Sarcopenia Diagnosis: Deep Transfer Learning versus Traditional Machine Learning

Carlos Sobral
carlosfssobral@gmail.com

Jose Silvestre Silva
jsilva@ci.uc.pt

Alexandra Andre
alexandra.andre@estescoimbra.pt

Jaime B. Santos
jaime@deec.uc.pt

University of Coimbra, Department of Electrical and Computer Engineering, Portugal

Portuguese Military Academy and Military Academy Research Center (CINAMIL), Lisbon, Portugal
Laboratory for Instrumentation, Biomedical Engineering and Radiation Physics (LIBPhys-UC), Coimbra, Portugal

Coimbra Health School, Coimbra, Portugal

University of Coimbra, CEMMPRE, Department of Electrical and Computer Engineering, Portugal

Abstract

Sarcopenia is a syndrome characterized by progressive and generalized loss of skeletal muscle mass and muscle strength. Through the analysis of ultrasound images, this work compares the effectiveness between traditional deep transfer learning and three traditional classifiers, FineKNN, CubicSVM and SubspaceKNN. The results showed that the deep learning transfer had the best final classification, 98.3%; the traditional classifier that presented the better performance was CubicSVM, with an efficiency of 97.9%, higher than the others, with SubspaceKNN achieving 97.1% and FineKNN reaching 96.5%.

Keywords: Sarcopenia, ultrasound, traditional classifiers, Deep Transfer Learning, Inception-v3, FineKNN, CubicSVM, KNN Subspace.

1 Introduction

Deep learning is a method that has the ability to "learn" automatically the mid and high-level features from untrained images. Since the creation of AlexNet, the winner of the "ImageNet" Large Scale Recognition Challenge" (ILSVRC) in 2012 [1], deep learning attracts attentions in the field of machine learning [2]. In 2013, deep learning was selected for the top 10 of the most innovative technologies [3]. Nowadays, this technology is successfully applied in several areas, such as medical imaging [4], using different anatomical structures, such as muscles [5].

Musculoskeletal tissue has as main task the contraction and energy production to carry out movement. It is composed by elongated cells, and at the level of its histology has a striated appearance. These cells have transverse striations, alternated by a light band and a dark band, containing inside the main components for the process of contraction and muscle relaxation, called sarcomeres. As time goes by, sarcomeres are degraded leading to sarcopenia.

In 2010, the European Working Group on Sarcopenia in Older People (EWGSOP) defined sarcopenia as "a syndrome characterized by progressive and generalized loss of skeletal muscle mass and muscle strength, with risk of adverse effects such as physical disability, poor quality of life and death" [6].

2 Methodology

This section aims not only to make a brief explanation of each step performed throughout the project, but also to indicate the workflow of it. The flow and the steps performed are illustrated in the following figure.

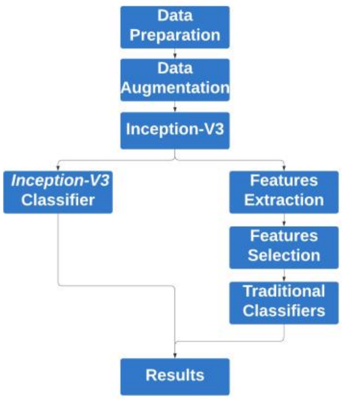

Figure 1: WorkFlow.

2.1 Data Preparation

From acquired images, several 40x40 ROI were extracted, since ROIs of larger size captured undesirable elements such as aponevroses. Once a rotated rectangular ROI presented the corners in black, which could negatively influence the final results, it was used circular ROIs to perform the rotations in the data augmentation process. In order to use this ROIs in the neural network, a resize was performed, changing its dimension to 299x299x3, since these are the input dimensions required by inception-v3 network. The ROI was replicated three times to fill the three layers of the RGB image.

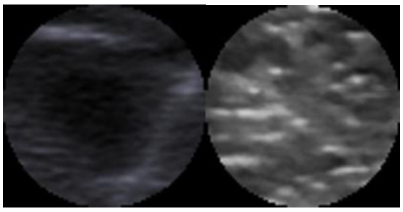

Figure 2: (a) Normal muscle ROI (b) Sarcopenic muscle ROI.

2.2 Data Augmentation

This is a mandatory process to be carried out when talking about Deep Transfer Learning and Machine Learning, since both need large amounts of data to be able to produce good accuracy. For the tests performed, three mathematical operations were used.

a) Rotations: 3 different angles were defined, being 10°, 20° and 30°. For this purpose, the matlab command "imrotate" was used. Having in mind these values, each image generated six new images.

b) Noise: In this operation was used the Gaussian noise formula in matlab:

$$x = b + \sqrt{c} * rand(size(b)) \quad (1)$$

In this expression b is the image (ROI) and c is a chosen value from one of three distinct values that were randomly selected. In this case 6,12 and 25 were used. Then, three new images were created.

c) Rotations + Noise: 18 new images were created, all obtained from the merging of rotation operations (clockwise and non-clockwise) and the addition of noise.

d) Flip: once the percentages of the training, validation and testing data were chosen, the number of images was doubled in each one steps with the "flip" operation, mirroring the images. For this, the matlab command "fliplr"was used.

2.3 Inception-V3

The neuronal network Inception-V3 was used for the analysis of the images. It has an architecture based on 2 types of factorization.

2.3.1 Factorization into small convolutions

The factorization of convolutions aims to reduce the number of connections/parameters without affecting the accuracy of the network. This allows the reduction of the size from convolution layers, namely a 5x5 convolution, which is replaced by two 3x3 convolutions. That leads to parameters reduction [7].

2.3.2 Factorization into Asymmetric Convolutions

This factorization uses the layers resulting from the previous process. These layers have their size reduced, namely a 3x3 layer, which is converted into two 1x3 and 3x1 layers. Using a 3×1 convolution followed by a 1×3 convolution is the equivalent of sliding a two-layer network with the same filter size as in a 3×3 convolution.

In order to reduce the computational costs of the network, it was found that the transformation into 1x7 and 7x1 layers provided better results [7].

2.3.3 Convolutional Layers

A convolutional layer contains a set of filters whose parameters need to be learned. Each filter is concatenated with the input volume to calculate an activation map, i.e., the filter goes through the entire width and height of the input. Finally, the product of the points between the input and the filter are calculated at each spatial position.

2.3.4 Pool Layers

The Pool layer is added after the convolutional layer. Specifically, after a nonlinearity (for example, ReLU) has been applied to the output of features maps by a convolutional layer. There are 2 types of pool layers:

a) Average pool: Calculates the average value of each patch on the features map.

b) Max pool: Calculates the maximum value for each patch of the features map.

The result of using this type of layers and creating sampled feature maps is a summarized version of the features found in the input.

2.5 Features Extraction

Loading dataset into the inception-V3 network, the "avg_pool" layer of this network was used to obtain the 2048 features of each image. Using the "activations" command present in Matlab it was possible to create an array with all the extracted features. Then this array was introduced into the feature extractor toolbox *FEAST* in order to select the best 400 features. This toolbox has several algorithms to perform the detection of the best features, to be used in the training of traditional classifiers.

3 Results

Datasets were obtained from 144 images (ROIs of the original images, 61 normal muscles and 83 sarcopenic muscles). The percentages assigned to the training, validation and testing steps were the following:

a) dataset A: 60% for training, 20% for validation and 20% for testing, which corresponds to 7200 images (4300 training, 1450 validation and 1450 test).

b) dataset B: 70% for training, 15% for validation and 15% for testing, which corresponds to 7200 images (5000 training, 1100 validation and 1100 test).

c) Dataset C: 80% for training, 10% for validation and 10% for testing, which corresponds to 7200 images (5750 training, 750 validation and 700 test).

Finally, three new datasets were obtained from 285 images (2 ROIs per original image), but 3 of the images only allowed to select 1 ROI since they have small muscles . The percentages assigned to the training, validation and testing steps were the following:

a) A2R dataset: 60% for training, 20% for validation and 20% for testing, which corresponds to 14350 images (8550 training, 2900 validation and 2900 test).

b) B2R dataset: 70% for training, 15% for validation and 15% for testing, which corresponds to 14350 images (9950 training, 2150 validation and 2150 test).

c) C2R dataset: 80% for training, 10% for validation and 10% for testing, which corresponds to 14350 images (11400 training, 1450 validation and 1450 test).

We used three different proportions of training, validation and test in order to verify the effect of the training percentages on the final results.

Tests were carried out on several traditional classifiers presented in the "classification learner" application of Matlab to identify which classifier would produce the best performance.

As input, these classifiers received the matrix with the features extracted from all the training images in the dataset and a matrix with the respective labels.

After training the classifiers, the results achieved by the best three classifiers were exported to be used later, in order to evaluate the test images of the dataset in question.

For the analysis with the inception-V3 network, the following values were chosen for the different hyper-parameters: Epochs = 5; LearningRateFactor=0.001;MiniBatchSize=100; WeightLearnRateFactor = 10; BiasLearnRateFactor = 5.

Method	Dataset A	Dataset B	Dataset C	Dataset A2R	Dataset B2R	Dataset C2R
CubicSVM	95.0%	93.3%	90.7%	97.1%	97.9%	97.9%
Fine KNN	91.9%	91.5%	89.6%	95.4%	96.5%	96.5%
SubspaceKNN	92.9%	92.1%	90.3%	95.9%	97.1%	96.4%
Inception-V3	98.3%	93.3%	90.9%	97.9%	98.0%	97.9%

Table 1: Accuracy values achieved by the four implemented methods.

Looking closely to the obtained results, it was concluded that traditional classifiers performance is lower than deep transfer learning since in datasets with only one ROI per image, the maximum classification value achieved by traditional classifiers was 95%, i.e., a difference of 3.2% compared to the best result obtained by inception-V3. With regard to datasets with double ROI per image, the classifications between both methods are similar. However, only Inception-V3 achieved an efficiency of 98%. These results suggests that deep transfer learning has a superior performance when compared to traditional classifiers.

3 Conclusions and Future Work

The datasets of this project were obtained through 144 original images and since both deep transfer learning and machine learning needs a large amounts of data, the acquisition of more images related to patients with and without the pathology can also allow an increase on accuracy. One of the aspects that can be studied in future work is the size defined for the ROIs, since having these with smaller sizes can also produce better results.

Acknowledgments

This research is sponsored by FEDER funds through the program COMPETE – Programa Operacional Factores de Competitividade – and by national funds through FCT – Fundação para a Ciência e a Tecnologia –, under the project UIDB/00285/2020.

References

[1] Teofilo F. Gonzalez. Handbook of approximation algorithms and metaheuristics. Handb. Approx. Algorithms Metaheuristics, pages 1–1432, 2007.

[2] Shengfeng Liu, Yi Wang, Xin Yang, Baiying Lei, Li Liu, Shawn Xiang Li, Dong Ni, and Tianfu Wang. Deep Learning in Medical Ultrasound Analysis: A Review. Engineering, 5(2):261–275, 2019.

[3] Ge Wang. A perspective on deep imaging. IEEE Access, 4:8914–8924, 2016.

[4] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Fran-cesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. van der Laak, Bram van Ginneken, andClara I. Sánchez. A survey on deep learning in medical image analysis. Med. Image Anal., 42:60–88, 2017.

[5] P. Burlina, N. Joshi, S. Billings, I. J. Wang, and J. Albayda. Unsupervised deep novelty detection: Application to muscle ultrasound and myositis screening. Proc. - Int. Symp. Biomed. Imaging, 1910–1914, 2019.

[6] Alfonso J, et al. Sarcopenia: Europeanconsensus on definition and diagnosis. Age Ageing, 39(4):412–423, 2010.

[7] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethin-king the Inception Architecture for Computer Vision. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2818–2826, 2016.