Transfer Learning for Improved Classification of Drivers in Atrial Fibrillation

Bram Hunt[1,2,3], Eugene Kwan[1,2,3], Tolga Tasdizen[4,5], Jake Bergquist[2,3,4], Matthias Lange[2,3,4], Benjamin Orkild[1,2,3], Robert S. MacLeod[1,2,4], Derek J. Dosdall[1,2,3,6], Ravi Ranjan[1,2,3]

[1]Department of Biomedical Engineering, University of Utah, SLC, UT, USA
[2]Nora Eccles Cardiovascular Research and Training Institute, University of Utah, SLC, UT, USA
[3]Division of Cardiovascular Medicine, University of Utah, SLC, UT, USA
[4]Scientific Computing and Imaging Institute, University of Utah, SLC, UT, USA
[5]Department of Electrical and Computer Engineering, University of Utah, SLC, UT, USA
[6]Division of Cardiothoracic Surgery, Department of Surgery, University of Utah, SLC, UT, USA

Abstract

Rotors and focal ectopies, known as "drivers", are potential mechanisms underlying persistent atrial fibrillation. Machine learning algorithms have been used to identify these drivers, but the small size of current driver datasets limits their performance. We hypothesized that pretraining with unsupervised learning on a large dataset of unlabeled electrograms would improve classifier accuracy on a smaller driver dataset. In this study, we used a SimCLR-based framework to pretrain an 18-layer residual neural network on a dataset of 113,075 unlabeled electrograms and found that the weighted testing accuracy significantly improved over a non-pretrained network (78.6±3.9% vs 71.9±3.3%, p=0.047). In addition to laying ground for development of superior driver detection algorithms, this finding supports broader application of transfer learning to deep learning applications for atrial fibrillation.

1. Introduction

Atrial fibrillation (AF) is the most common electrical disease of the heart, affecting over 50 million patients world-wide [1]. Pulmonary vein isolation, where the pulmonary veins are isolated from the left atrium substrate via ablation, is a standard treatment for AF, but is still associated with high recurrence with persistent atrial fibrillation (persAF) [2]. Non-pulmonary vein mechanisms in the form of focal ectopies and reentrant rotors, termed "drivers", have been proposed as reason for this recurrence [3]. As such, the automated identification of drivers in auxiliary treatments for persAF has developed recent interest. Machine learning algorithms, in particular, have seen recent success in driver identification and several algorithms are now available for this task with inputs from endocardial mapping, optical mapping, and ECGI [4]–[6].

However, despite broad utilization in machine learning literature, transfer learning remains unexplored as a means of improving the accuracies of driver detection algorithms. In transfer learning, a model is trained on a precursory task to establish a set of initialized parameters [7]. This pretraining can improve accuracy and reduce training time when the network is trained on a subsequent task. Many pretraining tasks are available for either labelled or unlabelled data, including denoising, image inpainting, and classification of readily available markers (e.g., age or sex).

Our laboratory has built an extensive dataset of endocardial voltage measurements with ultra-high density mapping catheters in a canine model. In this paper, we evaluate the hypothesis that contrastive pretraining will improve the accuracy of a deep learning model in identifying driving mechanisms from endocardial electrograms. Successful improvement in driver detection accuracy would be impetus for broader application or pretraining in driver detection as well as for non-driver related tasks.

2. Methods

For all studies, we adhered to the Guide for the Care and Use of Laboratory Animals. The Institutional Animal Care and Use Committee at the University of Utah approved the protocol.

Mapping Studies and Classification of Driving Sites. A paced canine model of persAF was used as described previously (10 female, 3 male, mongrel purpose-bred hound, 27-35 kg, 1-2 yrs.) [8]. In brief, we performed electrophysiological studies on the canines at 1, 3, and 6 months during periods of sustained, non-paced AF. In these studies, we captured endocardial electrograms with the Rhythmia mapping system and 64-electrode Orion basket catheter (Boston Scientific) at a sampling frequency of 954 Hz. For each study performed, we manually interrogated a subset of the highest frequency atrial AF electrograms at each major anatomical site to find

rotational and focal mechanisms.

Network Design. We split the 64-electrode electrograms from all studies into samples of 2 seconds each. We then transformed these electrograms into stacked images of 8×8 with 1907 channels encoding the time dimension. We used an 18-layer 3D ResNet as the base network for our model into which these electrograms could be fed.

Contrastive Learning. We pretrained our neural network with the SimCLR contrastive learning process [9]. In brief, we remove the final linear layer of the 3D ResNet and replace that layer with a linear layer feeding into a multilayer perceptron termed the projection head. This projection head utilized 1000x1000 interlayer linear multiplications, a single hidden layer, a ReLU function after the input layer, and 128 output features. Then, we train this altered network to encode images derived from the same original images (e.g., two crops of a base image, or a blurred image and its unblurred original) to the same latent space. Likewise, images derived from different original images are encoded to distant locations in the latent space. Key to this task are the random augmentations used: these augmentations are used to transform given original images into daughter images to be compared to one another. Augmentations are also chosen such that semantic meaning is retained after transformation, akin to rotations of the same image.

We used NT-Xent loss, a LARS optimizer with learning rate of 4.8, weight decay of 10^{-6}, and batch size of 4096. We trained for 200 epochs with a linear warmup for the first 10 epochs followed by a cosine decay schedule without restarts, terminating training upon plateau in loss reduction. However, given differences in shape and structure between 2D images and our 3D catheter electrograms, we use random cropping, random differentiation in the time dimension, and random blurring as our augmentation processes. Differentiation and blurring were performed with 50% chances of occurrence. For cropping, we reduced sample size by up to 75% with random aspect ratios ±25% of original. These crops were stretched and bilinearly interpolated to original image size. These transformations were performed to create pairs of augmented EGMs from a single original EGM in the contrastive process.

Classifier Training. After contrastive pretraining, we began fine-tuning our network on the classification task. Here, we detach the projection head and replace it with a linear layer with a binary output as seen in Figure 1. We utilized a 70-10-20 training-validation-testing split, performing hyperparameter sweeps with the training dataset and evaluating performance on the validation dataset. After completion of sweeping with both pretrained and non-pretraining networks, we identified networks with the lowest validation loss and evaluated their performance on the testing dataset.
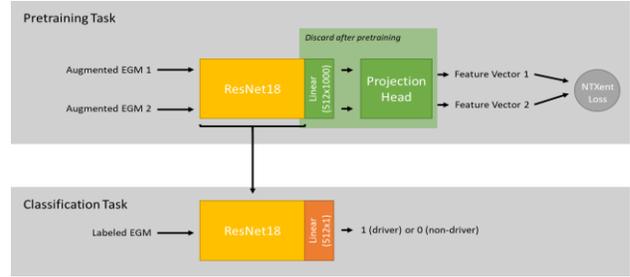

Figure 1. Diagram of model training paradigm. The deep neural network is first trained on the SimCLR task with the unlabeled dataset of AF electrograms. After completion of the pretraining task, the projection head is detached and replaced with a single linear layer leading to the binary output. The network is then trained on the driver classification process with the labeled electrogram dataset.

Interpretability Analysis. Interpretation analysis is rapidly becoming an essential prerequisite to clinical acceptance of deep neural networks due to the opaque rationale underlying network performance. As such, we used Grad-CAM visualizations of our model gradients and activations to inspect model decision making in testing results [10]. We used the 4th layer of the 3D ResNet for this visualization. In brief, Grad-CAM computes the elementwise changes in classification confidence via multiplication of the gradient and activations of an input image. These changes can be projected over the original image to highlight which regions were responsible for the greatest contributions to final class decision by the network.

Statistics. We used weighted accuracy for our classification to eliminate any effects of class imbalance. To compare performance between pretrained and non-pretrained networks, we bootstrapped the testing dataset to acquire distributions of testing accuracy. We compared these distributions with paired t-tests.

3. Results

In total, we obtained 113,572 64-electrode electrograms of 2-seconds each. From this, we examined 502 electrograms and classified them as contained a driver or not a driver. From our 502 samples used for driver identification, we manually found 272 non-drivers, 172 rotors, and 57 focal ectopies. Specific times where drivers began and terminated were noted. The remaining, unlabelled electrograms were used in the contrastive pretraining process. We reached a minimum in contrastive loss at 4e-4 after 184 epochs.

Classification. We obtained testing accuracies of 71.9% and 78.6% with the non-pretrained and pretrained networks respectively. To compare accuracies, we bootstrapped the testing dataset 10,000 times with replacement. From this, we found our pretrained network

to significantly increase in accuracy over the non-pretrained network (p=0.047). Our final classification training routine used a batch size of 250, 40 epochs, a learning rate of 5e-5, and the Adam optimizer.

Table 1. Training, validation, and testing results of the pretrained and non-pretrained network.

Network	Dataset	Loss	Weighted Accuracy	F1
Non-Pretrained	Training	0.0023	100.0%	1.00
	Validation	0.58	79.2%	0.79
	Testing	**0.70**	**71.9%**	**0.68**
Pretrained	Training	0.24	93.8%	0.93
	Validation	0.39	78.5%	0.78
	Testing	**0.50**	**78.6%**	**0.75**

Interpretability Analysis. With our final pretrained classification network, examined all correctly labelled driver images with Grad-CAM. We examined several input images from the testing dataset. Fig. 3 shows examples of these images with regions more important to the network classification highlighted. We generally observed highlighted regions to fall between the red lines where drivers were manually demarcated.

4. Discussion

We successfully demonstrated improvement in driver detection accuracy after contrastive pretraining of our neural network. This performance increase is not directly related to the electrical features of drivers, but rather to the network learning superior feature space representations of electrograms which are more amenable than raw electrograms for subsequent classification tasks. This amenability is derived from these representations possessing invariance to the augmentations used in the contrastive task – augmentations similar to common noise patterns and data restrictions (such as incomplete electrode contact) in endocardial data.

To our knowledge, our unlabeled endocardial electrogram dataset is the largest in AF deep learning literature. However, augmentation methods remain available which could have increased the effective size of our dataset. While our unlabeled dataset was sufficient for our pretraining task to improve our final classification, we did not explore data augmentation as a means of increasing the size of the unlabeled or labeled dataset. Such augmentations could improve network outcomes and should be investigated.

In this work, we utilized minimally preprocessed electrograms as input to our networks, using only QRS-subtraction and filtration of powerline noise. This allows for faster determination of drivers, reducing patient procedure time. Additionally, networks may be able to identify markers of drivers in raw electrograms that are otherwise undetectable to the human eye. Unnecessary removal of raw signal may degrade network performance.

Other machine learning algorithms in literature have achieved driver classification accuracies of up to 95%, exceeding our network performance [4], [6], [11]. We attribute this to differences in electrograms capture modality and use of phase maps rather than minimally processed electrograms. When limited to deep neural networks trained on high-density endocardial mapping datasets, we find comparable accuracies (~80%) [11]. Additionally, heterogeneity in dataset difficulty remains unexplored, preventing direct comparison of our results to others. Our innovation is in the introduction of pretraining to driver classification, where we show improvement in driver detection accuracy after pretraining. When combined with other advances in network design, data collection, and training routines, pretraining may lead deep learning algorithms to achieve results equal or superior to manual identification.
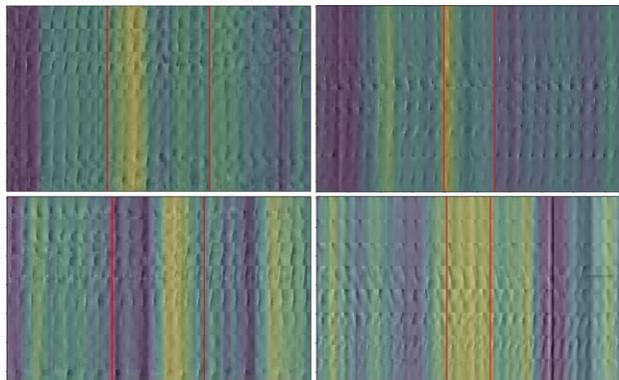

Figure 3. Example electrograms with drivers after Grad-CAM analysis. Regions more important to network identification are highlighted in yellow, and regions where drivers have been manually marked are indicated between the red lines.

We note synergy between the use of a basket catheter and 3D convolutional neural networks. Networks with 3D layers are able to preserve more spatial information than 2D and 1D networks via use of kernels which respect adjacencies of electrodes. When used with a basket catheter, these adjacencies are consistently respected across all measurements. Deforming catheters would not have such consistent adjacencies and would need alternate design to incorporate spatial conformation of electrodes. Drivers are identified by sequences of endocardial activations with spatial relationships, making this an important network design constraint.

Network Interpretability. When Grad-CAM was applied to our final pretrained network, we observed that many of the regions with greatest network attention

regarding identifying drivers correctly were located within manually annotated times of driver occurrence. Given this concurrence between automated and manual attention, we find it promising as to the underlying rationale utilized by the network.

Given the high functional capacity of the networks used, overfitting to spurious or unconsidered correlates of drivers is possible. The use of a testing dataset with entirely separate data from training data was intended to reduce this effect. With respect to Grad-CAM, attention on areas of electrograms not classified as drivers would be strong evidence of overfitting. This gives especial merit to the supposition that the network was learning features of drivers rather than noise or extraneous correlates.

5.   Conclusions

We confirmed our hypothesis, showing driver detection accuracy of our neural network to significantly increase after pretraining. Additionally, this pretraining process is non-specific to driver detection and can be explored as parameter initialization for other tasks.

6.   Limitations

We used a canine paced model of AF as source for our datasets, limiting the generalizability of our results to clinical electrophysiology. Additionally, unlabelled driving mechanisms may have been included in the contrastive pretraining dataset. However, since these mechanisms were unannotated, we expect this to have had little effect on classification accuracy.

Acknowledgments

References

[1]   G. Lippi, F. Sanchis-Gomar, and G. Cervellin, "Global epidemiology of atrial fibrillation: An increasing epidemic and public health challenge," *International Journal of Stroke*, vol. 16, no. 2, pp. 217–221, Feb. 2021, doi: 10.1177/1747493019897870.

[2]   J. E. Poole *et al.*, "Recurrence of Atrial Fibrillation After Catheter Ablation or Antiarrhythmic Drug Therapy in the CABANA Trial," *Journal of the American College of Cardiology*, vol. 75, no. 25, pp. 3105–3118, Jun. 2020, doi: 10.1016/j.jacc.2020.04.065.

[3]   J. Jalife, O. Berenfeld, and M. Mansour, "Mother rotors and fibrillatory conduction: a mechanism of atrial fibrillation," *Cardiovascular Research*, vol. 54, no. 2, pp. 204–216, May 2002, doi: 10.1016/S0008-6363(02)00223-7.

[4]   A. M. Zolotarev *et al.*, "Optical Mapping-Validated Machine Learning Improves Atrial Fibrillation Driver Detection by Multi-Electrode Mapping," *Circulation: Arrhythmia and Electrophysiology*, vol. 13, no. 10, p. e008249, Oct. 2020, doi: 10.1161/CIRCEP.119.008249.

[5]   M. Gutiérrez-Fernández-Calvillo, M. Á. Cámara-Vázquez, I. Hernández-Romero, M. S. Guillem, A. M. Climent, and Ó. Barquero-Pérez, "Non-Invasive Atrial Fibrillation Driver Localization Using Recurrent Neural Networks and Body Surface Potentials," in *2022 Computing in Cardiology (CinC)*, Sep. 2022, pp. 1–4. doi: 10.22489/CinC.2022.163.

[6]   M. I. Alhusseini *et al.*, "Machine Learning to Classify Intracardiac Electrical Patterns During Atrial Fibrillation," *Circulation: Arrhythmia and Electrophysiology*, vol. 13, no. 8, p. e008160, Aug. 2020, doi: 10.1161/CIRCEP.119.008160.

[7]   L. Torrey and J. Shavlik, "Transfer learning," in *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, IGI global, 2010, pp. 242–264.

[8]   D. J. Dosdall *et al.*, "Chronic atrial fibrillation causes left ventricular dysfunction in dogs but not goats: experience with dogs, goats, and pigs," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 305, no. 5, pp. H725–H731, 2013.

[9]   T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*, PMLR, 2020, pp. 1597–1607.

[10]   R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations From Deep Networks via Gradient-Based Localization".

[11]   G. R. Ríos-Muñoz, F. Fernández-Avilés, and Á. Arenal, "Convolutional Neural Networks for Mechanistic Driver Detection in Atrial Fibrillation," *International Journal of Molecular Sciences*, vol. 23, no. 8, p. 4216, 2022.

Address for correspondence:

Ravi Ranjan, MD, PhD.
30 N 1900 E Rm 4A100
Salt Lake City, UT 84132-2101, USA
E-mail: ravi.ranjan@hsc.utah.edu