IMA205 - Kaggle Challange Report

Cardiac Pathology Prediction

Luiz Augusto Facury de Souza

7th May 2023

Contents

1 Data Description

The data used for the Kaggle challenge consists of 150 MRIs and their corresponding segmentations of the Myocardium (MYO), Left Ventricle (LV), and Right Ventricle (RV). In addition, the height and weight of each patient and the classification of his heart (5 classes) were provided. The test set, composed of 50 samples drawn from the original data, does not have the segmentation of the LV. Therefore, it needs to be done because many of the features that will be extracted depend on the LV's mask. Finally, as each image has a different number of channels and sizes, it is difficult to feed it to some models, such as CNNs, so it is necessary to bypass this problem.

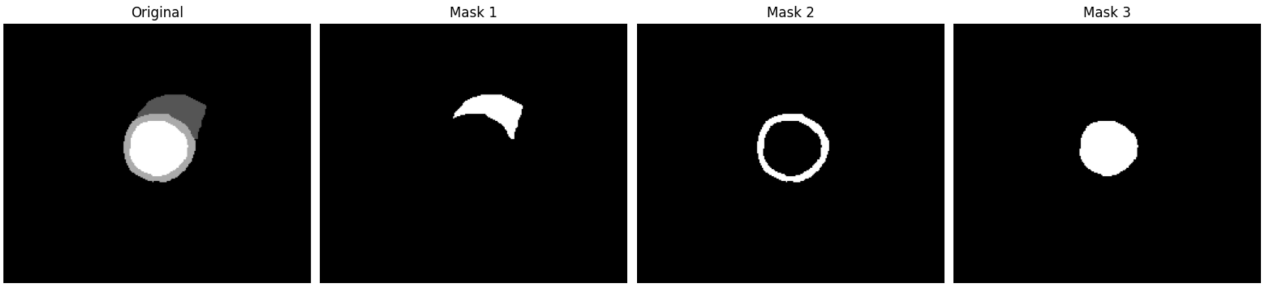

Figure 1: Segmentations of 1 slice, where 1 is RV, 2 is MYO and 3 is LV.

2 Segmentation

As seen in Figure 1, the LV is located inside the MYO, which allows for various techniques to be utilized for segmentation. While the U-net architecture suggested by Isensee et al. may be a viable option, standardizing the number of slices in each image is necessary before feeding the data to a neural network. Moreover, the images provided by the repository shared by Iakubovskii must have a width and height that are multiples of 32. However, simply adding channels of only 0s until all images have the same number of slices and resizing them to have multiples of 32 in their dimensions did not produce satisfactory results. As the authors did not specify how they overcame this problem, and did not provide comprehensive details regarding the architectures used, another approach was attempted.

A simpler approach was subsequently used, which involved utilizing the "binary_fill_holes" function from scipy, given that the LV is the area within the MYO. This function fills the hole inside the MYO, resulting in highly satisfactory outcomes, as depicted in the Python Notebook, with a Jaccard Score of 0.96 across all slices of all images for the train_ed set.

3 Feature Extraction

3.1 Approaches in the papers

To classify the heart, 3 papers were followed:

1. Wolterink et al. utilized 14 features: LV, RV, and myocardial volume at ED and ES (in ml), the LV and RV ejection fraction (EF), the ratio between RV and LV volume at ED and ES, and the ratio between myocardial and LV volume at ED and ES.

2. Isensee et al. extracted a series of instants and dynamic features from the segmentation maps.

3. Khened et al. used 11 features: Ejection Fraction of LV and RV, Volume of RV and LV at ED and ES, Mass of MYO at ED and Volume at ES, Height, and Weight

In addition Bernard et al. wrote a review of all those papers that classified the ACDC challange and compared them.

3.2 Implementation

Since Wolterink et al. is more detailed and specified, its features and model were first used. After that, some other modifications were derived from Isensee et al., while all the features used by Khened et al. were also used by Wolterink et al.. Therefore, as it can be seen in the code given, the features were extracted by:

1. Using the nibabel library, load all the images.

2. Obtain the volumes of the 3 masks at ED and ES by calculating the product of the spacing array obtained from the header of each image, and multiplying it by the sum of voxels in each 3D image.

3. Copmute the Ejection Fractions of LV and RV using $EJLV = \frac{LV_{ED}-LV_{ES}}{LV_{ED}}$ and $EJRV = \frac{RV_{ED}-RV_{ES}}{RV_{ED}}$, where $EJRV$ is the Ejection Fraction of RV, $EJLV$ is the Ejection Fraction of the LV, and YV_{EX} is the Volume of the mask YV at EX.

4. Compute the 4 ratios by dividing the Volumes of RV at ED by LV at ed, RV at ES by LV at ES, MYO at ED by LV at ED, and MYO at ES by LV at ES.

5. Concatenate the 12 obtained features with the patient's height and weight, in order to get all the 14 features presented in the paper.

After obtaining the 14 features from Wolterink et al., the Mosteller formula was employed to obtain the body surface area (BSA) of each patient, as used by Isensee et al., since the size of the masks also depends on the height and weight of the person, so it needs to be normalized to be more coherent. Therefore, the volumes were divided by the factor $BSA = \sqrt{\frac{Height*Weight}{3600}}$.

Regarding the other features used by Isensee et al., they were found not to be very useful, except for the myocardial thickness, which was implemented using the "canny" function from scipy. This function computes the contour of the myocardium to obtain the exterior and interior contour and the maximum, minimum, mean, and standard deviation of the myocardium's thickness across the slices of each image at ED and ES.

At the end of the process, a set of 22 features were utilized, which included the height and weight of each patient, the volumes of the masks at ED and ES normalized using the Mosteller formula, the Ejection Fractions, the Ratios mentioned previously, and the mean, max, min, and std of the myocardium's thickness. All of these variables have a sufficiently high correlation with the heart's Category, as demonstrated in the correlation matrix plotted in the notebook. While some variables, such as the minimum thickness of the myocardium, had lower correlations, removing them did not lead to any improvement in the models' performance, and hence, they were retained for the final models.

4 Classification

4.1 Approaches in the papers

1. Wolterink et al. employed a Random Forest with 1000 trees that were grown to full depth as the classifier model.

2. Isensee et al. utilized an ensemble of 50 multilayer perceptrons (MLPs) and a random forest to perform the classification.

3. Khened et al. employed a Random Forest with 100 trees.

4.2 Implementation

Since a random forest is being implemented, data normalization is not necessary, as the model simply partitions the data to make predictions. However, in order to test the MLP model, data normalization was applied.

Initially, a Random Forest Classifier (RF) was tested with 100 and 1000 trees, as implemented by Khened et al. and Wolterink et al.. Subsequently, an ensemble with MLP and RF was tested, similar to the approach used by Isensee et al..

Since there is not a large amount of data available, deep learning models tend not to work as well, especially in the case of tabular data, which is the type of data used in this project. This has been confirmed in studies such as Gorishniy et al. and Shwartz-Ziv and Armon. Therefore, Xgboost [2], a regularized version of the Gradient Boosting model, was implemented to compare with the other models, as it performs extremely well with tabular data and is also based on Decision Trees.

After selecting the models, a randomized grid search was conducted to find the optimal hyper-parameters for each model, using 10-fold cross-validation on the training data. Upon comparing the results of all implemented models, the RF model with 1000 trees achieved the highest accuracy in the cross-validation process, while the Xgboost and ensemble models did not yield satisfactory results. Thus, the model proposed by Wolterink et al. was deemed the best model, with the added features related to MYO thickness and normalization using the Mosteller formula.

5 Conclusion

Since we only had access to the accuracy of a subset of the test set, it was difficult to effectively predict how our model would perform in the private score. Therefore, the best way available to optimize it was using cross-validation, which yielded similar accuracy for different models and variables used. The highest accuracy achieved by the 10-Fold CV was approximately 0.88, which corresponded to 0.93 in the public score and 0.86 in the private score. Therefore, the results were close to the ones found in the cross-validation. However, there was a result that had an accuracy of 0.86 in the public score and 0.91 in the private score, which was not chosen because it was impossible to know if it would perform better in the private score than the one that obtained an accuracy of 0.93 in the public score.

In general, the models performed as expected, but it was difficult to improve the accuracy in relation to the one proposed by Wolterink et al., since none of the possible changes seemed to have a significant effect on the CV accuracy. Finally, it is possible to confirm that the extracted features are significant and able to well separate the data between the classes. Although the score is not perfect, it is relatively high, considering a complex real-life application with few data.

References

[1] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P.-A. Heng, I. Cetin, K. Lekadir, O. Camara, M. A. G. Ballester *et al.*, "Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved?" *IEEE transactions on medical imaging*, vol. 37, no. 11, pp. 2514–2525, 2018.

[2] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[3] Y. Gorishniy, I. Rubachev, V. Khrulkov, and A. Babenko, "Revisiting deep learning models for tabular data," *Advances in Neural Information Processing Systems*, vol. 34, pp. 18 932–18 943, 2021.

[4] P. Iakubovskii, "Segmentation models," https://github.com/qubvel/segmentation_models, 2019.

[5] F. Isensee, P. F. Jaeger, P. M. Full, I. Wolf, S. Engelhardt, and K. H. Maier-Hein, "Automatic cardiac disease assessment on cine-mri via time-series segmentation and domain specific features," in *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges: 8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017, Quebec City, Canada, September 10-14, 2017, Revised Selected Papers 8*. Springer, 2018, pp. 120–129.

[6] M. Khened, V. Alex, and G. Krishnamurthi, "Densely connected fully convolutional network for short-axis cardiac cine mr image segmentation and heart diagnosis using random forest," in *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges: 8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017, Quebec City, Canada, September 10-14, 2017, Revised Selected Papers 8*. Springer, 2018, pp. 140–151.

[7] R. Shwartz-Ziv and A. Armon, "Tabular data: Deep learning is not all you need," *Information Fusion*, vol. 81, pp. 84–90, 2022.

[8] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum, "Automatic segmentation and disease classification using cardiac cine mr images," in *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges: 8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017, Quebec City, Canada, September 10-14, 2017, Revised Selected Papers 8*. Springer, 2018, pp. 101–110.