SUPPLEMENTAL MATERIAL

Expanded Methods

Data acquisition

All ECGs were obtained using a General Electric MAC 5500 (GE Healthcare, Chicago, IL, United States) as part of routine clinical care. Using the MUSE ECG system (MUSE version 8, GE Healthcare, Chicago, IL, United States), we extracted raw 10 second 12-lead ECG data waveforms. From the raw data files, the median beats, consisting of 600 samples, sampled at 500 Hz, were used for further analysis.

Physician annotations were extracted from the ECGs using the text mining algorithm described above. In case of any form of annotated T-wave abnormalities, the ECGs were manually reviewed by two reviewers to determine presence and location of inverted T-waves. Disagreement was resolved by panel discussion. Inverted T-waves were defined as being more than 0.1 mV negative in lead I, II, aVL or V2-V6 in the absence of right bundle branch block. The peak-to-peak voltage of the QRS-complex was extracted from the median beat and used to determine the presence of low QRS voltage, which was defined as a peak-to-peak QRS voltages less than 0.5 mV in the extremity leads or less than 1 mV in the precordial leads. Conduction intervals were extracted from the MUSE ECG system.

Feature visualization

The final convolutional layer was used for visualization and all values for the guided backpropagation maps below zero were discarded. An overall visualization of important ECG segments was constructed with the following steps: (1) all median ECG beats and their corresponding per-patient normalized Guided Grad-CAM maps were aligned temporally by normalizing the PQ and QT intervals, (2) the mean and standard deviations of the ECG signal were derived within each group, (3) the proportion of the per-patient Guided Grad-CAM

maps above a threshold was calculated for each timepoint and plotted as a superimposed heatmap and (4) the overall heatmap was filtered using a 2nd order Savinsky-Golay filter. The threshold for an important feature was set at 5%. Only correctly classified PLN patients were used for feature visualization.

Deep neural network architecture

We constructed a deep convolutional neural network with exponentially dilated causal convolutions (Supplemental Figure 1). Based on the method described by Van Oord *et al*. and Franceschi *et al*., we built an architecture composed of several 1-dimensional causal convolution blocks.[1,2] These were followed by a 1D global max pooling layer squeezing the temporal dimension and finally a linear layer transforming the squeezed temporal information to one output logit. Each causal convolution block consists of a combination of causal convolutions, weight normalizations, leaky ReLUs and residual connections.[3-5] The dilation parameter used in the convolutional layer is exponentially doubled in each subsequent causal convolution block from 1 to 64. For the first 6 blocks, the number of output channels of the convolutional layers is kept constant at 128 and the final 7th block outputs 256 channels. The convolutional layer in the residual connection used a kernel size of 1 and was only applied in the case where the number of output channels differed from the number of input channels. All other convolutional layers used a kernel size of 3 and a value of 0.01 was used for the negative slope parameter of the leaky ReLU activation functions.

For training, we optimized the network parameters using a weighted binary focal loss function to handle class imbalance, and Adam with a learning rate of 0.0001 as the optimization algorithm.[6,7] This loss function reshapes the standard binary cross entropy such that it down-weights the loss assigned to well-classified examples. In addition, we assign a

weight to positive examples equal to the class imbalance ratio to force equal attribution to the loss of both class examples. The used batch size was 128. Network training was performed using the PyTorch package (version 1.3).[8]

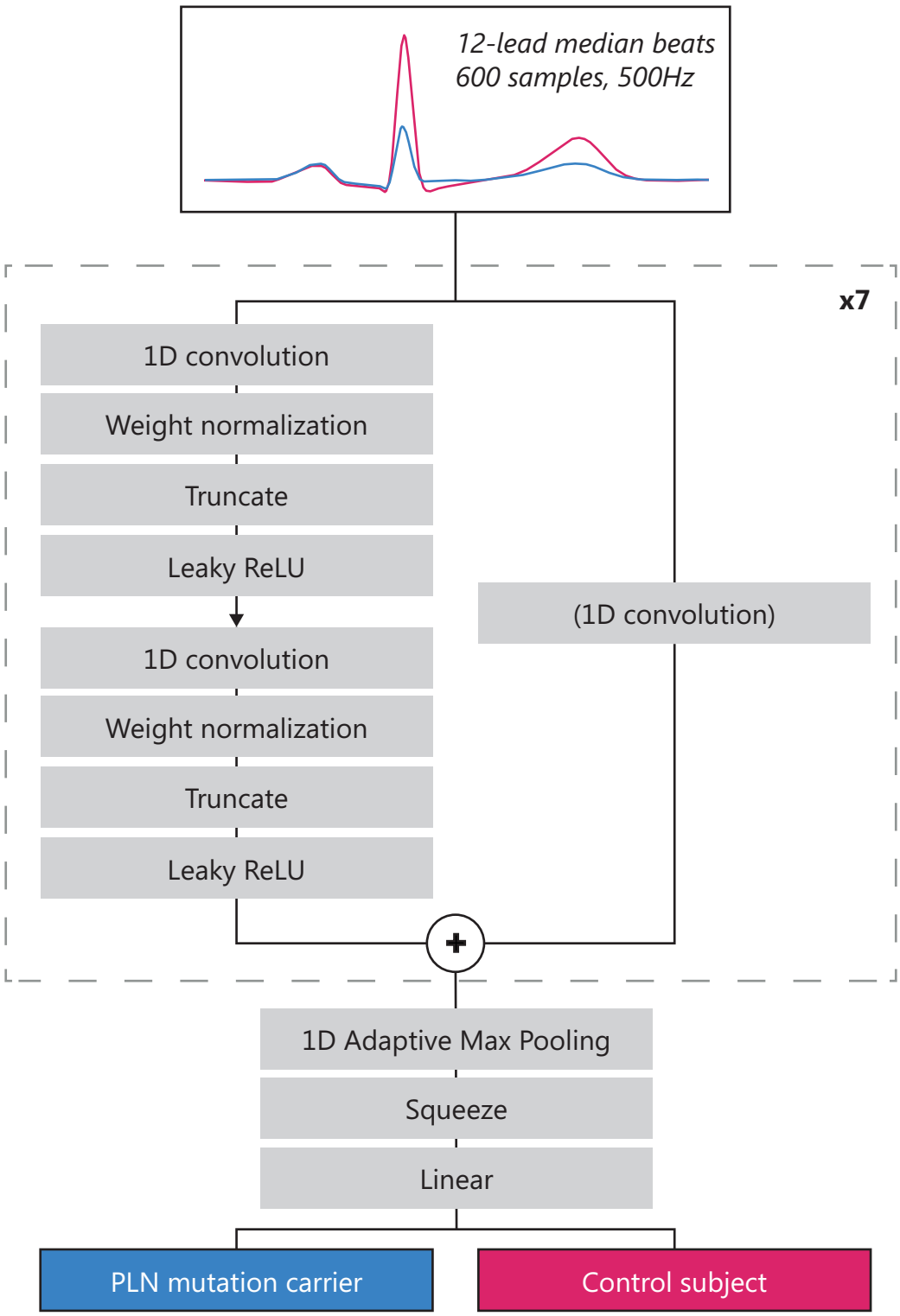
Eight-fold cross validation on the training dataset was used for optimization of the hyperparameters of the network. The following hyperparameters were optimized using a combination of grid search and manual tuning: learning rate, network depth, channel size and kernel size. Early stopping was performed when the validation loss did not decrease for 20 epochs. The simplest network with the highest geometric mean of area under the receiver operating curve and F2 score averaged over all folds was chosen and trained on the complete training dataset. The performance of this network was estimated on the test subset.

References

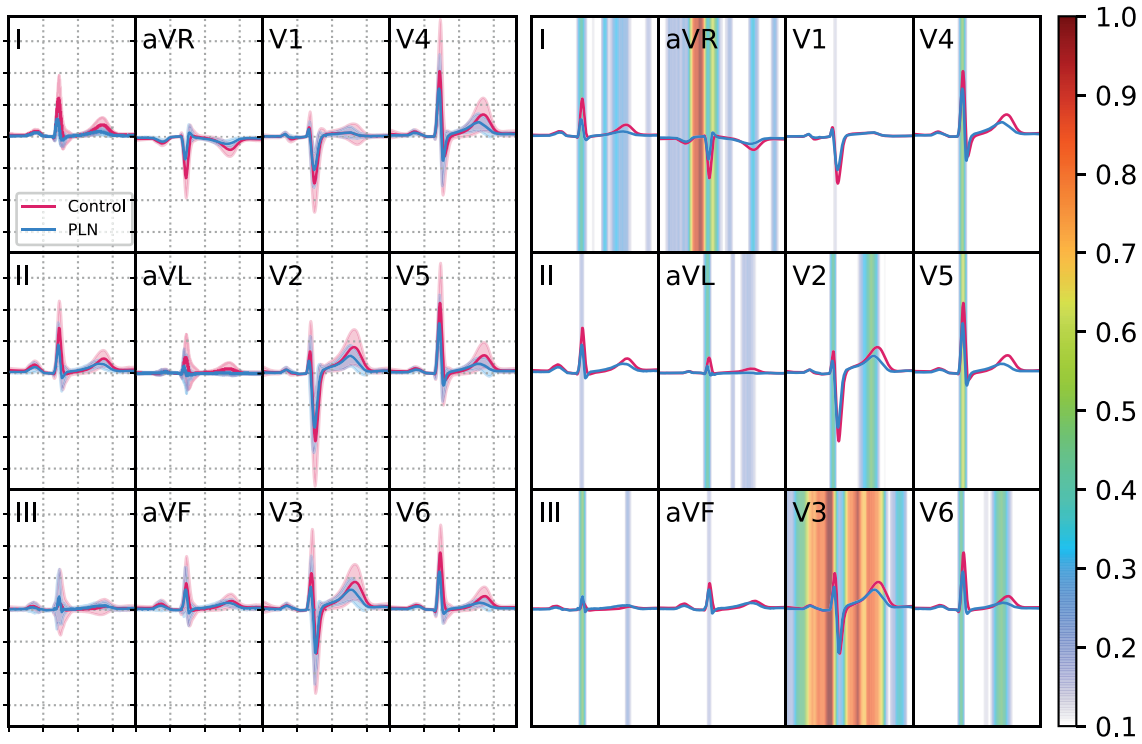
1. Oord A van den, Dieleman S, Zen H, Simonyan K, Vinyals O, Graves A, Kalchbrenner N, Senior A, Kavukcuoglu K. WaveNet: A Generative Model for Raw Audio. 2016;1–15. Available from: http://arxiv.org/abs/1609.03499
2. Franceschi J-Y, Dieuleveut A, Jaggi M. Unsupervised Scalable Representation Learning for Multivariate Time Series. 2019; Available from: http://arxiv.org/abs/1901.10738
3. Salimans T, Kingma DP. Weight normalization: A simple reparameterization to accelerate training of deep neural networks. *Adv Neural Inf Process Syst*. 2016;901–909.
4. Maas AL, Hannun AY, Ng AY. Rectifier nonlinearities improve neural network acoustic models. *ICML Work Deep Learn Audio, Speech Lang Process*. 2013;28.
5. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. *CoRR*.

2015;abs/1512.0.

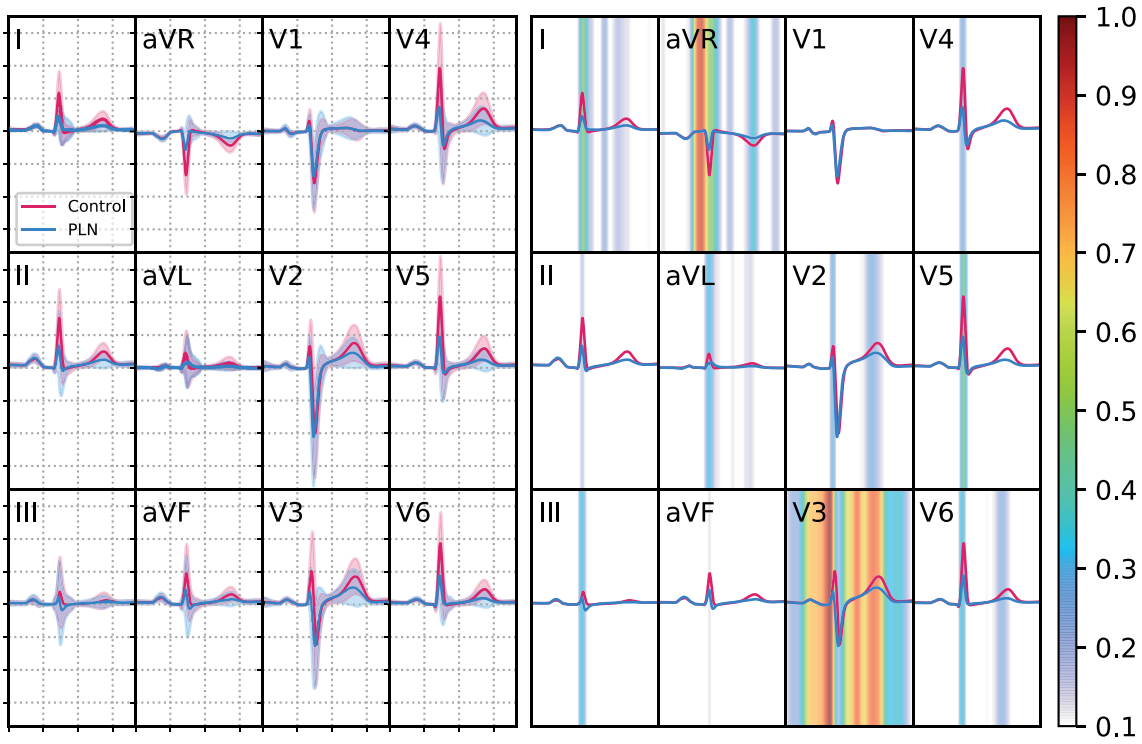
6. Lin TY, Goyal P, Girshick R, He K, Dollar P. Focal Loss for Dense Object Detection. *Proc IEEE Int Conf Comput Vis*. 2017;2017-Octob:2999–3007.
7. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. *AIP Conf Proc* 2014;1631:58–62.
8. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In: Advances in Neural Information Processing Systems. 2019.



Supplemental Figure 1. Overview of the deep neural network architecture used in this paper. Batches of 128 12-lead median ECG beats are used as the input during training. The 1D convolution on the right in only used when up- or downsampling is needed.



Supplemental Figure 2. Output of the Guided Grad-CAM visualization algorithm for presymptomatic PLN mutation carriers (n = 11) and their controls. Left: Mean of temporally normalized median 12-lead ECGs of both the presymptomatic PLN mutation carriers (blue) and control patients (red) with their respective standard deviations. Right: the same median ECG beat with the Guided Grad-CAM output of the DNN superimposed to indicate the importance of a specific temporal segment for the classification of the DNN. The colormap represents the proportion of patients where that region was important (i.e. had a Guided Grad-CAM value above the threshold). Only the feature maps of correctly classified PLN patients are used. Guided Grad-CAM: Guided Gradient Class Activation Mapping, PLN: phospholamban.



Supplemental Figure 3. Output of the Guided Grad-CAM visualization algorithm for symptomatic PLN mutation carriers (n = 75) and their controls. Left: Mean of temporally normalized median 12-lead ECGs of both the symptomatic PLN mutation carriers (blue) and control patients (red) with their respective standard deviations. Right: the same median ECG beat with the Guided Grad-CAM output of the DNN superimposed to indicate the importance of a specific temporal segment for the classification of the DNN. The colormap represents the proportion of patients where that region was important (i.e. had a Guided Grad-CAM value above the threshold). Only the feature maps of correctly classified PLN patients are used. Guided Grad-CAM: Guided Gradient Class Activation Mapping, PLN: phospholamban.