

Discovering and Visualizing Disease-specific Electrocardiogram Features Using Deep Learning: Proof-of-concept in Phospholamban Gene Mutation Carriers

Running title: *van de Leur & Taha et al.; Deep learning-based ECG feature detection in PLN*

Rutger R. van de Leur, MD^{1,2*}; Karim Taha, MD^{1,2*}; Max N. Bos, MSc^{1,3}; Jeroen F. van der Heijden, MD, PhD¹; Deepak Gupta, PhD³; Maarten J. Cramer, MD, PhD¹; Rutger J. Hassink, MD, PhD¹; Pim van der Harst, MD, PhD¹; Pieter A. Doevendans, MD, PhD^{1,2,4}; Folkert W.

Asselbergs, MD, PhD^{1,5}; René van Es, PhD¹



¹Department of Cardiology, University Medical Center Utrecht; ²Netherlands Heart Institute, Utrecht; ³Informatics Institute, University of Amsterdam, Amsterdam; ⁴Central Military Hospital, Utrecht, the Netherlands; ⁵Institute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, London, United Kingdom

*contributed equally

Correspondence:

René van Es, PhD

University Medical Center Utrecht

Heidelberglaan 100

3584 CX Utrecht

the Netherlands

Tel: 0031 88 757 3453

Email: r.vanes-2@umcutrecht.nl

Journal Subject Terms: Electrophysiology; Congenital Heart Disease; Cardiomyopathy; Machine Learning; Genetics

Abstract:

Background - Electrocardiogram (ECG) interpretation requires expertise and is mostly based on physician recognition of specific patterns, which may be challenging in rare cardiac diseases. Deep neural networks (DNN) can discover complex features in ECGs and may facilitate the detection of novel features which possibly play a pathophysiological role in relatively unknown diseases. Using a cohort of phospholamban (PLN) p.Arg14del mutation carriers, we aimed to investigate whether a novel DNN-based approach can identify established ECG features, but moreover we aimed to expand our knowledge on novel ECG features in these patients.

Methods - A DNN was developed on 12-lead median beat ECGs of 69 patients and 1380 matched controls and independently evaluated on 17 patients and 340 controls. Differentiating features were visualized using Guided Grad-CAM++. Novel ECG features were tested for their diagnostic value by adding them to a logistic regression model including established ECG features.

Results - The DNN showed excellent discriminatory performance with a c-statistic of 0.95 (95% confidence interval 0.91-0.99) and sensitivity and specificity of 0.82 and 0.93, respectively. Visualizations revealed established ECG features (low QRS voltages and T-wave inversions), specified these features (e.g. R and T-wave attenuation in V2/V3) and identified novel PLN-specific ECG features (e.g. increased PR-duration). The logistic regression baseline model improved significantly when augmented with the identified features ($p < 0.001$).

Conclusions - A DNN-based feature detection approach was able to discover and visualize disease-specific ECG features in PLN mutation carriers and revealed yet unidentified features. This novel approach may help advance diagnostic capabilities in daily practice.

Key words: ECG; dilated cardiomyopathy; arrhythmogenic right ventricular dysplasia/cardiomyopathy; electrocardiogram; deep neural network; deep learning; feature detection; phospholamban mutation

Nonstandard Abbreviations and Acronyms

AIC: Akaike information criterion

ARVC : Arrhythmogenic right ventricular cardiomyopathy

DCM: Dilated cardiomyopathy

DNN: Deep Neural Network

ECG: Electrocardiogram

Guided Grad-CAM: Guided Gradient Class Activation Mapping

HCM: Hypertrophic cardiomyopathy

LVAD: Left ventricular assist device

PLN: Phospholamban

ROC: receiver operating characteristic



Introduction

Interpretation of the electrocardiogram (ECG) requires expertise and is mainly based on physician recognition of patterns that are known to belong to a particular disorder. However, for rare and relatively unknown cardiac diseases, this may be challenging since ECG features are often unknown and require expert knowledge to recognize. By automating the discovery and expanding the knowledge on disease-specific ECG features, interpretation of ECGs by physicians could be improved. Such a support tool could be of particular importance when expert knowledge is not readily available or in research settings to automate the detection of disease-specific ECG features.

Recently, ECGs have been analyzed using deep neural networks (DNNs), which are computer algorithms that are based on the structure and functioning of the human brain.¹ Their layers can be trained to discover complex patterns in ECGs, without requiring hand-crafted

feature extraction. Several studies have applied DNNs for automated predictions from ECGs, and one recent study showed that it is feasible to diagnose hypertrophic cardiomyopathy (HCM) on the ECG.²⁻⁴ However, the methods used in these studies all require very large datasets, which are often not available for rare diseases. Furthermore, these previous studies all focus on prediction, but specific ECG patterns used by DNNs are rarely visualized.^{3,5-8} Visualization of such features takes advantage of the feature discovery embedded in DNNs and will help clinicians to interpret ECGs more accurately, and possibly facilitate discovery of novel features.

Cardiomyopathy-related genetic mutations are rare but are often associated with typical ECG features. An example is the deletion of three base pairs (c.40_42delAGA) in the phospholamban (*PLN*) gene, leading to the deletion of Arginine 14 in the PLN protein (p.Arg14del).⁹⁻¹¹ Prevalence of the PLN p.Arg14del mutation is estimated to be 0.07% in the northern regions of the Netherlands and is present in 12% of Dutch patients developing a phenotype of arrhythmogenic right ventricular cardiomyopathy (ARVC) and in 15% of patients developing dilated cardiomyopathy (DCM).¹¹⁻¹³ With regard to ECG characteristics in these mutation carriers, typical features that have previously been described are attenuated QRS-amplitudes and inverted T-waves in the right and left precordial leads.^{12,14,15}

Beside using DNNs merely for prediction or diagnosis, we hypothesize that DNNs can also be used for feature visualization itself. This will potentially enable discovery of novel ECG features that belong to a particular disease. In this study, we used a cohort of PLN mutation carriers to investigate whether a novel DNN-based approach can (i) identify the already well-established ECG features in these mutation carriers and (ii) possibly expand our knowledge on ECG features in these mutation carriers.

Methods

Data availability

The data used in this study are not publicly available due to privacy restrictions. The code for training the DNN and for generating the visualizations and tables in this paper is available upon request from the corresponding author.

Data source and study participants

The dataset consisted of 12-lead ECGs from patients between 18 and 85 years old acquired in the University Medical Center Utrecht (UMCU) from January 2000 to August 2019. All extracted data were de-identified in accordance with the EU General Data Protection Regulation and written informed consent was therefore not required by the ethical committee. All ECGs were interpreted by a physician as part of the clinical workflow and these free text annotations were structured using a text mining algorithm described before.³ We excluded all ECGs of insufficient quality and all ECGs with supraventricular and ventricular arrhythmias (excluding premature atrial and ventricular complexes), paced rhythms, undefined rhythms and signs of acute ischemia.

All index patients in the dataset who carry the genetic PLN p.Arg14del mutation and their relatives that tested positive, were identified. ECGs acquired after the implantation of a left ventricular assist device (LVAD) or heart transplantation were excluded. Only the first acquired ECG of each mutation carrier was used for development of the model.

The control group was derived from the remaining dataset and consisted of 365,173 ECGs of 147,098 unique patients. Per mutation carrier, 20 controls were matched using propensity score matching on age and sex. This number was chosen to have sufficient samples to train the DNN without having a too severe class imbalance. Only one ECG per control subject,

sampled without replacement, was used to make sure every subject was only used once. The matched groups were randomly split in an 80:20 manner to training and test sets.

Data acquisition

For all ECGs the median beats were exported from the MUSE ECG system (MUSE version 8, GE Healthcare, Chicago, IL, United States). The median beat data is constructed by aligning all QRS-complexes of the same shape (e.g. excluding premature ventricular complexes) and generating a representative QRS-complex by taking the median voltage.¹⁶ Acquisition and feature extraction of the included ECGs is described in more detail in the Supplemental Material.

Baseline logistic regression model

To demonstrate the capability of DNN in identifying novel relevant features, we first developed a baseline logistic regression model, only based on the established ECG features of PLN mutation carriers. The matching variables, age and sex, and the established PLN-specific ECG features (low QRS voltage and right (V2-V3) and left (V4-V6) precordial T-wave inversion) were included as predictors in the model.¹⁷ The model was trained on the training dataset and evaluated on the test set.

Deep neural network development

We constructed a deep convolutional neural network with exponentially dilated causal convolutions. The proposed architecture, inspired by the method described by Van Oord *et al.* and Franceschi *et al.*, comprises of several 1-dimensional dilated causal convolution blocks.^{18,19} Eight-fold cross validation on the training dataset was used for optimization of the hyperparameters of the network. 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. Network training was performed using the PyTorch package (version 1.3).²⁰ A detailed description of the architecture of the DNN can be found in the Expanded Methods and an overview of the network architecture is shown in Supplemental Figure 1.

Feature visualization

To identify the parts of the ECG that are important for the DNNs prediction, we applied Guided Gradient Class Activation Mapping ++ (Guided Grad-CAM++), a technique for explanations in convolutional neural networks, to 1-dimensional data.^{5,6} Guided Grad-CAM++ combines the fine-grained and lead-specific visualizations of guided backpropagation with the class-discriminative and global Grad-CAM technique. The median beat visualization methodology is described in more detail in the Supplemental Material.



Validation of newly identified features in an updated model

Based on inspection of the visualization output, we identified distinctive features with an arbitrary prevalence above 25%. The detected important features were translated to quantitative features (e.g. R-wave amplitude) and added to the baseline logistic regression model, starting with the most prevalent. If multiple similar features were found in leads belonging to the same region, the most prevalent feature in that region was used. Leads I, aVL and V4-V6 were grouped as lateral leads and II, III and aVF as inferior leads. To evaluate the added value of the newly identified ECG features, we determined if the nested baseline logistic regression model fit improved using the likelihood ratio test (LRT) and Akaike's information criterion (AIC).

Subgroup analyses

In subgroup analyses, we analyzed whether predictive performance and detected features differed between subsets of patients. Due to the small sample size, these exploratory subgroup analyses were performed on the combined training and test datasets. We investigated the

performance in presymptomatic PLN p.Arg14del mutation carriers. Presymptomatic was defined as: no cardiac symptoms as per judgement of the treating physician, no history of (non-)sustained ventricular arrhythmia, premature ventricular complex burden of <500 beats per 24 hours and left ventricular ejection fraction $\geq 45\%$.

Statistical analysis

The baseline characteristics were expressed as mean \pm standard deviation or median with interquartile range (IQR), where appropriate. Categorical variable differences were tested using the chi-square test or Fisher's exact test and continuous variables using the Student's t-test or Mann Whitney U test. Multiple testing correction was performed for the baseline characteristics using Bonferroni's method. The overall discriminatory performance of the DNN, baseline and updated models were assessed in the test set with the concordance-statistic (c-statistic) or area under the receiver operating characteristic (ROC) curve, sensitivities, specificities, positive and negative predictive values. The models were compared at a prespecified specificity of 94%. The 95% confidence intervals (CI) around the performance measures and odd's ratios were obtained using 2000 bootstrap samples. All statistical analyses were performed using R version 3.5 (R Foundation for Statistical Computing, Vienna, Austria).

Results

Study population

A total of 93 PLN p.Arg14del mutation carriers were identified, of which 86 were eligible for this study. Four patients were excluded as all their ECGs were acquired after LVAD or heart transplantation and three patients as all their ECGs were non-sinus rhythm. The control group

consisted of 135,353 patients after exclusions, of which 1,720 patients were matched. The flowchart is shown in Figure 1 and the baseline characteristics in Table 1.

Baseline logistic regression performance

The discriminative performance (by c-statistic) of the baseline logistic regression model was 0.84 (95% CI 0.73-0.92) in the test set. The most important predictor of the PLN mutation was the presence of low QRS voltage, followed by left precordial inverted T-waves. No significant effect of age, gender or right precordial negative T-waves was found.

Deep neural network performance

The cross-validated mean c-statistic, sensitivity, specificity and F2 score obtained in the training dataset were 0.86 ± 0.07 , 0.73 ± 0.13 , 0.91 ± 0.04 and 0.56 ± 0.05 , respectively. The c-statistic of the DNN, trained on the complete training dataset, was 0.95 (95% CI 0.91-0.99) in the independent test set. The mean ECG beats for the complete dataset with a superimposed Guided Grad-CAM visualization can be found in Figure 2. Figure 3 shows a representative example of a mutation carrier and a control subject that shows similar pre-established features (low QRS voltage and inverted T-waves) but is correctly identified by the DNN.

Feature detection

Based on the Guided Grad-CAM maps (Figure 2), we identified the following six most prevalent combined ECG segments: (i) R-waves in V2/V3 (58-99%), (ii) PR-interval (98%), (iii) T-waves in V2/V3 (36-89%), (iv) R-waves in I/aVL/V4-V6 (34-59%), (v) R-waves in II/III/aVF (22-46%) and (vi) T-waves in I/aVL/V6 (22-36%). Figure 4 shows correlation between the Grad-CAM maps and the human interpretation, on an individual level.

After inspection of the median beat and its standard deviation at these locations, the following most prevalent features per region were extracted from the ECG and added to the

baseline logistic regression model: (i) maximum R-wave amplitude in V3, (ii) PR interval, (iii) T-wave peak voltage in V3, (iv) maximum R-wave amplitude in V6, (v) maximum R-wave amplitude in III and (vi) T-wave peak voltage in I.

The updated logistic regression model's c-statistic was 0.91 (95% CI 0.83 - 0.97). The significantly associated baseline variables low QRS voltage and inverted left precordial T-waves remained significant in the updated model. The newly identified features were maximum R-wave amplitude in V3 and V6, the T-wave amplitude in I and V3 and the PR interval. The updated model had a better fit than the baseline model with an AIC of 388, compared to 461 for the baseline model (LRT $p < 0.001$). The performance measures of all three models are shown in Table 2. The odds ratios of the variables in the baseline and updated models are appreciated in Table 3. The summary measures for the quantitative translations of the newly identified features, that are added to the baseline logistic regression model, are shown in Table 4.

Subgroup analyses

Performance was higher for symptomatic than presymptomatic patients, with c-statistics of 0.97 (95% CI 0.95-0.98) and 0.95 (95% CI 0.91-0.98), respectively. Sensitivity was 86% for symptomatic patients (n = 75) and 64% for presymptomatic patients (n = 11), at a similar specificity of 94%. Guided Grad-CAM maps showed a difference in features between symptomatic and asymptomatic patients, where the prolonged PR-interval, attenuated R- and T-wave in V3 and attenuated T-wave in V6 were more important in presymptomatic patients while the overall attenuated R-waves were more prominent in symptomatic patients. The Guided Grad-CAM maps for presymptomatic and symptomatic patients are shown in Supplemental Figures 2 and 3.

Discussion

In this study we demonstrate a novel DNN-based end-to-end approach that allows for detection and visualization of disease-specific ECG features. To the best of our knowledge, this is the first time DNNs have successfully been applied as an ECG feature detector, in contrast to previously developed prediction algorithms. Using a unique combination of median ECG beats and visualizations, the algorithm was able to automatically reveal established ECG features in PLN p.Arg14del mutation carriers (low QRS voltages and T-wave inversions), specify these features (R and T-wave attenuation in V2 and V3) and find novel features (increased PR-duration). Applying this promising concept in more cardiac diseases (especially rare or unknown ones) can potentially support physicians while reviewing ECGs, thereby improving ECG interpretation in daily clinical practice.

Previous literature

Several studies showed that DNNs can be used to make predictions from ECGs with a high performance.^{2-4,7,8} An example is the recent study by Ko *et al.*, who developed a DNN to detect HCM, resulting in an AUC of 0.96.⁴ From a clinical point of view, this network is very attractive, because this would allow the clinician to easily and automatically distinguish HCM in a screening setting. However, clinical implementation of such a network is still challenging for several reasons. Firstly, such networks are often seen as “black boxes”, and, secondly, the validity of these high-dimensional networks in external datasets is still unproven.

Similarly, we developed a DNN that recognizes ECGs of a specific patient population (i.e. PLN mutation carriers) with high diagnostic performance.⁴ A different architecture was chosen, as it has an increased diagnostic performance in PLN mutation carriers and allows for

more detailed visualizations. Unique to our study is the use of hard outcome data and the focus on feature detection, which may directly support clinicians with ECG interpretation in daily practice. Moreover, we show that these features can be used in a relatively simple logistic regression model, which might be easier generalizable.

Disease-specific ECG features in PLN mutation carriers

This novel approach was validated in PLN mutation carriers, because typical ECG characteristics in these subjects have been described extensively before.^{10,12,14,15} PLN mutation carriers are at risk of developing an often biventricular phenotype of ARVC and/or DCM, and are typically characterized by subepicardial fibro-fatty replacement.²¹ This leads to an ECG with low QRS voltages, which can be seen both in the limb leads and in precordial leads.^{10,14} In addition, negative T-waves were previously described in both the right precordial leads and in the left precordial leads.^{12,15}

Using this novel approach, we could correctly identify all of these previously described ECG features (Figure 2) and show that the network also uses the pre-established features for diagnosis (Figure 4). In addition, we could specify the leads in which these features are typically present. With the visualization tool, we found attenuated R-waves to be particularly present in the lateral leads I, aVL and V6, and in the right precordial leads V2 and V3. While the low voltages in these mutation carriers are often measured as QRS peak-to-peak amplitude, we observed that these low voltages were only based on R-wave attenuation, while the S-wave seemed unaltered. Furthermore, we found attenuated/inverted T-waves to be typically present in leads V2, V3 and V6 (as described previously), but also in leads I and aVL. Besides the ECG characteristics that were already identified before in PLN mutation carriers, we also found an ECG feature, the PR-interval, that was not described before in these subjects. This was

confirmed in the updated logistic regression model. Interestingly, a recent meta-analysis of genome-wide association studies also showed an association between a locus in the PLN gene and PR-interval, which already suggested that PLN plays a role in atrio-ventricular conduction.²²

In an exploratory analysis, the DNN performed well in both presymptomatic and symptomatic mutation carriers. Our approach also suggested that particular features were more important in presymptomatic mutation carriers (PR-interval and R- and T-wave attenuation in V2 and V3), as compared to symptomatic carriers. This might indicate that our approach can be used in subgroups who are in different stages of a disease, to gain knowledge on the sequence in which ECG abnormalities naturally occur. In particular for PLN mutation carriers, it is important to gain knowledge on the first electrical changes, because this may improve early screening and risk stratification of presymptomatic mutation-carrying family members.

Employed methodology

The use of DNN for the analysis of data generally requires large amounts of balanced data but the group of PLN mutation carriers studied in this investigation contained only 86 patients. The focus on features detection instead of prediction in this paper, however, allowed the use of such small datasets, as we were able to reduce the highly dimensional DNN to a few important features. Moreover, to allow training on this extremely imbalanced dataset, while also correcting for age and sex differences between mutation carriers and controls, we applied propensity score matching.

In the present study we used ECG median beats as input for the DNN model, which allowed the network to focus on morphology rather than rhythm. The use of median beats prohibits detection of rhythm specific ECG features (e.g. premature contractions or heart rate

variability) and can also not be used for detection of beat-to-beat ECG variations. To our best knowledge, this is the first study in which median beats are used for deep learning.

Limitations

Firstly, although the proposed approach is feasible in small datasets, care should be taken while interpreting results derived from small cohort studies as findings may not hold up when evaluated on other cohorts. Especially the number of patients in the test set is a major limitation. To show clinical applicability of the ECG features and algorithm as described in this study, external validation studies should be performed. Secondly, for the PLN mutation carriers the clinical phenotype may be variable among mutation carriers. Therefore, it should be noted that this approach helps to visualize the most common ECG features on a group level, but important ECG features that are present in small subgroups may be missed. Subgroup analyses in more homogeneous subgroups (e.g. presymptomatic relatives) can be used to reveal important features in these specific subgroups. Thirdly, the ECGs of the control group were extracted from a large database in which additional patient specific characteristics are not available. Therefore, no comparisons or matching between both groups were possible to correct for other influencing factors. However, the case-control matching ratio of 1:20 used in this study presumably equalized the groups, and the detected features align with literature on other PLN mutation carriers. Fourthly, the conduction intervals and P-, QRS- and T-wave boundaries are based on the automated GE algorithm, which might cause inaccuracies. Boundary measurements on median beats have proven to be very accurate, however.²³ Fifthly, the proposed approach is not possible for ECGs with arrhythmias or acute ischemia, as these (temporary) conditions have a large influence on the morphology of the median beat. In this study, the algorithm is not intended to be used in these situations and only 3 patients were excluded for this reason. Finally, the

visualization technique used in this paper, Guided Grad-CAM++, is one of the most frequently used techniques for fine-grained heatmaps but has limitations of its own.⁶ For example, guided backpropagation might be independent on the choice of the model or data generating process.²⁴ Therefore, we validated the detected features in a logistic regression model and showed that Grad-CAM++ values agree with the pre-established PLN ECG features. Feature visualization in DNNs is a new and developing field and future research should focus on improving visualization techniques for DNNs and applying them in ECGs.


Future perspectives

Future studies should be conducted applying this novel approach to other less well characterized diseases, such as new genetic mutations, to discover novel ECG characteristics. The visualizations provide the end-user with feedback on the importance and location of detected ECG features. Moreover, future studies should consider elucidating the pathophysiological mechanisms of newly identified ECG features by using other experimental methods such as (non-)invasive electrophysiological mapping. The influence of the discovered ECG features on disease penetrance in asymptomatic carriers or progression of disease in symptomatic carriers should be examined with longitudinal ECG or outcome data. Finally, combining our approach and a DNN trained on other cohorts with a focus on screening, such as family members of mutation carriers or large healthy population cohorts, might be of interest in clinical practice. Detection and visualization of possible carrier status in the ECG even before the genetic diagnosis is done could determine which family members or healthy individuals require genetic testing or follow-up.

This study demonstrated a novel DNN-based end-to-end approach that allows for detection and visualization of disease-specific ECG features. In a cohort of PLN p.Arg14del

mutation carriers, the algorithm showed excellent diagnostic performance and revealed already established ECG features. Moreover, we were able to specify these features and to detect novel features. This novel way to use DNNs may help advance diagnostic capabilities in daily practice, especially in rare and new cardiac diseases.

Acknowledgments: We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research. Folkert Asselbergs is supported by UCL Hospitals NIHR Biomedical Research Centre.

Sources of Funding: This study was financed by the The Netherlands Organisation for Health Research and Development (ZonMw) with grant number 104021004, the Dutch Heart Foundation and the PLN Genetic Heart Disease Foundation. Part of this work is funded by  American Heart Association. Netherlands Cardiovascular Research Initiative, an initiative with support of the Dutch Heart Foundation (grant number CVON2015-12 eDETECT).

Disclosures: None

References:

1. Goodfellow I, Bengio Y, Courville A. Deep Learning. The MIT Press; 2016.
2. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, Ng AY. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med.* 2019;25.
3. Van de Leur RR, Blom LJ, Gavves E, Hof IE, Van der Heijden JF, Clappers NC, Doevendans PA, Hassink RJ, Van Es R. Automatic Triage of 12-Lead Electrocardiograms Using Deep Convolutional Neural Networks. *J Am Heart Assoc.* 2020;9.
4. Ko W-Y, Siontis KC, Attia ZI, Carter RE, Kapa S, Ommen SR, Demuth SJ, Ackerman MJ, Gersh BJ, Arruda-Olson AM, et al. Detection of Hypertrophic Cardiomyopathy Using a Convolutional Neural Network-Enabled Electrocardiogram. *J Am Coll Cardiol.* 2020;75:722–733.
5. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Proc IEEE Int Conf Comput*

Vis. 2017;2017-October:618–626.

6. Chattopadhyay A, Sarkar A, Howlader P, Balasubramanian VN. Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks. *Proc - 2018 IEEE Winter Conf Appl Comput Vision, WACV 2018*. 2018;839–847.
7. Raghunath S, Ulloa Cerna AE, Jing L, VanMaanen DP, Stough J, Hartzel DN, Leader JB, Kirchner HL, Stumpe MC, Hafez A, et al. Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network. *Nat Med*. 2020;26:886-891.
8. Kwon JM, Lee SY, Jeon KH, Lee Y, Kim KH, Park J, Oh BH, Lee MM. Deep Learning-Based Algorithm for Detecting Aortic Stenosis Using Electrocardiography. *J Am Heart Assoc*. 2020;9:e014717.
9. van der Zwaag PA, van Rijsingen IAW, de Rooter R, Nannenber EA, Groeneweg JA, Post JG, Hauer RNW, van Gelder IC, van den Berg MP, van der Harst P, et al. Recurrent and founder mutations in the Netherlands-Phospholamban p.Arg14del mutation causes arrhythmogenic cardiomyopathy. *Netherlands Hear J*. 2013;21:286–293.
10. Haghghi K, Kolokathis F, Gramolini AO, Waggoner JR, Pater L, Lynch RA, Fan GC, Tsiapras D, Parekh RR, Dorn GW, et al. A mutation in the human phospholamban gene, deleting arginine 14, results in lethal, hereditary cardiomyopathy. *Proc Natl Acad Sci U S A*. 2006;103:1388–1393.
11. Hof IE, van der Heijden JF, Kranias EG, Sanoudou D, de Boer RA, van Tintelen JP, van der Zwaag PA, Doevendans PA. Prevalence and cardiac phenotype of patients with a phospholamban mutation. *Netherlands Hear J*. 2019;27:64–69.
12. Van Der Zwaag PA, Van Rijsingen IAW, Asimaki A, Jongbloed JDH, Van Veldhuisen DJ, Wiesfeld ACP, Cox MGPI, Van Lochem LT, De Boer RA, Hofstra RMW, et al. Phospholamban R14del mutation in patients diagnosed with dilated cardiomyopathy or arrhythmogenic right ventricular cardiomyopathy: Evidence supporting the concept of arrhythmogenic cardiomyopathy. *Eur J Heart Fail*. 2012;14:1199–1207.
13. Milano A, Blom MT, Lodder EM, Van Hoeijen DA, Barc J, Koopmann TT, Bardai A, Beekman L, Lichtner P, Van Den Berg MP, et al. Sudden cardiac arrest and rare genetic variants in the community. *Circ Cardiovasc Genet*. 2016;9:147–153.
14. Posch MG, Perrot A, Geier C, Boldt LH, Schmidt G, Lehmkuhl HB, Hetzer R, Dietz R, Gutberlet M, Haverkamp W, Özcelik C. Genetic deletion of arginine 14 in phospholamban causes dilated cardiomyopathy with attenuated electrocardiographic R amplitudes. *Hear Rhythm*. 2009;6:480–486.
15. Groeneweg JA, Van Der Zwaag PA, Olde Nordkamp LRA, Bikker H, Jongbloed JDH, Jongbloed R, Wiesfeld ACP, Cox MGPI, Van Der Heijden JF, Atsma DE, et al. Arrhythmogenic right ventricular dysplasia/cardiomyopathy according to revised 2010 task force criteria with

inclusion of non-desmosomal phospholamban mutation carriers. *Am J Cardiol.* 2013;112:1197–1206.

16. GE Healthcare. Marquette 12SL ECG Analysis Program Physician's Guide. Chicago, United States: 2012.

17. Pearce N. Analysis of matched case-control studies. *BMJ.* 2016;352:1–4.

18. 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.

19. Franceschi J-Y, Dieuleveut A, Jaggi M. Unsupervised Scalable Representation Learning for Multivariate Time Series. 2019.

20. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems.* 2019.

21. Gho JMIH, Van Es R, Stathonikos N, Harakalova M, Te Rijdt WP, Suurmeijer AJH, Van Der Heijden JF, De Jonge N, Chamuleau SAJ, et al. High resolution systematic digital histological quantification of cardiac fibrosis and adipose tissue in phospholamban p.Arg14del mutation associated cardiomyopathy. *PLoS One.* 2014;9.

22. van Setten J, Verweij N, Mbarek H, Niemeijer MN, Trompet S, Arking DE, Brody JA, Gandin I, Grarup N, Hall LM, et al. Genome-wide association meta-analysis of 30,000 samples identifies seven novel loci for quantitative ECG traits. *Eur J Hum Genet.* 2019;27:952–962.

23. Willems JL, Zywiets C, Arnaud P, van Bommel JH, Degani R, Macfarlane PW. Influence of noise on wave boundary recognition by ECG measurement programs. Recommendations for preprocessing. *Comput Biomed Res.* 1987;20:543–562.

24. Adebayo J, Gilmer J, Muelly M, Goodfellow I, Hardt M, Kim B. Sanity checks for saliency maps. *Adv Neural Inf Process Syst.* 2018;2018-Decem:9505–9515.

Table 1. Baseline demographics and electrocardiogram characteristics of all patients and patients in the training and test splits, stratified by phospholamban mutation carriers and their matched controls.

	Overall			Train			Test		
	Controls	PLN	p-value	Controls	PLN	p-value	Controls	PLN	p-value
n	1720	86		1380	69		340	17	
Age , years, mean (SD)	44 (15)	44 (15)	1.0	44 (14)	44 (14)	1.0	42 (16)	42 (17)	1.0
Female sex , n (%)	1040 (61)	52 (61)	1.0	820 (59)	41 (59)	1.0	220 (65)	11 (65)	1.0
PR interval , ms, mean (SD)	151 (24)	162 (28)	0.001	151 (24)	161 (27)	0.001	149 (20)	164 (34)	0.005
QRS interval , ms, mean (SD)	93 (15)	93 (19)	0.74	94 (15)	93 (18)	0.59	93 (15)	94 (20)	0.72
QTc interval , ms, mean (SD)	422 (29)	429 (40)	0.050	422 (29)	427 (39)	0.17	423 (26)	434 (45)	0.086
Maximum voltage extremity leads , mV, mean (SD)	1.2 (0.38)	0.79 (0.43)	<0.001	1.2 (0.38)	0.81 (0.45)	<0.001	1.2 (0.40)	0.72 (0.39)	<0.001
Maximum voltage precordial leads , mV, mean (SD)	2.2 (0.80)	1.8 (0.74)	<0.001	2.2 (0.77)	1.8 (0.75)	<0.001	2.3 (0.89)	1.5 (0.69)	0.001
Low QRS voltage , n (%)	41 (2.4)	31 (36)	<0.001	27 (2.0)	22 (32)	<0.001	14 (4.1)	9 (53)	<0.001
T-wave morphology , n (%)			<0.001			<0.001			<0.001
Aspecific abnormalities	66 (3.8)	27 (31)		48 (3.5)	20 (29)		18 (5.3)	7 (41)	
Inverted in the extremity leads	33 (1.9)	14 (16)		25 (1.8)	12 (17)		8 (2.4)	2 (12)	
Inverted in the right precordial leads	34 (2.0)	10 (12)		28 (2.0)	9 (13)		6 (1.8)	1 (5.9)	
Inverted in the left precordial leads	49 (2.8)	22 (26)		41 (3.0)	20 (29)		8 (2.4)	2 (12)	

PLN: phospholamban

Table 2. Discriminatory performance of the baseline and updated logistic regressions models and the deep neural network in the independent test set. The baseline model includes the currently established electrocardiogram features of phospholamban mutation carriers. Features identified by the deep neural network, translated into quantitative measures, are added in the updated model for validation of these features.

	Logistic regression models		DNN
	Baseline	Updated	
C-statistic [95% CI]	0.84 [0.73-0.92]	0.91 [0.83-0.97]	0.95 [0.91-0.99]
Sensitivity	53%	76%	82%
Specificity	94%	93%	93%
Positive predictive value	33%	34%	37%
Negative predictive value	98%	99%	99%

DNN: deep neural network, CI: confidence interval.

Circulation: Arrhythmia
and Electrophysiology

Table 3. Odds ratios and 95% confidence intervals for the variables in the baseline and updated logistic regression models for prediction of PLN mutation carrier status in the training dataset. The baseline model includes the currently established electrocardiogram features of phospholamban mutation carriers. Features identified by the deep neural network, translated into quantitative measures, are added in the updated model for validation of these features.

	Baseline model	Updated model
Age per year increase	0.98 [0.97 – 1.0]	0.97 [0.94 – 0.99]
Male sex	0.96 [0.55 – 1.7]	1.17 [0.61 – 2.2]
Low QRS voltage	16.6 [8.1 – 34]	3.7 [1.5 – 9.0]
Left precordial inverted T-waves	7.4 [2.9 – 18]	4.5 [1.6 – 12]
Right precordial inverted T-waves	1.3 [0.37 – 4.2]	1.3 [0.33 – 4.9]
R-wave voltage in V3 per 1 mV increase		0.37 [0.15 – 0.86]
PR interval per 10 ms increase		1.2 [1.1 – 1.3]
T-wave voltage in V3 per 1 mV increase		1.13 [0.31 – 3.7]
R-wave voltage in V6 per 1 mV increase		0.073 [0.0025 – 0.20]
T-wave voltage in I per 1 mV increase		0.0013 [0.000013 – 0.06]
R-wave voltage in III per 1 mV increase		3.5 [1.4 – 9.0]

Table 4. Summary measures of the quantitative translations of the newly identified electrocardiogram features of phospholamban mutation carriers. Most prevalent newly identified features for predicting the phospholamban mutation, as identified by the visualizations of the deep neural network, were translated into quantitative measures and tested in the updated logistic regression model for validation.

	Controls	PLN	p-value
R-wave voltage in V3, mV, median [IQR]	0.72 [0.47 – 1.1]	0.31 [0.19 – 0.61]	<0.001
PR interval, ms, mean (SD)	151 (24)	162 (28)	<0.001
T-wave voltage in V3, mV, mean (SD)	0.46 (0.29)	0.28 (0.29)	<0.001
R-wave voltage in V6, mV, median [IQR]	0.66 [0.46 – 0.91]	0.28 [0.15 – 0.45]	<0.001
T-wave voltage in I, mV, mean (SD)	0.25 (0.15)	0.11 (0.15)	<0.001
R-wave voltage in III, mV, median [IQR]	0.38 [0.18 – 0.72]	0.22 [0.09 – 0.56]	<0.001

PLN: phospholamban, SD: standard deviation, IQR: interquartile range, mV: millivolt, ms: millisecond.

Circulation: Arrhythmia and Electrophysiology

Figure Legends:

Figure 1. Flowchart of the patient selection and model development process. ECG: electrocardiogram, HTx: heart transplantation, LVAD: left ventricular assist device, PLN: phospholamban.

Figure 2. Output of the Guided Grad-CAM visualization algorithm for all PLN mutation carriers and their controls. Left: Mean of temporally normalized median 12-lead ECGs of both the 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). Guided Grad-CAM: Guided Gradient Class Activation Mapping, PLN: phospholamban.

Figure 3. Representative examples of an ECG of a PLN mutation carrier (top panel) and a control subject (bottom panel) with their respective DNN probability score for having the PLN mutation. Note that the control subject ECG also exhibits the established PLN features (low QRS voltages and the presence of inverted T-waves in the left precordial leads) but is classified correctly as a control subject. The features as detected by the DNN (decreased R- and T-wave voltage in V3) can be used to distinguish the PLN mutation carriers and control subject. DNN: deep neural network, PLN: phospholamban.

Figure 4. Relationship of the mean Grad-CAM⁺⁺ importance value of the T-wave area with the human interpretation of the T-wave and of the QRS-complex area with the human classification of low QRS voltage in PLN patients. In the temporally aligned Grad-CAM⁺⁺ curves, the mean is taken for the area of the QRS-complex and the T-wave. A boxplot of the importance values (between 0 and 1) of that region for the network for predicting PLN are shown in relationship with the human interpretation of the corresponding segments. Grad-CAM⁺⁺: gradient class activation mapping ++.

What Is Known?

- Deep neural networks can be used to interpret raw electrocardiograms with high accuracy in for example hypertrophic cardiomyopathy patients
- Deep neural networks are able to discover complex features in ECG signals, but are considered ‘black box’ algorithms that lack interpretability

What the Study Adds?

- A deep learning-based feature detection approach is able to discover and visualize disease-specific ECG features
- In PLN p.Argdel14 mutation carriers, visualizations revealed established ECG features (low QRS voltages and T-wave inversions), specified these features (e.g. R and T-wave attenuation in V2/V3) and identified novel PLN-specific ECG features (e.g. increased PR-duration)

DATA SOURCE

PLN p.Arg14del gene mutation ECGs
(n = 1900, patients = 93)

Physician-annotated hospital ECGs
(n = 365.173, patients = 147.098)

Inclusion: first ECG
Exclusion: insufficient quality, non-sinus rhythm, signs of acute ischaemia, post-HTx and post-LVAD

PLN first ECGs
(n = 86)

Controls
(n = 135.353)



MATCHING

1:20 matching on age and sex

PLN first ECGs
(n = 86)

Matched controls
(n = 1720)

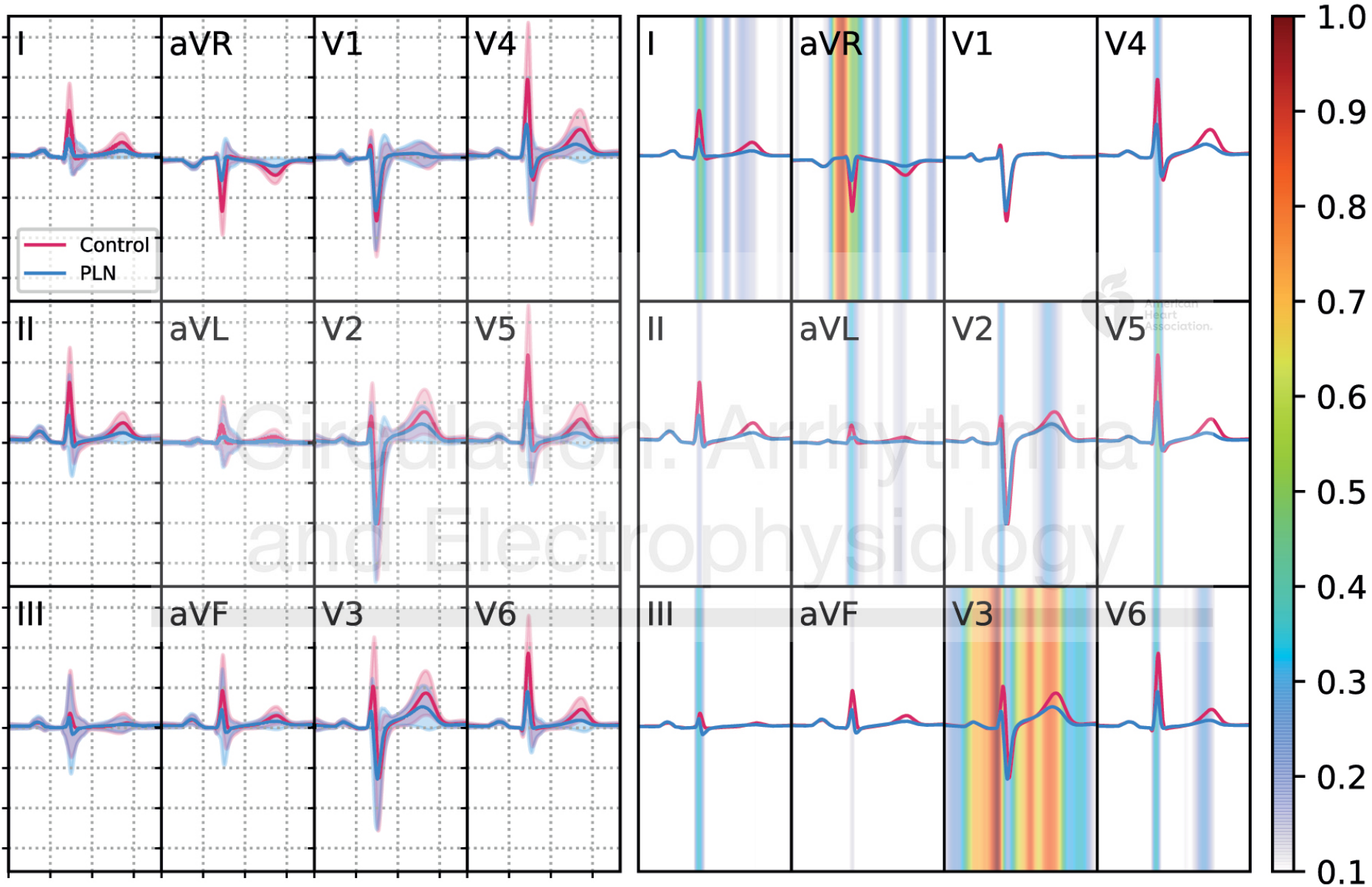
MODEL DEVELOPMENT

Training dataset
(PLN = 69, controls = 1380)

Testing dataset
(PLN = 17, controls = 340)

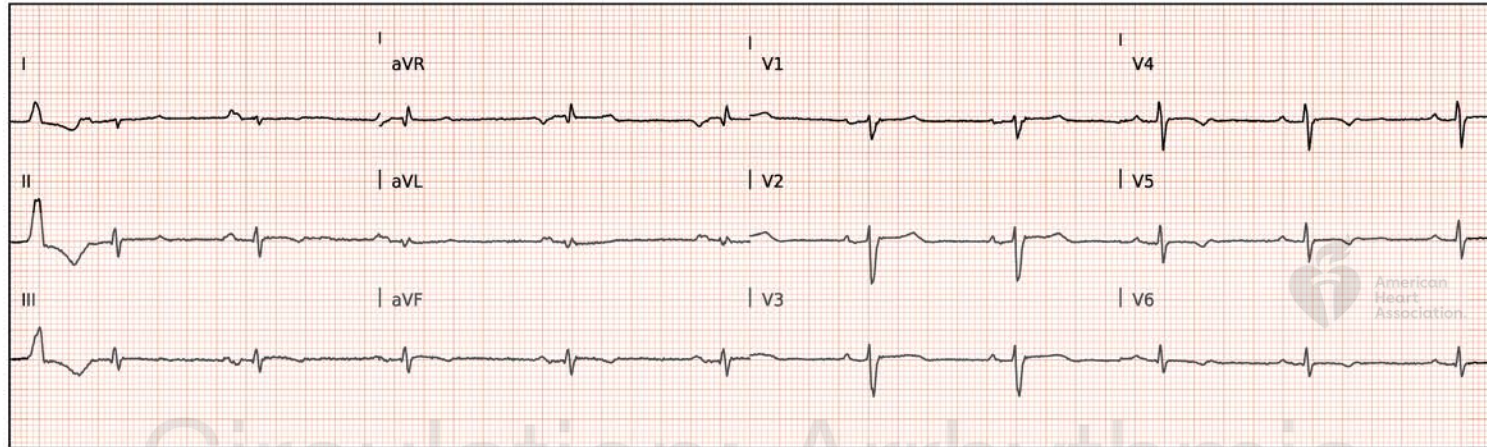
8-fold cross-validation





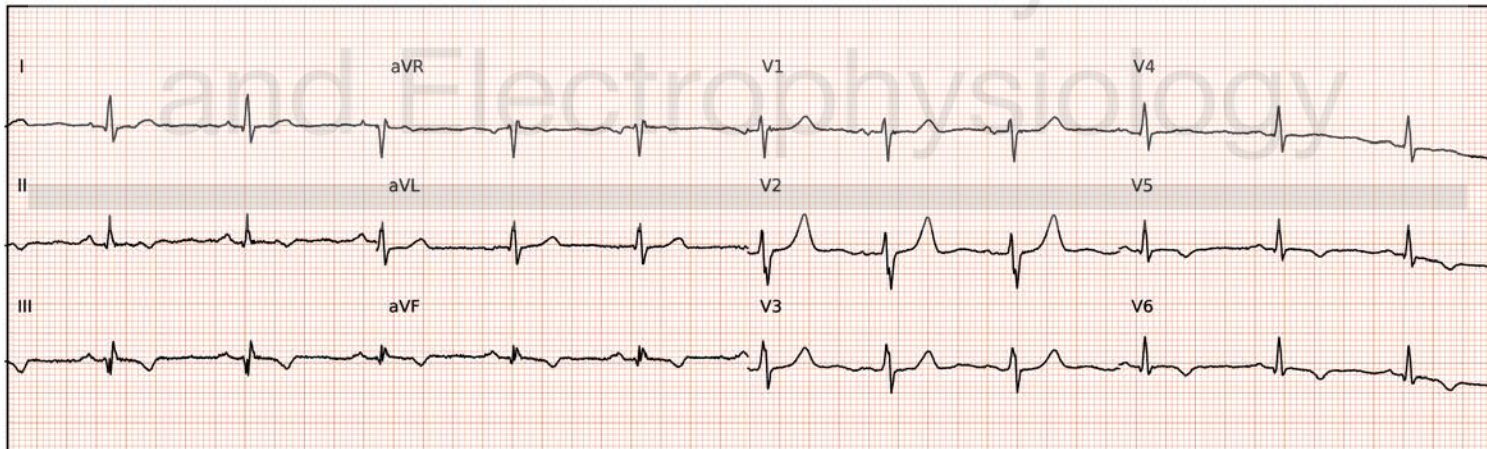
PLN mutation carrier

DNN estimated probability for having PLN: **99 %**



Control subject

DNN estimated probability for having PLN: **23 %**



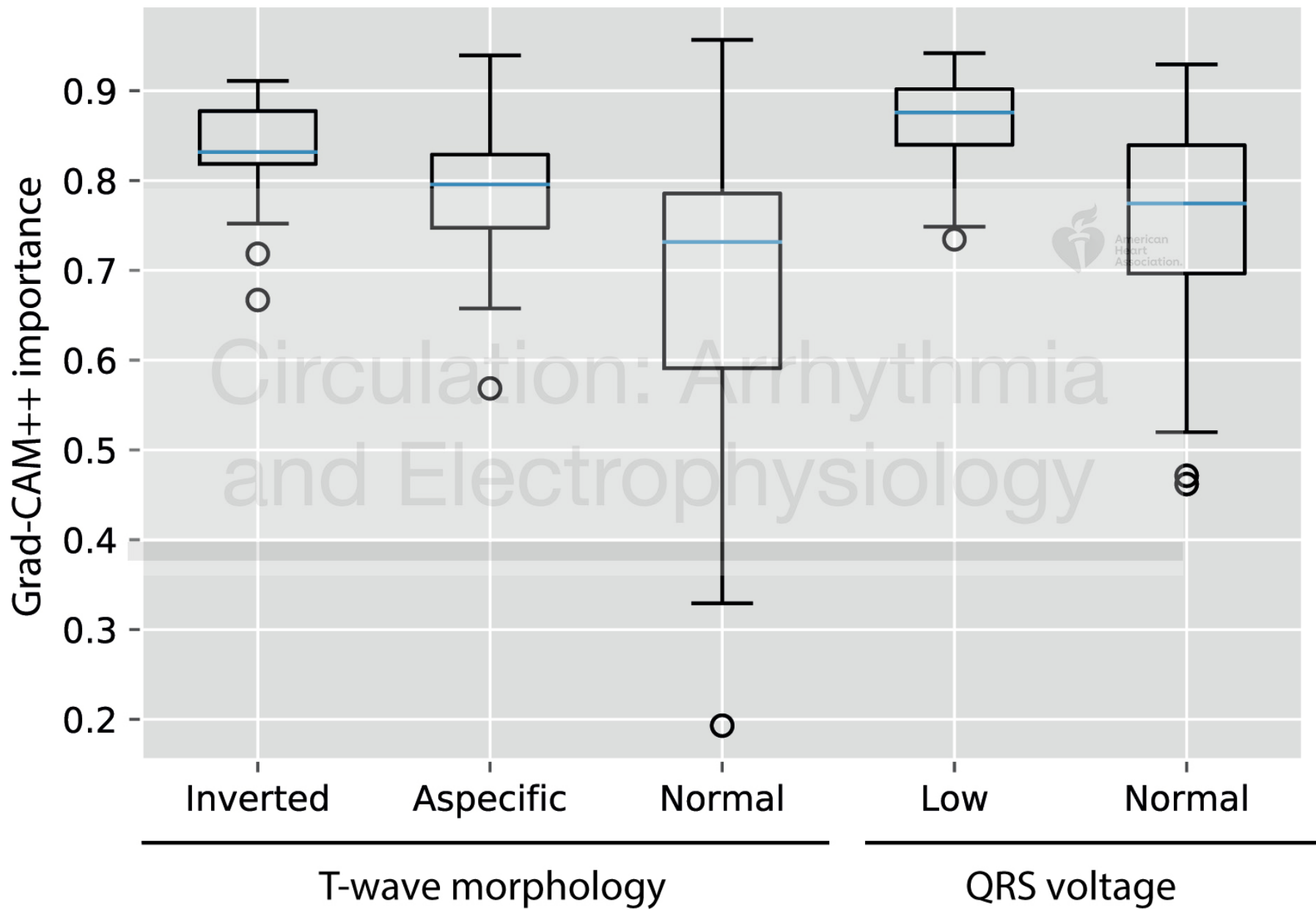
total time: 10.0 s sample freq.: 500 Hz

25.0 mm/s 10 mm/mV

$p < 0.001$

$p = 0.002$

$p < 0.001$

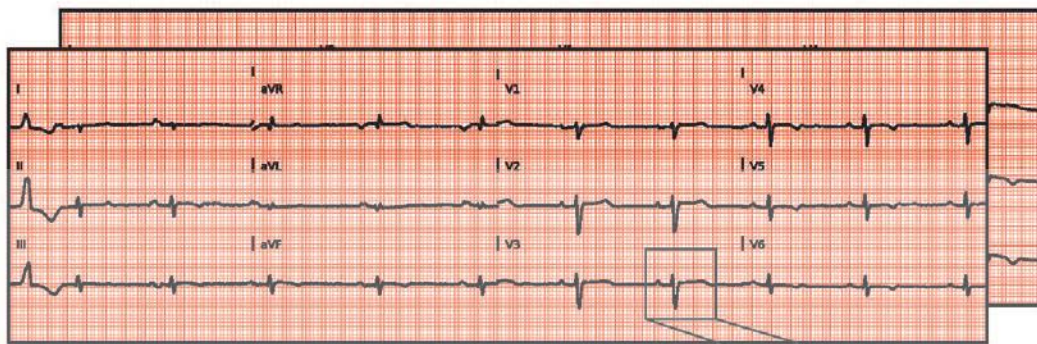


Deep learning as a disease-specific ECG feature detector

Selection of disease population and collection of ECGs

Recognition and validation of new ECG features

Example:
PLN p.Arg14del
 mutation carriers
 VS
 age and sex
 matched controls



1. Reduced R- and T-wave amplitude
2. T-wave inversion
3. Increased PR duration



Extraction of median ECG beats

Development of deep neural network

Visualization algorithm

