Deep Transfer Learning for Detection of Atrial Fibrillation Using Holter ECG Color Maps

Todor Stoyanov[1], Vessela Krasteva[1], Stefan Naydenov[2], Ramun Schmid[3], Irena Jekova[1]

[1]Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, Sofia, Bulgaria
[2]Medical University of Sofia, Sofia, Bulgaria
[3]Schiller AG, Baar, Switzerland

Abstract

This study aims to explore the potential of deep transfer learning of pretrained image recognition deep neural networks (DNN) in a new diagnostic modality of atrial fibrillation (AF) using Holter ECG color maps. These maps consist of stacked color bars, transforming the ECG waveform amplitude of sequential heartbeats over 30 s into a color code, similar to ECHOView™ (Schiller AG, Switzerland). Using 13,536 training, 6,624 validation and 11,400 test ECHOView color maps evenly split between AF and non-AF cases from the IRIDIA-AF Holter ECG monitoring database, we retrained, fine-tuned, and tested 18 image classification DNNs. Test accuracy of retrained DNNs (93.1–95.8%) was improved after fine-tunning (96.3–97.6%). Retrained EfficientNetV2B1 and finetuned MobileNetV2(α=1.4) were both accurate and fast.

ECHOView colormap images, composed of stacked heartbeat amplitudes, provide an interpretable visual modality for pretrained image classification networks that can be effectively retrained and fine-tuned to identify short AF episodes in Holter-ECG recordings.

1. Introduction

Long-term Holter-ECG monitoring is recommended for patients with suspected arrhythmias that may be transient and missed on routine resting 12-lead ECG [1]. Often, identification of short-term arrhythmic events is challenging in 24–72-hour recordings. Our recent clinical case series study on supraventricular arrhythmias [2] showed that cardiologists can rapidly analyse Holter-ECGs and detect transient events by interpreting compressed rhythm patters in ECHOView color maps, displayed by the Medilog DARWIN2 software (Schiller AG, Switzerland) [3]. Initially, the ECHOView-like principle, also referred to as Electrocardiomatrix (ECM) in some studies, was applied to visually check heart-rate variability, as well as visually detect atrial fibrillation (AF), atrial flutter, ischemic stroke, transient ischemic attacks, and congestive heart failure events, according to the review in [2]. Recently, image processing of ECM by 2D convolutional deep neural networks (DNNs) was applied with a focus on AF detection [4–6] and biometric identification [7].

This study aims to explore the potential of deep transfer learning, showing that several pretrained image recognition DNNs can be effectively retrained and fine-tuned to classify AF in Holter ECG recordings, using ECHOView-like color maps of the ECG rhythm over 30s.

2. Materials and methods

2.1. Holter ECG color maps

Data were sourced from the IRIDIA-AF database, a new ECG-Holter monitoring dataset for paroxysmal AF [8], including 167 records of 2-lead ECG (leads I, II) sampled at 200 Hz, with a total duration of 1609 hours (AF) and 6690 hours (total). The ECG signals were bandpass filtered (0.5–40 Hz) and transformed into ECHOView color maps similar to those presented by the Medilog DARWIN2 software (Schiller AG, Switzerland) [3]. From each lead, we generated images of stacked color bars, transforming the ECG waveform amplitude of sequential heartbeats into a “black-blue-white-orange-red” color code, as described in [2]. ECHOView image height (300 pixels) is set by the ECHOView segment resolution (1.5 s at 200 Hz), while the width varies with heart rate, representing the number of heartbeats over the ECG processing interval (30 s). Applying the nearest neighbor image resizing method, the original ECHOView images were rescaled to various fixed sizes suitable for DNN inputs, according to Figure 1.

Using the original IRIDIA-AF database annotations for AF presence and QRS positions, this study generated ECHOView images at random offsets, keeping a balanced 50:50 ratio of AF to non-AF cases, with an equal number of samples per record and lead. We extracted 13,536 training and 6,624 validation samples from the first 72 records (000–071), and 11,400 test samples from the remaining 95 records (072–166). The overall scheme for generating Holter ECG color maps is outlined in Figure 1.

 ISSN: 2325-887X DOI: 10.22489/CinC.2024.283

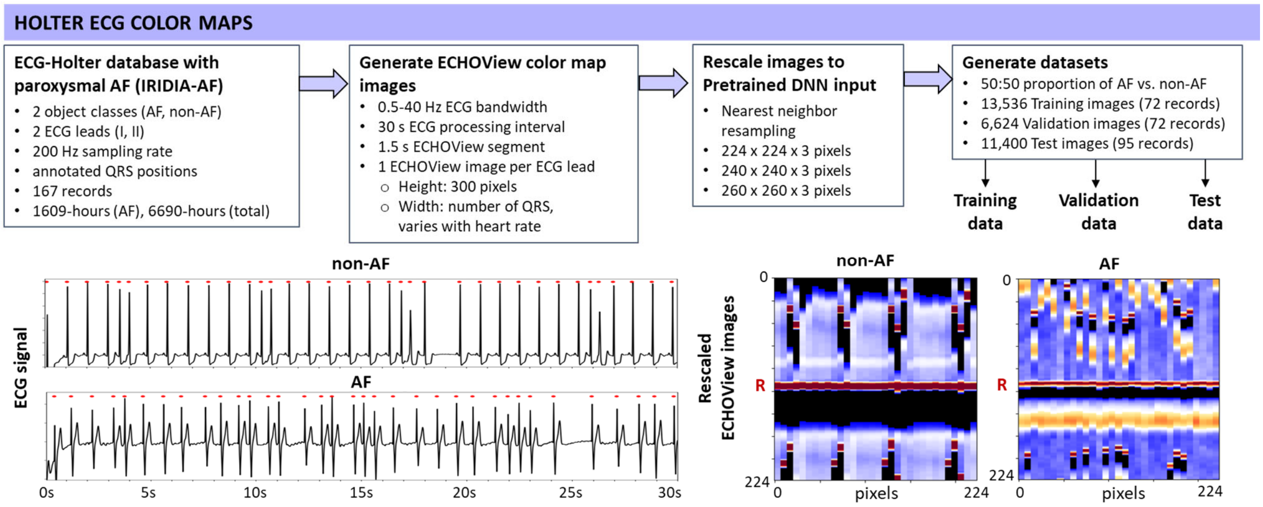

Figure 1. Processing scheme for generating ECHOView color map images from the Holter ECG IRIDIA-AF database.

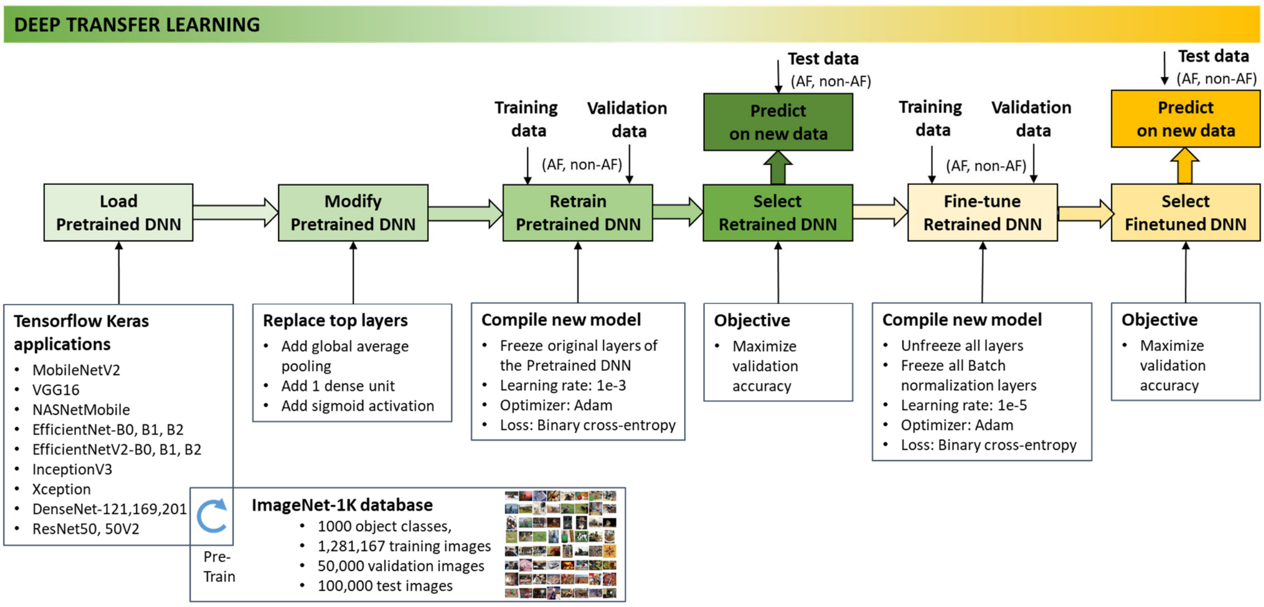

Figure 2. Deep transfer learning applied to retrain and fine-tune specific pretrained ImageNet DNNs for binary arrhythmia classification (AF vs. non-AF) using training, validation, and test data with ECHOView images of Holter ECG (Figure 1).

2.2. Deep transfer learning

This study on deep transfer learning investigates the power of large DNNs that have already been trained to extract information for 1000 object classes in >1 million natural images from the ImageNet-1K dataset created for the Large-Scale Visual Recognition Challenge [9]. To ensure cost-effectiveness, we selected pretrained DNNs [10–17] with limited parameters (<24 M) and input image dimensions (≤260 pixels). Following the workflow in Figure 2, the pretrained DNNs were loaded via Tensorflow Keras applications and adapted to the novel image recognition domain with an input of ECHOView Holter ECG color maps and binary output for the presence and absence of AF (AF, non-AF).

First, the pretrained DNN architectures were modified by replacing the top (classification) layers with a single dense unit and a sigmoid activation that converts the original DNN output from multi-class probabilities to a single probability for binary AF detection. Second, the new top layer of the modified model was trained to classify the training and validation data from our database. This step is named "Retrain Pretrained DNN" in the context that all original (base) layers of the pretrained DNN model were frozen to preserve the learned information from the source domain. Notably, training only a few top-layer weights is much faster and easier than training the entire DNN from scratch, given the limited data available for such large models. DNN retraining was done at the default learning rate (1e-3) while minimizing binary cross-entropy loss. Ultimately, the best retrained DNN model was selected with the objective of maximizing validation accuracy.

The final step applied fine-tuning of the retrained DNNs to gradually adapt all pretrained features with a very low learning rate (1e-5) to the new data. This can potentially achieve meaningful improvements for the training dataset, but there is a risk of overfitting on small-sized data. Therefore, we tested the fine-tuned models on new data to confirm generalizability. It is worth noting that batch normalization layers were not included in fine-tuning. Freezing these layers ensured that the statistics learned from the large original dataset remained intact and was not adversely affected by the smaller training dataset.

4. Results and discussion

ECG signal processing and deep transfer learning were implemented in Python 3.9.5 using Keras and Tensorflow 2.9.1 on a Persy Stinger workstation equipped with an NVIDIA RTX A5000 24 GB GPU. Eighteen pretrained DNNs were retrained and fine-tuned for up to 450 epochs, using early stopping after validation loss degradation. Figure 3 illustrates the training process, showing that the validation accuracy for all models peaked at 94.5–97.4% during retraining (with 513 to 2048 trainable parameters) and further increased to 98–98.8% with fine-tuning (slow training of 0.38M to 23.48M parameters). Detailed performances for the test dataset are presented in Table 1. They indicate that retrained DNNs provide 93.1–95.8% accuracy on new data and confirm the positive effect of fine-tunning with an additional accuracy improvement (up to 4%), reaching 96.3–97.6%. Fine-tuning improves the true positive rate the most (88.2–93.9% to 93.4–95.8%), and the true negative rate less (96.4–98.7% to 98.9–99.5%). Retrained and finetuned DNNs demonstrate robustness on independent test data with a tolerable accuracy drop (<1.5%) compared to the validation data.

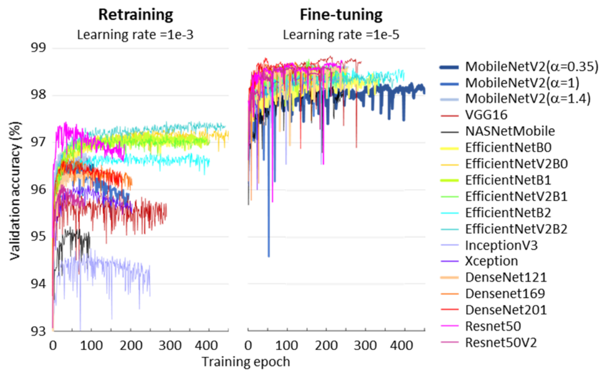

Figure 3. Retraining and fine-tuning of pretrained ImageNet DNNs on ECHOView images for AF detection.

Table 1. List of 18 pretrained DNNs used in this study, and the test AF detection performance of their retrained and finetuned versions. D: Input image size (DxDx3); TPR: True positive rate; TNR: True negative rate; Acc: Accuracy; Depth: number of layers or stages in the architecture, that might include different combinations of convolutional, depthwise convolutional and fully connected layers; α: width multiplier, scaling the number of filters in each layer of MobileNetV2.

Pretrained DNNs			**Retrained DNNs**				**Finetuned DNNs**			
Name	**Input D**	**Total Params (Depth)**	**Train Params**	**TPR (%)**	**TNR (%)**	**Acc (%)**	**Train Params**	**TPR (%)**	**TNR (%)**	**Acc (%)**
MobileNetV2(α=0.35) [10]	224	0.41 M (53)	1281	92.0	98.0	95.0	0.38 M	93.7	98.9	96.3
MobileNetV2(α=1) [10]	224	2.26 M (53)	1281	91.2	97.6	94.4	2.19 M	94.1	99.1	96.6
MobileNetV2(α=1.4) [10]	224	4.37 M (53)	1793	89.3	96.9	93.1	4.27 M	95.4	99.5	97.4
VGG16 [11]	224	14.72 M (16)	513	91.1	96.4	93.7	14.72 M	95.1	99.2	97.1
NASNetMobile [12]	224	4.27 M (88)	1057	88.2	97.5	92.9	4.20 M	93.4	99.1	96.3
EfficientNetB0 [13]	224	4.05 M (16)	1281	91.6	98.2	94.9	3.97 M	93.6	99.4	96.5
EfficientNetV2B0 [13]	224	5.92 M (11)	1281	92.2	98.5	95.3	5.78 M	94.0	99.1	96.5
EfficientNetB1 [13]	240	6.58 M (18)	1281	92.1	98.4	95.3	6.45 M	94.5	99.4	96.9
EfficientNetV2B1 [13]	240	6.93 M (16)	1281	93.9	98.7	96.3	6.79 M	95.3	99.4	97.3
EfficientNetB2 [13]	260	7.77 M (18)	1409	91.7	98.3	95.0	7.63 M	94.6	99.4	97.0
EfficientNetV2B2 [13]	260	8.77 M (19)	1409	93.3	98.3	95.8	8.61 M	93.7	99.3	96.5
InceptionV3 [14]	224	21.80 M (48)	2049	89.8	96.8	93.3	21.75 M	95.2	99.4	97.3
Xception [15]	224	20.86 M (71)	2049	91.2	97.3	94.2	20.75 M	94.4	99.0	96.7
DenseNet121 [16]	224	7.04 M (121)	1025	89.1	98.3	93.7	6.87 M	95.6	99.3	97.5
Densenet169 [16]	224	12.64 M (169)	1665	89.7	98.3	94.0	12.33 M	95.1	99.4	97.2
DenseNet201 [16]	224	18.32 M (201)	1921	91.6	98.7	95.1	17.87 M	95.8	99.5	97.6
Resnet50 [17]	224	23.59 M(50)	2049	92.1	98.4	95.2	23.48 M	95.4	99.3	97.3
Resnet50V2 [17]	224	23.57 M (50)	2049	89.8	97.6	93.7	23.48 M	94.3	99.3	96.8

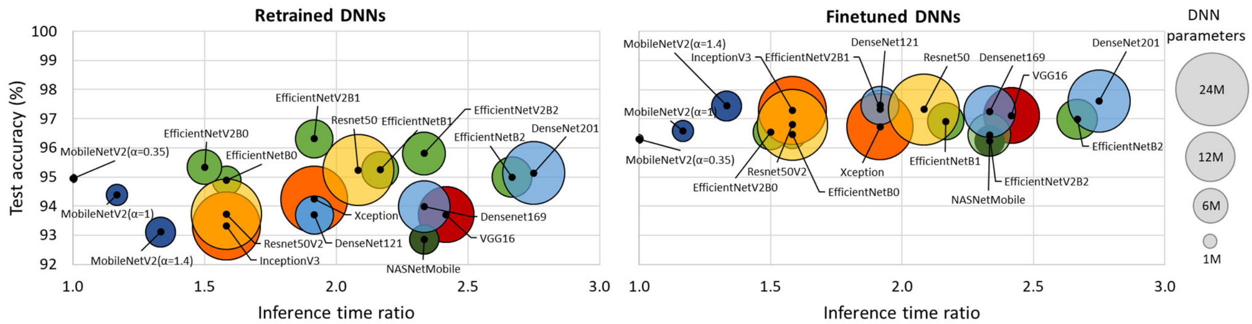

Figure 4. Comparison of retrained and finetuned DNNs in respect of their test accuracy for AF detection (y-axis) vs. computational cost, represented by the inference time ratio (the ratio of a model's inference time relative to the fastest model) (x-axis), and DNN parameters (bubble size).

Figure 4 compares the AF detection accuracy of all DNNs against their computational cost, which varies by up to 3x in inference time and ranges from 0.4M to 24M in parameters. EfficientNetV2B1, with a moderate computational cost (1.92), stands out as the most accurate retrained DNN (96.3%). The finetuned MobileNetV2(α=1.4) is both accurate (97.4%) and fast (1.33).

In conclusion, this study demonstrates the potential of deep transfer learning, comparing several retrained and fine-tuned image recognition DNNs to detect AF by using an input of ECHOView color maps, representative to short-duration AF rhythm in Holter ECG recordings.

Acknowledgments

This work was supported by the Bulgarian National Science Fund, grant number КП-06-Н42/3.

References

[1] G. Hindricks et al., “ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC”, *Eur Heart J*, vol. 42, pp. 373–498, 2021.

[2] S. Naydenov et al., “Recognition of supraventricular arrhythmias in holter ECG recordings by ECHOView color map: A case series study”, *J Cardiovasc Dev Dis*, vol. 10, Art. no. 360, 2023.

[3] Schiller’s Medilog Holter System. 2022. Available online: https://www.schiller.ch/schiller_images_pdfs/devices/holter/medilog-darwin2/2.500548_2022-03_en_medilog_holter-system_pp_lq.pdf (accessed on 22 August 2024).

[4] R. Salinas-Martínez et al., “Automatic detection of atrial fibrillation using electrocardiomatrix and convolutional neural network”, *Comput Cardiol*, vol. 47, doi: 10.22489/CinC.2020.170, pp.1–4, 2020.

[5] R. Salinas-Martínez et al., “Detection of brief episodes of atrial fibrillation based on electrocardiomatrix and convolutional neural network”, *Front Physiol*, vol. 12, Art. no. 673819, 2021.

[6] H. Lee and M. Shin, “Learning explainable time-morphology patterns for automatic arrhythmia classification from short single-lead ECGs”, *Sensors*, vol. 21, Art. no. 4331, 2021.

[7] C. Fuster-Barceló et al., “ELEKTRA: ELEKTRokardiomatrix application to biometric identification with convolutional neural networks”, *Neurocomput*, vol. 506, pp. 37–49, 2022.

[8] C. Gilon et al., “IRIDIA-AF, a large paroxysmal atrial fibrillation long-term electrocardiogram monitoring database”, *Sci Data*, vol. 10, Art. no. 714, 2023.

[9] O. Russakovsky et al., “ImageNet large scale visual recognition challenge”, *Int J Comput Vision*, vol. 115(3), pp. 211–252, 2015.

[10] M. Sandler et al., “MobileNetV2: Inverted residuals and linear bottlenecks”, in *Proc. IEEE Comput Soc Conf Comput Vis Pattern Recognit*, pp. 4510–4520, 2018.

[11] K. Simonyan and A. Zisserman. “Very deep convolutional networks for large-scale image recognition”, in *3rd Int Conf Learn Repres (ICLR)*, pp. 1–14, 2015.

[12] B. Zoph et al., “Learning transferable architectures for scalable image recognition”, in *Proc. IEEE Comput Soc Conf Comput Vis Pattern Recognit*, pp. 8697–8710, 2018.

[13] M. Tan and Q.V. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks”, in *Proc 36th Int Conf Mach Learn*, pp. 6105–6114, 2019.

[14] C. Szegedy et al., "Rethinking the inception architecture for computer vision." in *Proc. IEEE Comput Soc Conf Comput Vis Pattern Recognit*, pp. 2818–2826, 2016.

[15] F. Chollet, “Xception: Deep learning with depthwise separable convolutions”, in *Proc. IEEE Comput Soc Conf Comput Vis Pattern Recognit*, pp. 1800–1807, 2017.

[16] G. Huang et al., “Densely connected convolutional networks”, in *Proc. IEEE Comput Soc Conf Comput Vis Pattern Recognit*, pp. 2261–2269, 2017.

[17] K. He et al., “Deep residual learning for image recognition”, in *Proc. IEEE Comput Soc Conf Comput Vis Pattern Recognit*, pp. 770–778, 2016.

Address for correspondence:

Irena Jekova
Institute of Biophysics and Biomedical Engineering
Acad. G. Bonchev str., bl.105, 1113, Sofia, Bulgaria
irena@biomed.bas.bg