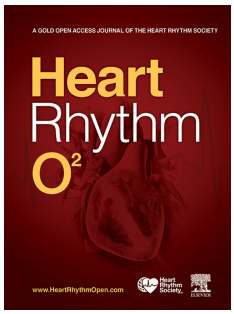
Journal Pre-proof



Contrastive Pretraining Improves Deep Learning Classification of Endocardial Electrograms in a Preclinical Model

Bram Hunt, Eugene Kwan, Jake Bergquist, PhD, James Brundage, Benjamin Orkild, Jiawei Dong, Eric Paccione, Kyoichiro Yazaki, MD, Rob S. MacLeod, PhD, Derek J. Dosdall, PhD, Tolga Tasdizen, PhD, Ravi Ranjan, MD, PhD

PII: S2666-5018(25)00016-9

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

Reference: HROO 641

To appear in: *Heart Rhythm O2*

Received Date: 8 May 2024

Revised Date: 15 January 2025

Accepted Date: 16 January 2025

Please cite this article as: Hunt B, Kwan E, Bergquist J, Brundage J, Orkild B, Dong J, Paccione E, Yazaki K, MacLeod RS, Dosdall DJ, Tasdizen T, Ranjan R, Contrastive Pretraining Improves Deep Learning Classification of Endocardial Electrograms in a Preclinical Model, *Heart Rhythm O2* (2025), doi: https://doi.org/10.1016/j.hroo.2025.01.008.

Contrastive Pretraining Improves Deep Learning Classification of Endocardial Electrograms in a Preclinical Model

Running Title: Pretraining for Endocardial Electrograms

Bram Hunt[1,2,3], Eugene Kwan[1,2,3], Jake Bergquist, PhD,[1,2,4], James Brundage[5], Benjamin Orkild[1,2,3], Jiawei Dong[1,2,3], Eric Paccione[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]

[1]Department of Biomedical Engineering, University of Utah, SLC, UT, USA
[2]Nora Eccles Harrison 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]Spencer Fox Eccles School of Medicine, University of Utah, SLC, UT, USA
[6]Division of Cardiothoracic Surgery, Department of Surgery, University of Utah, SLC, UT, USA
[7]Department of Electrical and Computer Engineering, University of Utah, SLC, UT, USA

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

Disclosures/Conflicts of Interest: This work was funded by NIH grant HL142913 (to RR), NIH/NHLBI 5F31HL162527 (to EK), NIH/NHLBI grant T32HL007576 (to JAB). Dr. Ranjan is a consultant for Abbott.

Abstract

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

Objective: We proposed that pretraining using unsupervised learning on a substantial dataset of unlabeled electrograms could enhance classifier accuracy when applied to a smaller driver dataset.

Methods: We utilized 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 intra-cardiac 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 to a non-pretrained network (80.8% vs. 62.5%). The pretrained network also demonstrated greater resilience to reductions in training dataset size, maintaining higher accuracy even with a 30% reduction in data. Grad-CAM 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, Intracardiac Electrograms, Atrial Fibrillation Drivers, Fibrillatory Mechanisms

1. Introduction

Atrial fibrillation (AF) is the most prevalent cardiac electrical disease, affecting over 50 million patients globally [1]. While pulmonary vein isolation via ablation remains a conventional treatment for AF, it is still associated with high recurrence in cases of persistent atrial fibrillation (persAF) [2]. Non-pulmonary 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 for persAF. Recent successes with machine learning algorithms for driver identification have been reported, leveraging data from endocardial mapping, optical mapping, and ECGI [4], [5], [6].

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 a preliminary task to establish a set of initialized parameters [7]. In some cases, this 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 (e.g., 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 datasets 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 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 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. Additionally, since ideal augmentation methods for EGM-based contrastive learning remain unclear, we analysed 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 dataset size.

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. A preliminary version of this work was presented as a conference proceeding with similar methodology [10].

Paced Canine Model. We utilized a paced-canine model of AF as described previously (n=19, mongrel purpose-bred hound, 27-35 kg, 1-2 yrs.) [11], [12]. Neurostimulators with screw-in bipolar pacing leads were implanted in the right atrium, either at the right atrial appendage or 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 persistent atrial fibrillation (AF). Persistent AF (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 the 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) 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 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 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 electrograms) 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 IV), and intubated. Anesthesia was maintained with vaporized isoflurane (1.5%-4%). Femoral vein access was achieved using 8.5 and 9 Fr sheaths (Abbott), and femoral artery access with 5 Fr sheaths for blood pressure monitoring. Heparin was not administered to minimize post-procedure bleeding. Transseptal access was obtained using NRG transseptal needles (Baylis) 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 electrograms 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 (BH) annotated rotational or focal activation patterns, with sites exhibiting ≥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 8×8 (electrodes × catheter splines) with 1,907 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 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 528×1,000 linear layer feeding into a multilayer perceptron termed the projection head. In order of application, this projection head consisted of a 1,000×1,000 linear layer, a ReLU, and a 1,000×128 linear layer to produce 128 output features. Then, we trained this 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.

This results in each image becoming a class unto themselves.

As our default pretraining hyperparameters, 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.

Augmentations. For our augmentations, we examined a set of traditional ML augmentations and physiologically relevant transformations. For our traditional augmentations, we used 1) cropping in spline and electrode dimensions followed by bilinear interpolation back to original size and 2) Gaussian blurring in electrode and spline dimensions with a kernel size of 3×3. Our cropping reduced image area in the spatial dimensions by up to 75% with random aspect ratios ±25% of 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 3-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 dataset splits, we randomly selected three animals and reserved their data for the testing classification dataset. 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, 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. Additionally, all networks were fine-tuned with two 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 dataset. Network parameters were saved at their validation loss minima, and the parameters for each network at those minima were used for the testing dataset evaluations. As a threshold for classification in the testing dataset, we took an average of the best threshold for the training and validation datasets. Those thresholds were found by optimizing against the receiver-operator-characteristic curves of the network outputs on those datasets.

Interpretability Analysis. Interpretation analysis has become an essential for clinical acceptance of deep neural networks due to the opaque rationale underlying network performance. To address this, we employed 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 final class decision by the network. For our studies, we used the 4th layer of our highest scoring 3D ResNet to perform these Grad-CAM visualizations.

3. Results

In total, we obtained 113,406 64-electrode EGMs of 2-seconds each. From this, we examined 709 EGMs per our methods above and classified them as contained a driver or non-driver. From the 709 samples used for driver identification, we manually found and labeled 396 non-drivers, 259 rotors, and 54 focal ectopies. Our classification training, validation, and testing data comprised 539 samples (290

non-driver and 349 driver), 60 samples (32 non-driver and 28 driver), and 110 samples (74 non-driver 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 dataset 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 dataset is shown in Supplementary Appendix 1.

Classification. The results of networks where all parameters were fine-tuned are shown in **Table 1**. Training curves for these networks are reported in Supplementary 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 dataset size is reduced by up to 30%.

Augmentation Comparison. The results of networks where only the final linear layer was fine-tuned are shown 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 dataset performance of our best network with all layers fine-tuned on the driver classification task per 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.

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 non-continuous conduction. A subsequent retrospective review confirmed that the site was a non-driver, 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 non-drivers. 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.

4. 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 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.

In addition to showing improved final accuracy with fine-tuning on the entire training dataset, our

best pretrained network showed higher performance than non-pretrained networks even as the size of the driver detection dataset 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 dataset 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 dataset is the largest in AF deep learning literature. However, augmentation methods remain available which could have increased the effective size of our driver 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 labeled dataset. Such augmentations could improve network outcomes and should be investigated.

In this work, we utilized minimally preprocessed EGMs 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 undiscovered markers of drivers in raw EGMs, and as such unnecessary removal of 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 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 dataset, we find comparable accuracy (~80%) [14]. Additionally, the heterogeneity in dataset "difficulty" among studies remains unknown, 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.

We note synergy between the use of a structured basket-style 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 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 impact 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 also see both Gaussian noising and 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 the 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 (e.g., 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 the comparison of fine-tuning on only the last layer. We attribute this to the contrastive augmentation set (i.e., 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 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 dataset 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 non-sensible network rationale. This was not observed, and instead we see network attention was significantly greater at times where EGMs were actively exhibiting driver behaviour. This implies 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 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 non-specific to driver detection and can be explored as parameter initialization for other electrophysiological tasks.

6. Limitations

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

Acknowledgments

This work was funded by NIH grant HL142913 (to RR), NIH/NHLBI 5F31HL162527 (to EK), NIH/NHLBI grant T32HL007576 (to JAB). Dr. Ranjan is a consultant for Abbott.

References

[1] G. Lippi, F. Sanchis-Gomar, and G. Cervellin, "Global Epidemiology of Atrial Fibrillation: An Increasing Epidemic and Public Health Challenge," *Int J 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," *J Am Coll Cardiol*, 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," *Cardiovasc Res*, vol. 54, no. 2, pp. 204–216, 2002.

[4] A. M. Zolotarev *et al.*, "Optical Mapping-Validated Machine Learning Improves Atrial Fibrillation Driver Detection by Multi-Electrode Mapping," *Circ Arrhythm Electrophysiol*, vol. 13, no. 10, 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," presented at the 2022 Computing in Cardiology, IEEE, 2022, pp. 1–4.

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

[7] E. S. Olivas, J. D. M. Guerrero, M. Martinez-Sober, J. R. Magdalena-Benedito, and L. Serrano, "Transfer Learning," in *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*, IGI global, 2009.

[8] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A Simple Framework for Contrastive Learning of Visual Representations," presented at the International Conference on Machine Learning, PMLR, 2020, pp. 1597–1607.

[9] B. Hunt, E. Kwan, D. Dosdall, R. S. MacLeod, and R. Ranjan, "Siamese Neural Networks for Small Dataset Classification of Electrograms," presented at the 2021 Computing in Cardiology, IEEE, 2021, pp. 1–4.

[10] B. Hunt *et al.*, "Transfer Learning for Improved Classification of Drivers in Atrial Fibrillation," presented at the 2023 Computing in Cardiology, IEEE, 2023, pp. 1–4.

[11] D. J. Dosdall *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 Physiol*, vol. 305, no. 5, pp. H725–H731, Sep. 2013, doi: 10.1152/ajpheart.00440.2013.

[12] B. Hunt *et al.*, "Are Drivers Recurring or Ephemeral? Observations from Serial Mapping of Persistent Atrial Fibrillation," *Europace*, p. euae269, Oct. 2024, doi: 10.1093/europace/euae269.

[13] 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," *Int J Comput Vis*, vol. 128, no. 2, pp. 336–359, Feb. 2020, doi: 10.1007/s11263-019-01228-7.

[14] G. R. Ríos-Muñoz, F. Fernández-Avilés, and Á. Arenal, "Convolutional Neural Networks for Mechanistic Driver Detection in Atrial Fibrillation," *Int J Mol Sci*, vol. 23, no. 8, p. 4216, 2022.

[15] E. Kwan *et al.*, "Diffuse functional and structural abnormalities in fibrosis: Potential structural basis for sustaining atrial fibrillation," *Heart Rhythm*, 2024, doi: 10.1016/j.hrthm.2024.10.060.

[16] E. Kwan *et al.*, "Functional and Structural Remodeling as Atrial Fibrillation Progresses in a Persistent Atrial Fibrillation Canine Model," *JACC: Clinical Electrophysiology*, 2024, doi: 10.1016/j.jacep.2024.10.001.

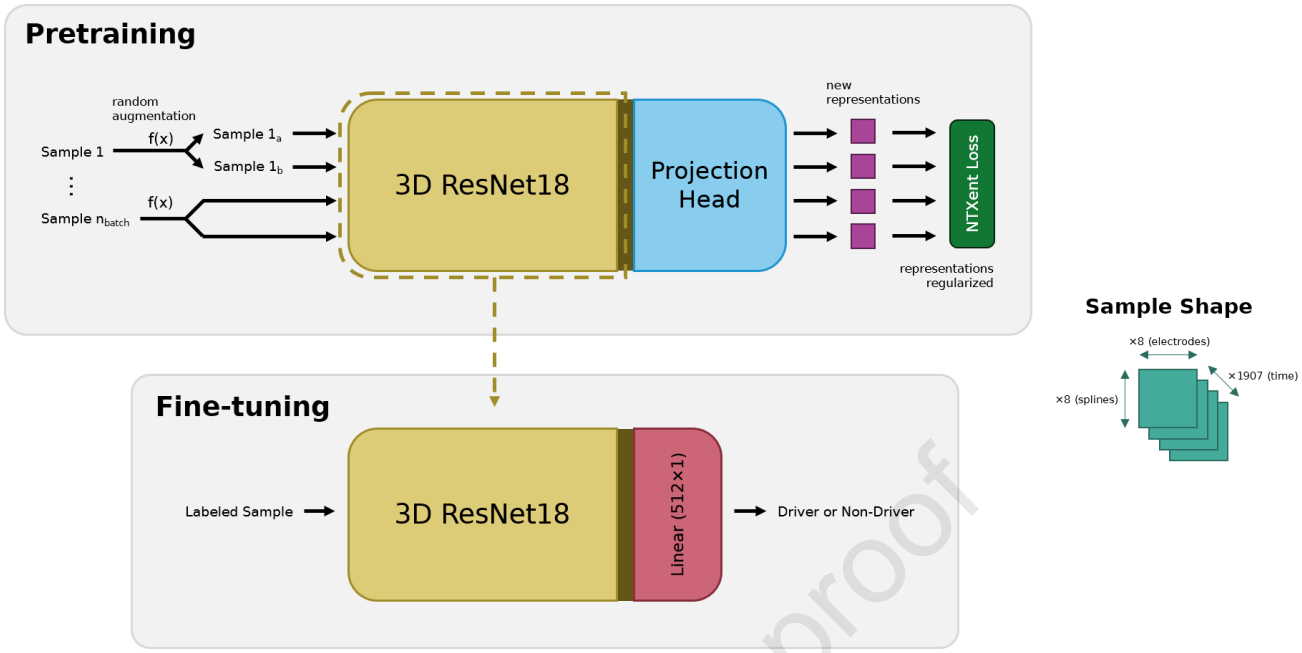


Figure 1. Model training schematic. First, the network is trained on the SimCLR task with unlabeled AF 64-electrodes EGMs. After the pretraining task is complete, the projection head is detached and replaced with a single linear layer leading to the binary output. The network is subsequently trained on the driver classification process with labeled 64-electrode EGM data.

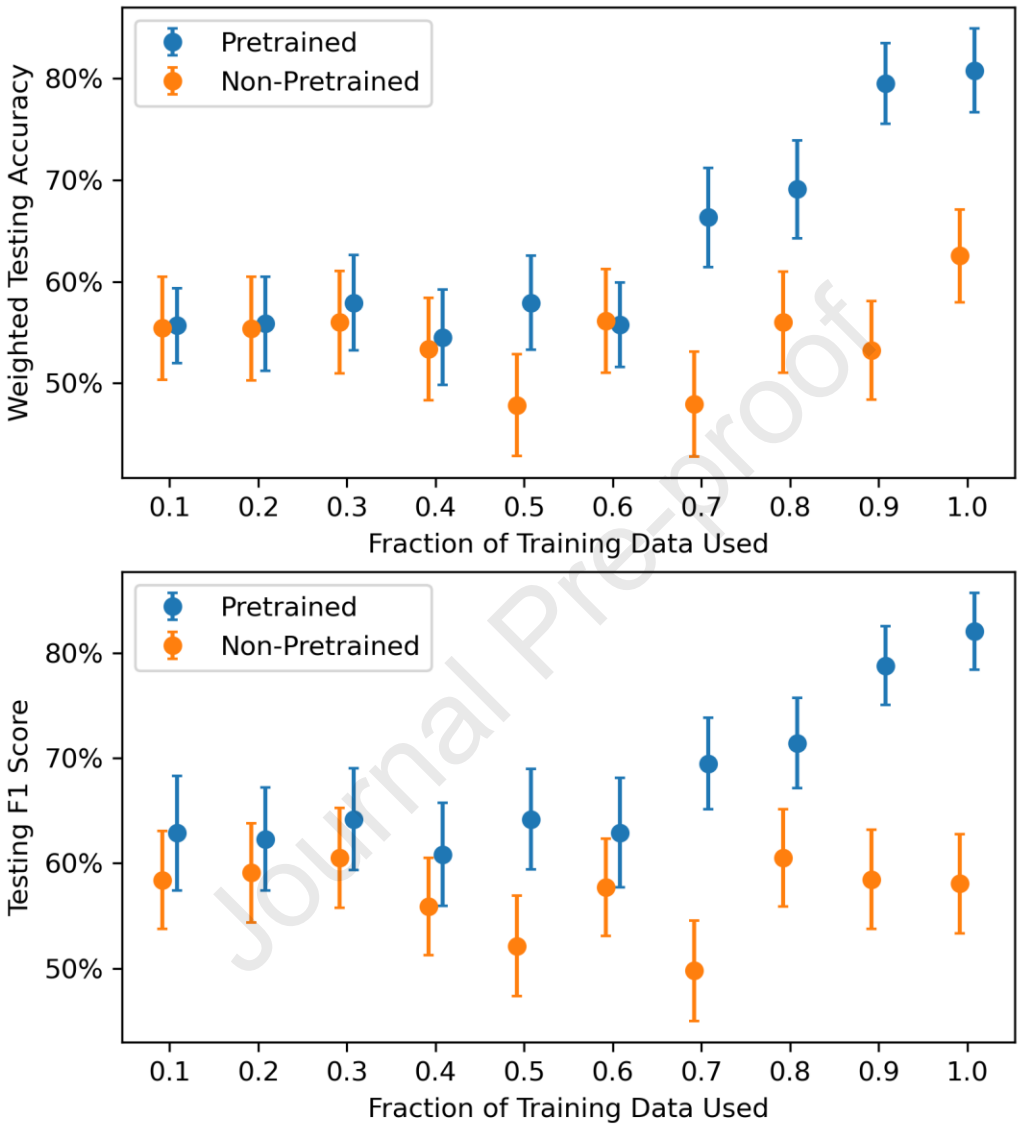


Figure 2. Testing accuracy and F1 score as a function of fraction of the training dataset used. The size of the testing and validation datasets size remained the same for all networks shown. Our pretrained network substantially outperformed the non-pretrained network until the size of the training dataset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

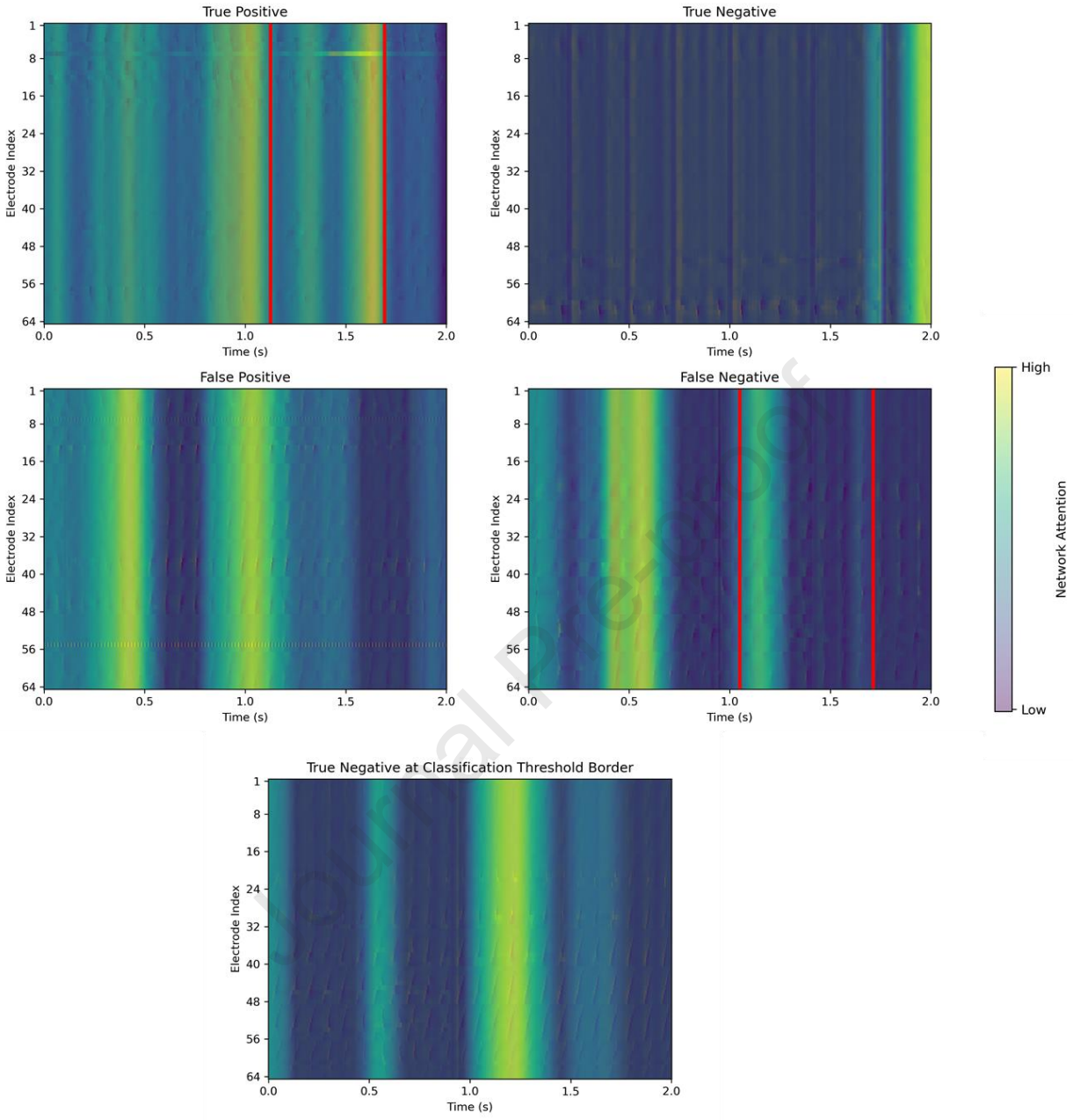


Figure 3. Example EGMs from testing dataset after Grad-CAM analysis of the best performing pretrained network with fine-tuning on all layers. The 3D tiles used in the dataset have been reshaped into 2D images for simpler visualization. Regions more important to network identification are highlighted 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

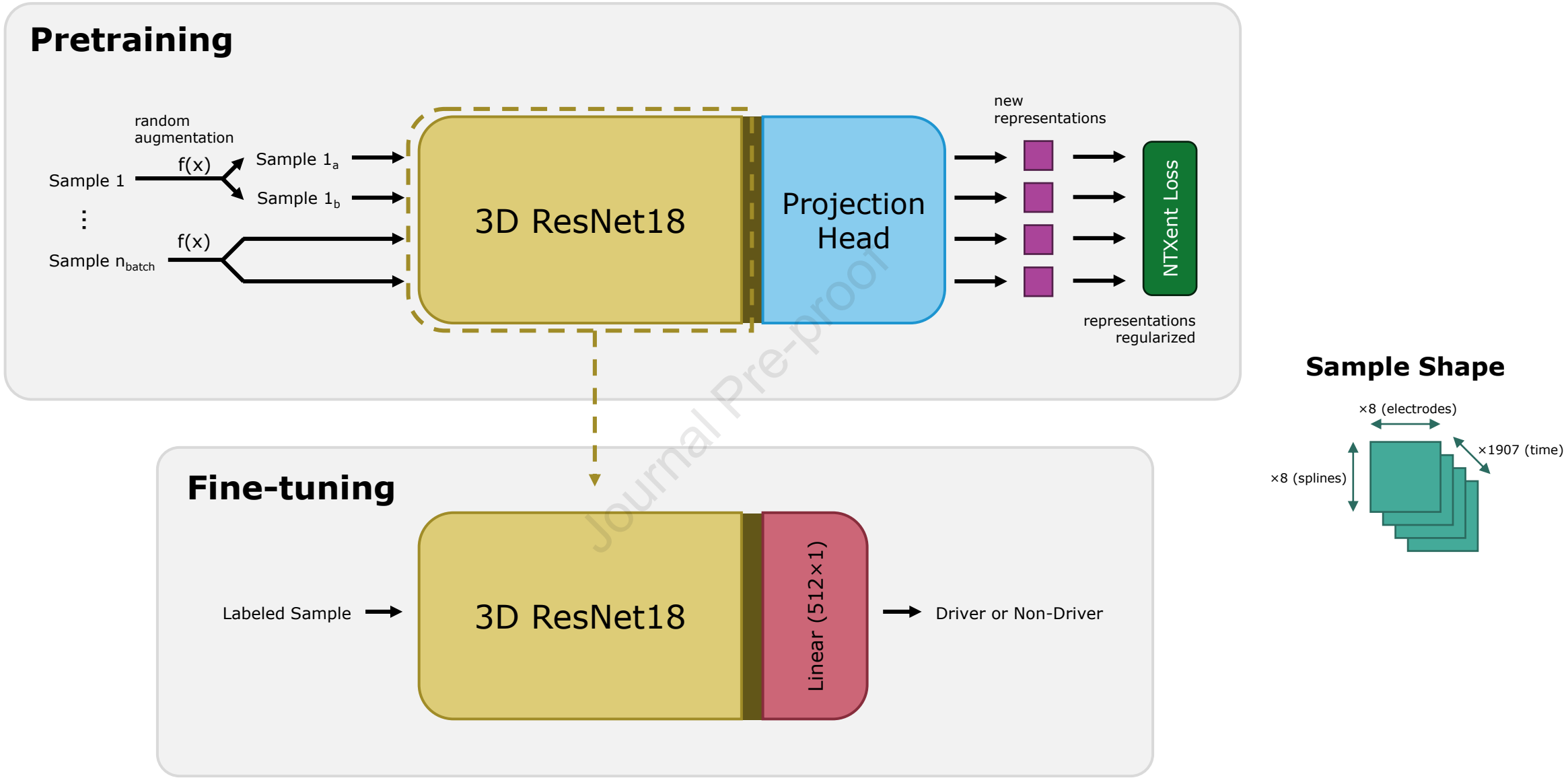
other regions. Shown are true positive, true negative, false positive, and false negative samples, and these samples had the highest network prediction confidence in comparison to other samples from the testing dataset.

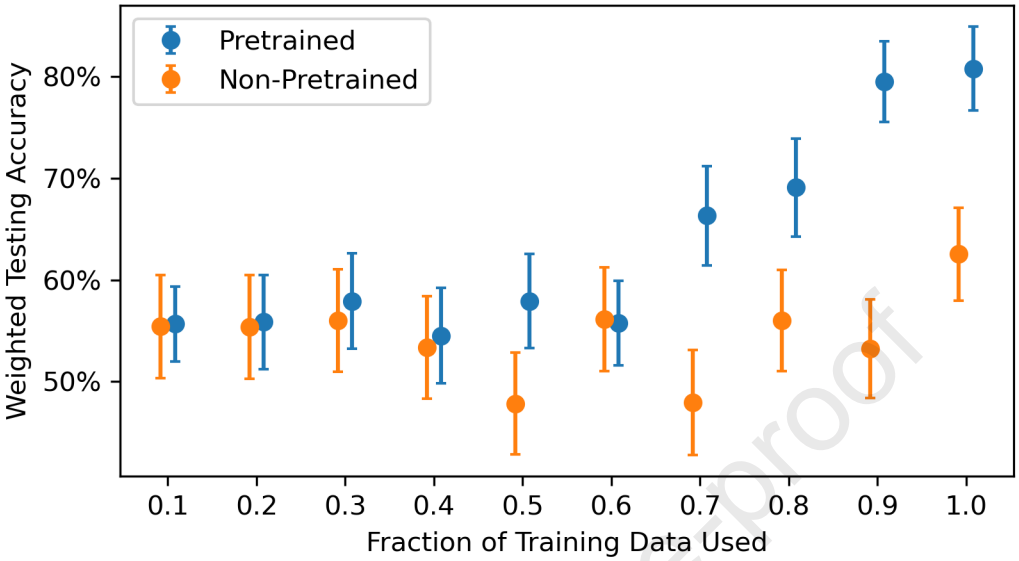
Table 1. Training, validation, and testing results of the pretrained and non-pretrained networks with fine-tuning on all layers. Best networks for each column are bolded. Rows are sorted by testing F1 score in descending order.

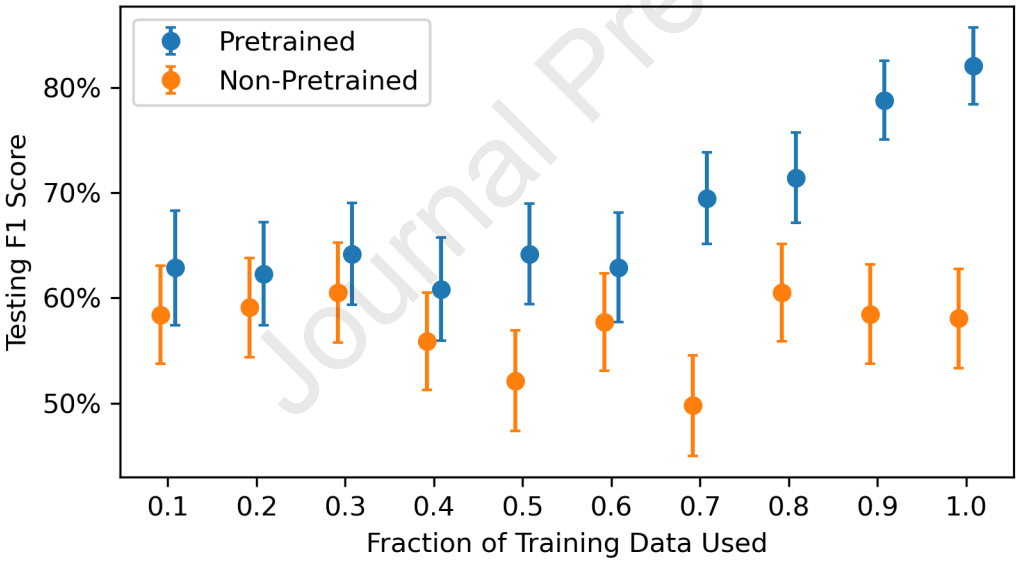
Augmentations	Train. Acc.	Train. F1	Train. Loss	Val. Acc.	Val. F1	Val. Loss	Test Acc.	Test F1	Test Loss
rotation, crop, Gaussian blur	**91.9%**	**0.922**	**0.240**	78.3%	0.780	0.548	**80.8%**	**0.820**	**0.480**
rotation, Gaussian noise, dV/dt	85.6%	0.859	0.396	80.0%	0.798	0.565	73.2%	0.764	0.581
crop, 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
N/A	74.4%	0.745	0.613	**83.3%**	**0.833**	**0.528**	62.5%	0.580	0.873

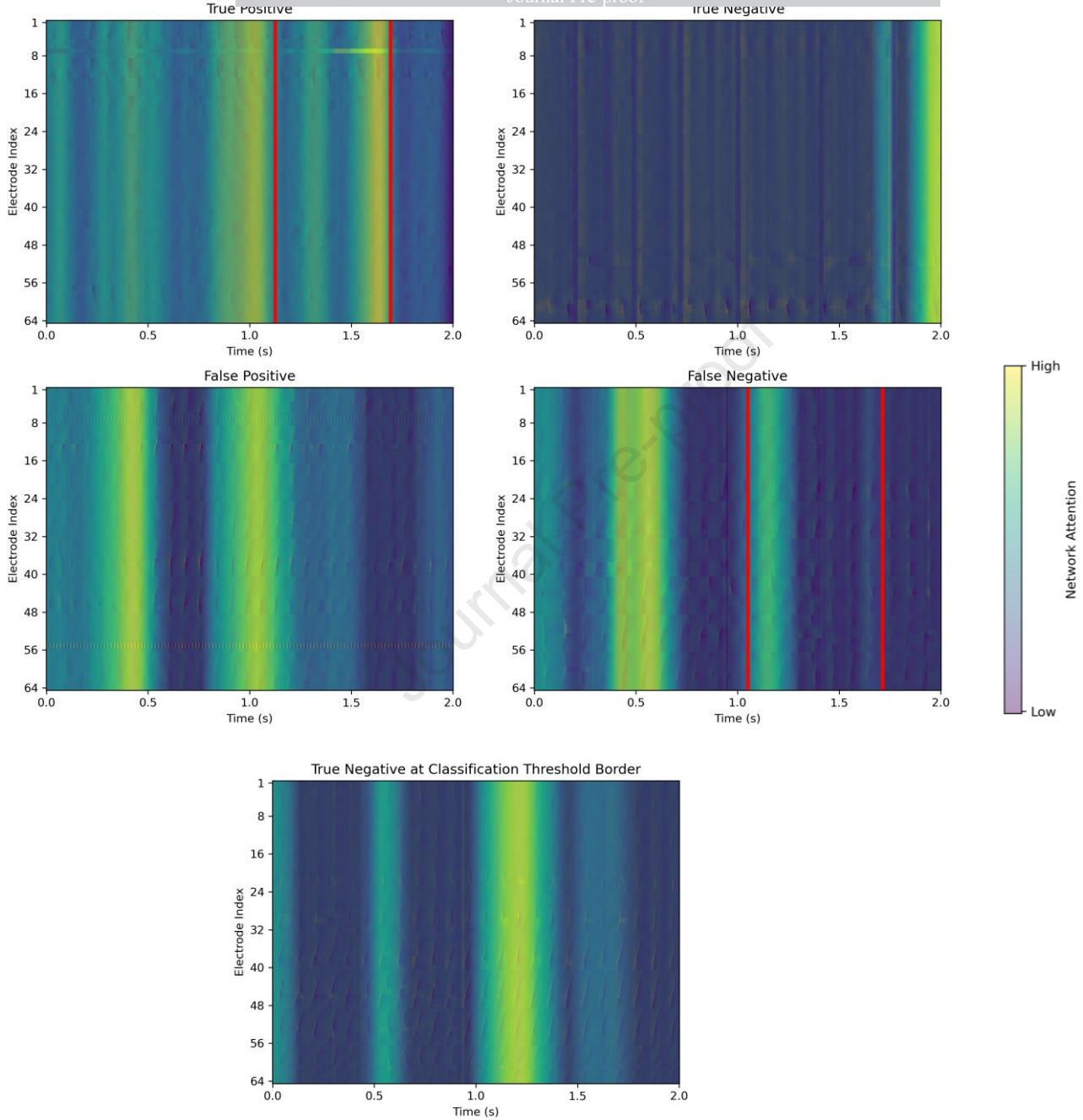
Table 2. Training, validation, and testing results of pretrained networks with fine-tuning on only the final layer. Best networks for each column are bolded. Rows are sorted by testing F1 score in descending order.

Augmentations	Train. Acc.	Train. F1	Train. Loss	Val. Acc.	Val. F1	Val. Loss	Test Acc.	Test F1	Test Loss
crop, Gaussian blur, 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









Key Findings:

- **Improved Accuracy**: Pretraining a neural network using a SimCLR-based framework on a large dataset 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 dataset size was reduced by up to 30%, indicating robustness and potential for application in scenarios with limited data.
- **Interpretability**: Grad-CAM 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.