Cédric VINCENT

IMA205 Challenge Report

Cardiac Pathology Prediction

I – Data

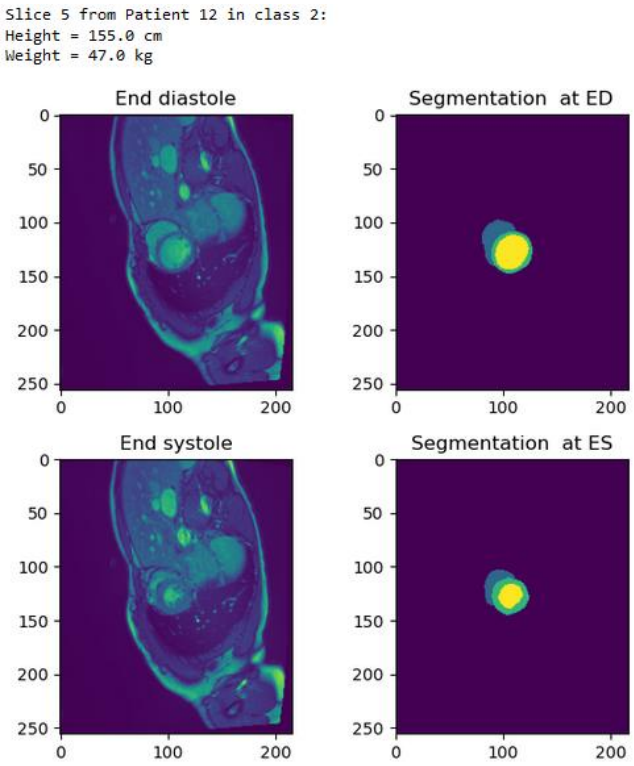
The available data is composed of 150 patient MRI scans, 100 for the training set and 50 for the test set. Each scan contains two 3D images, at end diastole (ED) and at end systole (ES), represented as multiple 2D slices. Each slice has a corresponding segmentation for the Left and Right ventricle cavities (LV and RV) and the Myocardium. In the test set, the Left ventricle cavity segmentation is missing. Patients' height and weight are also available.

The goal of the challenge is to successfully classify the test set into 5 possible diagnostics: '0' - Healthy control, '1' - Myocardial infarction, '2' - Dilated cardiomyopathy, '3' - Hypertrophic cardiomyopathy, '4' - Abnormal right ventricle. We have the training set true classes in order to train our model.

We store the training and test sets in python dictionaries containing for each patient the two MRI images, their segmentation, the patient's height and weight and the dimensions of the pixels of the scans, which vary for each patient because the modalities of acquisition are not the same. The pixels dimension on the slices plane (x, y) are the same in both directions, and orthogonally to this plane (z) the dimension, which corresponds to the distance between slices, is either 5 or 10mm.

The MRI scans are stored as NIfTI files, which we open using the nibabel python library. We access each 3D image by taking the data of the NIfTI file, and we access the pixel dimensions in the header of the file. Heigth and weight are stored in the metadata csv file, along with the target classes.

Figure 1: Visualization of a slice at ED and ES



II – Features extraction

[1] describes the process of diagnosis for each class which is based on several metrics. For example, patient diagnosed with abnormal right ventricle (class 4) have a "RV volume greater than 110 mL/m2 for men, and greater than 100 mL/m2 for women, or/and a RV ejection fraction below 40%".

Because the images are complex and contain much more information than what is useful, we will try to extract specific measures on the important organs in a manner similar to classical diagnosis. The segmentations of each organ are essential to this.

Some of the metrics used can be directly extracted, such as volumes and ejection fractions. Others such as visual analysis of contractions cannot be quantified. We can take inspiration from the literature [2] and add ratios between volumes. We also consider patients' height and weight. The 14 selected features are:

- LV (left ventricle cavity), RV (right ventricle cavity), and Myocardium **volumes** at ED (end diastole) and ES (end systole)
- LV and RD **ejection fractions:** $E_f = \frac{V_{ED} - V_{ES}}{V_{ED}}$
- **Ratios** between RV and LV volumes at ED and ES, and between Myocardium and LV volumes at ED and ES
- Patient **Height** and **Weight**

Volumes can be computed from the segmentations: we count the number of pixels in the relevant segmentation for each slice and then multiply by the scale (pixels dimensions) in all three dimensions to obtain the result in mm^3. Finally, we can normalize them from the patient size using the Body Surface Area (BSA), as mentioned in [1]. It Is defined with the Du Bois formula:

$$BSA = 0.007184 \times W^{0.425} \times H^{0.725}$$

This way, volumes and features depending on volumes are independent from the patient size. The other features are deduced from the volumes.

III – Features analysis

We begin by visualizing the distributions of the features. They have vastly different scales and distributions. Therefore, algorithms such as SVMs will necessitate them to be standardized.

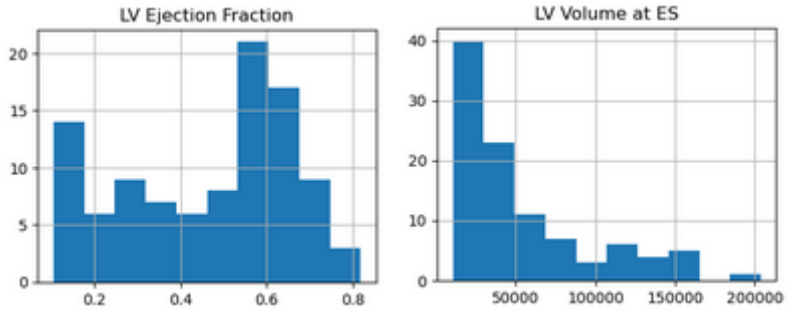


Figure 2: Two features with completly different scales and distributions

We then visualize the correlations between features with a correlation matrix:

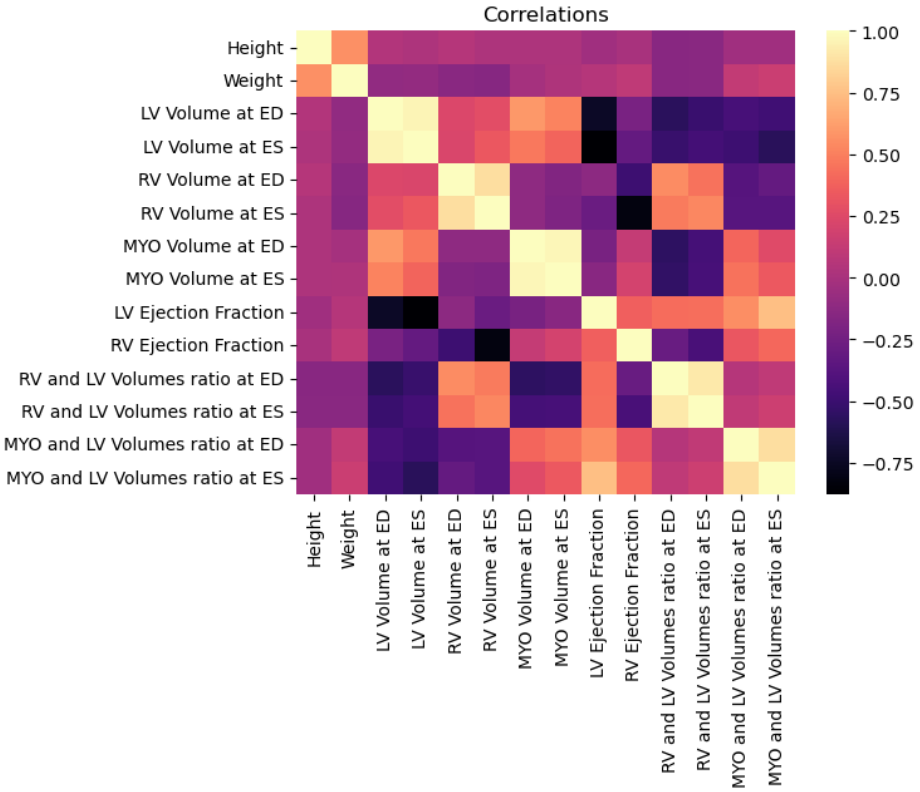
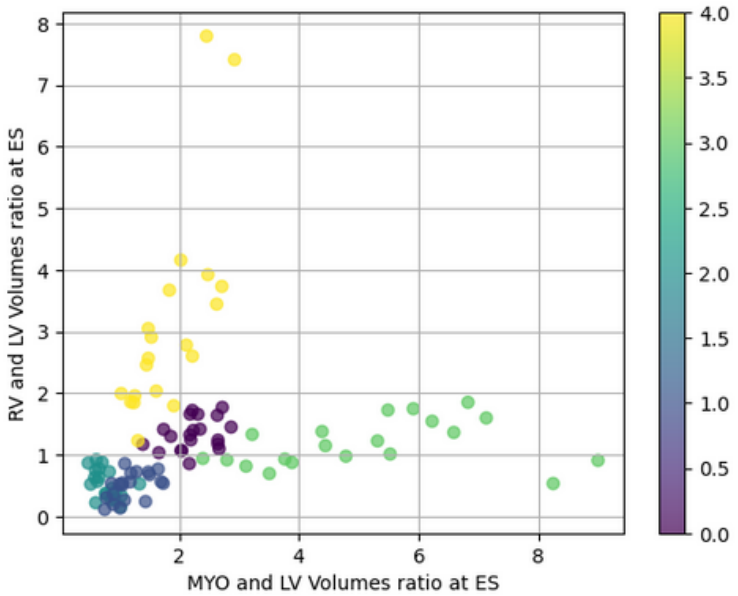


Figure 3: Correlation matrix of all features

As it could be expected, features at ED and ES are correlated. There is also an opposite correlation between volumes and ejection fractions.

By plotting features against each other (in an interactive plot), we can already see some classes which seem to be easily separable from the others: for example, '3' - Hypertrophic cardiomyopathy and '4' - Abnormal right ventricle in MYO and LV volumes ratio at ES and RV and LV volumes ratio at ES.



Decision trees are classifiers that can be interpreted: they enable us to view the influence of each feature. We will use them to view the importance of features relying on the LV segmentation and detect any feature without importance to see if we can reduce the dimension.

We obtain decent classification performances (0.82) with a single tree and visualize it using the graphviz library:

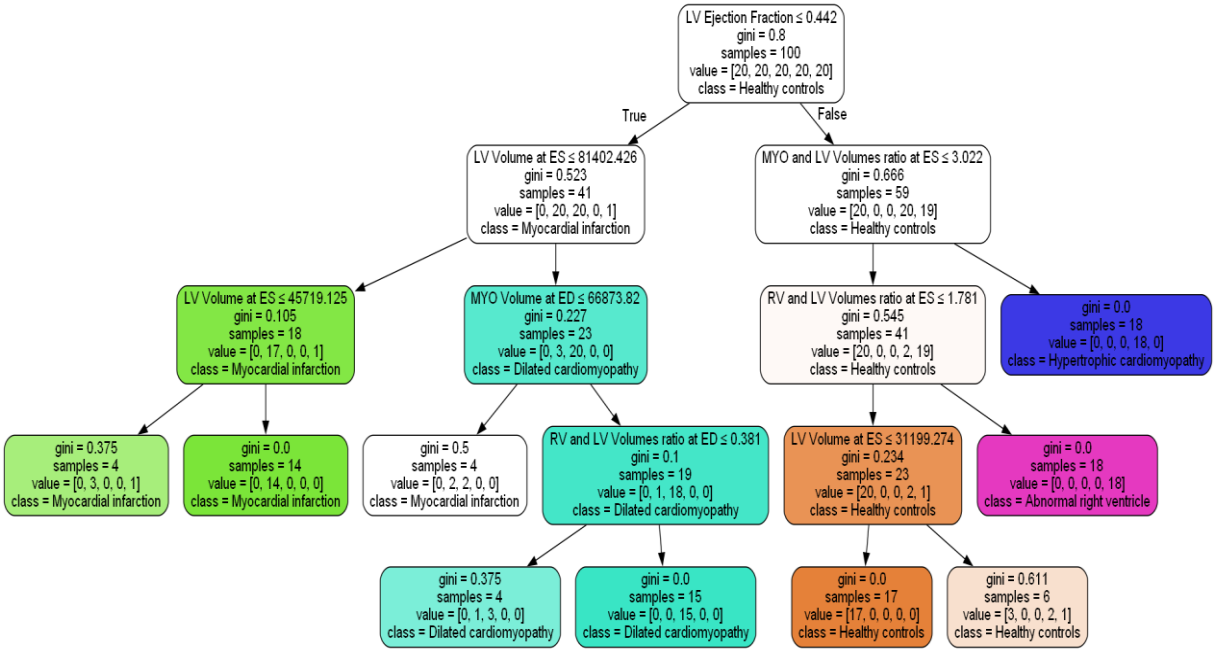


Figure 4: Resulting decision tree. All slpits but one are based on an LV-related feature

We can see that LV related features are the source of the most decisive splits between classes. We can confirm that using a random forest and visualize features importance:

```
Feature ranking:
1. Feature 3 representing LV Volume at ES (0.153645)
2. Feature 11 representing RV and LV Volumes ratio at ES (0.119265)
3. Feature 8 representing LV Ejection Fraction (0.118564)
4. Feature 10 representing RV and LV Volumes ratio at ED (0.117171)
5. Feature 13 representing MYO and LV Volumes ratio at ES (0.103105)
6. Feature 2 representing LV Volume at ED (0.075727)
7. Feature 12 representing MYO and LV Volumes ratio at ED (0.069549)
8. Feature 9 representing RV Ejection Fraction (0.065073)
9. Feature 5 representing RV Volume at ES (0.049786)
10. Feature 7 representing MYO Volume at ES (0.048601)
11. Feature 6 representing MYO Volume at ED (0.041553)
12. Feature 4 representing RV Volume at ED (0.024184)
13. Feature 1 representing Weight (0.007020)
14. Feature 0 representing Height (0.006756)
```

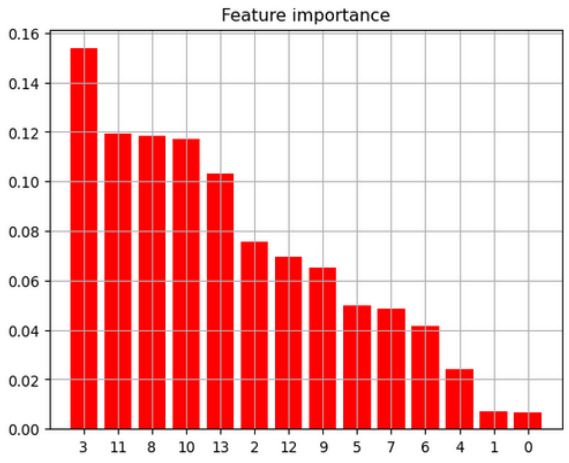


Figure 5: Feature importance. The 7 most important features are LV-related.

We search for roughly optimal parameters and obtain a 0.92 validation score. This model confirms that LV features are the most important in this classification task. Patients' height and weight appear to be irrelevant, probably because we have normalized the volumes, thus we remove them.

IV – Left ventricle segmentation

Based on the previous results, we conclude that LV segmentation on the test set is mandatory. Because we have access to the segmentations of other organs, we do not need a complex deep learning algorithm for the LV segmentation. We can instead use the Myocardium segmentation: the LV is always exactly inside it. We obtain the LV segmentation for each slice by filling the inside of the myocardium from its contours.

We obtain an almost perfect Dice score (0.996) on the training images. The errors happen when the myocardium is open:

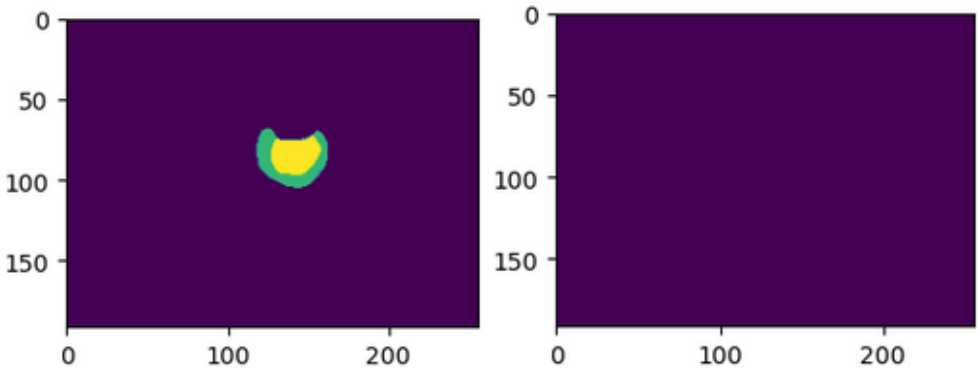


Figure 6: Right: true segmentation of an image which resulted in an error. The myocardium is in green and the LV in yellow. Left: The result of our segmentation on this image.

In this case, the obtained image for the LV segmentation is all black. We can check all such cases in the test set to see if any contains an open myocardium. Out of all the LV segmentation on the test set where no LV was found, there was either no myocardium or the myocardium was closed, which are all cases where there was no LV to segment in the image.

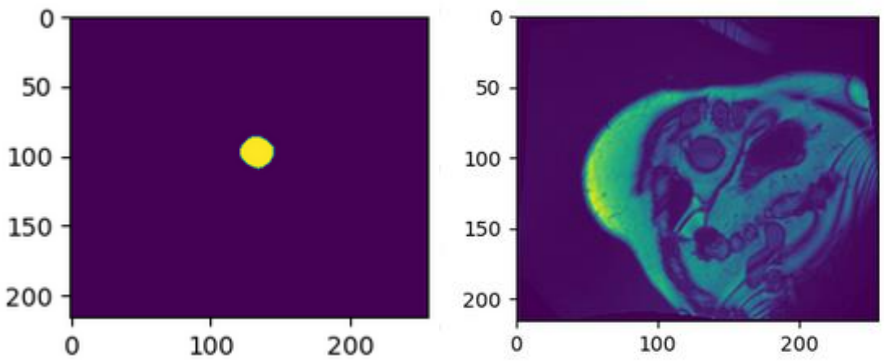


Figure 7: A slice in the test set which resulted in an all black LV segmentation. The myocardium (in yellow) is closed.

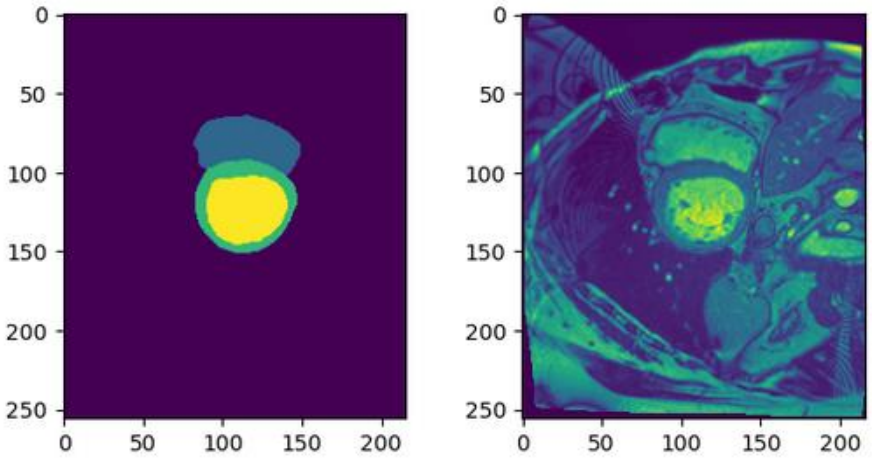


Figure 8: Result of our LV segmentation on an image from the test set.

V – Methods and model selection

Because the number of samples is low, we will use leave-one-out cross-validation to compare models to make sure the models have the most training samples to learn from while still having a fair evaluation of the performance.

We will use grid search cross-validation in order to find the optimal hyperparameters of the models suited for our evaluation method.

We begin by setting a benchmark with a linear classifier. We use a linear SVM and find $C = 0.41$ as the optimal regularization value. The validation accuracy is 0.91, comparable to the first random forest result. Most of the samples are correctly labeled, but there is some confusion between the classes 1 and 2 (Myocardial infarction and Dilated cardiomyopathy) and between the classes 0 and 3 and 4 (Healthy control, Hypertrophic cardiomyopathy, Abnormal right ventricle).

The results of a simple linear classifier are promising and show that the boundaries between classes are quite well delimited except for some samples. However, the model seems to be overfitting slightly which is something to be wary of with more complex models.

We now try to improve performances with non-linear boundaries. We have few samples and few features; therefore, we will first try SVMs which are capable of generalizing well even with a low number of training samples. Other advantages of SVMs are their efficiency in training and predictions.

In order to find the optimal hyperparameters, we perform two grid searches: one over a large space and one more precise. We find that the best kernel is the sigmoid kernel, with $C = 11.5$ and $\gamma = 0.1$. The validation score is 0.97, a strong improvement over previous results, and overfitting appear to be limited. The confusions have been greatly reduced.

Ensemble methods based on decision trees such as random forests are well suited for tabular data, and can efficiently learn from the training data even with few samples. We first try a simple bagging model: we achieve 0.95 accuracy for the cross-validation with 80 estimators each trained of 50% of the samples and 3 features. This model overfits too much: we can reduce that using random forests which limit correlation between trees.

With 10 estimators trained on 1 feature each, we obtain a 0.97 validation score, similar to the SVM. We observe that a lot of regularization is needed in order to avoid overfitting: the number of estimators is low and each tree is very simple.

VI – Results

Recap of the different models:

Model	Train score	Validation score
Linear SVM	0.94	0.91
SVM (sigmoid kernel)	0.979	0.97
Bagging (80 estimators)	0.99	0.95
Random Forest (10 estimators)	0.983	0.97

The SVM and Random Forest models both share similar cross-validation performances. Although the SVM is faster in training and prediction, the Random Forest has a slightly higher performance on the training set which could indicate better performances on the test set.

Finally, the model used for the submission is a Random Forest with 10 estimators having a number of minimum samples for a split equal to 4 and trained on 1 feature each. This model shows strong validation performances with 0.97 accuracy and does not seem to overfit. It is very fast to train and to make predictions.

For the submission, we compute and standardize features in the same way, having segmented the LV previously. We train the model on the whole train set and evaluate it on the test set.

VII - References

[1]: Bernard et al., "Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved?," in IEEE Transactions on Medical Imaging, vol. 37, no. 11, pp. 2514-2525, Nov. 2018, doi: 10.1109/TMI.2018.2837502.

[2]: Wolterink, J.M., Leiner, T., Viergever, M.A., Išgum, I. (2018). Automatic Segmentation and Disease Classification Using Cardiac Cine MR Images. In: , et al. Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges. STACOM 2017. Lecture Notes in Computer Science(), vol 10663. Springer, Cham. https://doi.org/10.1007/978-3-319-75541-0_11