Scar-Related Ventricular Arrhythmia Prediction from Imaging Using Explainable Deep Learning

Buntheng Ly, Sonny Finsterbach, Marta Nuñez-Garcia, Hubert Cochet, Maxime Sermesant

Scar-Related Ventricular Arrhythmia Prediction from Imaging using Explainable Deep Learning

Buntheng Ly[1], Sonny Finsterbach[2], Marta Nuñez-Garcia[2], Hubert Cochet[2], and Maxime Sermesant[1,2]

[1] Inria, Université Côte d'Azur, Epione team, Sophia Antipolis, France
[2] IHU Liryc, Université de Bordeaux, Bordeaux, France
Contact: `maxime.sermesant@inria.fr`

Abstract. The aim of this study is to create an automatic framework for sustained ventricular arrhythmia (VA) prediction using cardiac computed tomography (CT) images. We built an image processing pipeline and a deep learning network to explore the relation between post-infarct left ventricular myocardium thickness and previous occurrence of VA. Our pipeline generated a 2D myocardium thickness map (TM) from the 3D imaging input. Our network consisted of a conditional variational autoencoder (CVAE) and a classifier model. The CVAE was used to compress the TM into a low dimensional latent space, then the classifier utilised the latent variables to predict between healthy and VA patient. We studied the network on a large clinical database of 504 healthy and 182 VA patients. Using our method, we achieved a mean classification accuracy of 75%±4 on the testing dataset, compared to 71%±4 from the classification using the classical left ventricular ejection fraction (LVEF).

Keywords: Conditional-VAE · Sustained Ventricular Arrhythmia · CT cardiac imaging · Myocardium thickness · Image Classification

1 Introduction

VA is highly associated with sudden cardiac death (SCD), one of the major causes of death in the developed countries [10]. Increased risk of VA has been linked to the presence of left ventricular myocardial (LVMYO) scar. Although, the current gold standard for SCD risk stratification is the LVEF, most SCD events still occur in patients with higher LVEF than the cut-off value of 35%. Further evidence has shown LVMYO scar to be a better independent predictor of SCD than the LVEF. Nonetheless, the detection and characterisation of the scar region are still challenging, thus its application in clinical practice is limited. Recent studies have linked the potential of CT imaging in locating the LVMYO scar region through the assessment of wall thinning [9]. Moreover, the studies have also found VA substrate within or next to CT-defined scar region.

The advent of Deep Learning (DL) has given rise to several image classification networks of exceptional accuracy. Among the different network configurations, we investigate the embedding-classification architecture, which has been

successfully implemented in classification tasks for medical imaging [1,2,6]. The two-part architecture includes an embedding model and a classification model. Multiple DL architectures have been used as the embedding network, such as the deep residual network [1], the convolutional autoencoder [6], and the variational autoencoder [2]. In our framework, we used the CVAE architecture [13], which allows a more robust reconstruction of the input. Moreover, it enables to analyse the explainability of the results thanks to its generative nature.

In this paper, we studied the correlation between the myocardial wall thinning and the history of VA. To this end, we constructed an automatic pipeline to generate 2D TM from the imaging input and the CVAE-Class network to embed and classify the 2D TM.

2   Methods

Our pipeline used 3D cardiac CT image as the main input. Myocardial masks were automatically segmented and were used for TM computation. The 3D TM was then mapped into a uniform circular 2D image, which ensures the shape consistency between input images. The 2D TM was then fed the CVAE-Class network, where it was projected into the latent variables, which were then used for VA classification.

2.1   Image Processing

Fig. 1 outlines our automatic image processing pipeline. The main steps included the cardiac segmentation, the short-axis view (SAX) orientation, the TM calculation, the LV flattening, and the thickness value normalisation.

Image Segmentation To start with, we generated the ventricular masks from cardiac CT using a pre-trained Dual-UNet segmentation network proposed in [4]. The required ventricle masks included the right ventricular epicardium (RVEPI), the left ventricular epicardium (LVEPI) and endocardium (LVENDO). Then, the LVMYO could be extracted from LVEPI and LVENDO.

Short-Axis Orientation An automatically computed rigid transformation was applied to re-orient the ventricular masks in the standardised SAX view following the method presented in [11]. This step also included an isotropic resampling of the image and masks that along with the standardised SAX orientation highly facilitated tissue thickness calculation and the computation of the 2D TM map.

Thickness Map calculation The LVMYO thickness was calculated from the LVEPI and LVENDO masks using the Eulerian partial differential equation approach for tissue thickness calculation [14]. The python package implemented for [3] was made available by the author[3]. A 3D surface model of the LV was

---
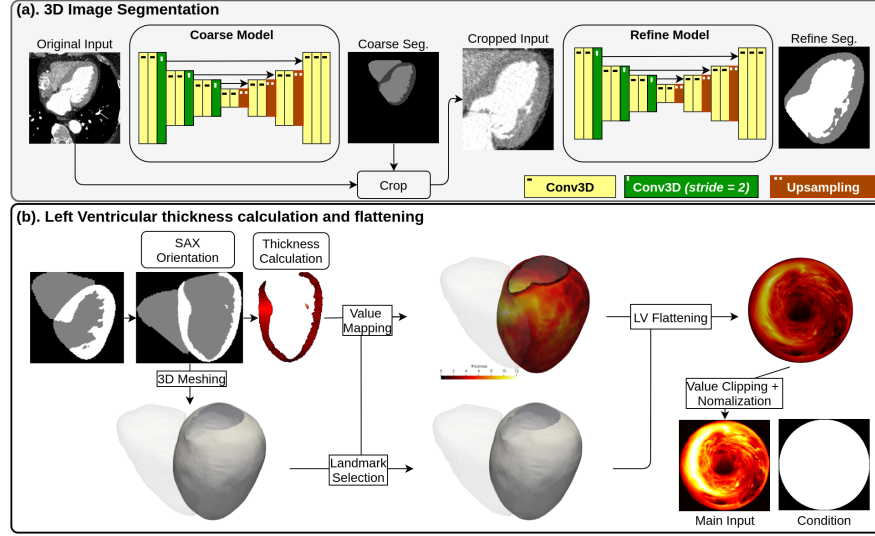
[3] https://pypi.org/project/pyezzi/

Fig. 1. Automatic pipeline to extract the 2D thickness map from 3D cardiac CT.

computed using the discrete marching cubes algorithm provided by the VTK package[4] and the thickness information was projected onto the mesh.

LV Flattening To transform the TM from 3D to 2D, we used the LV quasi-conformal flattening algorithm[5] based on the left atrial flattening proposed by [12]. In their LV extension, the authors proposed the manual selection of three landmarks: one at the centre of the apex and two points at the base. In this work, thanks to the standardised representation of the ventricles in SAX view, the landmark selection step could be automated.

Value clipping and Normalisation To prepare the TM as input to the DL model, we mapped the 2D TM mesh onto a 2D array in 256×256 resolution. We clipped the maximum thickness value at 10 mm, and divided the result by 10 to limit the values to $[0, 1]$. This pre-processing was done to generate a normalised map with increased separation between the regions with low value.

This processing pipeline reduced the input dimension from 3D into 2D, while still retained the 3D information in term of the wall thickness. Additionally, the uniform circular shape of the TM removed the variability of the LV shape in the different scan images, as the flattening landmarks ensured the rotation consistency between the inputs. Therefore, the model could be trained to focus on the specific characteristics of the TM, including the presence, the extent and the location of the thin regions.

---

[4] https://vtk.org/

[5] https://github.com/martanunez/LV_flattening

2.2   Model

The CVAE-Class network was a combination of a CVAE and a fully connected classifier model, as shown in figure 2. The network was built and trained using the Tensorflow 2.0 package[6].

Fig. 2. CVAE-Class network architecture. Please refer to the colour or the symbol in the bottom-left corner to distinguish the layer types. The number in each layer indicates the number of output filters or units.

CVAE Model The CVAE consisted of an encoder and a decoder branch. In the encoder branch, we used the concatenate layer to combine the main input image with the condition mask, following by a series of 1-stride and 2-stride convolution layers. The 2-stride layers progressively reduced the resolution from 256×256 to 16×16. At the bottleneck, we flattened the output features before applying the fully connected layers to generate the latent mean μ and log variance σ vector. Thereafter, we applied the Gaussian sampling method to generate the latent vector z. The latent dimension was set to 25.

In the decoder branch, we used a fully connected layer to increase the number of features of the latent vector z. Then, we reshaped the vector to the same shape as before the flattening layer. The decoder was then built using the sequences of regular and transpose convolution layer. We set the stride of the transpose layer to 2 to increase the resolution of the output. At each decoding level, we combined the condition mask and the main output using multiplication. A convolution layer with an adjusted number of stride was used to rectify the resolution difference. Finally, we multiplied the features with the conditional mask to get the final reconstruction output. We conditioned this VAE with the mask in order to prevent the network from using its parameters to correctly predict the circular shape of the map.

---

[6] https://www.tensorflow.org/

The CVAE was trained using the sum of the reconstruction loss (L_{recon}) and the Kullerback-Leibler divergence loss (L_{KL}), eq. 1. The L_{recon} was calculated using the weighted root mean square error, defined in eq 2, where N is the number of pixels, $\hat{y}$ denotes the decoder reconstruction output, y the input TM and m is the binary mask. The L_{KL} was calculated between the latent distribution $\mathcal{N}(\mu, \sigma^2)$ and the normal distribution $\mathcal{N}(0, 1)$. The β value was set to 0.01.

$$L_{CVAE} = L_{recon} + \beta L_{KL} \qquad (1)$$

$$L_{recon} = \sqrt{\frac{1}{N} \sum_i \sum_j (\hat{y}_{ij} - y_{ij})^2 \times (2m_{ij} - y_{ij})} \qquad (2)$$

Classifier Model The classifier model used the latent mean μ as input. The model was built using three blocks of two fully connected layers and a tanh activation layer. The numbers of units were set to 16, 8 and 4, respectively. To get the classification output (healthy and VA), we used a fully connected layer with 2 units followed by the softmax function.

The classifier was trained using the hinge loss, as defined in Eq. 3, where N denotes the number of classes, y denotes the ground truth and $\hat{y}$ denotes the classification score. Noted that the expected value of the ground truth for each class is between [-1, 1], while the classification score is between [0, 1].

$$L_{hinge} = \frac{1}{N} \sum_i^N \max\{1 - y_i * \hat{y}_i, 0\} \qquad (3)$$

3 Experimental Setup

Dataset A large series of CT scan datasets were retrospectively collected from Bordeaux University Hospital. The inclusion criteria were: history of myocardial infarction or ischaemic heart disease and available contrast-enhanced and ECG gated CT scan reconstructed in diastole. The medical records of these patients were reviewed to collect clinical characteristics, including demographics, the delay since myocardial infarction, the LVEF and the history of scar-related VA. Patients with the history of cardiac catheter ablation or ventricular surgery were not considered for inclusion. The LVEF was considered valid when measured with MRI or echocardiography closest to the time of CT within the limit of 1 year. VA events were qualified as past episodes of sustained ventricular tachycardia, ventricular fibrillation, aborted sudden cardiac arrest, or any appropriate therapy delivered by an implantable cardioverter defibrillator less then 1 year before the CT. VA occurring at the acute stage of the infarct (less than 1 month) were not considered as scar-related and were not analysed as such. Patients with incomplete medical records were also excluded from the analysis.

A total of 686 patients (age 73 ± 12, 83.4% men and 26.5% with VA) were included. We used the bootstrap method [8] to estimate the mean accuracy and

the standard error of our model. To account for the computation time of the deep learning model, the number of resampling was set to 25, which has been shown to be sufficient in gauging the randomness caused by the sampling data [7]. For each resampling, we randomly selected 100 cases (50 healthy and 50 VA patients) to be used as testing dataset.

Training Setup To balance between healthy and VA patients, we trained and validated the model with the same number of data from each class at each iteration.

The CVAE-Class networks were trained step-by-step, where we first trained the CVAE to adequately embed the input before optimising the classifier. Both models were optimised using Adam optimiser with the initial learning rate at $1e - 4$. The learning rate was reduced by half after 5 epochs with no validation improvement. The training was set to stop when the validation loss does not improve after 50 epochs. The model was rerun 5 times, afterward the model with the highest validation accuracy is selected.

Baseline model To compare our network performance, we also built a straightforward image classifier network using the architecture of the encoder and the classifier of the CVAE-Class network. Additionally, we run the test with a threshold model using the LVEF cut-off value. To adjust the value to our sample, we recalculated the best cut-off value to maximise VA classification accuracy in the training set, then applied the value to the testing data. We also trained a logistic regression model using the LVEF value.

4   Results

4.1   Accuracy

Table 1 shows the comparison of the classification accuracy of the CVAE-Class and the baseline models, at 25 resamplings. To adjust for the small resmapling size, the confidence intervals were calculated using the approximate bootstrap confidence interval method as proposed by [8]. The mean LVEF cut-off value was 39.4%.

The CVAE-Class model scored the highest accuracy, with an average of 75% with 4% standard deviation. The image classifier model achieved a mean accuracy of 73%. Lastly, the $LVEF_{LogReg}$ and $LVEF_{cut-off}$ model felt behind at 71% and 70%, respectively. The better performance of the CVAE-Class compared to straightforward image classifier model suggests that the latent variables generated by the CVAE network were relevant in improving the classification accuracy. Both models also showed better performance than the LVEF models, which further highlights the potential of the 2D TM in identifying patients with history of VA.

Table 1. The CVAE-Class and the baseline models classification accuracy.

Model	Mean Accuracy	St.Deviation	90% CI
CVAE-Class	**0.75**	**0.04**	**(0.74, 0.77)**
Img-Class	0.73	0.04	(0.72, 0.75)
LVEF$_{\text{LogReg}}$	0.71	0.05	(0.70, 0.73)
LVEF$_{\text{cut-off}}$	0.70	0.04	(0.69, 0.72)

4.2   Explainability

We tested the explainability of the CVAE-Class model using the GradCam++ method proposed in [5], which computes the classification attention map using the prediction score. This could help the user in understanding what was used by the network for its prediction.

Fig. 3 shows a series of attention maps including the true and false positive cases (Fig. 3.a and Fig 3.b) and the true and false negative cases (Fig. 3.c and Fig 3.d). As expected, we can see higher attention values in the thin regions, and higher positive prediction scores were placed on the cases with larger thin regions. In consequence, however, the model failed with the patient who presented wall thinning without VA history (Fig. 3.b) and vice versa when VA occurred in the absence of wall thinning (Fig. 3.d).

The generative nature of the model could also be used to explore the latent space and better understand the learned representation, as well as the differences between classes.

5   Discussion

Due to the compression of VAEs, the reconstructed TM did not contain some high frequency details, as shown in the Figure 4.a. While this could be due to the decoder branch, it could also suggest that the latent vectors did not hold necessary information to decode the missing details. Although, our current framework focused on the classification and did not directly use the reconstructed output, ensuring the latent vector contains heterogeneity information could lead to better classification accuracy.

We analysed the latent variables of the test data using the partial least square regression, Fig. 4.b, where we could observe partial overlap between the two classes. Adding constraints to increase the separation of the latent variables could allow for a more robust classification model. Moreover, in a well separated latent space, we can exploit the latent variables to generate a TM of the different classes. This would allow us to display the personalised features that influence the classification.

As our current model was only trained with the TM input, its accuracy was ultimately tied to the cases where the VA was related to the wall thinning, as
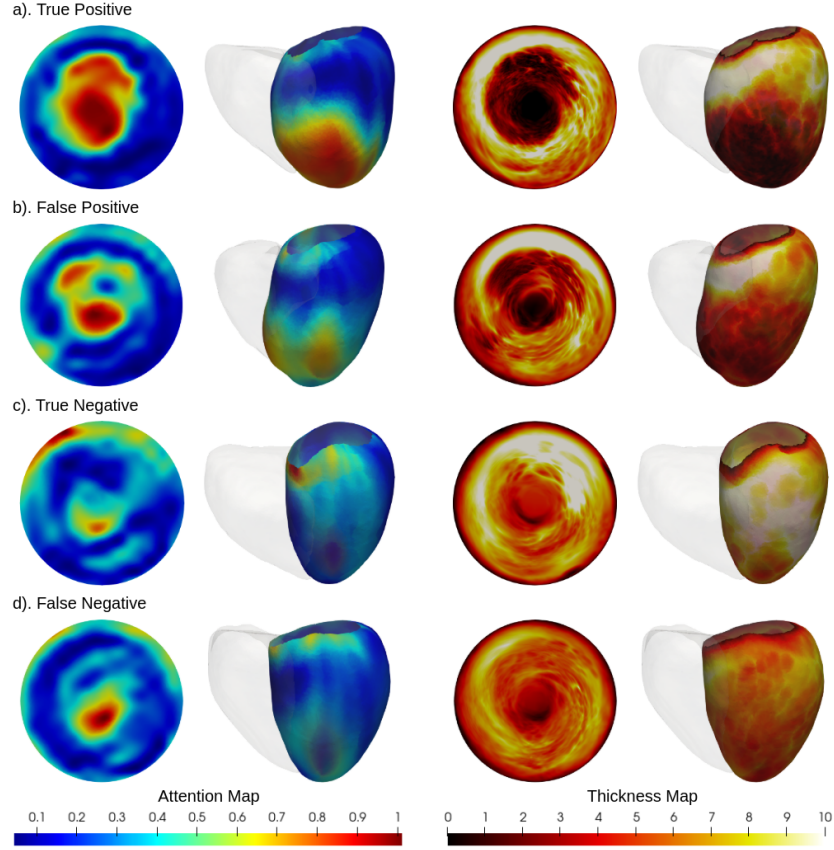
Fig. 3. The attention map and its corresponding thickness map. Higher gradients (Red) were placed on the thin regions of the thickness map, as expected, the scar was driving the decision of the model.

illustrated by the false predictions in Fig. 3 The subtle differences between the TM may be the key to a more accurate classification, therefore it is crucial to improve the latent representation of the TM. On the other hand, integrating the clinical information into the current framework can be the direct answer to the single-input limitation.

6   Conclusion

We proposed an automatic pipeline for VA classification using 3D CT images. Using a 2D thickness map, our model showed a clear increase in classification accuracy as compared to the current gold standard classification using LVEF value. We also provided evidence that the CVAE network generated meaningful
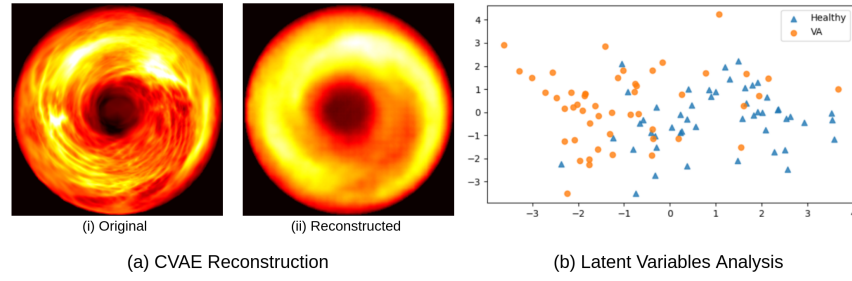
(i) Original   (ii) Reconstructed

(a) CVAE Reconstruction   (b) Latent Variables Analysis

Fig. 4. Image reconstruction (a) and latent variables analysis (b).

latent variables that improved the classification accuracy compared to a straight-forward image classification network. In the future work, we wish to explore further into the latent variables enhancement and clinical information integration to increase classification accuracy, as well as to exploit the generative capability of the network.

7 Acknowledgments

References

1. Abbet, C., Zlobec, I., Bozorgtabar, B., Thiran, J.P.: Divide-and-rule: Self-supervised learning for survival analysis in colorectal cancer. In: Martel, A.L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M.A., Zhou, S.K., Racoceanu, D., Joskowicz, L. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. Springer International Publishing, Cham (2020)
2. Biffi, C., Oktay, O., Tarroni, G., Bai, W., De Marvao, A., Doumou, G., Rajchl, M., Bedair, R., Prasad, S., Cook, S., O'Regan, D., Rueckert, D.: Learning interpretable anatomical features through deep generative models: Application to cardiac remodeling. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. pp. 464–471. Springer International Publishing, Cham (2018)
3. Cedilnik, N., Duchateau, J., Dubois, R., Sacher, F., Jaïs, P., Cochet, H., Sermesant, M.: Fast Personalized Electrophysiological Models from CT Images for Ventricular Tachycardia Ablation Planning. EP-Europace (Nov 2018)
4. Cedilnik, N., Duchateau, J., Sacher, F., Jaïs, P., Cochet, H., Sermesant, M.: Fully Automated Electrophysiological Model Personalisation Framework from CT Imaging. In: FIMH 2019 - 10th International Conference on Functional Imaging and Modeling of the Heart. pp. 325–333. Bordeaux, France (Jun 2019)

5. Chattopadhay, A., Sarkar, A., Howlader, P., Balasubramanian, V.N.: Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks. Proceedings - 2018 IEEE Winter Conference on Applications of Computer Vision, WACV 2018 pp. 839–847 (2018)
6. Dercksen, K., Bulten, W., Litjens, G.: Dealing with label scarcity in computational pathology: A use case in prostate cancer classification. arXiv (2019)
7. Efron, B., Rogosa, D., Tibshirani, R.: Resampling Methods of Estimation, vol. 19. Elsevier, second edition edn. (2015)
8. Efron, B., Tibshirani, R.J.: An Introduction to the Bootstrap. In: An Introduction to the Bootstrap. CRC Monographs on Statistics and Applied Probability, CRC Press (1993)
9. Mahida, S., Sacher, F., Dubois, R., Sermesant, M., Bogun, F., Haïssaguerre, M., Jaïs, P., Cochet, H.: Cardiac Imaging in Patients with Ventricular Tachycardia. Circulation (2017)
10. Nielsen, J.C., Lin, Y.J., de Oliveira Figueiredo, M.J., Shamloo, A.S., Alfie, A., Boveda, S., Dagres, N., Di Toro, D., Eckhardt, L.L., Ellenbogen, K., Hardy, C., Ikeda, T., Jaswal, A., Kaufman, E., Krahn, A., Kusano, K., Kutyifa, V., Lim, H.S., Lip, G.Y., Nava-Townsend, S., Pak, H.N., Diez, G.R., Sauer, W., Saxena, A., Svendsen, J.H., Vanegas, D., Vaseghi, M., Wilde, A., Jared Bunch, T., Buxton, A.E., Calvimontes, G., Chao, T.F., Eckardt, L., Estner, H., Gillis, A.M., Isa, R., Kautzner, J., Maury, P., Moss, J.D., Nam, G.B., Olshansky, B., Pava Molano, L.F., Pimentel, M., Prabhu, M., Tzou, W.S., Sommer, P., Swampillai, J., Vidal, A., Deneke, T., Hindricks, G., Leclercq, C.: European Heart Rhythm Association (EHRA)/Heart Rhythm Society (HRS)/Asia Pacific Heart Rhythm Society (APHRS)/Latin American Heart Rhythm Society (LAHRS) expert consensus on risk assessment in cardiac arrhythmias: Use the right tool for the right outcome,. Europace (2020)
11. Nuñez-Garcia, M., Cedilnik, N., Jia, S., Sermesant, M., Cochet, H.: Automatic multiplanar CT reformatting from trans-axial into left ventricle short-axis view. In: STACOM 2020 - 11th International Workshop on Statistical Atlases and Computational Models of the Heart. Lima, Peru (Oct 2020)
12. Nuñez-Garcia, M., Bernardino, G., Alarcón, F., Caixal, G., Mont, L., Camara, O., Butakoff, C.: Fast quasi-conformal regional flattening of the left atrium. IEEE Transactions on Visualization and Computer Graphics (2020)
13. Sohn, K., Yan, X., Lee, H.: Learning structured output representation using deep conditional generative models. Advances in Neural Information Processing Systems pp. 3483–3491 (2015)
14. Yezzi, A.J., Prince, J.L.: An eulerian pde approach for computing tissue thickness. IEEE Transactions on Medical Imaging (2003)