Automatic Segmentation of Skeletal Muscle Tissue in L3 CT Images Based on Random Forests and CNN Using Coarse Ground Truth Masks

Domingos B. S. Santos[1], **Gabriel F. L. Melo**[1] **and Thelmo de Araujo**[1]

[1]Universidade Estadual do Ceará - UECE

{domingos.bruno,gabriel.lins}@aluno.uece.br, thelmo.araujo@uece.br

Abstract. *Estimates of the composition of skeletal muscle tissue (SMT) and adipose tissues are important in the treatment of debilitating diseases, such as cancer, and in the control of overweight and obesity. Several studies have shown a high correlation between the percentage of SMT in computed tomography (CT) images corresponding to the cross-section at the level of the third lumbar vertebra (L3) and the percentage of this tissue in the whole body. A large number of models has been proposed to automatically segment CT images in order to estimate tissue compositions, many of them use supervised Machine Learning (ML) methods, such as neural networks, which require large amounts of* labeled *images, i.e., images and ground truth masks obtained from manual segmentation by human experts. These large labeled datasets are not easily available to the public, thus the present work proposes a methodology capable of performing the automatic segmentation of SMT in single-slice CT images (at L3) using only "coarse" segmentation masks as ground truth in the ML algorithms's training phases. By "coarse segmentation" we mean a* semiautomated *segmentation performed by a person without specialized knowledge of human anatomy. The proposed methodology oversegments the image into superpixels, which are classified by a Random Forest (RF) model. Then, a U-Net CNN refines the classification, using as input the pixels in the superpixel segments classified as SMT by the RF. The methodology achieved* 99.21% *of the accuracy obtained by the same CNN trained with golden standard ground truth masks, i.e., segmentation masks manually created by a medical expert.*

1. Introduction

Evaluation of the composition of Skeletal Muscle Tissue (SMT), Subcutaneous Adipose Tissue (SAT), and Visceral Adipose Tissue (VAT) may help to identify obesity (excess body fat accumulation), sarcopenia (low muscle mass), and sarcopenic obesity (muscle loss with excess body fat). These medical conditions may be relevant in the treatment of debilitating diseases, such as cancer, since they may be associated to muscle loss and fat accumulation [Schweitzer et al. 2016]. Furthermore, knowing body composition may help to understand survival rates in cancer patients [Caan et al. 2018a].

Anthropometric measurements, such as waist-to-hip ratio (WHR) and body mass index (BMI), are noninvasive methods widely used in routine clinic evaluation of body composition. However, these measures do not quantify the different types of fat and are not very accurate [Caan et al. 2018b], [Ding et al. 2016]. Body composition measurements using images, such as computed tomography scan (CT), are considered the golden standard in computing tissue compositions [Paris et al. 2020].

Many studies have shown the high correlation between the percentages of SMT, SAT, and VAT in the CT image corresponding to the region of the third lumbar vertebra (L3) and the percentages of these tissues in the whole body [Ross et al. 1992], [Schweitzer et al. 2015]. The obvious advantage of this approach is that only a single image (as opposed to the whole set of CT images) has to be analyzed in order to estimate tissue compositions.

In order to estimate the tissue composition in a CT-L3 image, one must calculate the number of pixels corresponding to this tissue. Counting pixels for each tissue depends on segmenting the image into the regions corresponding to the tissue. A fair amount of Image Processing and Machine Learning (ML) techniques have been developed to this purpose, many of which are *supervised* ML techniques, such as Neural Networks [Dabiri et al. 2019], that requires the corresponding segmentation masks, created by human experts, to be used as ground truth in the training phase of the algorithm.

Analysis of a CT image requires the manual segmentation of image regions by an expert with knowledge in human anatomy. Radiodensity information—given in Hounsfield units (HU) in DICOM images—may be used to segment fat and muscle tissues, but it cannot discriminate subcutaneous fat from visceral fat, or skeletal muscle from viscera. Therefore, the expert must use the knowledge about human anatomy to decide whether a pixel corresponds to a SMT or to a kidney, for example. The scenario may be even more difficult when SMT and viscera are touching in the image. These difficulties may explain why publicly available datasets with *labeled* CT-L3 images are rare. Until the completion of this paper, the authors were not aware of publicly available datasets with CT-L3 images labeled with SMT, SAT, and VAT masks.

In [Melo 2022], the authors obtained good results in segmenting SAT in CT-L3 images without using supervised ML methods. The proposed method was based only on the radiodensity information and the relative location of the SAT in the body image. If one were also able to segment the SMT, the segmentation of the VAT would be much easier.

Considering the difficulty to manually segment the SMT of a large number of CT-L3 images by an expert in human anatomy, we propose:

- A methodology to manually create a *coarse* segmentation of SMT in CT-L3 images, to be done by a user with low knowledge in human anatomy and in a fast way. The methodology uses the SLIC (Simple Linear Iterative Clustering) superpixel algorithm. A set of label masks shall then be created to be used as *coarse* ground truth in the supervised ML methods that follow.
- A methodology to classify the *superpixels* created by the SLIC algorithm into SMT and non-SMT classes, using Random Forests (RF), and the coarse masks as ground truth in the RF's training phase.
- A methodology to refine the segmentation obtained by the RF at pixel level, using a Convolutional Neural Network (CNN), which shall also use only the coarse masks as ground truth in its training phase.

It is worth stressing that, although segmentation masks created by a medical expert were available in the dataset used here, they were never used in the ML algorithms as ground truth. They were used only to measure the accuracy of the segmentation.

This work then proposes methodologies to automatically segment SMT in CT-L3

images using only coarse masks as ground truth in the ML algorithms. In order to analyze the quality the segmentation, our results were compared with the results obtained by the same CNN using the medical expert segmentation masks as ground truth.

2. Related Work

Caan et al. showed that skeletal muscle composition is an important predictor of mortality in cancer patients, since overweight (BMI between 25 and 30 kg/m^2) and obesity (BMI above 30 kg/m^2) are well stablished risk factors for cancer [Caan et al. 2018b]. Both overweight and obesity are associated with a chain of metabolic and endocrine changes related to cancer development, including insulin resistance, systemic inflammation, and hormone and growth level alterations. The study affirms that the sole use of BMI is not enough to identify the risk of mortality and suggests that patients with moderate levels of subcutaneous fat may be protected by higher levels of skeletal muscle. In another paper, Caan et al. evaluated the association of the percentages of subcutaneous, visceral, and intra/intermuscular fat tissues, estimated in CT images, with the general mortality in non-metastatic breast cancer patients [Caan et al. 2018a]. The authors affirm that sarcopenia, poor muscle quality (low muscle radiodensity), and excess adiposity have not yet been related to the prognosis in patients with nonmetastatic breast cancer, as opposed to their well established relation with a higher mortality in patients with metastatic breast cancer. Among the patients diagnosed with nonmetastatic breast cancer, those with sarcopenia presented a higher mortality rate than the ones without sarcopenia.

Accordingly to [Dabiri et al. 2019], in diseases like cancer, patients suffer from degenerative skeletal muscle loss, which may also happens in the advanced stages of aging. The authors proposed a novel model, based on deep neural networks, to automatically segment skeletal muscle tissue in axial CT images at the third lumbar vertebra (L3) and the fourth thoracic vertebra (T4). A two-branch neural network with two training stages is analyzed and its performance in three different datasets is measured. The best model achieved Jaccard indices ranging from 93.70% to 98%. In order to achieve such high accuracies, the authors applied a highly specialized CNN to large datasets; here we used a more general CNN with a modest size dataset, a much more common scenario.

An automated method to detect organs was developed in [Liu et al. 2018], based on CNN, superpixels, and Support Vector Machine (SVM). First, the SLIC algorithm was applied, then a feature vector was extracted from each superpixel and classified, by the SVM, in the class corresponding to the organ. The CNN receives the resulting pixels and refines the classification, finding the final segmentation of the organ. In a dataset of CT images, the liver was segmented by the model, which achieved a detection accuracy of 97.43%. The model was applied to a dataset with high resolution CT images containing the lungs, and obtained a detection accuracy of 97.93%. The authors also presorted the images into 10 different organ sizes. In our proposal, lower resolution images were used and the coarse ground truth masks were the only information provided for ML training.

3. Methodology

The methodology of this work is comprised of four steps, as shown in Figure 1, which are detailed in the following subsections. In the first step, the CT-L3 images are read and preprocessed: noise and artifacts are removed from the images using the technique

developed in [Melo 2022], which uses the radiodensity information of the original CT-L3 image; and an adaptive histogram equalization algorithm (CLAHE) is used to increase image contrast. The second step consists of oversegmenting the filtered image using the SLIC (Simple Linear Iterative Clustering) superpixel algorithm [Achanta et al. 2010]. By visual inspection, the superpixels corresponding to the SMT are manually segmented, resulting in a *coarse* segmentation mask. This coarse mask will be used in the training phase of both RF and CNN algorithms. In the third step, a RF classifier is trained (using the coarse segmentation masks as ground truth) in order to classify a superpixel as SMT or non-SMT. Finally, in the fourth step, a CNN is used to refine the segmentation obtained by the RF classifier: the pixels belonging to the superpixels classified as SMT by the RF algorithm are subjected to the CNN, using again the coarse segmentation masks as ground truth. A fast post-processing scheme is applied to return the final segmentation.

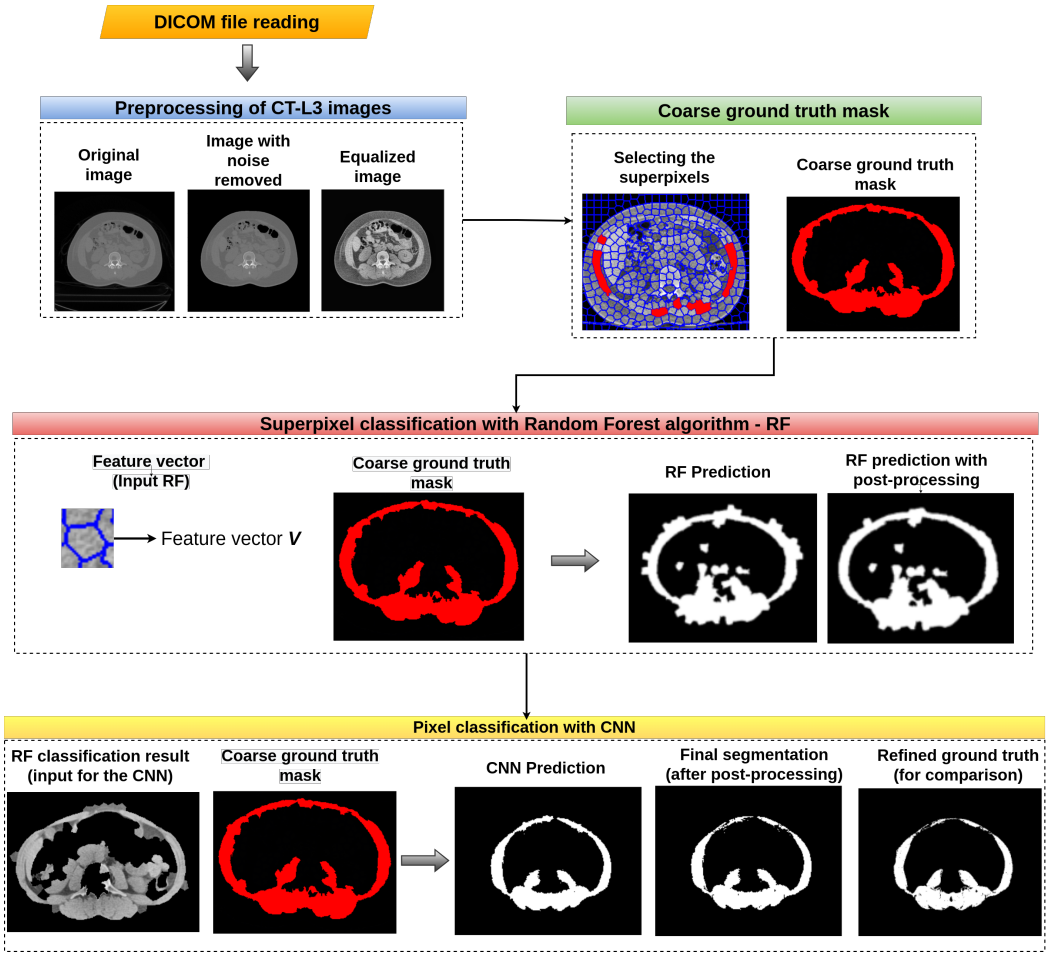


Figure 1: Diagram of the proposed methodology for skeletal muscle segmentation.

3.1. Dataset

The dataset used in this work consists of 203 anonymized CT-L3 images in DICOM format, with dimensions 512×512. The images are from female patients with ages between 23 and 39 years old. The dataset was kindly provided by Dr. Sara Maria M. Lima Verde, from the Bachelor Science in Nutrition Program at the State University of Ceará (UECE), in Fortaleza, Brazil. The dataset was approved for research uses by the Research Ethics

Committee of Instituto do Câncer do Ceará—Technical Report Nº4.617.764 published by the Health National Board (Conselho Nacional de Saúde) under Resolution Nº466/12.

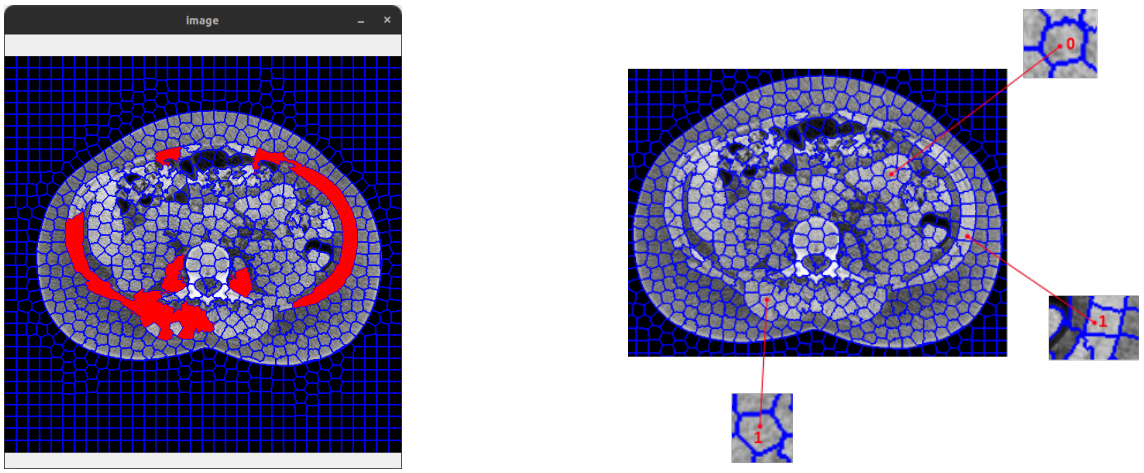
3.2. Preprocessing of CT-L3 images

Before the application of the SLIC superpixel algorithm, the CT-L3 images are preprocessed: First, all artifacts—such as the tomograph bed, clothes, and sheets—are removed, leaving only the region corresponding to the person's body. The technique applied was developed in [Melo 2022] and uses the radiodensity information in the original CT-L3 DICOM image. A binary air/non-air mask is created and the connected component with the largest area is kept, which corresponds to the person's body.

After performing noise reduction, the image contrast is increased by using the Contrast Limited Adaptive Histogram Equalization (CLAHE) technique in order to reinforce the objects's boundaries. CLAHE algorithm equalizes *local* histograms, using the clipped (limited) cumulative distribution function as the pixel transformation function. The resulting image has a much higher contrast and noise does not increase substantially.

3.3. Manual labeling with superpixels (*coarse labeling*)

Due to the difficulty to access publicly available CT-L3 images *with* the corresponding segmentation masks, a graphical interface was created to allow manual segmentation of the CT-L3 images. The interface receives the preprocessed image, as described in the latter paragraph, and applies the SLIC superpixel algorithm, showing the result to the user. The user selects, by mouse clicks, the superpixels corresponding to the SMT (the red regions in Figure 2a). In the end of the process, a binary mask of the SMT is created. Two are the reasons for calling this segmentation "coarse": There is no guarantee that the superpixels's boundaries generated by the SLIC algorithm perfectly coincide with the SMT's boundaries; and the fact that it was not performed by an expert on human anatomy.



(a) (b)

Figure 2: (a) Graphical interface for manual segmentation of SMT in CT-L3 DICOM images. (b) Superpixel segmentation; in the details, two superpixels are labeled as SMT (class 1) and one segment is labeled as non-SMT (class 0).

Using the superpixels as an aid to manually segment an image greatly reduces the workload of the human user for two reasons. First, because the number of superpixels

(one of the parameters of the SLIC algorithm) is much smaller than the number of image pixels: in this work, there is an average of 1200 superpixels as opposed the 262,144 pixels per image. The second reason is related to the time demanded by a human user to decide whether a single pixel close to the SMT boundary actually corresponds to it.

The graphical interface for manual segmentation was developed using Python's library *OpenCV* version 4.5.5.We used the SLIC algorithm implementation on the Python's library *Scikit-image* version 0.18.1. The default parameters were used in the SLIC algorithm, except for: `compactness=25`, `n_segments=1200`, and `max_iter=100`.

3.4. Classification of *superpixels* using RF

The binary segmentation task is split into two steps: In the first step, a Random Forest (RF) is trained in order to classify the *superpixels* generated by the SLIC algorithm into two classes (SMT and non-SMT), using the coarse ground truth. In a second step, the *pixels* on the superpixels classified as SMT by the RF feeds a Convolutional Neural Network (CNN) in order to be classified as SMT or non-SMT, using once more the coarse segmentation as ground truth. Let us emphasize: the RF classifies *superpixels* and the CNN classifies *pixels*; both use only the coarse segmentation masks as ground truth.

Seven commonly used classifiers were tested with the dataset: K-Nearest Neighbors (K-NN), MultiLayer Perceptron (MLP), Naive Bayes (NB), Support Vector Machine (SVM), Random Forest (RF), Decision Trees (DT), Logistic Regression (LR). All classifiers were implemented using the ML library *scikit-learn* with their respective default parameters.

The RF model achieved the best accuracy among the tested ML algorithms, therefore it was chosen as the superpixels classifier (see Table 1 in Section 4.1). A methodology was developed to automatically classify, using RF, the superpixels in two classes: SMT and non-SMT (see Figure 2b). Once the CT-L3 image is preprocessed and the superpixels are generated by the SLIC algorithm, the superpixels which largely intersect the muscle tissue (corresponding to the radiodensity of both skeletal muscle and viscera) are selected to be the input of the RF classifier. Using the coarse segmentation as ground truth, the RF is trained in order to classify the *muscle superpixels* into SMT and non-SMT classes.

A naive application of the RF classifier, using only a binary label (SMT and non-SMT), does not have a strong discriminant power, then 7-feature vectors are used as labels, following [Liu et al. 2018], to increase its discriminant power. Corresponding to each superpixel segment i, a feature vector S_i is created with the following coordinates:

$$S_i = [x_i, y_i, mean_i, var_i, std_i, med_i, mean\_edt_i]^T , \tag{1}$$

where (x_i, y_i) are the Cartesian coordinates of the superpixel's centroid; $mean_i$, var_i, std_i, and med_i are, respectively, the gray level intensity mean, variance, standard deviation, and median of the superpixel i. The last feature ($mean\_edt_i$) in the vector is the mean distance in a distance map (see Figure 3c), which is calculated using the Euclidean Distance Transform (EDT) and a binary mask formed by the pixels corresponding to the SAT and the all the pixels interior to it, excluding the region corresponding to the bone tissue (as seen in Figure 3b). The rationale for considering this distance map is that, in the CT-L3 images, the SMT pixels are closer to the SAT or to the bone tissue than are the viscera pixels. In other words, the probability of finding viscera in the CT-L3 image are

higher in the brighter region of Figure 3c than in its darker region. Using this feature in vector S_i improved the results obtained by the RF classifier.

(a) (b) (c)

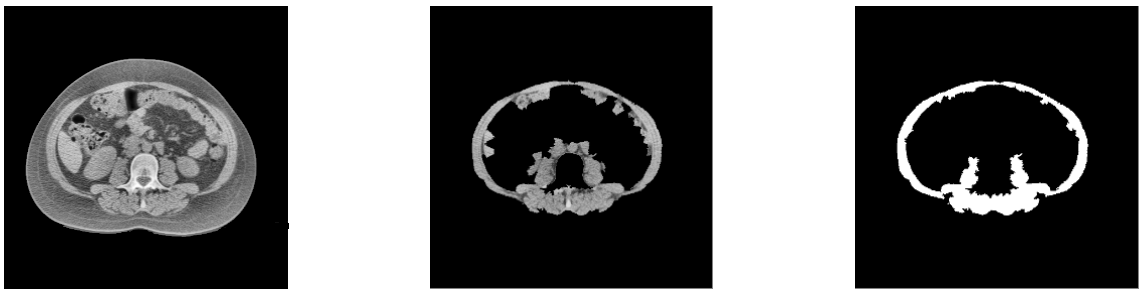
Figure 3: (a) CT-L3 image. (b) Binary mask. (c) Distance (EDT) map of the binary mask.

The superpixel binary classifier works as follows: All 203 CT-L3 images are manually segmented (as explained in Section 3.3), creating the coarse ground truth masks. For each superpixel (in all images), a feature vector S_i is created (as explained above) and the superpixel is classified as *class 1* (SMT) or *class 0* (non-SMT). This resulted in 16250 samples for class 1 and 71841 for class 0. Superpixels with mean radiodensity outside the muscle density range are discarded. This superpixel dataset is then balanced (16250 samples for each class) and split into training (67%) and testing (33%) sets.

After running the RF algorithm, a post-processing is applied, which consists on removing a superpixel segment classified by the RF as SMT if 50% or more of its area intersects the SAT mask, estimated by the method proposed in [Melo 2022].

3.5. *Pixel* segmentation using CNN

After the *superpixel* classification, the next step is to classify the *pixels* as SMT and non-SMT. In order to perform this refinement, a CNN is trained using as input the pixels inside the superpixels classified as SMT by the RF, and the coarse segmentation masks as ground truth. Known to be highly accurate, CNNs usually need large labeled datasets to perform well. Data augmentation methods are frequently used in order to compensate for datasets with modest size. Figure 4 shows (a) the CT-L3 image; (b) the pixels corresponding to the superpixels classified as SMT by the RF; (c) the coarse ground truth.



(a) (b) (c)

Figure 4: (a) An example of CT-L3 image. (b) The corresponding pixels in the superpixels classified as SMT by the RF. (c) The coarse ground truth.

U-Net is a CNN architecture proposed in [Ronneberger et al. 2015] for the specific task of medical image segmentation. The authors claim that U-Net may be trained with a relatively small set of images and still performs well. For these reasons, U-Net was the CNN architecture chosen for this work, and it was able to segment a image in less than a second, using a Graphics Processing Unit. U-Net architecture is summarized in Figure 5.

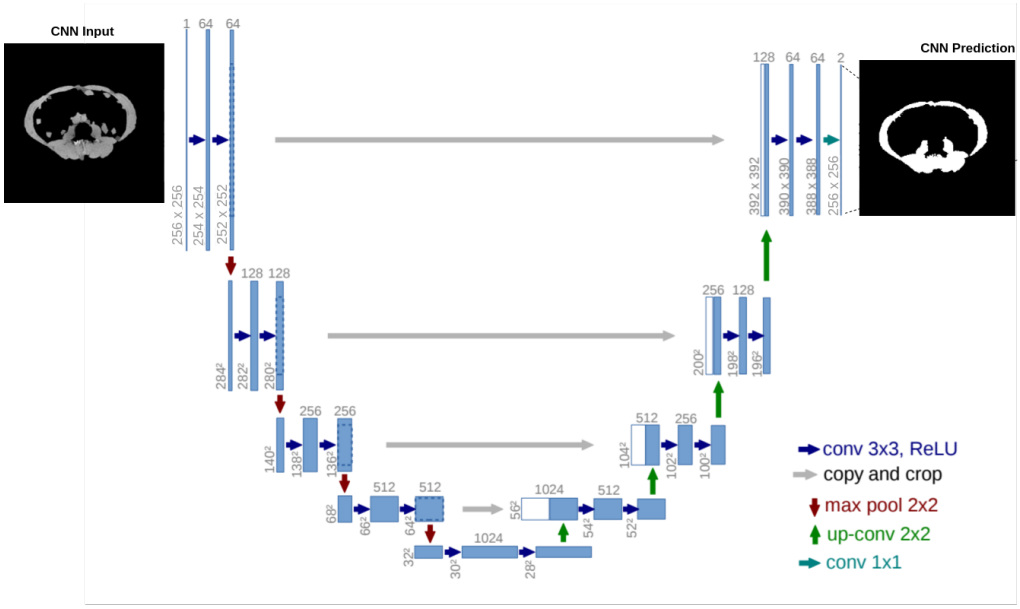


Figure 5: U-Net architecture.

U-Net was implemented in this work using TensorFlow library version 2.2.0. The size of the input images was reduced from 512×512 to 256×256, to reduce time and memory consumption on the training phase. 150 epochs were used with `BATCH_SIZE`$=10$. The *Adam* optimizer was used with a learning rate of 0.0001. The *Binary Cross Entropy* is the metric used for loss [Ruby and Yendapalli 2020], and *Intersection over Union* (IoU) is the segmentation quality measure [Cheng et al. 2021], given by:

$$\text{IoU} = \frac{\text{\# pixels in } A \cap B}{\text{\# pixels in } A \cup B},$$

where A is the reference image and B is the test image. IoU shall also be used as the accuracy measure for the final results obtained by the proposed method.

The 203-image dataset are then split into $80\%/20\%$ training/test sets. Even though U-Net requires a relatively small size training set, the 162 images available are still a very small set for a CNN. Then, a data augmentation technique is used in order to create a 8-fold size training set: a total of 1296 images. Data augmentation consisted of rotation, scaling, reflection, and shearing transforms using the following parameters: height_shift_range = 0.05, width_shift_range = 0.05, rotation_range = 0.1, shear_range = 0.05, zoom_range = 0.05, horizontal_flip = True, vertical_flip = True, fill_mode = 'nearest'. The 41 images in the testing set are used without data augmentation.

As a final post-processing, the pixels corresponding to the radiodensity of fat are removed from the resulting masks of the U-Net CNN.

4. Results and Discussion

In this section, the results obtained by the proposed methodology are described and compared with the segmentation results obtained by the U-Net CNN trained with the refined ground truth. Section 4.1 shows the results obtained by the various ML algorithms in the binary classification of superpixels and the confusion matrix of the algorithm with the best performance, viz., the RF algorithm. Section 4.2 presents the segmentation results of the proposed method, with a typical example, and compares the proposed method results with the ones obtained by the U-Net CNN trained with the refined ground truth.

4.1. Superpixel Classification

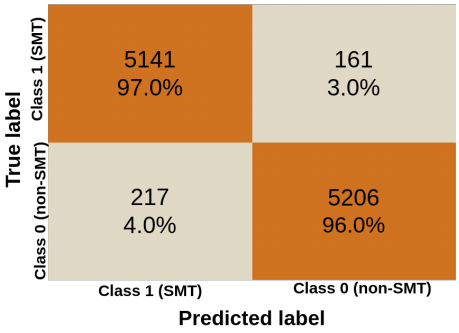
The intermediate step on the proposed methodology is to classify the superpixels created by the SLIC algorithm in two class: SMT and non-SMT. Seven commonly used classifiers were tested in order to select the one with the best performance: Random Forest (RF), Decision Trees (DT), K-Nearest Neighbors (K-NN), Naive Bayes (NB), Support Vector Machine (SVM), Logistic Regression (LR), and MultiLayer Perceptron (MLP). Table 1 shows their results for accuracy, precision, recall, F1-score, and AUC (Area Under the Curve). As seen in Table 1, the RF model achieved the best accuracy among the ML algorithms in the superpixel classification, thus it was chosen to perform this task.

Table 1: ML algorithms performances on superpixels classification.

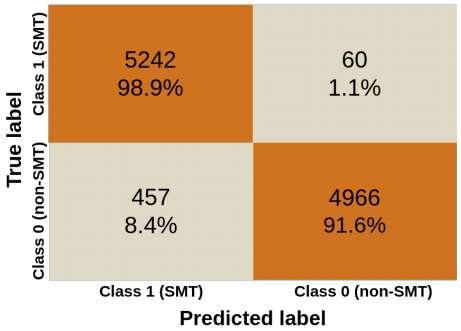
Algorithm	Accuracy	Precision	Recall	F1-score	AUC
RF	**96.47%**	**95.95%**	**96.96%**	**96.45%**	**96.48%**
DT	95.02%	94.76%	95.19%	94.97%	95.02%
K-NN	84.20%	82.42%	86.49%	84.41%	84.23%
NB	82.14%	80.03%	85.11%	82.49%	82.17%
SVM	77.65%	73.83%	84.89%	78.97%	77.74%
LR	75.97%	76.23%	74.66%	75.44%	75.95%
MLP	75.05%	67.25%	96.54%	79.28%	75.29%

Since false negative classification (i.e., classifying as non-SMT a superpixels truly corresponding to SMT) would reduce the number pixels of interest (SMT) to be used as input for the CNN, some experiments were made to adjust the RF threshold parameter. The probability threshold parameter in the RF algorithm is set to 0.50 as default, which means that, if the output is greater or equal to 0.50, the superpixel is classified as class 1 (SMT), otherwise it is classified as class 0 (non-SMT). After the experiments, an empirical threshold value of 0.20 reduced the false negative ratio while keeping a high accuracy rate. Figure 6 shows the confusion matrices using 0.50 and 0.20 as threshold values: the number of false negative results decreased from 161 to 60 (a reduction of 62.73%), while the accuracy decreased from 96.48% to 95.18% (a reduction of only 1.34%).

Figure 7 shows a typical example of results for the superpixel classification obtained by the RF algorithm. (a) is the CT-L3 image; (b) shows the mask formed by the superpixels classified as SMT by the RF algorithm; (c) shows the CT-L3 image masked by the previous binary mask, i.e., the *pixels* of the CT-L3 considered as SMT by the RF; (d) shows the corresponding coarse ground truth (manual segmentation using the graphical interface).

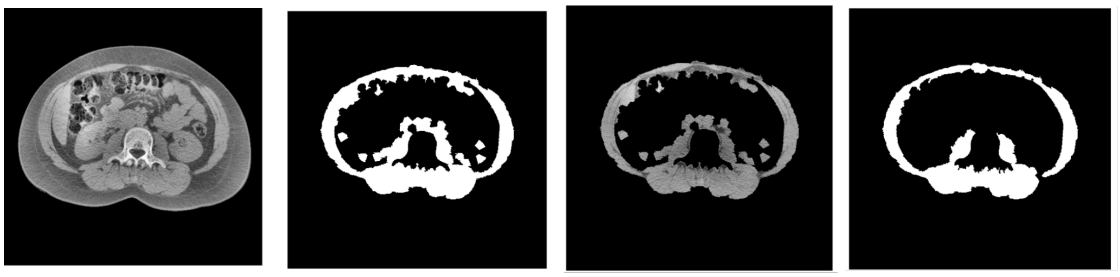


(a)



(b)

Figure 6: Confusion matrices with probability thresholds of 50% (a), and 20% (b).



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

Figure 7: (a) CT-L3 image. (b) Superpixel classification by the RF. (c) CT-L3 image masked by the binary mask on Figure 7b. (d) Coarse ground truth (manual segmentation).

4.2. Pixel Classification by the CNN

After the RF classifier obtained the binary masks composed by the superpixels classified as SMT, these masks are used to select the pixels in the corresponding CT-L3 images (the nonzero pixels in Figure 7c), which will, by their turn, be the inputs of the U-Net CNN. The CNN uses only the coarse segmentation as ground truth. A refined ground truth, manually segmented by a medical expert is used for efficiency analysis purposes only.

First, the U-Net CNN is trained and tested with *refined* ground truth masks. The CNN obtained an IoU of 86.15%. This value is considered the *ceiling accuracy*. Then, the same CNN is trained and tested using the coarse ground truth and obtained an IoU of 83.45%, which represents $\mathbf{96.87\%}$ of the ceiling accuracy.

The post-processing—i.e., removing the pixels classified as SMT but have the radiodensity corresponding to the fat tissue (as explained in Section 3.5)—is then applied to both results: The CNN trained with the refined ground truth obtained an improved IoU of 88.10% (the new ceiling accuracy), and the CNN trained with the coarse ground truth achieved an also improved IoU of 87.40%, which now represents $\mathbf{99.21\%}$ of the ceiling accuracy.

Figure 8 shows two examples of the segmentation results obtained by the CNN trained with the refined (third column) and coarse (second column) ground truth masks. The first column in Figure 8 shows the refined ground truth for visual inspection, and the fourth column shows the difference between the masks in the second and third columns.

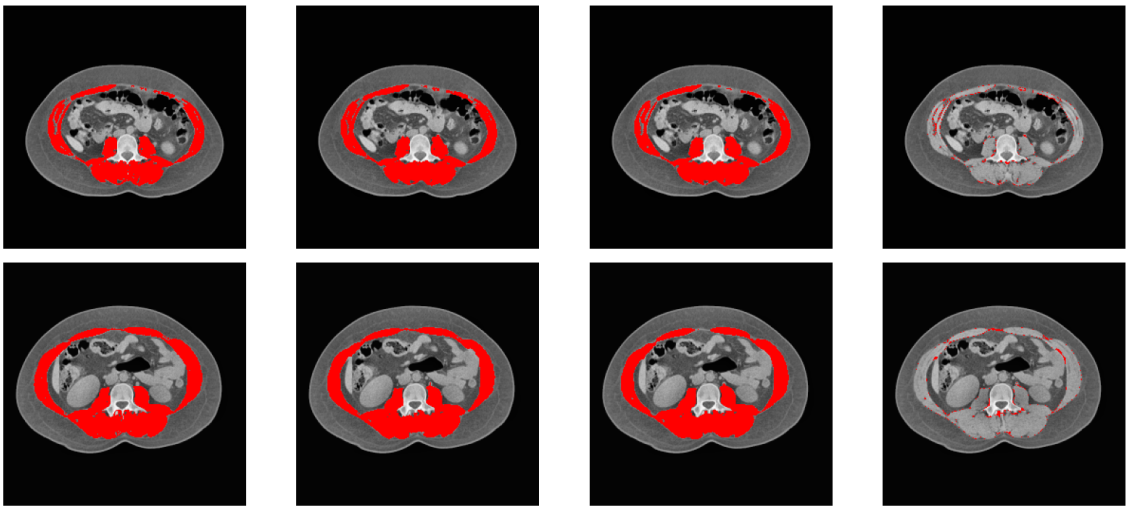


Figure 8: Column 1: ground truth created by the medical expert. Col. 2: results obtained by the proposed method. Col. 3: results obtained by the CNN trained with the refined masks as ground truth. Col. 4: difference between masks in columns 2 and 3.

5. Conclusion

Although ground truth masks manually created by medical experts are paramount to ML models in order to obtain top accuracies in the segmentation task of medical images, the IoU results presented in Section 4 (our methodology achieved $\mathbf{99.21\%}$ of the ceiling accuracy, i.e., the IoU obtained by the same CNN architecture using the refined ground truth) suggest that the proposed methodology may be successfully used in segmentation tasks of medical images lacking their corresponding ground truth masks created by a medical expert. Visual inspection of the examples in Figure 8 also corroborates this suggestion.

The methodology proposed here may be used in other ways: A modest size dataset of medical images with medical expert ground truth masks may be used with this methodology to automatize the segmentation process: superpixel classification (by the RF algorithm) followed by the pixel classification (by the CNN). Or yet, one may have an image set of reasonable size, but with only a modest size of ground truth masks manually segmented by a medical expert. These masks may be used to train a RF model to classify superpixels of a desirable object or tissue. Then, the images without ground truth may be have their superpixels classified by the RF and this coarse ground truth may feed a CNN for pixel classification.

In conclusion, the proposed methodology is promising and may be used in segmentation of medical images when the ground truth masks created by medical experts are not available, achieving results similar to the ones of a CNN trained with such masks.

Acknowledgments

This work was carried out with the support of Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). We also thank Dr. Sara Maria Moreira Lima Verde, who kindly provided the labeled dataset of CT-L3 images.

References

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Süsstrunk, S. (2010). SLIC superpixels. Technical report.

Caan, B. J. et al. (2018a). Association of muscle and adiposity measured by computed tomography with survival in patients with nonmetastatic breast cancer. *JAMA Oncology*, 4(6):798–804.

Caan, B. J. et al. (2018b). The importance of body composition in explaining the overweight paradox in cancer-counterpoint. *Cancer Research*, 78(8):1906–1912.

Cheng, B. et al. (2021). Boundary IoU: Improving object-centric image segmentation evaluation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15334–15342.

Dabiri, S. et al. (2019). Muscle segmentation in axial computed tomography (CT) images at the lumbar (L3) and thoracic (T4) levels for body composition analysis. *Computerized Medical Imaging and Graphics*, 75:47–55.

Ding, Z. et al. (2016). Association between high visceral fat area and postoperative complications in patients with Crohn's disease following primary surgery. *Colorectal Disease*, 18(2):163–172.

Liu, X. et al. (2018). Automatic organ segmentation for CT scans based on super-pixel and convolutional neural networks. *Journal of Digital Imaging*, 31(5):748–760.

Melo, G. F. L. (2022). Segmentação não supervisionada de tecido adiposo subcutâneo em imagens de tomografia computadorizada utilizando morfologia matemática e transformadas de distância. TCC, Universidade Estadual do Ceará, Fortaleza, Brazil.

Paris, M. T., Tandon, P., Heyland, D. K., Furberg, H., Premji, T., Low, G., and Mourtzakis, M. (2020). Automated body composition analysis of clinically acquired computed tomography scans using neural networks. *Clinical Nutrition*, 39(10):3049–3055.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer.

Ross, R. et al. (1992). Quantification of adipose tissue by MRI: relationship with anthropometric variables. *Journal of Applied Physiology*, 72(2):787–795.

Ruby, U. and Yendapalli, V. (2020). Binary cross entropy with deep learning technique for image classification. *Int. J. Adv. Trends Comput. Sci. Eng*, 9(10).

Schweitzer, L. et al. (2015). What is the best reference site for a single MRI slice to assess whole-body skeletal muscle and adipose tissue volumes in healthy adults? *The American Journal of Clinical Nutrition*, 102(1):58–65.

Schweitzer, L. et al. (2016). Estimation of skeletal muscle mass and visceral adipose tissue volume by a single magnetic resonance imaging slice in healthy elderly adults. *The Journal of Nutrition*, 146(10):2143–2148.