A multilayer CNN using the ECG, Age and Sex Predicts Ventricular Arrhythmias in the General Population

Julia Ramírez[1,2,3], Antonio Miguel[1], Stefan van Duijvenboden[3,4], Michele Orini[5], William J Young[3,6], Andrew Tinker[3], Pier D Lambiase[3,6], Patricia B Munroe[3], Juan Pablo Martínez[1,2]

[1] Instituto de Investigación en Ingeniería de Aragón, IIS-Aragón, University of Zaragoza, Spain
[2] Biomedical Research Networking Centre on Bioengineering, Biomaterials and Nanomedicine, Zaragoza, Spain
[3] Queen Mary University of London, London, United Kingdom
[4] University of Oxford, Oxford, United Kingdom
[5] University College London, London, United Kingdom
[6] St Bartholomew's Hospital, London, United Kingdom

Abstract

Life-threatening ventricular arrhythmias (LTVA) prediction in individuals without cardiovascular disease remains a major challenge. We tested the performance of a multilayer convolutional neural network (CNN) using ECG signals, age and sex. We split 86,603 individuals from the UK Biobank study into a training (90%) and a test (10%) set. In the training set, we trained a multilayer CNN using 15-second ECGs at rest from lead I, age and sex as inputs. The output was the probability of LTVA within a 12-year follow-up. The CNN model consisted of a four-layer CNN (128, 128, 256 and 256 channels, kernel sizes of 3) and a single attention layer. Age and sex were included as external inputs to the final layer. In the test set (0.9% LTVA events), the CNN's prediction led to a median AUC of 0.601, and a specificity of 0.287 for a sensitivity of 0.750. We set a threshold at the CNN's prediction value maximising the sum of specificity and sensitivity in the training set. Survival analyses showed a hazard ratio (HR) of 1.396 (P = 0.021) for individuals with a CNN's prediction value > threshold, versus those with a CNN's prediction value < threshold. A multilayer CNN model using 10-second ECG data from lead I, together with information on age and sex, can stratify individuals at risk of LTVA. Our findings support the potential utility of wearables for accessible screening in the general population.

1. Introduction

Life-threatening ventricular arrhythmias (LTVA) may occur before any warning or previous diagnosis of underlying heart disease in 50% of cases [1]. Population-based screening for LTVA is challenging because the yearly incidence in the general population is low [2, 3] and current strategies are heavily reliant on imaging assessment of ventricular function. There is a lack of easily accessible and non-invasive methodologies to accurately predict individuals at high risk.

The electrocardiogram (ECG) is an ideal candidate for large-scale screening, especially with the recent advances in the development and availability of wearable devices [4]. Advanced artificial intelligence (AI)-based methods, like convolutional neural networks (CNN) are able to identify subtle variations across the ECG signal that might indicate an unknown underlying cardiac pathology, but may be missed by previously proposed ECG indices [5, 6].

In this work, we trained a multilayer convolutional NN (CNN) using 10-second 1-lead ECGs, age and sex to predict LTVA in middle-age volunteers without cardiovascular disease.

2. Materials and Methods

2.1. Study design

We analysed ECG recordings from 86,603 individuals from the UK Biobank study (application 8256, the study was approved by an institutional review committee and all subjects gave informed consent) who either participated in an exercise bicycle test (N = 51,444) or in the imaging study (N = 35,159), whichever occurred first. Participants in the exercise test had a 15-second single lead (lead I) ECG recording prior to exercise, which was used in this analysis. Participants in the imaging study had a 12-lead 15-second resting ECG recorded, but only lead I was used in this work. Individuals were excluded if they had a prior diagnosis of cardiovascular disease or if the ECG peak-to-peak amplitude after pre-processing was lower than 0.25

mV or higher than 5 mV.

The primary endpoint was LTVA, defined as LTVA mortality or admission to hospital with an LTVA diagnosis. Follow-up was from the study inclusion date until March 25th, 2023 (median follow-up of 10.05 years).

Pre-processing included band-pass filtering with cut-off frequencies 0.5Hz to 40Hz. The 86,603 individuals were split into training (90%) and test (10%) sets. The training set was first up-sampled to the majority class (to compensate for the case and control imbalance). Then, it was divided into 10 equal parts for cross-validation (in each iteration, 90% of the training set was fed into the multilayer CNN, and the model was evaluated on the remaining 10% validation set).

2.2. Architecture of the multilayer CNN

We used the PyTorch library in Python [7] to train a multilayer CNN and leverage the distinct morphological patterns of the ECG with an attention layer.

The multilayer CNN consisted of three 1-dimensional CNNs, each followed by a rectified linear unit, with 128, 128 and 256 channels, kernel sizes of 3, groups of 1, 2 and 1, respectively, and a stride of 2 (Figure 1). Following these 3 layers, we included two parallel blocks, one including a 1-dimensional CNN with a rectified linear unit, and one including an attention mechanism, consisting of a convolutional block and a softmax. The sum of the outputs from the CNN and the attention blocks was then concatenated with the information from age and sex, and fed into a fully connected layer, followed by a softmax.

The 15-second ECG strip at rest (lead I) was split into 3 second windows, with a 50% overlap, which were the input to the CNN. Standardised age and sex (1 being male and -1 being female) were inputted into the model before the fully connected layer (Figure 1). The output of the multilayer CNN model is a value between 0 and 1 predicting the occurrence (or not) of LTVA within the follow-up period (the multilayer CNN's LTVA prediction, which can be interpreted as a probability ranging from 0 to 1).

2.3. Computation of other ECG Indices

We first calculated an average heartbeat using signal averaging. Then, we calculated previously proposed ECG indices. In particular, we derived the RR interval, the QRS duration, the QT interval, the T-peak-to-T-end (Tpe) interval, T-wave amplitude (TAmp) and the T-wave morphology variations (TMV) index [8]. The QRS duration and QT interval were measured as the interval between the QRS-onset and QRS-offset, and between the QRS-onset and the T-wave end, respectively, from the averaged heartbeat at rest. Then, we corrected the QT interval using Bazett formula [9]. The Tpe interval was calculated as

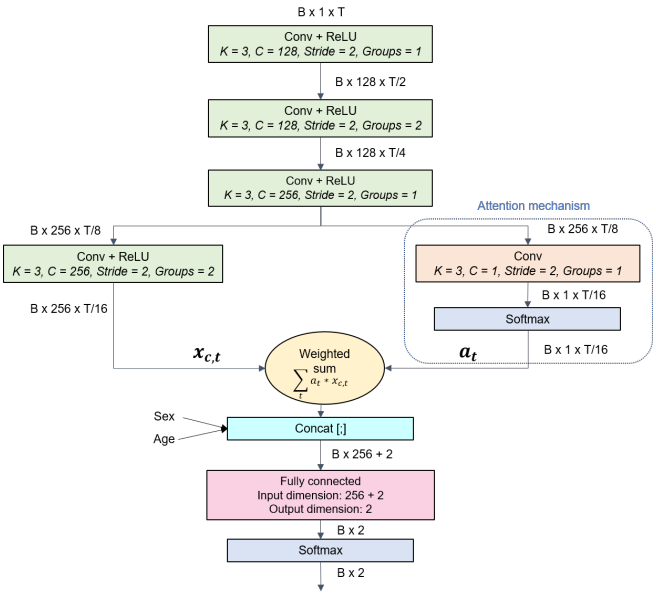

Figure 1. Architecture of the multilayer convolutional neural network.

the time interval from the peak to the end of the T-wave and TAmp measured the amplitude of the T-wave peak. Finally, we calculated TMV by comparing the average T-wave from each participant with their corresponding sex- and RR- normal T-wave morphology reference in lead I from a normal reference cohort, as previously described [8].

2.4. Statistical Analyses

We applied the logit transformation to the multilayer CNN's prediction to expand the range from 0-1 to -infinite to +infinite. Then, it was standardised, as the other ECG indices.

The area under the curve (AUC) was used to estimate the performance of the CNN model and of the previously proposed ECG indices. We trained the model three times to obtain mean and standard deviation AUC values. As we aim at developing a tool to be used as an initial step in identifying individuals that might benefit from early screening, we aimed for a high sensitivity. Therefore, for each of these parameters, we provide values of specificity for set sensitivity values of 0.6 to 0.95.

We derived Kaplan-Meier curves by setting a threshold at the cut-off value that maximised the sum of sensitivity and specificity in the training set, as in previous studies [8, 10], with a comparison of cumulative events performed by using log-rank tests, and plotted using the "survminer" package in R. Univariable Cox regression analyses were performed to determine the predictive value of the multilayer CNN model. Individuals who died from causes not

included in the primary end point, or individuals who did not reach the follow-up time, were censored at the time of death.

We repeated the analyses by stratifying the individuals in the test set by sex (men and women) and age (younger and older than 65 years old). Statistical analyses were performed using R version 4.0.2.

3. Results

In the test set, 8,729 (50.4%) individuals did not reach the full follow-up time and were censored from the study. From the remaining 8,591 subjects, 78 (0.9%) had an LTVA event.

After repeating the training of the multilayer CNN model three times (with different random partitions in the cross-validation step), we observed a range of variation across the three AUC values of 0.078 (minimum AUC of 0.590, maximum AUC of 0.668).

The median (interquartile range) AUC values obtained after bootstrapping in the training and in the test set for the multilayer CNN model and for the previously proposed ECG indices are shown in Figure 2. We observed that the multilayer CNN model outperformed all ECG indices, being the QRS duration the only index with an AUC higher than 0.5 in both training and test sets.

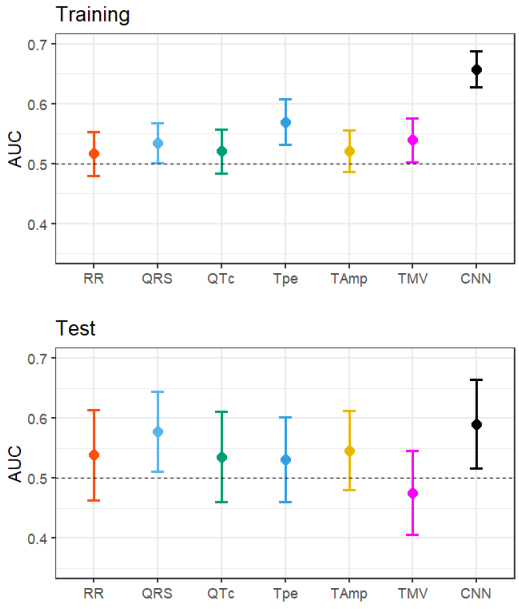

Figure 2. Median (interquartile range) area under the curve values of the multilayer convolutional neural network model and previously-proposed ECG indices.

We compared the specificity at increasing sensitivity levels for LVTA risk detection of the multilayer CNN model and of the QRS duration (being the only significant index, as shown in Figure 2) in the test set. As shown in Table 1, on average, the multilayer CNN showed higher specificity values than the QRS duration for the same sensitivity.

Sensitivity	QRS duration	CNN
0.600	0.434	0.493
0.650	0.413	0.389
0.700	0.366	0.389
0.750	0.310	0.287
0.800	0.262	0.195
0.850	0.236	0.156
0.900	0.146	0.078
0.950	0.014	0.033

Table 1. Specificity values of the QRS duration and the multilayer convolutional neural network model at different sensitivity values.

Individuals in the high-risk group defined by the optimal cut-off threshold of the multilayer CNN had a hazard ratio of 1.40 (P = 0.021) (Figure 3). The continuous values of CNN also remained significantly associated with LTVA risk (HR of 1.43 per standard deviation of CNN - 1 standard deviation was 0.231, P = 0.004). Regarding sex-stratified analyses, we only found significant HRs in women older than 65 years old (N = 1,636, 19 LTVA events). The multilayer CNN model showed a HR of 2.55 for women with a score > optimal threshold with respect to women with score < optimal threshold (P = 0.04).

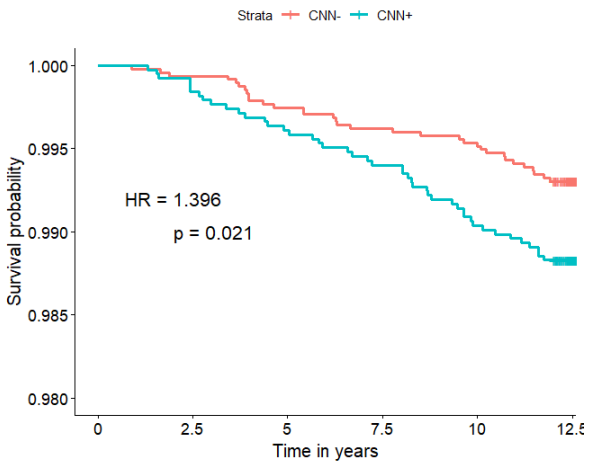

Figure 3. Survival curves for the two risk groups defined by the multilayer CNN model.

4. Discussion and Conclusions

In this study, we have trained a multilayer CNN using 15-second 1-lead ECGs, age and sex, and tested its long-term LTVA predictive value in an independent cohort of individuals without known cardiovascular disease from the UK Biobank. We demonstrate the ability of the multilayer CNN to predict LTVA risk outperforming previous ECG indices, like TMV, the QRS duration or the QT interval.

Prediction of long-term LTVA is a task particularly challenging due to the low number of events in this population [3]. Still, our multilayer CNN model was able to consistently provide robust LTVA predictive value, outperforming previously proposed strong LTVA predictors from the ECG, like TMV [8]. This confirms that a multilayer CNN using 15-second 1-lead ECGs, age and sex, is able to identify ECG features that could be markers of subclinical disease and that are associated with risk for ventricular arrhythmia. The investigation of the output from the attention layer will inform on the areas of the ECG that are contributing to LTVA risk according to the CNN model.

Sex is known to be the strongest risk factor in low-risk populations [3]. In our data, men had 2.7 fold risk than women. Yet, the multilayer CNN model proved to be particularly useful in predicting LTVA risk in older females. These findings have important implications for population screening in individuals at apparently low risk.

However, the number of LTVA cases in the study population was low, which may have limited the optimal performance of the multilayer CNN. In particular, there was a large variability in follow-up times across individuals, with only half the study population reaching 12 years follow-up. Future studies should evaluate the performance of CNN models in predicting LTVA risk at different follow-up times, as done previously [11,12], to optimise resources and provide a better clinical tool. Moreover, our analysis was limited to a single study with a predominance of individuals with European ancestry, so the generalizability to other populations and ancestries must be established.

In conclusion, a multilayer CNN model predicts long-term LTVA in a low-risk population, outperforming previous ECG indices. The model showed to be particularly useful in predicting events in older women. Our algorithm may then be used as a first selection step in LTVA screening to prioritise individuals for more advanced monitoring or treatment strategies. Future studies should investigate the predictive value of CNN models at different follow-up times, or when informed by additional co-morbidities.

Acknowledgements

This work was supported by projects PID2021-128972OA-100, TED2021-130459B-I00 and PID2022-140556OB-I00. JR acknowledges funding from the European Union-NextGenerationEU and fellowship RYC2021-031413-I from MCIN/AEI/10.13039/501100011033 and from the European Union "NextGenerationEU/PRTR". We also wish to acknowledge support by the MRC grant MR/N025083/1. PDL is supported by UCL/UCLH Biomedicine NIHR. AT and PBM acknowledge the NIHR Biomedical Research Centre at Barts and QMUL, UK.

References

[1] M. Hayashi et. al., "The spectrum of epidemiology underlying sudden cardiac death.," *Circ Res*, vol. 116, pp. 1887–1906, 2015.

[2] C. X. Wong et. al., "Epidemiology of sudden cardiac death: Global and regional perspectives.," *Heart, Lung and Circulation*, vol. 28, pp. 6–14, 2019.

[3] K. Zeppenfeld et. al., "2022 esc guidelines for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death: Developed by the task force for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death of the european society of cardiology (esc) endorsed by the association for european paediatric and congenital cardiology (aepc)," *EHJ*, vol. 43, pp. 3997–4126, 2022.

[4] M. V. Perez et. al., "Large-scale assessment of a smartwatch to identify atrial fibrillation.," *NEJM*, vol. 381, pp. 1909–1917, 2019.

[5] T. Hygrell et. al., "An artificial intelligence–based model for prediction of atrial fibrillation from single-lead sinus rhythm electrocardiograms facilitating screening," *EP Europace*, vol. 25, pp. 1332–1338, 2023.

[6] O. Akbilgic et. al., "Ecg-ai: electrocardiographic artificial intelligence model for prediction of heart failure," *European Heart Journal - Digital Health*, vol. 2, pp. 626–634, 2021.

[7] A. Paszke et. al., "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*, pp. 8024–8035, Curran Associates, Inc., 2019.

[8] J. Ramírez et. al., "Ecg t-wave morphologic variations predict ventricular arrhythmic risk in low- and moderate-risk populations," *JAHA*, vol. 11, p. e025897, 2022.

[9] H. C. Bazett, "An analysis of the time-relations of electrocardiograms," *Heart*, vol. 7, pp. 353–370, 1920.

[10] J. Ramírez et. al., "Cardiovascular predictive value and genetic basis of ventricular repolarization dynamics," *Circ Arrhythmia Electrophysiol*, vol. 12, p. e007549, 2019.

[11] P. C. Holm et.al , "Development and validation of a neural network-based survival model for mortality in ischemic heart disease," *medRxiv*, 2023.

[12] D. M. Popescu et. al., "Arrhythmic sudden death survival prediction using deep learning analysis of scarring in the heart," *Nat Cardiovasc Res*, vol. 1, p. 334–343, 2022.

Address for correspondence:

Julia Ramírez, Aragon Institute of Engineering Research
University of Zaragoza
Mariano Esquillor s/n, Lab 6.1.07, 50018, Zaragoza, Spain
Julia.Ramirez@unizar.es