



Article

Rhythm Analysis during Cardiopulmonary Resuscitation Using Convolutional Neural Networks

Iraia Isasi ^{1,*} , Unai Irusta ¹ , Elisabete Aramendi ¹ , Trygve Eftestøl ², Jo Kramer-Johansen ³  and Lars Wik ³

¹ Department of Communications Engineering, University of the Basque Country UPV/EHU, 48013 Bilbao, Spain; unai.irusta@ehu.eus (U.I.); elisabete.aramendi@ehu.eus (E.A.)

² Department of Electrical Engineering and Computer Science, University of Stavanger, 4036 Stavanger, Norway; trygve.eftestol@uis.no

³ Norwegian National Advisory Unit on Prehospital Emergency Medicine (NAKOS), Oslo University Hospital and University of Oslo, 0424 Oslo, Norway; jo.kramer-johansen@medisin.uio.no (J.K.-J.); lars.wik@medisin.uio.no (L.W.)

* Correspondence: irai.isasi@ehu.eus; Tel.: +34-946-01-73-86

Received: 15 May 2020; Accepted: 26 May 2020; Published: 27 May 2020



Abstract: Chest compressions during cardiopulmonary resuscitation (CPR) induce artifacts in the ECG that may provoke inaccurate rhythm classification by the algorithm of the defibrillator. The objective of this study was to design an algorithm to produce reliable shock/no-shock decisions during CPR using convolutional neural networks (CNN). A total of 3319 ECG segments of 9 s extracted during chest compressions were used, whereof 586 were shockable and 2733 nonshockable. Chest compression artifacts were removed using a Recursive Least Squares (RLS) filter, and the filtered ECG was fed to a CNN classifier with three convolutional blocks and two fully connected layers for the shock/no-shock classification. A 5-fold cross validation architecture was adopted to train/test the algorithm, and the process was repeated 100 times to statistically characterize the performance. The proposed architecture was compared to the most accurate algorithms that include handcrafted ECG features and a random forest classifier (baseline model). The median (90% confidence interval) sensitivity, specificity, accuracy and balanced accuracy of the method were 95.8% (94.6–96.8), 96.1% (95.8–96.5), 96.1% (95.7–96.4) and 96.0% (95.5–96.5), respectively. The proposed algorithm outperformed the baseline model by 0.6-points in accuracy. This new approach shows the potential of deep learning methods to provide reliable diagnosis of the cardiac rhythm without interrupting chest compression therapy.

Keywords: out-of-hospital cardiac arrest (OHCA); cardiopulmonary resuscitation (CPR); electrocardiogram (ECG); adaptive filter; deep learning; machine learning; convolutional neural network (CNN); random forest (RF) classifier

1. Introduction

Out of hospital cardiac arrest (OHCA) is one of the leading causes of death worldwide [1,2]. The two key life saving therapies are defibrillation (electric shock) when the cardiac rhythm is ventricular fibrillation (VF) or tachycardia (VT), and cardiopulmonary resuscitation (CPR) [3]. The defibrillator monitors the electrocardiogram (ECG), and includes a shock/no-shock algorithm that analyzes the patient's ECG to detect VF/VT [4]. The American Heart Association (AHA) has established the minimum accuracy requirements for these algorithms [5]. Shockable rhythms should be detected with a minimum sensitivity (Se) of 90% to properly identify defibrillation treatment conditions. The specificity (Sp) for detection of nonshockable rhythms must be above 95% to avoid unnecessary shocks that may damage the myocardium or deteriorate the quality of CPR.

The mechanical activity of chest compressions during CPR induces artifacts in the ECG that impede a reliable shock/no-shock decision by the defibrillator [6]. Therefore, defibrillators prompt the rescuers to stop chest compressions for rhythm analysis every 2 minutes [7,8]. These hands off (or no flow) intervals lead to intermittent periods with no cerebral and myocardial blood flow that deteriorate the patient's condition, and compromise survival [7,9–11]. Consequently, many biomedical engineering solutions have been proposed over the years to allow an AHA compliant shock/no-shock decision during CPR [12], but none of these solutions have yet a sufficient positive predictivity to be implemented in commercial defibrillators. These methods are based on adaptive filters to remove CPR artifacts. Adaptive filters are needed to address the time and frequency variability of the artifact and its spectral overlap with OHCA rhythms [13]. These filters use signals recorded by the defibrillator like compression depth (CD) or thoracic impedance (TI) to model the artifact [14,15]. Several adaptive approaches have been demonstrated including Wiener filters [16], Matching Pursuit Algorithms [17], Recursive Least Squares (RLS) [18], Least Mean Squares (LMS) [19], or Kalman filters [20,21]. Once the artifact is removed the ECG is analyzed using the shock/no-shock algorithms in defibrillators, or ad-hoc algorithms specially designed to analyze the filtered ECG [17,19,22]. The latter have shown the highest Se/Sp values by exploiting recent advances in ECG feature extraction and classical machine learning algorithms. ECG features are customarily computed in time, frequency or time-frequency domains [23–26]. These features have been efficiently combined using classical machine learning classification algorithms like support vector machines (SVM) or random forests (RF) [22,25,26].

Recently, deep learning approaches have proven to be superior to classical machine learning algorithms in many biomedical signal applications [27,28], including arrhythmia classification based on the ECG waveform [29–33]. Deep learning algorithms using convolutional neural networks (CNN) are end-to-end solutions in which the algorithm learns efficient internal representations of the data (features) and combines them to solve the classification task [34,35]. Deep learning algorithms have already been shown to outperform classical machine learning algorithms in some OHCA applications, such as detection of VF in artifact free ECG [30,36], or the detection of pulse [37]. However, deep learning has not been applied to design algorithms that give accurate shock/no-shock decisions during CPR.

The objective of this study was to design the first deep learning solution to discriminate shockable from nonshockable rhythms during CPR. The method comprises two stages, an adaptive RLS filter to remove CPR artifacts from the ECG followed by a CNN to classify the filtered ECG. The paper is organized as follows: the study dataset is detailed in Section 2, Section 3 describes the methodology including the CNN architecture and the evaluation procedure. The results are presented in Section 4, discussed in Section 5 and the main conclusions are presented in Section 6.

2. Materials

Data were extracted from a large prospective clinical trial designed to measure CPR quality during OHCA [38]. The study was conducted between March 2002 and September 2004 by the emergency services of London, Stockholm and Akershus (Norway). CPR was performed using prototype defibrillators based on HeartStart 4000 (Philips Medical Systems, Andover, MA, USA) together with a sternal CPR assist pad fitted with an accelerometer (ADXL202e, AnalogDevice, Norwood, Mass). The raw data for this study consisted of the ECG and TI signals acquired through the defibrillation pads and the CD signal derived from accelerometer data [16]. Defibrillator data was anonymized and converted to Matlab (MathWorks Inc, Natick, MA, USA) using a sampling rate of 250 Hz. The ECG had an amplitude resolution of 1.031 μV per least significant bit. A notch filter and a Hampel filter were used to remove powerline interferences and spiky artifacts from the ECG [37]. Finally, chest compressions instants (t_k) were automatically marked using a negative peak detector with a 1 cm threshold on the CD signal (see Figure 1, peak detection Th) [15].

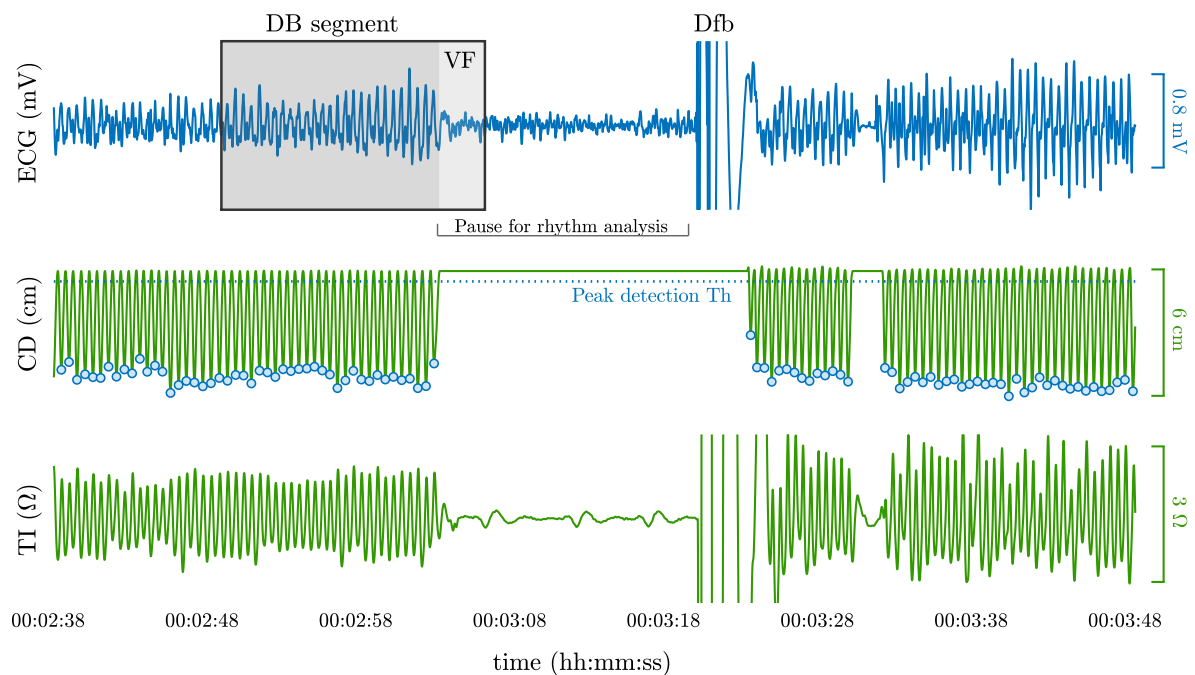


Figure 1. A 70 s interval from an OHCA episode showing the ECG, CD and TI signals. Activity shows CPR followed by a pause for rhythm analysis, the delivery of a defibrillation shock (Dfb) and immediate resumption of CPR. The interval highlighted in grey corresponds to a 15.5 s segment in the dataset. During the first 12.5 s of the segment chest compressions were delivered (see activity in TI and CD), and in the last 3 s there were no compressions and the ground truth rhythm (VF) for the whole segment could be annotated.

The rhythms in the OHCA episodes were originally annotated by two experienced resuscitation researchers/practitioners, a biomedical engineer and an anesthesiologist [38]. For the purpose of this study the rhythm annotations were grouped into shockable and nonshockable. Shockable rhythms comprised lethal ventricular arrhythmia, predominantly VF but also pulseless VT. Non-shockable rhythms included asystole (AS), the absence of electrical activity, and organized rhythms (ORG), or rhythms with visible QRS complexes. The OHCA episodes had median (interquartile range, IQR) durations of 26 min (17–33). From these episodes 15.5 s segments were automatically extracted following these criteria: unique rhythm type in the segment and an interval of 12.5 s with ongoing compressions followed or preceded by a 3 s interval without compressions. The 12.5 s interval with ongoing compressions was used to develop the shock/no-shock decision algorithm, and the 3 s segment was used to confirm the original rhythm annotation in an artifact free ECG. All the data were visually revised (double blind process by authors UI and TE) to ensure compliance with the extraction criteria and the correctness of the rhythm annotations. The annotated dataset contained 3319 segments from 272 OHCA patients, whereof 586 were shockable and 2733 (1192 AS and 1541 ORG) were nonshockable.

3. Methods

The shock/no-shock decision algorithms proposed in this study are composed of two stages. First, an adaptive RLS filter was used to remove chest compression artifacts from the ECG. Then shock/no-shock decision algorithms were designed to classify the filtered ECG using CNNs. In what follows $t = n \cdot T_s$, where $T_s = 4$ ms is the sampling period ($f_s = 250$ Hz), and n is the sample index.

3.1. CPR Artifact Suppressing Filter

CPR artifacts were suppressed using a state-of-the-art method [26,39] based on a RLS filter designed to remove periodic interferences [40]. The CPR artifact is modeled as a quasi-periodic

interference using a Fourier series truncated to N terms (harmonics). The fundamental frequency of the artifact is that of the chest compressions [19], which is assumed constant during a chest compression, but variable from compression to compression. This means that for an interval between two successive compressions at time points, t_{k-1} and t_k (see Figure 2), the frequency can be expressed as

$$f_0(n) = \frac{1}{t_k - t_{k-1}} \quad t_{k-1} \leq nT_s < t_k \quad (1)$$

and the N -term Fourier series representation is then:

$$\hat{s}_{\text{cpr}}(n) = A(n) \sum_{\ell=1}^N (a_{\ell}(n) \cos(\ell 2\pi f_0(n) T_s n) + b_{\ell}(n) \sin(\ell 2\pi f_0(n) T_s n)) \quad (2)$$

where $A(n)$ is an amplitude envelope which differentiates intervals with ($A = 1$) and without compressions ($A = 0$), N is the number of harmonics in the Fourier series and $f_0(n)$ is the instantaneous chest compression frequency.

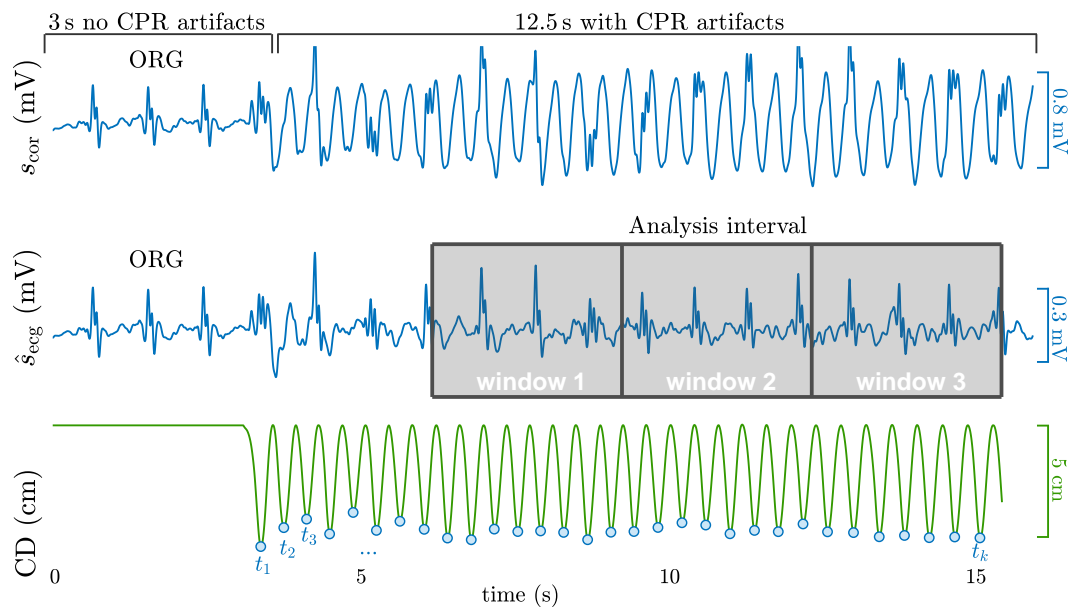


Figure 2. A 15.5 s segment from the study dataset corresponding to a patient in an organized rhythm is shown. In the initial 3 s interval without compressions three QRS complexes are visible, and the nonshockable rhythm annotation was confirmed. The following 12.5 s are corrupted by CPR artifacts (top panel) that conceal the underlying rhythm. The output of the adaptive filter, $\hat{s}_{\text{ecg}}(n)$, reveals the underlying rhythm during chest compressions. CPR activity and the chest compression instants (t_k) can be observed in the CD signal (bottom).

The RLS filter adaptively estimates the time-varying Fourier coefficients, $a_{\ell}(n)$ and $b_{\ell}(n)$, of the CPR artifact, $\hat{s}_{\text{cpr}}(n)$, by minimizing in each iteration the error between the corrupted ECG, $s_{\text{cor}}(n)$, and the estimated underlying ECG, $\hat{s}_{\text{ecg}}(n)$, only around the spectral components of the CPR artifact, that is $f_0(n)$ and its harmonics. The underlying ECG is estimated assuming an additive noise model, so $\hat{s}_{\text{ecg}}(n) = s_{\text{cor}}(n) - \hat{s}_{\text{cpr}}(n)$. A detailed description of the RLS filter equations is available in [39], and the values recommended there to suppress CPR artifacts were used in this study, that is, $N = 4$ and a forgetting factor of $\lambda = 0.999$ [39].

The shock/no-shock algorithms trained and evaluated in this study comprise algorithms based on CNNs (core methods of the paper), and a state of the art algorithm based on classical machine learning techniques used as a baseline model for comparison. In both cases, the algorithms were designed to analyze the filtered ECG in the interval from 3–12 s during compressions (see analysis

interval in Figure 2). That is, the algorithms use 9 s of the filtered ECG for a decision, excluding the initial 3 s to avoid RLS filtering transients [39]. The analysis interval was further divided into three non-overlapping analysis windows of 3 s (see Figure 2) and the shock/no-shock decision was obtained as the majority vote. The combination of consecutive analysis windows is a typical design practice in shock/no-shock decision algorithms for defibrillators [41,42], because it increases the reliability of the diagnosis by avoiding the effects of transient lower quality signal intervals, rhythm changes or filtering miss-adjustments.

Algorithm Based on CNNs

Figure 3 shows the architecture of the shock/no-shock decision algorithms based on CNNs. First the 3 s window of the filtered ECG is downsampled to 125 Hz, resulting in a 1-D signal of $N = 375$ samples, $\hat{s}_{\text{ecg}}(n)$. Then the CNN is composed of three convolutional blocks to extract the high level descriptors of the ECG, and two fully connected layers for classification. The b -th convolutional block consists of a convolutional layer with J_b filters of width I_b , followed by a batch normalization layer, a rectified linear unit (ReLU), a max-pooling layer ($K = 3$) and a dropout layer.

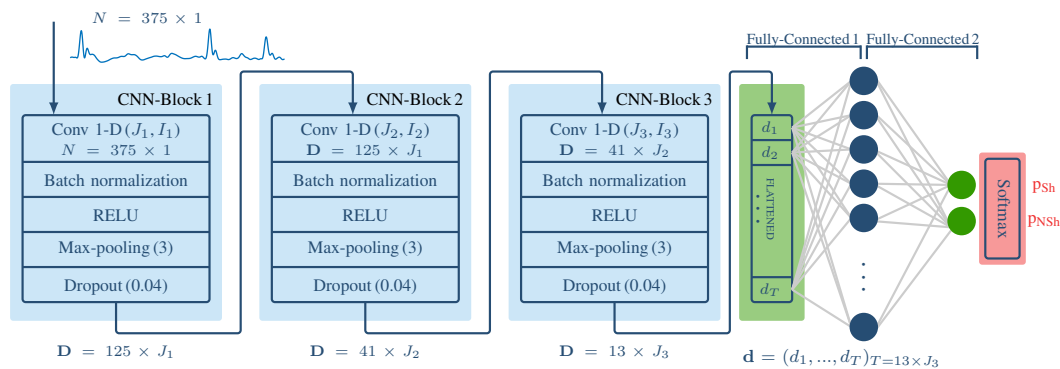


Figure 3. Architecture of the CNN-based shock/no-shock algorithm. It comprises two main stages: a CNN composed of three identical blocks and a classification stage based on two fully connected and a softmax layer.

3.2. Shock/No-Shock Decision Algorithms

Let us denote by $s_{b-1}(n, m)$ the output of block $b - 1$ (input to block b), where n is the time index and m the filter index. In the first block the input is $s_0(n, 1) = \hat{s}_{\text{ecg}}(n)$. The output of the Conv-1D layer at block b can be expressed as

$$c_b(n, m) = f \left(b_m + \sum_{\ell=1}^{J_{b-1}} \sum_{i=1}^{I_b} \omega_{\ell,i}^m s_{b-1}(n+i-1, \ell) \right) \quad (3)$$

where $\omega_{\ell,i}^m$ are the network weights (convolutional coefficients), and $f(x) = \max(x, 0)$ is the ReLU activation function that makes the network non-linear. The max-pooling layer selects the largest sample in blocks of K samples along the time index n to give the output of block b :

$$s_b(n, m) = \max \{ c_b(k, m) \}_{k=(n-1) \cdot K, \dots, n \cdot K} \quad (4)$$

Padding was applied before the convolutional and the max-pooling layers, so the only reduction of the dimensionality occurs at the max-pooling layers ($K = 3$). This means that the dimensions of the outputs at blocks $b = 1, 2, 3$ where $(125, J_1)$, $(41, J_2)$ and $(13, J_3)$, respectively and that the number of learnable parameters (ω, b) at block b where $J_b \times I_b + J_b$.

The dropout layer at the end of each block has a regularization effect, and is used only during training to avoid overfitting. It temporarily deactivates a randomly selected proportion of the

network's tunable parameters, and has been shown to improve performance by providing noisy inputs to the fully connected layers that help avoid overfitting [43].

The classification stage takes as input the flattened $13 \times J_3$ features and feeds them into two fully-connected layers. The first one is composed by 10 hidden units whereas the second one uses 2 neurons for the 2-class classification task. In the second fully-connected layer a softmax function is used to convert the output of the last two neurons into two values in the [0,1] range that can be interpreted as the likelihood that the 3 s window is shockable (p_{Sh}) or nonshockable (p_{NSh}).

The weights and biases of every layer were optimized using stochastic gradient descent with a momentum of 0.8. The initial learning rate was fixed to 0.02 and it was reduced by a factor of 0.8 at every epoch. The training data were fed into the CNN in batches of 256, and 20 epochs were used to train the networks [44]. During training data was augmented by splitting each 9 s training segment into overlapping 3 s windows with a linearly spaced start between 0 s and 6 s of the segment. To address class imbalance the augmented number of windows per segment during training was fixed to 100 for shockable and 40 for nonshockable rhythms, respectively. The binary cross entropy was used as loss function during network optimization (training):

$$L = \sum_i y_i \ln(p_{Sh_i}) + (1 - y_i) \ln(1 - p_{Sh_i}) \quad (5)$$

where $y_i = \{0 : NSh, 1 : Sh\}$ corresponds to the rhythm label of 3 s training window i .

Classical Machine Learning Shock/no-Shock Decision Algorithm for Baseline Comparison

The baseline machine learning shock/no-shock algorithm is a state of the art solution described in [25]. In short, the algorithm is based on multiresolution ECG analysis using the Stationary Wavelet Transform (SWT) for feature extraction, followed by a random forest (RF) classifier. The SWT decomposes the 3 s window into 7 sub-bands, and the denoised ECG is reconstructed using detail coefficients d_3 to d_7 , i.e. an analysis band of 0.98–31.25 Hz. The daubechies mother wavelet was used for the analysis as recommended in [26]. The selection of the mother wavelet was not a critical for this problem as shown in [26]. The denoised ECG, $s_{den}(n)$, and the detail coefficients d_3 – d_7 were used to obtain twenty five ECG features, selected using recursive feature elimination from a set of over 200 features (consult [25] for the details). The most relevant features were classical VF detection features like VFleak or x_4 [22,45] computed from s_{den} , and a rich set of features obtained from the detail coefficients $\{d_i\}_{i=3,\dots,7}$, such as: sample entropy (SampEn(d_i)), the mean and standard deviation of the absolute value of the signal ($\overline{|d_i|}$, $\sigma(|d_i|)$) and its slope ($\overline{|\Delta d_i|}$, $\sigma(|\Delta d_i|)$), and the Hjorth mobility (Hmb(d_i)) and complexity (Hmc(d_i)) indices [46]. A detailed description of the algorithm is found in [25], with a detailed bibliography for the computation of the features.

The parameters of the RF classifier were fixed to those recommended in [25], that is $B = 500$ trees, 5 predictors per split (standard in RF), and the minimum observations per leaf to 3 (to avoid growing excessively deep or overfit trees). To avoid class imbalance uniform priors were assigned and a cost function was introduced to penalize false shock classifications with a factor of 2.5 (similar to the shock/no-shock augmentation factor used in the CNN).

3.3. Evaluation

All the classification algorithms were trained/tested using 5-fold cross validation (CV). Folds were partitioned patient-wise to avoid training/test data leakage, and in a quasi-stratified way by ensuring that the shock/no-shock prevalences in all folds were at least 80% those of the whole dataset. The performance of the method was evaluated using the standard metrics for binary classification problems, taking the shockable class as positive and the nonshockable class as negative. For a 2×2 confusion matrix with true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) the performance metrics were

$$Se = \frac{TP}{TP + FN} \quad PPV = \frac{TP}{TP + FP} \quad (6)$$

$$Sp = \frac{TN}{TN + FP} \quad NPV = \frac{TN}{TN + FN} \quad (7)$$

$$Acc = \frac{TP+TN}{TP+FN+TN+FP} \quad BAC = \frac{1}{2}(Se + Sp) \quad (8)$$

The Balanced Accuracy (BAC) was used as target performance metric to ensure both shockable and nonshockable rhythms were accurately identified (as recommended by the AHA) despite the large class imbalance in the data.

4. Results

4.1. Parameters of the CNN Architecture

The effect of changing the main parameters of the CNN architecture was first studied taking the BAC as target performance metric (see Figure 4). Three parameters were studied: the number of blocks (B), the size of the filters (I), and the number of filters ($\mathbf{L} = (J_1, \dots, J_B)$). Four filter configurations were studied with decreasing number of filters (from dense to sparse): $\mathbf{L}_4 = (40, 30, 20, 10)$, $\mathbf{L}_3 = (32, 24, 16, 8)$, $\mathbf{L}_2 = (24, 18, 12, 6)$ and $\mathbf{L}_1 = (16, 12, 8, 4)$. The numbers in parentheses indicate the amount of filters from block 1 to block 4, so for architectures with 3 blocks and 2 blocks and \mathbf{L}_2 the number of filters would be (24,18,12) and (24,12), respectively.

The results of the analysis are shown in Figure 4, with the median BAC computed over the 5-fold CV partitions. The best classification results were obtained for 3 blocks. Adding a fourth block increases the complexity (number of trainable parameters) and slightly decreases the performance. Using only 2 blocks resulted in a large decrease in performance (over 1-point in BAC), or an overly simplistic model. The best results for a CNN with 3 blocks were obtained with a filter width of $I = 16$, and a filter configuration of $\mathbf{L} = (32, 24, 16)$. This was the CNN configuration adopted for the rest of the analyses.

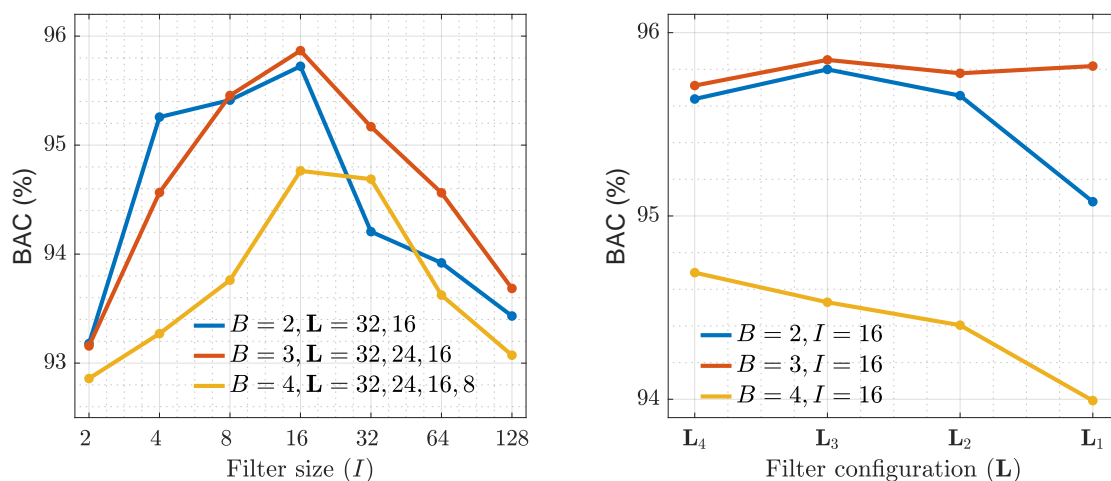


Figure 4. Performance of the CNN architecture for the configurable parameters of the network: the number of blocks (B), the filter size (I), and the filter configuration (\mathbf{L}). The left panel shows the effect of the filter size for networks with $\mathbf{L}_4 = (32, 24, 16, 8)$ filters. The right panel shows the effect of the filter configurations from dense (\mathbf{L}_4) to sparse (\mathbf{L}_1) for $I = 16$.

4.2. Comparison with the Baseline Machine Learning Model

The shock/no-shock decision algorithms using CNNs and the classical machine learning model were compared. Table 1 shows the results for all the performance metrics. The accuracies were compared using McNemar’s test in all 5-fold CV partitions, and the results were considered significant at the 95% level. The CNN model was significantly more accurate (median $p < 0.05$) than the baseline model. As shown in Table 1, the CNN model designed for 9 s improves the best baseline models in 0.6-points in BAC and Acc, and in both cases the algorithms presented balanced Se/Sp values because they were trained to avoid class imbalance. The predictivity is higher for the CNN solution, but the differences are only large for shockable rhythms (PPV) because shockable rhythms have a much lower prevalence in the dataset (1 to 5). The table shows the results for the 3 s windows (where CNN outperforms the baseline model), but also for the combination of three consecutive analyses (9 s). For short windows the algorithms do not meet the minimum 95% value recommended by the AHA for artifact free ECG, but combining diagnoses with a majority vote criterion considerably improves performance and brings both the CNN solution and the baseline algorithm above AHA specifications. The table also shows the shock/no-shock decision performance when the two subgroups of nonshockable rhythms were evaluated separately, AS and ORG rhythms. The results show that no-shock decisions were more inaccurate when the underlying rhythm was asystole. For 9 s segments the CNN architecture yielded results slightly above the AHA’s 95% Sp goal for AS, but the baseline model was marginally below.

Table 1. Performance metrics for the CNN and the baseline models. The results are shown as median and 90% confidence interval (CI).

Metric	3 s		9 s	
	CNN	Baseline	CNN	Baseline
Se	93.2 (92.2–94.0)	93.1 (92.6–93.6)	95.8 (94.6–96.8)	95.2 (94.7–95.7)
Sp	94.5 (94.1–94.9)	94.1 (93.9–94.3)	96.1 (95.8–96.5)	95.6 (95.2–95.9)
AS	93.1 (92.6–93.7)	92.5 (92.2–92.8)	95.4 (94.9–96.0)	94.5 (94.1–95.0)
ORG	95.6 (95.1–96.0)	95.3 (95.1–95.6)	96.8 (96.2–97.4)	96.4 (96.0–96.8)
BAC	93.8 (93.4–94.3)	93.6 (93.3–93.9)	96.0 (95.5–96.5)	95.4 (95.0–95.7)
Acc	94.3 (94.0–94.6)	93.9 (93.7–94.1)	96.1 (95.7–96.4)	95.5 (95.2–95.8)
PPV	78.5 (77.2–79.6)	77.2 (76.5–77.7)	84.3 (82.8–85.6)	82.2 (81.0–83.2)
NPV	98.5 (98.3–98.7)	98.5 (98.3–98.6)	99.1 (98.8–99.3)	98.9 (98.8–99.1)

4.3. Effect of the ECG Corruption Level on Classification

CPR artifacts during chest compressions present very different noise levels in the ECG depending on variables like the position of the hands relative to the pads and cables, pad placement, or environmental conditions [47,48]. These variables are difficult to control in a pre-hospital setting, but it is important to know what the observed corruption levels are, and how these corruption levels affect the shock/no-shock decisions. To estimate the signal-to-noise ratio (SNR) the underlying ECG was assumed to be stationary over the 15 s segments, and thus the power of the clean signal (P_{ecg}) was estimated in the 3 s interval without artifacts used to confirm the rhythm annotations. Then, CPR artifact estimated by the RLS filter was used to compute the power of the noise (P_{cpr}), and to obtain the SNR as:

$$SNR = 10 \cdot \log_{10} \left(\frac{P_{ecg}}{P_{cpr}} \right) \quad (\text{dB}) \quad (9)$$

The noise levels were divided into bins from very large corruption levels ($SNR < -18$ dB) to very low corruption levels ($SNR > 6$ dB). The distributions of noise levels and the classification results for the different noise conditions are shown in Figure 5 for shockable (a) and nonshockable (b)

rhythms. As expected the classification results improve as noise conditions improve, but noise affects the classification of shockable and nonshockable rhythms very differently. Nonshockable rhythms are detected with high specificity even in very noisy conditions, and the confidence in a nonshockable diagnosis (NPV) is high because the prevalence of nonshockable rhythms is 5/1 that of shockable rhythms. The sensitivity for shockable rhythms improves considerably as noise conditions improve, and was above the 90% value recommended by the AHA for $\text{SNR} > -10$ dB. However, the confidence on a shock diagnosis (PPV) is good only for $\text{SNR} > -6$ dB because of the lower prevalence of shockable rhythms. The SNR was significantly higher for nonshockable than for shockable rhythms ($p < 0.001$, Mann-Whitney U test), and in approximately 15% of shockable and nonshockable cases the noise level was negligible ($\text{SNR} > 25$ dB, see Figure 5). Although noise levels were lower in nonshockable rhythms, a high specificity was obtained regardless of the noise conditions. Even for the very noisy segments ($\text{SNR} < -12$ dB) the specificity was above 94%.

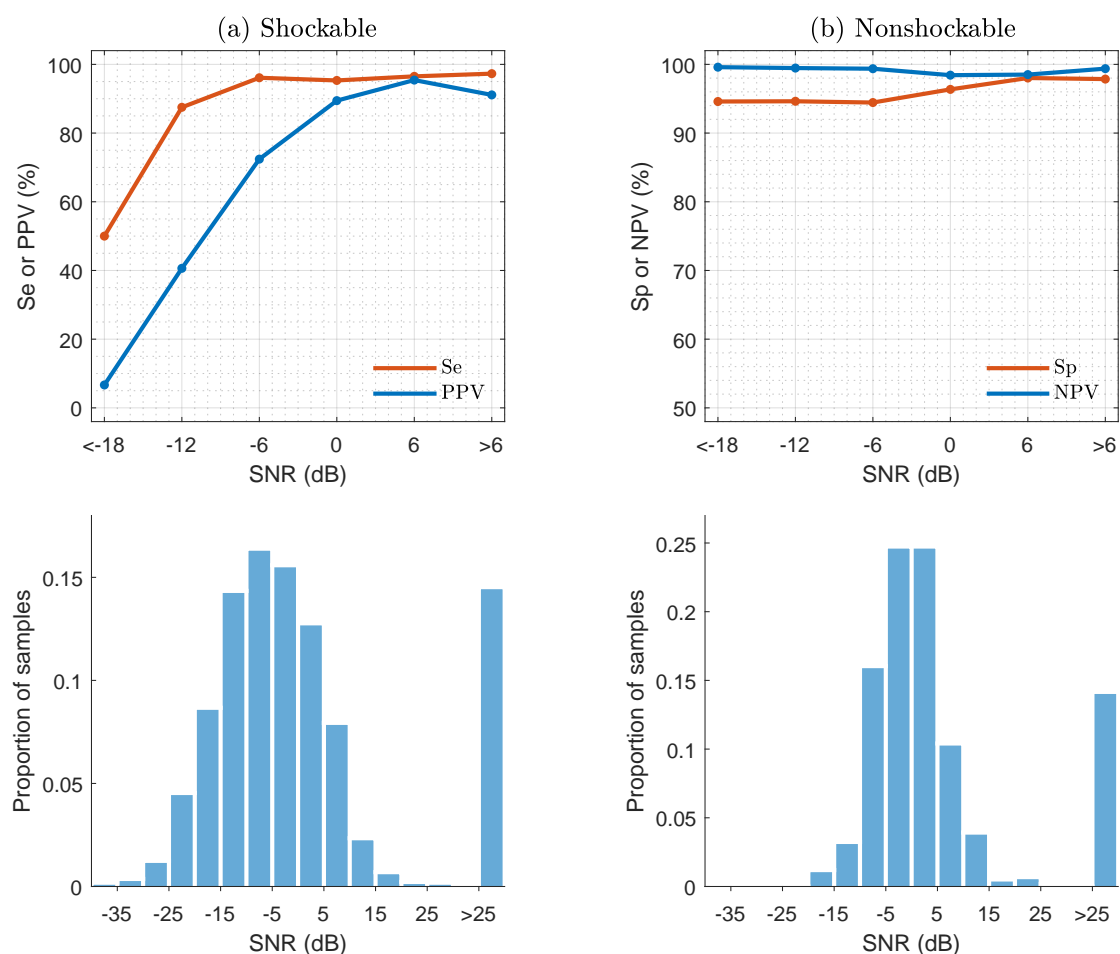


Figure 5. Median values of the performance metrics for shockable and nonshockable rhythms as a function of the SNR. The SNR levels were divided into 6 dB bins for the analysis from high (< -18 dB) to low (> 6 dB) corruption levels. The lower panels show the SNR distributions for shockable (a) and nonshockable rhythms (b).

4.4. Feature Extraction Using CNNs

For these experiments the 10 features at the output of the first fully connected layer were used as the features learned by the algorithm, these features will be named $\{f_i\}_{i=1,\dots,10}$. To evaluate feature extraction two experiments were conducted [36], and the results were compared to those obtained using the multiresolution features based on the SWT in the baseline model [25]. First, a dimensionality reduction experiment was conducted by projecting the feature space into a 2-D space using the

t-distributed stochastic neighbor embedding (t-SNE) algorithm [49]. The results were visually assessed, and are shown in Figure 6 for the f_i features and the handcrafted multiresolution features. The classes are shown in colors and the nonshockable rhythms are further divided into AS and ORG. As shown in the figure the CNN features produce better defined clusters than the handcrafted features in the 2D space. To numerically evaluate how the classes were clustered the Davies-Bouldin index (DBi) was computed to measure the separability of the clusters [50]. The experiment was repeated on 500 bootstrap replicas and the mean (standard deviation) BDi for the CNN and the handcrafted features were 2.28 (0.06) and 4.95 (0.17), respectively ($p < 0.05$, for the paired t-test) [51]. That is, the features learned by the CNN architecture resulted in a more efficient clustering of the classes, and thus to a better separability.

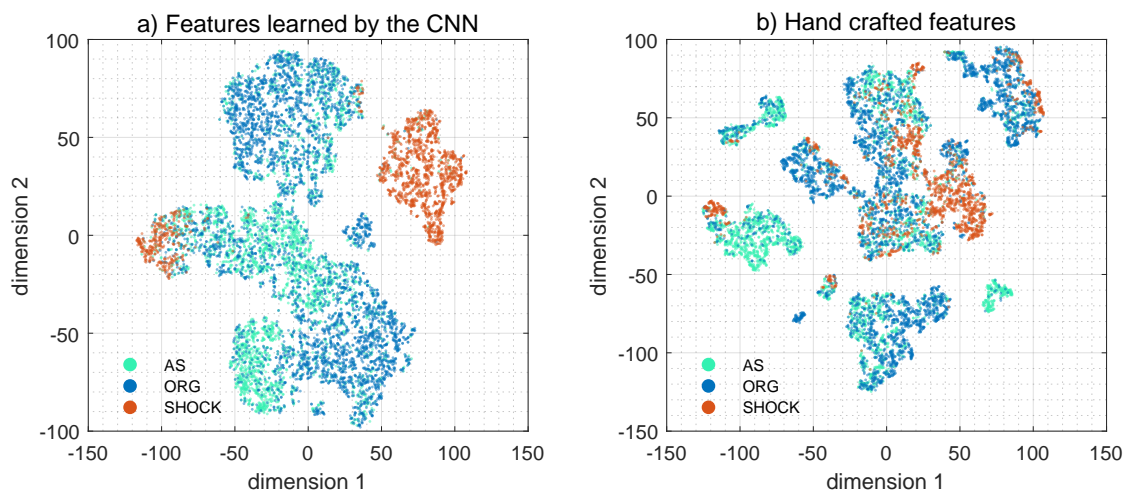


Figure 6. 2D map representation of the separability of the classes for the features learned by the CNN (a) and the handcrafted features (b). These maps were obtained using the t-SNE algorithm.

Second, the discriminating power of each feature was computed using the area under the receiver characteristics curve (AUC). The results were obtained over 500 bootstrap replicas to statistically characterize the AUCs and compare the AUC distributions for each feature (paired t-test). The results are shown in Table 2, which shows that the four top most discriminating features (f_6 , f_{10} , f_1 and f_5) had significantly higher AUCs ($p < 0.05$) than any of the handcrafted features. These results confirm the ability of the CNN to extract high quality discriminating features hidden in the signals.

Table 2. Mean (standard deviation) of the AUCs for the CNN features and the handcrafted features obtained using 500 bootstrap replicas of the data.

CNN Features		Handcrafted Features	
Feature	AUC	Feature	AUC
f_6	97.2 (1.1)	SampEn(d_3)	90.6 (2.0)
f_{10}	96.4 (1.6)	$\sigma(\Delta d_4)$	90.3 (1.7)
f_1	95.2 (2.6)	$\sigma(d_4)$	87.7 (1.8)
f_5	94.8 (2.3)	$\sigma(d_3)$	86.2 (2.3)
f_9	90.7 (3.7)	VFLLeak	85.9 (2.7)
f_3	81.2 (11.1)	SampEn(d_4)	84.8 (2.4)
f_8	75.2 (10.6)	$ \Delta d_3 $	84.6 (2.8)
f_4	73.9 (8.6)	x4	82.5 (3.6)
f_7	66.9 (6.2)	$\sigma(s_{den})$	82.4 (2.0)
f_2	59.3 (17.1)	SampEn(d_6)	80.6 (2.7)

4.5. Mixed Architectures

To further improve the BAC and accuracy of the CNN model three mixed architectures were also explored. First, the architecture of Figure 3 in which the softmax layer was replaced by a RF classifier to combine the best feature extraction (CNN) and classification (RF) of the the algorithms in Table 1, this solution was named CNN + RF . Second, a RF classifier fed with 25 handcrafted features and the 10 f_i features was tested to see if handcrafted features added information to the CNN features, this was named All-Features. Finally, a basic stacking solution [52] in which the outputs of the CNN+RF (based on f_i) and the baseline model (handcrafted features) were used to form a majority vote (6 analyses, two per window), this solution was called Stacked. The results for 9 s segments are shown in Table 3, which shows that by using more elaborate solutions the BAC and Acc could be further improved in 0.4 and 0.5-points, respectively, either using all features or stacking the classifiers.

Table 3. Performance metrics for 9 s segments of the mixed solutions. The results are shown as median and 90% confidence interval (CI).

Metric	CNN	Mixed Classification Solutions		
		CNN + RF	All-Features	Stacked
Se	95.8 (94.6–96.8)	95.3 (93.9–96.2)	95.6 (94.6–96.4)	96.1 (95.1–96.8)
Sp	96.1 (95.8–96.5)	96.7 (96.3–97.1)	96.8 (96.5–97.1)	96.7 (96.3–97.1)
AS	95.4 (94.9–96.0)	95.9 (95.4–96.5)	96.1 (95.6–96.6)	95.9 (95.3–96.4)
ORG	96.8 (96.2–97.4)	97.2 (96.7–97.7)	97.3 (96.9–97.7)	97.4 (96.9–97.9)
BAC	96.0 (95.5–96.5)	96.0 (95.3–96.5)	96.2 (95.7–96.7)	96.4 (95.9–96.8)
Acc	96.1 (95.7–96.4)	96.4 (96.0–96.7)	96.6 (96.3–96.9)	96.6 (96.3–96.9)
PPV	84.3 (82.8–85.6)	86.0 (84.6–87.4)	86.5 (85.3–87.8)	86.3 (84.8–87.5)
NPV	99.1 (98.8–99.3)	99.0 (98.7–99.2)	99.0 (98.8–99.2)	99.1 (98.9–99.3)

4.6. Analysis of Classification Errors

To conclude the analyses, the classification errors for the CNN based algorithm were identified. Some typical patterns leading to errors are shown in Figure 7. Most of the false positives are caused by the inability of the RLS filter to properly remove artifacts, leading to very disorganized filtering residuals that resemble a VF. Most false negatives occur at low SNR levels with compression rates around 100 min^{-1} . In these cases the filtered ECG still shows an organized activity locked to the compression frequency, incompatible with fast ventricular arrhythmia and thus classified as non-shockable. Interestingly, these errors can be related to the clustering analysis of Section 4.4. Most errors cluster around borderline AS/VF rhythms which appear in the center-left region of the 2D t-SNE map (Figure 6), and ORG/VF rhythms in a much lower proportion in the top-center.

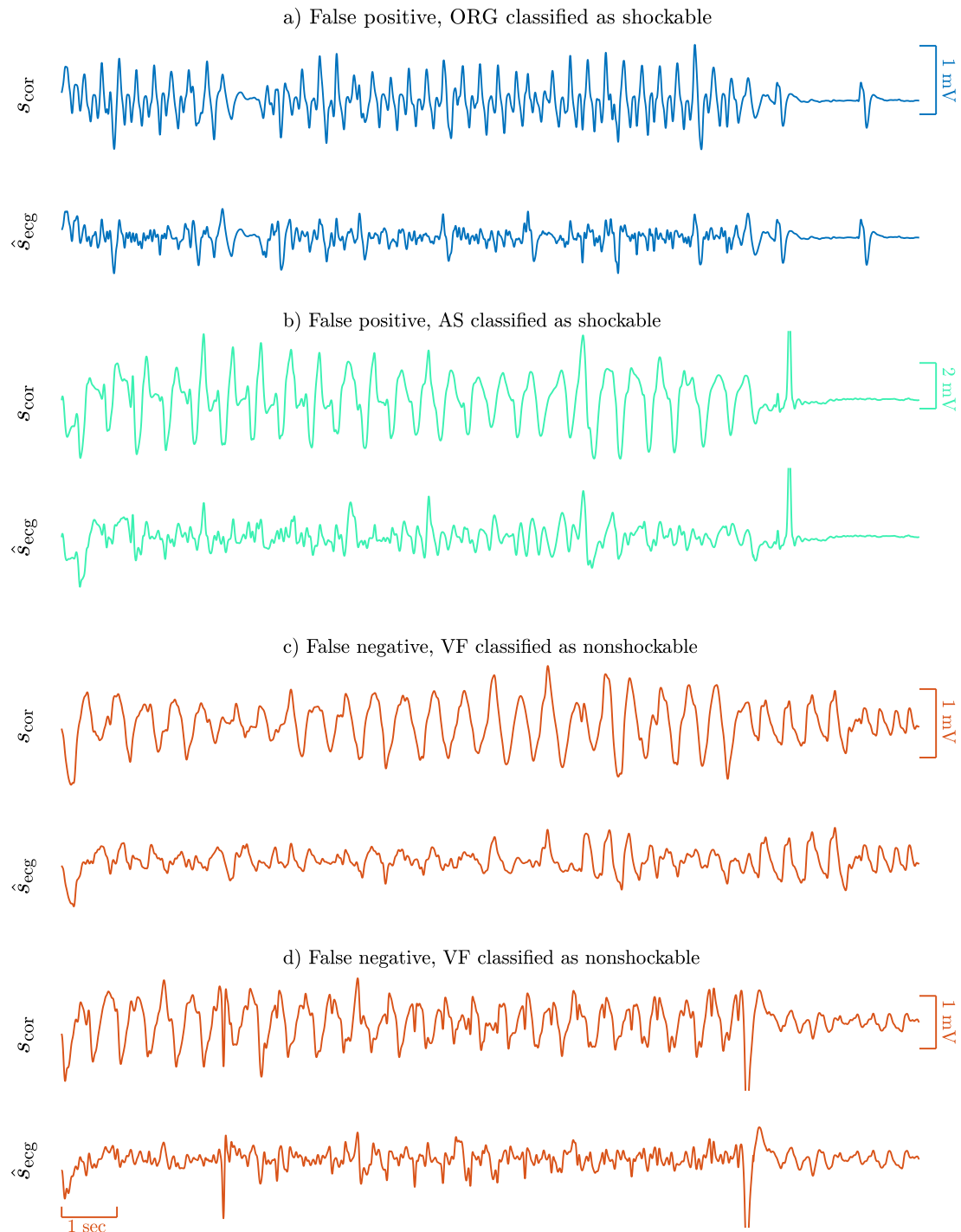


Figure 7. Examples of classification errors. The false positive examples (a,b) correspond to nonshockable rhythms classified as shockable (ORG panel a and AS panel b). The false negative examples (c,d) are shockable rhythms classified as nonshockable, and are shown in orange.

5. Discussion

This is, to the best of our knowledge, the first study that uses deep neural network models to discriminate between shockable and nonshockable rhythms during CPR. This algorithm consists of an adaptive RLS filter to remove CPR artifacts followed by a CNN to classify the filtered ECG. The algorithm designed for 9 s improves the performance of the classical machine learning algorithms by 0.6 points in BAC and Acc. This improvement is large considering that the best classical machine learning algorithms had accuracies over 95% and that they are based on more than 20 years of expert

knowledge on ECG feature engineering. Moreover, mixed solutions, obtained by either stacking classifiers or mixing handcrafted and CNN features, could yield further improvements in BAC and Acc, as shown by the preliminary experiments of Section 4.5.

One of the advantages of deep learning solutions is the capacity of the algorithms to learn discriminating features exploiting all information hidden in the ECG. This avoids the time-consuming feature extraction processes and, most importantly, improves the quality of the extracted features. The latter is well reflected by the AUCs on Table 2. Four of the ten features extracted by the deep learning architecture show a higher discrimination capacity than SampEn(d_3), which is the best handcrafted feature for shock/no-shock decisions during CPR in the available literature [25,26].

Two factors were key to improve the performance of the CNN based methods from the preliminary results communicated previously in [53]. First, the design and optimization of the parameters of the CNN to obtain a better model for classification. Second, increasing the size of the database by adding 1186 new annotated samples (a 55% increase in dataset size). These led to 0.5-points and 0.3-points increases in BAC and Acc respectively, of which 0.4-points and 0.1-points are attributable to the larger dataset. And there is further room for improvement from combining the knowledge gained from deep learning and handcrafted ECG feature extraction, basic examples are shown in Table 3 which added an extra 0.5-points in Acc. The performance of deep learning solutions improves as they are exposed to more data, whereas the accuracy of classical machine learning algorithms stagnate past a given sample size. The model presented in this study overfits when more than 3 CNN blocks are used (Figure 4) since from then on the number of trainable parameters is too large for the size of the available dataset. Adding more data would help to develop deeper networks and thus to the extraction of more sophisticated features. There is therefore room to improve the deep learning models for rhythm analysis during CPR, as more and more data is recorded every day and made available in centralized repositories. In research on OHCA, the Resuscitation Outcome Consortium (ROC) network provides the largest OHCA data repository, which includes recordings of eleven regional clinical centers. However, labeled OHCA data are scarce, and obtaining quality controlled rhythm annotations from clinicians is expensive and time consuming. As an alternative, semi-supervised learning could be an efficient way to augment training data and obtain better deep learning models in the future.

As Figure 6 shows, CNN features provide more separate clusters than the handcrafted features for the shock/no-shock classes. Moreover, the deep learning model shows a quite high separability between the features corresponding to AS, OR and shockable rhythms. Therefore, in the future CNN models could improve the accuracy of classical machine learning-based multiclass rhythm classifiers. These classifiers have been demonstrated for clean [24,52] and artifacted ECGs [25], and are multilabel classification algorithms that classify the ECG into the 5 OHCA rhythm types. These algorithms are important for research to analyze large sets of OHCA data [24], and could also help clinicians during OHCA treatment as clinical support tools. The best OHCA multiclass algorithms have unweighted mean sensitivities of 78% for clean ECG [24], and of 72% if the analysis is done during CPR [25]. There is therefore margin for improvement using methods based on deep learning if sufficiently large quality controlled annotated datasets become available.

6. Conclusions

This paper introduces the first shock/no-shock decision algorithm during CPR based on deep learning methods. This solution improves the accuracy of the best classical machine learning models based on handcrafted features, and is able to give a shock/no-shock diagnosis compliant with AHA recommendations for shockable and nonshockable rhythms. Moreover, deep learning algorithms have room for improvement if larger annotated datasets become available allowing the design of deeper networks. This may lead to the first practical solutions for rhythm analysis during CPR, eliminating the no-flow intervals for rhythm analysis and contributing to improve OHCA survival rates.

Author Contributions: Conceptualization, J.K.-J., L.W., I.I., U.I., E.A. and T.E.; methodology, I.I. and U.I.; software, I.I.; validation, U.I.; formal analysis, I.I. and U.I.; investigation, I.I., U.I., E.A., T.E., J.K.-J. and L.W.; resources, J.K.-J. and L.W.; data curation, U.I., E.A., T.E. and I.I.; writing—original draft preparation, I.I. and U.I.; writing—review and editing, I.I., U.I., E.A., T.E., J.K.-J. and L.W.; visualization, I.I. and U.I.; supervision, U.I. and E.A.; project administration, U.I. and E.A.; funding acquisition, U.I. and E.A. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Spanish Ministerio de Ciencia, Innovacion y Universidades through grant RTI2018-101475-BI00, jointly with the Fondo Europeo de Desarrollo Regional (FEDER), and by the Basque Government through grants IT1229-19 and PRE-2019-2-0066.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CPR	cardiopulmonary resuscitation
CNN	convolutional neural network
RLS	recursive least squares
OHCA	out of hospital cardiac arrest
ECG	electrocardiogram
VF	ventricular fibrillation
VT	ventricular tachycardia
ORG	organized
AS	asystole
AHA	American Heart Association
CD	compression depth
TI	thoracic impedance
LMS	least mean squares
SVM	support vector machine
RF	random forest
SWT	stationary wavelet transform
TP	true positive
TN	true negative
FP	false positive
FN	false negative
Se	sensitivity
Sp	specificity
Acc	accuracy
BAC	balanced accuracy
PPV	positive predictive value
NPV	negative predictive value
SNR	signal-to-noise ratio
t-SNE	t-distributed stochastic neighbour embedding
DBi	Davies-Bouldin index
AUC	area under the receiver characteristics curve
ROC	Resuscitation Outcome Consortium

References

1. Berdowski, J.; Berg, R.A.; Tijssen, J.G.; Koster, R.W. Global incidences of out-of-hospital cardiac arrest and survival rates: Systematic review of 67 prospective studies. *Resuscitation* **2010**, *81*, 1479–1487. [[CrossRef](#)] [[PubMed](#)]
2. Daya, M.R.; Schmicker, R.H.; Zive, D.M.; Rea, T.D.; Nichol, G.; Buick, J.E.; Brooks, S.; Christenson, J.; MacPhee, R.; Craig, A.; et al. Out-of-hospital cardiac arrest survival improving over time: Results from the Resuscitation Outcomes Consortium (ROC). *Resuscitation* **2015**, *91*, 108–115. [[CrossRef](#)] [[PubMed](#)]

3. Perkins, G.; Handley, A.; Koster, R.; Castrén, M.; Smyth, M.; Olasveengen, T.; Monsieurs, K.; Raffay, V.; Gärsner, J.; Wenzel, V.; et al. European Resuscitation Council Guidelines for Resuscitation 2015: Section 2. Adult basic life support and automated external defibrillation. *Resuscitation* **2015**, *95*, 81–99. [[CrossRef](#)]
4. Takata, T.S.; Page, R.L.; Joglar, J.A. Automated external defibrillators: Technical considerations and clinical promise. *Ann. Intern. Med.* **2001**, *135*, 990–998. [[CrossRef](#)] [[PubMed](#)]
5. Kerber, R.E.; Becker, L.B.; Bourland, J.D.; Cummins, R.O.; Hallstrom, A.P.; Michos, M.B.; Nichol, G.; Ornato, J.P.; Thies, W.H.; White, R.D.; et al. 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* **1997**, *95*, 1677–1682.
6. Snyder, D.; Morgan, C. Wide variation in cardiopulmonary resuscitation interruption intervals among commercially available automated external defibrillators may affect survival despite high defibrillation efficacy. *Crit. Care Med.* **2004**, *32*, S421–S424. [[CrossRef](#)]
7. Van Alem, A.P.; Sanou, B.T.; Koster, R.W. Interruption of cardiopulmonary resuscitation with the use of the automated external defibrillator in out-of-hospital cardiac arrest. *Ann. Emerg. Med.* **2003**, *42*, 449–457. [[CrossRef](#)]
8. Li, Y.; Tang, W. Techniques for artefact filtering from chest compression corrupted ECG signals: Good, but not enough. *Resuscitation* **2009**, *80*, 1219–1220. [[CrossRef](#)]
9. Sato, Y.; Weil, M.H.; Sun, S.; Tang, W.; Xie, J.; Noc, M.; Bisera, J. Adverse effects of interrupting precordial compression during cardiopulmonary resuscitation. *Crit. Care Med.* **1997**, *25*, 733–736. [[CrossRef](#)]
10. Eftestøl, T.; Sunde, K.; Steen, P.A. Effects of interrupting precordial compressions on the calculated probability of defibrillation success during out-of-hospital cardiac arrest. *Circulation* **2002**, *105*, 2270–2273. [[CrossRef](#)]
11. Edelson, D.P.; Abella, B.S.; Kramer-Johansen, J.; Wik, L.; Myklebust, H.; Barry, A.M.; Merchant, R.M.; Hoek, T.L.V.; Steen, P.A.; Becker, L.B. Effects of compression depth and pre-shock pauses predict defibrillation failure during cardiac arrest. *Resuscitation* **2006**, *71*, 137–145. [[CrossRef](#)] [[PubMed](#)]
12. Ruiz de Gauna, S.; Irusta, U.; Ruiz, J.; Ayala, U.; Aramendi, E.; Eftestøl, T. Rhythm analysis during cardiopulmonary resuscitation: Past, present, and future. *BioMed Res. Int.* **2014**, *2014*. [[CrossRef](#)] [[PubMed](#)]
13. Werther, T.; Klotz, A.; Granegger, M.; Baubin, M.; Feichtinger, H.G.; Amann, A.; Gilly, H. Strong corruption of electrocardiograms caused by cardiopulmonary resuscitation reduces efficiency of two-channel methods for removing motion artefacts in non-shockable rhythms. *Resuscitation* **2009**, *80*, 1301–1307. [[CrossRef](#)] [[PubMed](#)]
14. 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*, 1229–1233. [[CrossRef](#)] [[PubMed](#)]
15. Ayala, U.; Eftestøl, T.; Alonso, E.; Irusta, U.; Aramendi, E.; Wali, S.; Kramer-Johansen, J. Automatic detection of chest compressions for the assessment of CPR-quality parameters. *Resuscitation* **2014**, *85*, 957–963. [[CrossRef](#)]
16. Aase, S.O.; Eftestøl, T.; Husoy, J.; Sunde, K.; Steen, P.A. CPR artifact removal from human ECG using optimal multichannel filtering. *IEEE Trans. Biomed. Eng.* **2000**, *47*, 1440–1449. [[CrossRef](#)]
17. Eilevstjønn, J.; Eftestøl, T.; Aase, S.O.; Myklebust, H.; Husøy, J.H.; Steen, P.A. Feasibility of shock advice analysis during CPR through removal of CPR artefacts from the human ECG. *Resuscitation* **2004**, *61*, 131–141. [[CrossRef](#)]
18. Berger, R.D.; Palazzolo, J.; Halperin, H. Rhythm discrimination during uninterrupted CPR using motion artifact reduction system. *Resuscitation* **2007**, *75*, 145–152. [[CrossRef](#)]
19. Irusta, U.; Ruiz, J.; de Gauna, S.R.; 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 Trans. Biomed. Eng.* **2009**, *56*, 1052–1062. [[CrossRef](#)]
20. Rheinberger, K.; Steinberger, T.; Unterkofler, K.; Baubin, M.; Klotz, A.; Amann, A. Removal of CPR artifacts from the ventricular fibrillation ECG by adaptive regression on lagged reference signals. *IEEE Trans. Biomed. Eng.* **2007**, *55*, 130–137. [[CrossRef](#)]

21. Ruiz, J.; Irusta, U.; de Gauna, S.R.; Eftestøl, T. Cardiopulmonary resuscitation artefact suppression using a Kalman filter and the frequency of chest compressions as the reference signal. *Resuscitation* **2010**, *81*, 1087–1094. [[CrossRef](#)] [[PubMed](#)]
22. Ayala, U.; Irusta, U.; Ruiz, J.; Eftestøl, T.; Kramer-Johansen, J.; Alonso-Atienza, F.; Alonso, E.; González-Otero, D. A reliable method for rhythm analysis during cardiopulmonary resuscitation. *BioMed Res. Int.* **2014**, *2014*, 872470. [[CrossRef](#)] [[PubMed](#)]
23. Figuera, C.; Irusta, U.; Morgado, E.; Aramendi, E.; Ayala, U.; Wik, L.; Kramer-Johansen, J.; Eftestøl, T.; Alonso-Atienza, F. Machine learning techniques for the detection of shockable rhythms in automated external defibrillators. *PLOS ONE* **2016**, *11*, e0159654. [[CrossRef](#)] [[PubMed](#)]
24. Rad, A.B.; Eftestøl, T.; Egan, K.; Irusta, U.; Kvaløy, J.T.; Kramer-Johansen, J.; Wik, L.; Katsaggelos, A.K. ECG-based classification of resuscitation cardiac rhythms for retrospective data analysis. *IEEE Trans. Biomed. Eng.* **2017**, *64*, 2411–2418. [[CrossRef](#)] [[PubMed](#)]
25. Isasi, I.; Irusta, U.; Rad, A.B.; Aramendi, E.; Zabihi, M.; Eftestøl, T.; Kramer-Johansen, J.; Wik, L. Automatic Cardiac Rhythm Classification With Concurrent Manual Chest Compressions. *IEEE Access* **2019**, *7*, 115147–115159. [[CrossRef](#)]
26. Isasi, I.; Irusta, U.; Elola, A.; Aramendi, E.; Ayala, U.; Alonso, E.; Kramer-Johansen, J.; Eftestøl, T. A machine learning shock decision algorithm for use during piston-driven chest compressions. *IEEE Trans. Biomed. Eng.* **2018**, *66*, 1752–1760. [[CrossRef](#)]
27. Rim, B.; Sung, N.J.; Min, S.; Hong, M. Deep Learning in Physiological Signal Data: A Survey. *Sensors* **2020**, *20*, 969. [[CrossRef](#)]
28. Hong, S.; Zhou, Y.; Shang, J.; Xiao, C.; Sun, J. Opportunities and Challenges in Deep Learning Methods on Electrocardiogram Data: A Systematic Review. *arXiv* **2019**, arXiv:2001.01550.
29. Kiranyaz, S.; Ince, T.; Gabbouj, M. Real-time patient-specific ECG classification by 1-D convolutional neural networks. *IEEE Trans. Biomed. Eng.* **2015**, *63*, 664–675. [[CrossRef](#)]
30. Acharya, U.R.; Fujita, H.; Lih, O.S.; Hagiwara, Y.; Tan, J.H.; Adam, M. Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network. *Inf. Sci.* **2017**, *405*, 81–90. [[CrossRef](#)]
31. Al Rahhal, M.M.; Bazi, Y.; Al Zuair, M.; Othman, E.; BenJdira, B. Convolutional neural networks for electrocardiogram classification. *J. Med. Biol. Eng.* **2018**, *38*, 1014–1025.
32. Xia, Y.; Wulan, N.; Wang, K.; Zhang, H. Detecting atrial fibrillation by deep convolutional neural networks. *Comput. Biol. Med.* **2018**, *93*, 84–92. [[CrossRef](#)] [[PubMed](#)]
33. Hannun, A.Y.; Rajpurkar, P.; Haghpanahi, M.; Tison, G.H.; Bourn, C.; Turakhia, M.P.; Ng, A.Y. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat. Med.* **2019**, *25*, 65. [[CrossRef](#)] [[PubMed](#)]
34. Pourbabaee, B.; Roshtkhari, M.J.; Khorasani, K. Deep convolutional neural networks and learning ECG features for screening paroxysmal atrial fibrillation patients. *IEEE Trans. Syst. Man. Cybern. Syst.* **2018**, *48*, 2095–2104. [[CrossRef](#)]
35. Andreotti, F.; Carr, O.; Pimentel, M.A.; Mahdi, A.; De Vos, M. Comparing feature-based classifiers and convolutional neural networks to detect arrhythmia from short segments of ECG. In Proceedings of the 2017 Computing in Cardiology (CinC), Rennes, France, 24–27 September 2017; pp. 1–4.
36. Picon, A.; Irusta, U.; Álvarez-Gila, A.; Aramendi, E.; Alonso-Atienza, F.; Figuera, C.; Ayala, U.; Garrote, E.; Wik, L.; Kramer-Johansen, J.; et al. Mixed convolutional and long short-term memory network for the detection of lethal ventricular arrhythmia. *PLoS ONE* **2019**, *14*, e0216756. [[CrossRef](#)]
37. Elola, A.; Aramendi, E.; Irusta, U.; Picón, A.; Alonso, E.; Owens, P.; Idris, A. Deep neural networks for ECG-based pulse detection during out-of-hospital cardiac arrest. *Entropy* **2019**, *21*, 305. [[CrossRef](#)]
38. Wik, L.; Kramer-Johansen, J.; Myklebust, H.; Sørebo, H.; Svensson, L.; Fellows, B.; Steen, P.A. Quality of cardiopulmonary resuscitation during out-of-hospital cardiac arrest. *JAMA* **2005**, *293*, 299–304. [[CrossRef](#)]
39. Isasi, I.; Irusta, U.; Aramendi, E.; Ayala, U.; Alonso, E.; Kramer-Johansen, J.; Eftestøl, T. A multistage algorithm for ECG rhythm analysis during piston-driven mechanical chest compressions. *IEEE Trans. Biomed. Eng.* **2018**, *66*, 263–272. [[CrossRef](#)]
40. Xiao, Y.; Ma, L.; Ward, R.K. Fast RLS Fourier analyzers capable of accommodating frequency mismatch. *Signal Process.* **2007**, *87*, 2197–2212. [[CrossRef](#)]

41. Atkins, D.L.; Scott, W.A.; Blaurock, A.D.; Law, I.H.; Dick II, M.; Geheb, F.; Sobh, J.; Brewer, J.E. Sensitivity and specificity of an automated external defibrillator algorithm designed for pediatric patients. *Resuscitation* **2008**, *76*, 168–174.
42. Kwok, H.; Coult, J.; Drton, M.; Rea, T.D.; Sherman, L. Adaptive rhythm sequencing: A method for dynamic rhythm classification during CPR. *Resuscitation* **2015**, *91*, 26–31. [[CrossRef](#)] [[PubMed](#)]
43. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
44. Sutskever, I.; Martens, J.; Dahl, G.; Hinton, G. On the importance of initialization and momentum in deep learning. In Proceedings of the International conference on machine learning, Atlanta, GA, USA, 16–21 June 2013; pp. 1139–1147.
45. Kuo, S. Computer detection of ventricular fibrillation. *Proc. Comput. Cardiol. IEEE Comput. Soc.* **1978**, 347–349.
46. Gonzalez, L.; Walker, K.; Challa, S.; Bent, B. Monitoring a skipped heartbeat: A real-time premature ventricular contraction (pvc) monitor. In Proceedings of the 2016 IEEE Virtual Conference on Applications of Commercial Sensors (VCACS), Raleigh, NC, USA, 15 June 2016–15 January 2017; pp. 1–7.
47. Langhelle, A.; Eftestøl, T.; Myklebust, H.; Eriksen, M.; Holten, B.T.; Steen, P.A. Reducing CPR artefacts in ventricular fibrillation in vitro. *Resuscitation* **2001**, *48*, 279–291. [[CrossRef](#)]
48. Fitzgibbon, E.; Berger, R.; Tsitlik, J.; Halperin, H.R. Determination of the noise source in the electrocardiogram during cardiopulmonary resuscitation. *Crit. Care Med.* **2002**, *30*, S148–S153.
49. Maaten, L.v.d.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
50. Davies, D. DW 1979. A Cluster Separation Measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **1979**, *PAMI-1*, 224–227.
51. Vesanto, J.; Himberg, J.; Alhoniemi, E.; Parhankangas, J. SOM toolbox for Matlab 5. *Hels. Univ. Technol.* **2000**, 216, 57.
52. Rad, A.B.; Eftestøl, T.; Irusta, U.; Kvaløy, J.T.; Wik, L.; Kramer-Johansen, J.; Katsaggelos, A.K.; Engan, K. An automatic system for the comprehensive retrospective analysis of cardiac rhythms in resuscitation episodes. *Resuscitation* **2018**, *122*, 6–12.
53. Isasi Liñero, I.; López de Larruzea, L.; Irusta Zarandona, U.; Aramendi Ecenarro, E. Diagnóstico del ritmo cardíaco durante la resucitación cardiopulmonar mediante técnicas de aprendizaje profundo. *Actas del XXXVII Congreso Anual de la Sociedad Española de Ingeniería Biomédica* **2019**, *37*, 37–40.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).