classifiers and their performance analysis. The main parts of the system are described below.



Fig. 1. Overview of the 3D modeling approach

Data Acquisition and ROI Selection. All CT images were in NIFTI (Neuroimaging Informatics Technology Initiative) format and 3D images were extracted. The masks of each subject were also provided in NIFTI format. The masks were applied to extract the 3D lung images of each subject. Then, a threshold value was defined in order to extract only the bronchi within the lung. Based on the threshold, ROIs were selected (see Figure 1). Then, the statistical features of mean, higher order moments (skewness and kurtosis), homogeneity, energy, and entropy of the whole bronchi pixel value distribution of each individual subject were calculated.

Estimation of Mean. For quantitative data analysis, the ROIs on multiple image slices were extracted within the bronchi. Then the whole 3D bronchi ROI mean of signal intensity pixel values was computed by weighted (by pixel number) average of slice values (using all pixel values within the multi-slice ROIs encompassing the bronchi). The histograms of signal intensity were plotted with the mean values.

Higher Order Moments and Fractal Dimension Analysis. The whole bronchi pixel value distribution higher order moments (skewness and kurtosis) and fractal dimension of pixel values were calculated within the bronchi ROI using the Equations (1) and (3) for each image slice covering the entire lung. The fractal dimension was estimated after linear fitting of $\log A(\varepsilon)$ vs. $log(A(\varepsilon))$.

Homogeneity and Energy Analysis. Image texture analysis was performed on pixel values within the bronchi at all 2D slices within the 3D ROI. The normalized GLCM was calculated for each 2D slice, and based on the GLCM obtained, the two feature measures of energy and homogeneity were computed (Equations (5) and (6)).

2.3 Slice Extraction Approach

In this method, a focus was taken on individual slices. Using a threshold and performing averages over attributes computed on each slice the final feature values for the particular image was obtained. First the input dataset is preprocessed: slice extraction, ROI generation using a mask and ROIs selection based on threshold values are done; then, using Texture analysis on each ROI features are extracted by averaging slice-wise values; finally, a feature vector is computed, followed by the application of multiple classifiers and their performance analysis. The block diagram is presented in Figure 2.



Fig. 2. Overview of the slice extraction approach

Individual Slice Consideration. Since the images were in NIFTI format (where a single image has several slices), so in this approach, we have taken individual slices for further processing and finally performed a weighted average among all slices. Figure 3 shows a slice of Tuberculosis CT scan image.

After slice extraction, an ROI for each slice was defined based on the given provided mask [3]. Figure 4 shows a mask for image depicted in Figure 3.

Observing each slice pattern, a threshold value of 15000 (pixel area) was chosen to ensure that no slices with meaningful information would be missed. Here, meaningful information corresponds to some dots being present in the ROI.

Feature Extraction. There are two parts for the feature extraction task: Texture Analysis on each ROI and Averaging Attribute Value.

For texture analysis, the Gray Level Co-occurrence Matrix (GLCM) [7] based on texture features from the selected slices was computed. For each ROI, 96dimensional features were extracted.

For present work, 16 offsets are considered for generation of GLCM as shown below. The value 1, 2, 3, 4 represent the pixel distance from the point of interest in each of the four directions: 0, 45, 90 and 135 degrees. So altogether 16



Fig. 3. A slice of Tuberculosis CT scan



Fig. 4. Provided Mask image of the lung

offsets are available which eventually generates 16 co-occurrence matrices. Now, from each of this co-occurrence matrix, 06 features are computed namely: Contrast, Correlation, Energy, Homogeneity, Entropy and Mean. Finally, 16*6 = 96 dimensional feature vector is generated.

- (a) (0 1); (0 2); (0 3); (0 4); \rightarrow at 0 degree direction
- (b) (-1 1); (-2 2); (-3 3); (-4 4); \rightarrow at 45 degree direction
- (c) $(-1 \ 0); (-2 \ 0); (-3 \ 0); (-4 \ 0); \rightarrow \text{ at } 90 \text{ degree direction}$
- (d) $(-1 1); (-2 2); (-3 3); (-4 4) \rightarrow \text{ at } 135 \text{ degree direction}$

Averaging attribute values for all slices was done to generate the final feature vector, obtaining the correspondent 96 average feature values (for contrast, correlation, energy, homogeneity, entropy and mean).

2.4 Classifiers

For each of the previous approaches, several machine learning algorithms were used to build classification models. The tested classifiers 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) also experimented.

3 Experiment and Submitted Runs

As already mentioned, there are 3 different sub-tasks: multi-drug resistant (MDR) detection, tuberculosis type (TBT) classification and severity scoring (SVR). This section describes the datasets and introduces the evaluation measures and system configuration.

3.1 Datasets Description

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

Patients	Training set	Test set
DS	134	101
MDR	125	113
Total	259	214

Table 1. MDR dataset: mumber of patients per class

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 this the classes are given. Table 2 presents the details for training and test sets. For training set, the total number of patients 677, but 1008 chest CT scans of TB patients along with the TB type. And some patients include more than one scan. Similar to the test set, patients 317 but chest CT scans image 505.

 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)

The third sub-task consisted of chest CT scans of TB patients with the corresponding severity score (1 to 5) and the severity level ("low" and "high"). Table 3 presents the details for training and test sets.

Table 3. SVR dataset: number of patients p	ber cl	ass
--	--------	-----

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

Moreover, lung segmentation extracted automatically [3] were also provided. In our work, we used this segmentation to restrict the region of interest of the lungs. Figure 5 shows sample slices of the Computerized Tomography (CT) images with segmented lungs.



Fig. 5. Sample slices of CT images with segmented lungs [9]

3.2 Evaluation Metrics and System Configuration

For MDR task 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. Cohen's Kappa coefficient (Kappa) is used for measuring the system performance of TBT task. Kappa statistic is measure inter-rater agreement for qualitative (categorical) items. The performance of the system of SVR task was used Root Means Square Error (RMSE). RMSE presents the sample standard deviation of the differences between observed values and predicted values.

To evaluate the system, a stratified k-fold cross-validation approach was used. The value of k was chosen experimentally. Regarding resources, all experiments were carried out using MATLAB 2017b software and Weka 3.8.1 toolkit [6] 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 for the ImageCLEF 2018 Tuberculosis classification challenge (10 runs for each task), 23 runs were uploaded. The runs were selected based on best accuracy, AUC, Kappa coefficient and RMSE measures (calculated with a 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 tasks, respectively.

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
$3D \mod H + Slice extraction$	Vote (LDA, SL)	MDR-Run-10-Mix
Slice extraction	LDA	MDR-Run-07-Sk

Table 4. MDR subtask: top five submitted runs.

Table 5. MDR subtask: top five submitted runs.

Method	Classifier	Run Name
3D modeling	RF	TBT-Run-02-Mohan
3D modeling	\mathbf{RF}	TBT-Run-05-Mohan
3D modeling	\mathbf{RF}	TBT-Run-03-Mohan
$3D \mod H + Slice extraction$	\mathbf{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	\mathbf{RF}	SVR-Run-02-Mix
3D modeling	\mathbf{RF}	SVR-Run-05-Mohan

3.4 Results on Test Set

Tables 7, 8 and 9 shows the performance measures and rank for the best 3 submitted runs for MDR, TBT and SVR subtasks, respectively. As can be seen, very competitive results were achieved for MDR subtask and for TBT subtask positions 5 and 7 of the top 10 were obtained. For SVR subtask a 3rd place was reached.

Also, the best run over the test set was also the best obtained for the training set. For the MDR subtask, it uses the 3D modeling approach (Section 2.2) with patient personal data and the Simple Logistic (SL) classifier algorithm.

For TBT subtask, it uses the 3D modeling approach (Section 2.2) with the machine learning Random Forest (RF) algorithm (with the following parameters: numFeatures=20, numIterations=1500 and seed=20) and for the SVR subtask, the 3D modeling approach with the multi-layer perception algorithm (MLP) with parameter trainingTime=100.

 Table 7. MDR subtask: performance on the test set based on the Area Under the Curve (AUC)

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: performance on the test set based on Cohen's Kappa coefficient (Kappa)

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: performance on the test set based on Root Mean Square Error (RMSE)

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

4 Conclusion

This work presents texture analysis based approaches to categorize different TB images for 3 different problems: multi-drug resistance (MDR) detection, tuberculosis type (TBT) classification and severity scoring (SVR). Two different texture analyses, one based on 3D modeling and another based on slice extraction were proposed. Their individual and combined performances were tested using different machine learning classifiers.

Though in terms of the accuracy both approaches are very competitive, using the TB tasks ImageCLEF 2018 performance measures (AUC for MDR, Kappa for TBT and RMSE for SVR subtasks), using 3D modeling features seems more promising.

In future, we will 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

- M.-C. Asselin, J. P. O'Connor, R. Boellaard, N. A. Thacker, and A. Jackson. Quantifying heterogeneity in human tumours using mri and pet. *European journal* of cancer, 48(4):447–455, 2012.
- 2. J. Baish and R. Jain. Fractals and cancer. Cancer Research, 60(14):104, 2000.
- Y. Dicente Cid, O. A. Jiménez del Toro, A. Depeursinge, and H. Müller. Efficient and fully automatic segmentation of the lungs in ct volumes. In O. Goksel, O. A. Jiménez del Toro, A. Foncubierta-Rodríguez, and H. Müller, editors, *Proceedings of the VISCERAL Anatomy Grand Challenge at the 2015 IEEE ISBI*, CEUR Workshop Proceedings, pages 31–35. CEUR-WS, May 2015.
- 4. Y. Dicente Cid, V. Liauchuk, V. Kovalev, , and H. Müller. Overview of ImageCLEFtuberculosis 2018 - detecting multi-drug resistance, classifying tuberculosis type, and assessing severity score. In *CLEF2018 Working Notes*, CEUR Workshop Proceedings, Avignon, France, September 10-14 2018. CEUR-WS.org http://ceur-ws.org>.
- 5. B. G and P. F. Fractals and pathology. *Journal of Biostatistics and Biometric Applications*, 1(1):104, 2015.
- M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The WEKA data mining software: an update. *SIGKDD Explorations*, 11(1):10–18, 2009.
- R. M. Haralick, K. Shanmugam, et al. Textural features for image classification. IEEE Transactions on systems, man, and cybernetics, SMC-3(6):610-621, 1973.
- 8. B. Ionescu, H. Müller, M. Villegas, A. G. S. de Herrera, C. Eickhoff, V. Andrearczyk, Y. D. Cid, V. Liauchuk, V. Kovalev, S. A. Hasan, Y. Ling, O. Farri, J. Liu, M. Lungren, D.-T. Dang-Nguyen, L. Piras, M. Riegler, L. Zhou, M. Lux, and C. Gurrin. 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, September 10-14 2018. LNCS Lecture Notes in Computer Science, Springer.
- V. Liauchuk and V. Kovalev. Imageclef 2017: Supervoxels and co-occurrence for tuberculosis ct image classification. In CLEF2017 Working Notes. CEUR Workshop Proceedings, Dublin, Ireland, CEUR-WS. org; http://ceur-ws. org¿(September 11-14 2017), 2017.
- A. Materka, M. Strzelecki, et al. Texture analysis methods-a review. *Technical university of lodz, institute of electronics, COST B11 report, Brussels*, pages 9–11, 1998.

- H. Müller, P. Clough, T. Deselaers, B. Caputo, and I. CLEF. Experimental evaluation in visual information retrieval. *The Information Retrieval Series*, 32:1–554, 2010.
- 12. W. H. Organization et al. Global tuberculosis report 2016. World Health Organization, 2016.
- S. Peleg, J. Naor, R. Hartley, and D. Avnir. Multiple resolution texture analysis and classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(4):518–523, 1984.
- 14. X. Zhang, J. Cui, W. Wang, and C. Lin. A study for texture feature extraction of high-resolution satellite images based on a direction measure and gray level co-occurrence matrix fusion algorithm. *Sensors*, 17(7):1474, 2017.