Image Segmentation for Lung Nodule Identification and Classification

Erin Bratu[1] and Mirza Khan[1]

Vanderbilt University, Nashville TN 37212, USA

Abstract. Lung nodules are commonly found in chest imaging. Methods to accurately identify and measure these nodules are increasingly important to improve time to diagnosis and initiation of therapeutic interventions. For malignant nodules, we wish to minimize the possibility of false negatives and to classify nodules as malignant metastatic or malignant lung cancer nodules. In this report, we detail two methods of segmentation implemented using the Insight Segmentation and Registration Toolkit (ITK). We selected the Markov random fields and watershed segmentation methods to compare the results of using two different classes of segmentations, region-based and intensity-based, respectively. Using the results from these two methods, we used Dice coefficients and statistical analysis to determine significance in classifying based on malignancy type and diagnostic method.

Keywords: Image segmentation · ITK · VTK · Markov Random Fields · Watershed.

1 Introduction

Lung, breast, and colorectal cancers are three of the most common forms of cancer, with an estimated 650,000 new cases expected in 2019 [15]. Both lung and metastatic breast or colorectal cancers are commonly associated with lung nodule formation. Manual identification and segmentation of these nodules can be a tedious, time-consuming task, and once segmented, the lung nodules must be differentiated between those formed from malignant lung cancer and those from malignant metastatic cancers. Developing an automated methodology to segment and classify lung nodules could reduce the time to diagnosis and initiation of treatment.

There are two primary objectives for the implementation of such methods. The first is accurate identification and segmentation of nodules, with a focus on minimizing the risk of false negatives. The second is to determine if automated segmentation methods are capable of extracting nodule features that would indicate malignancy and, if so, type of malignancy. We also wish to study whether there is any correlation between the results of the automated segmentations and the diagnostic method used. For the purposes of this project, we have implemented two segmentation methods using ITK. The first uses Markov random fields (MRF), a region-based method that classifies images based on

pixel neighborhoods. The second is the watershed segmentation algorithm, an intensity-based method that classifies pixels according to their paths of steepest descent such that those that share the same minima represent an object of interest. We collected results from our segmentations and compared them to ground truth data provided with our dataset to calculate Dice coefficients. These coefficients were then analyzed to determine if there is any statistical significance between the segmentation results, malignancy type, and diagnostic method.

1.1 Deviations From Proposal

This project originally proposed to study the use of automated segmentation methods to classify benign and malignant lung nodules. Not all of the available CT scans had accompanying diagnoses and diagnostic method data, which automatically limited the potential dataset to approximately ten percent of the original. This change is further discussed in Section 2.1. Additionally, we proposed to compare the capabilities of a classical segmentation method versus a more novel one, but we instead opted to implement one segmentation method from two different classes of segmentation methods, specifically the region-based and intensity-based classes.

1.2 Team Responsibilities

Both members of the team were responsible for the literature review of their selected segmentation method, as well as the analysis and review of their results. Both members worked collaboratively on the preparation of the presentation material and final report. Erin was responsible for implementing the MRF segmentation method as the group's region-based segmentation method. She also prepared a C++ program to calculate Dice scores for the results and a multi-viewport VTK visualization. Mirza was responsible for implementing an intensity-based segmentation method, specifically the Watershed method. He also selected and pre-processed the dataset used in this project from the meta-data available with the LIDC-IDRI dataset and performed statistical analysis on the calculated Dice scores.

2 Methods

2.1 Data Acquisition

Meta-data related to Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) images were obtained. We limited the scope of this study to patients with lung nodules as identified by the experts of the original dataset. Of these, nodule type was classified into two broad categories: primary lung cancer and metastatic cancer. The primary lung cancer type included squamous cell carcinoma, small cell carcinoma and adenocarcinoma. This information was also retained for downstream analysis as the lung cancer type

may also influence the lung nodule segmentation based on nodular characteristics of each cancer type. The metastatic lesions were selected from cases of colorectal cancer and breast cancer. Both of these malignancies comprise most of the metastatic cancer types in the LIDC-IDRI lung nodule cases. Additionally, both are commonly associated with metastatic lung involvement. A random subset of ten primary lung cancer cases and ten metastatic lung cases were selected for this study. Images and manual segmentations corresponding to these twenty total cases were then acquired from the publicly available LIDC-IDRI collection [1, 2, 5].

Manual Nodule Segmentations Manual segmentations were converted from the imaging and coordinate information present in the JSON file format into image files. Each nodule was represented in an individual file, so a single image containing all the nodules for a case was created using an addition filter in Slicer. These manually segmented image files were then used as the ground truth for comparisons with the automated and semi-automated segmentations we had constructed.

2.2 Lung Segmentation from Thoracic Images

As the task was to identify nodules only present in the lung, we began by first performing lung segmentation to create a lung mask (Fig. 1A). To retain lung tissue only, and to exclude areas outside of the body, a binary threshold of -200 to -800 was applied using the ITK BinaryThresholdImageFilter class, which was followed by a fillhole procedure using ITK's BinaryFillholeImageFilter. The lung was determined to be the largest connected component of the resulting volume. Thus, the largest connected component was selected by first applying the ConnectedComponentImageFilter followed by the LabelShapeKeepNObjectsImageFilter. The latter allowed us to select the largest connected component based upon the component containing the largest number of pixels. Using ITK's MaskImageFilter, we were then able to apply this lung mask to the original input image to retain only the lung for the subsequent segmentation tasks.

2.3 Markov Random Fields Segmentation

The pipeline for the MRF segmentation is shown in Figure 1(C). No preliminary filtering was applied to the input images for the results of the MRF segmentation presented in this paper. We found that with the current implementation, the lung mask either had no significant benefit or produced worse results than when using the original image. This held true for other attempted methods of filtering the images, including implementations of Otsu thresholding and Bayesian classification.

The MRFImageFilter class required membership functions derived from a labeled prior image, neither of which was readily available in the LIDC-IDRI

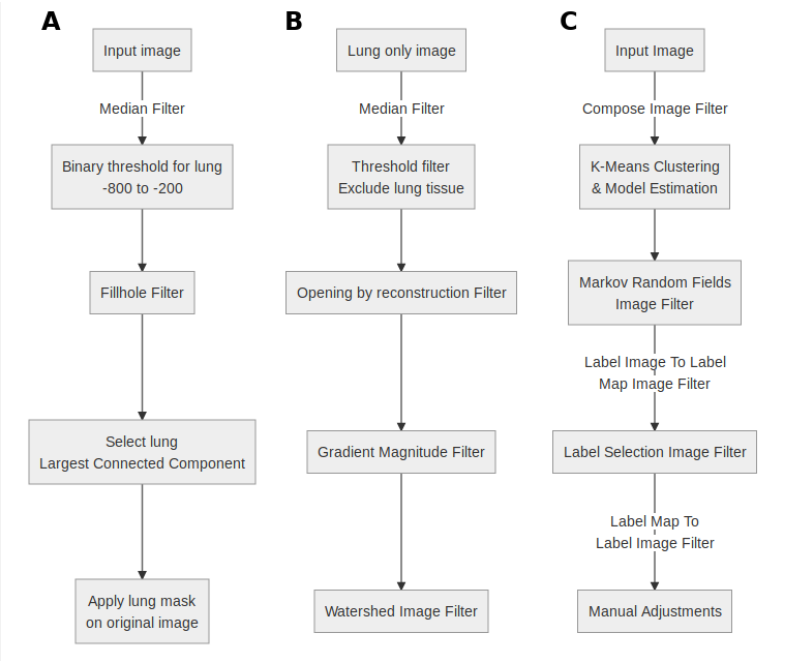


Fig. 1. Workflow diagrams for (A) lung mask creation, (B) watershed segmentation, and (C) MRF segmentation.

dataset. We use the ITK ImageKMeansModelEstimator class to create the labeled image and its associated membership functions. The membership functions were determined based upon distances to cluster centroids using the ITK Statistics DistanceToCentroidMembershipFunction class. We specified offset addition and multiplication parameters as 0.01, which would attempt to separate classes if the given number of classes had not been met.

The membership functions were used to initialize the MRFImageFilter classifier, which used an ITK ImageClassifierBase object. The classifier's decision rule was set using the ITK Statistics MinimumDecisionRule class. After initializing the classifier, the neighborhood size and neighbors' respective weights were set for the MRFImageFilter object. A small neighborhood with the most immediate voxels weighing the most was recommended for the segmentation of small objects [12]. A neighborhood size of 3x3x3 voxels was selected, and the chosen weights are given in Fig 2. The highest weight is assigned to neighbors in the same slice as the current pixel, with the next highest weight belonging to the same pixel location in the previous and next slices. All other pixels in the neighborhood are given the same weight. Once all of the parameters are set, the filter is updated, and the output image is written to a new file.

To extract the desired label, the output image must first be converted to an ITK LabelObject. This is done using the ITK LabelImageToLabelMapFilter.

1.3	1.3	1.3
1.3	1.5	1.3
1.3	1.3	1.3

1.7	1.7	1.7
1.7	0.0	1.7
1.7	1.7	1.7

1.3	1.3	1.3
1.3	1.5	1.3
1.3	1.3	1.3

Fig. 2. Neighbor weights in three consecutive slices. Pixel with weight 0.0 is center pixel.

The LabelSelectionLabelMapFilter is then used to extract the specified class. This label is then converted to an ITK Image object using the ITK LabelMapToLabelImageFilter. It is then written to a new file, as the final step of the automated segmentation.

Manual Adjustments Although the nodules were well segmented for most cases in the dataset, the MRFImageFilter was unable to distinguish the nodules from much of the surrounding organ tissue. For this reason, we made manual adjustments to remove the large pieces of extraneous tissue. We used the Segment Editor module in 3D Slicer to perform this portion of the project. The Dice coefficients for the MRF segmentation were calculated using these manually adjusted segmentations.

2.4 Watershed Segmentation

The lung mask was applied to our input image to retain only the lung from the original input volume. Given that the watershed algorithm is highly sensitive to noise, smoothing was applied using the ITK MedianImageFilter class [11, 4]. To exclude lung tissue, the BinaryThresholdImageFilter class was applied to retain areas with intensity greater than -200.

Opening by reconstruction was then used to remove some of the fine noise and small non-nodular objects. This method applies an erosion procedure followed by dilation. For this operation, an ITK BinaryBallStructuringElement was used as the kernel for the ITK OpeningByReconstructionImageFilter class. Different radius measures were trialed in the BinaryBallStructuringElement in order to exclude small non-nodular objects, such as vessels, while being able to preserve lung nodules. One of the benefits of using the opening by reconstruction method is that it maintains the edges of objects [13, 4]. The GradientMagnitudeImageFilter class was then applied as a height function [14]. The resulting image is then provided as an input to the WatershedImageFilter class. To mitigate oversegmentation, different watershed levels and thresholds were attempted. Ultimately, we selected a radius of 0.8, level of 0.4, and threshold of 0.004 as parameters for the segmentation task.

Watershed Variant Additionally, we assessed another variant of the watershed algorithm available in ITK via the MorphologicalWatershedImageFilter class to

see if this may provide more suitable results. Distance mapping methods, such as the SignedMaurerDistanceMapImageFilter and DanielssonDistanceMapImageFilter available in ITK were also evaluated. Ostensibly, these findings appeared comparable to what we had obtained using the original WatershedImageFilter class.

2.5 VTK

We attempted to produce a multiple-viewport VTK implementation, which was intended to be our primary method of displaying results. This was implemented using the basic pipeline for viewing images in VTK. To create multiple viewports, the SetViewPort method in the vtkRenderer was used. The multiple viewports work well, but there is an unresolved problem with being able to scroll through image slices. Issues with displaying images of certain data types have been encountered in previous assignments, which may be the problem here.

2.6 Dice Score Calculation

The automated and semi-automated segmentations generated using the methods described above were compared to the "ground truth" manual segmentations created by expert radiologists. To ensure that each image occupied similar physical space, the origin and spacing of the automated and manual segmentations were both changed to 0 and 1 mm, respectively. These changes were made using the ITK ChangeInformationImageFilter class. Upon this change, we were able to then calculate Dice coefficients comparing the manual and automated segmentations. The LabelOverlapMeasuresImageFilter class was used to compute the Dice coefficients.

Statistical Analysis Multiple linear regression was used to study the relationship of characteristics of each case on the Dice coefficient. This method was used to explore if we may be able to segment certain nodules better than others based upon certain clinical features. These clinical variables included malignancy type, total number of nodules and method of cancer diagnosis. Our outcome variable was the Dice score for each image. This procedure was separately applied on Dice scores from each of our segmentation methods, Markov Random Fields and Watershed transformation. An *a priori* threshold for statistical significance was set as $\alpha = 0.05$. Statistical analysis was performed using the linear model functions available in R.

3 Usage Instructions

3.1 Markov Random Fields

The following instructions refer to usage following successful download and compilation of the code in the relevant repository. In the following commands, inputDataFile refers to the .NRRD file on which one wishes to perform segmentation.

numberOfClasses is the number of labels you would like the segmentation to produce, and labelNumber refers to the label that contains the information you are interested in. caseId refers to the 4-character ID used in the LIDC-IDRI dataset to distinguish individual CT scans.

The next several commands refer to the usage of this repository directly from the command line. In order, these commands perform the following actions: segmentation of an image, extraction of a label from a segmented image, calculation of the Dice coefficient of a segmentation, and visualization of the ground truth and semi-automatic segmentations.

```
./SegmentationDriver 1 inputDataFile numberOfClasses
./SegmentationDriver 2 inputDataFile numberOfClasses labelNumber
./SegmentationDriver 3 caseId
./SegmentationDriver 4 caseId
```

Additionally, most of these options can be run on the ACCRE cluster using slurm scripts. To perform segmentation and extract a label, use `sbatch segmentation.slurm inputDataFile numberOfClasses labelNumber`. If the most likely label is known beforehand, the bash script `slurm_control.sh` can be used to segment and extract the specified label from all available image files. To calculate the Dice coefficients for all data files, use `sbatch similarity.slurm`.

3.2 Watershed Segmentation

Upon compilation and using similar naming conventions as those described above, the first argument of the watershed transformation application `Watershed` takes the input thoracic image file, `inputDataFile`, while the second argument is the name for the output image, `outputDataFile`:

```
./Watershed inputDataFile outputDataFile
```

For the Dice coefficient calculation, the `SimCalc` application takes the "ground truth" manual segmentation image, `manualSeg`, as the first argument and the automated segmentation image, `mySeg`, as the second argument:

```
./SimCalc manualSeg mySeg
```

4 Results and Discussion

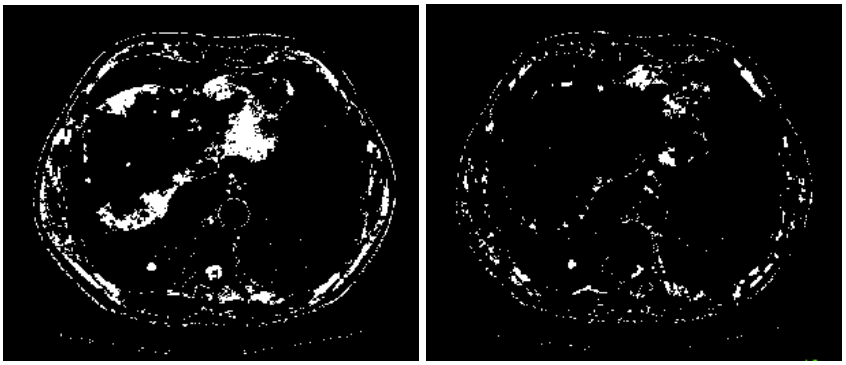
4.1 Markov Random Fields

The results of segmentation using two different numbers of classes are shown in Figure 3. One slice of the segmentation using four classes for case ID 0129 is shown in Figure 3(a). Figure 3(b) shows the same slice when using 64 classes for the segmentation. These examples demonstrate one of the primary problems with this implementation of Markov Random Fields segmentation. Due to the similarity in intensity of the nodules and the organ tissue around the lungs, these two groups were often classified as the same tissue. Despite this problem, the nodules appeared to be well-segmented, so we performed a manual segmentation to remove the organ tissue. We opted to use a lower number of classes for the

results used in the statistical analysis for two reasons. Firstly, the organ tissue was still largely connected, making its removal faster than when more classes were used. Secondly, the segmentation of actual nodules tended to be worse as the number of classes increased. These properties made the use of the higher-class segmentations undesirable when using the manual process, as it took longer to produce worse results.

Dice coefficients following the manual adjustments ranged from 0.00 to 0.95, with an average coefficient of 0.71. Two cases produced coefficients below 0.3, for which the segmentation method either separated the nodules across multiple classes or classified them as the same as the lung cavity. The nodule in the segmentation with a score of 0.95 was very smooth and almost spherical, leading us to believe that the quality of this implementation is very dependent upon the shape and size of nodules.

Figure 2 shows a three-dimensional comparison of the results following the manual adjustments to the ground truth segmentations provided with the dataset. The evaluation of only the nodule segmentations showed promising results for further development of the pipeline. The segmentation captured two false positives, which can be seen in the lower and upper left hand sides of Figure 2(b). One nodule, in the center of the left hand side of the image, was not captured in this label. Most other nodules were well segmented with regard to size and location.

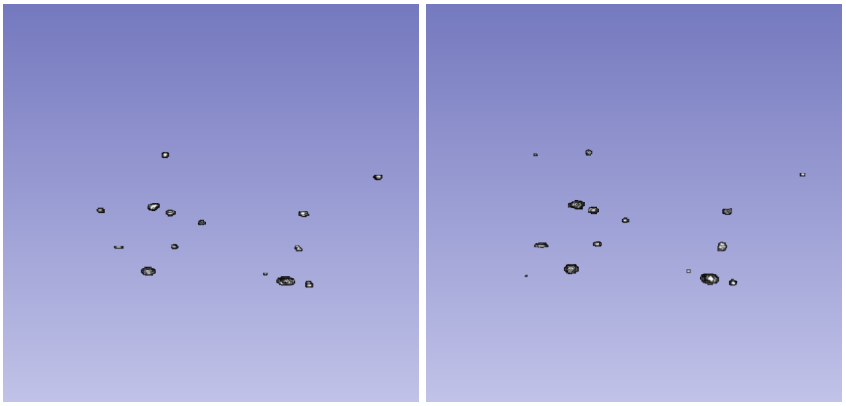


(a) Classes: 4. Label: 1. (b) Classes: 64. Label: 49.

Fig. 3. Segmentation using the MRF method and a varied number of classes before any manual adjustment.

4.2 Watershed Segmentation

Using the example of case ID 0129, Figure 5 demonstrates efforts to optimize the lung nodule segmentation task using the watershed transformation. Figure 5A



(a) Ground truth segmentation. (b) Semi-automated segmentation.

Fig. 4. 3D view of lung nodules segmented using the semi-automatic MRF method compared to the ground truth segmentations provided in the LIDC-IDRI dataset.

shows the resulting the gradient magnitude image. This height function is then applied to the watershed algorithm. Figures 5B-C show how varying the watershed level parameter, among other parameters, influences the segmentation and poses the potential for oversegmentation.

We attempted different radii values for the BinaryBallStructuringElement for our Opening by reconstruction method. Similarly, different watershed transformation parameters, namely level and threshold, were evaluated for risk of over- and undersegmentation. Too low of a radius or watershed level posed the risk of potentially capturing a significant amount of artifact, such as vessels, other cartilage and non-lung tissue. By contrast, too large of a radius or watershed level carried the potential for false negative nodule detection and segmentation. We favored minimizing the possibility of false negatives at the expense of increasing our false positives given the hazardous clinical implications of a false negative malignant lung nodule.

Upon optimizing the parameters for the watershed transformation, Dice scores were found to range from 0.00 to 0.016 with a mean Dice score of 0.003. Visual inspection of the automated segmentations revealed that nodules were successfully identified, but the presence of artifact and noise persisted and largely explained the low Dice scores. Manual cleaning of the automated segmentations was attempted to offset some of the non-nodular artifact. This process appeared to be subject to user error as it failed to yield significant and robust improvements in the computed Dice coefficients.

4.3 Statistical Analysis

Multiple linear regression was used for statistical inference, and not as a means of prediction of Dice scores. This method was used to assess the utility of clinical

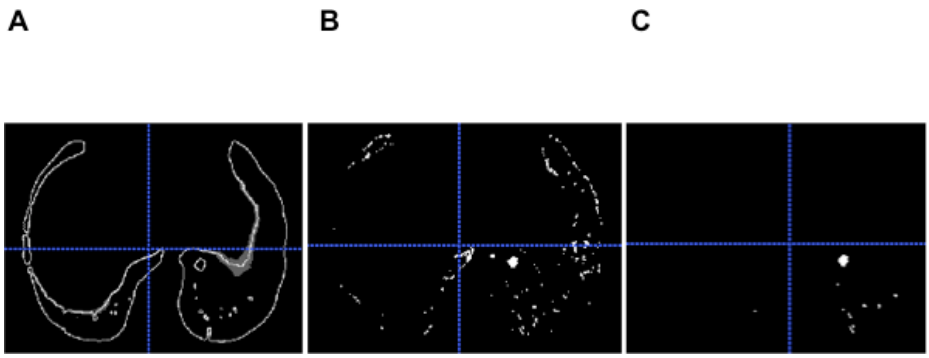


Fig. 5. Views of the watershed segmentation procedure: (A) computation of the gradient magnitude, (B) watershed segmentation using a level of 0.4, and (C) watershed segmentation using a level of 0.5.

features in determining each algorithm's ability to perform segmentation that is commensurate with the expert manual segmentations. The variance explained, R^2 for both algorithms was 0.24 and 0.59 for the Markov Random Field and Watershed algorithms, respectively. This demonstrates that these clinical features fail to explain much of the variance. Other variables may need to be assessed to better explain our outcome variable. We do find that diagnosis determined by progression or response was statistically significant for the Watershed algorithm ($p = 0.011$), whereas all other remaining variables showed no difference. Only two of our twenty cases were diagnosed by progression or response to treatment. Thus, further studies are needed to elucidate the meaning and generalizability of this finding.

4.4 Improvements and Future Work

There is much room for improvement within our implementations. One of the primary issues with the MRF segmentation method was an inability to differentiate the nodules from organ tissue in an unfiltered image, as well as difficulty classifying nodules into the same group when using the lung mask. These issues may be due to a bad initial classification and model estimation. Although there are many implementations of unsupervised MRF methods for tumor segmentation [3, 9], the recommended method for the ITK implementation uses a Gaussian model estimator trained using a labeled prior image [12]. Creating a better prior image may lead to better results and allow use of the lung mask to reduce the inclusion of organ tissue in the segmentation. Our watershed segmentation may improve with the use of more advanced and robust denoising methods, height functions and watershed parameter tuning. The methods may help to reduce the number of regional minima and mitigate the potential for oversegmentation. Furthermore, future studies should assess if different cancer types may be best segmented using different parameters or pre-processing steps. Other studies

have also demonstrated greater success using variants of the watershed method, such as marker-controlled watershed, or an ensemble of procedures, including marker-controlled watershed, active countours and Markov random field [8, 10].

As such, another consideration for continuation of the project is to construct a pipeline that implements both segmentation methods, performed either in tandem or sequentially. The weaknesses of each method appear to affect different features in the images, leading us to believe a combination of the two methods may produce better results. Following our presentation, we tested a very basic form of this combination by masking the initial results from each method, which did lead to a better segmentation. Additionally, making individual or combined processes iterative may lead to better results, where the models for the MRF filter are re-estimated after each segmentation. As an alternative to this iterative approach, literature shows that implementation of these methods with deep learning techniques produces promising results for segmentation of tumors [6, 7].

5 Conclusion

Lung nodule segmentation is a necessary and important task given the clinical implications of early detection of malignancy and the importance of minimizing false negative results. Upon manual review, we found that each of our segmentation methods works well to identify known lung nodules. Yet, non-lung tissue and artifact were also captured. Thus, noise and oversegmentation were persistent errors despite our best efforts to correct for this. Our statistical analysis found that the malignancy type and the number of lung nodules were not associated with our segmentation ability, as determined by Dice scores. Future work should utilize an ensemble of these segmentation methods, which may work better than either segmentation method alone.

References

1. Armato III, S., McLennan, G., Bidaut, L., McNitt-Gray, M., Meyer, C., Reeves, A., Clarke, L.: Data From LIDC-IDRI. The Cancer Imaging Archive. http://doi.org/10.7937/K9/TCIA.2015.LO9QL9SX (2015)
2. Armato III, S., McLennan G., Bidaut L., McNitt-Gray M., Meyer C., Reeves A., Zhao B., Aberle D., Henschke C., Hoffman E., Kazerooni E., MacMahon H., van Beek E., Yankelevitz D., et al.: The Lung Image Data
base Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans. Medical Physics **38**: 915–931 (2011).
3. Ashraf, A., Gavenonis, S., Daye, D., Mies, C., Rosen, M., Kontos, D.: A Multichannel Markov Random Field Framework for Tumor Segmentation With an Application to Classification of Gene Expression-Based Breast Cancer Recurrence Risk. IEEE Transactions on Medical Imaging **32**(4), 637–648 (2013) doi: 10.1109/TMI.2012.2219589
4. Beare, R., Lehmann, G.: The watershed transform in ITK - discussion and new developments. The Insight Journal. 92 (2006).

5. Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., Tarbox, L., Prior, F.: The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository, Journal of Digital Imaging **26**(6), 1045–1057 (2013)
6. Das, A., Acharya, U., Panda, S., Sabut, S.: Deep Learning Based Liver Cancer Detection Using Watershed Transform and Gaussian Mixture Model Techniques. Cognitive Systems Research **54** 165–175 (2019)
7. Liu, Z., Li, X., Luo, P., Loy, C., Tang, X.: Deep Learning Markov Random Field for Semantic Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **40**(8) 1814–1828 (2018)
8. Ignatious, S., Joseph, R., John, J., Prahladan, A.: Computer Aided Lung Cancer Detection and Tumor Staging in CT image using Image Processing. IJCA. **128**, 29–33 (2015). https://doi.org/10.5120/ijca2015906607.
9. Ruan, S., Bloyet, D., Revenu, M., Dou, W., Liao, Q.: Cerebral Magnetic Resonance Image Segmentation Using Fuzzy Markov Random Fields. In: Proceedings IEEE International Symposium on Biomedical Imaging, 237–240. Washington, DC, USA (2002)
10. Tan, Y., Schwartz, L.H., Zhao, B.: Segmentation of lung lesions on CT scans using watershed, active contours, and Markov random field. Med Phys. 40, 043502 (2013). https://doi.org/10.1118/1.4793409.
11. ITK - Segmentation & Registration Toolkit, https://itk.org/ITK/applications/WatershedSegmentationExample.html. Last accessed 29 Nov 2019.
12. ITK itk::MRFImageFilter < TInputImage, TClassifiedImage > Class Template Reference, https://itk.org/Doxygen/html/classitk_1_1MRFImageFilter.html. Last accessed 30 Nov 2019.
13. ITK: itk::OpeningByReconstructionImageFilter < TInputImage, TOutputImage, TKernel > Class Template Reference, https://itk.org/Doxygen/html/classitk_1_1 OpeningByReconstructionImageFilter.html. Last accessed 30 Nov 2019.
14. ITK: itk::WatershedImageFilter < TInputImage > Class Template Reference, https://itk.org/Doxygen/html/classitk_1_1WatershedImageFilter.html. Last accessed 30 Nov 2019.
15. National Cancer Institute Common Cancer Types, https://www.cancer.gov/types/common-cancers. Last accessed 6 December 2019.