# Ultrasound Versus Elastography in the Study of Sarcopenia

Luís Lopes
AndreLuis_lopes2000@hotmail.com
Department of Physics, University of Coimbra, Portugal

Alexandra André
alexandra.andre@estescoimbra.pt
Coimbra Health School, Coimbra, Portugal

Jaime B. Santos
jaime@deec.uc.pt
CEMMPRE, ARISE, Department of Electrical and Computers Engineering, University of Coimbra, Portugal

José Silvestre Silva
jose.silva@academiamilitar.pt
CINAMIL & Portuguese Military Academy, Portugal
LIBPhys-UC & LA-Real, University of Coimbra, Portugal

## Abstract

Using medical images from the rectus femoris muscle acquired through ultrasound and elastography, the present work analyses the classification of sarcopenia using modern deep learning architectures and conventional machine learning models. The dataset consists of 180 medical images collected from 30 people with ages ranging from 20 to 75.

The study explores a variety of models, including deep learning models like DenseNet 121, VGG16, VGG19, ResNet50, and Inception V3, as well as traditional models like logistic regression and neural networks. The performance of the neural network model is in line with deep learning models. The Neural Network achieved the best performance with an F1 score of 99.81%. This study demonstrates improved performance when just using ultrasound images as the dataset and a traditional machine learning model for classification, providing insights into diagnostic tools for early intervention and enhanced care of an aging global population, shedding light on their potential to classify sarcopenia properly.

**Keywords:** Sarcopenia, ultrasound, elastography, traditional classifiers, Deep Learning

## 1 Introduction

Sarcopenia's impact on muscle health, particularly in an aging population, has boost research into detection and intervention strategies. Medical imaging methods like ultrasonography and elastography offer non-invasive insights into muscle characteristics [2] [1].

Models based on deep learning and machine learning techniques are able to identify patterns and specific characteristics of sarcopenia from large amounts of data, which can lead to more accurate classifications. Automating the classification process can be faster and more efficient than the manual process, saving time and resources [3].

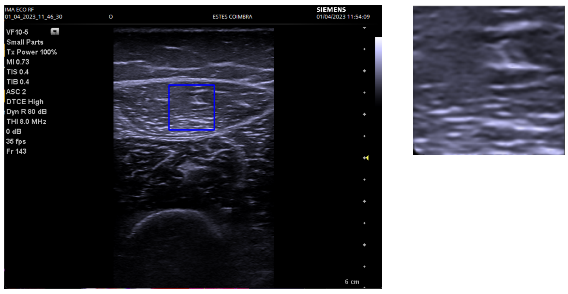
This work explores the potential for classifying sarcopenia using machine learning and deep learning models. The dataset, comprising 180 images from the rectus femoris muscle of 30 participants aged 20 to 75, went thorough preprocessing. Traditional machine learning models and advanced deep learning architectures, including VGG16 and others, are evaluated. By comparing these approaches, the research seeks to illuminate effective pathways for early sarcopenia diagnosis and intervention, ultimately contributing to improved healthcare for an aging population.

## 2 Methodology

The methodology used in this study applies a methodical approach to data preparation, preprocessing, and data augmentation. It also makes use of conventional classification approaches and deep learning Convolutional Neural Networks (CNNs). The following is a summary of the methodology's essential phases.

### 2.1 Data Preparation

The rectus femoris muscle images from the dataset, were acquired by ultrasonography and elastography. Thirty volunteers, ranging in age from 20 to 75, provided the data points needed to create a sample that was representative of each age group. A region of interest (ROI) used as input to the models was extracted from the acquired images, as shown in Figure 1.



(a) (b)

Figure 1: Comparison between the images obtained by the ultrasound scanner and the region of interest extracted from the previous image (not in real sizes).

### 2.2 Preprocessing and Data Augmentation

Recognizing the critical role of data volume in traditional machine learning, it was employed the sliding window technique to increase the dataset size. This approach effectively expanded the machine learning input by segmenting the images, thereby augmenting the number of samples available for analysis.

In contrast, the approach to augmenting the input data for deep learning models was done by introducing variations to the input images: Horizontal flips were applied to introduce mirror images, Gaussian noise was added to simulate real-world variability, and rotations were executed at six distinct angles (-90, -45, -15, 15, 45 and 90 degrees).

### 2.3 Traditional Classification

The conventional classification techniques included feature extraction, normalization, and selection. In order to compare how well different classifiers performed in identifying sarcopenia from the acquired images, the classifiers' performance was analyzed.

Throughout the optimization process, several functionalities and approaches were systematically explored. At several stages of optimization, these functions, which are essential to the classification process, were assessed and compared. They had a crucial role in determining the classifiers' learning behavior and discriminating abilities, which affected their ability to distinguish between images with and without sarcopenia. This thorough analysis allowed the understanding how several factors affected the performance of the classifiers and ultimately helped to correctly identify sarcopenia in the images.

### 2.4 Deep Learning

Transfer learning was used in the pre-trained models like DenseNet 121, VGG16, VGG19, ResNet50, and Inception V3. To increase feature relevance, attention-related mechanisms were included.

This research included a wide range of features and properties essential to model optimization. A thorough investigation and comparison of variables including batch size, loss functions, optimization techniques, and learning rate schedulers was conducted. The model's learning dynamics and convergence were considerably impacted by the selection of these components. This research intended to identify the ideal configuration that would enable the accurate diagnosis of sarcopenia in the images by methodically evaluating and contrasting these elements.

## 3 Results and Discussion

### 3.1 Datasets

We used datasets: A, B, and C. The dataset A has 89 principal Regions of Interest (ROIs) from ultrasound scans; 46 of these ROIs were from elderly adults with sarcopenia and 43 from young, healthy people.

The dataset B was 91 primary ROIs from elastography including 42 from young, healthy individuals and 49 from elderly persons with sarcopenia.

With 137 principal ROIs, Dataset C consists of ultrasonography and elastography data, where 85 of these come from young, healthy people in their 20s, and 95 from older people with sarcopenia.

The sliding window technique became a key technique for dataset analysis in the area of machine learning. The most effective combinations were found by a methodical investigation of various window widths and step sizes. The dataset A performed better with a 38 point window and 3 point step size. While dataset B showed improved outcomes with a 35-point window and a 3-point step size. On the other hand, the combined dataset C showed the best results with a window size of 38 and a step size of 3. The datasets are presented in Table 1 (see ML rows).

Several data augmentation techniques were employed in the effort to fully use the potential of the models. These techniques aimed to increase the input's diversity in order to produce input that was more like real-world scenarios and so increase the model's adaptability. The details about these datasets are in Table 1 (see DL rows).

Table 1: Datasets used for both machine learning (ML) and deep learning (DL) models.

| | Datasets | Number of images | Ratio (unhealthy:healthy) |
|---|---|---|---|
| | A | 12 509 | 45:55 |
| **ML** | B | 8 612 | 42:58 |
| | C | 19 460 | 43:57 |
| | A | 22 784 | 52:48 |
| **DL** | B | 23 296 | 54:46 |
| | C | 46 080 | 53:47 |

### 3.2 Results

In the current work, it is present a comparison of five classic classifiers and five Deep Learning Models. The computing requirements of testing different models has an impact on the choice of classifier in the deep learning context. In order to optimize, a base set of functions was initially created. Then, in each step, different functions were compared, and the one that produced the highest performance was chosen, modifying the basic set of functions. In machine learning, the hyperparameters were first optimized using Bayes search, and then the functions for feature normalization, feature selection, and feature quantity were compared. Following deep learning model tuning, a number of features and properties were assessed. A thorough evaluation and comparison of batch size, loss functions, optimization methods, and learning rate schedulers was conducted. The usage of attention processes was also evaluated. The final performance of each classifier is present in Table 2.

Table 2: Classifier's final performance using datasets A, B and C.

| | F1_score (%) | | |
|---|---|---|---|
| **Model** | **A** | **B** | **C** |
| **Neural Network** | 99.81 | 99.59 | 99.20 |
| **Nu Support Vector** | 94.03 | 93.32 | 92.84 |
| **Logistic Regression** | 96.84 | 96.16 | 92.90 |
| **Stochastic Gradient Descent Classifier** | 97.85 | 96.79 | 95.01 |
| **Support Vector Machine** | 99.04 | 98.34 | 96.81 |
| **DenseNet 121** | 96.94 | 97.63 | 97.63 |
| **VGG 16** | 99.02 | 98.55 | 99.12 |
| **VGG 19** | 99.33 | 99.12 | 98.71 |
| **ResNet 50** | 95.36 | 94.32 | 93.22 |
| **Inception V3** | 93.50 | 89.13 | 88.35 |

A number of reasons may have contributed to machine learning classifiers' superior performance. Notably, given that classical machine learning is more effective with smaller, simpler datasets, the dataset's size and complexity may have been better suited for it.

It's possible that hyperparameter adjustment also had an impact. The machine learning models were subjected to refined hyperparameter optimization, resulting in configurations that take advantage of the features in the dataset. Due to the high computational power required by the deep learning models, the hyperparameter tuning was not as rigorous as the one made for the machine learning classifiers.

While typical machine learning classifiers performed more effectively when it came to categorizing cases of sarcopenia, an interesting finding regarding the effectiveness of deep learning models was made. Some deep learning models, including VGG16 and VGG19, showed high performance without requiring intensive optimization efforts. These models repeatedly demonstrated their promise as reliable classifiers across a variety of datasets by achieving an F1 score of 99%. They did have the capacity to reach comparable results more quickly, highlighting the importance of their inbuilt feature extraction abilities. It is also worth mentioning that the models ResNet 50 and Inception V3 reached a better performance without the attention mechanism. This finding could be related to their architecture that facilitates the propagation of features through different layers, thus not needing an attention mechanism.

Finally, both machine learning and deep learning models had comparable performance across all datasets. Overall, dataset A resulted in the best performances and dataset C in the lowest. In most models, the decrease in performance isn't relevant, but in models like Inception V3 is noticeable. This could be due to difficulties in defining elastography images, leading to suboptimal performance in datasets containing elastography images.

## 4 Conclusions and Future Work

In this research, it was possible to define sarcopenia in both ultrasound and elastography images with an average F1 score above 90% and in some models reaching 99%. These results show the model's ability to enhance the diagnostic of sarcopenia, facilitating more effective healthcare interventions in the context of an aging population.

The short dataset employed is the primary factor accounting for the higher performance of traditional machine learning classifiers. The dataset utilized may have produced models with lower performance than anticipated because the deep learning models require a very large dataset in order to efficiently train them.

To elect the best method for classifying sarcopenia, future studies may incorporate more thorough deep learning model exploration, including various architectures and configurations. Additionally, increasing the focus on hyperparameter optimization and domain-specific modifications has the potential to improve classification.

## 5 Acknowledgements

This research is sponsored by national funds through FCT – Fundação para a Ciência e a Tecnologia, under the project UIDB/00285/2020 and LA/P/0112/2020

## References

[1] Giuseppe Sergi *et al*. Imaging of sarcopenia. *European Journal of Radiology*, 85:1519–1524, 2016. ISSN 0720-048X. doi: 10.1016/J.EJRAD.2016.04.009.

[2] Roger A. Fielding *et al*. Sarcopenia: an undiagnosed condition in older adults. current consensus definition: prevalence, etiology, and consequences. international working group on sarcopenia. *Journal of the American Medical Directors Association*, 12:249–256, 2011. ISSN 1538-9375. doi: 10.1016/J.JAMDA.2011.01.003. URL `https://pubmed.ncbi.nlm.nih.gov/21527165/`.

[3] Shengfeng Liu *et al*. Deep learning in medical ultrasound analysis: A review. *Engineering*, 5:261–275, 4 2019. ISSN 2095-8099. doi: 10.1016/J.ENG.2018.11.020.