

RESEARCH ARTICLE

Mixed convolutional and long short-term memory network for the detection of lethal ventricular arrhythmia

Artzai Picon¹, Unai Irusta²*, Aitor Álvarez-Gila¹, Elisabete Aramendi², Felipe Alonso-Atienza^{3,4}, Carlos Figuera^{3,4}, Unai Ayala⁵, Estibaliz Garrote¹, Lars Wik⁶, Jo Kramer-Johansen⁶, Trygve Eftestøl⁷

1 Computer Vision Group, Tecnalia Research & Innovation, Derio, Spain, **2** Department of Communications Engineering, University of the Basque Country UPV/EHU, Bilbao, Spain, **3** Department of Signal Theory and Communications, Rey Juan Carlos University, Madrid, Spain, **4** Client Solutions Advanced Analytics, BBVA, Madrid, Spain, **5** Electronics and Computing Department, Mondragon Unibertsitatea, Faculty of Engineering (MU-ENG), Mondragón, Spain, **6** Norwegian National Advisory Unit on Prehospital Emergency Medicine (NAKOS), Oslo University Hospital and University of Oslo, Oslo, Norway, **7** Department of Electrical Engineering and Computer Science, University of Stavanger, Stavanger, Norway

☯ These authors contributed equally to this work.

* unai.irusta@ehu.eus



OPEN ACCESS

Citation: Picon A, Irusta U, Álvarez-Gila A, Aramendi E, Alonso-Atienza F, Figuera C, et al. (2019) Mixed convolutional and long short-term memory network for the detection of lethal ventricular arrhythmia. PLoS ONE 14(5): e0216756. <https://doi.org/10.1371/journal.pone.0216756>

Editor: Steve Lin, St. Michael's Hospital, CANADA

Received: January 21, 2019

Accepted: April 26, 2019

Published: May 20, 2019

Copyright: © 2019 Picon et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The preprocessed databases are the public and ohca databases containing the preprocessed ECGs. The data can be found in: - CUDB database: http://www.tsc.urjc.es/~felipe.alonso/data_ohca/cudb.mat - VFDB database: http://www.tsc.urjc.es/~felipe.alonso/data_ohca/vfdb.mat - AHADB: we used files 8201 - 8210, which can be acquired in <https://www.ecri.org/american-heart-association-ecg-database-usb> - OHCADB: this database is a proprietary database. The owner has stated that access to the data for scientific purposes will be granted but that cannot

Abstract

Early defibrillation by an automated external defibrillator (AED) is key for the survival of out-of-hospital cardiac arrest (OHCA) patients. ECG feature extraction and machine learning have been successfully used to detect ventricular fibrillation (VF) in AED shock decision algorithms. Recently, deep learning architectures based on 1D Convolutional Neural Networks (CNN) have been proposed for this task. This study introduces a deep learning architecture based on 1D-CNN layers and a Long Short-Term Memory (LSTM) network for the detection of VF. Two datasets were used, one from public repositories of Holter recordings captured at the onset of the arrhythmia, and a second from OHCA patients obtained minutes after the onset of the arrest. Data was partitioned patient-wise into training (80%) to design the classifiers, and test (20%) to report the results. The proposed architecture was compared to 1D-CNN only deep learners, and to a classical approach based on VF-detection features and a support vector machine (SVM) classifier. The algorithms were evaluated in terms of balanced accuracy (BAC), the unweighted mean of the sensitivity (Se) and specificity (Sp). The BAC, Se, and Sp of the architecture for 4-s ECG segments was 99.3%, 99.7%, and 98.9% for the public data, and 98.0%, 99.2%, and 96.7% for OHCA data. The proposed architecture outperformed all other classifiers by at least 0.3-points in BAC in the public data, and by 2.2-points in the OHCA data. The architecture met the 95% Sp and 90% Se requirements of the American Heart Association in both datasets for segment lengths as short as 3-s. This is, to the best of our knowledge, the most accurate VF detection algorithm to date, especially on OHCA data, and it would enable an accurate shock no shock diagnosis in a very short time.

be made publicly available. Unai Irusta will act as proxy to facilitate access upon request to Laerdal Medical. Requests should be handled via unai.irusta@ehu.eus.

Funding: This study was supported by the Ministerio de Economía, Industria y Competitividad, Gobierno de España (ES) (TEC-2015-64678-R) to UI and EA and by Euskal Herriko Unibertsitatea (ES) (GIU17/031) to UI and EA. The funders, Tecnalia Research and Innovation and Banco Bilbao Vizcaya Argentaria (BBVA), provided support in the form of salaries for authors AP, AA, FAA, CF, EG, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the author contributions section.

Competing interests: The funders, Tecnalia Research and Innovation and BBVA, provided support in the form of salaries for authors AP, AA, FAA, CF, EG. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

Introduction

Worldwide out-of-hospital cardiac arrest (OHCA) has an average incidence of 55 cases per 100 000 person-year [1], and constitutes one of the leading causes of death in the industrialized world. Lethal ventricular arrhythmia such as pulseless ventricular tachycardia (VT) and ventricular fibrillation (VF) are the most frequent trigger of cardiac arrest [2]. In fact, ventricular arrhythmia are observed as initial rhythm in up to 75% of cases if recorded right after the collapse by an on-site automated external defibrillator (AED) [3]. An electric defibrillation shock is the only effective therapy to revert VF/VT and restore a normal rhythm. AEDs are simple devices designed for untrained lay people, and thus incorporate a shock/no-shock decision algorithm based on the analysis of the patient's electrocardiogram (ECG) [4]. When the algorithm detects a shockable rhythm (VF/VT) the device delivers a defibrillation shock to restore a perfusing rhythm, otherwise the AED recommends continuation of cardiopulmonary resuscitation (CPR). The American Heart Association (AHA) established the minimum accuracy requirements for shock decision algorithms to ensure a safe use of the device, and that the adequate therapy is administered [5]. Shockable rhythms should be detected with a sensitivity (positive class) above 90%. The specificity for nonshockable rhythms (negative class) should be above 99% in case of normal sinus rhythms, and above 95% for other arrhythmia like atrial fibrillation, supraventricular tachycardia, ideoventricular rhythms, heart blocks, or bradycardia.

The development of automatic shock decision algorithms has been an active field of research for decades. The ECG of lethal ventricular arrhythmia present some distinctive characteristics such as more waveform irregularity, the absence of narrow or wide QRS complexes (the ECG waveform deflections associated to ventricular contraction), smaller bandwidths, and higher ventricular rates. Consequently, initial efforts focused on the development of ECG features to identify VF, including features to quantify amplitudes [6, 7], waveform correlation [8], heart rate [9, 10], slope [11, 12], spectral [13–15] and time-frequency characteristics [16–18], or the complexity of the VF waveform [19–21]. Many of these features have been systematically reviewed and compared [22–25] using data from public ECG repositories available through the physionet initiative [26]. Current findings, however, support the use of ECG from OHCA patients gathered by defibrillators during treatment because the ECGs may be very different both for shockable and nonshockable rhythms [27]. Recently, efforts have focused on efficiently combining ECG features for VF detection through state of the art machine learning techniques like support vector machines (SVM) or random forests [27–33]. These latter studies explore the limits of the accuracy attainable by a feature extraction approach using both data from public repositories [28, 29, 31, 32], and from OHCA cases [27, 30].

Machine learning paradigms based on feature extraction and feature selection strongly depend on the design of the features. An alternative approach is to exploit all available knowledge from the signals and to let the machine learning algorithm learn and select those features. This is the rationale behind deep learning, and in particular convolutional neural networks (CNN), and their one dimensional counterparts 1D-CNN, which have been recently introduced for patient specific heartbeat classification [34], arrhythmia classification [35], detection of myocardial infarction [36], detection of ectopic beats [37], and the detection arrhythmia [35] including atrial fibrillation [38] and VF [39]. ECG-based arrhythmia classification using CNNs has been shown to be more accurate than that of clinicians [40, 41], and 1-D CNNs have outperformed classical feature extraction-classification paradigms in ECG beat classification [42]. When there are time dependencies and variable segment lengths in the data Recurrent Neural Networks (RNN) are an efficient deep learning approach [43], and in particular some of its variants like Long Short-Term Memory (LSTM) networks [44, 45]. In fact, LSTMs

have been proven accurate in diagnostics [46], but also for the detection of anomalous heart-beat contractions using the ECG [47].

This study introduces a deep learning approach that combines a CNN and an LSTM network for the detection of lethal ventricular arrhythmia using data from both public ECG repositories and from defibrillators used to treat OHCA patients. The paper is organized as follows. The methodology is described in materials and methods, including a description of the data, the mixed CNN+LSTM architecture proposed in this paper, and a classical machine learning approach using SVMs used for baseline comparisons. The results section compares the baseline classifier and 3 deep learning architectures based only on CNNs to our mixed CNN+LSTM architecture, analyzes the influence of shortening the analysis segment length, and evaluates the classification features learned by our deep learner. In the discussion the results are put in the context of previous findings on the detection of lethal ventricular arrhythmia, with emphasis on the importance of using OHCA data in these studies.

Materials and methods

Data preparation

The data used in this study come from a previous study on classical machine learning approaches for the detection of ventricular arrhythmia [27]. What follows is an abridged description of the data, the ECG preprocessing, and data partitioning for training and testing. For further details consult [27].

Datasets. The data comprise two datasets, one originated from public repositories of arrhythmia (public database), and a second dataset from OHCA data recorded by monitor-defibrillators during treatment (OHCA dataset). Fig 1 shows four characteristic examples of 4-s ECG segments from both databases. The examples illustrate the differences between the ECGs found in public databases, which come from continuously monitored patients (Holter records), and the rhythms seen by a monitor-defibrillator attached to patient several minutes after the onset of cardiac arrest.

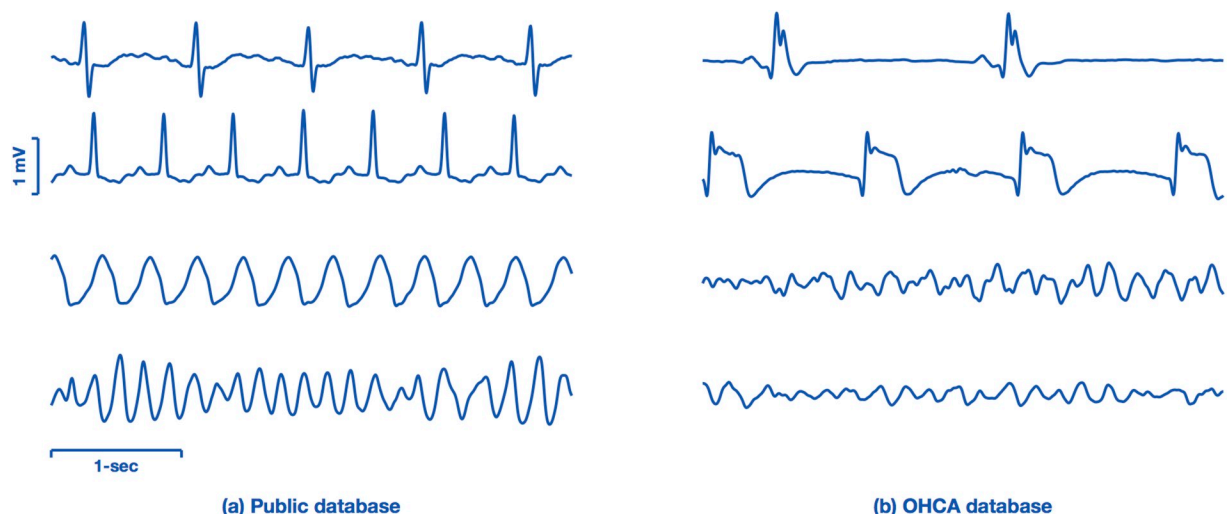


Fig 1. ECG samples from the public and OHCA datasets. The examples on the left panel show ECGs from public databases, ranging from normal (top) to VT and VF in the bottom most examples. The ECGs on the right come from the OHCA dataset, and the two topmost cases are organized rhythms (without pulse) but with very aberrant QRS complexes. The two examples in the bottom are VF observed minutes after the arrest occurred, and have smaller amplitudes and fibrillation frequencies than those observed in the public databases (bottom-left).

<https://doi.org/10.1371/journal.pone.0216756.g001>

The public dataset was formed using three databases. It includes the complete record set from the physionet MIT-BIH Malignant Ventricular Arrhythmia Database (VFDB) and the Creighton University Ventricular Tachyarrhythmia Database (CUDB), and ten episodes (8201-8210) from the series 1 American Heart Association ECG Database (AHADB) [26]. When more than one ECG channel were available (VFDB and AHADB) only the first channel was extracted to avoid redundancies, as done in recent studies on this topic [27, 39, 48]. The records included arrhythmia annotations in the Physionet's scheme, and all databases included recordings with lethal ventricular arrhythmia from the onset of the arrhythmia. The sample rate of all databases was $f_s = 250$ Hz.

The OHCA database originated in a multicenter OHCA study conducted in Akershus (Norway), Stockholm (Sweden), and London (UK) between 2002-2004 [49, 50]. The original studies were approved by the regional ethics committees and complied with the Helsinki declaration. Rhythms were annotated by expert clinicians into five classes [30]: VF, VT, asystole (AS), and organized rhythms with pulse (Pulsed Rhythms, PR) and without pulse (Pulseless Electrical Activity, PEA). ECG segments of 10 s duration and without CPR artifacts were included in the dataset. The ECG was acquired through the defibrillation pads with a resolution of $1.031 \mu\text{V}$ per least significant bit and a sampling frequency of 500 Hz. The ECG was resampled to $f_s = 250$ Hz.

ECG preprocessing and data labeling. The ECG was filtered in all datasets following a typical preprocessing introduced in [24] which consists of: mean subtraction, a moving average filter, and bandpass filtering in the typical 1 – 30 Hz AED monitoring bandwidth [51, 52]. This is the typical preprocessing found in AEDs, although the cutoff frequencies may differ across AED models. Preprocessing removes low frequency noise (movement, respiration, . . .) and high frequency noise (power line interference) that may confound the classification algorithms.

Ground truth rhythm labels were reviewed by consensus among two experienced biomedical engineers specialized in cardiac arrest rhythms. The process included: identification of device saturation intervals, annotation of noise intervals, identification of very low amplitude VF (peak amplitude $< 200 \mu\text{V}$), and slow VT (rate under 150 bpm), and identification of asystole defined as rhythms with very low rates (under 12 bpm) and/or peak-to-peak amplitude ($< 100 \mu\text{V}$).

The ECG signals were then divided into non-overlapping segments of 2, 3, 4, and 8-s, to analyze the accuracy of the shock decision methods in terms of the ECG segment length. Segments with noise, low-quality signal or artifacts were discarded. Following the AHA statement, only segments with a unique rhythm label were considered. Slow VT and fine VF were discarded because the benefits of defibrillation are unclear for these rhythms [5]. Consequently, they cannot be unequivocally labeled as shockable or nonshockable. Finally, segments labeled as AS were also discarded [27, 29], because normally AS is identified before the shock decision algorithm using simple thresholding based on the amplitude/power of the ECG segment [11, 53]. In total 86.5% of available data was used from the public dataset, 96.3% of the AHADB, 82.6% of the VFDB and 85.2% of the CUDB. This was the subset used by Figuera et al [27], and the curated public datasets with noise and unusable signal annotations are available from there.

Table 1 shows the final datasets for the different segment lengths grouped as shockable and nonshockable. Shockable rhythms include VF, VT, and ventricular flutter, and nonshockable rhythms include organized rhythms (PEA/PR) from the OHCA database, and normal sinus rhythm, supraventricular tachycardia, atrial fibrillation, heart blocks, or ectopic ventricular activity from public databases.

Table 1. Dataset used for shock decision algorithms for different segment lengths grouped by shockable (Sh) and nonshockable (NSh) segments.

Database	2-s segments		3-s segments		4-s segments		8-s segments	
	Sh	NSh	Sh	NSh	Sh	NSh	Sh	NSh
Public	7407	29365	4842	19450	3578	14495	1696	7086
vfdb	3318	15757	2149	10427	1586	7761	746	3780
cudb	1510	6075	981	4015	716	2986	323	1446
ahadb	2579	7533	1712	5008	1276	3748	627	1860
OHCA	1700	3235	1020	1941	680	1294	340	647

<https://doi.org/10.1371/journal.pone.0216756.t001>

Data partitioning. Data was partitioned patient-wise into two sets, 80% of patients to train the classification algorithms and 20% to report the performance of the algorithms. The split was done randomly on each of the four databases separately (VFDB, CUDB, AHADB, and OHCA), and it is the same split used in the study that originated the datasets [27]. Then, all four training sets were merged into a single dataset to train the models. The results were evaluated independently for two test sets, the public test set obtained merging the three test set from the public databases (VFDB, CUDB, AHADB), and the OHCA test set.

Deep learning architecture

Proposed architecture. Fig 2 shows the architecture of our deep learning classifier, which comprises three stages: a convolutional block with two 1D-CNN stages, an LSTM network, and a classification stage. The convolutional block extracts the high level descriptors from the ECG signal, and creates a feature map representing the ECG signal that maintains its temporal order. The long and short time relations in the feature map are integrated by the LSTM network that outputs Q features that are used by a dense neural network that outputs $p_{Sh} \in [0, 1]$, the likelihood that the ECG corresponds to a shockable rhythm. p_{Sh} was used for the shock/no-shock decision.

1D-CNN. The convolutional block is composed of two convolutional sub-blocks with a similar architecture. Each sub-block starts with a convolutional layer of $J = 32$ filters of size $M = 3$, followed by a rectified linear unit (ReLU) as nonlinear activation function. The ReLU unit is followed by a max-pooling unit of size $K = 7$ to reduce the dimensionality. The two convolutional subblocks result in a time-compressed signal rich enough for feature extraction.

Mathematically the architecture (in the forward direction) of the network is explained as follows. Let us represent the input signal segment as a column vector of N ordered samples by:

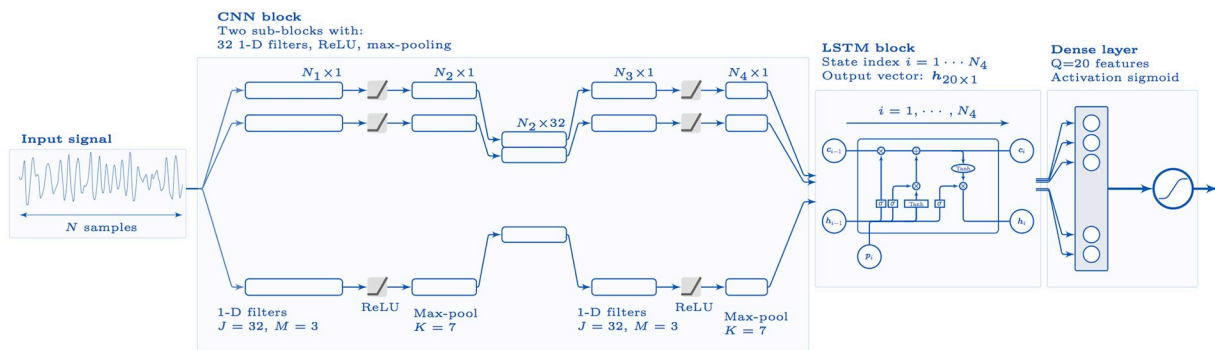


Fig 2. Architecture of the deep learning network. The architecture has three blocks: a convolutional block (CNN), a recursive block (LSTM), and a final decision stage based on a neural network.

<https://doi.org/10.1371/journal.pone.0216756.g002>

$\mathbf{s} = \{s_i\}_{i=1, \dots, N} = (s_1, s_2, \dots, s_N)^T$, where $N = L \cdot f_s$ for a segment of length L seconds. A convolutional layer with $j = 1, \dots, J$ filters of size M and ReLU activation function $f(\cdot)$, has the following outputs:

$$c_i^j = f\left(b_j + \sum_{m=1}^M \omega_m^j \cdot s_{i+m-1}^j\right) \quad \text{with } f(\cdot) = \max\{0, \cdot\}, \tag{1}$$

where ω_m^j is the m -th coefficient in the j -th filter of the network, and s^j its input signal. At the max-pooling stage each output signal is down-sampled by a factor of K , the pool size, by applying maximum filter to non-overlapping K sample segments:

$$p_i^j = \max\{c_k^j\}_{k=(i-1) \cdot K + 1, \dots, i \cdot K} \tag{2}$$

The temporal length, N_ℓ , at each of the four phases ($\ell = 1, 2, 3, 4$) of the two-stage CNN network is given by (see Fig 2):

$$N_\ell = \text{floor}\left(\frac{N_{\ell-1} + 2p - M}{s}\right) + 1 \tag{3}$$

where $N_0 = N$ is the length of the input signal, $p = 0$ is the padding, M is the kernel size (filter size, max-pooling size), and s is the stride which is $s = 1$ for the convolutional layers and $s = K$ for the max-pooling layers. So for an 8-s segment the successive lengths would be ($M = 3$ and $K = 7$):

$$N = 8 \cdot f_s = 2000 \Rightarrow N_1 = 1998 \Rightarrow N_2 = 285 \Rightarrow N_3 = 283 \Rightarrow N_4 = 40 \tag{4}$$

and the input to the LSTM network would have a temporal index of $i = 1, \dots, 40$.

LSTM. The output of the convolutional network feeds the LSTM network in which the vector with the current state (value) of the input signal is $\mathbf{p}_i = (p_i^1, \dots, p_i^J)^T$ with $i = 1, \dots, N_4$. The LSTM network maps the current output signal state $\mathbf{h}_i = (h_i^1, \dots, h_i^Q)^T$ to the previous states of the input and output by means of a network of consecutive recurrent cells that consist of various gates with sigmoid activation functions $\sigma(\cdot)$.

The equations for the LSTM network in the forward direction are [43–45]:

$$\mathbf{f}_i = \sigma(\mathbb{W}_f \mathbf{p}_i + \mathbb{R}_f \mathbf{h}_{i-1} + \mathbf{b}_f) \tag{5}$$

$$\mathbf{i}_i = \sigma(\mathbb{W}_i \mathbf{p}_i + \mathbb{R}_i \mathbf{h}_{i-1} + \mathbf{b}_i) \tag{6}$$

$$\mathbf{o}_i = \sigma(\mathbb{W}_o \mathbf{p}_i + \mathbb{R}_o \mathbf{h}_{i-1} + \mathbf{b}_o) \tag{7}$$

$$\mathbf{c}_i = \mathbf{f}_i \odot \mathbf{c}_{i-1} + \mathbf{i}_i \odot \tanh(\mathbb{W}_c \mathbf{p}_i + \mathbb{R}_c \mathbf{h}_{i-1} + \mathbf{b}_c) \tag{8}$$

$$\mathbf{h}_i = \mathbf{o}_i \odot \tanh(\mathbf{c}_i) \tag{9}$$

where the $\mathbb{W}_{Q \times J}$ and $\mathbb{R}_{Q \times Q}$ are the input and recurrence weight matrices, respectively, and $\mathbf{b}_{Q \times 1}$ are the bias vectors. The number of output features of our network is $Q = 20$. The $\mathbf{f}_i, \mathbf{i}_i, \mathbf{o}_i, \mathbf{c}_i$ vectors correspond to the forget, input, output, and state vectors of the i -th cell in the LSTM architecture, and \odot denotes Hadamard or element-wise product.

Fully connected layer: The 20 feature output vector of the LSTM network, \mathbf{h}_i for $i = N_4$, is densely connected to a classical neural network with one sigmoid output layer that outputs p_{Sh} . Finally, p_{Sh} is compared to a threshold to output a binary shock/no-shock decision. The

decision threshold that maximized the balanced accuracy of the training set was selected once the model was optimized.

At the training stage, the network was optimized to minimize the binary cross-entropy loss function (L), defined as:

$$L = \sum_i \eta_i [y_i \ln(p_{sh_i}) + (1 - y_i) \ln(1 - p_{sh_i})] \tag{10}$$

where η_i is the weight of instance i . Weights were assigned to address class imbalance and to give equal weight to shockable and non-shockable classes during training. Optimization was done using Stochastic Gradient Descent with Nesterov acceleration, a learning rate of 10^{-3} , a learning rate decay of 10^{-6} , and a momentum of 0.9. These are recommended/typical values used in Stochastic Gradient Descent from the specialized literature on the topic [54, 55]. One of the advantages of our design is that the three blocks of the network are optimized end-to-end to simultaneously extract the high level feature description of the signal (convolutional block), its temporal relationships (recurrent network block), and the arrhythmia classification (classification block).

CNN only networks. A number of CNN only networks have been previously introduced for heartbeat [34, 56] and arrhythmia classification [39]. These approaches proved that it is possible to obtain high level descriptors of the ECG using only 1-D convolutional blocks, although they differed on the characteristics of the networks such as depth, filter lengths or max-pooling size. We have implemented three representative examples (see Table 2) and compared their performance to our architecture, in which an additional LSTM block is added to better characterize the temporal relations of the arrhythmia.

All the experiments on the deep learning models were done using a Tesla V100-DGXS-16Gb GPU, and were based on the Keras framework [57] with Tensorflow backend [58]. The CNN only and the proposed CNN-LSTM architectures were trained using 600 epochs, at each epoch all training data were fed to the models in batches of 256.

Baseline learner. A SVM classifier with Gaussian kernel [59] was optimized to obtain a baseline performance for the shock decision algorithm. SVMs for the detection of VF have been demonstrated before on data from public databases [27, 29, 31], and on OHCA studies [27]. In the feature extraction phase 30 classical VF detection features were computed, comprising all the analysis domains of the ECG. The code for feature extraction is available through [27]. Two configurations of the SVM were tested, with all 30 features, and with the top 8 features as selected in [27] using ensemble methods. These will be termed SVM_{all} and SVM_{sel} in what follows.

Our dataset can thus be represented by a set of instance-label pairs $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\} \in \mathbb{R}^k \times \{\pm 1\}$, where $y_i = 1$ for shockable and $y_i = -1$ for nonshockable rhythms, and the vector \mathbf{x}_i contains the values of the K features for ECG segment i . The decision function of the SVM is found by solving the following maximization

Table 2. Architectures of the CNN only models used for comparison.

CNN	Layers	Filters	Filter size	Max-pool size	Dense layers
Kiranyaz et al [34]	2	32, 16	15, 15	6	1
Zubair et al [56]	3	32, 16, 8	5, 5, 5	2	1
Acharya et al [39]	4	3, 5, 10, 10	5, 5, 5, 4	2	2

<https://doi.org/10.1371/journal.pone.0216756.t002>

problem [59]:

$$\max_{\alpha_i} \left\{ \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j \exp(-\gamma \| \mathbf{x}_i - \mathbf{x}_j \|^2) \right\} \tag{11}$$

$$\text{s.t. : } 0 \leq \alpha_i \leq C \quad \forall i, \quad \text{and} \quad \sum_{i=1}^n \alpha_i y_i = 0 \tag{12}$$

After solving the maximization problem, the vectors that have non-zero Lagrange multipliers (α_i) are called support vectors. The SVM decision function depends only on the support vectors, and can be written as:

$$f(\mathbf{x}) = \text{sign} \left[\sum_{i=1}^{N_s} \alpha_i y_i \exp(-\gamma \| \mathbf{x} - \mathbf{x}_i \|^2) + b \right] \tag{13}$$

where the N_s support vectors and the threshold b are determined in the optimization phase. A rhythm will be classified as shockable for $f(\mathbf{x}) = 1$, or nonshockable for $f(\mathbf{x}) = -1$.

The gaussian kernel SVM has two optimization hyper-parameters, C is the soft margin parameter and γ the width of the gaussian kernel. Bayesian optimization was used to determine the optimal hyper-parameters of the SVM in the training data, and the hyperparameter search region was bound to $10^{-3} \leq C \leq 2 \cdot 10^2$ and $10^{-3} \leq \gamma \leq 2 \cdot 10^2$. Class imbalance was addressed by weighting the instances in the training data to match the total weight of the shockable and nonshockable classes. Finally, at most 0.5% of instances were discarded as outliers during training, this number is roughly equivalent to applying the 3- σ rule for a normal distribution [60].

Performance metrics and evaluation

The performance of the classification algorithms was evaluated using the test sets. Shock decision algorithms are binary classifiers, so results are reported in terms of sensitivity (Se) and specificity (Sp), positive and negative predictive values (PPV and NPV), and two measures of global accuracy, total accuracy (Acc) and balanced accuracy (BAC), which is defined as:

$$\text{BAC} = \frac{1}{2} (\text{Se} + \text{Sp}).$$

The AHA statement gives a similar importance to the detection of shockable (Se) and non-shockable (Sp) rhythms, by setting the minimum performance goals for Se and Sp at 90% and 95%, respectively. Consequently, we ranked our solutions by their BAC, which is simply the unweighted mean of sensitivities for a two-class problem. BAC is less sensitive to class imbalance than total accuracy [30]. Since our data is strongly imbalanced (roughly by a proportion of 4/1 in all datasets), we assigned weights as inverses of the class prevalences to the training instances on the optimization phase of all the classifiers. The McNemar test was used to test the null hypothesis that the accuracy of our model was greater than that of the other models at the 95% confidence interval, that is a $p < .05$ was considered significant [61].

The discriminating power of the features learned by our CNN+LSTM architecture was evaluated using a Receiver Operating Curve (ROC) analysis in the test set. In our architecture, those features are the 20 components of the output signal state at time $i = N_4$, $\mathbf{h}_{N_4} \in \mathbb{R}^Q$ of the LSTM block. We will denote those features by lstm_j for $j = 1, \dots, 20$. The CNN+LSTM model was trained using the training data, and the trained model was used to compute the LSTM

features in the training and test sets. The value of the LSTM features output by the model for the test set was used for the ROC curve analysis. The features were ranked using the Area Under the Curve (AUC) of the ROC analysis.

Results

Performance of the classification algorithms

Table 3 shows the performance metrics for 4-s segments in the test dataset, both for the public and OHCA databases. As shown in the table the mixed deep learning architecture proposed in this study presents the best performance, and the McNemar test showed that the accuracy of our model was significantly better than that of the rest of the classifiers. The differences between our solution and the best CNN and SVM solutions are only marginal for the public dataset, but very large for the OHCA database with differences in BAC and Acc of over 2.2-points and 1.4-points, respectively. A 1.4-point increase in Acc from a baseline of 96.1% (for SVM_{sel}) means that over 35% of the errors are now corrected.

For 4-s segments, the time needed to train our model (CNN+LSTM) in our high-end architecture was 1 hour and 3 min. The CNN only models were much quicker to train, with training times below 5 min. The time needed to classify a test sample (without parallelization) using the trained models was of 245 ms for our model, and below 14 ms for the CNN only models. All the deep learning models could be easily implemented and used in real time in low end hardware like the Field Programmable Gate Arrays (FPGA) and low end micro-processors used in AEDs [62].

Dependence with the length of the analysis segment

The BAC of the algorithm as the analysis segment is shortened is shown in Fig 3. The algorithm's accuracy is preserved for ECG analysis segments longer than 3-s, with a BAC above 99% in the public database and above 97% for the OHCA database. Although the accuracy is still good for 2—segments, the performance of the shock decision algorithms degrades. As shown in the figure, the proposed architecture outperforms the baseline SVM classifier ($p < .05$ for the McNemar test for all segment lengths) specially in the OHCA database. And it performs better than the best CNN only solution ($p < .05$) for longer segment lengths, and similarly for small segment lengths (see S1 Fig for the comparison with the rest of the CNN models).

Our results show that a deep learning architecture can accurately detect shockable rhythms with very short ECG segments, even on data from OHCA patients. Furthermore, the algorithm

Table 3. Performance metrics for 4-s segments. The the p-values are for the pairwise comparison of the accuracy of the proposed model to each of the other classifiers. The proposed network had a highest accuracy and the results were significant at the 95% confidence level (indicated by a[†]).

Method	Public datasets				OHCA datasets			
	Se/Sp	BAC/Acc	PPV/NPV	p-val	Se/Sp	BAC/Acc	PPV/NPV	p-val
Proposed network	99.7/98.9	99.3/99.1	96.3/99.9	-	99.2/96.7	98.0/97.5	93.6/99.6	-
Deep learners								
Kiranyaz et al [34]	99.3/98.4	98.9/98.6	94.6/99.8	< .05 [†]	97.7/93.8	95.8/95.1	88.4/98.8	< .05 [†]
Zubair et al [56]	95.6/96.2	95.9/96.1	87.5/98.7	< .05 [†]	93.9/89.4	91.7/90.9	81.0/96.8	< .05 [†]
Acharya et al [39]	98.4/98.1	98.2/98.2	93.6/99.5	< .05 [†]	94.7/91.6	93.2/92.6	84.5/97.3	< .05 [†]
Baseline SVM								
SVM _{all}	98.5/98.3	98.4/98.4	94.3/99.6	< .05 [†]	91.7/96.0	93.9/94.6	91.7/96.0	< .05 [†]
SVM _{sel}	99.2/98.7	99.0/98.8	95.5/99.8	< .05 [†]	93.9/97.1	95.5/96.1	93.9/97.1	< .05 [†]

<https://doi.org/10.1371/journal.pone.0216756.t003>

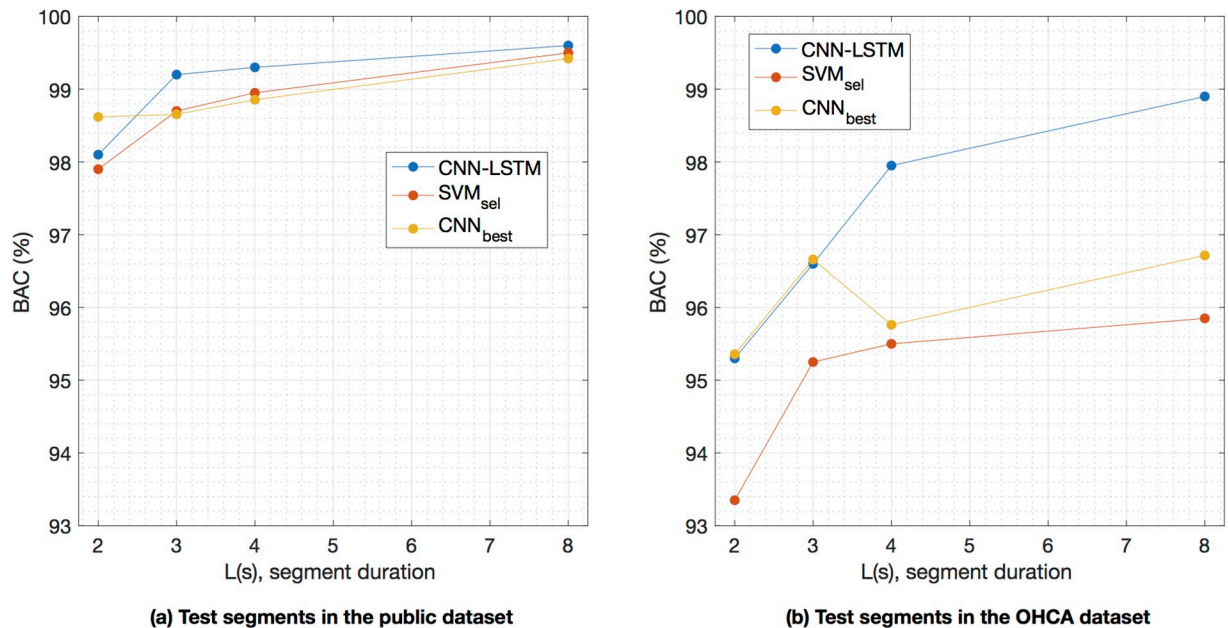


Fig 3. Performance for different segment lengths. Performance of the proposed network and the best baseline SVM and CNN only solution [34] for different segment lengths. The performance is worse for the OHCA database. The proposed solution outperforms the baseline SVM in both databases for all segment lengths ($p < .05$), and the best CNN solution ($p < .05$) except for short segment lengths.

<https://doi.org/10.1371/journal.pone.0216756.g003>

is AHA compliant for all segment lengths and datasets (except 2-s segments on OHCA data), as shown in Fig 4. The sensitivity and specificity were always above their respective minimum AHA goals of 90% and 95%.

Learned and designed features

One of the salient characteristics of deep-learning architectures is their ability to automatically learn the discrimination features. Table 4 shows the AUCs of the $lstm_j$ features on the test dataset for 4-s segments. Features are ordered using the AUC for the complete dataset (public +OHCA) because the features presented a mixed ranking when evaluated on the public and OHCA datasets separately. Six features had global AUCs above 95%, and are therefore very good individual shock/no-shock decision features. In most cases the AUCs for the public database were larger than for the OHCA database, which confirms that accurate shock/no-shock decisions are more difficult for OHCA rhythms. However the network was also able to learn several features that are better suited for OHCA data, such as $lstm_9$ or $lstm_1$.

We then applied dimensionality reduction techniques to visually assess and compare the classical VF-detection features to the features learned by our deep learning architecture. We choose the t-Distributed Stochastic Neighbor Embedding (t-SNE) dimensionality reduction technique [63, 64], which is a nonparametric nonlinear technique well suited for 2-D representations that has been shown to outperform many typical dimensionality reduction techniques in problems ranging from hand-digit recognition to genomic data clustering [63, 65]. In our configuration we reduced 30 classical features [27] and 20 LSTM features computed using 4—segments to a 2-D map suitable for visualization using scatterplots. The results for the public and OHCA datasets are shown in Fig 5, respectively. It is visually apparent that LSTM features produce a better and more separable clustering than the classical features. In fact, the Davies-Bouldin index to measure the separability of the clusters [66] was significantly lower for the

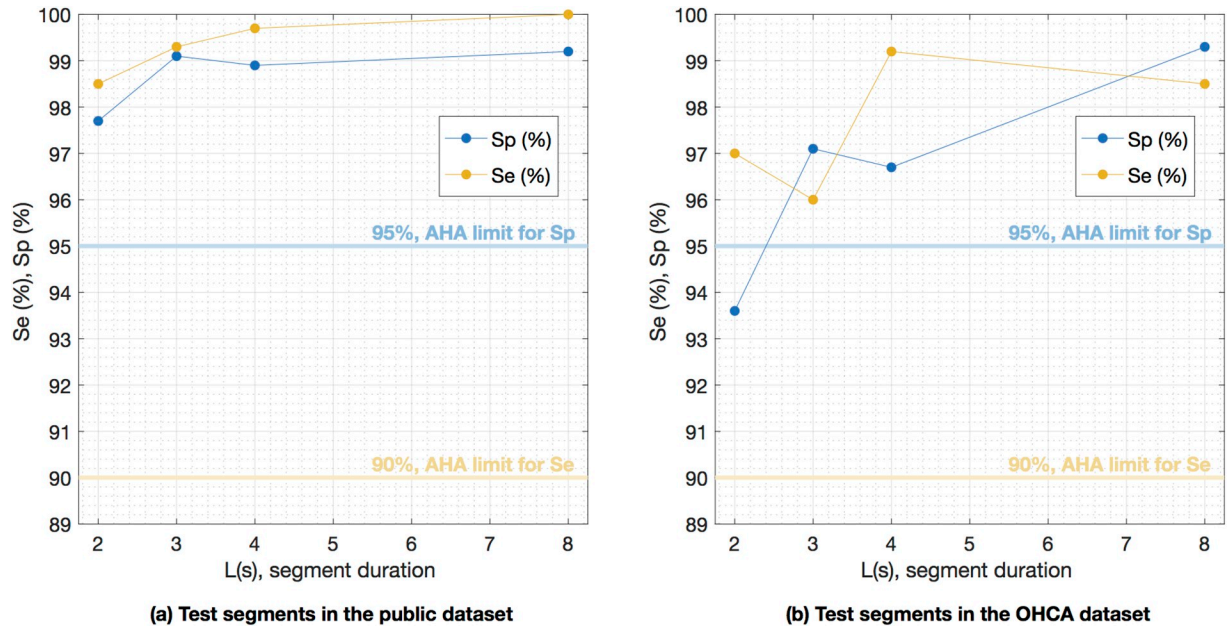


Fig 4. Sensitivity and specificity for different segment lengths. Sensitivity and specificity of the proposed deep learning network for different segment lengths for the public and OHCA databases. The algorithm meets AHA’s minimum 90% sensitivity for all segment lengths, and the 95% specificity for segment lengths longer than 2 s.

<https://doi.org/10.1371/journal.pone.0216756.g004>

learned features on 1000 bootstrap replicas of the experiment. The values mean (standard deviation) values were 1.073 (.006) for the LSTM features and 1.472 (.013) for the hand-crafted features ($p < .05$). These numbers confirm that the LSTM features produced a better clustering of the two classes than the classical VF-detection features.

Finally, we redesigned the classical SVM learner but fed it with the LSTM features instead, in line with a recent proposal that uses learned features from a 1-D CNN to detect VF [48]. The objective was to show if, for a classical machine learning approach based on SVM, the accuracy improved using the LSTM features instead of the classical VF-features. And if so, to determine the minimum number of LSTM features needed to improve those classification

Table 4. AUCs on the test set of the LSTM features of proposed architecture (see Fig 2). The model was trained in the training set for 4-s ECG segments, and this model was used to compute the LSTM features in the test set. The features are ordered by descending AUCs on the complete test set (public+OHCA).

Feature	Dataset			Feature	Dataset		
	Complete	Public	OHCA		Complete	Public	OHCA
lstm ₄	99.12	99.34	96.96	lstm ₅	91.19	91.10	92.57
lstm ₆	98.81	99.19	95.13	lstm ₃	85.91	85.44	90.24
lstm ₁₉	98.53	99.34	94.99	lstm ₈	85.84	85.93	88.43
lstm ₁₄	96.64	96.72	97.33	lstm ₁₀	83.15	84.35	82.93
lstm ₁₁	95.39	95.62	90.95	lstm ₁₇	80.19	81.40	72.90
lstm ₁₅	95.29	97.13	78.30	lstm ₉	79.80	78.39	88.32
lstm ₁₆	94.47	94.89	92.24	lstm ₁	78.91	76.87	90.97
lstm ₁₈	94.19	94.35	92.83	lstm ₁₂	75.47	75.61	83.91
lstm ₁₃	92.49	93.19	88.70	lstm ₂	66.97	67.84	56.51
lstm ₁₇	91.52	81.40	72.90	lstm ₂₀	56.80	57.00	50.87

<https://doi.org/10.1371/journal.pone.0216756.t004>

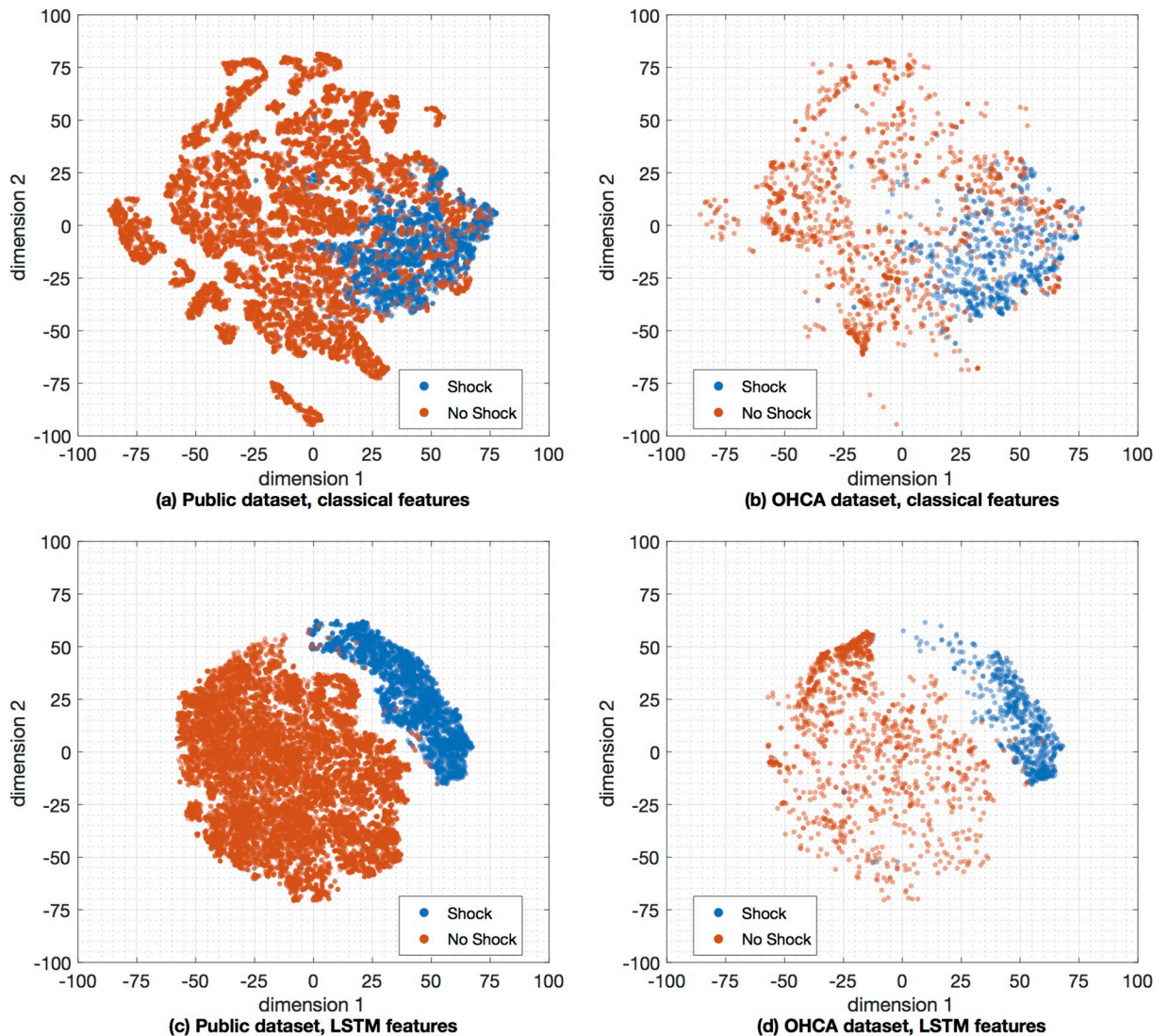


Fig 5. Visualization by t-SNE of the separability for the learned and classical VF-detection features. Two dimensional map visualization using t-SNE with the Barnes-Hut algorithm based on 30 classical VF-detection features [27] (top), and the 20 LSTM features (bottom). All experiments were done for the 4-s segments, and the visualization is separated for the public (left) and OHCA (right) datasets.

<https://doi.org/10.1371/journal.pone.0216756.g005>

results. All the experiments were done using the features obtained from the 4-s segments, and the SVM classifiers with LSTM features were optimized and evaluated using the same procedures used for the baseline SVM. First, on the training data a simple linear discriminant analysis classifier was used to select the best combination of LSTM features for all the possible feature sets of $K = 1, 2, \dots, 20$ features. Then, for each value of K the SVM classifier was optimized on the training data, and its performance was evaluated in the test set. The results are shown in Fig 6, compared to the best baseline SVM learner based on classical VF-features. The SVM based on LSTM features outperformed the SVM based on classical features for as few as $K = 4$ LSTM features, and using the LSTM features improved the accuracy particularly for the most challenging data, the OHCA dataset.

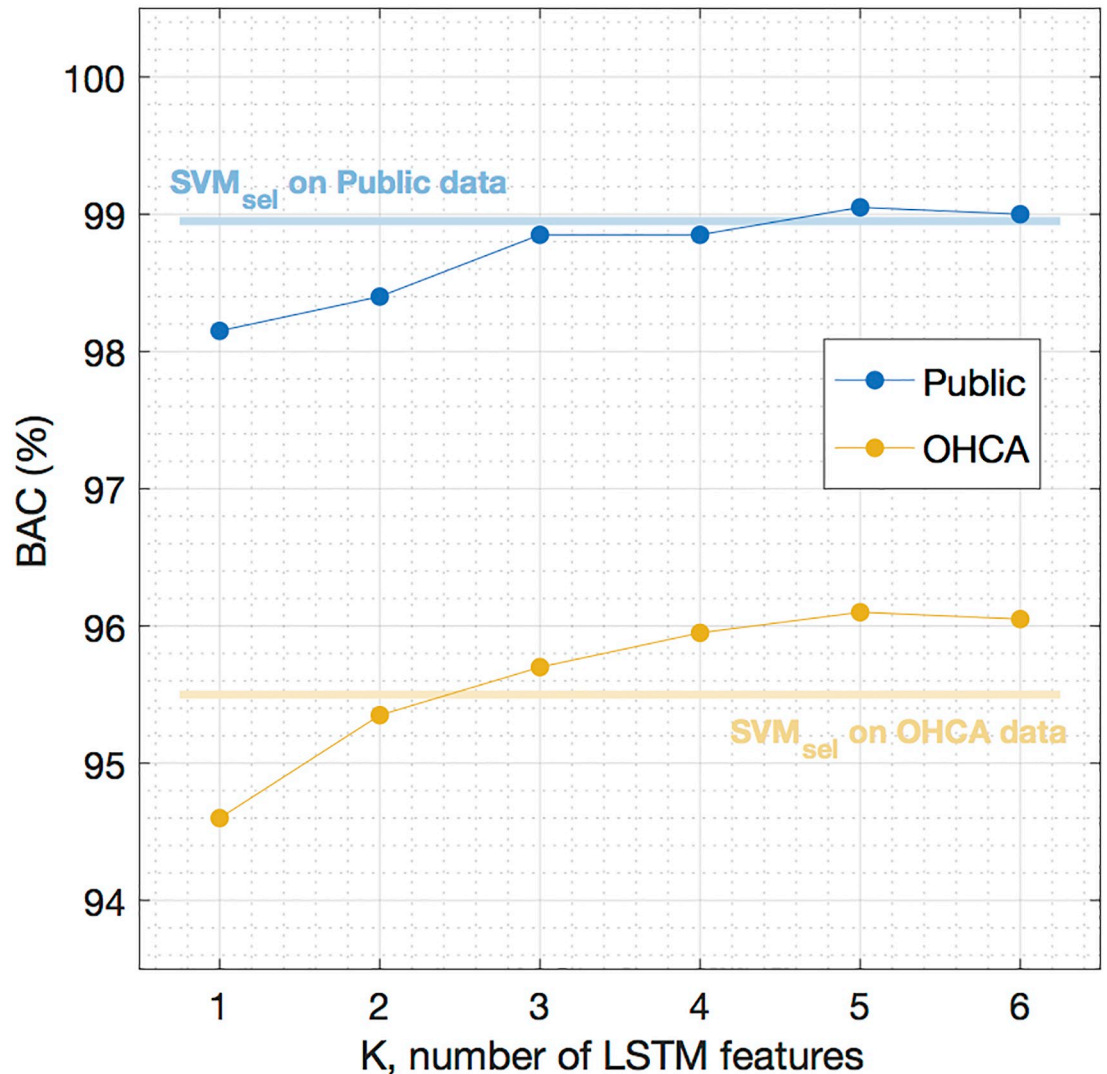


Fig 6. LSTM features used in an SVM classifier. Performance of an SVM classifier based on LSTM features, compared to the best baseline classifier using classical features, SVM_{sel} .

<https://doi.org/10.1371/journal.pone.0216756.g006>

Sources of classification errors

Fig 7 shows eight 4-s segments from the OHCA dataset incorrectly classified by our deep learning architecture. The examples shown in the figure illustrate the difficulty of classifying certain borderline OHCA rhythms. The four examples on the left (Fig 7(a)) show rhythms labeled as shockable that were misclassified as nonshockable by our algorithm. The first missed shockable rhythm presents several bursts of high amplitude activity, which result in large slopes in the ECG that are sometimes interpreted as QR segments, although no distinct QRS complexes are visible. In the second example the rhythm is ventricular and regular, but the ventricular rate is around 150 bpm, a threshold used frequently to separate intermediate VT from shockable fast VT. Finally the two bottom-most examples correspond to the late stages of VF in which the amplitude and fibrillation frequencies are lower, and the rhythm starts to degenerate into asystole.

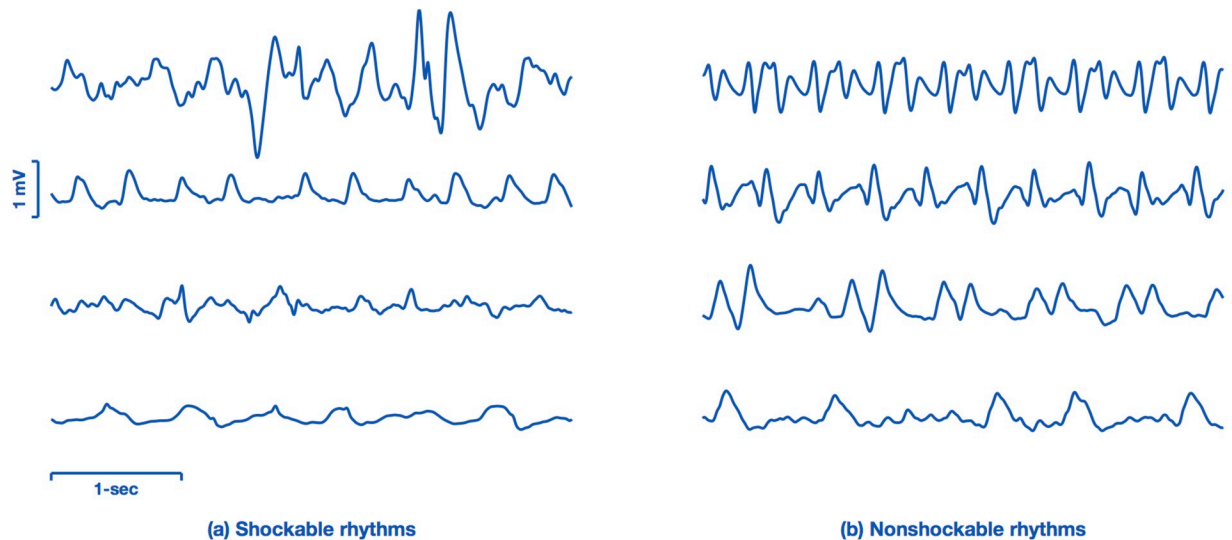


Fig 7. Examples of classification errors from the OHCA set. Examples of 4-s segments corresponding to borderline rhythms that were misclassified by the deep learning architecture. Segments corresponding to shockable rhythms are shown in the left, and segments corresponding to nonshockable rhythms in the right.

<https://doi.org/10.1371/journal.pone.0216756.g007>

The four examples on the right (Fig 7(b)) show rhythms labeled as nonshockable that were misclassified as shockable by our algorithm. The two topmost examples of missed nonshockable rhythms correspond to two supraventricular rhythms but with very wide QRS complexes and unstable heart rate, which result in a more irregular ECG than normally seen for nonshockable rhythms. The bottom most examples correspond to PEA cases with no supraventricular activity, but low rates. In the bottom most example the ventricular rate is very slow (around 60 bpm), but there is a high frequency quivering between ventricular contractions that produces a very irregular ECG.

Discussion

This is, to the best of our knowledge, the first study that uses OHCA data to develop and evaluate a deep learning architecture for the detection of shockable arrhythmia during cardiac arrest. In addition, we have for the first time analyzed the influence of the segment length on the performance of the deep learning classifier, and we have done an in-depth analysis of the features learned by our architecture when compared to the classical features used for the detection of VF over the past two decades. When compared to other deep learning approaches and classical machine learning approaches our architecture showed the best accuracy for the shock/no-shock decisions, specially of OHCA data. Our solution would therefore increase the security of AED algorithms by avoiding inappropriate shocks, and would ensure shocks are delivered when the patient is in VF. Furthermore, the algorithm fits within the computational and memory constraints of the low-end hardware of an AED. Therefore it could be implemented simply by changing the shock decision algorithm of the AED, i.e. a low cost software change.

Shock advice algorithms in current AED are safe, they are tested following the AHA framework [5] and IEC standards. Testing of those algorithms is done using AHA compliant vendor proprietary rhythm libraries, and the algorithms meet the AHA requirements for Se/Sp. In fact, many of those algorithms have been recently updated to be safely used in children using pediatric rhythm libraries [67, 68]. The characteristics of the rhythms in these libraries are

similar to the publicly available databases, and are frequently obtained from in-hospital electrophysiology studies [11, 68]. However, when the algorithms are used to diagnose rhythms in OHCA patients the Se/Sp values reported are lower. For instance, a recently described AED algorithm [11] presented Se/Sp of 99.5/98.0% when tested on proprietary rhythm databases, but when tested with OHCA data in the ventilation pauses during 30:2 CPR the Se/Sp decreased to 93.8/95.9%, respectively [69]. Similar results were found for another AED algorithm with OHCA rhythms and short ECG analysis segments of 3-5 s [70]. So further research on VF detection using OHCA data is needed, and the deep learning approach proposed in this study is a contribution in that direction.

Two recent papers have introduced deep learning architectures based on 1D-CNNs for the detection of shockable arrhythmia [39, 48]. Both studies are based on data from public ECG repositories, in particular the malignant ventricular arrhythmia were obtained from the CUIDB and VFDB databases. As shown by our results (see Table 3) the detection of shockable arrhythmia is easier on data from public repositories than in OHCA data, a result that had already been proved using classical machine learning algorithms [27]. OHCA data is recorded using defibrillators in a pre-hospital setting, and by the time the emergency services arrive the condition of the myocardium may have considerably deteriorated. This implies that the nonshockable rhythms recorded during OHCA most frequently correspond to pulseless patients, have irregular rates, and aberrant QRS complexes in patients with conduction problems [71]. VF is also recorded minutes after its onset, and by that time VF amplitude, fibrillation frequency, and waveform complexity have normally deteriorated [72, 73]. Furthermore, in OHCA there are frequent rhythm transitions from shockable to non-shockable rhythms as a result of the therapies applied by the emergency services [74], so borderline rhythms are more frequent. Therefore, although using ECGs from public databases is a standard practice and provides a good benchmark for the development and comparative assessment of VF detection methods [23, 24], it is desirable to use OHCA data to develop and test VF-detection methods. Our results show that the BAC were 1-3 points lower for the OHCA data than for the data from public ECG repositories, depending on the classifier (SVM or deep learner) and the segment length used for classification.

The design of our deep learning architecture combines 1D-CNN blocks with an LSTM network. The solution considerably improves the performance of some previous deep learning approaches [34, 56], of which some were specifically introduced to detect lethal ventricular arrhythmia [39]. Those architectures were based only on convolutional networks, and did not include the LSTM stage introduced in this study to integrate the temporal relations in the ECG. The benefits of adding an LSTM block are larger for longer segment lengths (see Fig 3) because then the LSTM network efficiently captures the temporal relations of the arrhythmia. An advantage of using an LSTM is that the network is able to learn the long term temporal relations of the ECG. LSTMs can capture different time scales and can capture subtle ECG nuances for arrhythmia classification [75, 76]. As shown in Table 3, the addition of an LSTM block improves the VF-detection accuracy of the deep learning architectures based only on CNN blocks. The increase in BAC over the best CNN only architecture is 0.5-points for public data, but over 2-points for the more challenging OHCA data. Acharya et al [39] reported an Se and Sp of 91.0% and 95.3% (BAC, 93.3%) for the public datasets in the first study that introduced CNN networks for the detection of lethal ventricular arrhythmia. Our implementation of their network yielded a much higher BAC of 98.2% on the public datasets, but we used a 4-s segment and cleaned the data of artifacts, intervals with loss of signal, asystole, and intermediate rhythms. These differences increase the BAC since classification is easier on longer segment lengths (see Fig 3), and noisy data is a confounding factor for the design of the classifiers. The recent contribution by Nguyen et al [48] used 8-s ECG segments and data cleaning, and

reported Se and Sp of 97.1% and 99.4% (BAC, 98.3%). Their results are therefore similar, when compared on the same conditions, to the values we obtained for the model proposed by Acharya et al. In any case, the addition of the LSTM block results in at least 1-point increase in BAC for the public datasets over those two architectures. It is important to emphasize that in practice an AED instructs the rescuer to stop CPR and avoid movements to analyze the rhythm, so the ECG observed by the device is free of artifacts. In the event of a low quality signal or loss of signal the device would not analyze the ECG. VF detection algorithms should therefore be tested under these conditions, as recommended by the AHA [5].

Most studies introducing or comparing algorithms for the detection of malignant ventricular arrhythmia use 4- or 8-s analysis segments [23, 24, 27–29, 31, 32, 48]. The duration of the interruptions in CPR therapy associated to an AED analysis ranges from 5.2 – 28.4 s depending on the AED model [77]. An important factor in this time is the ECG segment length needed by VF-detection algorithms to classify the rhythm as shockable or non-shockable. Consequently, shortening the analysis segment would shorten hands-off intervals (without CPR) during AED use, and shorter hands-off intervals increase the OHCA survival rates [78]. Moreover, reducing the analysis segment to around 3-s would enable the continuous monitoring of the rhythm during 30:2 CPR (30 compressions and 2 ventilations) because the rhythm could be analyzed during the ventilation pauses [69]. Our results for the OHCA data show that AHA requirements for Se and Sp are met for analysis segments of at least 3-s. When we tested a 2-s analysis segment the Sp on the OHCA data fell to 93.6%, below the 95% AHA target, and there was an overall decrease in BAC from 98.9% for 8-s segments to 95.2% for 2-s segments. In contrast, on the public data the BAC decreased only from 99.6% for 8-s segments, to 98.1% for 2-s segments, and the Se and Sp were above 97.5% for all the tested segment lengths (see Figs 3 and 4). These results suggest that capturing the characteristics of OHCA arrhythmia is difficult using short segment lengths, specially for nonshockable rhythms in which heart rates are frequently very slow and QRS complexes may be very wide due to aberrant ventricular conduction. It is therefore important to use OHCA data to explore the limits of accurate VF detection in cardiac arrest patients. In any case, our deep learning architecture was capable of an AHA compliant diagnosis for segment lengths as short as 3-s, shorter than the typical AED analysis duration which is above 5-s [77].

Finally, we bench-marked the results of our deep learning architecture against a state of the art machine learning algorithm using SVM and a large set of well known hand crafted VF-detection features. Our deep learning architecture outperformed the classical machine learning approach on both the public and the OHCA datasets and for all segment lengths, a trend observed in several other ECG classification problems [34, 37, 40, 41]. The increase in BAC was marginal for the public dataset, for instance for 4-s segments the BAC was only 0.3-points higher. However, for the OHCA there was a substantial increase in BAC which was 1.5 to 3-points larger depending on the segment length. Our deep learning architecture learned a set of 20 features output by the LSTM network. These features showed better separability, and performed better than the classical hand-crafted features when fed to the SVM classifier. Deep learning approaches can therefore be conceived and exploited to obtain improved ECG classification features, as proposed by Nguyen et al [48].

Deep learning solutions have proven effective and accurate for VF detection during OHCA. In the future several other clinical decision support algorithms during OHCA could benefit from using deep learning solutions. Some areas of current research include the prediction of shock success [79, 80], the analysis of the ECG during manual and mechanical chest compressions [81, 82], multiclass rhythm classification [30], or the detection of pulse [83]. In fact some deep learning solutions have been recently proposed for instance for the detection of pulse [84]. Moreover, given the difficulties in gathering annotated OHCA data it would be of great

interest to investigate the possibility of using these initial approaches in solutions with improved data efficiency, such as few shot learning solutions.

Conclusion

We have designed a mixed CNN and LSTM deep learning architecture for the detection of lethal ventricular arrhythmia. Our results show that this network outperforms the classical machine learning algorithms and the deep learning architectures based only CNN layers that had been proposed to date for this task. Furthermore, our results are tested on public repositories of ECG data but also on OHCA data which is harder to classify. Our architecture enabled an AHA compliant shock/no-shock decision for analysis segments as short as 3-s. Finally, the computational demands of our model fit in the low-end hardware of an AED, and implementing it on an AED would be low cost since it only involves a software change to the device.

Supporting information

S1 Fig. Performance of all deep learners for different ECG segment lengths. In the figure CNN₁ refers to Kiranaz et al [34], CNN₂ to Zubair et al [56], and CNN₃ to Acharya et al [39]. The networks CNN_{2,3} perform worse for longer segment length, they are more complex and need to adjust more weights (see Table 2) and perform better with more instances in the dataset. The advantages of using an LSTM block diminish for shorter segment lengths, as the LSTM is not able to capture the long temporal relations in the arrhythmia. (TIFF)

Author Contributions

Conceptualization: Artzai Picon, Unai Irusta, Aitor Álvarez-Gila, Elisabete Aramendi, Felipe Alonso-Atienza, Carlos Figuera, Unai Ayala, Estibaliz Garrote.

Data curation: Unai Irusta, Felipe Alonso-Atienza, Carlos Figuera, Unai Ayala, Lars Wik, Jo Kramer-Johansen, Trygve Eftestøl.

Formal analysis: Artzai Picon, Aitor Álvarez-Gila.

Investigation: Artzai Picon, Elisabete Aramendi, Trygve Eftestøl.

Methodology: Artzai Picon, Unai Irusta, Aitor Álvarez-Gila, Elisabete Aramendi, Estibaliz Garrote, Trygve Eftestøl.

Resources: Lars Wik, Jo Kramer-Johansen.

Software: Artzai Picon, Unai Irusta, Felipe Alonso-Atienza, Carlos Figuera, Unai Ayala.

Validation: Unai Irusta.

Writing – original draft: Artzai Picon, Unai Irusta.

Writing – review & editing: Artzai Picon, Aitor Álvarez-Gila, Elisabete Aramendi, Felipe Alonso-Atienza, Carlos Figuera, Unai Ayala, Estibaliz Garrote, Lars Wik, Jo Kramer-Johansen, Trygve Eftestøl.

References

1. Berdowski J, Berg RA, Tijssen JGP, Koster RW. Global incidences of out-of-hospital cardiac arrest and survival rates: Systematic review of 67 prospective studies. *Resuscitation*. 2010; 81(11):1479–87. <https://doi.org/10.1016/j.resuscitation.2010.08.006> PMID: 20828914

2. Luu M, Stevenson WG, Stevenson LW, Baron K, Walden J. Diverse mechanisms of unexpected cardiac arrest in advanced heart failure. *Circulation*. 1989; 80(6):1675–80. <https://doi.org/10.1161/01.CIR.80.6.1675> PMID: 2598430
3. Berdowski J, Blom MT, Bardai A, Tan HL, Tijssen JGP, Koster RW. Impact of onsite or dispatched automated external defibrillator use on survival after out-of-hospital cardiac arrest. *Circulation*. 2011; 124(20):2225–32. <https://doi.org/10.1161/CIRCULATIONAHA.110.015545> PMID: 22007075
4. Takata TS, Page RL, Joglar JA. Automated external defibrillators: technical considerations and clinical promise. *Ann Intern Med*. 2001; 135(11):990–8. <https://doi.org/10.7326/0003-4819-135-11-200112040-00011> PMID: 11730400
5. Kerber RE, Becker LB, Bourland JD, Cummins RO, Hallstrom AP, Michos MB, et al. Automatic External Defibrillators for Public Access Defibrillation: Recommendations for Specifying and Reporting Arrhythmia Analysis Algorithm Performance, Incorporating New Waveforms, and Enhancing Safety. *Circulation*. 1997; 95(6):1677–1682. <https://doi.org/10.1161/01.CIR.95.6.1677> PMID: 9118556
6. Jekova I, Krasteva V. Real time detection of ventricular fibrillation and tachycardia. *Physiological Measurement*. 2004; 25(5):1167. <https://doi.org/10.1088/0967-3334/25/5/007> PMID: 15535182
7. Anas E, Lee S, Hasan M. Sequential algorithm for life threatening cardiac pathologies detection based on mean signal strength and EMD functions. *BioMedical Engineering OnLine*. 2010; 9(1):43. <https://doi.org/10.1186/1475-925X-9-43> PMID: 20815909
8. Chen S, Thakor NV, Mower MM. Ventricular fibrillation detection by a regression test on the autocorrelation function. *Medical and Biological Engineering and Computing*. 1987; 25(3):241–9. <https://doi.org/10.1007/BF02447420> PMID: 3329694
9. Thakor NV, Zhu YS, Pan KY. Ventricular tachycardia and fibrillation detection by a sequential hypothesis testing algorithm. *IEEE Transactions on Biomedical Engineering*. 1990; 37(9):837–43. <https://doi.org/10.1109/10.58594> PMID: 2227970
10. Arafat M, Chowdhury A, Hasan M. A simple time domain algorithm for the detection of ventricular fibrillation in electrocardiogram. *Signal, Image and Video Processing*. 2011; 5:1–10. <https://doi.org/10.1007/s11760-009-0136-1>
11. Irusta U, Ruiz J, Aramendi E, de Gauna SR, Ayala U, Alonso E. A high-temporal resolution algorithm to discriminate shockable from nonshockable rhythms in adults and children. *Resuscitation*. 2012; 83(9):1090–1097. <https://doi.org/10.1016/j.resuscitation.2012.01.032> PMID: 22322285
12. Ayala U, Irusta U, Ruiz J, Eftestøl T, Kramer-Johansen J, Alonso-Atienza F, et al. A Reliable Method for Rhythm Analysis during Cardiopulmonary Resuscitation. *BioMed Research International*. 2014; 2014. <https://doi.org/10.1155/2014/872470>
13. Kuo S, Dillman R. Computer detection of ventricular fibrillation. In: *Computers in Cardiology*; 1978. p. 2747–2750.
14. Barro S, Ruiz R, Cabello D, Mira J. Algorithmic sequential decision making in the frequency domain for life threatening ventricular arrhythmias and imitative artifacts: a diagnostic system. *Journal of Biomedical Engineering*. 1989; 11(4):320–8. [https://doi.org/10.1016/0141-5425\(89\)90067-8](https://doi.org/10.1016/0141-5425(89)90067-8) PMID: 2755113
15. Minami K, Nakajima H, Toyoshima T. Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network. *IEEE Transactions on Biomedical Engineering*. 1999; 46(2):179–185. <https://doi.org/10.1109/10.740880> PMID: 9932339
16. Khadra L, Al-Fahoum A, Al-Nashash H. Detection of life-threatening cardiac arrhythmias using the wavelet transformation. *Medical and Biological Engineering and Computing*. 1997; 35(6):626–632. <https://doi.org/10.1007/BF02510970> PMID: 9538538
17. Li Y, Bisera J, Weil MH, Tang W. An Algorithm Used for Ventricular Fibrillation Detection Without Interrupting Chest Compression. *IEEE Transactions on Biomedical Engineering*. 2012; 59(1):78–86. <https://doi.org/10.1109/TBME.2011.2118755> PMID: 21342836
18. Mjihad A, Rosado-Muñoz A, Bataller-Mompeán M, Francés-Villora J, Guerrero-Martínez J. Ventricular Fibrillation and Tachycardia detection from surface ECG using time-frequency representation images as input dataset for machine learning. *Computer Methods and Programs in Biomedicine*. 2017; 141:119–127. <https://doi.org/10.1016/j.cmpb.2017.02.010> PMID: 28241963
19. Zhang XS, Zhu YS, Thakor NV, Wang ZZ. Detecting ventricular tachycardia and fibrillation by complexity measure. *IEEE Transactions on Biomedical Engineering*. 1999; 46(5):548–55. <https://doi.org/10.1109/10.759055> PMID: 10230133
20. Amann A, Tratnig R, Unterkofler K. Detecting ventricular fibrillation by time-delay methods. *IEEE Transactions on Biomedical Engineering*. 2007; 54(1):174–177. <https://doi.org/10.1109/TBME.2006.880909> PMID: 17260872

21. Chicote B, Irusta U, Alcaraz R, Rieta JJ, Aramendi E, Isasi I, et al. Application of Entropy-Based Features to Predict Defibrillation Outcome in Cardiac Arrest. *Entropy*. 2016; 18(9):313. <https://doi.org/10.3390/e18090313>
22. Clayton RH, Murray A, Campbell RW. Comparison of four techniques for recognition of ventricular fibrillation from the surface ECG. *Medical and Biological Engineering and Computing*. 1993; 31(2):111–7. <https://doi.org/10.1007/BF02446668> PMID: 8331990
23. Jekova I. Comparison of five algorithms for the detection of ventricular fibrillation from the surface ECG. *Physiological Measurement*. 2000; 21:429–39. <https://doi.org/10.1088/0967-3334/21/4/301> PMID: 11110242
24. Amann A, Tratnig R, Unterkofler K. Reliability of old and new ventricular fibrillation detection algorithms for automated external defibrillators. *Biomedical Engineering Online*. 2005; 4(60). <https://doi.org/10.1186/1475-925X-4-60> PMID: 16253134
25. Aramendi E, Irusta U, Pastor E, Bodegas A, Benito F. ECG spectral and morphological parameters reviewed and updated to detect adult and paediatric life-threatening arrhythmia. *Physiological Measurement*. 2010; 31(6):749. <https://doi.org/10.1088/0967-3334/31/6/002> PMID: 20410557
26. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*. 2000; 101(23):E215–20. <https://doi.org/10.1161/01.CIR.101.23.e215> PMID: 10851218
27. Figuera C, Irusta U, Morgado E, Aramendi E, Ayala U, Wik L, et al. Machine Learning Techniques for the Detection of Shockable Rhythms in Automated External Defibrillators. *PLoS One*. 2016; 11(7): e0159654. <https://doi.org/10.1371/journal.pone.0159654> PMID: 27441719
28. Alonso-Atienza F, Morgado E, Fernandez-Martinez L, Garcia-Alberola A, Rojo-Alvarez J. Detection of life-threatening arrhythmias using feature selection and support vector machines. *IEEE Transactions on Biomedical Engineering*. 2014; 61(3):832–840. <https://doi.org/10.1109/TBME.2013.2290800> PMID: 24239968
29. Li Q, Rajagopalan C, Clifford GD. Ventricular Fibrillation and Tachycardia Classification Using a Machine Learning Approach. *IEEE Transactions on Biomedical Engineering*. 2014; 61(6):1607–1613. <https://doi.org/10.1109/TBME.2013.2275000> PMID: 23899591
30. Rad AB, Eftestøl T, Engan K, Irusta U, Kvaløy JT, Kramer-Johansen J, et al. ECG-Based Classification of Resuscitation Cardiac Rhythms for Retrospective Data Analysis. *IEEE Transactions on Biomedical Engineering*. 2017; 64(10):2411–2418. <https://doi.org/10.1109/TBME.2017.2688380> PMID: 28371771
31. Nguyen MT, Nguyen BV, Kim K. Shockable Rhythm Diagnosis for Automated External Defibrillators Using a Modified Variational Mode Decomposition Technique. *IEEE Transactions on Industrial Informatics*. 2017; 13(6):3037–3046. <https://doi.org/10.1109/TII.2017.2740435>
32. Cheng P, Dong X. Life-Threatening Ventricular Arrhythmia Detection With Personalized Features. *IEEE Access*. 2017; 5:14195–14203. <https://doi.org/10.1109/ACCESS.2017.2723258>
33. Tripathy RK, Sharma LN, Dandapat S. Detection of Shockable Ventricular Arrhythmia using Variational Mode Decomposition. *Journal of medical systems*. 2016; 40:79. <https://doi.org/10.1007/s10916-016-0441-5>
34. Kiranyaz S, Ince T, Gabbouj M. Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks. *IEEE Trans Biomed Eng*. 2016; 63(3):664–75. <https://doi.org/10.1109/TBME.2015.2468589> PMID: 26285054
35. Acharya UR, Fujita H, Lih OS, Hagiwara Y, Tan JH, Adam M. Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network. *Information sciences*. 2017; 405:81–90. <https://doi.org/10.1016/j.ins.2017.04.012>
36. Acharya UR, Fujita H, Lih OS, Hagiwara Y, Tan JH, Adam M. Application of Deep Convolutional Neural Network for Automated Detection of Myocardial Infarction Using ECG Signals. *Information Sciences*. 2017;. <https://doi.org/10.1016/j.ins.2017.04.012>
37. Al Rahhal MM, Bazi Y, Al Zuair M, Othman E, BenJdira B. Convolutional Neural Networks for Electrocardiogram Classification. *Journal of Medical and Biological Engineering*. 2018; p.1–12.
38. Xia Y, Wulan N, Wang K, Zhang H. Detecting atrial fibrillation by deep convolutional neural networks. *Computers in biology and medicine*. 2018; 93:84–92. <https://doi.org/10.1016/j.combiomed.2017.12.007> PMID: 29291535
39. Acharya UR, Fujita H, Oh SL, Raghavendra U, Tan JH, Adam M, et al. Automated identification of shockable and non-shockable life-threatening ventricular arrhythmias using convolutional neural network. *Future Generation Computer Systems*. 2018; 79:952–959. <https://doi.org/10.1016/j.future.2017.08.039>
40. Rajpurkar P, Hannun AY, Haghpanahi M, Bourn C, Ng AY. Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks. *arXiv preprint arXiv:170701836*. 2017;.

41. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature medicine*. 2019; 25:65–69. <https://doi.org/10.1038/s41591-018-0268-3> PMID: 30617320
42. Ince T, Zabihi M, Kiranyaz S, Gabbouj M. Learned vs. Hand-Designed Features for ECG Beat Classification: A Comprehensive Study. In: *EMBECC & NBC 2017*. Springer; 2017. p. 551–554.
43. Graves A. Supervised Sequence Labelling with Recurrent Neural Networks. vol. 385 of *Studies in Computational Intelligence*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012.
44. Hochreiter S, Schmidhuber J. Long Short-Term Memory. *Neural Computation*. 1997; 9(8):1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735> PMID: 9377276
45. Gers FA, Schmidhuber J, Cummins F. Learning to forget: continual prediction with LSTM. *Neural computation*. 2000; 12:2451–2471. <https://doi.org/10.1162/089976600300015015> PMID: 11032042
46. Lipton ZC, Kale DC, Elkan C, Wetzell R. Learning to diagnose with LSTM recurrent neural networks. arXiv preprint arXiv:151103677. 2015;.
47. Chauhan S, Vig L. Anomaly detection in ECG time signals via deep long short-term memory networks. 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA); 2015. p. 1–7.
48. Nguyen MT, Nguyen BV, Kim K. Deep Feature Learning for Sudden Cardiac Arrest Detection in Automated External Defibrillators. *Scientific reports*. 2018; 8:17196. <https://doi.org/10.1038/s41598-018-33424-9> PMID: 30464177
49. Wik L, Kramer-Johansen J, Myklebust H, Sørebo H, Svensson L, Fellows B, et al. Quality of cardiopulmonary resuscitation during out-of-hospital cardiac arrest. *JAMA*. 2005; 293(3):299–304. <https://doi.org/10.1001/jama.293.3.299> PMID: 15657322
50. Kramer-Johansen J, Myklebust H, Wik L, Fellows B, Svensson L, Sørebo H, et al. Quality of out-of-hospital cardiopulmonary resuscitation with real time automated feedback: a prospective interventional study. *Resuscitation*. 2006; 71(3):283–292. <https://doi.org/10.1016/j.resuscitation.2006.05.011> PMID: 17070980
51. Li Y, Bisera J, Geheb F, Tang W, Weil MH. Identifying potentially shockable rhythms without interrupting cardiopulmonary resuscitation. *Critical Care Medicine*. 2008; 36(1):198–203. <https://doi.org/10.1097/01.CCM.0000295589.64729.6B> PMID: 18090359
52. Krasteva V, Jekova I, Dotsinsky I, Didon JP. Shock advisory system for heart rhythm analysis during cardiopulmonary resuscitation using a single ECG input of automated external defibrillators. *Annals of Biomedical Engineering*. 2010; 38(4):1326–1336. <https://doi.org/10.1007/s10439-009-9885-9> PMID: 20069371
53. Jekova I, Krasteva V, Ménétré S, Stoyanov T, Christov I, Fleischhackl R, et al. Bench study of the accuracy of a commercial AED arrhythmia analysis algorithm in the presence of electromagnetic interferences. *Physiological Measurement*. 2009; 30(7):695–705. <https://doi.org/10.1088/0967-3334/30/7/012> PMID: 19525573
54. Sutskever I, Martens J, Dahl GE, Hinton GE. On the importance of initialization and momentum in deep learning. *ICML*. 2013; 28(5):1139–1147.
55. Ripoll VJR, Wojdel A, Romero E, Ramos P, Brugada J. ECG assessment based on neural networks with pretraining. *Applied Soft Computing*. 2016; 49:399–406. <https://doi.org/10.1016/j.asoc.2016.08.013>
56. Zubair M, Kim J, Yoon C. An automated ECG beat classification system using convolutional neural networks. 2016 IEEE Conference on IT Convergence and Security (ICITCS); 2016. p. 1–5.
57. Chollet, François and others. Keras; 2015. <https://keras.io>.
58. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems; 2015. <https://www.tensorflow.org/>.
59. Scholkopf B, Sung KK, Burges CJC, Girosi F, Niyogi P, Poggio T, et al. Comparing support vector machines with Gaussian kernels to radial basis function classifiers. *IEEE Transactions on Signal Processing*. 1997; 45(11):2758–2765. <https://doi.org/10.1109/78.650102>
60. Maronna RA, Martin DR, Yohai VJ. *Robust Statistics: Theory and Methods*. John Wiley and Sons; 2006.
61. Fagerland MW, Lydersen S, Laake P. The McNemar test for binary matched-pairs data: mid-p and asymptotic are better than exact conditional. *BMC medical research methodology*. 2013; 13:91. <https://doi.org/10.1186/1471-2288-13-91> PMID: 23848987
62. Ukidave Y, Kaeli D, Gupta U, Keville K. Performance of the NVIDIA Jetson TK1 in HPC. 2015 IEEE International Conference on Cluster Computing; 2015. p. 533–534.
63. van der Maaten L, Hinton G. Visualizing data using t-SNE. *Journal of Machine Learning Research*. 2008; 9(2579-2605):85.

64. van der Maaten L. Accelerating t-SNE using Tree-Based Algorithms. *Journal of Machine Learning Research*. 2014; 15:3221–3245.
65. Taskesen E, Reinders MJT. 2D Representation of Transcriptomes by t-SNE Exposes Relatedness between Human Tissues. *PloS one*. 2016; 11:e0149853. <https://doi.org/10.1371/journal.pone.0149853> PMID: 26906061
66. Davies DL, Bouldin DW. A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1979; PAMI-1(2):224–227. <https://doi.org/10.1109/TPAMI.1979.4766909>
67. Atkins DL, Scott WA, Blaufox AD, Law IH, Dick M, Geheb F, et al. Sensitivity and specificity of an automated external defibrillator algorithm designed for pediatric patients. *Resuscitation*. 2008; 76:168–174. <https://doi.org/10.1016/j.resuscitation.2007.06.032> PMID: 17765384
68. Irusta U, Ruiz J. An algorithm to discriminate supraventricular from ventricular tachycardia in automated external defibrillators valid for adult and paediatric patients. *Resuscitation*. 2009; 80(11):1229–1233. <https://doi.org/10.1016/j.resuscitation.2009.07.013> PMID: 19716221
69. Ayala U, Irusta U, Ruiz J, de Gauna SR, González-Otero D, Alonso E, et al. Fully automatic rhythm analysis during chest compression pauses. *Resuscitation*. 2015; 89:25–30. <https://doi.org/10.1016/j.resuscitation.2014.11.022> PMID: 25619441
70. Didon JP, Krasteva V, Ménétré S, Stoyanov T, Jekova I. Shock advisory system with minimal delay triggering after end of chest compressions: accuracy and gained hands-off time. *Resuscitation*. 2011; 82 Suppl 2:S8–15. [https://doi.org/10.1016/S0300-9572\(11\)70145-9](https://doi.org/10.1016/S0300-9572(11)70145-9) PMID: 22208180
71. Mehta C, Brady W. Pulseless electrical activity in cardiac arrest: electrocardiographic presentations and management considerations based on the electrocardiogram. *American Journal of Emergency Medicine*. 2012; 30(1):236–239. <https://doi.org/10.1016/j.ajem.2010.08.017> PMID: 20970286
72. Brown CG, Dzwonczyk R, Martin DR. Physiologic measurement of the ventricular fibrillation ECG signal: estimating the duration of ventricular fibrillation. *Annals of Emergency Medicine*. 1993; 22(1):70–74. [https://doi.org/10.1016/S0196-0644\(05\)80253-7](https://doi.org/10.1016/S0196-0644(05)80253-7) PMID: 8424618
73. Sherman LD, Rea TD, Waters JD, Menegazzi JJ, Callaway CW. Logarithm of the absolute correlations of the ECG waveform estimates duration of ventricular fibrillation and predicts successful defibrillation. *Resuscitation*. 2008; 78(3):346–354. <https://doi.org/10.1016/j.resuscitation.2008.04.009> PMID: 18597916
74. Nordseth T, Bergum D, Edelson DP, Olasveengen TM, Eftestøl T, Wiseth R, et al. Clinical state transitions during advanced life support (ALS) in in-hospital cardiac arrest. *Resuscitation*. 2013; 84:1238–1244. <https://doi.org/10.1016/j.resuscitation.2013.04.010> PMID: 23603153
75. Oh SL, Ng EYK, Tan RS, Acharya UR. Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats. *Computers in biology and medicine*. 2018; 102:278–287. <https://doi.org/10.1016/j.combiomed.2018.06.002> PMID: 29903630
76. Yildirim O. A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. *Computers in biology and medicine*. 2018; 96:189–202. <https://doi.org/10.1016/j.combiomed.2018.03.016> PMID: 29614430
77. Snyder D, Morgan C. Wide variation in cardiopulmonary resuscitation interruption intervals among commercially available automated external defibrillators may affect survival despite high defibrillation efficacy. *Critical Care Medicine*. 2004; 32(9 Suppl):S421–S424. <https://doi.org/10.1097/01.CCM.0000134265.35871.2B> PMID: 15508671
78. Edelson DP, Abella BS, Kramer-Johansen J, Wik L, Myklebust H, Barry AM, et al. Effects of compression depth and pre-shock pauses predict defibrillation failure during cardiac arrest. *Resuscitation*. 2006; 71:137–145. <https://doi.org/10.1016/j.resuscitation.2006.04.008> PMID: 16982127
79. Chicote B, Irusta U, Aramendi E, Alcaraz R, Rieta J, Isasi I, et al. Fuzzy and sample entropies as predictors of patient survival using short ventricular fibrillation recordings during out of hospital cardiac arrest. *Entropy*. 2018; 20(8):591. <https://doi.org/10.3390/e20080591>
80. Shandilya S, Kurz MC, Ward KR, Najarian K. Integration of Attributes from Non-Linear Characterization of Cardiovascular Time-Series for Prediction of Defibrillation Outcomes. *PloS one*. 2016; 11(1): e0141313. <https://doi.org/10.1371/journal.pone.0141313> PMID: 26741805
81. Irusta U, Ruiz J, de Gauna SR, Eftestøl T, Kramer-Johansen J. A least mean-square filter for the estimation of the cardiopulmonary resuscitation artifact based on the frequency of the compressions. *IEEE Transactions on Biomedical Engineering*. 2009; 56(4):1052–1062. <https://doi.org/10.1109/TBME.2008.2010329> PMID: 19150778
82. Isasi I, Irusta U, Aramendi E, Ayala U, Alonso E, Kramer-Johansen J, et al. A Multistage Algorithm for ECG Rhythm Analysis During Piston-Driven Mechanical Chest Compressions. *IEEE Transactions on Biomedical Engineering*. 2019; 66(1):263–272. <https://doi.org/10.1109/TBME.2018.2827304> PMID: 29993407

83. Alonso E, Aramendi E, Daya M, Irusta U, Chicote B, Russell JK, et al. Circulation detection using the electrocardiogram and the thoracic impedance acquired by defibrillation pads. *Resuscitation*. 2016; 99:56–62. <https://doi.org/10.1016/j.resuscitation.2015.11.014> PMID: 26705970
84. Eloia A, Aramendi E, Irusta U, Picón A, Alonso E, Owens P, et al. Deep Neural Networks for ECG-Based Pulse Detection during Out-of-Hospital Cardiac Arrest. *Entropy*. 2019; 21(3):305. <https://doi.org/10.3390/e21030305>