Automatic evaluation of liver cirrhosis in lung cancer patients receiving radiotherapy

Gabriel Battcock
9618026

Second semester report

MPhys project
School of Physics and Astronomy
The University of Manchester
April 20, 2020

Automatic evaluation of liver cirrhosis in lung cancer patients receiving radiotherapy

Gabriel Battcock

Abstract

1 Introduction

1.1 Computed tomography

Computed tomography (CT) is a process where a narrow beam of X-rays are targeted at a patient and rotated around body to create layers of the body. Different tissue in the body show up as different shades of gray, making internal organs distinct [1]. By stacking the of the CT scan together, a 3D image of the body can be produced and regions identified.

CT scans are used in diagnosis of various diseases including cancers, heart disease and head injuries. After lung cancer has been diagnosed and patients are receiving treatment, CT scans are taken from the top of the torso to below the lung. This data is used to locate tumours for treatment using radiotherapy, and to track the change in their size [2]. With a wealth of data, it would be useful to learn more about diseases that may be present in patients who have lung cancer.

Directly below the lung in the human body is the liver (Fig. 1). To make sure that the whole lung has been imaged, the CT scan is taken until the middle of the torso, including most, if not all of the liver. With access to the CT scans of lung cancer patients, one can study the liver well, and potentially make a connection between health of the lungs, and health of the liver.

1.2 Liver cirrhosis

When the liver tries to repair itself, it creates scar tissue. A build up this tissue is known as liver cirrhosis. Cirrhosis can cause many problems in the body including high blood pressure (portal hypertension), jaundice and enlargement of the spleen [4].

Liver cirrhosis is best diagnosed using an invasive biopsy. This poses many risks and problems, with sampling errors and biased results arising from such a small sample being taken [5]. Other non-invasive tests have been created like liver-stiffness measurements but have disadvantages that make them undesirable [6].

A more recent development is using the use of computer programs to help with the diagnosis. By evaluating the roughness of the surface of the liver compared to the boundary of a healthy, smooth liver, a liver surface nodu-

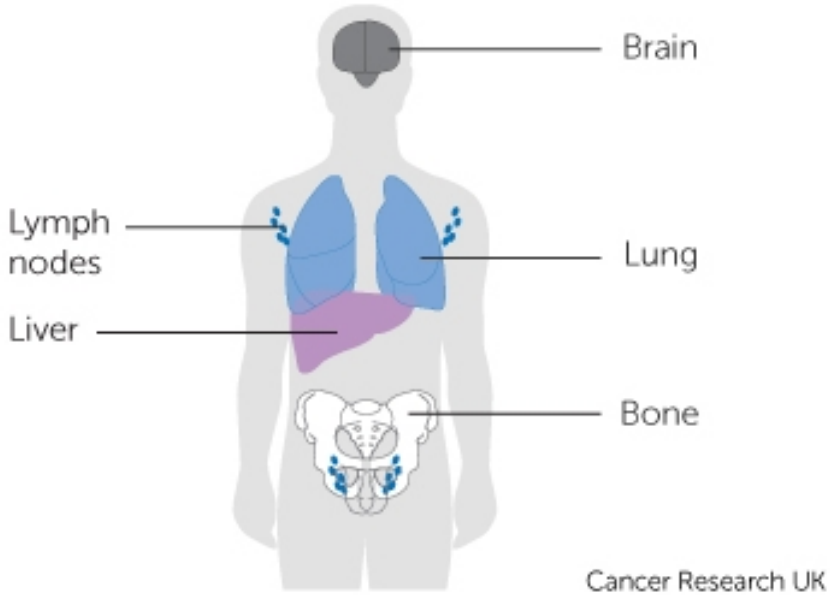


Figure 1: Visual of organs in the body, showing that the liver is located directly below the lungs, taken from Cancer research UK [3].

larity (LSN) score can be obtained. This is an important biomarker for liver cirrhosis. As written in A D Smith et al., the LSN score can *"accurately differentiate cirrhotic from noncirrhotic livers and is highly reproducible"* [6]. With this method, a region of interest (ROI) is selected by doctor, and the program automatically draws a polynomial line for where the surface should be, and where it actually is.

1.3 Motivation

By using the data available from lung cancer patients, the aim of the project was to automatically evaluate livers in lung cancer patients. Three main steps were proposed for this: automatically finding a 'good' slice of liver to use; using a pre-trained transfer learning neural network to automatically segment the liver; and a program to give the patient an LSN score. By creating an automated system, a check for liver cirrhosis can be performed as a safety measure. Furthermore, this can help with further research into a relation between lung cancer and liver cirrhosis.

These sections were divided between myself and Theo by sharing the first section, Theo taking the neural network, and I worked on the LSN score.

2 Theory

2.1 CT scans

CT scans are taken using narrow beams of x-rays rotated around the body in thin slices, transverse to the longitudinal axis of the body. A patient lies on a bed that moves through a annulus shaped device that which emits and detects the radiation which is then interpreted as a slice by a computer. These slices can vary in thickness, between 1-10 mm. Slices are saved as a DICOM file (standing for digital imaging and communication in medecine). Each file includes a header file which includes important metadata, such as patient ID, slice location, slice thickness and other information about the patient and the scan [7]. These data are invaluable when using scans for analysis.

CT images have dimension 512 x 512 pixels, and for medical images have high resolution of approximately 0.5 mm across the transverse plane, with a pixel being approximately 1 mm in length [8]. Each pixel in of the scan gives a value given in a dimensionless unit called the Hounsfield Unit (HU). This is defined on -1000 HU for air and 0 HU as distilled water [9]. Thus, values for other tissue and body parts can be found using:

$$HU = \frac{\mu_X - \mu_{water}}{\mu_{water} - \mu_{air}} \times 1000 \tag{1}$$

where μ_i is the linear attenuation coefficient, a measure of how easily a material can be penetrated by radiation. Typical values of internal organs can be seen in Tab. 1. The liver and fat have distinct values and thus can be distinguished from each other well.

Table 1: Table listing HU values for various organs and material that are in close proximity to the liver [10] [11].

Organ	HU value
Liver	50.4
Midline anterior subcutaneous fat	113.4
Lung	-498
Bone	570
Soft tissue	56

To convert from HU values to grey scale, on range [0,255], a process called windowing is used. This is composed of two parameters; window-width and window-level. Window-width refers to the range of values that are converted to grey scale. A window can be wide or narrow depending on the organ or tissue being examined [12] [13]. Window-level is the midpoint of of the window range being displayed. Changing these values together can highlight features of the scan that are not immediately apparent, as seen in Fig. 2.

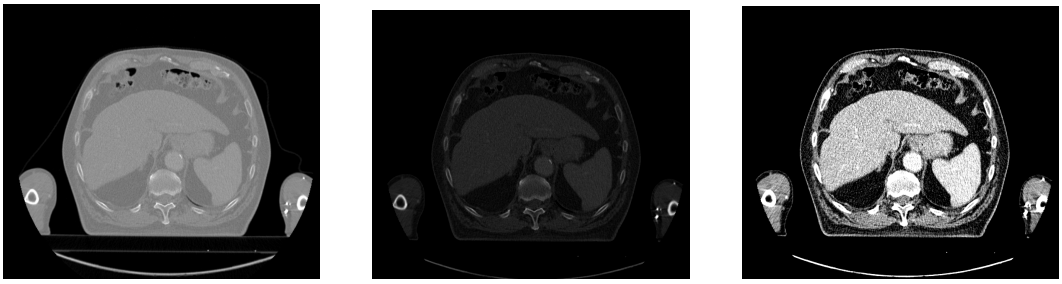


Figure 2: This is an example of what changing window and level values has on an image. The image on the left is the original, showing the liver and fat that look similar in terms of grey scale. The middle image is of the same liver, with a higher window-level value. The right image is of the liver with a narrower window-width. In this image, there is a very obvious boundary between the fat and the liver that is not as clear in the original.

2.2 Liver segmentation

In order to train a neural network to auto segment the liver,and to calculate an LSN score, a 'good' slice of the CT scan has to be found. A 'good' slice is one which has a border with fat, as it has a good contrast for training a neural network and for finding the liver boundary for calculating an LSN score. This was found in the middle of the liver. Furthermore, only the outer boundary of the liver was wanted due to the shape of liver having a slope which blurs the boundary in the middle of the body. Slices that contain gas, such as the bottom of the lung or bowl gas is best to be avoided.

To make a quantitative assessment on liver cirrhosis based on a CT scan, a metric called the liver surface nodularity score is used. As the liver builds up more scar tissue, the liver surface becomes wrinkly (fibrosis). The textured surface can be detected by a CT scan. An LSN score can be calculated using two lines drawn on the boundary of the liver. The first is a smooth polynomial

line for where the boundary of a healthy liver would be. The second line is the detected edge of the liver, calculated using active snake contours. The LSN score is defined as the mean pixel-by-pixel distance (10th of a millimeter), between the polynomial line and the detected boundary [6] [14] [15]. By grouping scores together, an assessment of the level of fibrosis the liver is, although the value of classification differs in the literature.

To find the two boundaries of the liver previously outlined, active contour algorithms are used. These models are used to accurately describe the boundaries of objects in images, and can be manipulated for the purposes of detecting the liver surface from a CT scan. In image processing, energy is a way of measuring the uniformity of an image.These work by finding the minimum energy of line, with constraints of the energy of the image that guide it towards boundaries [16]. As explained in Kass et al., the energy functional of the snake, E^*_{snake} is:

$$E^*_{\text{snake}} = \int_0^1 E_{\text{int}}(\boldsymbol{v}(s)) + E_{\text{image}}(\boldsymbol{v}(s)) + E_{\text{con}}(\boldsymbol{v}(s))ds \tag{2}$$

where E_{int} is the internal energy of the line, E_{image} is the energy of the image, E_{con} is the energy of the constraints and $\boldsymbol{v}(s)$ is the parametric equation for the snake. The energy of the image consists of energy of lines and edges. Thus, finding the minimum of this functional, given appropriate parameters, will find the boundary of the given image. These parameters include starting boundary conditions for the snake, smoothness and attraction to edges.

Another way to to find the boundary on images is by using morphological active contours. These are similar to active contour explained prior, but have a better stability [17].

2.3 Neural network

Talk a little bit about transfer learning neural networks, but only enough to explain why they were used.

3 Results

3.1 Slice finding

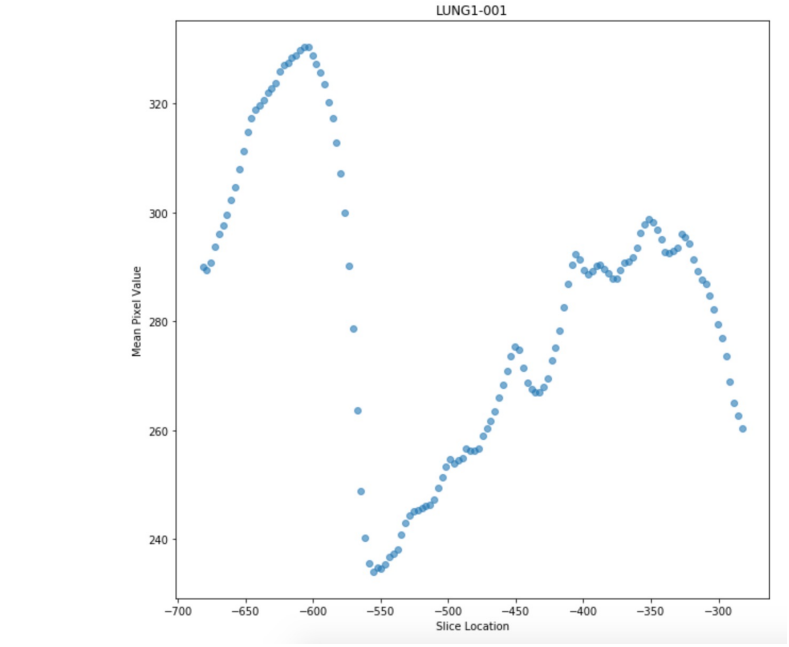
To investigate liver cirrhosis in lung cancer patients, an open source data set of lung cancer patients from the Cancer Imaging Archive was used [18]. This

included CT scans of 422 patients. Files were read into a Python file, and manipulated using PyDICOM package [19].

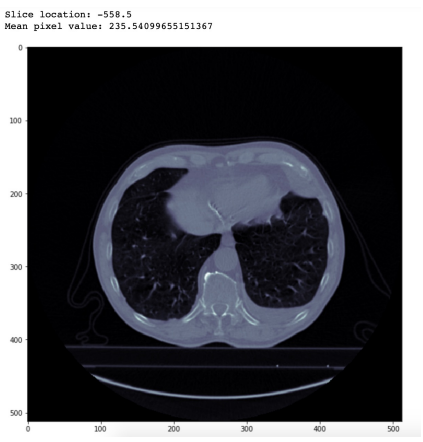
Using these data, two methods of finding a good slice of the liver were proposed. The initial, naive approach was to find the liver using the mean pixel value of slice against the position in the body to find where the lungs ended, and thus the liver. The second approach was to use connected components to find the bottom of the lung and thus the liver.

The naive approach using mean pixel value was discussed after viewing looking at slices of the CT scan. Lungs appear dark due to their composition of mostly air, whereas the liver is bright and large (Tab. 1). After plotting the mean pixel value against slice position, clear maxima and minima were found (Fig. 3a). These maxima corresponded nicely to a region inside of the liver, but not always a slice with a clear boundary with fat that could be used for segmentation . An example of finding initial approach can be seen in Fig. 3. After looking through for the maximum mean pixel slice for patients in the data set, the program would sometimes return a slice of the shoulder blades, due to bone being thick in this region while also having a high HU value. To get around this, the first 10 cm of the CT scan would not be included. This is valid for CT scans of lung cancer patients as they should all start around the same position at the top of the torso.

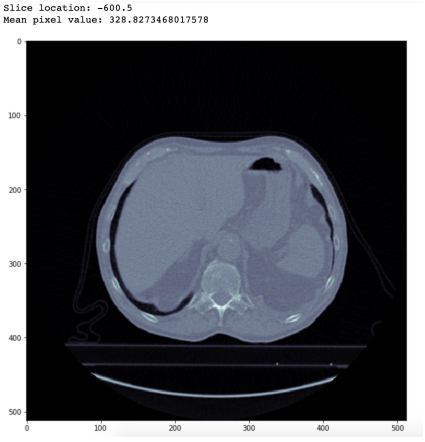
Due to the naive approach not consistently finding a slice of the liver with a clear boundary with fat, a second method was tried. Connected components in image processing is a method of labeling sections of an image by first creating a binary image from a threshold value, below which pixel will be zero, and above will be one. Pixels of the same that share an edge or corner are considered to be connected. This can be transferred up to 3D when taking into account voxels that border in the same manner. By applying this to a full CT scan, the lungs were found as the second largest connected component. Using the bottom of the lung as a marker for where the liver is, slices were found. However, this had the same problem of naive approach, where a reliable slice that could used for segmentation. This is due to variability in the shape of the lung and the shape of the liver. Lungs can extend low into the torso, resulting in a slice of the liver that is too low down. Avoiding this by a technique of blurring and redefining can end up with a liver that borders the lung, which is not ideal for segmentation. **Finding a biomarker for a 'good' slice of the liver is a difficult.**



(a)



(b)



(c)

Figure 3: Representation of finding the liver from a patient's full CT scan using the mean pixel value. Fig. a shows a plot of the normalized mean pixel value against the slice location. The global minimum and global maximum are Fig. b and Fig. c respectively. Fig. b shows the middle of the lungs, while Fig. c shows a slice of the liver.

Discuss slightly the use of Slicer to segment the images at this point, how it is useful for the n-n and for the line finding. Once a slice of the liver was found for a patient, a manual segmentation was performed using Slicer [20]. A painting tool was used to highlight the boundary of the liver. The segmentation was saved as a binary Nifti (Neuroimaging Informatics Technology Initiative) file, as a 3D array. These were converted to a 2D Numpy array using Nibabel package [21]. Files of the binary map and original slice were saved together as npy files, useful for training the neural network and for finding drawing a smooth polynomial line on the boundary of the liver.

3.2 Transfer learning neural network

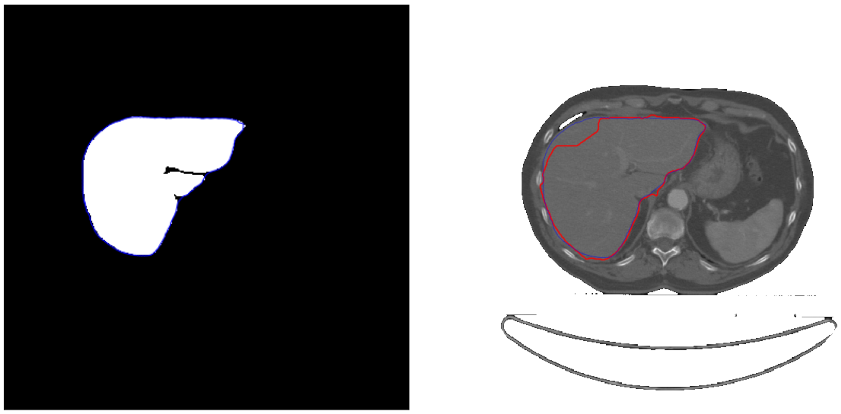
Discuss very briefly what numbers were used and what the results are. Don't really know what to discuss here. What optimizing functions were used, what value of loss we get, show some results of the segmentation.

3.3 Calculating LSN score

Discuss how the boundary was found for the mask (quite easy). show a few plots and make a table of the LSN values. Talk about errors To calculate the LSN score, two lines were drawn on the boundary of the liver, as described in Sec. 2.2. To handle the data used, a class was created for the patient which include the binary mask file and original CT image of the liver. Within this, functions were created for finding the smooth polynomial line and the true boundary of the liver.

The smooth polynomial was found using the 2D binary mask from the segmentation. Using an active contour, a smooth line was found for the whole liver, from which a section could be isolated to calculate the LSN score (Fig. 4a).

To find the true boundary of the liver a geodesic active contour (GAC) was used on the CT image. First, the background of the image was removed using a threshold to set dark pixel to NaN so that it did not interfere with the contouring. For the GAC to find the minima of the image, the slice has to be preprocessed into the form of an inverse Gaussian gradient. This applies a Gaussian filter to blur the image, then take the inverse of the first derivative of the image. The resulting array has areas of local uniformity around one and areas close to a boundary as close to zero [22].



(a) **(b)**

Figure 4: Boundaries drawn onto the arrays. Fig. a shows a blue polynomial line fit to the segmentation of the liver. Fig. b shows the true boundary of the liver segmented using a geodesic active contour on top of the blue polynomial lines, and the original CT scan image.

The contour algorithm works by using an initial snake which then expands or contracts trying to find the minimum value from the parameters set. For the boundary of the liver, the initial contour was a small circle set in the centre of the liver. A threshold value between HU values for the liver and fat was set so that the initial contour, after 'ballooning' (expanding from the centre), can converge toward, marking out the boundary of the liver. This consistently found the boundary of the liver boundary, but would sometimes get stuck at seemingly random sections, as in Fig. 4b.

Using the two boundaries found, the LSN was calculated on a section of the liver. The most ideal section to use one with a clear boundary with fat, usually at the front of the body, the top of the CT scan. To measure the mean pixel distance between the two segmentations, two lines of with the same ends points followed the respective boundaries. The distance in terms of pixels was measured for each point on the lines, and the mean was taken giving the LSN value. Having to manually attach the end points of the line meant the process could not be fully automated for the whole data set, so only a few scores are given in Tab. 2.

Table 2: Various LSN score calculated for patients.

Patient ID	LSN score
LUNG1-002	1.21
LUNG1-027	1.85
LUNG1-049	1.54

4 Discussion

5 Conclusion

References

[1] J. Broder, *Chapter 9 - Imaging of Nontraumatic Abdominal Conditions*. Saint Louis: W.B. Saunders, 2011, pp. 445–577. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9781416061137100092

[2] N. C. Purandare and V. Rangarajan, "Imaging of lung cancer: Implications on staging and management," *The Indian journal of radiology & imaging*, vol. 25, no. 2, pp. 109–120, Apr-Jun 2015. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/25969634

[3] (2020, 04). [Online]. Available: https://www.cancerresearchuk.org/what-is-cancer/how-cancer-can-spread/where-cancer-can-spread

[4] (2020, 04). [Online]. Available: https://www.mayoclinic.org/diseases-conditions/cirrhosis/symptoms-causes/syc-20351487

[5] A. Huber, L. Ebner, J. T. Heverhagen, and A. Christe, "State-of-the-art imaging of liver fibrosis and cirrhosis: A comprehensive review of current applications and future perspectives," *European journal of radiology open*, vol. 2, pp. 90–100, 05 2015. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/26937441

[6] A. D. Smith, C. R. Branch, K. Zand, C. Subramony, H. Zhang, K. Thaggard, R. Hosch, J. Bryan, A. Vasanji, M. Griswold, and X. Zhang, "Liver surface nodularity quantification from routine ct images as a biomarker for detection and evaluation of cirrhosis," *Radiology*, vol. 280, no. 3, pp. 771–781, 2020/04/20 2016. [Online]. Available: https://doi.org/10.1148/radiol.2016151542

[7] D. R. Varma, "Managing dicom images: Tips and tricks for the radiologist," *The Indian journal of radiology & imaging*, vol. 22, no. 1, pp. 4–13, 01 2012. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/22623808

[8] E. Lin and A. Alessio, "What are the basic concepts of temporal, contrast, and spatial resolution in cardiac ct?" *Journal of cardiovascular computed tomography*, vol. 3, no. 6, pp. 403–408, Nov-Dec 2009. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/19717355

[9] (2019, 12). [Online]. Available: https://radiopaedia.org/articles/hounsfield-unit?lang=gb

[10] R. Lamba, J. P. McGahan, M. T. Corwin, C.-S. Li, T. Tran, J. A. Seibert, and J. M. Boone, "Ct hounsfield numbers of soft tissues on unenhanced abdominal ct scans: variability between two different manufacturers' mdct scanners," *AJR. American journal of roentgenology*, vol. 203, no. 5, pp. 1013–1020, 11 2014. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/25341139

[11] J. Broder, *Chapter 9 - Imaging of Nontraumatic Abdominal Conditions*. Saint Louis: W.B. Saunders, 2011, pp. 445–577. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9781416061137100092

[12] (2019, 12). [Online]. Available: https://radiopaedia.org/articles/windowing-ct?lang=gbv

[13] Z. Xue, S. Antani, L. R. Long, D. Demner-Fushman, and G. R. Thoma, "Window classification of brain ct images in biomedical articles," *AMIA ... Annual Symposium proceedings. AMIA Symposium*, vol. 2012, pp. 1023–1029, 2012. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/23304378

[14] R. Sartoris, P.-E. Rautou, L. Elkrief, G. Pollorsi, F. Durand, D. Valla, L. Spahr, S. Terraz, O. Soubrane, F. Cauchy, V. Vilgrain, and M. Ronot, "Quantification of liver surface nodularity at ct: Utility for detection of portal hypertension," *Radiology*, vol. 289, no. 3, pp. 698–707, 2020/04/24 2018. [Online]. Available: https://doi.org/10.1148/radiol.2018181131

[15] T.-H. Kim, J. E. Kim, J.-H. Ryu, and C.-W. Jeong, "Development of liver surface nodularity quantification program and its clinical application in nonalcoholic fatty liver disease," *Scientific Reports*, vol. 9, no. 1, p. 9994, 2019. [Online]. Available: https://doi.org/10.1038/s41598-019-46442-y

[16] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988. [Online]. Available: https://doi.org/10.1007/BF00133570

[17] P. Márquez-Neila, L. Baumela, and L. Alvarez, "A morphological approach to curvature-based evolution of curves and surfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 1, pp. 2–17, 2014.

[18] W. L. R. V. E. L. R. T. H. P. C. G. P. . L. P. A. H. J. W. L. W. L. R. V. E. L. R. T. H. P. C. G. P. . . . . L. P. . Aerts, H. J. W. L., "Data from nsclc-radiomics," 2019, [Data set].

[19] D. Mason, "Su-e-t-33: Pydicom: An open source dicom library," *Medical Physics*, vol. 38, no. 6Part10, pp. 3493–3493, 2020/04/24 2011. [Online]. Available: https://doi.org/10.1118/1.3611983

[20] (2020). [Online]. Available: https://www.slicer.org/

[21] M. Brett, C. J. Markiewicz, M. Hanke, M.-A. Côté, B. Cipollini, P. McCarthy, D. Jarecka, C. P. Cheng, Y. O. Halchenko, M. Cottaar, S. Ghosh, E. Larson, D. Wassermann, S. Gerhard, G. R. Lee, H.-T. Wang, E. Kastman, J. Kaczmarzyk, R. Guidotti, O. Duek, A. Rokem, C. Madison, F. C. Morency, B. Moloney, M. Goncalves, R. Markello, C. Riddell, C. Burns, J. Millman, A. Gramfort, J. Leppäkangas, A. Sólon, J. J. van den Bosch, R. D. Vincent, H. Braun, K. Subramaniam, K. J. Gorgolewski, P. R. Raamana, B. N. Nichols, E. M. Baker, S. Hayashi, B. Pinsard, C. Haselgrove, M. Hymers, O. Esteban, S. Koudoro, N. N. Oosterhof, B. Amirbekian, I. Nimmo-Smith, L. Nguyen, S. Reddigari, S. St-Jean, E. Panfilov, E. Garyfallidis, G. Varoquaux, J. H. Legarreta, K. S. Hahn, O. P. Hinds, B. Fauber, J.-B. Poline, J. Stutters, K. Jordan, M. Cieslak, M. E. Moreno, V. Haenel, Y. Schwartz, Z. Baratz, B. C. Darwin, B. Thirion, D. Papadopoulos Orfanos, F. Pérez-García, I. Solovey, I. Gonzalez, J. Palasubramaniam, J. Lecher, K. Leinweber, K. Raktivan, P. Fischer, P. Gervais, S. Gadde, T. Ballinger, T. Roos, V. R. Reddam, and freec84. (2020, apr) nipy/nibabel: 3.1.0. Zenodo. [Online]. Available: https://doi.org/10.5281/zenodo.3757992

[22] [Online]. Available: https://scikit-image.org/docs/dev/api/skimage.segmentation.html#skimage.segmentation.inverse_gaussian_gradient