Automatic evaluation of the ejection fraction on echocardiography images

Alexandre Aubry[1], Luc Duong[1]

[1] Department of software and IT engineering, École de technologie supérieure, Montreal, Canada

Abstract— **The ejection fraction measures the amount of blood pumped by the heart. Echocardiography is the most common imaging modality to assess the ejection fraction but can be challenging to interpret. This study proposes an automatic method to evaluate the ejection fraction from echocardiography images. First, several supervised learning model with different convolutional neural networks and a regression model were proposed for the segmentation of the ventricle. This last regression model was built with different layers of neurons. We also build a U-Net model to segment the ventricle. Subsequently, a linear regression model was designed to provide an estimation of the ejection fraction from the difference of area between the ventricle at the systole and the diastole. The model was evaluated on a subset of the EchoNet-Dynamic dataset. The four models described in this article study the ejection fraction of the hearts with another approach than the previous EchoNet-Dynamic studies. The results of this study could be used for the development of prediction models for cardiac motion and eventually to assist in the assessment of cardiovascular diseases.**

Keywords— **Echocardiography, Ejection fraction, Deep neural networks, Transfer learning, segmentation**

I. Introduction

Echocardiography is a technique that uses ultrasound to visualize in real time the different parts and chambers of the beating heart. Echocardiography is safe since it does not involve ionizing radiation, with very fast and simple acquisition and thus, widely available in the medical field. For instance, this allows the detection of cardiac disorders, such as cardiomyopathy.

The last few years have seen the development of computer vision techniques applied to medical imagery [1].

Many different studies have been conducted on cardiac medical imaging. Computer vision models based on convolutional neural networks (CNN) achieve good results on view identification, segmentation, heart chamber volume, ventricle mass and even diagnosis of cardiomyopathy and hypertension [2]. Most computer vision models rely on the training using the ImageNet dataset, which contains natural images, and are not specifically tailored for medical images. RadImageNet has been recently proposed and focus only on greyscale medical images [10]. Another model, the EchoNet-Dynamic was developed for echocardiography images. In addition to predicting the different characteristics of the heart under study, EchoNet-Dynamic can predict the phenotype of the patient such as age, sex, weight. These are characteristics that are difficult for a human to predict with an echocardiography [3]. Three methods for estimating the ejection fraction from echocardiography videos are described in this work. These three models use ResNet-18 3D convolution networks. Their best model has a mean absolute error of 5.44%, knowing that a specialist makes an error of 4-5% for the ejection fraction [4]. In 2017, a study works on ejection fraction with computer vision applied to echocardiograms. First, a neural network locates the ventricle, this allows finding the diastole and systole images. A segmentation network segments the ventricle to then calculate its volume [5]. In 2021, a study estimated the ejection fractions of several datasets including EchoNet-Dynamic using a U-Net based model (DPS-Net). This study investigates healthy hearts and hearts with dysfunctions. Accurate results were obtained for ventricle segmentation and ejection fraction calculation [6]. In 2020, a study used view classification, cardiac cycle synchronization, segmentation and landmark extraction, to calculate left ventricular volume and ejection fraction [7]

The goal of our study is to investigate transfer learning on different deep neural networks and to estimate the ejection fraction. In order to segment the ventricle at the end of diastole and systole, four models are considered:

- Transfer learning with VGG16 (features extraction)
- Transfer learning with VGG16 (fine tuning)
- U-Net
- DenseNet121 trained on the RadimageNet dataset

Then, with the coordinates of the boundary of a ventricle at his minimum and maximum volume, a regression model is introduced to estimate the ejection fraction of the heart.

II. Material and methods

A. EchoNet-Dynamic

EchoNet-Dynamic consists of 10036 cardiac ultrasound videos. Each of the videos can contain from 50 to several hundred images. These 10036 videos are from 10036 different individuals who performed an echocardiogram between

2006 and 2018 at Stanford University Hospital. These videos were made to study the movement of the heart and the variations of volume of the chambers constituting it. For EchoNet-Dynamic, we have an apical-4-chambers view, it means we see four chambers of the heart as illustrated in Figure 1. The videos present a view of the left ventricle of the patient. Each sequence depicts the ventricle in motion varying with the heartbeats. For each video of the dataset, we have access to different characteristics: [4]

- The ejection fraction
- Volumes of the left ventricle at the end of diastole and systole
- Image's dimension (112x112 pixels for each)
- Number of frames per seconds
- The coordinates of the boundary of the ventricles at the end of diastole and systole (see Image 1)

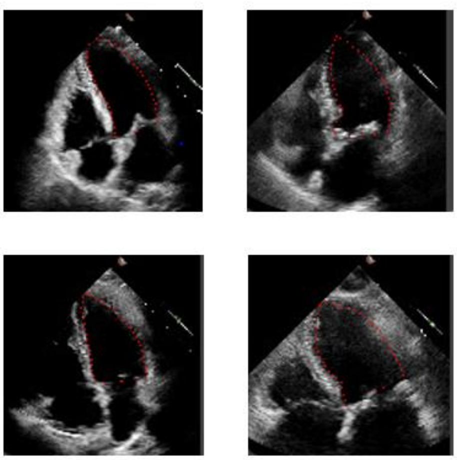

Figure 1 : Representation of the boundary (ground truth) of the left ventricle in red

B. Description of the segmentation architecture

Two models proposed are based on the convolutional neural network VGG16. VGG16 has 16 layers trained on ImageNet. To build our model, transfer learning was applied and image features are extracted [9].

Also, we evaluated DenseNet121 trained on the RadImageNet dataset. This dataset contains 1.35 million of grayscale medical images and is better suited for medical image segmentation than ImageNet. With transfer learning, we can apply this model with the weights trained on RadimageNet, on echocardiography images and extract important features [10].

III. FEATURES EXTRACTION FROM IMAGES

The purpose of this section is to present the models we set up to extract features from the end-diastolic and end-systolic images.

A. Transfer learning with VGG16

With Python, we imported the convolutional neural network VGG16 trained on ImageNet. We applied it to the images of the ventricle at the end of diastole and systole that we were isolated from the EchoNet-Dynamic dataset. As a result, for each image the network extracts features and put it in a vector. So, for each ventricle images, we have a corresponding representative vector. [9]

B. Fine tuning with VGG16

We used another segmentation model with VGG16. We also used transfer learning, but with fine-tuning. It means that in order to extract the important features, we begin to adjust the weights with the images of our dataset. Like the precedent model, the head of the network is a regression model that we build. We perform a first training of the network with a small learning rate and by freezing the weights of VGG16. Indeed, our model is built with a part already trained (VGG16) and a regression model with random weights. With this first training, we taught a little bit the regression model with our data, without modifying the weights of VGG16. Then, the network is ready for the second training where all the weights can be modified. With this type of transfer learning, the model is more adapted to our data. Then, like the previous model, we obtained a feature vector for each image. [9]

C. Transfer learning with DenseNet 121 trained on RadimageNet

For this segmentation model, we imported the convolutional neural network DenseNet121 trained on the RadimageNet dataset. The input layer of this network has a dimension of 224x224x3, and the images of our dataset have a dimension of 112x112x3. We had to resize our images and their corresponding masks. Then, we extracted the features of the ventricle with the same method as VGG16. Indeed, with the CNN we obtained a feature vector. [10]

IV. Segmentation of the ventricle

The purpose of this part is to determine the segmentation coordinates of the ventricles from the extracted features.

A. Segmentation of the ventricle using a regression model

We need a regression model to find the segmentation coordinates of the ventricle in an image after feature extraction. A new sequential model was built in Keras for this regression model involved in the segmentation and the calculation of the ejection fraction. This model is composed of four layers of 128 neurons and an activation with the ReLU function, and a final layer with 42 neurons and an activation with a linear function. The weights of the model are adjusted with the Adam optimizer, and the model is evaluated by mean square error.

For the ventricle at the end of diastole and end of systole, we separately created two regression models with this architecture for each model presented previously in Section III. Each model was trained using the data from EchoNet-Dynamic. The input of the two models is a feature vector of a ventricle image. The output of the first model is the list of the abscissa of the coordinates that segment the ventricle, and the output of the second one is the list of the ordinate. We used this method for the end of diastole and end of systole. Then, we put the abscissa and ordinate together to have the full list of coordinates.

To conclude, once the model is trained and validate, we took test images, put them in one of our models that extracts features. Then, we put the feature vectors in two regression models to estimate the boundary of the ventricles from the images, for the end of diastole and systole. Figure 2 illustrates the segmentation (in red) of a ventricle at the end of diastole.

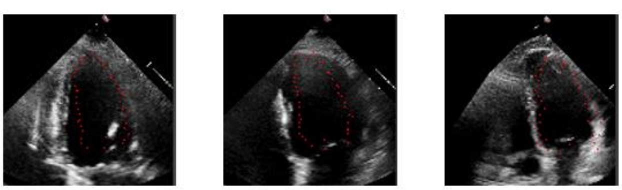

Figure 2 : Representation of the segmentation using VGG16 (red points)

B. Segmentation with U-Net

We also used U-Net[8] to segment the ventricle. We trained U-Net with the data of EchoNet-Dynamic. The input of U-Net is a ventricle images from our dataset, the output is a binary segmentation mask, that indicate for each pixel if it is contains by the ventricle or not (see Figure 3). In Figure 3, the part in blue is the segmentation of a ventricle made by the U-Net model that we trained, and the red points represent the ground truth made by the specialists in EchoNet-Dynamic. [8]

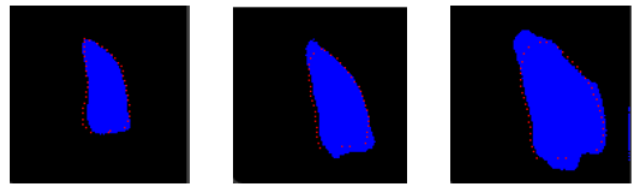

Figure 3 : Representation of the segmentation of a ventricle with U-Net (blue) and its ground truth (red).

V. Estimation of ejection fraction using the segmentation

Once we have models that segment a ventricle with its images at the end of diastole and systole, we can build other models that determine the ejection fraction. For each of our models, we plot learning curve in order to avoid overfitting and underfitting.

First, with the list of coordinates of the boundary that we determine, we can easily find the number of pixels inside a ventricle. For a video in EchoNet-Dynamic, we have the segmentation of the ventricle at the end of diastole, and the segmentation at the end of systole. So, for these two different times, we know the number of pixels inside the ventricle, and we calculate the difference. Thus, for a video of the dataset, we know the difference of pixels in the ventricle between the end of diastole and end of diastole. With this, we trained a another linear regression model, which takes as input the difference of pixels between the end of diastole and end of systole. The output is the value of the ejection fraction. To train the model, we used the theoretical ejection fraction values (ground truth) that are indicated in the dataset.

Then we can test the complete model from the beginning on test images (not used during training). To summarize, these images of tests go through one of the segmentation models, then we determined the difference of pixels between the end of diastole and systole, and with this value, a regression model determined the ejection fraction.

VI. Results and conclusion

In this part, we will present the results and compare them to the results obtained by echocardiograms specialists [4].

In the figure 1-4, we can see the evolution of MSE for ejection fraction, with the training sample size. It corresponds to the four segmentation models that we built. We can see on these curves, that at the beginning, the precision improves and then is stabilized with the increase of the training sample size.

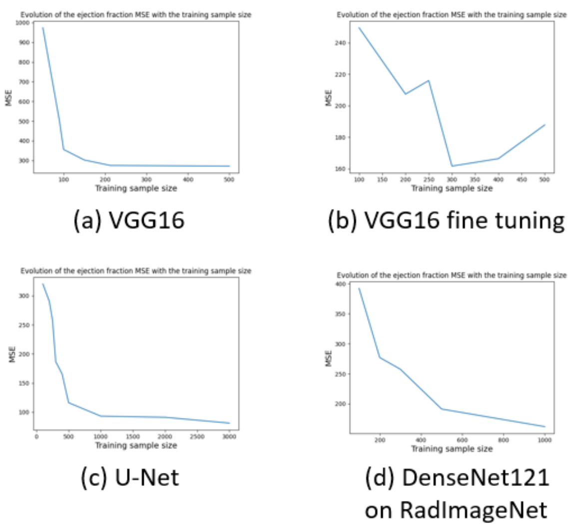

Figure 4 : Evolution of the ejection fraction MSE with the training sample size for our four models

The best result is obtained when the segmentation model is U-Net, we reach a mean absolute error of 9,27%. It's also in this model where we have the best segmentations. We obtain better results when the segmentation model is using VGG16 fine tuning, than the classical transfer learning with VGG16, which is logical. And we obtain slightly better results with DenseNet121 trained on RadimageNet than VGG16 with fine tuning as segmentation model.

We compare our best model (with U-Net) with the result of the Stanford university and the values obtain by specialists. our model reaches a mean absolute error of 9,27%, while an echocardiogram specialist estimate ejection fraction with a mean absolute error of 5%. [4] Our best result was reach by training the model on 3000 images and testing it on 521 images.

With our different models, we outline the importance of a accurate segmentation to have a good ejection fraction estimation. It was also the most challenging part, because in a lot of images the boundary is not very clear.

In conclusion, we proposed a two stage method for the evaluation of the ejection fraction of the heart using EchoNet-Dynamic. We built three different models with transfer learning and one model that use an encoder and a decoder (U-Net). These models were link to regression models that we built and studied with learning curves. With these models, we obtained some precise results, not far human performances. Also, a pretrained network specifically tailored for medical images (RadImageNet) was evaluated for transfer learning. Such approach might be very useful in future works.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Andre Esteva, Katherine Chou, Serena Yeung, Nikhil Naik, Ali Madani, Ali Mottaghi, Yun Liu, Eric Topol, Jeff Dean and Richard Socher, "Deep learning-enabled medical computer vision," npj Digital Medicine, 2021.
2. Jeffrey Zhang, et al., "Fully automated echocardiogram interpretation in clinical practice", Circulation, 2018.
3. Amirata Ghorbani, et al., "Deep learning interpretation of echocardiograms", npj Digital Medicine, 2020.
4. David Ouyang, Bryan He, Amirata Ghorbani, Matt P. Lungren, "EchoNet-Dynamic: A large new cardiac motion video data resource for medical machine learning", NeurIPS, 2019
5. AbdulWahab Kabani and Mahmoud R. El-Sakka, "Ejection fraction estimation using a wide convolutional neural network" ICIAR, 2017
6. Xin Liu, Yiting Fan, Shuang Li, "Deep learning-based automated left ventricular ejection fraction assessment using 2-D echocardiography", The American Journal of Physiology-Heart and Circulatory Physiology, 2021
7. Erik Smistad , Andreas Østvik, "Real-time automatic ejection fraction and foreshortening detection using deep learning", IEEE transactions on ultrasonics, ferroelectrics, and frequency control, 2020.
8. Olaf Ronneberger, Philipp Fischer, and Thomas Brox "U-Net: convolutional networks for biomedical image segmentation", MICCAI, 2015.
9. Karen Simonyan and Andrew Zisserman "Very deep convolutional networks for large-scale image recognition", ICLR, 2015.
10. Xueyan Mei, Zelong Liu, Philip M. Robson, Brett Marinelli, "RadImageNet: an open radiologic deep learning research dataset for effective transfer learning", Radiology: Artificial intelligence, 2022.