

# Feature-Based Approach for Severity Scoring of Lung Tuberculosis from CT Images

Kirill Bogomasov, Ludmila Himmelspach, Gerhard Klassen, Martha Tatusch,  
and Stefan Conrad

Heinrich-Heine-Universität Düsseldorf, Institut für Informatik  
Universitätsstraße 1, 40225 Düsseldorf, Germany  
{bogomasov,himmelspach,klassen,tatusch,conrad}@cs.uni-duesseldorf.de

**Abstract.** Nowadays tuberculosis is still a widespread disease that causes worldwide more than one million deaths and ten million new infections every year. As part of ImageCLEF 2018, we investigated whether the severity of the disease can be determined from CT scans, only. We therefore extracted features from the images which we then tested with several classifiers. Afterwards we chose the best combinations of different feature sets and classification models. Our best approach is based on three features, namely cavitation, cavity tissue, and infection ratio. Combined with random forests we achieved rank 10 regarding the RMSE measure.

**Keywords:** feature extraction · image classification · lung tuberculosis

## 1 Introduction

In 2018 tuberculosis was listed as one of the top ten causes of death worldwide [1]. Depending on the severity degree of the disease different medical treatments are necessary. To this day the distinction of the severity degree has been executed by medical experts based on diverse information including the results of the mycobacterial culture test, pleural fluid and cerebrospinal fluid (CSF) analyses, lesion patterns in radiological images of the lungs, patient's age, duration of treatment and others [11]. In patients with tuberculosis, computed tomography (CT) is often performed for analyzing the lesion patterns in the lungs. The human-based analysis of the existing data is an expensive and time-consuming task. Additionally, a manual classification can be error-prone. In contrast to the manual examination of CT scans, a computer-based method could lower the error rate and simplify the procedure.

In this paper we present a feature-based approach for severity scoring of lung tuberculosis exclusively based on CT scans. This work is a contribution to the severity score tuberculosis task of ImageCLEF 2018 [6], [9]. Besides the tuberculosis degree determination, the main goal of our approach was to create a descriptive classification framework that provides information about the influence of different kinds of irregularities in lungs on severity scores.

## 2 Feature Extraction

We extracted features from CT images assuming that medical experts look for irregularities in the lungs that are typical for tuberculosis while analyzing CT scans in the context of severity score determination. In the medical literature, different kinds of irregularities and lesions in the lung associated with pulmonary tuberculosis are described. Our feature choice is for the most part based on this description. In this section, feature extraction methods are described that we used in our approach. Since the most of our feature extraction methods worked on binary images, we binarized all CT images using IsoData method [16] in a preprocessing step. Hereby we also used lung masks which were extracted by the algorithm that was published in [7].

### 2.1 Lung Calcification

Calcification is significant to the disease pattern of tuberculosis[2]. We assume that the identification and quantification of chalk within lung lobes can be a meaningful feature. Hounsfield Units (HU)[4] of chalk vary around 700 depending on its density. These HU values overlap with those of bones (300 HU - 1500 HU), which are often located in the boundary area of the masks. Hence, a simple thresholding approach is not sufficient. In order to avoid misclassification, we therefore had to adapt the size of the masks as long as parts of the bones were contained. Finally, pixel with the value  $\geq 700$  are counted because we regard those as calcifications. In detail our approach contains the following steps:

(1) Slice-wise CT scan preparation: Set values below 700 to  $-3024$  (no density)

(2) Slice-wise boundary analysis:

2.1 Boundary identification in mask slices:

$$boundary[i] = mask[i] - erode(mask[i]), \text{ with } i \in \{0, slices(mask)\}$$

2.2 Boundary extraction from CT scans:

$$bscan[i] = boundary[i] * scan[i]$$

2.3 Adaption of masks:

```
while (max(bscan[i])  $\geq$  700) do  
    mask[i] = erode(mask[i])  
    Step 2.2  
end while
```

(3) Summation of pixels with value  $\geq 700$  in  $mask[i] * scan[i]$

With *erode* [10], the function which executes an erosion on a binarized image slice.

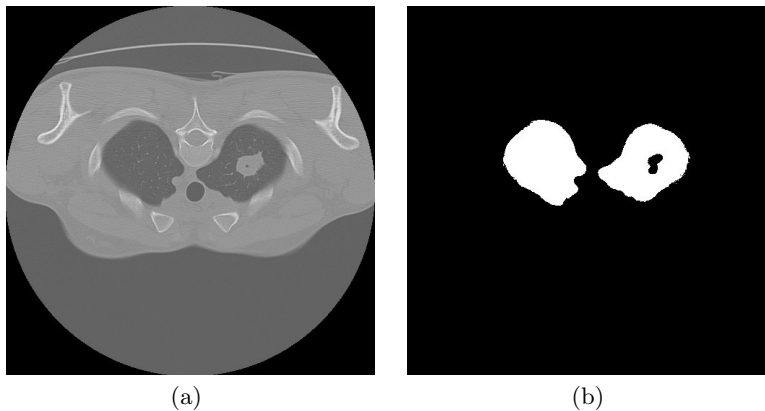
## 2.2 Lung Wateriness

Water accumulation is a potential concomitant of an existing tuberculosis disease. Nevertheless we assumed that an existing tuberculosis infection weakens the immune system of the patient and may lead to trace-diseases. There are several indispositions that are associated with fluid retention within the lungs or in the pleural space between the lungs and the ribs, such as pleural effusion[19]. Our assumption was that water effusion is a clear evidence for an advanced level of infection or traces of a cured serious illness. The process of searching the water retention was based on the HU[4]. The searching algorithm is the same as in Section 2.1.

## 2.3 Pulmonary Cavities

One of the classic indicators of lung tuberculosis are the pulmonary cavities which occur in 50 percent of patients [14]. According to [13] pulmonary cavitation formed as a result of tuberculosis is a site of very high mycobacterial burden. They may lead to transmission of the infection to other humans and they are associated with emergence of drug resistance. Furthermore, in [12], the authors reported about the relationship between the cavity wall thickness combined with the diameter of the lesion and the malignancy of the disease. Therefore, we extracted the size of pulmonary cavities and cavity walls from CT images as further features for severity scoring of lung tuberculosis. Although pulmonary cavities may also occur as a result of other diseases like lung cancer [14], the presence of additional pulmonary diseases in patients with lung tuberculosis may increase the degree of tuberculosis.

We extracted the pulmonary cavities from single CT image slices as dark spots completely surrounded by light tissue. Since we wanted to avoid finding



**Fig. 1.** An example of missing pulmonary cavitation in the mask: (a) CT scan showing cavitation in the left lung, (b) corresponding lung mask.

similar structures in other parts of CT images than lungs, we used the lung masks that were provided by the organizers of the task using the approach described in [7]. Due to variations of Hounsfield Units in different regions of cavities and cavity walls in different CT images, first, we binarized the CT images as described above. We had to make some adjustments to the lung masks because they did not cover the entire lung and often the cavities were cut out from the masks (compare Figure 1). Therefore, we closed all holes in the masks. Since we also closed the holes that correctly indicated bronchi, we cut out the middle part of the lung masks to avoid incorrectly recognizing bronchi as pulmonary cavities.

After processing the lung masks we performed the pulmonary cavitation search in binarized CT images as follows: First, we removed all objects smaller than 20 pixels because it is unlikely for a cavitation to be of such small size and analyzing such objects would unnecessarily require processing time. Since bronchioles scanned across and shadows caused by breathing and body movements could be falsely recognized as cavities, in the second step, we closed all holes that were smaller than two pixels. Obviously, the internal parts of undesired objects could be larger than 2 pixels, but, on the other side, we had to prevent erroneously discarding parts of real cavities. Because cavity walls are usually thicker than bronchiole walls, in the third step, we performed morphological opening with a  $2 \times 2$  square to discard undesired objects remained after the second step. We considered all holes that were completely surrounded by walls as pulmonary cavities after performing these three preprocessing steps. For performance reasons we estimated the volumes of pulmonary cavities by simply summing up the pixels of found cavities and cavity walls, respectively, over all CT scan slices in the file.

## 2.4 Infection Ratio

Pulmonary tuberculosis is an infectious lung disease whose bacilli spread through the lungs and cause lung tissue damage. Depending on the type of tuberculosis different types of lesions occur in the lungs (see Figure 2). That makes it difficult to estimate the amount of the affected part of the lungs automatically. Since the infection of the most tuberculosis types cause a thickening of the lung tissue



**Fig. 2.** Different types of lesions in the lung.

which can be recognized in CT scans, we simply estimated the ratio of the lung tissue to the entire lung volume. In our approach we did not differentiate between healthy and affected lung tissue which is a difficult task, our approach is based on the assumption that the lung tissue ratio compared to the lung volume is smaller in healthy persons than in persons suffering from tuberculosis. In order to highlight the lung lesions, we first binarized the CT images as described above. After the binarization we simply counted the number of white pixels and related it to the number of pixels in the lung mask [7].

## 2.5 Hounsfield Histograms

Since its introduction in 1972 [17] the technology of X-Ray computed tomography (CT) has been continuously refined. Over time several different devices with different parameter sets were developed [5]. The different technology and the parameters do not only concern the distance and time between images, but also the Hounsfield Units represented in the final image [15]. That leads to the problem, that the same object can have different Hounsfield Units on different images [5]. As there is no information provided what hardware and what parameters were involved in creating the scans for the dataset, it is difficult to look for certain Hounsfield Units which are comparable throughout all scans.

In order to overcome this problem we decided to compare intervals of Hounsfield Units with the help of histograms. As the intervals of Hounsfield Units of different tissues overlap, it is difficult to determine reasonable bins. Therefore we divided the interval of  $[-1024, 3000]$  into 20 equal sized bins. In the classification task every bin has been regarded as a single feature.

## 2.6 Lung Shape Comparison

Since we assume that the degree of pulmonary tuberculosis can correlate with the overall health of a patient, we considered the shape of the lungs, as well. We also assume that the difference between the shapes of the two lungs can provide information about the patient's health. To obtain a comparison measure, it is sufficient to look at the masks.

Note, that the following procedure needs all slices to be processed separately. In order to compare the different lungs, all masks have to be divided into two separate lung masks. Afterwards the contours of the relevant regions are calculated. In [18] a border following method is introduced to describe the contour of an object. The silhouettes of the lungs are calculated and stored in a vector of points using the findContours-method of OpenCV [3]. They can then be matched with OpenCV's matchShapes-method. A measure for the match can be computed by the following equation:

$$I(A, B) = \sum_{i=1}^7 |m_i^A - m_i^B| \quad (1)$$

with

$$m_i^A = \text{sign}(h_i^A) \cdot \log(h_i^A) \quad \text{and} \quad m_i^B = \text{sign}(h_i^B) \cdot \log(h_i^B),$$

where  $h_i^A$  and  $h_i^B$  are the Hu Moments of the contours  $A$  and  $B$  [8].

Now there are comparison values for each slice of the CT Scan but it is desired to receive one representative measure for the match. So finally, the average value over all slices has to be calculated.

### 3 Classification Methods

We used different classifiers for predicting the severity score and the severity level of tuberculosis using the feature set obtained from the feature extraction step described in Section 2. In the training phase of classification models, we performed feature selection based on the cross-validation mean square error for severity score on the training set. We tried different classification methods including the multi-class support vector machine (SVM) with RBF kernel, the k-nearest-neighbor (kNN) algorithm, and the multi-layer perceptron classifier with different parameter settings. Below we describe the best classification models with respect to the mean square error on the training set and the way of predicting the severity score and the severity level of tuberculosis.

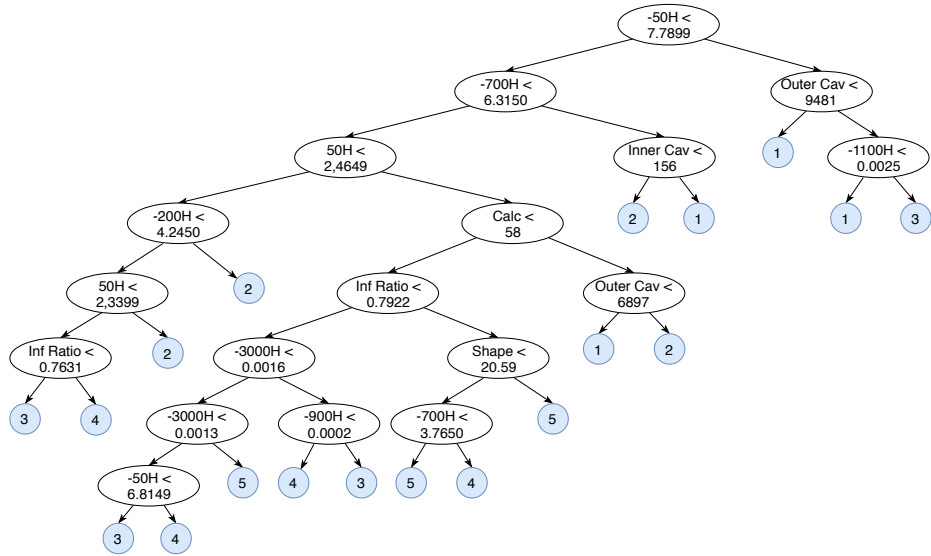
#### 3.1 Decision Tree

Using the Chi Square method it turned out that 13 of the 17 features are the most meaningful. Apparently the histograms of the higher Hounsfield Units are not very informative. Therefore, all features have been considered for the decision tree, except for the histograms of ranges with values greater than 50.

In Figure 3 the structure of the resulting decision tree is demonstrated. The numbers followed by an H stand for the ranges of the histograms. The assigned classes of the leaf nodes are highlighted in blue. Arrows to the left indicate that the condition of the parent node applies. Arrows to the right represent the non-applicability. The classifier was fitted using the Gini Impurity, a minimum fraction of 3% per leaf and a minimum quality gain of 0.01. The structure shows that most scores of the same severity class ("LOW"/"HIGH") share similar features. But the scores 3 and 4 are often separated by only one property, as well. According to the paths of the decision tree, the distinction between score 3 and 4 is the most difficult task and probably represents the biggest source of errors regarding the AUC score, since both degrees belong to different severity classes.

#### 3.2 Random Forests and Linear Regression

Using random forests and linear regression as classifiers we interpreted the prediction of the severity score and the severity level of tuberculosis as a regression problem. In our first attempt, we converted the severity level into the numbers 0



**Fig. 3.** Illustration of the decision tree using 13 features.

and 1, where 1 means "HIGH" and 0 means "LOW" severity level. We used the random forest classifier with the maximum depth value of 2 because the larger values led to overfitting of the classification model. We calculated the severity scores from the predicted severity levels by dividing the severity level values into intervals.

In the second test, we trained two separate classification models for the severity score and the severity level prediction. Since the severity score and the severity level values mismatched for some data items, we adjusted the severity score values depending on the corresponding severity level values at the extreme boundaries of the severity level. In particular, we set the severity score values to 1 if the corresponding severity level values were higher than 0.95. If the severity level values were below 0.22, we set the corresponding severity score values to 5 regardless the values that were predicted for the severity score before. Here, we also used the random forest classifier with the maximum depth value of 3. Additionally, we submitted a linear regression model which we trained on a subset of the training set for which we achieved the best results with respect to the cross-validation mean square error for the severity score. We used only a subset of the training set because linear regression is sensitive to outliers that we assumed in the training set due to variations in the mean square errors in different cross-validation runs.

Due to performance variations for the severity level in different cross-validation runs for the random forest model trained in the second test, we assumed that the classification model for prediction of the severity level overfitted on the training set. Therefore, in the third test, we reduced the maximum depth value to 2 for

the random forest model. Furthermore, we refrained from the adjustment of the severity score values based on the severity level.

## 4 Evaluation and Results

A maximum of 10 runs could be submitted by each group per subtask in the ImageCLEF 2018 TB task. This section shows the final performance results of our feature-based approach in the severity scoring challenge (subtask 3). The final ranking was based on the root mean square error (RMSE) for the severity score. Table 1 summarizes the results for different runs of our approach ordered by the ranking provided by the subtask organizers. Additionally, we listed the results of the best runs with respect to the root mean square error (RMSE) and the Area Under the ROC Curve (AUC) submitted in the competition.

The best run of our approach was obtained by the random forest classification model with severity score adjustment described in the second test in Section 3.2 (indicated as Rnd\_Frst\_depth\_3 in the table) using only three features: size of cavity, size of cavity tissue, and the infection ratio. Although our best run achieved the tenth rank regarding the RMSE measure, it was ranked sixteenth according to the AUC measure. In order to improve the results regarding the AUC measure too, we performed feature selection based on the cross-validation AUC value for the severity level on the training set using the same classification method. The so selected features were calcification, infection ratio, size of cavity, and the third, the sixth and the tenth bins of the histogram. Although this run was only ranked on the 25th placed regarding the RMSE, it achieved the eight place regarding the AUC measure.

Our second best run was achieved by the random forest classification model with the maximum depth value of 2 without severity score adjustment and the linear regression model trained on the subset of the training set (indicated as Rnd\_Frst\_depth\_2 and Lin\_Reg\_part in the table) on the same feature subset as our best run. The performance results of our approach that calculated the severity score from the predicted severity level values by the random forest classification model (indicated as Rnd\_Frst\_score\_by\_level in the table) were the worst among the regression based approaches described in Section 3.2. Although the AUC value for the severity level was the same as for our best run, the RMSE value for the severity score calculated based on the severity level was much worse than for the separate severity scores prediction model.

The decision trees unfortunately performed worst. They only achieved the ranks 32-34 regarding the RMSE measure. The severity class was determined on the basis of the received scores. For this, two methods were used. In the first approach, the values 1, 2 and 3 represent the class "HIGH", and 4 and 5 belong to class "LOW". In the other method the probability  $p$  of a high severity was calculated by the formula  $p = \frac{5-\hat{y}}{4}$ , where  $\hat{y}$  stands for the predicted severity score of the decision tree. The results showed that the first method scored significantly better AUC values. Our best decision tree even reached the ninth rank in regard to the AUC measure.



**Table 1.** Results for our top 5 runs for Subtask 3 – Severity scoring.

Classification model	Features	RMSE	Rank <sub>RMSE</sub>	AUC	Rank <sub>AUC</sub>
–	–	0.7840	1	0.7025	6
–	–	0.8934	5	0.7708	1
Rnd_Frst_depth_3	cav., cav. tissue, inf. ratio	0.9626	10	0.6484	16
Rnd_Frst_depth_2	cav., cav. tissue, inf. ratio	0.9768	13	0.6620	13
Lin_Reg_part	cav., cav. tissue, inf. ratio	0.9768	14	0.6507	15
Rnd_Frst_depth_3	calc., inf. ratio, cav., hist. bins 3,6,10	1.1046	25	0.6862	8
Rnd_Frst_score_by_level	cav., cav. tissue, inf. ratio	1.2040	29	0.6484	17

## 5 Conclusion

In this paper we have shown that our feature-based approach is competitive to other participants of the ImageCLEF 2018 challenge [6]. With our best methods we achieved rank 10 regarding the RMSE and rank 8 regarding the AUC measure. Almost all features in our approach were extracted using the lung masks provided by the organizers of the task. These masks were created by an automatic segmentation algorithm [7] that failed to recognize especially large lesions in the lungs in some cases. Consequently, our feature extraction algorithms also failed to work in such cases. Therefore, we assume that an optimization of the masks could lead to a more precise feature extraction and improvement of the final results of our approach. As the reproduction of Hounsfield Units of CT scanners may vary, further information about the hardware and the used parameters could lead to an improvement of the results. These could also be used to determine reasonable bins for the Hounsfield histograms.

Finally, the feature choice in our approach was for the most part based on own observations that could be approved by medical studies published in the literature in recent years. We believe that we could improve the feature extraction and consequently the final results of our approach by consulting medical experts specialized in treatment of pulmonary tuberculosis.

## References

1. World health organization. Website (2018), <http://www.who.int/en/newsroom/fact-sheets/detail/tuberculosis>; visited on 31. May 2018.
2. Ball, R., Greene, C., Camp, J., Rowntree, L.: Calcification in tuberculosis of the suprarenal glands: Roentgenographic study in addison’s disease. *Journal of the American Medical Association* **98**(12), 954–961 (1932)
3. Bradski, G.: The OpenCV Library. *Dr. Dobb’s Journal of Software Tools* (2000)
4. Brooks, R.: A quantitative theory of the hounsfield unit and its application to dual energy scanning. *Journal of Computer Assisted Tomography* **1**(4), 487–493 (1977)
5. Cropp, R.J., Seslija, P., Tso, D., Thakur, Y.: Scanner and kVp dependence of measured CT numbers in the ACR CT phantom. *Journal of Applied Clinical Medical Physics* **14**(6), 338–349 (2013)

6. Dicente Cid, Y., Liauchuk, V., Kovalev, V., Müller, H.: Overview of ImageCLEF-tuberculosis 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)
7. Dicente Cid, Y., Jiménez del Toro, O.A., Depeursinge, A., Müller, H.: Efficient and fully automatic segmentation of the lungs in ct volumes. In: Proceedings of the VISCERAL Anatomy Grand Challenge at the 2015 IEEE International Symposium on Biomedical Imaging (ISBI). pp. 31–35. CEUR-WS (2015)
8. Hu, M.K.: Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory* **8**(2), 179–187 (1962)
9. Ionescu, B., Müller, H., Villegas, M., García Seco de Herrera, A., Eickhoff, C., Andreczyk, V., Dicente Cid, Y., 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), LNCS Lecture Notes in Computer Science, Springer, Avignon, France (September 10-14 2018)
10. Jankowski, M.: Erosion, dilation and related operators. In: Proceedings of 8th International Mathematica Symposium (2006)
11. Koegelenberg, C., A Balkema, C., Jooste, Y., Taljaard, J., Irusen, E.: Validation of a severity-of-illness score in patients with tuberculosis requiring intensive care unit admission. *South African Medical Journal* **105**(5), 389–392 (2015)
12. Nin, C., de Souza, V., Alves, G., do Amaral, R., Irion, K., Marchiori, E., Hochhegger, B.: Solitary lung cavities: Ct findings in malignant and non-malignant disease. *Clinical radiology* **71**(11), 1132–1136 (2016)
13. Ong, C.W., Elkington, P.T., Friedland, J.S.: Tuberculosis, pulmonary cavitation, and matrix metalloproteinases. *American journal of respiratory and critical care medicine* **190**(1), 9–18 (2014)
14. Parkar, A., Kandiah, P.: Differential diagnosis of cavitary lung lesions. *Journal of the Belgian Society of Radiology* **100**(1) (2016)
15. Richter, A., Hu, Q., Steglich, D., Baier, K., Wilbert, J., Guckenberger, M., Flentje, M.: Investigation of the usability of conebeam ct data sets for dose calculation. *Radiation Oncology* **3**(1), 42 (2008)
16. Ridler, T., Calvard, S.: Picture Thresholding Using an Iterative Selection Method. *IEEE Transactions on Systems, Man and Cybernetics* **8**(8), 630–632 (1978)
17. Subburaj, K.: CT Scanning –Techniques and Applications. IntechOpen (2011)
18. Suzuki, S., Abe, K.: Topological structural analysis of digitized binary images by border following. *Computer Vision, Graphics, and Image Processing* **30**(1), 32–46 (1985)
19. Vorster, M., Allwood, B., Diacon, A., Koegelenberg, C.: Tuberculous pleural effusions: Advances and controversies. *Journal of Thoracic Disease* **7**(6), 981–991 (2015)