Contrastive pretraining improves deep learning classification of endocardial electrograms in a preclinical model

Bram Hunt, BS,[1,2,3] Eugene Kwan, PhD,[1,2,3] Jake Bergquist, PhD,[1,2,4] James Brundage, MD,[5] Benjamin Orkild, BS,[1,2,3] Jiawei Dong, PhD,[1,2,3] Eric Paccione, MS,[1,2,3] Kyoichiro Yazaki, MD,[1,2,3] Rob S. MacLeod, PhD,[1,2,4] Derek J. Dosdall, PhD,[1,2,3,6] Tolga Tasdizen, PhD,[4,7] Ravi Ranjan, MD, PhD[1,2,3,4]

From the [1]Department of Biomedical Engineering, University of Utah, Salt Lake City, Utah, [2]Nora Eccles Harrison Cardiovascular Research and Training Institute, University of Utah, Salt Lake City, Utah, [3]Division of Cardiovascular Medicine, Department of Internal Medicine, University of Utah, Salt Lake City, Utah, [4]Scientific Computing and Imaging Institute, University of Utah, Salt Lake City, Utah, [5]Spencer Fox Eccles School of Medicine, University of Utah, Salt Lake City, Utah, [6]Division of Cardiothoracic Surgery, Department of Surgery, University of Utah, Salt Lake City, Utah, and [7]Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, Utah.

BACKGROUND Rotors and focal ectopies, or "drivers," are hypothesized mechanisms of persistent atrial fibrillation (AF). Machine learning algorithms have been used to identify these drivers, but the limited size of current driver data sets constrains their performance.

OBJECTIVE We proposed that pretraining using unsupervised learning on a substantial data set of unlabeled electrograms could enhance classifier accuracy when applied to a smaller driver data set.

METHODS We used a SimCLR-based framework to pretrain a residual neural network on 113,000 unlabeled 64-electrode measurements from a canine model of AF. The network was then fine-tuned to identify drivers from intracardiac electrograms. Various augmentations, including cropping, Gaussian blurring, and rotation, were applied during pretraining to improve the robustness of the learned representations.

RESULTS Pretraining significantly improved driver detection accuracy compared with a non-pretrained network (80.8% vs 62.5%). The pretrained network also demonstrated greater resilience to reductions in training data set size, maintaining higher accuracy even with a 30% reduction in data. Gradient-weighted Class Activation Mapping analysis revealed that the network's attention aligned well with manually annotated driver regions, suggesting that the network learned meaningful features for driver detection.

CONCLUSION This study demonstrates that contrastive pretraining can enhance the accuracy of driver detection algorithms in AF. The findings support the broader application of transfer learning to other electrogram-based tasks, potentially improving outcomes in clinical electrophysiology.

KEYWORDS Atrial fibrillation; Deep learning; Machine learning; Pretraining; Endocardial electrograms; Atrial fibrillation driving mechanisms; Pretraining augmentation

(Heart Rhythm O^2 2025;6:473–480) Published by Elsevier Inc. on behalf of Heart Rhythm Society.

Introduction

Atrial fibrillation (AF) is the most prevalent cardiac electrical disease, affecting more than 50 million patients globally.[1] While pulmonary vein isolation via ablation remains a conventional treatment of AF, it is still associated with high recurrence in cases of persistent atrial fibrillation (persAF).[2] Nonpulmonary vein mechanisms in the form of focal ectopies and reentrant rotors, termed "drivers," have been proposed as the reason for this recurrence.[3] As such, there is a growing interest in the automated identification of drivers in auxiliary treatments of persAF. Recent successes with machine learning algorithms for driver identification have been reported, leveraging data from endocardial mapping, optical mapping, and electrocardiographic imaging.[4–6]

Despite broad utilization in the 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 a preliminary task to establish a set of initialized parameters.[7] In some cases, this

Address reprint requests and correspondence: Dr Ravi Ranjan, Division of Cardiovascular Medicine, Department of Internal Medicine, University of Utah, 30 N 1900 E, Room 4A100, Salt Lake City, UT 84132. E-mail address: ravi.ranjan@hsc.utah.edu.

2666-5018/Published by Elsevier Inc. on behalf of Heart Rhythm Society. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

https://doi.org/10.1016/j.hroo.2025.01.008

KEY FINDINGS

- *Improved accuracy:* Pretraining a neural network using a SimCLR-based framework on a large data set of unlabeled electrograms significantly improved driver detection accuracy from 62.5% to 80.8%.
- *Effective augmentations:* The most effective augmentations for pretraining were cropping, Gaussian blurring, and catheter rotation, which helped the network learn invariant features crucial for accurate classification.
- *Robustness to data reduction:* The pretrained network maintained higher performance even when the training data set size was reduced by up to 30%, indicating robustness and potential for application in scenarios with limited data.
- *Interpretability:* Gradient-weighted Class Activation Mapping visualizations showed that the network's attention aligned well with manually annotated driver regions, suggesting that the network learned meaningful features related to driver identification.
- *Potential for broader application:* The pretraining approach is not specific to driver detection and could be applied to other electrophysiological tasks, enhancing the generalizability and utility of the method.

pretraining may improve accuracy and expedite training when a model is subsequently fine-tuned on a separate task, even when the later task has few training samples. Many pretraining tasks are available for either labeled or unlabeled data, including denoising, image inpainting, and classification of readily available markers (eg, age or sex).

The SimCLR process is an unsupervised contrastive pretraining technique in which networks are trained to produce semantically meaningful representations of data.[8] SimCLR maximizes agreement between differently augmented views of the same input image while minimizing agreement between views of dissimilar images. Augmentation choice is key in this process. Generally, chosen augmentations apply a common noise type or highlight important features in data. Examples of augmentations for image data sets include rotation, cropping, and color jittering. While contrastive learning frameworks have been tested on endocardial electrograms (EGMs), there are no standard augmentations for the endocardial EGM and it is unclear what types of augmentations are most effective in such contrastive learning processes.[9]

Our laboratory has built an extensive data set of endocardial voltage measurements with ultra-high-density mapping catheters in a canine model. In this article, we evaluate the hypothesis that contrastive pretraining will improve the accuracy of a deep learning model in identifying driving mechanisms from endocardial EGMs. Successful improvement in driver detection accuracy would be impetus for broader application of pretraining in driver detection as well as for non–driver-related tasks. In addition, since ideal augmentation methods for EGM-based contrastive learning remain unclear, we analyzed the effect of several different augmentations on testing performance. We also examined the comparative resilience of our best pretrained and non-pretrained networks against reductions in training data set size.

Methods

For all studies, we adhered to the *Guide for the Care and Use of Laboratory Animals*. The Institutional Animal Care and Use Committee of the University of Utah approved the protocol. A preliminary version of this work was presented as a conference proceeding with similar methodology.[10]

Paced canine model

We used a paced canine model of AF as described previously (N = 19; mongrel purpose-bred hound; weight 27–35 kg; age 1–2 years).[11,12] Neurostimulators with screw-in bipolar pacing leads were implanted in the right atrium, either at the right atrial appendage or at the lateral wall. Initially, pacing was set at 50 Hz for 1 second every other second. Every 1–2 weeks, pacing was paused, and electrocardiogram recordings were taken to check for perAF. *persAF* was defined as AF lasting more than 20 minutes after stopping external stimulation. Once persAF was confirmed, the pacing interval was adjusted to 1 second of pacing per minute to ensure AF reinitiation if the animal reverted to sinus rhythm. Most animals developed persAF within 3 weeks of starting the pacing protocol. Weekly electrocardiogram recordings were used to confirm sustained AF. No spontaneous termination of AF was observed during local EGM recordings, and animals maintained AF for an average of 6 months.

Mapping studies

Serial electrophysiological studies were conducted at 1-, 3-, and 6-months post-AF induction. Pacing was halted, and the Rhythmia mapping system with a 64-electrode high-density Orion catheter (Boston Scientific, Marlborough, MA) was used to create detailed endocardial maps of sustained AF. Geometries were generated using the Rhythmia mapping system's internal impedance mapping software, combined with magnetic tracking and intracardiac echocardiography (ICE) guidance to ensure comprehensive atrial coverage.

The catheter was positioned at stable sites within the atria to record AF activity, ensuring good contact and acquiring 4 minutes of 64-electrode EGMs per site. Contact was optimized by adjusting the catheter to maximize voltage at each site, verified through real-time ICE observation and 3-dimensional (3D) atrial geometry positioning using fluoroscopy and ICE. This process was repeated across the left and right atria, requiring 26–34 individual recordings (104–136 minutes of EGMs) per study. All EGMs were captured at a sampling frequency of 953.7 Hz.

Animals were fasted for at least 12 hours before procedures, sedated with propofol (5–8 mg/kg intravenous), and intubated. Anesthesia was maintained with vaporized isoflurane (1.5%–4%). Femoral vein access was achieved using 8.5- and 9-F sheaths (Abbott, Abbott Park, IL) and femoral artery access with 5-F sheaths for blood pressure monitoring. Heparin was not administered to minimize postprocedure bleeding. Transseptal access was obtained using NRG transseptal needles (Baylis Medical Company, Mississauga, ON, Canada) under ICE and fluoroscopy guidance. All animals were in AF at the start of the procedures and maintained AF throughout the mapping studies. Pacing was resumed after each study.

Driver identification

Unipolar atrial AF EGMs were analyzed to identify rotational or focal mechanisms driving AF. The atrium was divided into 15 major anatomical sites, and recordings from each site were inspected. The 2-second recording with the highest dominant frequency was selected, and QRS-T artifacts were removed using average beat subtraction, with powerline noise filtered out using a 60-Hz notch filter. The first temporal derivative of these cleaned signals was used to create activation sequence videos. For sites with multiple recordings, the one with the higher dominant frequency was analyzed. A reviewer (B.H.) annotated rotational or focal activation patterns, with sites exhibiting $\geq$3 consecutive rotations or focal activations classified as "drivers."

Data structure and network design

We split the 64-electrode EGMs from all studies into samples of 2 seconds each. We then transformed these EGMs into stacked images of 8 $\times$ 8 (electrodes $\times$ catheter splines) with 1907 channels encoding the time dimension. This choice of stacked images enabled us to retain the spatial relationships between splines and electrodes; adjacent electrodes on different splines would be adjacent in the 3D representation. As default, we used an 18-layer 3D residual network (ResNet) as the base network for our model.

Contrastive learning

We pretrained our neural network with the SimCLR contrastive learning process.[8] In brief, we replace the final linear layer of the 3D ResNet with a 528 $\times$ 1000 linear layer feeding into a multilayer perceptron termed the *projection head*. In order of application, this projection head consisted of a 1000 $\times$ 1000 linear layer, a rectified linear unit, and a 1000 $\times$ 128 linear layer to produce 128 output features. Then, we trained this network to encode images derived from the same original images (eg, 2 crops of a base image or a blurred image and its unblurred original image) 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 with one another. Augmentations are also chosen such that semantic meaning is retained after transformation, akin to rotations of the same image. This results in each image becoming a class unto itself.

As our default pretraining hyperparameters, we used normalized temperature-scaled cross-entropy loss, a layer-wise adaptive rate scaling optimizer with a learning rate of 4.8, a weight decay of 10^{-6}, and a batch size of 4096. We trained for 200 epochs with a linear warm-up for the first 10 epochs followed by a cosine decay schedule without restarts, terminating training upon plateau in loss reduction.

Augmentations

For our augmentations, we examined a set of traditional machine learning augmentations and physiologically relevant transformations. For our traditional augmentations, we used (1) cropping in spline and electrode dimensions followed by bilinear interpolation back to the original size and (2) Gaussian blurring in electrode and spline dimensions with a kernel size of 3 $\times$ 3. Our cropping reduced image area in the spatial dimensions by up to 75% with random aspect ratios $\pm$25% of the original. For our physiologically based augmentations, we used (1) rotation of the catheter by rearrangement of the spline dimensions, (2) the addition of Gaussian electrical noise with a random standard deviation between 0% and 100% of the standard deviation of the target signal, and (3) differentiation of the time signal (dV/dt). With the exception of random rotation and cropping, augmentations were applied with a 50% chance of occurrence. We evaluated all 10 three-sample combinations of these augmentations in the contrastive learning process with subsequent fine-tuning on the driver classification task.

Classifier training

After contrastive pretraining, we fine-tuned network parameters on the driver classification task. Here, we detach the projection head and replace it with a linear layer with a binary output as seen in Figure 1. For our data set splits, we randomly selected 3 animals and reserved their data for the testing classification data set. EGMs from those animals were not used in the pretraining task. The rest of the data were used for training with 10% of the data reserved for validation. For all networks (pretrained and non-pretrained), we performed a training grid search with 7 logarithmically spaced learning rates varying between 0.0001 and 0.1. We used a batch size of 256, a weight decay of 10^{-6}, and a training length of 100 epochs. In this training process, we did not use any augmentations. Early stopping was used if validation loss did not improve for 20 epochs. In addition, all networks were fine-tuned with 2 different methods: first, by updating all parameters, and second, by updating only the parameters of the final linear layer. After completing the training grid searches with both pretrained and non-pretraining networks, we identified networks with the lowest validation loss and evaluated their performance on the testing data set. Network parameters were saved at their validation loss minima, and the parameters for each network at those minima were used for the testing data set evaluations. As a

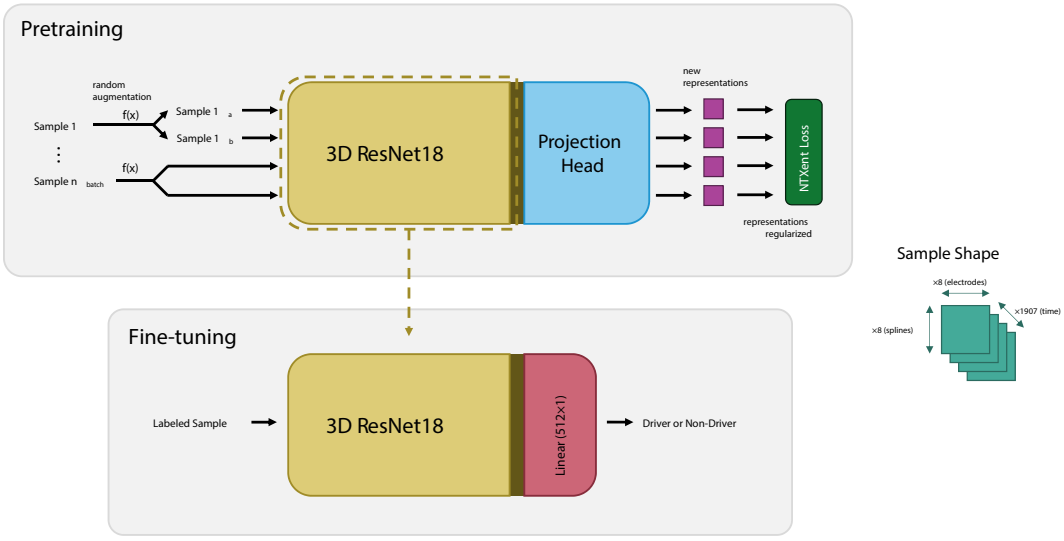

Figure 1 Model training schematic. First, the network is trained on the SimCLR task with unlabeled atrial fibrillation 64-electrode electrograms (EGMs). After the pretraining task is complete, the projection head is detached and replaced with a single linear layer leading to the binary output. The network is subsequently trained on the driver classification task with labeled 64-electrode EGM data. 3D = 3-dimensional.

threshold for classification in the testing data set, we took an average of the best threshold for the training and validation data sets. Those thresholds were found by optimizing against the receiver operating characteristic curves of the network outputs on those data sets.

Interpretability analysis

Interpretability analysis has become an essential for clinical acceptance of deep neural networks because of the opaque rationale underlying network performance. To address this, we used Gradient-weighted Class Activation Mapping (Grad-CAM) visualizations, examining model gradients and activations to scrutinize decision-making processes in testing outcomes.[13] In brief, Grad-CAM computes the elementwise changes in classification confidence via multiplication of the gradient and activations of an input image. These changes are then laid onto the original image to highlight which elements were responsible for the greatest contributions to the final class decision by the network. For our studies, we used the fourth layer of our highest scoring 3D ResNet to perform these Grad-CAM visualizations.

Results

In total, we obtained 113,406 sixty-four–electrode EGMs of 2 seconds each. From this, we examined 709 EGMs according to the methods described above and classified them as containing a driver or nondriver. From the 709 samples used for driver identification, we manually found and labeled 396 nondrivers, 259 rotors, and 54 focal ectopies. Our classification training, validation, and testing data comprised 639 samples (290 nondriver and 349 driver), 60 samples (32 nondriver and 28 driver), and 110 samples (74 nondriver and 36 driver) respectively. Specific times where drivers began and terminated were noted. The remaining unlabeled EGMs were used in the contrastive pretraining process. All EGMs from animals in the testing data set for driver classification were excluded from both the pretraining and driver classification training tasks. The number of EGMs used from each animal and study for each data set is reported in Online Supplemental Appendix 1.

Classification

The results of networks where all parameters were fine-tuned are presented in Table 1. Training curves for these networks are reported in Online Supplemental Appendix 2. We found our best pretrained network (augmentations: rotation, cropping, and Gaussian blurring) outperformed the non-pretrained network in testing accuracy (80.8% vs 62.5%). Per the training and validation optimized thresholds, we classified all outputs above 0.664 as a driver. As seen in Figure 2, this relationship was seen to be durable even when training data set size is reduced by up to 30%.

Augmentation comparison

The results of networks where only the final linear layer was fine-tuned are presented in Table 2. The best augmentation set was crop, Gaussian blurring, and differentiation, with cropping being present in all top 6 of the 10 augmentation combinations. Notably, our best network with only fine-tuning in the last layer reaches near parity in accuracy against the fully trained control network (62.0% vs 62.5%).

Interpretability analysis

For our Grad-CAM analysis, we examined the testing data set performance of our best network with all layers fine-tuned on the driver classification task as discussed above. Figure 3 shows example EGM images with regions more important to the network classification highlighted. We generally observed that these highlighted regions corresponded to the time intervals where drivers were manually annotated.

Table 1 Training (Train.), validation (Val.), and testing (Test) results of the pretrained and non-pretrained networks with fine-tuning on all layers

Augmentations	Train. accuracy	Train. F1 score	Train. loss	Val. accuracy	Val. F1 score	Val. loss	Test accuracy	Test F1 score	Test loss
Rotation, Gaussian blurring and cropping	**91.9%**	**0.922**	**0.240**	78.3%	0.780	0.548	**80.8%**	**0.820**	**0.480**
Rotation, Gaussian noise, dV/dt	85.6%	0.859	0.396	80.0%	0.798	0.565	73.2%	0.764	0.581
Crop, Gaussian blur, Gaussian noise	80.3%	0.805	0.477	78.3%	0.782	0.543	73.2%	0.757	0.610
Rotation, Gaussian blur, dV/dt	78.2%	0.788	0.545	80.0%	0.796	0.565	69.6%	0.752	0.588
Rotation, crop, Gaussian noise	79.5%	0.793	0.678	83.3%	0.833	0.541	74.7%	0.744	0.979
Rotation, crop, dV/dt	76.2%	0.767	0.596	80.0%	0.798	0.537	68.4%	0.719	0.662
Rotation, Gaussian blur, Gaussian noise	75.2%	0.755	0.616	76.7%	0.766	0.591	65.7%	0.680	0.821
Gaussian blur, Gaussian noise, dV/dt	75.4%	0.757	0.593	81.7%	0.816	0.568	63.6%	0.675	0.724
Crop, Gaussian noise, dV/dt	71.2%	0.710	0.671	81.7%	0.815	0.587	63.9%	0.600	0.840
Crop, Gaussian blur, dV/dt	71.6%	0.715	0.668	81.7%	0.817	0.582	64.6%	0.598	0.834
No pretraining	74.4%	0.745	0.613	**83.3%**	**0.833**	**0.528**	62.5%	0.580	0.873

Best networks for each column are in boldface. Rows are sorted by testing F1 score in descending order.
dV/dt = differentiation of the time signal.

Network errors in classification appeared to be reasonable upon closer inspection. For the false-positive site with the highest network confidence (network output: 0.979/1.00), the network's attention in the first 500 ms highlighted a set of reentries in the left atrial appendage. During the initial review, these reentries were noted but classified as noncontinuous

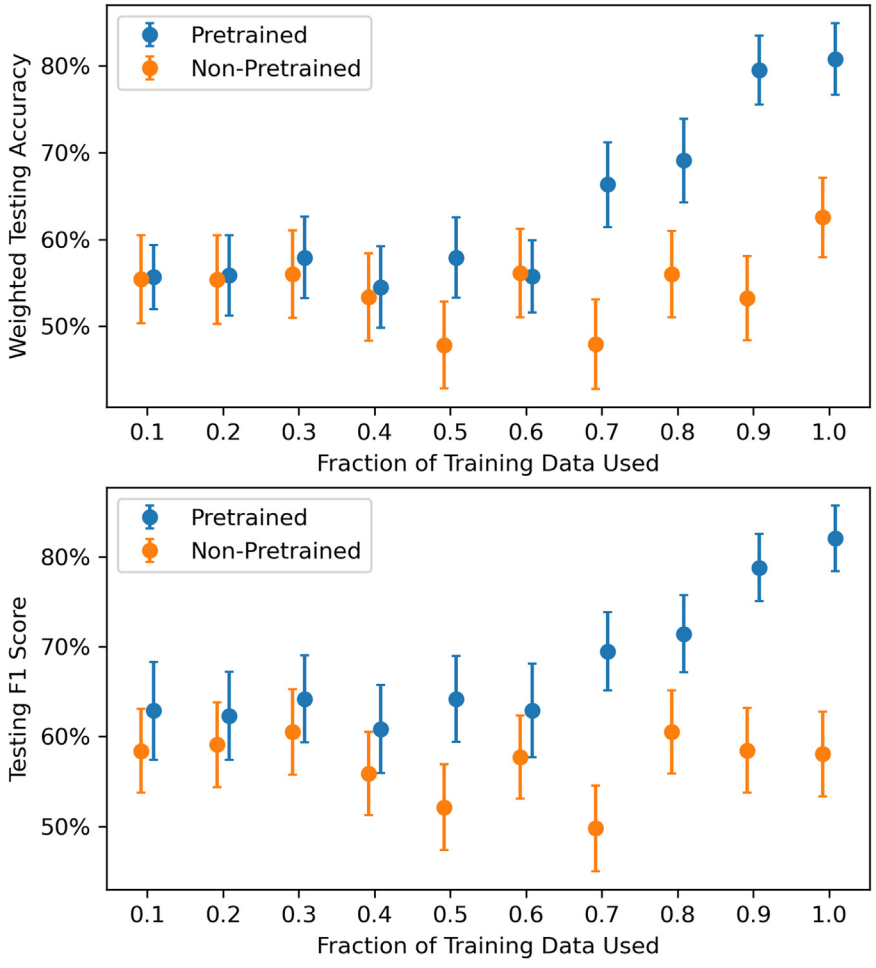

Figure 2 Testing accuracy and F1 score as a function of fraction of the training data set used. The size of the testing and validation data sets remained the same for all networks shown. Our pretrained network substantially outperformed the non-pretrained network until the size of the training data set was reduced by 40%.

Table 2 Training (Train.), validation (Val.), and testing (Test) results of the pretrained networks with fine-tuning on only the final layer

Augmentations	Train. accuracy	Train. F1 score	Train. loss	Val. accuracy	Val. F1 score	Val. loss	Test accuracy	Test F1 score	Test loss
Cropping, Gaussian blurring, dV/dt	70.1%	0.711	0.643	68.3%	0.683	0.656	62.0%	**0.683**	0.680
Rotation, crop, dV/dt	71.5%	0.722	**0.630**	70.0%	0.699	0.661	61.3%	0.679	**0.644**
Crop, Gaussian blur, Gaussian noise	**74.4%**	**0.749**	0.631	63.3%	0.632	0.727	**64.5%**	0.639	0.790
Rotation, crop, Gaussian noise	73.7%	0.746	0.641	63.3%	0.632	0.749	59.5%	0.633	0.812
Rotation, crop, Gaussian blur	73.2%	0.722	0.646	61.7%	0.616	0.739	63.2%	0.590	0.854
Crop, Gaussian noise, dV/dt	67.9%	0.688	0.689	**75.0%**	**0.748**	**0.634**	56.9%	0.577	0.851
Gaussian blur, Gaussian noise, dV/dt	62.2%	0.630	0.755	63.3%	0.605	0.720	57.6%	0.577	0.882
Rotation, Gaussian blur, dV/dt	59.6%	0.603	0.760	71.7%	0.710	0.700	59.1%	0.543	0.849
Rotation, Gaussian blur, Gaussian noise	63.7%	0.621	0.725	56.7%	0.566	0.752	59.2%	0.502	0.859
Rotation, Gaussian noise, dV/dt	63.4%	0.591	0.745	63.3%	0.627	0.735	55.2%	0.415	0.856

Best networks for each column are in boldface. Rows are sorted by testing F1 score in descending order.
dV/dt = differentiation of the time signal.

conduction. A subsequent retrospective review confirmed that the site was a nondriver, indicating that the network had identified a challenging classification scenario.

For the false-negative site with the highest network confidence (0.485/1.000), we observed a set of focal activations at the 1-second mark. Although the network focused on these activations, it classified them as nondrivers. These focal activations were near the edge of the catheter's perspective, and the network may have reasonably interpreted them as planar conduction.

In the true-negative measurement with confidence closest to the classification threshold (0.659/1.000), potential reentries were observed around the left superior pulmonary vein. However, a critical section of the reentrant pathway was determined to be noise. Network attention was aligned with the times where these reentries were observed, suggesting that the network made a similar conclusion regarding the noise.

Discussion

We successfully demonstrated improvement in driver detection accuracy after contrastive pretraining of our neural network, and we attribute this improvement to the initialization, which provided a superior feature space representation of EGMs. We propose that these representations were more amenable than raw EGMs for subsequent classification tasks. This amenability is derived from these representations having 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.

In addition to showing improved final accuracy with fine-tuning on the entire training data set, our best pretrained network showed higher performance than non-pretrained networks even as the size of the driver detection data set was reduced by up to 30%. This is additional evidence in support of the pretrained network having developed a representation space requiring less network complexity to successfully separate classes. The improved robustness of the pretrained network against data set reduction is also promising for application to electrophysiological tasks with even fewer data available for training. Future work should investigate further.

To our knowledge, our unlabeled endocardial EGM data set is the largest in the AF deep learning literature. However, augmentation methods remain available, which could have increased the effective size of our driver data set. While our unlabeled data set 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 labeled data set. Such augmentations could improve network outcomes and should be investigated.

In this work, we used minimally preprocessed EGMs as input to our networks by using only QRS subtraction and filtration of powerline noise. This allows faster determination of drivers, reducing patient procedure time. In addition, networks may be able to identify undiscovered markers of drivers in raw EGMs, and as such, unnecessary removal of the raw signal may degrade network performance. Our QRS subtraction and powerline filtration may have reduced network performance, however; future studies may evaluate whether these adjustments had a deleterious impact.

Other machine learning algorithms in the literature have achieved driver classification accuracies of up to 95%, exceeding our network performance.[4,6] We attribute this to differences in EGM capture modality and use of phase maps rather than minimally processed EGMs. When limited to deep neural networks trained on a high-density endocardial mapping data set, we find comparable accuracy (~80%).[14] In addition, the heterogeneity in data set "difficulty" among studies remains unknown, preventing a direct comparison of our results with 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.

We note synergy between the use of a structured basket-style catheter and 3D convolutional neural networks. Networks

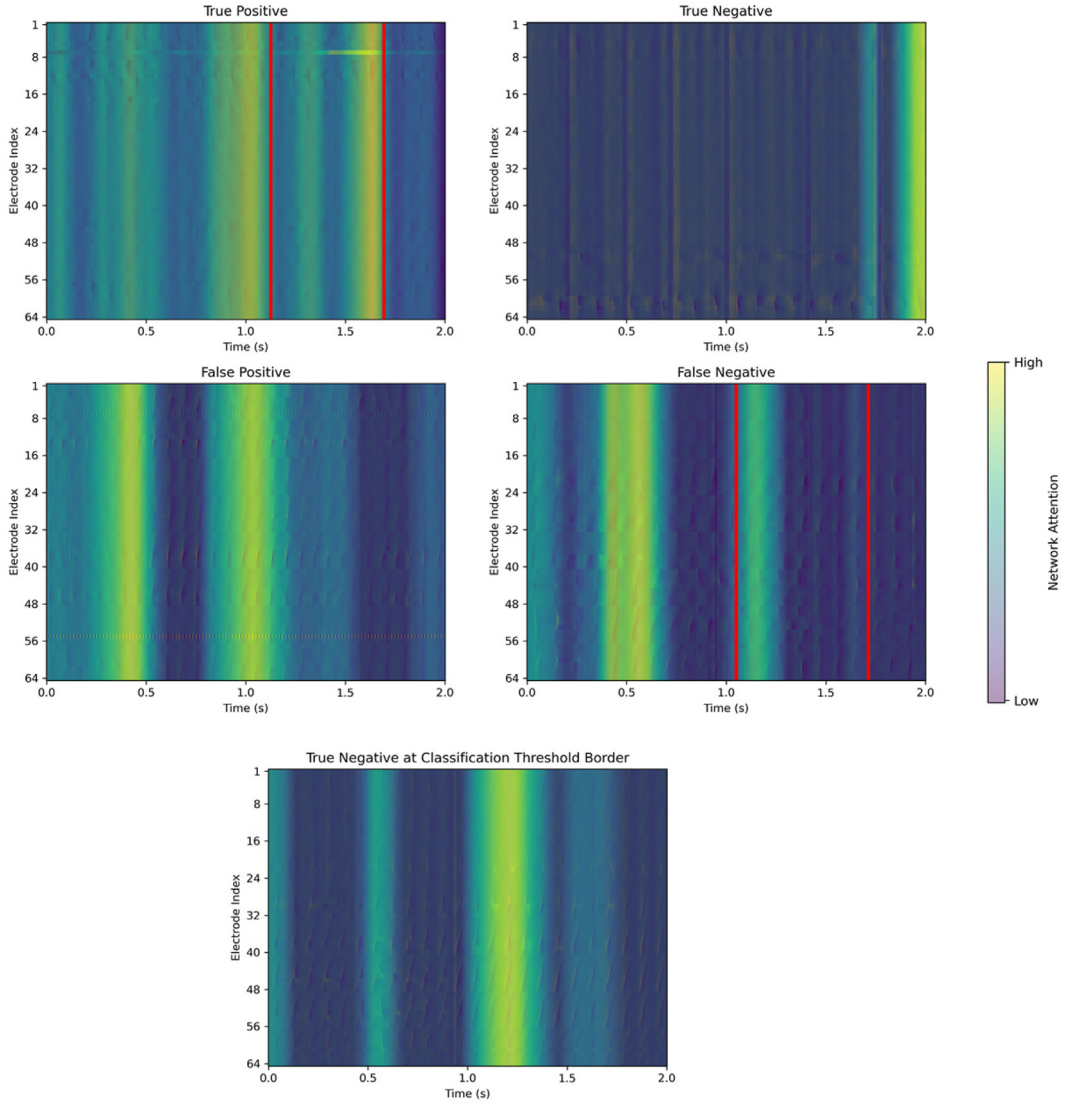

Figure 3 Example electrograms from testing data set after Gradient-weighted Class Activation Mapping (Grad-CAM) analysis of the best performing pretrained network with fine-tuning on all layers. The 3-dimensional tiles used in the data set have been reshaped into 2-dimensional images for simpler visualization. Regions more important to network identification are highlighted in *yellow*, and regions where drivers have been manually marked are indicated between *red lines*. Regions where drivers were manually marked tended to have more network attention than other regions. Shown are true-positive, true-negative, false-positive, and false-negative samples, and these samples had the highest network prediction confidence in comparison to other samples from the testing data set.

with 3D layers are able to preserve more spatial information than 2D and 1D networks via use of kernels that respect electrode adjacency. When used with a basket-style catheter, these adjacencies are consistently respected across 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.

Impact of augmentation choice

As expected, the choice of augmentation was highly influential for network performance in the driver classification task. Cropping was notably impactful, consistently being in all top performing pretrained networks. This may be explained by the cropping augmentation requiring networks to identify EGMs by a broad number of spatial features rather than by focusing on a narrow subset of those features. These spatial features have importance for their utility in capturing structural information of the arrhythmia, which may affect the presence or absence of drivers in a location.[15,16] Catheter rotation was also impactful, similarly being seen in top performing networks. This is unique in that it is a domain-specific augmentation rather than a common machine learning augmentation. This demonstrates the importance of considering augmentations appropriate to subdomains rather than relying on out-of-the-box augmentations for transfer learning tasks. In this case, given that the Orion catheter is rotational symmetric without any known information being stored in this rotation, we consider it a noise source for which networks should be invariant.

In addition to the above, we found both Gaussian noising and Gaussian blurring to be impactful augmentations in the

pretraining process. Similar to rotation, we assert that this reflects a need for the final representation space to become invariant to common electrical signal noise and far-field effects. Both additive Gaussian noising and Gaussian blurring are methods of artificially introducing or enhancing that noise. More sophisticated methods of introducing far-field noises into the contrastive process (eg, random introduction of QRS complexes to the signal) may result in improved accuracy gains.

Curiously, our worst performing pretrained network in the full parameter fine-tuning obtained the highest F1 score in comparison of fine-tuning on only the last layer. We attribute this to the contrastive augmentation set (ie, cropping, Gaussian blurring, and dV/dt) producing a strong initialization but being particularly ill-suited for full parameter optimization—perhaps creating "bad habits" that force the network to rely on features that do not generalize well.

Network interpretability

Understanding the rationale behind network decision making is highly difficult given the opacity of the relationship between the function and output of a network. However, an assessment of network decision making is critical for maximizing the likelihood that network performance has external validity. When we examined our best pretrained network with fine-tuning on all layers under Grad-CAM, we observed that many regions with the greatest network attention were located within manually annotated times of driver occurrence. This concurrence between automated and manual attention is promising, as this implies the network followed an identification process similar to the manual annotation of the underlying data.

Given the high functional capacity of the networks used, overfitting to spurious or unconsidered correlates of drivers is a possibility. The use of a testing data set with entirely separate data from training data was intended to reduce this effect. In practice, we observe our testing and validation accuracies to be generally lower than our training accuracies, demonstrating modest overfitting. With respect to Grad-CAM, attention on areas of EGMs not classified as drivers would be strong evidence of overfitting and nonsensible network rationale. This was not observed, and instead we see that network attention was significantly greater at times where EGMs were actively exhibiting driver behavior. This implies that the network was learning features of drivers rather than noise or extraneous correlates.

Limitations

We used a paced canine model of AF as the source for our data sets, potentially limiting the generalizability of our results to clinical electrophysiology.

Conclusion

We confirmed our hypothesis, showing driver detection accuracy of our neural network to increase after pretraining. Our best augmentation set for contrastive pretraining was cropping in the spline and electrode dimensions, rotation of the catheter, and the addition of random Gaussian blurring. The relative importance of catheter rotation to network performance shows domain-specific augmentation choice to be key when designing the contrastive pretraining process. Finally, our pretraining process is nonspecific to driver detection and can be explored as parameter initialization for other electrophysiological tasks.

Funding Sources: This work was funded by the National Institutes of Health (NIH) grant HL142913 (to Dr Ranjan), the NIH/National Heart, Lung, and Blood Institute (NHLBI) grant 5F31HL162527 (to Dr Kwan), and the NIH/NHLBI grant T32HL007576 (to Dr Bergquist). The Rhythmia Mapping System was provided (on loan) by Boston Scientific.

Disclosures: Dr Ranjan is a consultant for Abbott. The rest of the authors have no conflicts of interest.

Authorship: All authors attest they meet the current ICMJE criteria for authorship.

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

References

1. Lippi G, Sanchis-Gomar F, Cervellin G. Global epidemiology of atrial fibrillation: an increasing epidemic and public health challenge. Int J Stroke16:217–221.
2. Poole PE, Bahnson TD, Monahan KH, et al. Recurrence of atrial fibrillation after catheter ablation or antiarrhythmic drug therapy in the CABANA trial. J Am Coll Cardiol75:3105–3118.
3. Jalife J, Berenfeld O, Mansour M. Mother rotors and fibrillatory conduction: a mechanism of atrial fibrillation. Cardiovasc Res54:204–216.
4. Zolotarev AM, Hansen BJ, Ivanova EA, et al. Optical mapping-validated machine learning improves atrial fibrillation driver detection by multi-electrode mapping. Circ Arrhythm Electrophysiol13:e008249.
5. Gutiérrez-Fernández-Calvillo M, Cámara-Vázquez MÁ, Hernández-Romero I, Guillem MS, Climent AM, Barquero-Pérez Ó. Non-invasive atrial fibrillation driver localization using recurrent neural networks and body surface potentials. In: 2022 Computing in Cardiology. Tampere, Finland: IEEE; 2022:1–4.
6. Alhusseini MI, Abuzaid F, Rogers AJ, et al. Machine learning to classify intracardiac electrical patterns during atrial fibrillation: machine learning of atrial fibrillation. Circ Arrhythm Electrophysiol13:e008160.
7. Torrey L, Shavlik J. Transfer learning. In: Klinger K, Snavely J, Brehm M, eds. Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques. Hershey, PA: IGI Global; 2009. 242–264.
8. Chen T, Kornblith S, Norouzi M, Hinton G. A simple framework for contrastive learning of visual representations. In: . Proceedings of the 37th International Conference on Machine Learning, vol 119. Vienna, Austria: PMLR; 2020:1597–1607.
9. Hunt B, Kwan E, Dosdall D, MacLeod RS, Ranjan R. Siamese neural networks for small dataset classification of electrograms. In: 2021 Computing in Cardiology. Brno, Czech Republic: IEEE; 2021:1–4.
10. Hunt B, Kwan E, Tasdizen T, et al. Transfer learning for improved classification of drivers in atrial fibrillation. In: 2023 Computing in Cardiology. Atlanta, GA: IEEE; 2023:1–4.
11. Dosdall DJ, Ranjan R, Higuchi K, et al. Chronic atrial fibrillation causes left ventricular dysfunction in dogs but not goats: experience with dogs, goats, and pigs. Am J Physiol Heart Circ Physiol305:H725–H731.
12. Hunt B, Kwan E, Paccione E, et al. Are drivers recurring or ephemeral? Observations from serial mapping of persistent atrial fibrillation. Europace26:euae269.
13. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. Int J Comput Vis128:336–359.
14. Ríos-Muñoz GR, Fernández-Avilés F, Arenal Á. Convolutional neural networks for mechanistic driver detection in atrial fibrillation. Int J Mol Sci23:4216.
15. Kwan E, Ghafoori E, Good W, et al. Diffuse functional and structural abnormalities in fibrosis: potential structural basis for sustaining atrial fibrillation. Heart Rhythm 2024; https://doi.org/10.1016/j.hrthm.2024.10.060.
16. Kwan E, Hunt B, Paccione E, et al. Functional and structural remodeling as atrial fibrillation progresses in a persistent atrial fibrillation canine model. JACC Clin Electrophysiol 2024; https://doi.org/10.1016/j.jacep.2024.10.001.