Towards computer-aided diagnosis of sacroillitis from CT scans: segmentation, classification, and preliminary results

CASMIP Lab: Computer Aided Surgery and Medical Image Processing Laboratory

School of Computer Science and Engineering
MSc Laboratory Project

Eliel Hojman

Advisor: Prof. Leo Joskowicz
Collaborators: Dr. Arnaldo Mayer, PhD, Prof. Iris Eshed, PhD MD, Sheba Medical Center.

October 18, 2016

Abstract

Sacroiliitis is the inflammation of the sacroiliac joint, characterized by a lower back pain. Early diagnosis of the condition can lead to preventive treatment which can significantly improve the patient's quality of life in the long run. However, if diagnosed late, the damage in the joint is irreversible, and the patient may suffer from chronic pain in the lower back as well as limited mobility. The detection of the condition in its early stages is a difficult task, as the differences between a healthy and inflammated joint in the early stages are very subtle. In this report we describe a method for the automatic segmentation and diagnosis of the sacroiliac joint (SIJ) in CT images. The segmentation is computed using the min-cut algorithm, and the classification is obtained with the bag-of-words method. Our experimental results on 300 clinical CT images yield a successful segmentation for 61.6% of the cases and a classification accuracy of 55.3% in the healthy joints and of 63.6% for the pathological cases. These preliminary results indicate that computer aided diagnosis may be a useful tool for assisting clinicians in the early diagnosis of sacroiliitis.

1 Introduction

Computer-aided diagnosis (CAD) has become one of the major research subjects in medical imaging and diagnostic radiology [18]. The idea behind CAD is to use the diagnose of the computer not as a final decision, but as a "second opinion" for the radiologist to consider [19]. The weight that the radiologist will give to this "second opinion" will be directly correlated with the performance of the computer. In this report we develop a CAD method for the segmentation and classification of the medical pathology known as sacroiliitis. Below we detail the characteristics of the condition and describe the goals of the project.

1.1 Clinical background

The sacroiliac joint (SIJ) is the joint between the sacrum and the ilium, its function is to transmit all of the forces of the upper body to the pelvis and legs, and also act as a shock-absorbing structure. It is capable of only some small movement.

Sacroiliitis is the inflammation of the sacroiliac joint, it is characterized by lower back pain which can spread to the buttocks, legs, groin and sometimes even feet. As the symptoms are similar to those of more common conditions, such as an herniated disc, this makes the accurate diagnosis of sacroiliitis a difficult task [7].

According to the New York grading criteria[1], a known standard for diagnosing the stiffness of the SIJ, medically referred as ankylosis spondylitis (AS), there are five different grades that can be assigned when reading a medical image showing the joint:

- Grade 0: Normal, no disease
- Grade 1: Some blurring of the joint margins, suspicious
- Grade 2: Minimal sclerosis with some erosion
- Grade 3: Severe erosions
- Grade 4: Complete ankylosis

1.2 Project goals

Early diagnosis of sacroiliitis can lead to very successful preventive treatments which can delay and even stop the degeneration of the joint. However, the only way to identify the condition in its early stage is by a specialized radiologist, which needs to analyze in depth the results of the CT scan, a process that can take several minutes. Due to the small number of specialists and the large

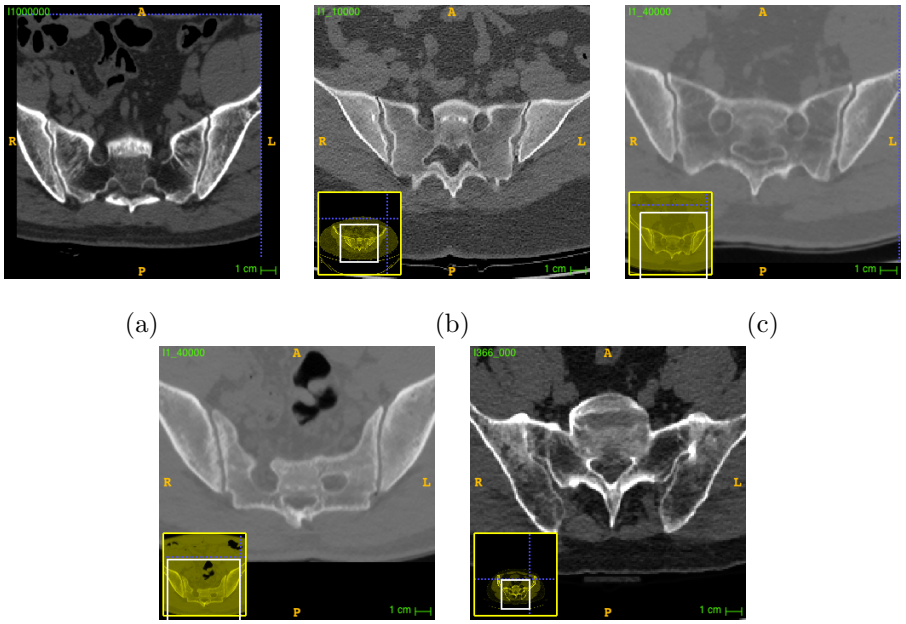


Figure 1: (a) - Grade 0, (b) - Grade 1, blurred edges both on the right and left side, (c) - Grade 2, Erosion on the right joint, (d) - Grade 3, Many erosions clearly identified in the right joint, (e) - Grade 4, Complete stiffness both on the left and right joints.

amount of patients, it is not uncommon that some patients are left undiagnosed. These patients usually come back years later presenting chronic lower back pain or any other symptoms, but unfortunately the damage on the joint has already been done, and may now be irreversible.
The goal of our project is to aid the work of the radiologist by segmenting the region of interest, and identifying image patches which show features proper of the disease. After analyzing all of the patches, the joint is classified as healthy or sick.

2 Related work

The incidence percentage of ankylosing spondylitis (AS), the highest grade given to sacroiliitis, is 0.1% to 1.4%, depending on the population studied [8]. The higher rate being found among Caucasians. It has also been found that AS affects more men that women, with a ratio of 2 to 1 [10], and that it is the cause for 4% to 5% of all the patients with chronic low back pain [9]. Most of the patients which suffer from AS develop their first symptoms at an age younger

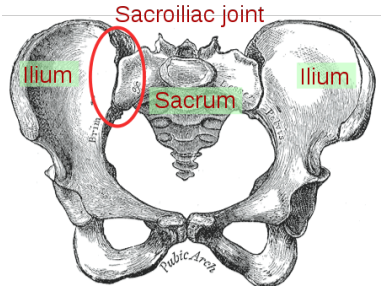


Figure 2: Sacroiliac joint

than 30.

The first criteria for diagnosing sacroiliitis was developed in in 1961 in Rome [17] and it did not include any kind of imaging of the SIJ to make a diagnosis, the disease was only diagnosed according to the restriction of motion of the lumbar spine. Later in 1966, in New York, radiographic evidence of sacroiliac inflammation was included as a requisite for the diagnosis [12], and the grading system mentioned above was introduced. The last modification of the New York criteria occurred in 1984 [13], which it introduced two new clinical criteria for diagnosing sacroiliitis, those are lower back pain for a period of three months and restriction of chest expansion.

Today image is crucial for the diagnosis and classification of sacoriliitis. And for this purpose each imaging technique has its advantages and disadvantages. Conventional x-ray still remains the most utilized method in the clinical practice for the detection of AS [11], but it is known to have low sensitivity for detecting the early stages of the disease [14]. CT in comparisson to standard x-ray has shown to be more sensitive in detecting bone erosions and joint space narrowing. It is comparable to MRI on the detection bone erosion, but superior for evaluating bone sclerosis. On the other hand MRI has the advantage of having better image quality, absence of radiation and a better detection of the cartilage of the sacroiliac joint[16].

3 Methods

We present next a method for the segmentation and classification of the sacroiliitis. Our method receives as input a CT scan, and after being processed, it is given a diagnosis on sacroiliitis based on a classifier which was trained using the bag of words model. The full flow for processing the CT scan is described below:

1. Segment the joint area

2. Extract image patches from the boundary of the sacrum and ilium bone

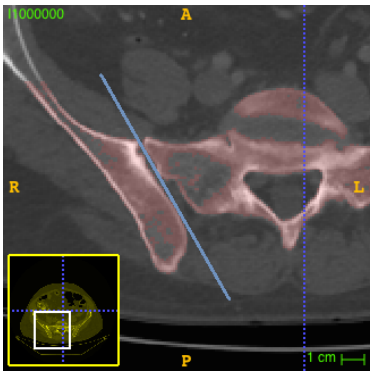


Figure 3: Min Cut algorithm

and the joint

3. Use the trained classifier to label each patch as either healthy or sick.

4. Process all of the labels of the image patches to deliver a diagnosis of the joint.

3.1 Segmentation

The objective of the segmentation is to identify the location of the boundaries between the sacrum/ilium and the joint. The algorithm is designed to process both the right and the left side of the sacrum. The main steps are:

1. Skeleton segmentation

2. Hips bone segmentation

3. Sacrum/ilium bone segmentation through min-cut algorithm

4. Segmentation of sacroiliac joint boundaries in the axial CT slices

5. Extraction of image patches in the sacroiliac joint region of interest

The min-cut algorithm was chosen for the task of segmenting the bone structures of the SIJ because given the similarity of the gray values of the sacrum, ilium and joint, the edge detection methods are expected to perform poorly. To segment the structures a connectivity graph of all the relevant voxels is used, the cut is effectuated through the darker voxels corresponding to the joint, thus segmenting the different bones structures.

3.1.1 Hips segmentation

The first step of the algorithm is the identification of the hips ROI. The segmentation of the skeleton is obtained using the following heuristic. For all values

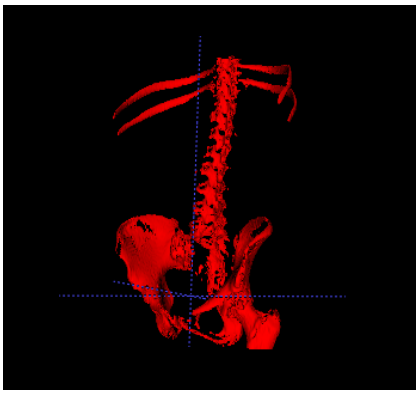


Figure 4: Skeleton segmentation

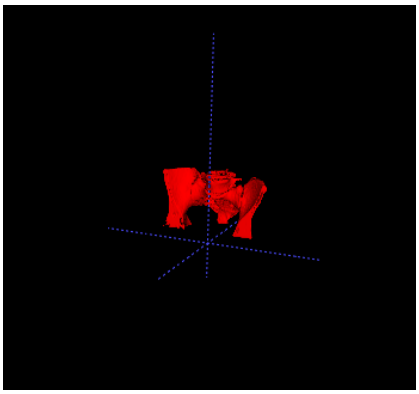


Figure 5: Hips segmentation

from 150 to 500 with step of 20 all of the voxels which their gray value is above the specified threshold and below a constant threshol,d which was set to be 1300, are taken. On all of these voxels the number of connected components is calculated, the lower threshold is then set according to the smallest number of connected components delivered. Once the minimum threshold is set dilation and erosion of 7 voxels in all of the three coordinates is performed, the segmentation of the skeleton is the resulting connected component with the largest number of voxels.

As the hips are much wider than the spine, the start of the bone can be identified by searching for a sudden change in the width of the convex hull of the skeleton pixels previously identified. Starting from the bottom of the body, every 10 axial slices the width of the convex hull is measured, if a difference of at least 60% is found between two consecutive measures, the corresponding slice is set to be the start of the hips.

As the CT scans which are zoomed into the joint area usually leave out of the image part of the hips, for this cases we expect a difference of only 30% between the convex hulls. The zoomed CTs are identified by searching for pixels which belong to the hips structure and are located at the border of the image.

3.1.2 Sacrum Ilium segmentation

The second step is the segmentation of the sacrum and ilium bone. Each is segmented separately, for both the right and left sides.

The segmentation is obtained using the min-cut algorithm [2]. A graph connecting all of the voxels obtained from the hips segmentation is constructed and the weights of the edges are assigned according to the gray value of the voxels. As the area of the joint separating the sacrum from the ilium has lower gray values, the min-cut algorithm cuts the connectivity graph exactly through it (Figure: 3). In order to execute the algorithm some "source" and "sink"

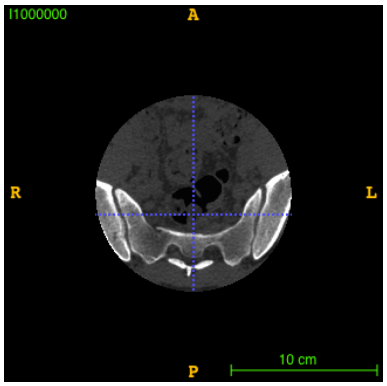


Figure 6: CT scan in which part of the hips are left out

nodes must be first assigned, these are the nodes which are constrained to be left in different connected components after the cut has been effectuated. These are the voxels which are sure to belong to the sacrum and to the ilium.

For the construction of the connected graph each voxel is connected to 10 limiting neighbors, the eight neighboring voxels from the same z plane and two voxels from the above and below plane in the same xy location. The weight of an edge is set according to the minimum gray value of the pair of connecting voxels, for the edges connecting the voxels among the z plane their weight is elevated by the power of two, this is done in order to discourage large jumps in the xy location of the cut. When this penalty was not added the cut was not continuous, but rather jumping in each slice, searching for the areas with low nodes density.

For the identification of which voxels are sure to belong to the sacrum and ilium three different techniques were used.

1. Location. An area is delimited in the center of the CT scan which is considered safe to define it as part of the sacrum. The same is repeated for the ilium (Figure: 7).

2. Identifying components. The previous selection is extended by considering the connected components in each of the axial images. If any of the pixels of a component was already assigned to the sacrum or to the ilium, all of the rest of the pixels of the component are added to the corresponding selection (Figure: 8).

3. Closed ring. For the cases when the connected component contains both pixels from the ilium and the sacrum, a close ring inside of the bone structure is considered for the extension. If any of the pixels in the closed ring belong to only the sacrum or the ilium all of the pixels of the closed ring are added to the selection. (Figure: 9).

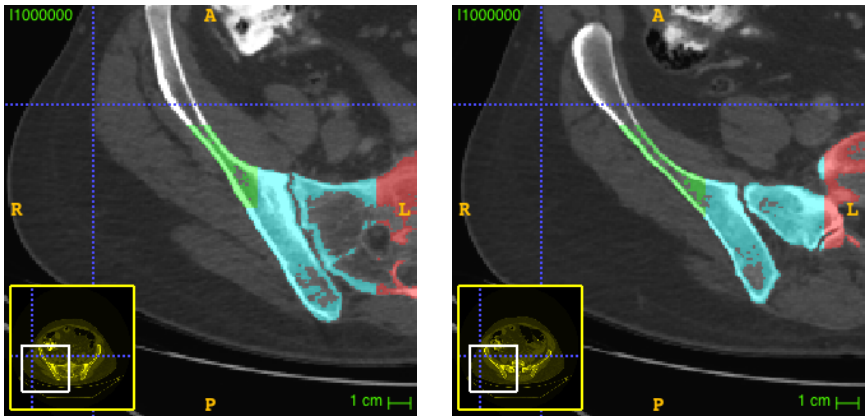


Figure 7: Skyblue are the still unidentified voxels, red are the voxels assigned to the sacrum, green are the ones assigned to the ilium

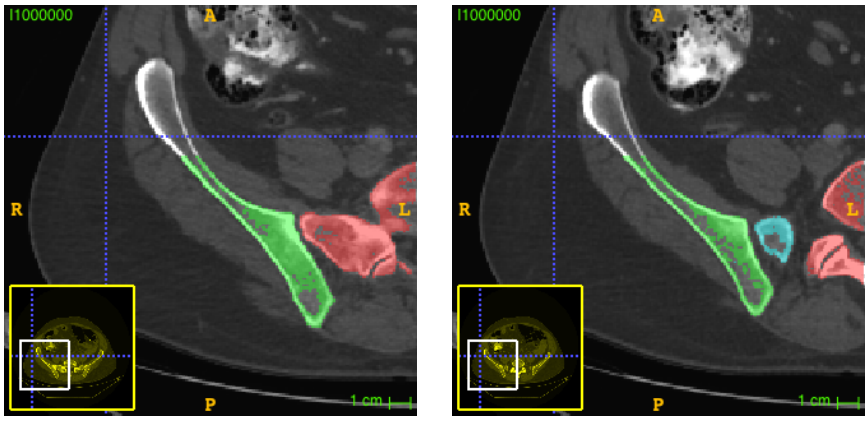


Figure 8: Example of extending the previous defined values to all of the component

After doing the automatic identification for some voxels, the min-cut algorithm is executed and the label of each unidentified voxel is obtained (Figure: 10).

3.1.3 Region of interest identification

The next step is to identify the area of the joint in contact with the bones. The joint begins in the front of the hips and extends along the space between the sacrum and the ilium. Its location can be identified by searching for a narrow area between the sacrum and ilium which has its boundaries running along in parallel one in front of the other.

The segmentation of the relevant boundary is computed by dilating the opposite bone of the one being work on, for example, when searching the boundary

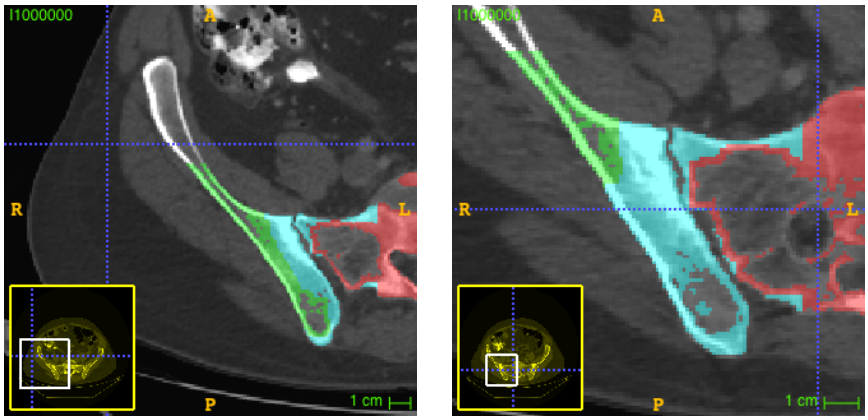


Figure 9: Notice the closed rings areas both for the sacrum and the ilium

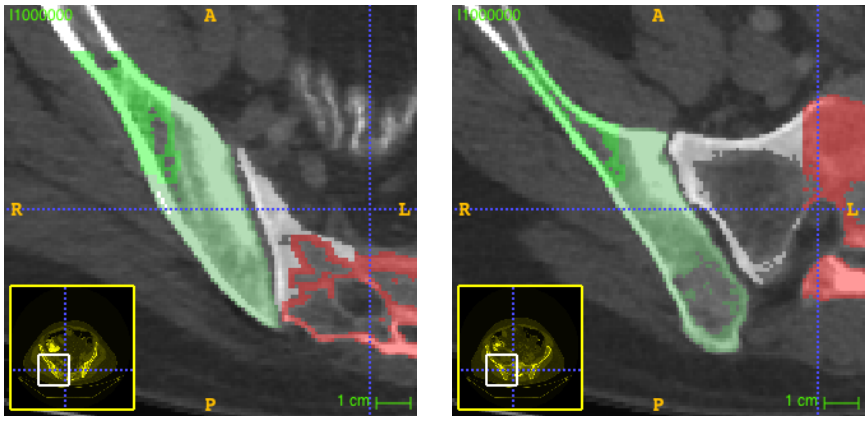


Figure 10: The dark green and pink voxels are the resulting labels for the previously unidentified nodes of the mincut algorithm

of the ilium, the previously found segmentation of the sacrum is dilated. The dilation being computed is of 3.5mm. The traversal distance along the joint being considered is of 35mm. Both of these distances were advised by the radiologist (Figure: 11).

It is from the lines showed in figure 12 that the images to be used for the classification are extracted.

3.1.4 Extracting the patches

The final step of our segmentation is the extraction of the patches along the boundaries of the joint. After segmenting the relevant boundaries, each axial slice is analyzed and for each boundary a number of points are selected. The selection is performed by taking one point along the boundary every 2mm. Each patch is constructed by taking an area of 5×5mm around each selected point.

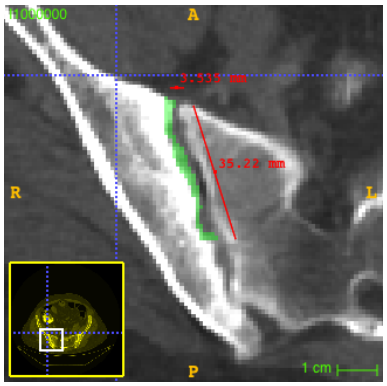


Figure 11: Intersection between the extend sacrum and the ilium voxels

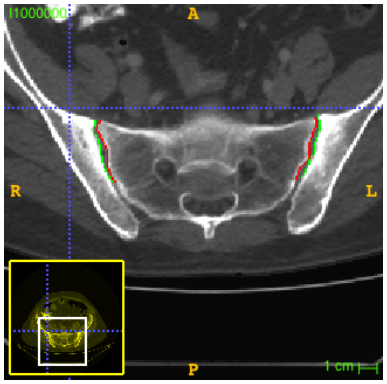
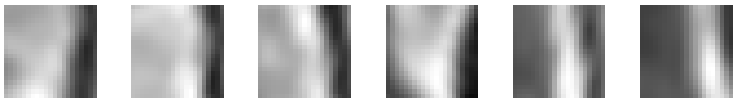


Figure 12: Final segmentation to be used for extracting image patches

This area is then resized to 21×21 pixels.

As the bone area being shown in the patches is complex and has its own patterns and particularities, there could be a case where regions inside the bone may look similar as a eroded region. To limit these misidentifications it was considered best to work with features which are variant to rotation. As result of this all of the image patches inputted into the classifier should be aligned in the same relative angle. To achieve this the patches are processed in the following way:

- Rotate the images so that the boundary of the bone is set in a vertical orientation.
- Flip the images, if needed, so that the boundary of the bone will be located on the right of the image.



(a) (b) (c) (d) (e) (e)

Figure 13: Examples of automatic obtained images

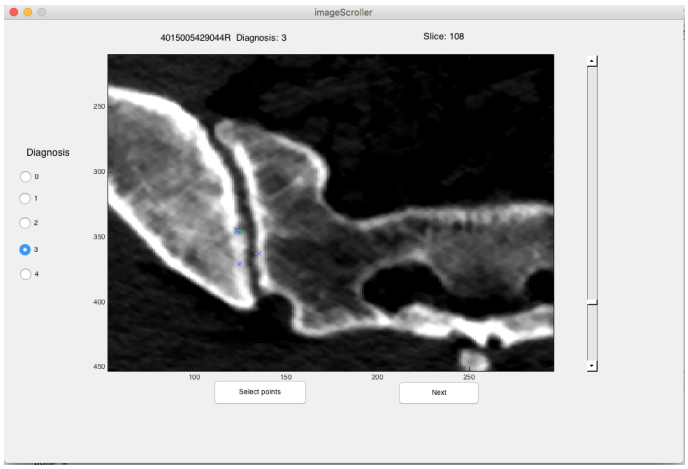


Figure 14: Screenshot of the created GUI

3.2 GUI for selecting points of interest

To create the trainset a GUI was designed in MATLAB to enable sorting among the already segmented joints and selecting points which characterize the disease. The selection of the points was performed by the author after receiving indications from Dr. Eshed on how to identify the points of interest.

The display area focuses automatically on the segmentation of the boundaries of the bones acquired previously and the scroll bar enables the user to select different axial images. The radio-buttons specify which grade should be given to the selected points. Once all of the points on a SIJ have been selected, the "Next" button displays the next SIJ segmentation.

The points are stored as a struct holding the xyz coordinates, the side (left or right), the assigned grade, and the name of the original CT scan file from where it was taken. This information is used for constructing the image patches.

The selection of the points of interest was performed only for the 91 successfully segmented joints which were diagnosed by the radiologist to have the disease.

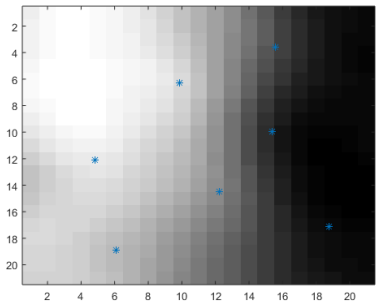


Figure 15: Detection of points of interest

3.3 Classification

The classification was performed using the bag-of-words method [3]. This method is composed of three main steps.

1. Detection and description of image patches. The detection of interesting points was computed using the minimum eigenvalue algorithm for corner detection [6], which proved to deliver very good results. A SURF descriptor was then used to extract the features of the location [4].

2. Assign the patch descriptors to a set of predetermined clusters. The computed centers of the clusters represent the words of the vocabulary. The number of words in the vocabulary needs to be large enough so as to hold all the relevant descriptors from the categories, but not large enough so as to make the classifier overfit to the noise in the images. After some validation procedure using a trainset and testset and trying difference vocabulary sizes, the vocabulary size was set to 200.

3. Create a classifier where each feature is the number of times each word appeared on the image. For this purpose a SVM [5] classifier was used.

3.4 Joint classification

The previously described classifier works only on single images, however, the main goal of the project is the classification of a joint. As explained above a joint is modeled as a large amount of images, therefore to obtain the classification of the joint, the predictions obtained for all of the patches need to be summarized. If the image classifier would have a 100% accuracy on the healthy images, it would be enough to identify one image as "sick" to state that the joint has the pathology. As this is not the case, something needs to compensate for the error margin in the image classification.

A joint delivers in average 500 image patches, for healthy joints all of the images should be classified as healthy, but for sick joints it could be that only in 10 patches the inflammation is noticeable, and all of the rest look as healthy patches. It is for this reason that in order to obtain a better classification of the joints more rigorousness is needed when classifying sick images.

The implementation of the SVM delivers a score value expressing the similarity of the image being classified to one of the categories of the classifier. It then performs the classification by selecting the category with the largest score. To reduce the misclassification of healthy images a fixed handicap is subtracted from the score of the sick category. This way only images which are strongly identified as sick are now labeled as such, all the rest is classified as healthy.

4 Experimental study

4.1 Dataset

The dataset is composed of 150 CT scans, which translates to 300 different SIJ, each with its corresponding diagnosis according to Dr. Eshed. The 300 datasets consist of 119 joints diagnosed as grade 0, 51 as grade 1, 70 as grade 2, 47 as grade 3 and 13 as grade 4.

4.2 Segmentation Results

The evaluation of the segmentation was performed manually. Two different categories were defined, successful and unsuccessful. The evaluation was performed by selecting axial images on different slices and verifying that the cut obtained was along the location of the joint for at least 90% of the images, if for the rest 10% the error was minimal, the segmentation was considered to be successful. The reason a perfect segmentation is not necessary is because the extraction of relevant image patches from the boundaries of the sacrum and ilium can still be achieved even when there is a small error in the location of the cut.

For the 300 joints of the dataset, a successful segmentation was obtained for 61.6% of them. This amounts to 82 for grade 0, 36 for grade 1, 38 for grade 2, 23 for grade 3 and 0 for grade 4, the success rate for each grade is 68.9%, 70.5%, 54.2%, 48.9% and 0% correspondingly.

4.3 Image classifciation results

The trainset for the classifier of images was build using 108 points with grade 1, 189 points with grade 2 and 199 points with grade 3, amounting a total of 496 "sick" patches. The "healthy" patches were taken automatically as stated in 3.1.4 from 7 healthy joints, for a total of 3175 "healthy" patches. As explained

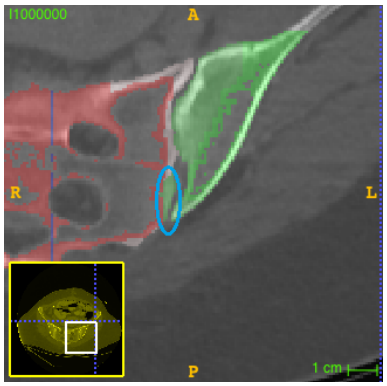


Figure 16: Tolerated misssegmentation

in section 3.4 specificity holds a higher weight than sensitivity, it is for this reason that much more healthy patches were used when training the classifier.

The classifier was trained using 80% of the patches, and its performance evaluated using the rest 20%. The accuracy for the training and test set can be seen in Tables 1 and 2.

Table 1: Classifier accuracy for training set

	Grade 0	**Grade 1**	**Grade 1**	**Grade 1**
Healthy	**80.7%**	15.2%	16.8%	19.5%
Sick	19.3%	**84.8%**	**83.2%**	**80.5%**

Table 2: Classifier accuracy for test set

	Healthy	**Grade 1**	**Grade 2**	**Grade 3**
Healthy	**78.5%**	33.9%	24.2%	31.4%
Sick	21.5%	**66.1%**	**75.8%**	**68.6%**

4.4 Joint classification results

As explained in 3.4, the classification accuracy of the healthy images needs to be boosted. To achieve this a threshold is set so as to deliver an accuracy of 98% for the classification of healthy images. The results show that if the score of the SVM for the sick classfication is subtracted 1.4, a 97.9% accuracy is achieved for the healthy images and a 13.1% for the sick ones. As the possibility of misclassifying a healthy image still exist, the condition that at least 2.3% of all the patches should be classified as sick in order to classify the analyzed joint as sick is added. This threshold was set by taking half of the segmented joints to be used as trainset and setting the threshold value which delivers the best

results. The threshold was then validated using the other half of the dataset. Tables 3 and 4 show the final results on the joint classification.

Table 3: Joint classification - Train set accuracy

	Grade 0	**Grade 1**	**Grade 2**	**Grade 3**
Healthy	**69.3**	45.8	46.6	39.3
Sick	30.7	**54.2**	**53.4**	**60.7**

Table 4: Joint classification - Test set accuracy

	Grade 0	**Grade 1**	**Grade 2**	**Grade 3**
Healthy	**55.3**	40.2	42.9	25.1
Sick	44.7	**59.8**	**57.1**	**74.9**

As can be seen in the tables above, the train and test set results are not tightly correlated, as it can be seen by the accuracy fall from 69.3% to 55.3% in the grade 0 classification, but for all of the grades we still manage to achieve an accuracy of above 50%, which is no minor task given the fact that the classification is computed over a large amount of images. The best results for the sick joints are found for the Grade 3, which is what we would expect as those are the joints which present the largest amount of erosions.

5 Conclusions

The overall results of the joint classification are not as impressive as we hoped them to be. However, we think we have marked the path on how the automatic segmentation and classification of the sacroiliitis could work, and managed to achieve results which are better of those of a random guess for such a challenging task.

In order to improve the results of the segmentation we could first try to automatically identify the cases which have been badly segmented, and rerun the algorithm for the segmentation with different parameters. The classification of the segmentations could use features such as the weight of the cut from the min-cut algorithm, as well as the length and direction of the detected boundaries of the sacrum and ilium.

For the classification we could try to increase the size of the training set of the "sick" images. Another idea would be to retrieve some more points of interest in each image. Instead of only using the corners detection, we could detect the edge of the boundaries and have some points aligned on top of it, and

derive from there new SURF descriptors. These points could be trained using another classifier and the results of both combined.

References

[1] van der Linden S, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis: a proposal for modification of the New York criteria. Arthritis Rheum 1984;27(4):361368. CrossRef, Medline

[2] Boykov, Yuri, and Vladimir Kolmogorov. "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision." Pattern Analysis and Machine Intelligence, IEEE Transactions on 26.9 (2004): 1124-1137.

[3] Csurka, Gabriella, et al. "Visual categorization with bags of keypoints." Workshop on statistical learning in computer vision, ECCV. Vol. 1. No. 1-22. 2004.

[4] Bay, Herbert, et al. "Speeded-up robust features (SURF)." Computer vision and image understanding 110.3 (2008) 346-359.

[5] Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks." Machine learning 20.3 (1995): 273-297.

[6] Shi, J., and C. Tomasi, "Good Features to Track," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 1994, pp. 593600.

[7] http://www.mayoclinic.org/diseases-conditions/sacroiliitis/symptoms-causes/dxc-20166359

[8] Lawrence R, Helmick C, Arnett F, et al: Estimates of the prevalence of arthritis and selected musculoskeletal disorders in the United States. Arthritis Rheum 1998;41:778-799

[9] Calin A, Garrett S, Whitelock H, et al: A new approach to defining functional ability in ankylosing spondylitis: The development of the Bath Ankylosing Spondylitis Functional Index. J Rheumatol 1994;21:2281-2285.

[10] Feldtkeller E, Khan MA, van der Heijde D, van der Linden S, Braun J. Age at disease onset and diagnosis delay in HLA-B27 negative vs. positive patients with ankylosing spondylitis

[11] Braun, Jrgen, and Joachim Sieper. "Ankylosing spondylitis." The Lancet 369.9570 (2007): 1379-1390.

[12] Bennett PH, Burch TA. Population studies of the rheumatic diseases. Amsterdam: Excerpta Medica Foundation, 1968: 45657

[13] van der Linden S, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modifi cation of the New York criteria. Arthritis Rheum 1984; 27: 36168

[14] Montandon, Cristiano, et al. "Sacroiliitis: imaging evaluation." Radiologia Brasileira 40.1 (2007): 53-60.

[15] Braun J, Sieper J, Bollow M. Imaging of sacroiliitis. Clin Rheumatol 2000;19:5157

[16] Puhakka KB, Jurik AG, Egund N, et al. Imaging of sacroiliitis in early seronegative spondylarthropathy. Assessment of abnormalities by MR in comparison with radiography and CT. Acta Radiol 2003;44:218229

[17] Kellgren, Jonas Henrik. The epidemiology of chronic rheumatism: a symposium. Eds. Maurice Rutherford Jeffrey, and John Ball. Vol. 2. FA Davis Company, 1963.

[18] Giger ML, Huo Z, Kupinski MA, Vyborny CJ. Computer-aided diagnosis in mammography. In: Fitzpatrick JM, Sonka M, editors. The Handbook of Medical Imaging, volume 2 Medical Imaging Processing and Analysis. SPIE; 2000. pp. 9151004

[19] Doi K. Computer-Aided Diagnosis in Medical Imaging: Historical Review, Current Status and Future Potential. Computerized medical imaging and graphics: the official journal of the Computerized Medical Imaging Society. 2007;31(4-5):198-211. doi:10.1016