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[2], 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, 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 and intracardiac electrocardiograms (ECG and IECG) 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 label decoder. It is capable to recognize well between of 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 the electrophysiology 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 (VFA) 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 VFA difficult for common methods based on the 1st derivative, thresholding or wavelet transform. Our main goal is to provide simple deep neural network framework for beat-to-beat automatic 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 helps with retrospective analysis of AA patterns in long-term recordings obtained by interventional EP examination.

2. Material and Methods

2.1. Data

Dataset consist of 12-lead surface ECG and 5-lead IECGs recorded from decapolar diagnostic catheter placed in the CS. Data were recorder by St. Jude EP Workmate 4.3 system (2000 Hz, 72 μV/LSB) during indicated intervential EP procedure. Raw ECG and IECG signals were

extracted from 100 consecutive patients followed by the bipolar lead montage of IECGs. All of those recordings underwent clinical evaluation by an EP professional targeting sinus rhythm, stimulated rhythms, and either clinically manifested or co-present arrhythmic events. Regions of interest were manually segmented and exported as a single strip in which the beginning and the end of all present AAs in the CS leads were manually delineated. 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 in the dataset is of two distinct origins. The first is given by a limited duration of AA compared to a resting phase giving only 7.62 % of positive samples. The second is caused by extremely imbalanced distribution of arrhythmia types, in consequence, skewed distribution of easy, moderate and hard to classify AA segments.

2.2. Data preprocessing

EP recording system had a high-pass filter with cut-off frequency of 0.5 Hz and 50 Hz notch filter active during recording. Offline data were undersampled with sampling frequency of 500 Hz using decimation and anti-aliasing FIR filter. Before passing into the model, the standard deviation of IECG signals were normalized to 1 mV and manual annotations were encoded to one-hot sequences. Train-validation split into 8:2 ratio was performed by a stratified greedy-based sampling strategy [2] in order to get a non-overlapping subsets of patients with approximately the same distribution of clinical arrhythmias.

2.3. Model Architecture

To address aforementioned task, we have adopted proven concept of CNN encoder based on residual network (ResNet) with preactivation [3] and grouped layer normalization [4] as a regularization technique independent of a batch size. As a layer normalization introduces another hyperparameter (number of groups) we have performed small 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. [5] in order to reduce computation complexity and memory consumption caused by commonly used dilated convolutions. 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 label sequence. Detailed schema of the model is depicted in the Figure 2.

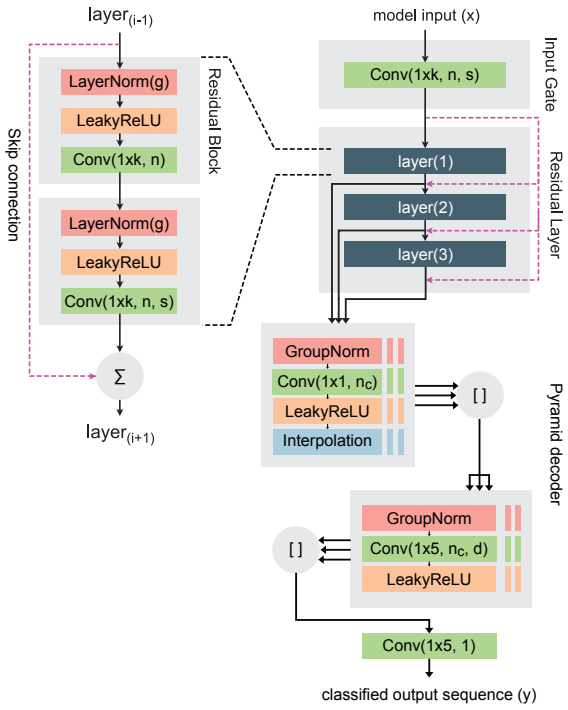

Figure 1: The architecture of the CNN model. k, n, s – filter size, number of filters and stride of the ResNet encoder; n_c – number of convolutional filters in the pyramid decoder layer with dilatation factor d; g – number of groups of the layer normalization.

2.4. Augmentation pipeline

To overcome overfitting problems occurring in datasets of small size, we have adopted several more preprocessing techniques applied during training phase. The augmentation pipline consisted of Gaussian noise addition (randomly 0–30 % of signal RMS); voltage scaling (randomly ± 50 %); temporal scaling (randomly ± 30 %); temporal shifting (randomly ± 40 %). The effect of individual transformations on the model performance has been evaluated by an ablation study using the best performing model.

2.5. Loss Function

Dice loss (DS) or its generalized variant [6] in the case of class imbalance are commonly used as a loss functions for segmentation tasks [7]. For better control over how the loss behaves in case of rare or hard to segment examples Focal Tversky Loss [8] based on Tversky index can be introduce instead. For a binary segmentation and continuous output variable it can be defined as:

$$TL = \left(1 - \frac{y_n t_n}{y_n t_n + \alpha[(1-y_n)t_n] + \beta[y_n(1-t_n)]}\right)^{\gamma}, \tag{1}$$

where y_n and $t_n \in \{0, 1\}$ are elements of model output and target sequence, respectively, at sample n; α and β are weighting factors of the Tversky index and γ controls non-linearity of the Tversky loss.

For $\alpha = \beta = 0.5$ and $\gamma = 1$ the loss simplifies to a dice loss. When $\gamma > 1$ the TL gradient becomes steeper for inaccurately classified samples which can result in better optimization in case of class imbalance. A simple grid search was performed to find the best performative settings of TL parameters for our model.

2.6. Training setup

Weights and biases of convolutional layers were initialized with Kaiming [9] and constant ($c = 0$) initialization. The latter method was also used for the normalization layers with $c = 1$ and $c = 0$ for weights and biases, respectively. Model was trained with Adam optimizer [10] with $\beta = \{0.9, 0.999\}$ and decoupled weight regularization decay $\lambda = 10^{-6}$ [11]. Optimal initial learning rate α_0 of 0.01 was found by exponential LR warm-up phase. LR schedule strategy was based on the reduce on plateau method with threshold of 0.01 and decaying factor $10^{-1}\alpha_0$. Scheduler control variable was the dice score computed on the evaluation data subset each epoch. 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 dice score (DS), precision or recall. The results for training and validation sets are marked by subsripts T and V.

Table 2: The results of the best performing model on the training end validation subset using all five CS leads independently and as a single averaged CS lead.

Model	DS_T	DS_V
Best model, 5-lead CS	0.888	0.875
Best model, single-lead CS (avg)	0.818	0.801

The impact of loss function parameters on the performance is listed in the Table 3. Although parameters α and β are able to modulate precision and recall according to one needs, it is without substantial benefit in segmentation of "hard" samples. A profit in one metric is lower then a drop is the second one for each tested pair of parameters. Neither modulation of γ led to better performance than with dice loss.

Table 3: The results of the best performing model depending on the hyperparameters of the Focal Tversky Loss $\alpha, \beta and \gamma$.

α	β	γ	DS_T	DS_V	Precision	Recall
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	0.0.871	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

The ablation study (Table 4) shows the highest decrease in DS metric (by 0.023) 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 is well regularized and doesn't tend to overfit.

Table 4: The results of the augmentation techniques ablation study for the best performing model.

Augmentation type	DS_T	DS_V
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

One of the major improvement in DS was reached by appropriate settings of the number of groups in normalization layers (Figure 3). The most performative setup was given by exponentially increasing number of groups according to

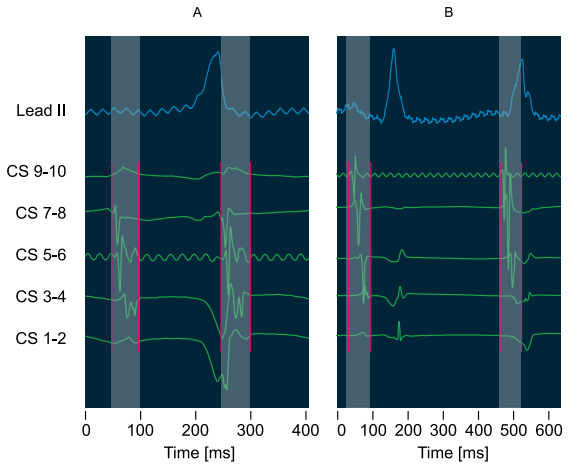

Figure 2: Examples the results of atrial signal segmentation on signals containing mixture of atrial and ventricular activity. Segmentation borders are marked by magenta lines, manual references are depicted by gray rectangles. A – recording with atrial flutter; B – recording with normal sinus beat followed by premature atrial beat.

number of filters in the residual block. If set to a constant number in all blocks, either too few or too many groups led to significant drop in performance.

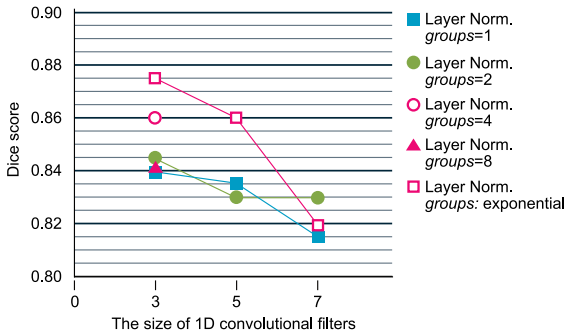

Figure 3: The relationship between convolutional filter size and number of groups of the layer normalization. Layer normalization with a single group only is identical to a group normalization. Empty rectangles have variable group size growing exponentially with number of filters, specifically: 2, 2, 2, 4, 4, 8, 8, 16

3.1. Limitations

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

4. Conclusions

Presented CNN architecture provides fast and reliable segmentation of the AA and will be used by our team for a non-trivial segmentation tasks in the future. We also believe that provided methodology addressing common machine learning issues such as class imbalance and generalization on small training samples makes a contribution to knowledge.

References

[1] Walters TE, Teh AW, Spence S, Kistler PM, Morton JB, Kalman JM. 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. URL https://www.sciencedirect.com/science/article/pii/S1547527113014823.

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

[3] He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks, 2016.

[4] Ba JL, Kiros JR, Hinton GE. Layer normalization, 2016.

[5] Wu H, Zhang J, Huang K, Liang K, Yu Y. Fastfcn: Rethinking dilated convolution in the backbone for semantic segmentation. CoRR 2019;abs/1903.11816. URL http://arxiv.org/abs/1903.11816.

[6] Sudre CH, Li W, Vercauteren T, Ourselin S, Cardoso MJ. 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.

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

[8] Salehi SSM, Erdogmus D, Gholipour A. Tversky loss function for image segmentation using 3d fully convolutional deep networks. CoRR 2017;abs/1706.05721. URL http://arxiv.org/abs/1706.05721.

[9] He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. CoRR 2015;abs/1502.01852. URL http://arxiv.org/abs/1502.01852.

[10] Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. arXiv preprint arXiv14126980 2014;.

[11] Loshchilov I, Hutter F. Decoupled Weight Decay Regularization. arXiv preprint arXiv171105101 2017;.

Address for correspondence:

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