Segmentation of Atrial Electrical Activity in Intracardiac Electrograms (IECGs) Using Convolutional Neural Network (CNN) Trained on Small Imbalanced Dataset

Jakub Hejc[1,2], David Pospisil[3], Petra Novotna[4], Martin Pesl[1,5,6], Oto Janousek [4], Marina Ronzhina[4], Zdenek Starek[1,5]

[1] ICRC, St. Anne's University Hospital, Brno, Czech Republic
[2] Department of Pediatric, Children's Hospital, University Hospital Brno, Brno, Czech Republic,
[3] Department of Internal Medicine and Cardiology, University Hospital Brno and Faculty of Medicine of Masaryk University, Brno, Czech Republic,
[4] Department of Biomedical Engineering, Brno University of Technology, Brno, Czech Republic ,
[5] 1st Department of Internal Medicine, Cardio-Angiology, Faculty of Medicine, Masaryk University, Brno, Czech Republic
[6] Department of Biology, Faculty of Medicine, Masaryk University, Brno, Czech Republic

Abstract

Timing pattern of intracardiac atrial activity recorded by multipolar catheter in the coronary sinus (CS) provides insightful information about the type and approximate origin of common non-complex arrhythmias. Depending on the anatomy of the CS, the atrial activity can be substantially disturbed by ventricular far field complex preventing accurate segmentation by convential methods. In this paper, we present small clinically validated database of 326 surface 12-lead and intracardiac electrograms (ECG and IEGs) and a simple deep learning framework for semantic beat-to-beat segmentation of atrial activity in CS recordings. The model is based on a residual convolutional neural network (CNN) combined with pyramidal upsampling decoder. It is capable to recognize well between atrial and ventricular signals recorded by decapolar CS catheter in multiple arrhythmic scenarios reaching dice score of 0.875 on evaluation dataset. To address a dataset size and imbalance issues, we have adopted several preprocessing and learning techniques with adequate evaluation of its impact on the model performance.

1. Introduction

Analysis of the intracardiac atrial activity (AA) provides detailed insight into conduction patterns of various arrhythmias. Interval timings and regularity of the AA belong amongst the most decriptive factors in the differential diagnostics and mapping of non-complex tachycardias performed in an electrophysiology (EP) lab. It has been also used to estimate dynamics patterns during atrial fibrillation (AF) as a way for understanding its pathophysiology [1]. While the latter approach needs to be applicable primarily to signals recorded by mapping or ablation catheters, the differential diagnostics of "convential" arrhythmias is usually done with a multipolar diagnostic catheters placed in prespecified positions such as His bundle and coronary sinus (CS). In both cases, catheters lie in close proximity to the ventricles resulting in measurement of ventricular far field (VFF) activity, which can superimpose AA. The magnitude of the interference depends on the anatomical placement of both the CS and the catheter. In some cases it can make the distinction between AA and VFF difficult for common methods based on the 1st derivative, thresholding or wavelet transform [2]. Our main goal is to provide simple deep neural network framework for beat-to-beat segmentation of the AA in CS recordings followed by a database of arrhythmic ECG signals with annotated electrical activation of the left atria in the future. We are aware of the limitations of the framework and will discuss it in later sections. Still, such a tool can substantially help with retrospective analysis of AA patterns in long-term recordings obtained within an interventional EP examination.

2. Material and Methods

2.1. Data

The data comes from 100 consecutive pre-adults indicated for an interventional EP procedure within the Children's Hospital, University Hospital Brno, Brno, Czech

Republic. Of them, we have selected 326 short-term segments with clinically manifested or co-present arrhythmic events validated and confirmed by an electrophysiology professional. The dataset consists of 12-lead surface ECGs and 5-lead IEGs recorded by St. Jude WorkMate 4.3 EP system (2000 Hz, 72 nV/LSB). IEGs were recorded via decapolar diagnostic catheter placed in the CS, and were used for manual segmentantion of the onset and offset of present atrial beats. Clinical characteristics of the database is provided in Table 1.

Table 1: Baseline clinical characteristics of the database. Values are presented as frequency or median (interquartile range).

Parameter	$n = 100$
Age [years]	14.0 (12.0–17.0)
Sex: females	48
The length of the strip [s]	8.5 (6.4–12.2)
Number of strips	326
Sinus rhythm	191
Atrial premature beat	47
Ventricular premature beat	40
Atrial fibrillation	33
Atrial flutter	5
AV node re-entry tachycardia	45
AV re-entry tachycardia	40
Junctional rhythm	22
Right bundle branch block	7
Left bundle branch block	9
Atrial stimulation	19
Ventricular stimulation	35
Focal atrial tachycardia	8
Ventricular pre-excitation	58
1st degree AV block	3

Class imbalance here is of two distinct origins. One can be attributed to various arrhythmia types leading to skewed distribution of easy-, moderate- or hard-to-segment atrial beats. Second one is given by usually short duration of the atrial beat (on average 7.62 % in our dataset) with respect to the total length of the signal.

2.2. Data preprocessing

Built-in hardware high-pass (cut-off 0.5 Hz) and notch filter (50 Hz) were active during data acquisition. Recorded signals were undersampled with sampling frequency of 500 Hz using decimation and anti-aliasing FIR filter. Before passing into the model, the signal units were normalized to 1 mV. Train-validation split with 8:2 ratio was performed by a stratified greedy-based sampling strategy [3] in order to get a non-overlapping subsets of patients with approximately the same distribution of clinical arrhythmias.

To overcome overfitting problems, that may occur with small-sized dataset, we have adopted few additional preprocessing techniques applied during the training. The augmentation pipeline consisted of Gaussian noise addition (randomly 0–30 % of signal RMS); voltage scaling (randomly ±50 %); temporal scaling (randomly ±30 %); and temporal shifting (randomly ±40 %). The effect of individual augmentations on the performance has been evaluated by an ablation study with the best performing model.

2.3. Model Architecture

We have adopted a proven concept of CNN encoder based on residual network (ResNet) with pre-activation [4] and Group Normalization [5] as regularization technique independent of a batch size. As the Group Normalization introduces another hyperparameter (number of groups), we have performed brief grid search study to evaluate the effect of this variable on the model performance. The encoder itself consists of only 3-layer layout with 2 residual blocks and 1D convolutional filters in every layer, leaving the amount of model parameters substantially small. The decoder is made by Joint Upsampling Module proposed by Huikai et al. [6] in order to reduce computation complexity while preserving decoder receptive field. The input of the encoder is made of 5-lead IEGs (or single lead if applicable) arranged as multichannel 1D tensor. The output of each ResNet layer is passed into the decoder as an individual input. The output of the model is one-hot encoded atrial activity sequence. A schema of the model is depicted in Figure 1.

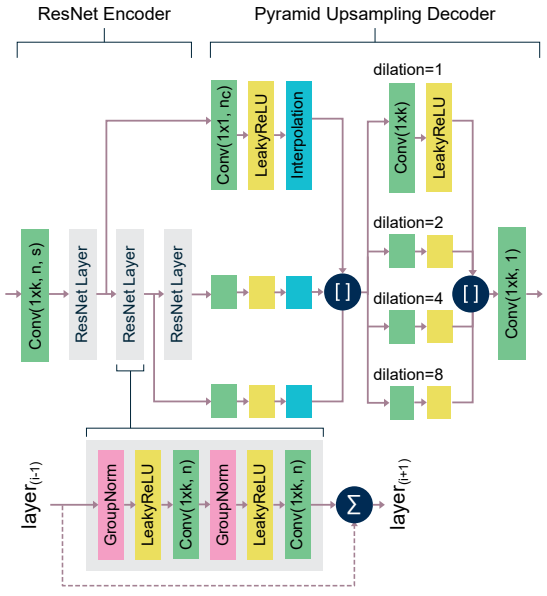

Figure 1: The architecture of the CNN model. k, n, s – kernel size, number of filters, and stride of convolutional layers.

2.4. Training setup

Dice loss, or its generalized variant [7] in the case of class imbalance, are commonly used as a loss functions for segmentation tasks [8]. For better control over how the loss behaves in case of rare or hard to segment examples, Focal Tversky Loss [9] based on Tversky Index can be introduced instead. For binary segmentation and continuous output variable, it can be defined as:

$$\mathbf{TL} = \left\{ \mathbf{1} - \frac{\hat{\mathbf{Y}} \odot \mathbf{T}}{\hat{\mathbf{Y}} \odot \mathbf{T} + \alpha[(\mathbf{1} - \hat{\mathbf{Y}}) \odot \mathbf{T}] + \beta[\hat{\mathbf{Y}} \odot (\mathbf{1} - \mathbf{T})]} \right\}^{\gamma}, \quad (1)$$

where $\odot$ denotes point-wise multiplication; $\hat{\mathbf{Y}}$ and $\mathbf{T}$ are tensors representing model output scores and ground truth binary targets, respectively; α and β are weighting factors of Tversky index and γ controls non-linear behaviour of Tversky loss. For $\alpha = \beta = 0.5$ and $\gamma = 1$ the loss simplifies to Dice loss. When $\gamma > 1$, gradient of loss function becomes steeper for inaccurately classified samples, which can result in better optimization in case of class imbalance. To assess this hypothesis, a grid search has been performed to determine the best performative settings of the loss along with our model.

Weights and biases of convolutional layers were initialized with Kaiming [10] and constant ($c = 0$) initialization. The latter method was also used for normalization layers with $c = 1$ and $c = 0$ for weights and biases, respectively. Model was trained with Adam optimizer with $\beta = \{0.9, 0.999\}$ and decoupled weight regularization [11] decay $\lambda = 10^{-6}$. Optimal initial learning rate α_0 was found by exponential warm-up search. Schedule strategy was based on reducing learning rate on plateau with threshold of 10^{-2} and decaying factor $10^{-1}\alpha_0$. Scheduler control variable was F_1 score computed on evaluation subset. Mini-batch size was set to 16 as an optimum estimate given by a coarse grid search.

3. Results and discussion

The model performance was evaluated by F_1 score or positive predictive value (P^+) and sensitivity (Se) when appropriate. Overall results for the most successful model are listed in Table 2.

Table 2: Model performance on 5-lead and average-lead CS recordings.

Model	Train F_1	Val. F_1
Best model, 5-lead CS	**0.888**	**0.875**
Best model, single-lead CS (avg.)	0.818	0.801

F_1 score of 0.875 on the validation subset was reached by a model with small filter kernels ($k = 3$), exponentially increasing number of groups in normalization layers, and combined with conventional Dice loss. Since we were limited by globally stated reference points, single-lead performance was evaluated using an average of all CS leads only. A drop in F_1 score here might have been, to some extent, expected considering what happens with e.g. highly disorganized fibrillatory waves after averaging.

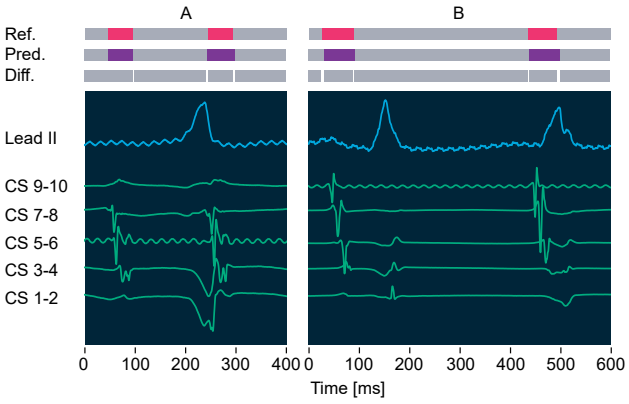

Figure 2: Examples of the results of atrial signal segmentation on recordings containing mixture of atrial and ventricular activity. Ref. – manual references; Pred. – model output; Diff. – difference between reference and prediction; A – atrial flutter; B – normal sinus beat followed by premature atrial beat.

An impact of loss function parameters on the performance is listed in Table 3. Although parameters α and β are able to modulate P^+ and Se according to one needs, it is without substantial benefit in segmentation of "hard" samples. A gain in one metric never exceeded expanses of the second one for each tested pair of parameters. Neither modulation of γ led to a better performance compared to common Dice loss.

Table 3: Model performance based on the hyper-parameters of the Focal Tversky Loss α, β and γ.

α	β	γ	Tr. F_1	Val. F_1	Val. P^+	Val. Se
0.1	0.9	1	0.749	0.745	0.633	0.615
0.3	0.7	1	0.798	0.783	0.902	0.692
0.5	**0.5**	**1**	**0.888**	**0.875**	**0.878**	**0.872**
0.7	0.3	1	0.871	0.859	0.804	0.922
0.9	0.1	1	0.833	0.780	0.675	0.971
0.7	0.3	2	0.813	0.818	0.827	0.809
0.5	0.5	2	0.869	0.844	0.824	0.866
0.3	0.7	2	0.739	0.733	0.896	0.620
0.5	0.5	0.5	0.733	0.738	0.861	0.657

An ablation study (Table 4) shows the highest decrease in F_1 score (by 0.026) after switching off the scaling of temporal axis which has probably the highest impact on the AA morphology and is able to mimic slow conduction zones and far field signals. Interestingly, even without any augmentation the model was well regularized and did not tend to overfit too much.

Table 4: Ablation study of augmentation techniques used for training; **w/o = without particular augmentation**.

Augmentation type	Train F_1	Val. F_1
Gaussian noise w/o	0.868	0.854
Voltage random scaling w/o	0.862	0.852
Temporal random scaling w/o	**0.831**	**0.829**
Temporal shift w/o	0.827	0.803

As can be seen in Figure 3, Group normalization can be a source of performance drop if not set properly. Either too few or too many groups did more harm than good. This has been also reported by others in deep-learning community. The most performative setup (0.875) was given by exponentially increasing number of groups according to number of filters in the residual block. Interestingly, the model worked best with the smallest filter size ($k = 3$) possible. This may however change after fine tuning the remaining hyperparameters of the model.

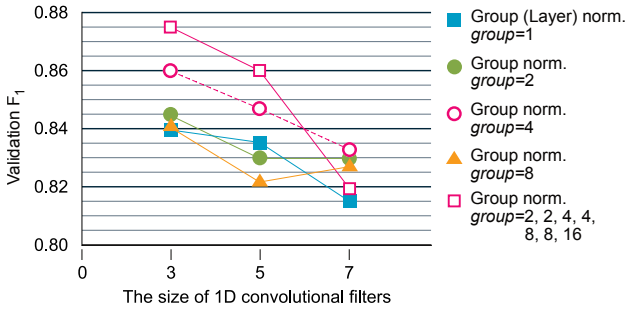

Figure 3: The relationship between convolution kernel size and number of groups in Group normalization. Models depicted by magenta outline squares have variable group size growing exponentially with number of filters in corresponding layer.

3.1. Limitations

As training samples originate from the CS recordings the model may suffer a generalization drop when used with signals recorded within different heart cavities and structures. Another limitation is given by a fixed number of leads required by the model. Also, the model is not capable to recognize ventricular activity, which is essential for the interval measurement during the EP study. All the mentioned limitations will be dealt with in our future work.

4. Conclusion

Presented CNN architecture provides reliable segmentation of atrial beats in signals from coronary sinus and will be used by our team for non-trivial segmentation tasks. We also believe that provided study brings to bear some practical considerations on handling small imbalanced dataset.

5. Acknowledgements

Supported by Ministry of Health, Czech Republic - conceptual development of research organization (FNBr, 65269705) and SV/MUNI/A/1437/2020.

References

[1] Walters TE, Teh AW, Spence S, et al. Relationship Between the Electrocardiographic Atrial Fibrillation Cycle Length and Left Atrial Remodeling: A Detailed Electroanatomic Mapping Study. Heart Rhythm 2014;11(4):670–676. ISSN 1547-5271.

[2] Sayadi O, Puppala D, Ishaque N, et al. A Novel Method to Capture the Onset of Dynamic Electrocardiographic Ischemic Changes and its Implications to Arrhythmia Susceptibility. Journal of the American Heart Association 2014; 3(5).

[3] Sechidis K, Tsoumakas G, Vlahavas I. On the stratification of multi-label data. Machine Learning and Knowledge Discovery in Databases 2011;145–158.

[4] He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. In Leibe B, Matas J, Sebe N, Welling M (eds.), Computer Vision – ECCV 2016. Cham: Springer International Publishing. ISBN 978-3-319-46493-0, 2016; 630–645.

[5] Wu Y, He K. Group Normalization. International Journal of Computer Vision 2019;128(3):742–755.

[6] Chen L, Papandreou G, Kokkinos I, et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Transactions on Pattern Analysis Machine Intelligence 2018;40(04):834–848. ISSN 1939-3539.

[7] Sudre CH, Li W, Vercauteren T, et al. Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer, 2017; 240–248.

[8] Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep Learning, volume 1. MIT press Cambridge, 2016.

[9] Salehi SSM, Erdogmus D, Gholipour A. Tversky Loss Function for Image Segmentation Using 3D Fully Convolutional Deep Networks. CoRR 2017;abs/1706.05721.

[10] He K, Zhang X, Ren S, Sun J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In 2015 IEEE International Conference on Computer Vision (ICCV). 2015; 1026–1034.

[11] Loshchilov I, Hutter F. Decoupled Weight Decay Regularization. In International Conference on Learning Representations. 2019; .

Address for correspondence:

Jakub Hejc
ICRC – St. Anne's University Hospital
Brno, Czech Republic
jakub.hejc@fnusa.cz