

Received July 8, 2020, accepted August 9, 2020, date of publication August 24, 2020, date of current version September 3, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3018704

Shock Decision Algorithms for Automated External Defibrillators Based on Convolutional Networks

XABIER JAUREGUIBEITIA¹, GORKA ZUBIA¹, UNAI IRUSTA¹, (Member, IEEE), ELISABETE ARAMENDI¹, (Member, IEEE), BEATRIZ CHICOTE², DANIEL ALONSO³, ANDIMA LARREA⁴, AND CARLOS CORCUERA⁴

¹Department of Communications Engineering, University of the Basque Country (UPV/EHU), 48013 Bilbao, Spain

²Department of Smart Manufacturing, IK4-Lortek, 20240 Ordizia, Spain

³Emergency Medical System (Emergentziak-Osakidetza), Basque Health Service, Hospital Donostia, 20014 Donostia, Spain

⁴Emergency Medical System (Emergentziak-Osakidetza), Basque Health Service, 48010 Bilbao, Spain

Corresponding author: Unai Irusta (unai.irusta@ehu.eus)

This work was supported in part by the Spanish Ministerio de Ciencia, Innovación y Universidades, jointly with the Fondo Europeo de Desarrollo Regional (FEDER), under Project RTI2018-101475-BI00, and in part by the Basque Government under Grant IT1229-19 and Grant PRE-2019-1-0209.

ABSTRACT Automated External Defibrillators (AED) incorporate a shock decision algorithm that analyzes the patient's electrocardiogram (EKG), allowing lay persons to provide life saving defibrillation therapy to out-of-hospital cardiac arrest (OHCA) patients. The most accurate shock decision algorithms are based on deep learning, but these algorithms have not been trained and tested using OHCA data. In this study we propose novel deep learning architectures for shock decision algorithms based on convolutional and residual networks. EKG electronic recordings from a cohort of 852 OHCA cases (4216 AED EKG analyses) were used in the study. EKGs were annotated by a pool of six expert clinicians resulting in 3718 nonshockable and 498 shockable EKGs. Data were partitioned patient wise in a stratified way to train and test the models using 10-fold cross validation, and the procedure was repeated 100 times for statistical evaluation. Performance was assessed using sensitivity (shockable), specificity (non-shockable) and accuracy, and the analysis was conducted for EKG segments of decreasing duration. The best model had median (interdecile range) accuracies of 98.6 (98.5–98.7)%, 98.4 (98.2–98.6)%, 98.2 (97.9–98.4)%, and 97.6 (97.4–97.8)%, for 4, 3, 2 and 1 second EKG segments, respectively. The minimum 90% sensitivity and 95% specificity requirements established by the American Heart Association for shock decision algorithms were met, and the best model presented significantly greater accuracy ($p < 0.05$ McNemar test) than previous deep learning solutions for all segment durations. Moreover, the first AHA compliant shock decision algorithm using 1-s segments was demonstrated. This should contribute to a combined optimization of defibrillation and cardiopulmonary resuscitation therapy to improve OHCA survival.

INDEX TERMS Automated external defibrillator (AED), electrocardiogram (EKG), convolutional neural networks (CNN), deep learning, ventricular fibrillation (VF), residual networks.

I. INTRODUCTION

Cardiac arrest is the unexpected sudden cessation of the cardiac function, and occurs mostly in a pre-hospital setting. Out-of-hospital cardiac arrest (OHCA) constitutes a major global health problem. Only in the US one thousand OHCA events are estimated to occur daily, with survival rates around 10% [1]. Two therapies are key for OHCA survival: defibrillation, to restore the normal function of the heart; and

cardiopulmonary resuscitation (CPR), to generate an artificial blood flow and deliver oxygen to the vital organs when the heart cannot be defibrillated [2]. Electrical defibrillation can be provided by non-medical staff through automated external defibrillators (AEDs), which are equipped with a shock decision algorithm that automatically interprets the patient's electrocardiogram (EKG). These algorithms must have a high sensitivity (S_e) to detect shockable heart rhythms, i.e. malignant ventricular arrhythmia like ventricular fibrillation (VF) and tachycardia (VT). The specificity (S_p) must also be high to avoid inappropriate shocks that may deteriorate the rhythm

The associate editor coordinating the review of this manuscript and approving it for publication was Carmen C. Y. Poon.

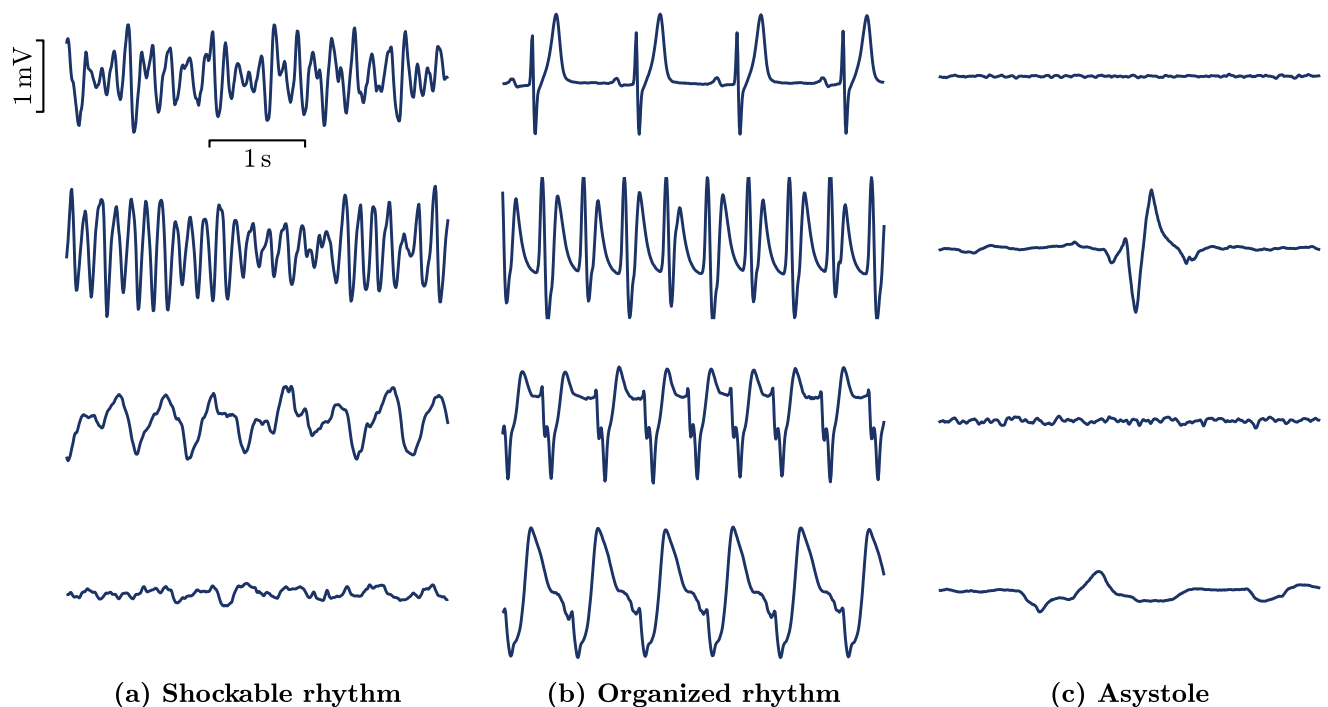


FIGURE 1. EKG samples from the OHCA study dataset. These samples illustrate the variability in waveforms and EKG characteristics of VF, ORG and AS rhythms. VF varies from coarse in amplitude and dominant frequency (top) to almost AS like waveforms (bottom). ORG rhythms can present well defined and narrow QRS complexes at normal heart-rates (top) but also aberrant QRS waveforms (bottom). Typically AS is characterized by a flat line (top), but it can present some activity in the form of isolated heartbeats, tremors and disorganized low amplitude activity that can be confused with VF.

and cause myocardial damage [3]. To guarantee a safe and efficient use of the device, the American Heart Association (AHA) establishes that the Se and Sp of these algorithms must be above 90% and 95%, respectively, when tested on OHCA data [3]. In addition, CPR must be interrupted to analyze the patient's EKG because chest compression artifacts in the EKG may confound the algorithm [4]. These time periods without chest compressions (no-flow intervals) for AED rhythm analysis can take from 5 s to 30 s [5], and have a negative impact on OHCA survival [6]. Therefore, there is a need to develop accurate shock decision algorithms capable of interpreting very short EKG segments to minimize interruptions in CPR [7], [8].

Research on shock decision algorithms has been framed traditionally as a VF detection problem [9]. Initial advances included analyses from EKG signal processing experts in the time, frequency and time-frequency domains [10]–[12], from which classification features were proposed and heuristic decision algorithms designed [13]. Later, machine learning algorithms like support vector machines or ensemble methods were introduced [14]–[16]. These approaches effectively combined the systematic and comprehensive extraction of EKG features [17], and the selection of the optimal feature subsets for VF detection [18]. Over the last years, deep learning has superseded traditional machine learning in many physiological signal analysis realms [19], [20]. This includes various EKG applications [21], ranging from heartbeat classification [22] to the detection of arrhythmia like atrial fibrillation [23], or even general cardiologist-level arrhythmia classification [24]. Recently, deep learning methods for

VF detection have been described, either using fully convolutional neural networks (CNN) [25], CNNs in combination with ensemble methods [26], or CNNs mixed with recurrent networks to identify the time dependencies in the data [27].

One of the caveats of deep learning solutions is the need for large annotated datasets to adjust the thousands or even millions of trainable network parameters. OHCA data with quality controlled rhythm annotations is scarce [17]. Consequently, most machine and deep learning solutions for VF detection have been demonstrated using Holter EKG data, recorded at the onset of the arrest, and available from public repositories like the MIT-BIH database [14], [15], [28]. However, the performance of shock decision algorithms degrades when trained/tested using OHCA data from defibrillators [18], [27], which are generally recorded minutes after the onset of the arrest. As the arrest progresses myocardial perfusion deteriorates [29], and so does the electrical activity of the heart. VF waveform characteristics like amplitude, dominant frequency and waveform complexity decrease over time [30]. Nonshockable rhythms with organized electrical activity (ORG), present lower heart-rates and more aberrant heartbeat waveforms (longer QRS complex durations) [31], [32]. Finally asystole (AS), the absence of electrical activity, becomes prevalent [31], [33], and the recommendation is not to shock AS (Sp > 95%) and to resume CPR immediately [2], [3]. Fig 1 shows some examples extracted from the study dataset that illustrate the variability in waveform morphology in OHCA rhythms.

There are few studies on shock decision algorithms using OHCA data gathered from defibrillators [7], [8], [18].

TABLE 1. Number of samples (prevalence in the dataset) of the analysis intervals and the extracted non-overlapping segments. Data is grouped into shockable and nonshockable rhythms, and it is also stratified into rhythm types within each class.

	Intervals	Segments			
		4 s	3 s	2 s	1 s
Shockable	498 (11.8%)	2187 (18.4%)	2982 (18.2%)	4606 (17.9%)	9461 (17.7%)
VF	461 (10.9%)	2056 (17.3%)	2801 (17.1%)	4328 (16.8%)	8883 (16.6%)
VT	37 (0.9%)	131 (1.1%)	181 (1.1%)	278 (1.1%)	578 (1.1%)
Nonshockable	3718 (88.2%)	9670 (81.6%)	13429 (81.8%)	21121 (82.1%)	44064 (82.3%)
ORG	1300 (30.8%)	4130 (34.8%)	5715 (34.8%)	8895 (34.6%)	18434 (34.4%)
AS	2418 (57.4%)	5540 (46.7%)	7714 (47.0%)	12226 (47.5%)	25630 (47.9%)

Among the VF detection algorithms based on deep learning, only one study included OHCA data [27], and all the studies excluded AS from their datasets [25]–[27]. But as shown in Fig. 1 the EKG may present electrical activity during AS that can be confused with a shockable rhythm, leading to an electrical shock that would worsen the prognosis of the patient. Moreover, AHA compliant algorithms using very short EKG segments (less than 3 seconds) have not been demonstrated, and they could be of importance to optimize defibrillation/CPR therapies. This study covers those knowledge gaps. First, it is supported on a large dataset of OHCA rhythms that includes AS, which was obtained using AEDs during treatment and annotated by a pool of expert clinicians. Second, the dataset was used to develop and test optimized convolutional network architectures for shock/no-shock classification capable of analyzing the rhythm using EKG segments as short as 1-s.

II. MATERIALS

The study dataset was collected between June 2013 and December 2015 by the Emergency Medical Services (EMS) of the Basque Country. It comprises OHCA cases treated by emergency medical technicians working in Basic Life Support (BLS) ambulances [34]. The Basque EMS serves a 2.2 million population with an estimated annual incidence of 39 EMS treated cases per 100,000 inhabitants [35]. Around 60% of the cases treated by the BLS ambulances in the study period were included, totaling 852 defibrillator files obtained from Lifepak 1000 AEDs (Stryker, Kalamazoo MI, US). The electronic data comprised the defibrillator messages, with treatment information like AED analysis intervals, and the signals recorded from the defibrillator pads: the EKG to analyze the heart rhythm, and the thorax impedance to monitor CPR activity. The Lifepak 1000 AED records the EKG with a sampling frequency of $F_s = 125$ Hz, a resolution of $4.8 \mu\text{V}$ per least significant bit, and a typical AED bandwidth of $0.5 - 21$ Hz [36]. This narrow EKG bandwidth, typical of AEDs, ensures very low levels of the main EKG noise sources such as baseline wander or power line interference.

Defibrillator data were converted from a proprietary file format to an open Matlab (Mathworks, Natick MA, US) format, and a custom tool was prepared to annotate the rhythm in the analysis periods of the AED. Six clinicians specialized in OHCA treatment blindly revised the EKG and annotated

the rhythm. The clinicians adhered to the following rhythm definitions [37], [38]: VF (disorganized ventricular rhythms with amplitudes above $200 \mu\text{V}$) and VT (regular ventricular rhythms with rates above 150 min^{-1}) in the shockable class; and ORG (rhythms with visible QRS complexes and heart rates above 12 min^{-1}), and AS (rhythms with peak-to-peak amplitudes below $100 \mu\text{V}$ or heart rates below 12 min^{-1}) in the nonshockable class. The majority vote from the pool of six clinicians was adopted as the final ground truth rhythm annotation. The final composition of the annotated dataset is shown in Table 1. A total of 4216 AED analysis intervals were reviewed, from which 498 were shockable and 3718 nonshockable. The inter-rater agreement measured using the Fleiss kappa score (κ) [39] was excellent among the six clinicians, with $\kappa = 0.895$. Moreover, the sub-pool of three clinicians with highest agreement had a $\kappa = 0.955$. The quality control of the annotations guaranteed robust ground truth shock/no-shock labels for the development and evaluation of the deep learning classification models.

The AED analysis intervals were extended to cover the full EKG interval without artifacts, that is from cessation of CPR to resumption of CPR or defibrillation. If rhythm transitions occurred the interval was extended up to those transitions, to ensure a unique rhythm per interval. The intervals included time delays to push the rhythm analysis button, pre-shock charging, and delays in resumption of CPR which typically occur during OHCA treatment. The median (interquartile range, IQR) duration of the extended intervals was 11.8 (9.1 – 16.6) s for all rhythms, 20.1 (17.6 – 22.5) s for shockable and 11.2 (8.9 – 14.6) s for nonshockable rhythms. These intervals were then divided into non-overlapping segments of 512 (4.096 s), 384 (3.072 s), 256 (2.048 s) and 128 (1.024 s) samples. Fig. 2 shows how an AED analysis interval was first extended and then divided into segments of different duration. In what follows these segments will be denoted by their approximate durations: 4 s, 3 s, 2 s and 1 s. Table 1 lists the number and proportions of segments grouped into shock/no-shock categories and further into rhythm types.

III. METHODS

A. CONVOLUTIONAL NETWORKS

For classification each segment can be represented as (x_i, y_i) , where $x_i \in \mathbb{R}^{1 \times D}$ is the signal segment of D samples, and

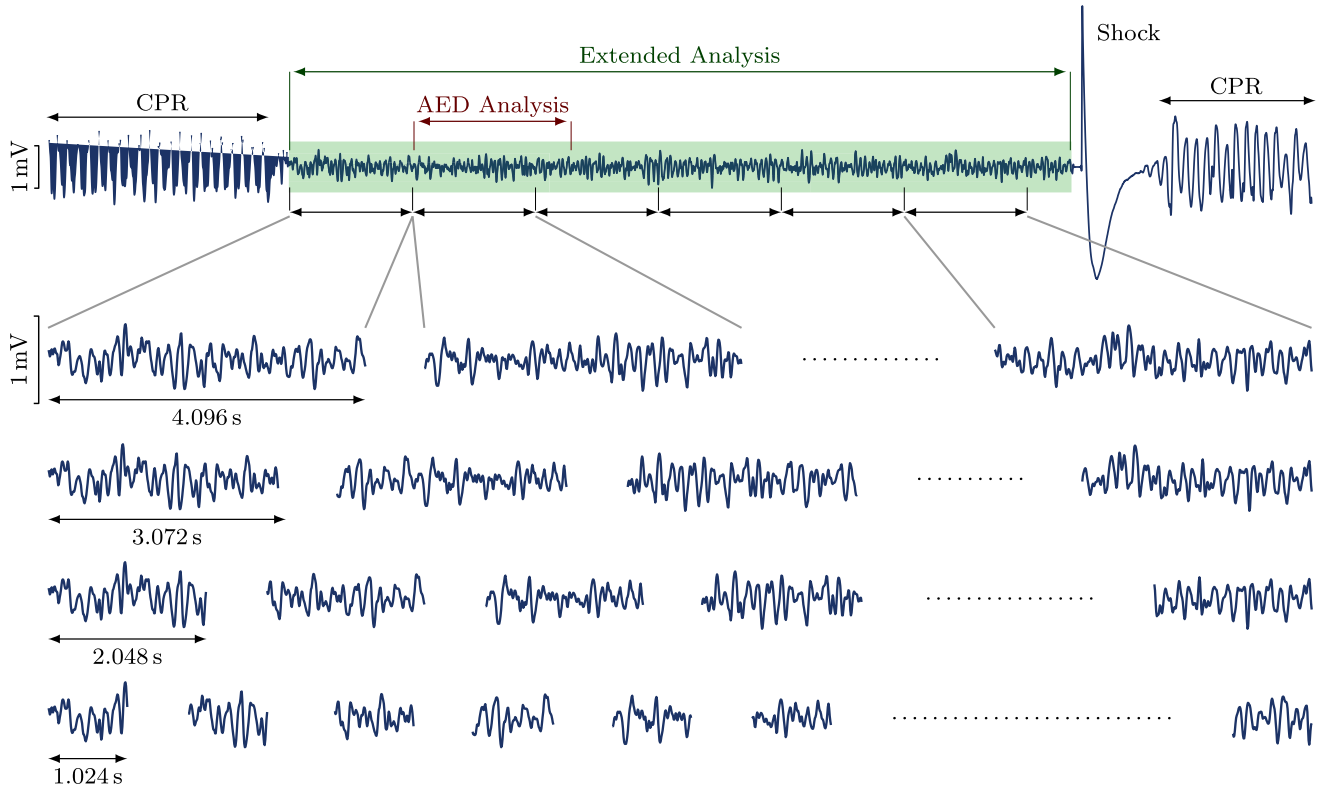


FIGURE 2. Example of the extraction of the non-overlapping EKG segments from the extended analysis interval. The figure shows a shock delivered to a patient in VF. First CPR is stopped to analyze the rhythm (start of extended analysis interval), then after approximately 4 s the rescuer pushes the AED analysis button and VF is identified, then the AED is charged and finally the shock is delivered (end of extended analysis segment). After the shock CPR is resumed.

$y_i \in \{0, 1\}$ is the class label inherited from the analysis interval ($0 \equiv$ no-shock, and $1 \equiv$ shock). The data were partitioned into train and test (see section III-B) to optimize the convolutional models, and to obtain the performance metrics, respectively. Two architectures were studied (see Fig. 3), a CNN network and a residual network (ResNet).

1) CNN ARCHITECTURE

The CNN architecture consists of four blocks, each comprised of a convolutional layer, a batch normalization layer (BN), a max-pooling layer, and a rectified linear unit (ReLU) non-linear activation layer. The convolutional layers linearly transform an input $X \in \mathbb{R}^{M \times D}$, consisting of M signal channels of D samples to produce an output $Y \in \mathbb{R}^{N \times D'}$. The input is convolved with a set of N filters $W^n \in \mathbb{R}^{M \times L}$, where L is the filter size, and then shifted channel-wise by a bias b^n . The individual elements $y_{n,d}$ of Y can be obtained as

$$y_{n,d} = b^n + \sum_{m=1}^M \sum_{l=1}^L w_{m,l}^n x_{m,d+l-1} \quad (1)$$

where the filter weights $w_{m,l}^n$ and biases b^n are learnable parameters, adjusted during training. In our network, all convolutional layers had a filter size L of 16 and an increasing number of filters $N = \{8, 16, 32, 64\}$. Input channels were zero-padded symmetrically so $D' = D$, and bias terms were removed. The BN layers adjust the output of the preceding

layer so that the data distribution forwarded to the next block does not depend on complex cross-layer weight interactions. This lessens the need for fine parameter tuning and weight initialization, and speeds up training by allowing the use of larger learning rates [40]. It also improves generalization, reducing overfitting. For an input batch \mathcal{I} with data samples $i \in \{1, \dots, I\}$, a BN layer computes the channel-wise means $\mu_{\mathcal{I},n}$ and variances $\sigma_{\mathcal{I},n}^2$ and normalizes each channel by:

$$\hat{y}_{n,d}^i = \frac{y_{n,d}^i - \mu_{\mathcal{I},n}}{\sqrt{\sigma_{\mathcal{I},n}^2 + \epsilon}} \quad (2)$$

with ϵ a small value added for numerical stability. The normalized channels are then scaled and shifted to make the most use of the following non-linear activation. The outputs $z_{n,d}^b$ are thus given as

$$z_{n,d}^b = \gamma_n \hat{y}_{n,d}^b + \beta_n \quad (3)$$

where γ_n and β_n are trainable parameters. After training, channel-wise means and variances were computed for the full training set, and then used during test to normalize the data. Max-pooling layers downsample input data by selecting the largest element in blocks of K elements along the time-dimension d , such that:

$$p_{n,d'} = \max\{z_{n,d}\}_{d=(d'-1)K+1, \dots, d'K} \quad (4)$$

where $K = 2$ for all blocks in our architecture. This reduces the computational burden of the network by increasing the

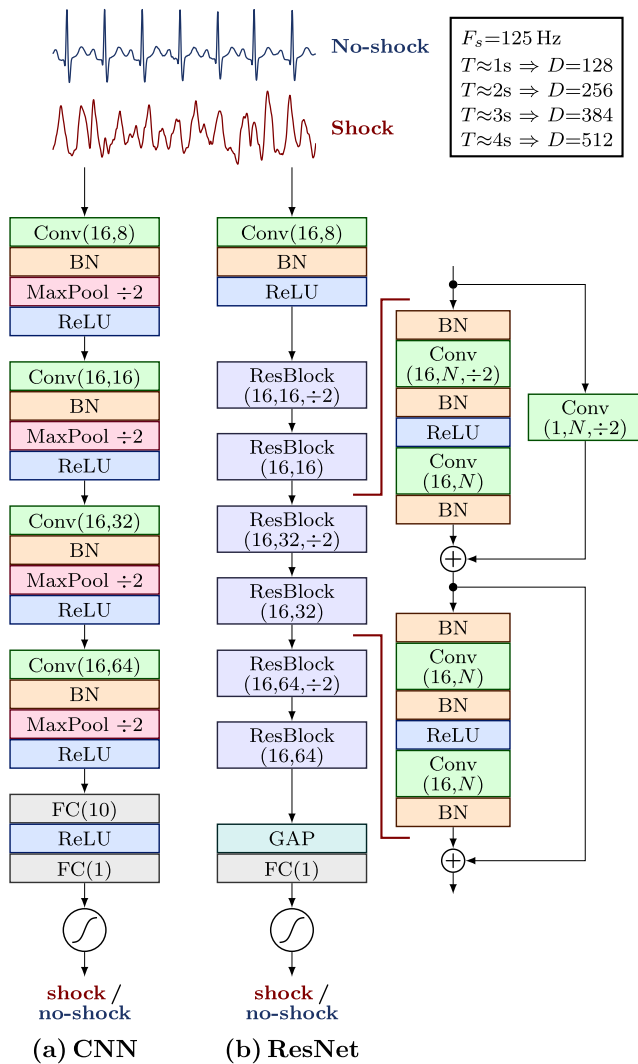


FIGURE 3. Architecture of the convolutional networks designed for the shock decision algorithms. The left-most architecture is a fully CNN architecture, and the right most one is the ResNet architecture (with its expanded residual block to the right).

field of view of the filters without increasing their size. Finally, the ReLU layers introduce nonlinearity in the network through the activation function $f(x) = \max\{0, x\}$, which allows learning complex nonlinear mappings.

The output of the last convolutional block ($N = 64$ filters) was flattened to produce a feature vector $f_i(x_i)$ that mapped the input signal segment $x_i \in \mathbb{R}^{1 \times D}$ to a feature embedding $f_i \in \mathbb{R}^{1 \times J}$, where $J = (D/2^4) \cdot N$. The feature vector was input to a dense network composed of two fully connected (FC) layers with 10 neurons (ReLU activation) and 1 neuron (sigmoid activation for classification), respectively, to produce the binary shock/no-shock decision.

2) ResNet ARCHITECTURE

The second architecture we studied was a residual network or ResNet [41], which addresses the problem of performance degradation as layers are stacked and the depth of the network increases. Given $\mathcal{H}(x)$, the transformation of a series of

stacked layers to the input data x , residual network design argues that it is easier to map the residual transformation $\mathcal{F}(x) := \mathcal{H}(x) - x$. This is achieved in practice by enabling a secondary shortcut path which directly connects the input x to the main path's output. An element-wise addition of both paths is then performed for an effective mapping of $\mathcal{F}(x) + x = \mathcal{H}(x)$.

Our ResNet architecture (See Fig. 3b) was designed to mimic that of the CNN, increasing the network's depth while maintaining a recognizable structure. Each convolutional block was replaced by two residual blocks, preserving spatial length D and channel depths N (see Fig. 3b). For each residual block the main path consisted of a sequence of convolutional, BN and ReLU layers, following the improved pre-activation configuration described in Han *et al.* [42]. Pooling layers were replaced by strided convolutions, which skip every other step in the filtering process. When length and depth had to be adjusted, the shortcut path included a strided convolution to produce a linear projection of the input. Finally, the hidden fully connected layer was replaced by a global average pooling (GAP) layer which outputs the mean value of each input channel. This was meant to reduce overfitting and improve robustness to spatial translations of data [43].

3) CONFIGURATION OF THE NETWORK OPTIMIZER

The training process was optimized using stochastic gradient descent (SGD) with momentum ($m = 0.9$), and an initial learning rate of 0.05. The mini-batch size was set to 256 for segments of 4 s. For the other segment durations, the batch size was adjusted to produce a similar amount of training iterations, that is, 341, 512 and 1024 for 3 s, 2 s and 1 s respectively. As few as 10 epochs were sufficient for the networks to converge. A piecewise learning rate decay factor of 0.5 per epoch was applied, allowing the solver to transition from rougher to finer optimization steps over a short training process. All calculations were performed using the Matlab's deep learning toolbox, on a multi-GPU setting over 4 Nvidia GeForce RTX 2080 Ti GPUs.

B. DATA PARTITIONING AND MODEL EVALUATION

Both architectures were evaluated using a 10-fold cross-validation (CV) strategy. Data partitioning was conducted patient-wise and in a quasi-stratified way, so that the analysis intervals contained in each fold would present a rhythm distribution close to that of the whole dataset. Patient-wise data partitioning ensured that the training and test patients did not overlap, and thus may have different chest configurations, physiological characteristics and/or defibrillator pad placements. Since the results could depend on how data were partitioned, 100 different random¹ CV partitions were used. This allowed the statistical characterization of the performance of the models.

Data augmentation was used to increase the available number of shockable segments. This addressed two problems.

¹The Australian National University's Quantum Random Numbers Server was used: <https://qrng.anu.edu.au>

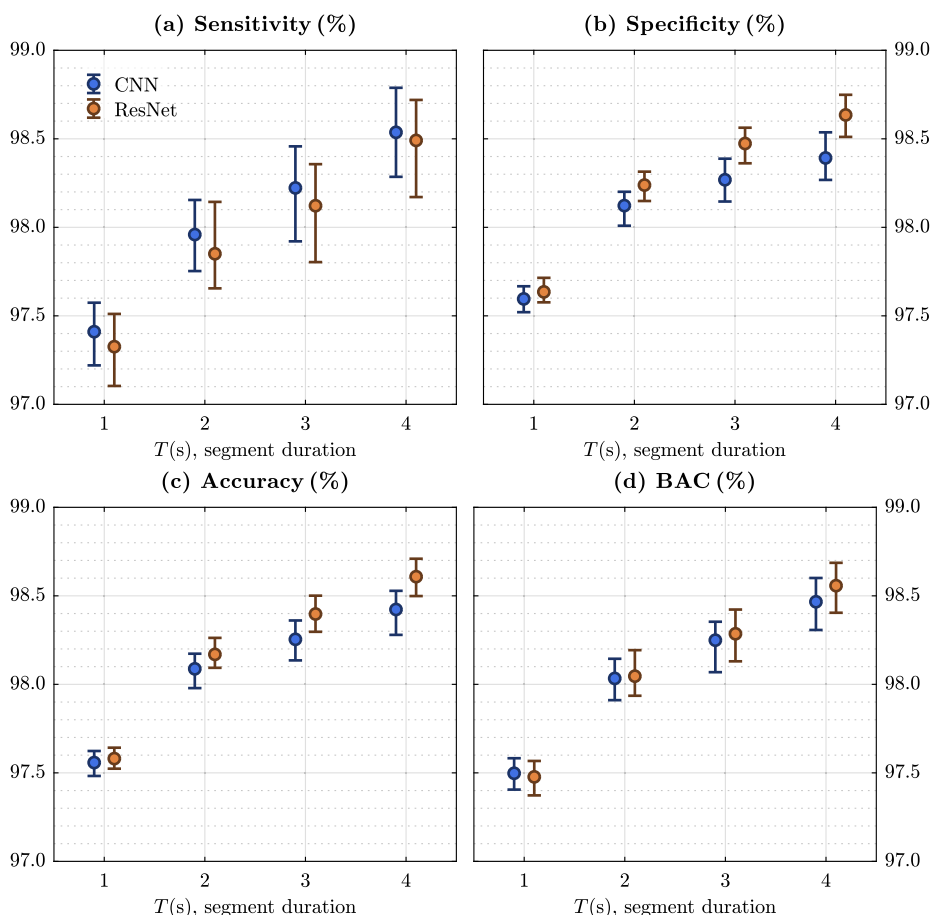


FIGURE 4. Median (interdecile range, IDR) performance scores for the CNN and ResNet models as a function of the segment duration.

First, the large class imbalance of the dataset, with a nonshockable to shockable class proportion in excess of 4:1 (see Table 1). Second, the low number of VT samples. An almost four-fold increase in the number of available segments was achieved by extracting segments with a 75% overlap. Augmented data were used only during training; all testing was carried out using the original non-overlapping segments.

A shock decision algorithm is a binary classification problem, in which shockable rhythms are the positive class and nonshockable rhythms the negative class. The standard performance metrics are Se, Sp and accuracy (Acc). However, in OHCA nonshockable rhythm prevalences are much higher than those of shockable rhythms. Since the AHA requires high Se/Sp values, the balanced accuracy (BAC) was also computed [17]:

$$BAC = \frac{1}{2}(Se + Sp) \tag{5}$$

IV. RESULTS

A. PERFORMANCE OF THE MODELS

Fig. 4 shows the performance metrics for each model as a function of the segment duration. Performance dropped as the duration of the segment shortened. Still, AHA performance goals were met for all segment durations, with median Acc and BAC scores above 97.4% for segments as short as 1 s.

CNN models achieved slightly higher median Se, although the differences were not significant. The ResNet models had higher Sp and therefore higher Acc, given the much larger prevalence of nonshockable rhythms in the dataset. When compared using the McNemar test, the ResNet models were significantly more accurate than the CNN models ($p < 0.05$) only for 4 s and 3 s.

Table 2 shows the detailed shock/no-shock classification results for the different rhythm types. The worst performance was obtained for VT, which was by far the least prevalent rhythm. However, all models met the AHA’s 75% sensitivity performance goal for VT. The differences between the CNN and ResNet models were small. ResNet models outperformed CNN models in ORG rhythms for longer segments, and the converse occurred for VT. For shorter segments both models performed similarly in all rhythm types.

B. DATA CLUSTERIZATION AND ERROR SOURCES

A dimensionality reduction technique was applied to visualize how each network separated the data, and to identify potential sources of error. Models were trained and tested using the full dataset. Then, the activations of the last convolutional block were retained and projected to a 2-D map using the t-distributed stochastic neighbor embedding algorithm (t-SNE), a non-linear dimensionality reduction method

well suited for high-dimensional data representation [44]. Fig. 5 shows the resulting t-SNE scatter plots for both networks and 4 s segments.

VF samples formed a well-separated cluster, similar for both architectures, ending in a boundary connecting to AS (borderline AS/VF rhythms). The boundary between both regions could be intuitively associated to the amplitude thresholds commonly accepted for the definition of coarse VF ($>200 \mu\text{V}$) and AS ($<100 \mu\text{V}$) [37], [38]. Moreover, this region also showed an overlap between VF and ORG rhythms, associated to ventricular arrhythmia with slower ventricular rates and ORG rhythms with aberrant QRS complexes (borderline ideoventricular rhythms).

It is noteworthy that although the networks were not trained to differentiate AS and ORG, the 2-D maps show a good clusterization of AS and ORG rhythms. The borderline cases could be intuitively associated to the heart-rate thresholds customarily used in the definitions of AS ($<12 \text{ min}^{-1}$) and ORG ($>12 \text{ min}^{-1}$) [17], [45]. Finally VT samples were spread along the VF cluster, and were therefore not well separated.

The largest concentration of shock/no-shock classification errors occurred in the interface of the AS/ORG/VF clusters shown in Fig. 5. In order to better appreciate the reasons for these errors the different rhythm classes were parametrized using standard measures applicable to each rhythm type [8], [32], [37], [46]. For AS we computed the mean power of the signal [8], [37], for ORG rhythms the mean heart-rate [32] and for VF the amplitude spectrum area (AMSA), which is a weighted sum of spectral amplitudes correlated to myocardial perfusion [46]. Fig. 6 shows the t-SNE plot of the CNN model graded by the typical parameter of each rhythm type. As shown in the figure, VF and AS mix when the power in AS is moderate-to-high ($> 10^{-3} \text{ mW}$), or AMSA in VF is low, indicating those VF rhythms have lower fibrillation frequencies and lower amplitudes. ORG rhythms mix with VF when heart-rates are moderate and amplitudes are low (AS/VF/ORG boundary). These samples correspond, in most cases, to ideoventricular rhythms, that

is slow ventricular rhythms with wide QRS complexes. This is better appreciated in the examples of classification errors shown in Fig. 7. The figure also shows why shock decisions are specially challenging for very short EKG: short transient low amplitude intervals during VF or short periods of fast disorganized activity during nonshockable rhythms may result in an incorrect decisions (see lower panel in Fig. 7).

C. COMPARISONS WITH PREVIOUS MODELS

The convolutional models developed in this study were compared with the published deep learning algorithms for shock decision [22], [25]–[27]. These architectures include a CNN combined with a recurrent network [27], and three designs based on CNNs: Acharya *et al.* [25] for shock decision, Kiranyaz *et al.* [22] for heartbeat classification (better suited for short segments), and Nguyen *et al.* [26] with a multi-channel input using the EKG and two components obtained from the variational mode decomposition [47] of the EKG. For the analyses, all the models were trained and tested as described in section III-B, and using the optimizer described in section III-A3. This allowed a much faster training and yielded better preliminary results. BN layers were also added at the output of each convolutional layer, which reduced overfitting and improved convergence.

Fig. 8 shows the accuracy and BAC of our best model (ResNet) compared to those of the previous deep learning algorithms. Our best model had a significantly higher accuracy ($p < 0.05$ in the McNemar test) for all segment durations. The accuracy of the ResNet model was greater by 0.27-points (16.5% of remaining errors corrected) for 4 s, 0.22-points (12.0%) for 3 s, 0.25-points (12.1%) for 2 s and 0.56-points (18.7%) for 1 s. The second best model was a recent design combining convolutional networks and a recursive network, but its performance degraded for short EKG segments.

V. DISCUSSION

This study presents a comprehensive analysis of shock decision algorithms based on convolutional networks, introducing

TABLE 2. Median (interdecile range, IDR) Se/Sp scores for the different rhythm types. The Se/Sp values refer to a binary shock decision, and not to a multi-class rhythm type classification.

	Sensitivity (%) ^a		Specificity (%) ^b	
	VF	VT	ORG	AS
CNN				
$T = 4 \text{ s}$	99.0 (98.8-99.2)	90.8 (87.0-93.1)	98.0 (97.7-98.4)	98.7 (98.6-98.8)
$T = 3 \text{ s}$	98.8 (98.6-98.9)	89.0 (85.6-91.7)	98.0 (97.8-98.3)	98.5 (98.4-98.5)
$T = 2 \text{ s}$	98.6 (98.5-98.8)	87.4 (84.9-90.1)	98.0 (97.7-98.2)	98.2 (98.2-98.3)
$T = 1 \text{ s}$	98.2 (98.1-98.3)	84.6 (82.2-87.0)	97.3 (97.1-97.4)	97.8 (97.7-97.9)
ResNet				
$T = 4 \text{ s}$	99.1 (98.8-99.2)	89.3 (85.1-92.4)	98.5 (98.3-98.8)	98.7 (98.6-98.8)
$T = 3 \text{ s}$	98.8 (98.6-99.0)	87.8 (84.5-90.6)	98.4 (98.2-98.6)	98.5 (98.4-98.6)
$T = 2 \text{ s}$	98.6 (98.5-98.8)	86.3 (82.7-89.4)	98.1 (97.9-98.3)	98.3 (98.2-98.4)
$T = 1 \text{ s}$	98.2 (98.0-98.3)	84.5 (82.4-86.7)	97.3 (97.2-97.5)	97.9 (97.8-97.9)

^a Minimum AHA recommendation: $\text{Se}_{\text{VF}} > 90\%$, $\text{Se}_{\text{VT}} > 75\%$

^b Minimum AHA recommendation: $\text{Sp}_{\text{AS}} > 95\%$, $\text{Sp}_{\text{ORG}} > 95\%$

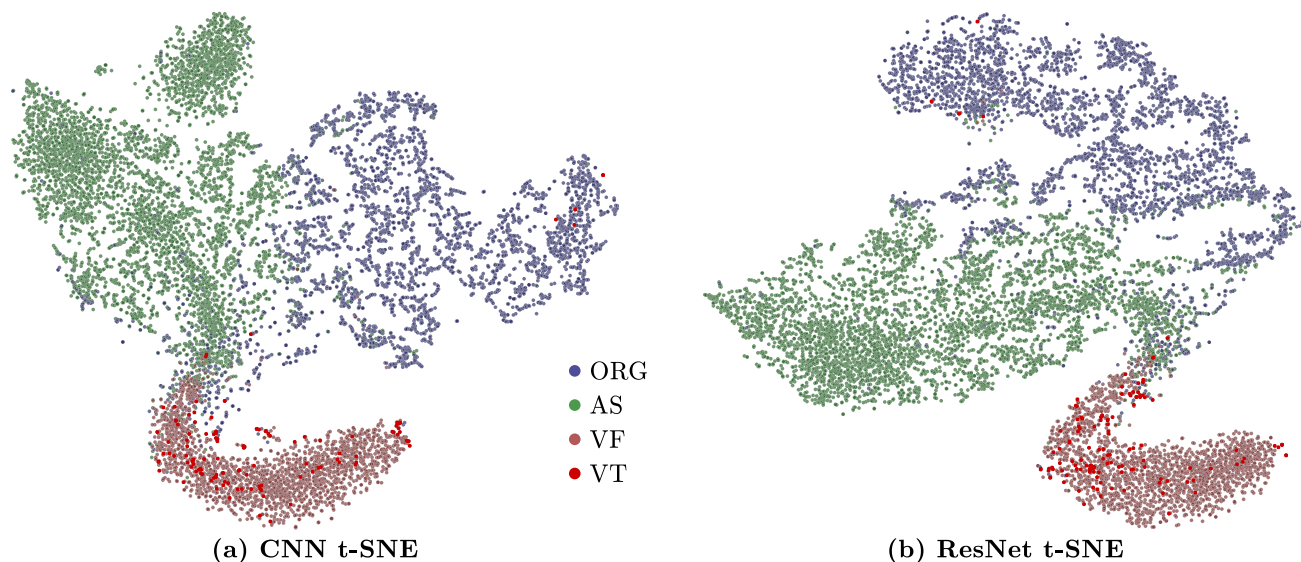


FIGURE 5. Two-dimensional representation of the feature maps extracted by the CNN (left) and ResNet (right) models for the 4 s segments. Different color codes were used to highlight the rhythm types.

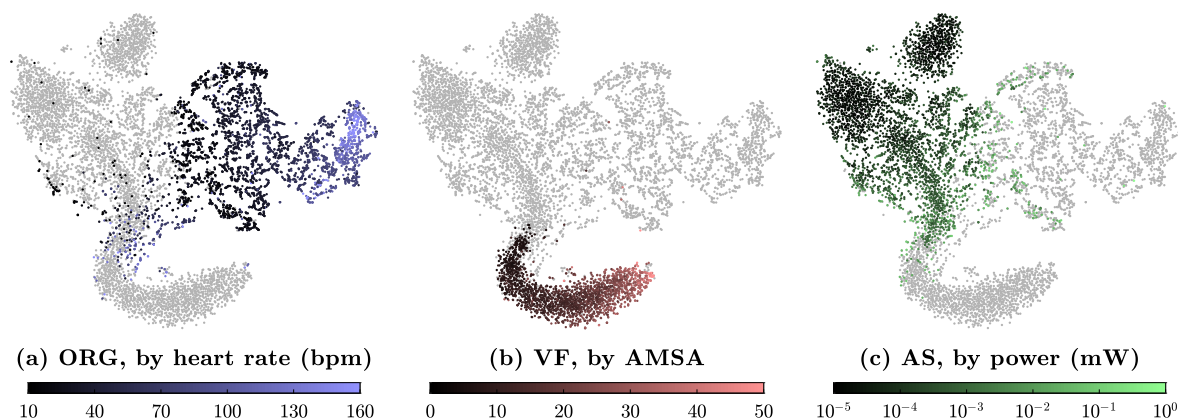


FIGURE 6. 2-D mappings based on the t-SNE algorithm for the different rhythm types. The color grading indicates the value of a typical rhythm characteristic to appreciate the reasons for the mixture of the rhythms and the classification errors.

a new CNN architecture and a ResNet architecture to assess the benefits of deeper networks for this task. Our results show that a rhythm analysis compliant with AHA specifications is possible using EKG segments as short as 1-s, what would contribute to shorten interruptions in CPR for the analysis of the rhythm. Most importantly, all the experiments were done using OHCA data gathered using AEDs, and with quality controlled rhythm annotations, thus ensuring the results are meaningful for the clinical scenario at hand.

Our models had a significantly better accuracy ($p < 0.05$) than the next best deep learning models of previous studies (see Fig. 8). The accuracy of our best model was 0.2-points greater for any segment duration than that of the next best model, an architecture mixing convolutional and recursive layers, and was greater by 0.4-points than any previous convolutional design. This means that at least 10% (20% for convolutional designs) of errors were corrected. Moreover,

the advantages were larger for 1 s segments (see Fig 8), for which AHA compliant shock decision algorithms had not yet been demonstrated. Two reasons may explain this improvement. First, the architecture was inspired by a recent approach using ResNets for general purpose EKG arrhythmia detectors [24]. We used filters of width 16, which produced the best preliminary results, and an increase in channels per block in powers of 2 (with max-pooling layers of 2), and a smaller sampling rate better adapted to AED EKG bandwidths than Acharya *et al.* [25] or Kiranyaz *et al.* [22]. The use of narrower filters with larger sampling rates as in [25] may produce poorer representations, and using very wide filters with larger downsampling (max-pooling) as in [26] produces representations that are too coarse. Second, we introduced BN layers [40], which were not present in previous solutions [22], [25], [26]. These layers stabilize the learning process and add a regularization effect, thus avoiding overfitting

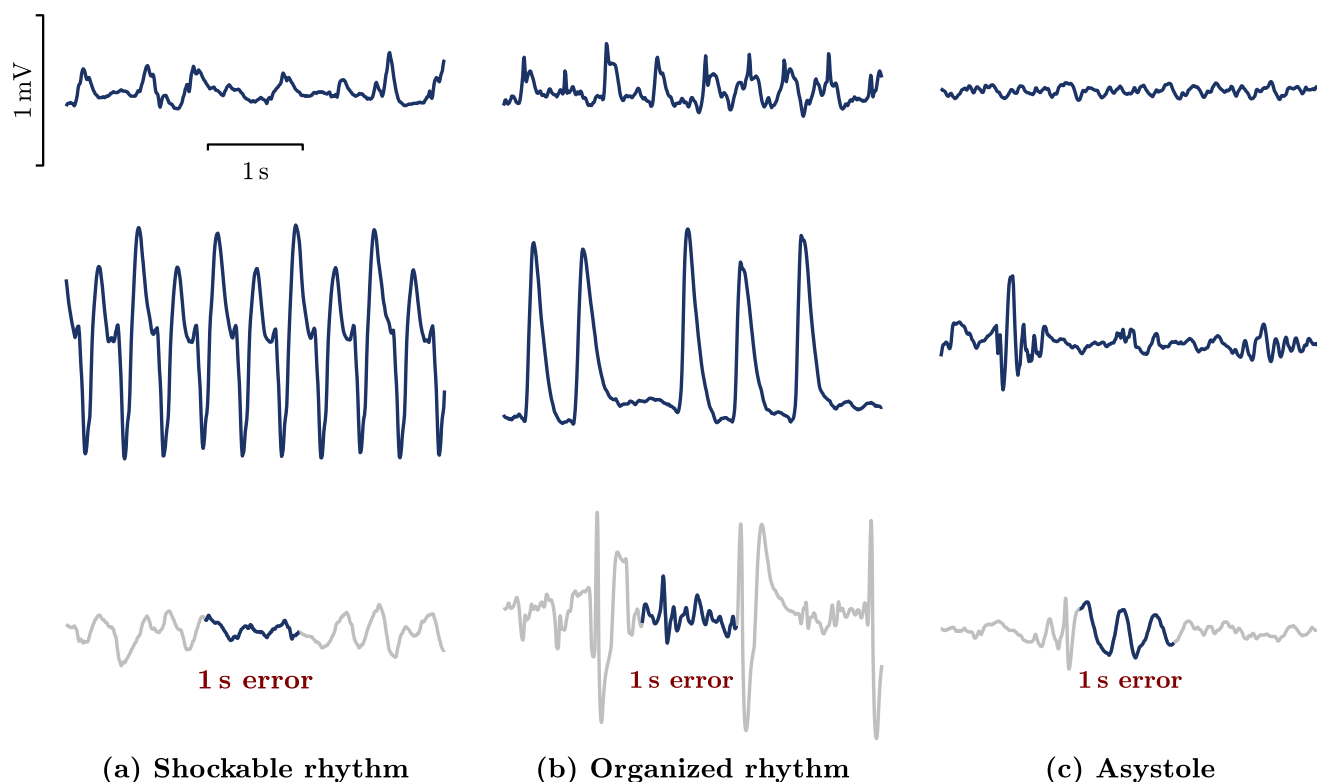


FIGURE 7. Examples of classification errors grouped by rhythm types, for 4 s segments (top) and 1 s segments (low). In the examples below the 1 s segments with classification errors are highlighted, illustrating that a larger field for view (4 s) results in a greater accuracy (see Table 2).

and allowing larger learning rates and fewer epochs to train the networks. These design characteristics were particularly important as the segment duration increased because they allowed the network to learn subtle signal details that differentiate borderline shockable and nonshockable rhythms (see Fig. 5), and explain why our ResNet design was only significantly more accurate ($p < 0.05$) than our CNN design for segment durations of 4 s and 3 s.

From a clinical perspective, our models met, for all segment durations, the minimum Se/Sp values recommended by the AHA. Even for VT, the least prevalent rhythm class, the algorithms were over 9-points (see Table 2) past the minimum 75% AHA recommendation. Our results show, for the first time, that an AHA compliant shock decision is possible analyzing 1-s EKG segments, which is over 3-times shorter than the typical segment duration used in AEDs [7], [8], [48], [49]. Even for 1-s segments, Se was above 97% and Sp was above 97.5%, that is 7-points and 2.5-points over the minimum values established by the AHA, respectively. This ensures a safe (high Sp) and efficient (high Se) use of the algorithm for very short EKG segments, opening the possibility of a combined optimization of CPR therapy with AED use. In addition to shortening no-flow intervals for rhythm analysis, two other possibilities open up. First, a fully automatic rhythm analysis during 30:2 CPR (30 compressions to 2 ventilations) during ventilation intervals [50], an application that involves the use of existing algorithms

to detect pauses in compressions based on the impedance recorded by AEDs [51], [52]. Second, the algorithm could be used to improve current methods for rhythm analysis during CPR [38], which have an unsafe positive predictivity due to an excess of false positives. So when a shockable rhythm is suspected during CPR, the AED could instruct the rescuer to stop CPR for a short confirmatory analysis [7], and then use a very short EKG segment without artifacts to confirm the decision using a very accurate algorithm.

Typically AED memory and computation resources are limited because the devices are equipped with low end processors that handle many other tasks in parallel [53]. So it is important to design simple shock decision algorithms, with low computation and storage demands unlike those of CNN architectures. In order to simplify our architecture we developed thin-CNN and thin-ResNet solutions based on depth-wise separable convolutional layers [54], which separate the filtering process in two steps: first, a 1-D filter of size L is applied to each of the M input channels; and then, N linear combinations of the results are computed as output channels. This reduces the number of trainable weights from $M \cdot N \cdot L$ to $M \cdot L + N \cdot M$. The CNN architecture could be further simplified by replacing the hidden fully connected layer by a global average pooling layer, as done in the ResNet architecture (see Fig. 3), which removes the $D/2^4 \cdot 64 \cdot 10 + 10$ weights of the FC layers. These modifications reduced the number of weights by at least 90% (worst case of 1 s segments).

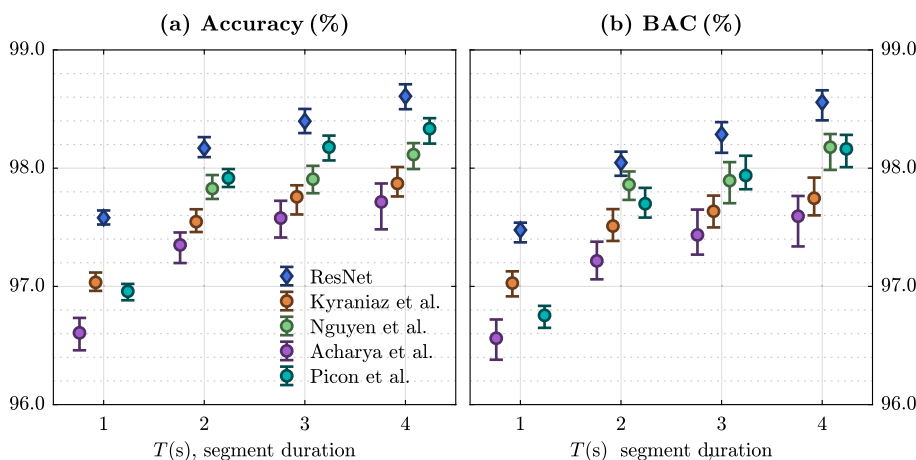


FIGURE 8. Comparison of the accuracy and BAC of the ResNet model of Fig. 3b to the deep learning solutions proposed in Kiranyaz et al [22], Acharya et al [25], Nguyen et al [26] and Picon et al [27]. Note that Nguyen et al [26] cannot be implemented for $T = 1$ s due to filter and max-pooling sizes.

In addition, the number of element-wise products in the convolutional layers is reduced from $N \cdot L \cdot D \cdot M$ to $L \cdot D \cdot M + N \cdot D \cdot M$, a reduction of over 90%, which could make a big difference for low-end hardware not specially designed for matrix computations. Fig. 9 shows the results for the thin solutions compared to the complete solutions of Fig 3. The thin solutions were significantly less accurate than the complete solutions ($p < 0.05$), but there were no significant differences in accuracy between the thin-ResNet and the best previous deep learning solutions. Moreover, the thin-ResNet outperforms all previous convolutional solutions. Our thin solutions are considerably lighter than all previous deep learning solutions [22], [25], [26]. Their performance is above AHA specifications even for segments as short as 1-s, making them an implementable deep learning solution in the type of microprocessors customarily found in AEDs.

Finally, this study originates from a dataset of EKG from OHCA patients recorded using AEDs, and annotated by a

pool of specialized clinicians. By including all OHCA events in a two and a half year period of data collection, we made sure the dataset included a diverse set of over 850 patients that properly represented the differences in gender or age, chest configurations and physiological characteristics. Moreover, by partitioning the data patient-wise we also made sure that different patients were used to learn the characteristics of the OHCA rhythm classes, and to test the accuracy of our algorithms. And very importantly, we made sure our dataset included all OHCA rhythm types, including AS. Our results show that shock decisions for AS were as difficult as for ORG rhythms (see Table 2). All previous studies using deep learning methods had dismissed the use of AS in the design of their algorithms, arguing that AS could be safely identified using a simple power-threshold [26], [27]. Our clustering analysis shows (see Fig. 5) that borderline AS/VF are rhythms pose one of the greatest challenges to the accuracy of the algorithms, and that AS cannot be safely identified using

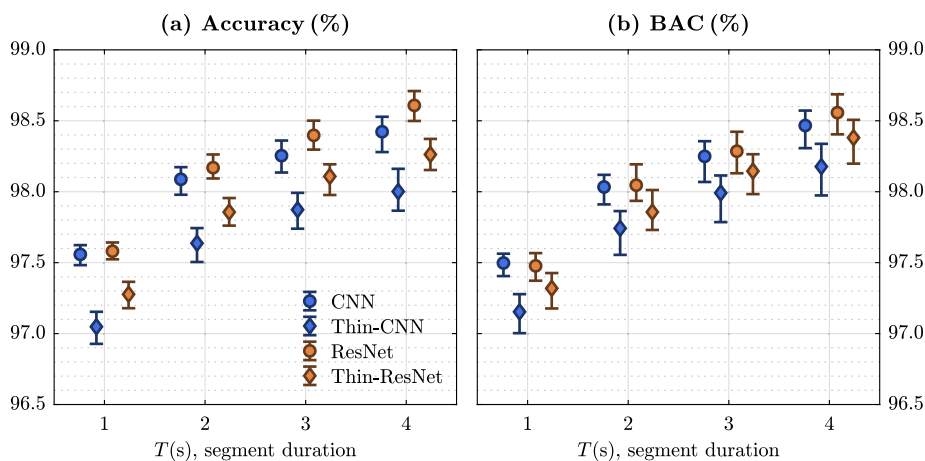


FIGURE 9. Performance degradation for the thin models compared to the complete architectures introduced in this study. The thin architectures are better suited for low-end processors by reducing the storage needs and number of products to compute by at least 90%.

a power threshold (see Fig 6). In OHCA research, obtaining quality audited annotated signal datasets involves large time and money investments to recruit pools of specialized clinicians [36], [45], [48]. For instance, the annotation and quality review of the data for this project was a year long project involving data collection, the development of easily deployable annotation tools, and collecting and auditing all the clinician's annotations [34]. Since the need for large quality audited datasets is one of the limitations of deep learning solutions, more complex algorithmic applications like the classification of OHCA rhythms into 5-classes (4 rhythm types plus presence of pulse) using deep learning may be a challenge [17], [38]. However, our data dimensionality reduction analysis shows that our architectures found richer structures in the data than the shock/no-shock classes. Samples were grouped into three well defined clusters corresponding to the three main rhythm types (VF/AS/ORG), and graded by meaningful waveform characteristics (see Fig. 6). So it is likely that in the future deep learning solutions for multi-class OHCA rhythm annotation could be developed using smaller OHCA datasets and the networks developed in this study as base models for transfer learning [55].

This study has some limitations. First, data were collected using a single device model, and other AED models may have different EKG acquisition circuitry. However, the two fundamental EKG acquisition characteristics of the device, the EKG bandwidth (0.5 – 21 Hz) and amplitude resolution, are typical of AEDs [36]. Consequently, the algorithms would most likely be usable in any AED after EKG resampling ($F_s = 125$ Hz), or fine-tuning of the convolutional networks (with small datasets) if the data had a more restrictive bandwidth. If larger bandwidths were used, such as in monitor-defibrillators, the EKG could be first filtered to the 0.5 – 21 Hz band, and then fed to the convolutional networks. Second, the effect of typical EKG noise sources, such as baseline wander or power-line interference, were not analyzed. However, given the AED analysis bandwidth, these noise sources would be considerably attenuated. The EKG data used for the study were taken from the device after pre-processing, and all the AED analysis intervals of the patient cohort were used. Only intervals with chest compression activity were discarded because CPR must be interrupted for AED rhythm analysis [4]. That is, data were not discarded if the typical EKG noise sources were present, thus ensuring the algorithms are accurate regardless of these noise sources.

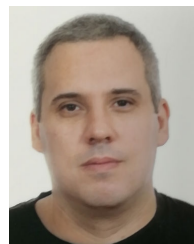
VI. CONCLUSION

New convolutional architectures were proposed for AED shock decision algorithms. The algorithms were trained and tested using OHCA data recorded using AEDs and annotated by a pool of specialized clinicians. The accuracy of our methods improves that of previous solutions, and we demonstrated the possibility of an AHA compliant shock decision with EKG segments as short as 1 s. This should contribute to a combined optimization of defibrillation and CPR to improve OHCA survival.

REFERENCES

- [1] E. J. Benjamin et al., "Heart disease and stroke statistics-2019 update: A report from the American heart association," *Circulation*, vol. 139, pp. e56–e528, Mar. 2019.
- [2] R. W. Neumar et al., "Part 1: Executive summary: 2015 American heart association guidelines update for cardiopulmonary resuscitation and emergency cardiovascular care.," *Circulation*, vol. 132, no. 18, pp. S315–S367, Nov. 2015.
- [3] R. E. Kerber, L. B. Becker, J. D. Bourland, R. O. Cummins, A. P. Hallstrom, M. B. Michos, G. Nichol, J. P. Ornato, W. H. Thies, R. D. White, and B. D. Zuckerman, "Automatic external defibrillators for public access defibrillation: Recommendations for specifying and reporting arrhythmia analysis algorithm performance, incorporating new waveforms, and enhancing safety. A statement for health professionals from the American heart association task force on automatic external defibrillation, subcommittee on AED safety and efficacy," *Circulation*, vol. 95, no. 6, pp. 1677–1682, 1997.
- [4] S. R. de Gauna, U. Irusta, J. Ruiz, U. Ayala, E. Aramendi, and T. Eftestøl, "Rhythm analysis during cardiopulmonary resuscitation: Past, present, and future," *Biomed. Res. Int.*, vol. 2014, Jan. 2014, Art. no. 386010.
- [5] D. Snyder and C. Morgan, "Wide variation in cardiopulmonary resuscitation interruption intervals among commercially available automated external defibrillators may affect survival despite high defibrillation efficacy," *Crit. Care Med.*, vol. 32, pp. S421–S424, Sep. 2004.
- [6] S. Cheskes, R. H. Schmicker, J. Christenson, D. D. Salcido, T. Rea, J. Powell, D. P. Edelson, R. Sell, S. May, J. J. Menegazzi, L. Van Ottingham, M. Olsufka, S. Pennington, J. Simonini, R. A. Berg, I. Stiell, A. Idris, B. Bigham, and L. Morrison, "Perishock pause: An independent predictor of survival from out-of-hospital shockable cardiac arrest.," *Circulation*, vol. 124, no. 1, pp. 58–66, Jul. 2011.
- [7] J.-P. Didon, V. Krasteva, S. Ménétré, T. Stoyanov, and I. Jekova, "Shock advisory system with minimal delay triggering after end of chest compressions: Accuracy and gained hands-off time," *Resuscitation*, vol. 82, no. 2, pp. S8–S15, Dec. 2011.
- [8] U. Irusta, J. Ruiz, E. Aramendi, S. R. de Gauna, U. Ayala, and E. Alonso, "A high-temporal resolution algorithm to discriminate shockable from nonshockable rhythms in adults and children," *Resuscitation*, vol. 83, no. 9, pp. 1090–1097, Sep. 2012.
- [9] R. H. Clayton, A. Murray, and R. W. Campbell, "Comparison of four techniques for recognition of ventricular fibrillation from the surface ECG," *Med. Biol. Eng. Comput.*, vol. 31, no. 2, pp. 111–117, Mar. 1993.
- [10] N. V. Thakor, Y. S. Zhu, and K. Y. Pan, "Ventricular tachycardia and fibrillation detection by a sequential hypothesis testing algorithm," *IEEE Trans. Biomed. Eng.*, vol. 37, no. 9, pp. 837–843, Sep. 1990.
- [11] S. Barro, R. Ruiz, D. Cabello, and J. Mira, "Algorithmic sequential decision making in the frequency domain for life threatening ventricular arrhythmias and imitative artifacts: A diagnostic system," *J. Biomed. Eng.*, vol. 11, no. 4, pp. 320–328, Jul. 1989.
- [12] L. Khadra, A. S. Al-Fahoum, and H. Al-Nashash, "Detection of life-threatening cardiac arrhythmias using the wavelet transformation," *Med. Biol. Eng. Comput.*, vol. 35, no. 6, pp. 626–632, Nov. 1997.
- [13] A. Amann, R. Tratnig, and K. Unterkofler, "Reliability of old and new ventricular fibrillation detection algorithms for automated external defibrillators," *Biomed. Eng. Online*, vol. 4, no. 1, p. 60, Oct. 2005.
- [14] Q. Li, C. Rajagopalan, and G. D. Clifford, "Ventricular fibrillation and tachycardia classification using a machine learning approach," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 6, pp. 1607–1613, Jun. 2014.
- [15] F. Alonso-Atienza, E. Morgado, L. Fernandez-Martinez, A. Garcia-Alberola, and J. L. Rojo-Alvarez, "Detection of life-threatening arrhythmias using feature selection and support vector machines," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 3, pp. 832–840, Mar. 2014.
- [16] R. K. Tripathy, L. N. Sharma, and S. Dandapat, "Detection of shockable ventricular arrhythmia using variational mode decomposition," *J. Med. Syst.*, vol. 40, no. 4, p. 79, Apr. 2016.
- [17] A. B. Rad, T. Eftestøl, K. Engan, U. Irusta, J. T. Kvaloy, J. Kramer-Johansen, L. Wik, and A. K. Katsaggelos, "ECG-based classification of resuscitation cardiac rhythms for retrospective data analysis," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 10, pp. 2411–2418, Oct. 2017.
- [18] C. Figuera, U. Irusta, E. Morgado, E. Aramendi, U. Ayala, L. Wik, J. Kramer-Johansen, T. Eftestøl, and F. Alonso-Atienza, "Machine learning techniques for the detection of shockable rhythms in automated external defibrillators," *PLoS ONE*, vol. 11, no. 7, Jul. 2016, Art. no. e0159654.

- [19] O. Faust, Y. Hagiwara, T. J. Hong, O. S. Lih, and U. R. Acharya, "Deep learning for healthcare applications based on physiological signals: A review," *Comput. Methods Programs Biomed.*, vol. 161, pp. 1–13, Jul. 2018.
- [20] B. Rim, N.-J. Sung, S. Min, and M. Hong, "Deep learning in physiological signal data: A survey," *Sensors*, vol. 20, no. 4, p. 969, Feb. 2020.
- [21] S. Hong, Y. Zhou, J. Shang, C. Xiao, and J. Sun, "Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review," 2019, *arXiv:2001.01550*. [Online]. Available: <http://arxiv.org/abs/2001.01550>
- [22] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 664–675, Mar. 2016.
- [23] R. S. Andersen, A. Peimankar, and S. Puthusserypady, "A deep learning approach for real-time detection of atrial fibrillation," *Expert Syst. Appl.*, vol. 115, pp. 465–473, Jan. 2019.
- [24] A. Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Med.*, vol. 25, no. 1, pp. 65–69, Jan. 2019.
- [25] U. R. Acharya, H. Fujita, S. L. Oh, U. Raghavendra, J. H. Tan, M. Adam, A. Gertych, and Y. Hagiwara, "Automated identification of shockable and non-shockable life-threatening ventricular arrhythmias using convolutional neural network," *Future Gener. Comput. Syst.*, vol. 79, pp. 952–959, Feb. 2018.
- [26] M. T. Nguyen, B. V. Nguyen, and K. Kim, "Deep feature learning for sudden cardiac arrest detection in automated external defibrillators," *Sci. Rep.*, vol. 8, no. 1, p. 17196, Nov. 2018.
- [27] A. Picon, U. Irusta, A. Álvarez-Gila, E. Aramendi, F. Alonso-Atienza, C. Figuera, U. Ayala, E. Garrote, L. Wik, J. Kramer-Johansen, and T. Eftestøl, "Mixed convolutional and long short-term memory network for the detection of lethal ventricular arrhythmia," *PLoS ONE*, vol. 14, no. 5, May 2019, Art. no. e0216756.
- [28] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. E215–E220, Jun. 2000.
- [29] J. C. Reynolds, D. D. Salcido, and J. J. Menegazzi, "Correlation between coronary perfusion pressure and quantitative ECG waveform measures during resuscitation of prolonged ventricular fibrillation," *Resuscitation*, vol. 83, no. 12, pp. 1497–1502, Dec. 2012.
- [30] B. Chicote, U. Irusta, E. Aramendi, R. Alcaraz, J. Rieta, I. Isasi, D. Alonso, M. Baqueriza, and K. Ibarguren, "Fuzzy and sample entropies as predictors of patient survival using short ventricular fibrillation recordings during out of hospital cardiac arrest," *Entropy*, vol. 20, no. 8, p. 591, Aug. 2018.
- [31] M. Hulleman, H. Mes, M. T. Blom, and R. W. Koster, "Conduction disorders in bradysystolic out-of-hospital cardiac arrest," *Resuscitation*, vol. 106, pp. 113–119, Sep. 2016.
- [32] A. Elola, E. Aramendi, U. Irusta, J. Del Ser, E. Alonso, and M. Daya, "ECG-based pulse detection during cardiac arrest using random forest classifier," *Med. Biol. Eng. Comput.*, vol. 57, no. 2, pp. 453–462, Feb. 2019.
- [33] J. P. Nolan, J. Soar, G. B. Smith, C. Gwinnutt, F. Parrott, S. Power, D. A. Harrison, E. Nixon, and K. Rowan, "Incidence and outcome of in-hospital cardiac arrest in the united kingdom national cardiac arrest audit," *Resuscitation*, vol. 85, no. 8, pp. 987–992, Aug. 2014.
- [34] C. Corcuera, G. Galdos, M. Olabarria, J. Veintemillas, A. Larrea, J. M. Bastida, K. Ibarguren, D. Alonso, B. Chicote, E. Aramendi, and U. Irusta, "Assessment of the diagnoses of automated external defibrillators operated by basic life support personnel," *Resuscitation*, vol. 118, p. e1, Sep. 2017.
- [35] K. Ibarguren, J. M. Unanue, D. Alonso, I. Vaqueriza, U. Irusta, E. Aramendi, and B. Chicote, "Difference in survival from pre-hospital cardiac arrest between cities and villages in the basque autonomous community," *Resuscitation*, vol. 96, p. 114, Nov. 2015.
- [36] E. Atkinson, B. Mikysa, J. A. Conway, M. Parker, K. Christian, J. Deshpande, T. K. Knilans, J. Smith, C. Walker, R. E. Stickney, D. R. Hampton, and M. F. Hazinski, "Specificity and sensitivity of automated external defibrillator rhythm analysis in infants and children," *Ann. Emergency Med.*, vol. 42, no. 2, pp. 185–196, Aug. 2003.
- [37] U. Ayala, U. Irusta, J. Ruiz, T. Eftestøl, J. Kramer-Johansen, F. Alonso-Atienza, E. Alonso, and D. González-Otero, "A reliable method for rhythm analysis during cardiopulmonary resuscitation," *BioMed. Res. Int.*, vol. 2014, Mar. 2014, Art. no. 872470.
- [38] I. Isasi, U. Irusta, A. B. Rad, E. Aramendi, M. Zabihi, T. Eftestøl, J. Kramer-Johansen, and L. Wik, "Automatic cardiac rhythm classification with concurrent manual chest compressions," *IEEE Access*, vol. 7, pp. 115147–115159, 2019.
- [39] T. R. Nichols, P. M. Wisner, G. Cripe, and L. Gulabchand, "Putting the kappa statistic to use," *Qual. Assurance J.*, vol. 13, nos. 3–4, pp. 57–61, Jul. 2010.
- [40] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: <https://arxiv.org/abs/1502.03167>
- [41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [42] D. Han, J. Kim, and J. Kim, "Deep pyramidal residual networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2017, pp. 6307–6315.
- [43] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013, *arXiv:1312.4400*. [Online]. Available: <http://arxiv.org/abs/1312.4400>
- [44] V. Laurens and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 2605, pp. 2579–2605, 2008.
- [45] L. Wik, J. Kramer-Johansen, H. Myklebust, H. Sørebo, L. Svensson, B. Fellows, and P. A. Steen, "Quality of cardiopulmonary resuscitation during out-of-hospital cardiac arrest," *J. Amer. Med. Assoc.*, vol. 293, pp. 299–304, Jan. 2005.
- [46] G. Ristagno, T. Mauri, G. Cesana, Y. Li, A. Finzi, F. Fumagalli, G. Rossi, N. Grieco, M. Migliori, A. Andreassi, R. Latini, C. Fornari, and A. Pesenti, "Amplitude spectrum area to guide defibrillation: A validation on 1617 patients with ventricular fibrillation," *Circulation*, vol. 131, pp. 478–487, Feb. 2015.
- [47] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014.
- [48] D. L. Atkins, W. A. Scott, A. D. Blaurox, I. H. Law, M. Dick, II, F. Geheb, J. Sobh, and J. E. Brewer, "Sensitivity and specificity of an automated external defibrillator algorithm designed for pediatric patients," *Resuscitation*, vol. 76, no. 2, pp. 168–174, Feb. 2008.
- [49] F. Cecchin, D. B. Jorgenson, C. I. Berul, J. C. Perry, A. A. Zimmerman, B. W. Duncan, F. M. Lupinetti, D. Snyder, T. D. Lyster, G. L. Rosenthal, B. Cross, and D. L. Atkins, "Is arrhythmia detection by automatic external defibrillator accurate for children? Sensitivity and specificity of an automatic external defibrillator algorithm in 696 pediatric arrhythmias," *Circulation*, vol. 103, no. 20, pp. 2483–2488, May 2001.
- [50] J. Ruiz, U. Ayala, S. R. de Gauna, U. Irusta, D. González-Otero, E. Alonso, J. Kramer-Johansen, and T. Eftestøl, "Feasibility of automated rhythm assessment in chest compression pauses during cardiopulmonary resuscitation," *Resuscitation*, vol. 84, no. 9, pp. 1223–1228, Sep. 2013.
- [51] U. Ayala, U. Irusta, J. Ruiz, S. R. de Gauna, D. González-Otero, E. Alonso, J. Kramer-Johansen, H. Naas, and T. Eftestøl, "Fully automatic rhythm analysis during chest compression pauses," *Resuscitation*, vol. 89, pp. 25–30, Apr. 2015.
- [52] J. Coult, J. Blackwood, T. D. Rea, P. J. Kudenchuk, and H. Kwok, "A method to detect presence of chest compressions during resuscitation using transthoracic impedance," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 3, pp. 768–774, Mar. 2020.
- [53] T. S. Takata, R. L. Page, and J. A. Joglar, "Automated external defibrillators: Technical considerations and clinical promise," *Ann. Internal Med.*, vol. 135, pp. 990–998, Dec. 2001.
- [54] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE*, Jul. 2017, pp. 1800–1807.
- [55] G. Van Steenkiste, G. van Loon, and G. Crevecoeur, "Transfer learning in ECG classification from human to horse using a novel parallel neural network architecture," *Sci. Rep.*, vol. 10, no. 1, p. 186, Jan. 2020.



XABIER JAUREGUIBEITIA received the M.Sc. degree in telecommunications engineering from the University of the Basque Country (UPV/EHU), in 2016, where he is currently pursuing the Ph.D. degree. He is currently working on the applications of machine learning and deep-learning methods to develop automatic detection and classification algorithms for pre-hospital cardiac arrest treatment. His work was financially supported by a competitive grant from the Basque Government, a part of the BioRes Research Group, UPV/EHU. His fields of expertise are biomedical signal processing, machine learning, and deep learning.



research within the Biores Research Group, UPV/EHU.

GORKA ZUBIA was born in Barakaldo, Spain, in 1996. He received the B.Sc. degree in telecommunications engineering from the University of the Basque Country (UPV/EHU), in 2019, where he is currently pursuing the M.Sc. degree in telecommunications engineering. His research interests include statistical learning theory, deep learning, and biomedical signal processing. He was awarded the Ikasiker Research Collaboration Fellowship by the Basque Government to conduct



Research Group, recognized as outstanding by Basque Science System. His fields of expertise are biomedical signal processing, machine learning, and data management in the fields of pre-hospital emergency medicine and cardiac arrest. In this field, he has collaborated with leading European and U.S. clinicians, and he has published over 40 articles in SCI-IF journals and contributed over 100 communications in scientific conferences.

UNAI IRUSTA (Member, IEEE) was born in Bilbao, Spain, in 1973. He received the M.Sc. degree (Hons.) in telecommunications engineering from the University of the Basque Country (UPV/EHU) in 1998. After working with international companies in the mobile communications sector, he joined as an Assistant Professor with the UPV/EHU, in 2003, where he has been an Associate Professor, since 2011. He is a Founding Member of the Bioengineering and Resuscitation



of expertise are statistical signal processing, biomedical signal processing, and data management in the fields of pre-hospital emergency medicine and cardiac arrest. In this field, she has collaborated with leading researchers of the Regional Outcomes Consortium (ROC), and she has published over 30 articles in SCI-IF journals and contributed over 100 communications in scientific conferences.

ELISABETE ARAMENDI (Member, IEEE) was born in Azkoitia, Spain, in 1969. She received the M.Sc. degree in telecommunications engineering from the University of the Basque Country (UPV/EHU) in 1993. She has been an Assistant Professor with the UPV/EHU, since 1994, and an Associate Professor, since 2002. She is a Founding Member and the Director of the Bioengineering and Resuscitation Research Group, recognized as outstanding by Basque Science System. Her fields



Smart Manufacturing. Her research was focused on developing new processing techniques for shock outcome prediction in the fields of pre-hospital emergency medicine and cardiac arrest.

BEATRIZ CHICOTE received the B.Sc. degree in industrial electronics engineering, the M.Sc. degree in advanced electronic systems, and the Ph.D. degree in electronics and telecommunication from the University of the Basque Country (UPV/EHU), in 2013, 2015, and 2019, respectively. In 2019, she joined the Lortek Technological Center, IK4-Lortek, a member of the Basque Research and Technology Alliance, where she is currently a member of the Digital Department of



to the European registry (Eureka initiative). He has participated in numerous research projects from competitive calls, has published eight articles in SCI-IF journals, and has participated in many national conferences on emergency medicine. His fields of research include the management of cardiac arrest patients, new technologies in the treatment of cardiac arrest, and the management of cardiac arrest registries.

DANIEL ALONSO was born in Donostia, Spain, in 1961. He received the degree in nursing from the University of the Basque Country (UPV/EHU), in 1982. He is currently with the Basque Emergency Medical System (Emergentziak-Osakidetza), Basque Health Service, Hospital Donostia, where he is also the Scientific Coordinator of the Basque Cardiac Arrest Registry. He is a part of the Spanish initiative for a nation-wide cardiac arrest registry and contributes



Emergency Medicine and Emergencies (SEMES), Basque Country, and the Spanish Committee of International Trauma Life Support (ITLS). He is an expert in the field of pre-hospital emergency medicine, where he has taught many courses in workshops, participated in forums, contributed SCI-IF articles, and many contributions to scientific conferences.

ANDIMA LARREA was born in Etxebarri, Spain, in 1966. He received the degree in medicine and surgery from the University of the Basque Country (UPV/EHU), in 1992, and the degree in epidemiology and public health, in 1999. He has been working as an Associate Doctor with the Emergency Unit of the Basque Government's Department of Health, Basque Health Service, since 1993. He is a member of the CPR Group. He is a member of the Board of Directors of the Spanish Society of



OHCA.

...