Computer versus cardiologist: Is a machine learning algorithm able to outperform an expert in diagnosing a phospholamban p.Arg14del mutation on the electrocardiogram?

Hidde Bleijendaal, MD,[*†1] Lucas A. Ramos, MSc,[†‡1] Ricardo R. Lopes, MSc,[‡] Tom E. Verstraelen, MD,[*] Sarah W.E. Baalman, MD,[*] Marinka D. Oudkerk Pool, MSc,[*] Fleur V.Y. Tjong, MD, PhD,[*] Francisco M. Melgarejo-Meseguer, PhD,[§] F. Javier Gimeno-Blanes, PhD,[‖] Juan R. Gimeno-Blanes, MD, PhD,[§¶] Ahmad S. Amin, MD, PhD,[*¶] Michiel M. Winter, MD, PhD,[*] Henk A. Marquering, PhD,[‡] Wouter E.M. Kok, MD, PhD,[*] Aeilko H. Zwinderman, PhD,[†] Arthur A.M. Wilde, MD, PhD, FHRS,[*¶] Yigal M. Pinto, MD, PhD[*]

From the [*]Amsterdam UMC, University of Amsterdam, Heart Center, Department of Clinical and Experimental Cardiology, Amsterdam Cardiovascular Sciences, Amsterdam, The Netherlands, [†]Amsterdam UMC, University of Amsterdam, Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam, The Netherlands, [‡]Amsterdam UMC, University of Amsterdam, Department of Biomedical Engineering and Physics, Amsterdam, The Netherlands, [§]Virgen de Arrixaca Hospital, El Palmar, Spain, [‖]Miguel Hernández University, Elche, Alicante, Spain, and [¶]Member of the European Reference Network for Rare and Low Prevalence Complex Diseases of the Heart (ERN GUARD-Heart).

BACKGROUND Phospholamban (PLN) *p.Arg14del* mutation carriers are known to develop dilated and/or arrhythmogenic cardiomyopathy, and typical electrocardiographic (ECG) features have been identified for diagnosis. Machine learning is a powerful tool used in ECG analysis and has shown to outperform cardiologists.

OBJECTIVES We aimed to develop machine learning and deep learning models to diagnose PLN *p.Arg14del* cardiomyopathy using ECGs and evaluate their accuracy compared to an expert cardiologist.

METHODS We included 155 adult PLN mutation carriers and 155 age- and sex-matched control subjects. Twenty-one PLN mutation carriers (13.4%) were classified as symptomatic (symptoms of heart failure or malignant ventricular arrhythmias). The data set was split into training and testing sets using 4-fold cross-validation. Multiple models were developed to discriminate between PLN mutation carriers and control subjects. For comparison, expert cardiologists classified the same data set. The best performing models were validated using an external PLN *p.Arg14del* mutation carrier data set from Murcia, Spain (n = 50). We applied occlusion maps to visualize the most contributing ECG regions.

RESULTS In terms of specificity, expert cardiologists (0.99) outperformed all models (range 0.53–0.81). In terms of accuracy and sensitivity, experts (0.28 and 0.64) were outperformed by all models (sensitivity range 0.65–0.81). T-wave morphology was most important for classification of PLN *p.Arg14del carriers*. External validation showed comparable results, with the best model outperforming experts.

CONCLUSION This study shows that machine learning can outperform experienced cardiologists in the diagnosis of PLN *p.Arg14del* cardiomyopathy and suggests that the shape of the T wave is of added importance to this diagnosis.

KEYWORDS Cardiomyopathy; Deep learning; ECG analysis; Genetic heart disease; Machine learning; Phospholamban

(Heart Rhythm 2021;18:79–87)

Funding sources: This study was supported by ERN GUARD-Heart and the University of Amsterdam Research Priority Area Medical Integromics (to Dr Bleijendaal) and Medolution ITEA3, project number 14003 (to Mr Ramos). Disclosures: The authors have no conflicts of interest to disclose. [1]Shared first authorship. **Address reprint requests and correspondence:** Dr Hidde Bleijendaal, Amsterdam UMC, University of Amsterdam, Heart Center, Department of Clinical and Experimental Cardiology, Amsterdam Cardiovascular Sciences, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands. E-mail address: h.bleijendaal@amsterdamumc.nl.

1547-5271/

https://doi.org/10.1016/j.hrthm.2020.08.021

Introduction

Phospholamban (PLN) is a transmembrane sarcoplasmic reticulum phosphoprotein and is a major regulator of calcium homeostasis in cardiomyocytes. Mutations in the gene encoding this protein are known to cause cardiomyopathy, including arrhythmogenic cardiomyopathy (ACM) and dilated cardiomyopathy.[1] Carriers of mutations in PLN are at increased risk of developing malignant ventricular arrhythmias and end-stage heart failure, leading to high mortality.[2–4]

Low QRS voltages have been reported as the electrocardiographic (ECG) hallmark in PLN mutation carriers.[5] Two large Dutch cohort studies reported low-voltage ECGs in 46% and 41% in 52 (van der Zwaag et al[1]) and 295 (van Rijsingen et al[2]) patients, respectively. Additionally, repolarization changes on the ECG, in particular T-wave inversions in the lateral leads, are frequently seen in PLN *p.Arg14del* mutation carriers. van Rijsingen et al[2] reported T-wave inversions in 40%, while van der Zwaag et al[1] reported T-wave inversions in 57%. A Canadian cohort study by Cheung et al[6] reported 53% in 50 patients. Additionally, PLN is known to cause ACM and one of the diagnostic criteria for ACM is frequent ventricular extrasystoles (>500/24 h).[7] This was present in 48% of the carriers in the van Rijsingen cohort[2] and in 65% of the Holter that were evaluated by van der Zwaag et al.[1]

PLN *p.Arg14del* cardiomyopathy is a rare disease, with a prevalence of 0.08%–0.38% in selected cardiomyopathy cohorts.[8] Other PLN gene mutations have been described, mostly in case reports and small cohorts, while Hof et al[8] reported data of over a thousand *p.Arg14del* mutation carriers in the Netherlands alone, making *p.Arg14del* the most common PLN mutation in the literature to date.[4] Most general cardiologists do not routinely see patients with PLN-associated cardiomyopathy and consequently may not recognize the ECG features associated with this disease. The standard for diagnosing a PLN *p.Arg14del* mutation is genetic testing. However, when a patient is suspected of having a gene mutation causing structural heart disease, the ECG can increase (or decrease) the probability of having a mutation, assisting the clinician in early decision making regarding the diagnosis and possible therapy. Early diagnosis is of major importance because PLN-associated cardiomyopathy is among the most malignant cardiomyopathies necessitating early ICD implantation.[2,7]

In the past few years, the use of machine learning (ML) and, more specific, deep learning (DL) methods in medicine has increased significantly.[9] An advantage of DL is that it can automatically learn features from raw data, allowing the discovery of previously unknown relationships.[10] Within cardiology, DL is used for the detection of a variety of cardiac arrhythmias, such as atrial fibrillation, in which the models outperform cardiologists, thereby positioning DL as a powerful tool for ECG analysis.[9,11] The increased accuracy of DL models often comes with the downside of the lack of interpretability. However, new techniques have been developed, making it possible to visualize the features a DL model uses and thus can be used to identify new features.[12,13]

In this study, we aimed to develop ML and DL models and study their accuracy compared to expert cardiologists in diagnosing PLN *p.Arg14del* cardiomyopathy on an ECG. We aimed to present a proof of concept to show how ML-enabled ECG analysis is of added value, specifically when it concerns a rare disease that is often missed simply because it is rarely seen.

Moreover, we aimed to identify specific regions of ECGs that could give insights for improving diagnosis of this disease and be used for better understanding of PLN mutation cardiomyopathy in general.

Methods

Data collection and labeling

We collected ECGs from all patients that were stored in the ECG database (MUSE, GE Healthcare, Chicago, Illinois) of the Amsterdam University Medical Centers (UMC), location Academic Medical Center, during the period from 1998 up to and including 2018. To minimize the amount of non-PLN mutation–related cardiovascular pathology that could potentially influence the ECG, we included only ECGs from patients aged 18–60 years. From this database, we extracted all patients known to have a PLN *p.Arg14del* mutation. A mutation carrier was defined as symptomatic when they suffered from either an arrhythmic event (sustained ventricular tachycardia or ventricular fibrillation) or a symptomatic episode caused by heart failure (New York Heart Association class 2 or higher, as defined by clinical staff). This information was provided by the national PLN registry, and informed consent for reuse of patient information has been obtained. ECGs were excluded if they were made on the emergency ward or during hospitalization in a clinical ward to exclude the possible effect of acute cardiac disease on the ECG. As a control group, we selected ECGs from patients aged between 18 and 60 years who underwent general noncardiovascular preoperative screening at the outpatient clinic of the Amsterdam UMC, location Academic Medical Center, after which we randomly selected a subgroup to match the population with PLN according to age and sex, to ensure the same distribution for each group. For both groups, only the first recorded ECG for each patient was used. Figure 1 shows a diagram with the PLN and control group selection process.

We excluded all ECGs that were considered technically inadequate according to an experienced investigator (H.B.) (limb lead reversal, loss of signal on 1 or more leads, and high amount of noise of 2 or more leads, making analysis impossible) or that had any other rhythm than sinus rhythm. ECGs were labeled as “PLN” or “control” on the basis of the presence of a PLN Arg14.del gene mutation. This data set was named the Amsterdam data set to discriminate from the external validation set. External validation was performed on a population of PLN *p.Arg14del* mutation

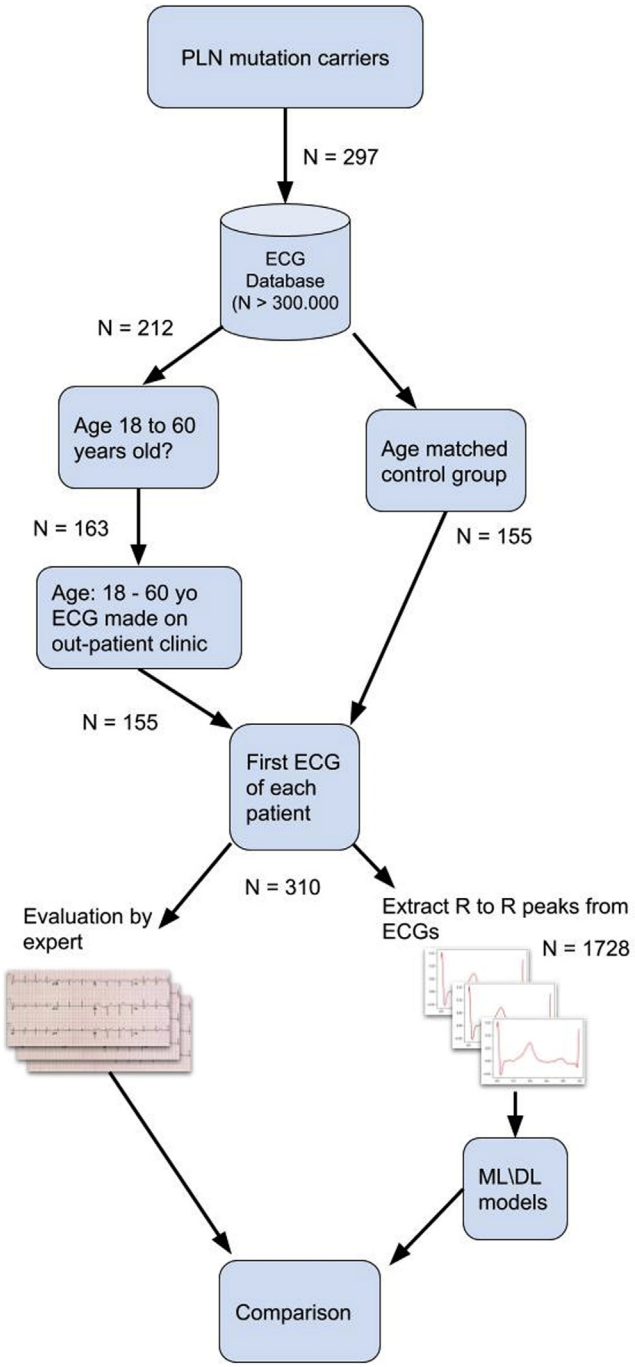

Figure 1 Data cleaning process from patient selection to model development. DL = deep learning; ECG = electrocardiographic; ML = machine learning.

carriers from the Virgen de Arrixaca Hospital in Murcia, Spain. From the local ECG database, a random set of non-PLN mutation carriers in this hospital was selected as a control group. This external validation set was named the Murcia data set.

The study was approved and the requirement for informed consent was waived by the Medical Ethics Commission of the Amsterdam UMC on 22-11-2018 (registration number W18_371#18.425).

Evaluation by expert cardiologists

All ECGs included were anonymized and visually evaluated separately by 2 cardiologists with expertise in PLN-associated cardiomyopathy (A.A.M.W. and W.E.M.K.). The experts classified the ECGs in PLN or non-PLN and were not informed of the ratio between carriers and noncarriers. For ECG classification, they used known ECG features, as described in Introduction (low QRS voltages, T-wave inversion, and frequent extrasystoles).

Data preprocessing and development of ML models

To increase the amount of training data, we extracted all beats from each 10-second ECG available and used them as individual samples during training. Details about the data preprocessing are given in Online Supplemental Methods Section 1.1 and Online Supplemental Figure 1. Patients were randomly split into training, validation, and testing sets by using 4-fold cross-validation stratified for carriers and controls. Initially, 3-folds are separate for training and 1-fold is left aside for testing. From the 3-folds used for training, 20% is separated as a validation set to be used to assess network performance during training and hyperparameter optimization. All heartbeats from each individual patient were kept in either the training set or the testing set in the initial split to prevent data leakage. For testing, only 1 beat was used per patient as reference. We did not choose a beat on one of the edges of the ECG because of the high probability of it containing noise. For creating the models, we followed 2 approaches, defined below.

Our first approach—the wavelet ML-based approach—consisted of applying a wavelet transform for each individual beat, since wavelets have been broadly and successfully used in multiple ECG applications.[14,15] More details about wavelets and their implementation can be found in Online Supplemental Methods Section 1.2. The output of the wavelet transformation (of size (64 × 8)) was flattened and used as input to train ML classifiers—logistic regression, support vector machine (SVM), multilayer perceptron, random forest, and extreme gradient boosting (XGB)—by following the approach of Kumar et al.[15]

In our second approach–DL-based approach—we implemented 1- and 2-dimensional (1D and 2D) convolutional neural networks (CNNs) and long short-term memory (LSTM) networks by using the R-to-R peak as input. For each type of network (CNN and LSTM), we implemented 2 approaches (using 1D and 2D convolutions), namely, approach A and approach B. Details about these approaches and their implementations are available in Online Supplemental Methods Section 1.3.

Statistical analysis

For model evaluation, we reported the average accuracy, sensitivity, specificity, and area under the receiver operating characteristic (ROC) curve. We deemed the best performing models the ones with the highest accuracy and sensitivity because of greater importance of missing true-positive

Table 1 Description of the Amsterdam data set

Variable	PLN (n = 155)	Control (n = 155)
Age (y)*	39 (28–50)	39 (28–50)
Sex: male*	63 (41)	63 (41)
Ventricular rate (beats/min)	68 (60–75)	65 (57–73)
Atrial rate (beats/min)	68 (60–75)	66 (57–73)
QRS duration (ms)	86 (80–94)	94 (84–104)
QT interval (ms)	388 (368–406)	400 (374–426)
QT corrected interval (ms)	407 (394–424)	410 (401–429)
P-wave axis (deg)	55 (37–66)	48 (33– 61)
R-wave axis (deg)	48 (2–75)	34 (3–63)
T-wave axis (deg)	46 (1–63)	38 (20–58)

Values are presented as median (interquartile range) or as n (%).
*Variables were used to match samples from the control group.

(PLN) patients. We used the McNemar test to check whether the difference between the models and the experts was statistically significant.[16]

Visualization of ECG features

For our best performing model, we created visualization plots to visualize the parts of the ECG that were most relevant for classification of patients with PLN in our DL model; we used "occlusion maps" for this purpose.[17] We generated occlusion maps by systematically occluding parts of the heartbeat signal. We split the (8 × 256) input signal into 16 parts of (8 × 16) and occluded a region by setting all its values to zero. Then, we applied the trained model to the signal with the occluded region and evaluated the loss of model performance. The higher the loss of performance, the more important the occluded region.

Reproducibility and open access

Given its sensitive nature, the data used in this study are not publicly available. All the code used in this study, however, is available at the following GitHub page: https://github.com/L-Ramos/CardiologyAI.

Results

Among the 297 known PLN *p.Arg14del* mutation carriers in Amsterdam UMC, 155 were eligible for inclusion in this study (see Figure 1 for a flow diagram). Among PLN mutation carriers, 13.5% were symptomatic at the time the ECG was recorded. The mean age in this group was 39 years (interquartile range 28–50 years), and 63 (41%) were male. Baseline ECG characteristics are summarized in Table 1.

Performance of ML and DL models compared to expert cardiologists

Table 2 presents the results for both the experts and ML and DL models averaged over the 4-folds. Experts 1 and 2 had an accuracy of 0.65 and 0.63, a sensitivity of 0.32 and 0.27, and a specificity of 0.97 and 0.99, respectively (Table 2). Despite showing a slightly higher accuracy, expert 1 had also larger standard deviation than did expert 2. Figure 2 shows ROC curves for the selection of the best performing models and the results of the best performing expert; the ROC curves for other models are shown in Online Supplemental Figure 2. Figure 3 shows an example of an ECG correctly classified as PLN by both the 1D CNN and the experts. For accessing interrater reliability between the 2 cardiologists, we computed the Cohen's κ score,[18] which was 0.65 for the Amsterdam data set, indicating a substantial agreement between the experts. For the Murcia data set, κ was 0.27, which indicates a fair agreement between the experts.

ML-based approach

The different wavelet ML models showed comparable results. The wavelet SVM (accuracy 0.76) and wavelet logistic regression (accuracy 0.76) can be marked as the 2 best performing models. In terms of sensitivity, the wavelet ML model also outperformed the cardiologist (0.72 vs 0.31). In terms of specificity, the cardiologist outperformed the wavelet ML model (0.99 vs 0.81).

Table 2 Performance (in terms of accuracy, sensitivity, and specificity) of the experts and ML and DL models on the Amsterdam data set

Model	Sensitivity	Specificity	Accuracy	AUC
Expert 1	0.32 ± 0.01	0.97 ± 0.02	0.65 ± 0.06	0.65 ± 0.06
Expert 2	0.25 ± 0.05	1.0 ± 0.00	0.63 ± 0.03	0.63 ± 0.02
1D CNN—approach A	0.65 ± 0.02	0.67 ± 0.07	0.65 ± 0.04	0.74 ± 0.03
2D CNN—approach B	0.77 ± 0.03	0.67 ± 0.09	0.72 ± 0.03	0.78 ± 0.03
1D LSTM network—approach A	0.65 ± 0.13	0.59 ± 0.18	0.62 ± 0.05	0.72 ± 0.09
2D LSTM network—approach B	0.81 ± 0.08	0.53 ± 0.12	0.67 ± 0.08	0.74 ± 0.09
Wavelet—MLP	0.70 ± 0.05	0.76 ± 0.03	0.73 ± 0.02	0.78 ± 0.02
Wavelet—SVM	0.71 ± 0.05	0.81 ± 0.06	0.76 ± 0.05	0.80 ± 0.06
Wavelet—LR	0.72 ± 0.07	0.79 ± 0.06	0.76 ± 0.05	0.80 ± 0.06
Wavelet—KNN	0.69 ± 0.05	0.77 ± 0.07	0.74 ± 0.06	0.76 ± 0.06
Wavelet—RFC	0.69 ± 0.5	0.80 ± 0.07	0.75 ± 0.06	0.83 ± 0.03
Wavelet—XGB	0.69 ± 0.03	0.81 ± 0.00	0.75 ± 0.02	0.82 ± 0.02

Values are presented as mean ± SD. These values are computed over 4-folds.
1D = 1-dimensional; 2D = 2-dimensional; AUC = area under the receiver operating characteristic curve; CNN = convolutional neural network; DL = deep learning; KNN = k-nearest neighbor; LR = logistic regression; LSTM = long short-term memory; ML = machine learning; MLP = multilayer perceptron; RFC = random forest classifier; SVM = support vector machine; XGB = extreme gradient boosting.

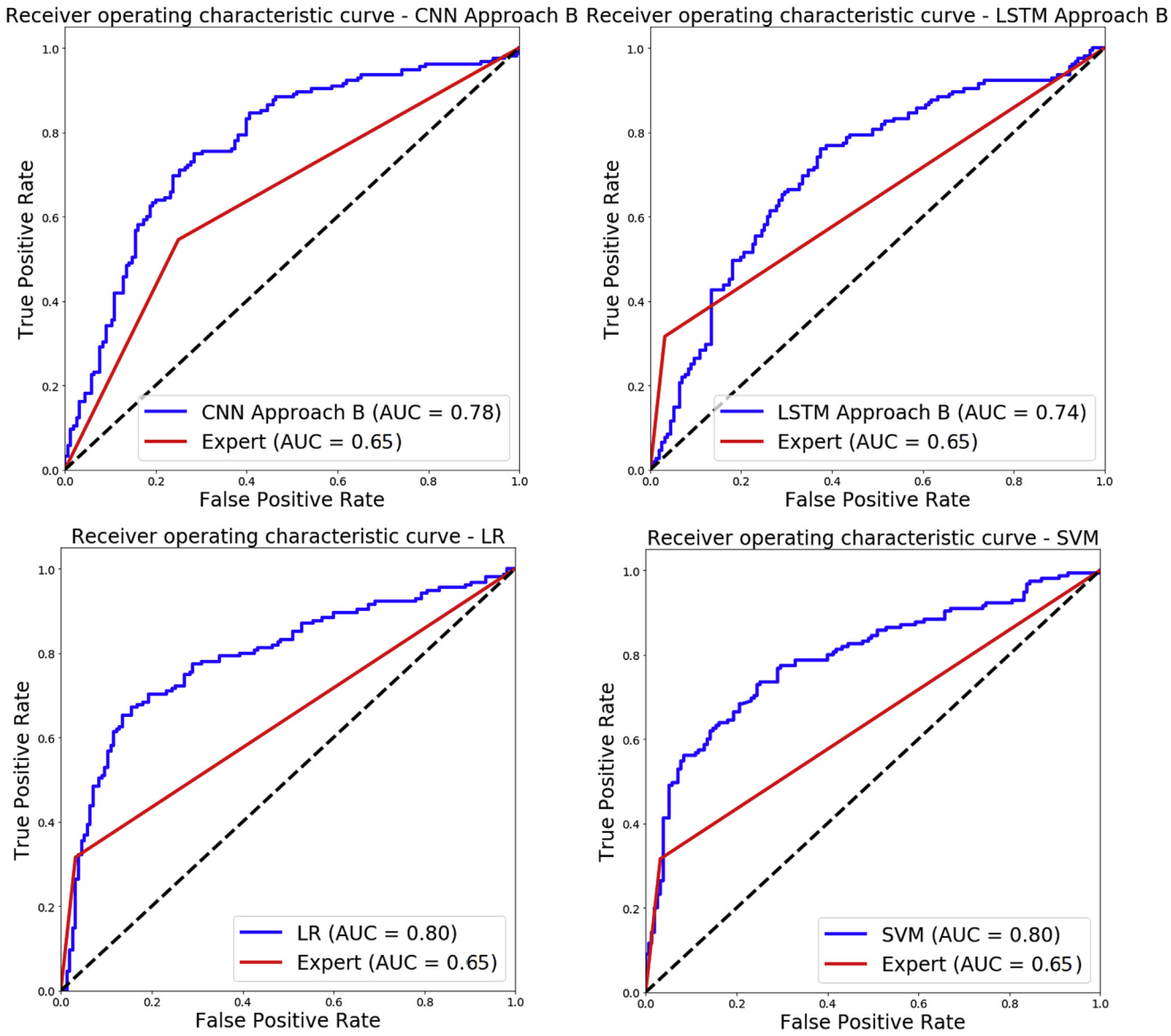

Figure 2 Receiver operating characteristic curves for the best performing expert and the 4 best performing models on the Amsterdam data set. AUC = area under the receiver operating characteristic curve; CNN = convolutional neural network; LR = logistic regression; LSTM = long short-term memory; SVM = support vector machine.

DL-based approach

The DL model performing best on test data was the 2D CNN with approach B, with an accuracy of 0.72, outperforming both the expert cardiologists, with a standard deviation comparable to that of the experts. In terms of sensitivity, this CNN also outperformed both experts (0.77 vs 0.31 of the expert with highest sensitivity). In terms of specificity, the experts outperformed the CNNs (0.99 vs 0.67). Using the McNemar test, we compared the best performing model (CNN with approach B) with the expert with the highest accuracy. The χ^2 statistic was 2.125 and the P value was .145 for the Amsterdam data set.

ML and DL models have multiple hyperparameters to be optimized, and all the parameters used are listed in Online Supplemental Table 1.

External validation on the Murcia data set

The results of the external validation for expert cardiologists, the 2 best performing wavelet ML models, and the 2 best performing DL models are presented in Table 3 (standard deviation is not available since the trained models were used for inference in the whole set). In terms of accuracy, the CNN with approach B performed slightly better on the Murcia data set than did the expert with highest accuracy (0.68 vs 0.65). Our wavelet ML models showed the highest sensitivity (0.96) vs the CNN and LSTM network (0.64 and 0.48); however, both wavelet ML models showed poor specificity (0.20). A comparison between the ROC curves for the best performing expert and those for the best ML/DL approaches for the Murcia data set is presented in Online Supplemental Figure 3. Using the McNemar test for

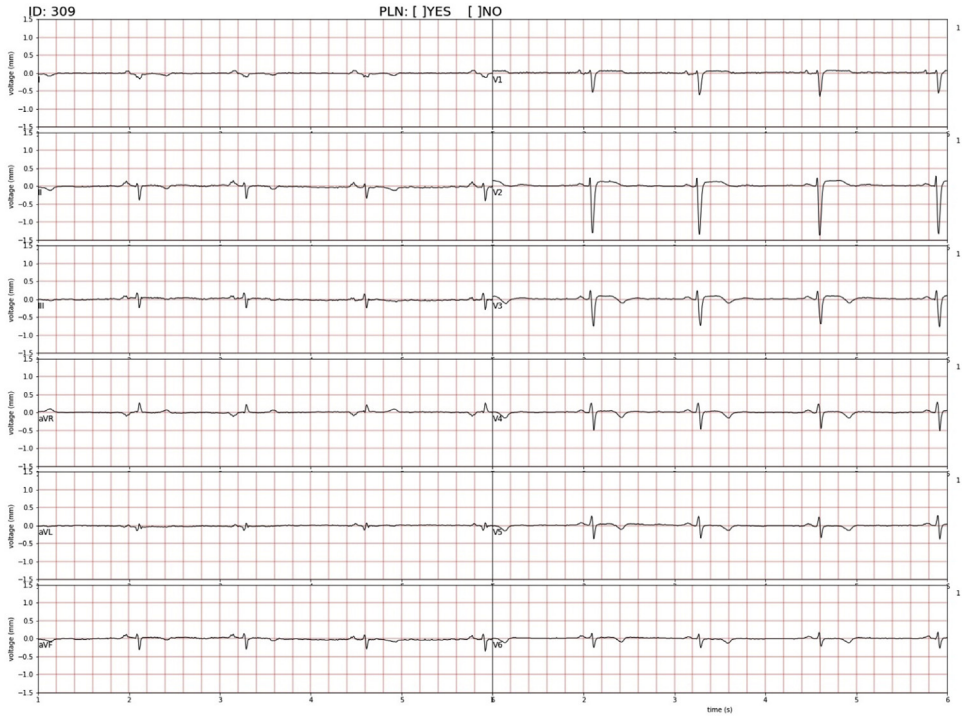

Figure 3 Example of an electrocardiogram (ECG) that both the experts and the convolutional neural network labeled correctly as "phospholamban (PLN)." This example shows the typical ECG features that the experts use to detect PLN: low QRS voltages on the limb leads and T-wave inversion in leads V_3 through V_6.

comparison of the CNN with approach B and expert 1, the χ^2 statistic was found to be 4.114 ($P = .043$).

Visualization of ECG features

Figure 4 shows 4 examples of the ECG regions from which the model extracted features so as to classify the specific ECG sample in either PLN or control patients. In 63% of the true positives, our results showed that the T wave was the most important part for the model; an example is shown in Figure 4A. In 14.2%, the model did not use a specific part of the signal but used the whole signal (see Figure 4B). In the majority of true negatives, the model used the whole signal for classification (56%), and in only 3%, the T wave was the most prominent ECG feature. An overview of the ECG features used is given in Online Supplemental Table 2.

Table 3 External validation on the Murcia data set

Model	Sensitivity	Specificity	Accuracy	AUC
Expert 1	0.55	0.75	0.65	0.65
Expert 2	0.18	0.91	0.56	0.55
2D CNN—approach B	0.64	0.72	0.68	0.70
Wavelet—LR	0.96	0.20	0.58	0.58
Wavelet—SVM	0.96	0.20	0.58	0.58
2D LSTM network—approach B	0.48	0.68	0.58	0.63

Values shown are averaged over 4-folds.

2D = 2-dimensional; AUC = area under the receiver operating characteristic curve; CNN = convolutional neural network; LR = logistic regression; LSTM = long short-term memory; SVM = support vector machine.

Discussion

Among all our models, the 2D CNN with approach B outperformed expert cardiologists in accuracy and sensitivity on both the Amsterdam and the Murcia data sets. In terms of specificity, cardiologists were superior in the identification of PLN mutation carriers on the ECG. This suggests that neither the use of ML/DL nor the assessment of an expert cardiologist for the diagnosis of a PLN *p.Arg14del* mutation on the ECG is superior to each other.

Performance of the models

On the Amsterdam data set, the wavelet SVM showed the highest accuracy (0.76), while 2D CNN with approach B, which had an accuracy of 0.72 on the Amsterdam data set, performed best on the Murcia data set, with an accuracy of 0.68 as compared with 0.58 from the wavelet SVM. It is clear from our results that wavelet models did not generalize well for a different population, which might indicate that the features extracted by the discrete wavelet transform might not be informative enough across different data sets. 2D CNN with approach B resulted in the best DL models, where the learned convolutional kernels were shared among all leads, instead of learning individual kernels per lead (approach A). The standard deviation for accuracy and sensitivity for 2D CNN with approach B was also one of the lowest, showing that the model generalized well across different folds.

Comparison with previous studies

This study is the first to evaluate the diagnosis of a PLN *p.Arg14del* mutation by using solely the ECG. Also, our

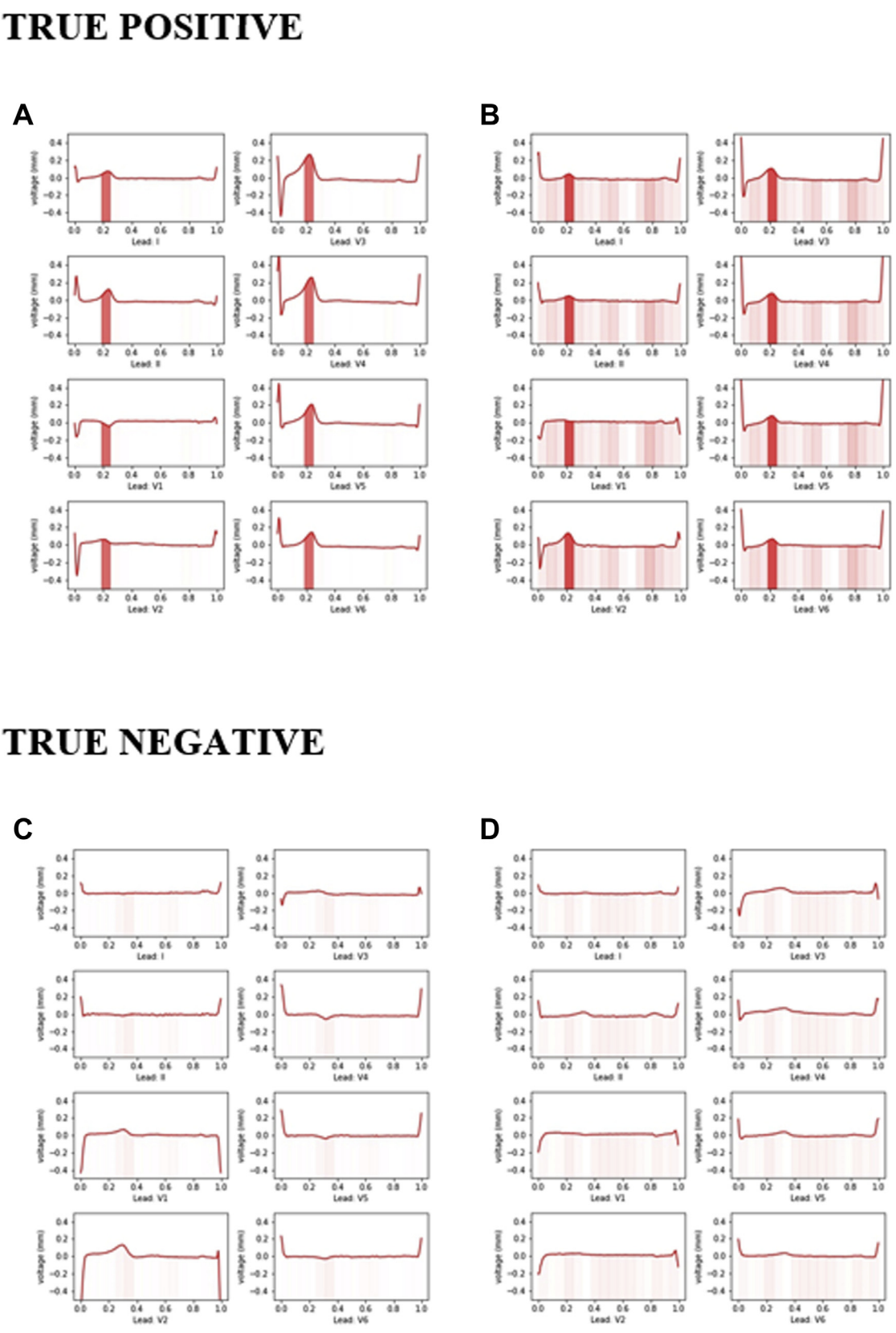

Figure 4 Examples of visualization of electrocardiographic (ECG) features that the convolutional neural network used for classification by using occlusion maps on unique ECGs. **A and B:** Phospholamban (PLN) ECGs correctly classified by the model as PLN. **C and D:** ECGs form control while were correctly classified as non-PLN. The *red highlighted areas* are the parts of the signal that the model used to classify. If no specific area was highlighted, this means the model used the whole signal for classification.

study is the first to use both ML and DL to prove this concept. In the field of cardiogenetics and ECG analysis, Hermans et al[19] recently added T-wave morphology characterizations to age, sex, and corrected QT interval in an SVM and improved the diagnosis of long QT syndrome on the ECG. In the Amsterdam data set of our study, wavelet ML models also proved to perform with the highest accuracy. However, Hermans et al[19] did not use a DL approach, in which the model learns features by itself.

A recent study from Mayo Clinic developed an ML model to diagnose hypertrophic cardiomyopathy.[20] Their model outperform ours; however, their data set is much larger, and furthermore, hypertrophic cardiomyopathy is a diagnosis based on a, for example, echocardiographic phenotype, not solely based on the prevalence of a gene mutation.[21] Therefore, to our knowledge, this study is the first to use DL to detect a genetically proven structural heart disease by solely looking at the ECG in a data set including asymptomatic mutation carriers.

Several studies have used ML- or DL-based ECG analysis to diagnose cardiac arrhythmia, with a higher accuracy than our ML models. An example is the DL model of Hannun et al,[11] which is a neural network for the automatic detection of cardiac arrhythmia on the ECG and which was trained on a much higher number of ECGs than we used in our study. PLN is a rare disease worldwide, and it would be impossible to reach the same number of patients as they have included.

Visualization of ECG features in the DL model

Many techniques have been developed to interpret these models and give insight into their decision process. To our knowledge, we are the first to use a DL-based approach to identify ECG features associated with genetic structural heart disease. In the majority of our correctly classified PLN ECGs, the model used the T wave as its most important ECG feature. Although low QRS voltages are seen as the main ECG feature, T-wave inversions are also common in PLN *p.Arg14del* mutation carriers.[1,2,5,6] More focused research will be needed to further elaborate these findings and to identify more specific and potentially new ECG features.

Clinical interpretation

To implement our models in a clinical setting, first their performance has to increase. The criterion standard for the diagnosis of a PLN *p.Arg14del* mutation is genetic testing. Models such as ours are unlikely to replace genetic testing as a whole, but can serve as a risk stratification tool to predict which patients do need genetic testing. This is currently done by (expert) physicians, and now that our models have shown to outperform the sensitivity of expert cardiologists, our results could contribute to improved and earlier diagnosis of this progressive genetic cardiomyopathy. Because sensitivity of our models is higher than that of the experts, the models are better at diagnosing PLN mutation carriers, compared with the assessment of the expert cardiologist, in the current setting. When looking at specificity, it is the other way around; the experts outperform the models, with almost a maximum specificity and therefore often correctly classifying an ECG as a PLN *p.Arg14del* mutation carrier.

PLN is not the only genetic heart disease that has a "typical" phenotype on the ECG. Other diseases such as long QT syndrome, hypertrophic cardiomyopathy, and Brugada syndrome are only a few examples of syndromes in which a gene mutation can lead to a clinically severe and life-threatening syndrome. This study suggests that an ML/DL-based approach could also be used for the diagnosis of these inherited cardiac syndromes.

Limitations

This study is performed on a (relatively) small data set. It is known that DL is a technique that is highly dependent on the amount of data it is trained on. Because PLN-associated cardiomyopathy is a rare disease, it is difficult to bring together a much larger number of patients with PLN. We augmented our data by using multiple beats from a single ECG as individual samples. Moreover, we decided to first use patients only from our own center to prove the concept that it is possible to predict the carrier status of a specific mutation leading to heart disease by using ML/DL-based ECG analysis.

Also, for this analysis we chose to evaluate genetically proven carriers of the PLN *p.Arg14del* mutation, which were either symptomatic or asymptomatic. This was done to identify possible ECG features, which are present in both these groups and not only in symptomatic patients.

The main goal of this study was to evaluate the predictive value for the ECG for the diagnosis of PLN-associated cardiomyopathy. Therefore, we did not include basic demographics such as age and sex, especially because of the risk of bias given these parameters could influence the ECG by itself.[22]

Conclusion

This study has shown that ML and DL can improve the diagnosis of PLN *p.Arg14del* cardiomyopathy, and our results find regions of the surface ECG that are related to PLN *p.Arg14del* mutations and therefore suggest that the T wave is of added importance to diagnose PLN mutation–caused heart disease even before they become symptomatic.

Acknowledgments

We thank Willeke van der Stuijt, MD and Virginnio Proost, MD (Amsterdam UMC) for their help with this study.

Appendix

Supplementary data

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.hrthm.2020.08.021.

References

1. van Der Zwaag PA, van Rijsingen IAW, Asimaki A, 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.
2. van Rijsingen IAW, van der Zwaag PA, Groeneweg JA, et al. Outcome in phospholamban R14del carriers. Circ Cardiovasc Genet 2014;7:455–465.
3. Bosman LP, Verstraelen TE, van Lint FHM, et al. The Netherlands Arrhythmogenic Cardiomyopathy Registry: design and status update. Neth Heart J 2019; 27:480–486.
4. van der Zwaag PA, van Rijsingen IAW, de Ruiter R, et al. Recurrent and founder mutations in the Netherlands—phospholamban p.Arg14del mutation causes arrhythmogenic cardiomyopathy. Neth Heart J 2013;21:286–293.
5. Posch MG, Perrot A, Geier C, et al. Genetic deletion of arginine 14 in phospholamban causes dilated cardiomyopathy with attenuated electrocardiographic R amplitudes. Heart Rhythm 2009;6:480–486.
6. Cheung CC, Healey JS, Hamilton R, et al. Phospholamban cardiomyopathy: a Canadian perspective on a unique population. Neth Heart J 2019;27:208–213.
7. Towbin JA, McKenna WJ, Abrams DJ, et al. 2019 HRS expert consensus statement on evaluation, risk stratification, and management of arrhythmogenic cardiomyopathy. Heart Rhythm 2019;16:e301–e372.
8. Hof IE, van der Heijden JF, Kranias EG, et al. Prevalence and cardiac phenotype of patients with a phospholamban mutation. Neth Heart J 2019;27:64–69.
9. Krittanawong C, Johnson KW, Rosenson RS, et al. Deep learning for cardiovascularmedicine: a practical primer. Eur Heart J 2019;40:2058–2073.
10. Lecun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–444.
11. Hannun AY, Rajpurkar P, Haghpanahi M, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat Med 2019;25:65–69.
12. Chattopadhay A, Sarkar A, Howlader P, et al. Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV; 2018; p. 839–847.
13. Baalman SWE, Schroevers FE, Oakley AJ, et al. A morphology based deep learning model for atrial fibrillation detection using single cycle electrocardiographic samples. Int J Cardiol 2020;316:130–136.
14. Martis RJ, Acharya UR, Min LC. ECG beat classification using PCA, LDA, ICA and discrete wavelet transform. Biomed Signal Process Control 2013; 8:437–448.
15. Kumar M, Pachori RB, Acharya UR. Characterization of coronary artery disease using flexible analytic wavelet transform applied on ECG signals. Entropy 2017; 19:488.
16. McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika 1947;12:153–157.
17. Zeiler MD, Fergus R. Visualizing and understanding convolutional networks,in Proc. Eur. Conf. Comput. Vis. New York, NY, USA: Springer; 2014. p. 818–833.
18. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960; 20:37–46.
19. Hermans BJM, Bennis FC, Vink AS, et al. Improving long QT syndrome diagnosis by a polynomial-based T-wave morphology characterization. Heart Rhythm 2020;17:752–758.
20. Ko WY, Siontis KC, Attia ZI, et al. Detection of hypertrophic cardiomyopathy using a convolutional neural network-enabled electrocardiogram. J Am Coll Cardiol 2020;75:722–733.
21. Authors/Task Force members, Elliott PM, Anastasakis A, Borger MA, et al. 2014 ESC Guidelines on diagnosis and management of hypertrophic cardiomyopathy: the Task Force for the Diagnosis and Management of Hypertrophic Cardiomyopathy of the European Society of Cardiology (ESC). Eur Heart J 2014; 35:2733–2779.
22. Attia ZI, Friedman PA, Noseworthy PA, et al. Age and sex estimation using artificial intelligence from standard 12-lead ECGs. Circ Arrhythm Electrophysiol 2019;12:e007284.