



**HAL**  
open science

## Constraint-Based Model in Multimodal Learning to Improve Ventricular Arrhythmia Prediction

Evariste Njomgue Fotso, Buntheng Ly, Hubert Cochet, Maxime Sermesant

► **To cite this version:**

Evariste Njomgue Fotso, Buntheng Ly, Hubert Cochet, Maxime Sermesant. Constraint-Based Model in Multimodal Learning to Improve Ventricular Arrhythmia Prediction. *Statistical Atlases and Computational Modeling of the Heart (STACOM)*, Oct 2024, Marrakech, Maroc, Morocco. hal-04843518

**HAL Id: hal-04843518**

<https://inria.hal.science/hal-04843518v1>

Submitted on 17 Dec 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Constraint-Based Model in Multimodal Learning to Improve Ventricular Arrhythmia Prediction

Evariste Njomgue Fotso<sup>1</sup>, Buntheng Ly<sup>2</sup>, Hubert Cochet<sup>2</sup>, and  
Maxime Sermesant<sup>1,2</sup>

<sup>1</sup> Inria, Université Côte d’Azur, Epione team, Sophia Antipolis, France

<sup>2</sup> IHU Liryc, Université de Bordeaux, Bordeaux, France

Contact: maxime.sermesant@inria.fr

**Abstract.** Cardiac disease evaluation depends on multiple diagnostic modalities: electrocardiogram (ECG) to diagnose abnormal heart rhythms, and imaging modalities such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT) and echocardiography to detect signs of structural abnormalities. Each of these modalities brings complementary information for a better diagnosis of cardiac dysfunction. However, training a machine learning (ML) model with data from multiple modalities is a challenging task, as it increases the dimension space, while keeping constant the number of samples. In fact, as the dimension of the input space increases, the volume of data required for accurate generalisation grows exponentially. In this work, we address this issue, for the application of Ventricular Arrhythmia (VA) prediction, based on the combined clinical and CT imaging features, where we constrained the learning process on medical images (CT) based on the prior knowledge acquired from clinical data. The VA classifier is fed with features extracted from a 3D myocardium thickness map (TM) of the left ventricle. The TM is generated by our pipeline from the imaging input and a Graph Convolutional Network is used as the feature extractor of the 3D TM. We introduce a novel Sequential Fusion method and evaluate its performance against traditional Early Fusion techniques and single-modality models. The cross-validation results show that the Sequential Fusion model achieved the highest average scores of  $80.7\% \pm 4.4$  Sensitivity and  $73.1\% \pm 6.0$  F1 score, outperforming the Early Fusion model at  $65.0\% \pm 8.9$  Sensitivity and  $63.1\% \pm 6.3$  F1 score. Both fusion models achieved better scores than the single-modality models, where the average Sensitivity and F1 score are  $62.8\% \pm 10.1$ ;  $52.1\% \pm 6.5$  for the clinical data modality and  $62.9\% \pm 6.3$ ;  $60.7\% \pm 5.3$  for the medical images modality.

**Keywords:** multimodal deep learning · arrhythmia · imaging

## 1 Introduction

Sudden Cardiac Death (SCD), when the heart stops beating suddenly, is recognised as a high-priority public health topic. SCD is the most common cause of death worldwide, accounting for 4.25 million deaths every year [1]. Among

causes of SCD, ischaemic heart disease (also called coronary heart disease) is the most common cause [1]. Ischaemic heart disease is a significant risk factor for the development of Ventricular Arrhythmia, such as Ventricular Tachycardia (VT) and Ventricular Fibrillation (VF). Despite intensive research on the topic, over the past 25 years, the absolute number of Cardiovascular disease (CVD) cases has increased in Europe and in the EU, with increases in the number of new CVD cases found in most countries [2]. The most significant advance in the prevention of SCD has been the development of the implantable cardioverter-defibrillator (ICD) [3]. On the one hand people are still dying of SCD and on the other hand, the current prescription guidelines of ICDs which is left ventricular ejection fraction (LVEF)  $<30-35\%$  [7] capture only a mere 20% all SCD [8].

To enhance early risk identification of SCD, Dakun Lai et al. have proposed a machine learning model on measurable arrhythmic markers derived from ECG signals [11]. O’Mahony et al., have presented an individualised risk estimates for SCD in hypertrophied cardiomyopathy patients, considering various clinical parameters [10]. Global Electric Heterogeneity (GEH) parameters, quantifying the abnormal electrophysiological substrate, have also been identified as independent factors associated with SCD risk, improving risk prediction when combined with other clinical characteristics [12]. For paediatric patients, the HCM Risk-Kids model estimates the risk of SCD and highlights the importance of considering both clinical and genetic factors in risk assessment [13]. In order to improve the current clinical criteria for ICD candidacy, a combination (fusion) of clinical characteristics with other markers may significantly improve risk stratification [9], but feature fusion is still a challenging research topic in machine learning [5]. In the literature, feature concatenation is the primary technique used for data fusion in multimodal classification [15]. This concatenation is typically performed either at the beginning of the learning process (Early Fusion) or at the end of the learning process (Late Fusion).

In this study, we propose a novel fusion technique for multimodal VA classification (positive VA+ or negative VA-). The technique can also handle the class imbalance issue which is particularly prevalent in medical diagnosis, where the occurrence of positive cases (VA+) is significantly rarer than negative cases. Our Sequential Fusion technique progressively constrains the learning process on higher dimensional data (medical images modality), based on knowledge acquired on lower dimensional data (clinical data). The knowledge acquired on clinical data, called prior modality, is used in the loss function of the learning process on medical images modality. We also study the contribution of each modality in our fusion technique.

## 2 Method

### 2.1 Sequential Fusion Model

Inspired by constraints-based models, we propose the Sequential Fusion model. A constraints-based model refers to a type of modelling approach that incorporates constraints to guide the learning process or restrict the possible solutions. These

constraints can be mathematical expressions, logical statements, or any other form of limitation that helps narrow down the solution space and ensure that the model produces results that adhere to certain predefined criteria.

To setup a Sequential Fusion model with two modalities (fig. 1), a first classifier is trained only with the first modality (the prior modality). The classification result on the prior modality is used to compute the prior modality constraint as defined in the next chapter. This constraint is then added into the learning process with the second modality to get the final output.

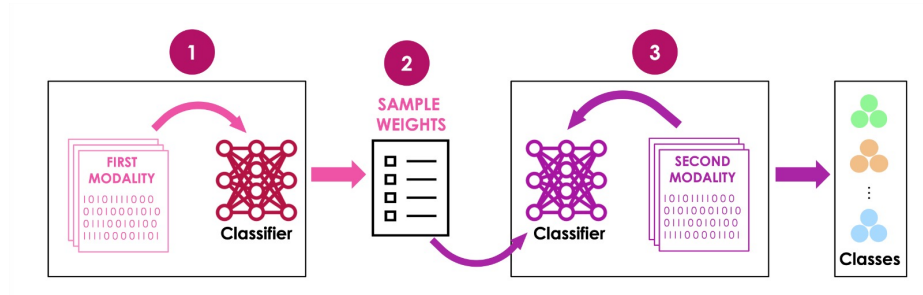


Fig. 1: Sequential Fusion Model: general setting.

## 2.2 Prior Modality Constraint Formulation

The first model is trained using only clinical data which is the prior modality: it provides prior knowledge for the final classifier on imaging data from the classification errors on the clinical data. We account for this prior knowledge with the following weighting strategy.

Let us consider two sets  $s_1$  and  $s_2$ , where  $s_1$  is the set of well-classified samples and  $s_2$  the set of misclassified samples using the prior modality. The prior modality weights ( $prior\_weight_i$ ) is computed using the eq. 1, where  $prior\_weight_i$  is the weight for each set  $s_i$ ,  $n$  is the total number of samples,  $k$  is the number of sets (two in our setting) and  $n_{s_i}$  is the number of samples in each set  $s_i$ .

$$prior\_weight_i = \frac{n}{k * n_{s_i}} \quad (1)$$

The prior modality weights ( $prior\_weight_i$ ) are the mathematical formulation of the prior-knowledge, where more importance is given to misclassified samples by the prior modality so that the final classifier is more focused on these samples.

Moreover, in our setting, the data set is unbalanced. To tackle the unbalanced class issue, we use the same samples weighting technique, to give higher weight to minority class (VA-) and lower weight to the majority class (VA+).

To control both of these effects, we define two hyper-parameters  $\alpha, \beta \in [0, 1]^2$  and two weighting strategies: same weighting, defined in eq. 2, and stratified weighting, defined in eq. 3.

$$w_l = \{ \alpha * class\_weight_j + \beta * prior\_weight_i \} \quad (2)$$

$$w_l = \begin{cases} \alpha * class\_weight_j & \text{for } s_1(i = 1) \\ \alpha * class\_weight_j + \beta * prior\_weight_i & \text{for } s_2(i = 2) \end{cases} \quad (3)$$

In equations 2 and 3,  $class\_weight_j$  corresponds to weights which correct the imbalance effect. The weight of each class  $j$  ( $class\_weight_j$ ) is computed by the same formulation in eq. 1, where  $n$  is the total number of samples,  $k$  is the number of classes (two in our setting) and  $n_{c_j}$  is the number of samples in each class  $c_j$ . In eq. 1,  $n_{s_i}$  is replaced by  $n_{c_j}$  for the computation of  $class\_weight_j$ .

In the same weighting strategy (eq. 2), equal importance is given to both the effects of unbalanced class and misclassified samples with the prior modality, whereas in stratified weighting strategy (eq. 3) more importance is given to positive (VA+) class and misclassified samples by the prior modality.

The loss function (binary cross-entropy) of the learning process on the medical images modality is then defined in eq. 4, where  $y_l$  is the true label of the  $l$ -th sample (1 for the positive class VA+ and 0 for the negative class VA- and  $p_l$  is the predicted probability of the  $l$ -th sample belonging to the positive class.

$$Loss_{BCE} = w_l * y_l \log(p_l) + w_l * (1 - y_l) \log(1 - p_l) \quad (4)$$

The Early Fusion model is trained with both weighting strategies and only the best one in terms of our performance metrics is kept (F1 score and Sensitivity). The novelty of our approach lies in these weights ( $prior\_weight_i$ ), which is computed based on the knowledge acquired from the prior modality. This formulation can be generalised in multi-class classification with any additional constraints required in the learning process.

### 3 Materials

#### 3.1 Data set and Features Extraction

We used a retrospective data set of myocardial infarction patients collected at a hospital between 2010 and 2020. Patients with history of surgical procedure on the LV were excluded from the cohort. In this work, patient’s cardiac CT image, with the scan date more than 1 month after the infarction, and clinical descriptions were collected. We classified as VA+ the patients, who have experienced sustained VT, VF and abort cardiac arrest. Finally, we obtained a study data set of 600 patients, with 165 and 435 of VA+ and VA- patients, respectively. Table 1 gives details on available clinical data modality.

To extract the thickness map features from CT images, we used the automatic pipeline proposed by a cardiac imaging group in 2022. [4].

Table 1: Clinical Data Modality (N=600)

Feature Name	Statistics: [min, max] mean or proportion
LVEF %	[10, 78] 45.07±13.36
Age (years)	[27, 99] 72.69±12.09
Smoking	270 (45%)
Dyslipidaemia	436 (73%)
Diabete	159 (27%)
Hypertension	394 (66%)
Sex	100 Females / 500 Males
VA	Positive (VA+) 165 (28%)

**Description:** LVEF % - Left Ventricular Ejection Fraction, Age - Patient age at the scan date.

### 3.2 Experimental setup

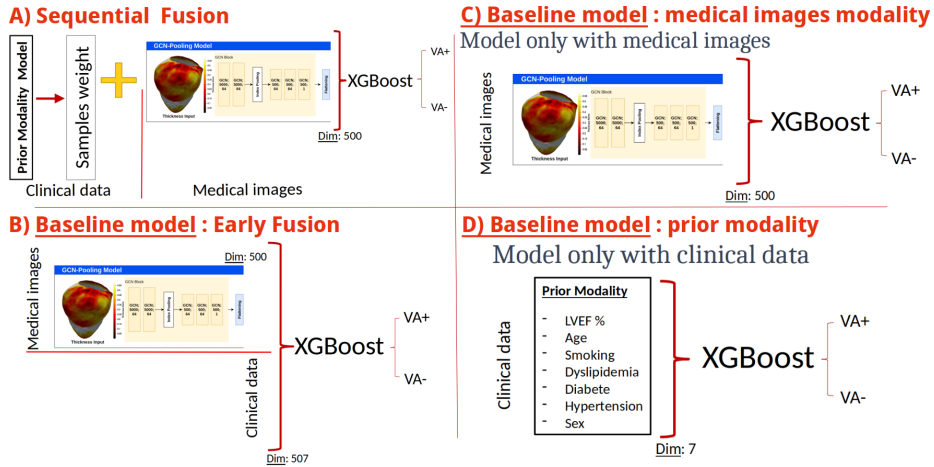


Fig. 2: Different model configurations compared.

The Early Fusion model (see fig. 2) is the concatenation of extracted thickness features with patients’ clinical data. The GCN-Pooling Model is used to extract thickness features. The data tensor is flattened and then given as input to a XGBoost classifier. The Early Fusion is the baseline model with both modalities.

The Sequential Fusion and Early Fusion models are compared with two others models used as a baseline with single modality. The first single modality baseline model is the prior model train with XGBoost classifier and only with clinical data (prior modality). The second single modality baseline model is the GCN-Pooling Model [4] where the “FCN block” is replaced by an XGBoost classifier; this model

is trained only with the medical images modality. We use the XGBoost model implemented in the `xgboost` python package [14] as the base classifier. Same hyper-parameters are used for the four models with the number of estimators set to 300 and the maximum depth is set to 6. The default values, as proposed by the [14] package are kept for the other hyper-parameters.

To isolate the impact of multimodal learning, we employed the XGBoost classifier across our model settings, acknowledging that the architecture of deep learning models (such as "FCN block") can vary depending on the input data. This approach ensures that any observed differences in performance are attributable to the multimodal learning techniques rather than variations in classifier architecture. We chose to utilise the TM rather than raw CT images for the VA classifier, as myocardial scar characterisation—derived from the TM—is a well-recognised substrate for VA [16] [17].

### 3.3 Controlled cross-validation setup

In the realm of machine learning, the quality and structure of the training data significantly impact the performance and generalisation of a machine learning model. In our context, in addition to keeping the same ratio VA- over VA+ in each split, we must also check that each split (training, validation and testing set) has the same data distribution as the whole data set (used as reference). We use the T-test (uni-variate hypothesis testing on each patients' clinical data) to only keep triple splits (training, validation and testing set) which have the same distribution as our reference. The  $p$ -value significance threshold for the T-tests was set to 0.05.

To compare the four models in our setting we have selected Sensitivity and F1 Score for performance metrics. To measure the variability of our performance metrics, 10 cross-validations are performed, then we compute the mean and standard deviation of each metric.

## 4 Results

As shown in Table. 2, the Sequential Fusion model outperforms the Early Fusion model in term of both metrics F1 Score (+10%) and Sensitivity (+15%). On the F1 Score, it performs 21% better than the prior modality model and 12% better than the Medical images modality model. Looking at the Sensitivity metric, the Sequential Fusion performs 18% better than the prior modality model and 18% better than the Medical images modality model.

Table 2: Model performance results. Best performance metrics in **bold**.

Model	Accuracy	F1 Score	Sensitivity	Specificity
(A) Sequential Fusion	<b>0.833</b> : $\pm 0.04$	<b>0.731</b> : $\pm 0.06$	<b>0.807</b> : $\pm 0.04$	<b>0.844</b> : $\pm 0.06$
(B) Early Fusion	0.790: $\pm 0.03$	0.631: $\pm 0.06$	0.650: $\pm 0.08$	<b>0.844</b> : $\pm 0.04$
(C) Medical Images Modality	0.773: $\pm 0.04$	0.607: $\pm 0.05$	0.629: $\pm 0.06$	0.828: $\pm 0.05$
(D) Prior modality	0.682: $\pm 0.04$	0.521: $\pm 0.06$	0.628: $\pm 0.10$	0.703: $\pm 0.05$

## 5 Discussion

The objective of modalities fusion is to get the best from each modality according to the purpose of interest. When we look at the details, with fig. 3-A, which shows samples well-classified by each modality (statistics on tests set), we can see that the Sequential Fusion is able to capture almost all knowledge from medical images modality. However, some samples well-classified by the prior modality are still misclassified by the Sequential Fusion. This point highlights the fact that there is still a room of progress to fuse modalities. When comparing the classification results of Early Fusion and Sequential Fusion on positive samples (VA+), we observe an interesting point: Sequential Fusion captures nearly all the correctly classified VA+ samples identified by Early Fusion, as shown in fig. 3-B. In addition, we can also observed that some samples are misclassified by all models. Those samples may required another modality to improve the Sequential Fusion model.

Fig. 4 gives an insight into the contribution of the medical images modality to the classification of VA+ samples, by comparing the decision tree of prior modality and Sequential Fusion. In fact, we can notice that the medical images modality has improved the classification of positive samples (VA+) with Left Ventricular (LV) Ejection Fraction greater than 46.5%. This point is in agreement with the fact that cardiac imaging techniques have allowed improved SCD risk stratification, especially in the group of samples with an LVEF > 35% [6].



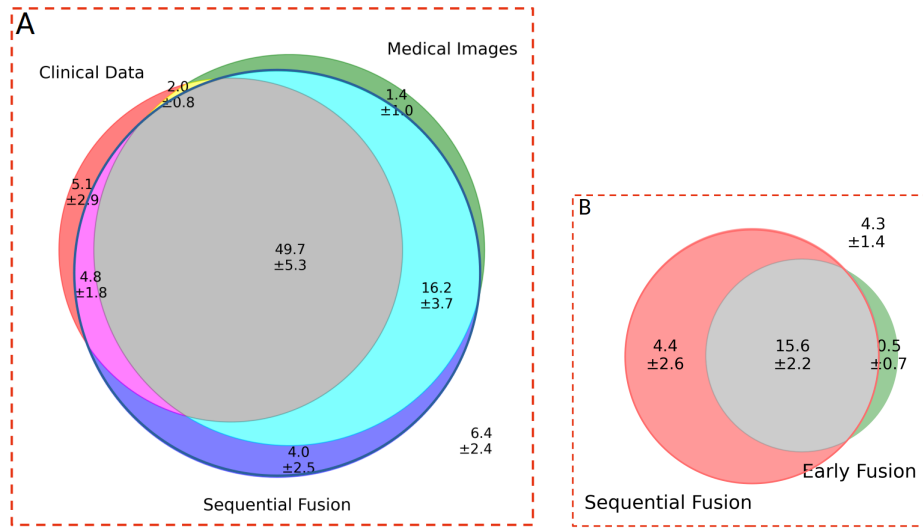


Fig. 3: **A**: Venn diagram comparing well-classified samples by each modalities and Sequential Fusion model. Values are means and standard deviations on tests set. The clinical data classifier is red, the medical images classifier is green, and the Sequential Fusion classifier is blue. Overlapping regions are samples well-classified by corresponding overlapping classifiers. Overlapping regions colour are the mix of corresponding colours red, green or blue, except the intersection of all models which is in gray. Out of circles is the statistic of misclassified samples by all models. **B**: Venn diagram comparing well-classified positive samples (VA+) by Early Fusion and Sequential Fusion models. Sequential Fusion classifier is in light red, Early Fusion in light green and the intersection in gray.

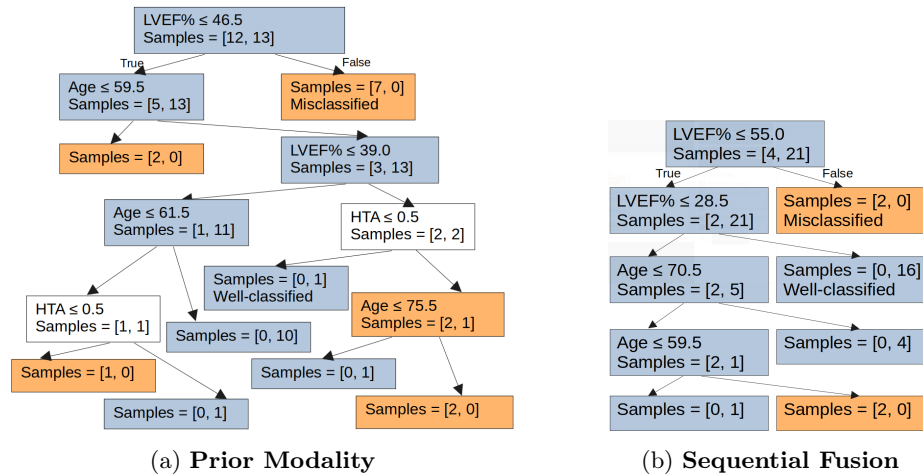


Fig. 4: Classification decision tree, of positive samples (VA+), on test set.

## 6 Conclusion

We have presented a novel technique to fuse two modalities by sequential learning on each modality. The main novelty of this technique is the weights computation based on knowledge acquired from the prior modality. Our Sequential Fusion outperforms the Early Fusion with the ability to get most information from both modalities. When we look at misclassified samples by the Sequential Fusion and misclassified samples by all modalities, there is potential for growth, on the one hand, for fusion techniques, and for the inclusion of additional modalities on the other hand. For future work on features fusion techniques, we believe that constraining progressively higher dimensional data based on lower dimensional data can improve the robustness of the classification model.

## 7 Acknowledgments.

This work has been supported by the French government through France 2030, the National Research Agency (ANR) Investments in the Future with 3IA Côte d’Azur (ANR-19-P3IA-000) and LIRYC (ANR-10-IAHU-04).

## References

1. Srinivasan N.T., Schilling R.J.: Sudden Cardiac Death and Arrhythmias. *Arrhythm Electrophysiol* **7**(2), 111-117 (2018)
2. Wilkins, E., Wilson, L., Wickramasinghe, K., Bhatnagar, P., Leal, J., Luengo-Fernandez, R., Burns, R., Rayner, M., Townsend, N.: *European Cardiovascular Disease Statistics*. European Heart Network (2017)
3. Kumaraswamy Nanthakumar, Andrew E. Epstein, G. Neal Kay, Vance J. Plumb, Douglas S. Lee: Prophylactic implantable cardioverter-defibrillator therapy in patients with left ventricular systolic dysfunction: A pooled analysis of 10 primary prevention trials. *Journal of the American College of Cardiology* **44**(11), 2166-2172 (2004)
4. Buntheng Ly, Sonny Finsterbach, Marta Nuñez-Garcia, Pierre Jais, Damien Garreau, Hubert Cochet, Maxime Sermesant: Interpretable Prediction of Post-Infarct Ventricular Arrhythmia Using Graph Convolutional Network. *Lecture Notes in Computer Science*, vol 13593. Springer, Cham. STACOM (2022)
5. Li, Y., Yang, M., Zhang, Z.: A survey of multi-view representation learning. *IEEE Trans Knowl Data Eng* **31**, 1863–83 (2019)
6. Van der Bijl, P., Bax, J.J.: Imaging for risk stratification of sudden cardiac death. *Herzschr Elektrophys* **33**, 261–267 (2022)
7. Russo, A. M., Stainback, R. F., Bailey, S. R., Epstein, A. E., Heidenreich, P. A., Jessup, M., Kapa, S., Kremers, M. S., Lindsay, B. D., Stevenson, L. W.: ACCF/HRS/AHA/ASE/HFSA/SCAI/SCCT/SCMR 2013 appropriate use criteria for implantable cardioverter-defibrillators and cardiac resynchronization therapy. *Journal of the American College of Cardiology*, **61**(12), 1318–1368 (2013)
8. Wellens, H. J., Schwartz, P. J., Lindemans, F. W., Buxton, A. E., Goldberger, J. J., Hohnloser, S. H., Huikuri, H. V., Kääb, S., La Rovere, M. T., Malik, M., Myerburg, R. J., Simoons, M. L., Swedberg, K., Tijssen, J., Voors, A. A., Wilde, A. A.: Risk

- stratification for sudden cardiac death: current status and challenges for the future. *European heart journal*, **35**(25), 1642–1651 (2014)
9. Nikolaos Dages, Gerhard Hindricks: Risk stratification after myocardial infarction: is left ventricular ejection fraction enough to prevent sudden cardiac death?. *European Heart Journal* **34**(26), 1964–1971 (2013)
  10. O’Mahony C, Jichi F, Pavlou M, Monserrat L, Anastasakis A, Rapezzi C, Biagini E, Gimeno JR, Limongelli G, McKenna WJ, Omar RZ, Elliott PM: A novel clinical risk prediction model for sudden cardiac death in hypertrophic cardiomyopathy (HCM Risk-SCD). *European Heart J* 2013, (2014)
  11. D. Lai, Y. Zhang, X. Zhang, Y. Su, M. B. Bin Heyat: An Automated Strategy for Early Risk Identification of Sudden Cardiac Death by Using Machine Learning Approach on Measurable Arrhythmic Risk Markers. *IEEE Access* 7, 94701-94716 (2019)
  12. Waks, J. W., Sitlani, C. M., Soliman, E. Z., Kabir, M., Ghafoori, E., Biggs, M. L., Henrikson, C. A., Sotoodehnia, N., Biering-Sørensen, T., Agarwal, S. K., Siscovick, D. S., Post, W. S., Solomon, S. D., Buxton, A. E., Josephson, M. E., Tereshchenko, L. G.: Global Electric Heterogeneity Risk Score for Prediction of Sudden Cardiac Death in the General Population: The Atherosclerosis Risk in Communities (ARIC) and Cardiovascular Health (CHS) Studies. *Circulation* **133**(23), 2222–2234 (2016)
  13. Norrish, G., Ding, T., Field, E., Ziolkowska, L., Olivotto, I., Limongelli, G., Anastasakis, A., Weintraub, R., Biagini, E., Ragni, L., Prendiville, T., Duignan, S., McLeod, K., Ilina, M., Fernández, A., Bökenkamp, R., Baban, A., Kubuš, P., Daubeney, P. E. F., Sarquella-Brugada, G., ... Kaski, J. P.: Development of a Novel Risk Prediction Model for Sudden Cardiac Death in Childhood Hypertrophic Cardiomyopathy (HCM Risk-Kids). *JAMA cardiology* **4**(9), 918–927 (2019)
  14. Tianqi Chen and Carlos Guestrin: XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD ’16)*, 785–794 (2016)
  15. William C. Sleeman, Rishabh Kapoor, Preetam Ghosh: Multimodal Classification: Current Landscape, Taxonomy and Future Directions. *ACM Computing Surveys* **55**(7), 1-31 (2022)
  16. Noordman, A. B. P., Maass, A. H., Groenveld, H., Mulder, B. A., Rienstra, M., Blaauw, Y.: Myocardial Scar Characterization and Future Ventricular Arrhythmia in Patients With Ischemic Cardiomyopathy and an Implantable Cardioverter-Defibrillator. *Frontiers in cardiovascular medicine* **8**, (2021)
  17. Pandozi, C., Mariani, M. V., Chimenti, C., Maestrini, V., Filomena, D., Magnocavallo, M., Straito, M., Piro, A., Russo, M., Galeazzi, M., Ficili, S., Colivicchi, F., Severino, P., Mancone, M., Fedele, F., Lavalle, C.: The scar: the wind in the perfect storm-insights into the mysterious living tissue originating ventricular arrhythmias. *Journal of interventional cardiac electrophysiology* **66**(1), 27–38 (2023)