Journal of the American Heart Association

ORIGINAL RESEARCH

Deep Neural Network Approach for Continuous ECG-Based Automated External Defibrillator Shock Advisory System During Cardiopulmonary Resuscitation

Shirin Hajeb-M, PhD; Alicia Cascella, MS; Matt Valentine, MS; K. H. Chon, PhD

BACKGROUND: Because chest compressions induce artifacts in the ECG, current automated external defibrillators instruct the user to stop cardiopulmonary resuscitation (CPR) while an automated rhythm analysis is performed. It has been shown that minimizing interruptions in CPR increases the chance of survival.

METHODS AND RESULTS: The objective of this study was to apply a deep-learning algorithm using convolutional layers, residual networks, and bidirectional long short-term memory method to classify shockable versus nonshockable rhythms in the presence and absence of CPR artifact. Forty subjects' data from Physionet with 1131 shockable and 2741 nonshockable samples contaminated with 43 different CPR artifacts that were acquired from a commercial automated external defibrillator during asystole were used. We had separate data as train and test sets. Using our deep neural network model, the sensitivity and specificity of the shock versus no-shock decision for the entire data set over the 4-fold cross-validation sets were 95.21% and 86.03%, respectively. This result was based on the training and testing of the model using ECG data in both the presence and the absence of CPR artifact. For ECG without CPR artifact, the sensitivity was 99.04% and the specificity was 95.2%. A sensitivity of 94.21% and a specificity of 86.14% were obtained for ECG with CPR artifact. In addition to 4-fold cross-validation sets, we also examined leave-one-subject-out validation. The sensitivity and specificity for the case of leave-one-subject-out validation were 92.71% and 97.6%, respectively.

CONCLUSIONS: The proposed trained model can make shock versus nonshock decision in automated external defibrillators, regardless of CPR status. The results meet the American Heart Association's sensitivity requirement (>90%).

Key Words: automated external defibrillator ■ cardiopulmonary resuscitation–contaminated ECG ■ deep neural network ■ long short-term memory ■ out-of-hospital cardiac arrest ■ residual block ■ shock advisory system

Immediate high-quality cardiopulmonary resuscitation (CPR) can double or triple the survival rate in the event of an out-of-hospital cardiac arrest. During a cardiac arrest, an intervention that can restart regular heart beats is electrical defibrillation (controlled electrical shock). Early defibrillation along with continuous CPR can lead to a return of spontaneous circulation while minimizing end-organ damage.[1,2] The decision of whether a heart rhythm is shockable or nonshockable depends on the near real-time analysis of the ECG. In parallel, to minimize depriving oxygen to the brain, CPR delivery needs to be continuously performed. However, accurately deciphering the rhythms derived from ECG during CPR is challenging because chest compressions induce large artifacts in ECG waveforms. To overcome this, the current automated

Correspondence to: Shirin Hajeb-M, PhD, Biomedical Engineering Department, A.B. Bronwell Building, Room 209, University of Connecticut, Storrs, CT 06269-3247. E-mail: shirin.hajeb@uconn.edu

For Sources of Funding and Disclosures, see page 10.

JAHA is available at: www.ahajournals.org/journal/jaha

CLINICAL PERSPECTIVE

What Is New?

- In the present study, a new algorithm for making reliable shock/no-shock decision in presence of cardiopulmonary resuscitation was developed.
- Unlike most of the previous approaches, no additional reference signal is required.
- The approach successfully made accurate shock decision in the presence of cardiopulmonary resuscitation, with a sensitivity of 94% and a specificity of 86%, meeting the sensitivity requirement of the American Heart Association.
- To the best of our knowledge, this is one of the first studies that has used deep learning to make such a classification with the goal of removing pauses during cardiopulmonary resuscitation.

What Are the Clinical Implications?

- This approach may allow automated external defibrillators to decipher the ECG rhythms and make accurate decisions about cardioversion/defibrillation without the need for stopping cardiopulmonary resuscitation.

Nonstandard Abbreviations and Acronyms

BiLSTM	bidirectional long short-term memory
CNN	convolutional neural network
CUDB	Creighton University tachyarrhythmia database
DNN	deep neural network
SDDB	sudden cardiac death Holter database
VFDB	malignant ventricular fibrillation database
MIT-BIH	Massachusetts institude of technology-Beth Israel hospital

external defibrillators (AEDs) instruct the user not to deliver CPR during the rhythm analysis period.[3] According to Waalewijn et al,[4] survival probability from shockable cardiac arrhythmias, such as ventricular fibrillation, decreases by 10% to 12% for every minute the electrical shock is delayed. However, it was determined that if CPR could be continuously performed, the survival rate would decrease only 3% to 4% per minute while electrical shock was delayed.[4,5] Therefore, an accurate shock advisory system that could handle CPR artifacts in ECG waveforms would be of great value.

Recent studies to mitigate CPR artifact in the analysis of arrhythmia classification can be categorized into 2 different approaches.

For the first category, the removal of CPR artifacts was attempted using a reference signal (such as compression depth, chest pressure, chest displacement, chest acceleration, or thoracic impedance), which is highly correlated with CPR artifact.[6,7] Using the selected reference signal, various adaptive filtering methods, such as Kalman filters,[8,9] least mean square filters,[10–12] Gabor filters,[13] or recursive least square[14] methods, have been used to suppress the CPR artifact before analyzing the ECG rhythms.

Although most AEDs do not have the hardware capability to record reference signals, unfortunately only a few algorithms have been developed without the need for reference signals.[15–18] These methods do not always work well when the frequencies of ECG rhythm overlap with that of the CPR artifact. Furthermore, these filtering approaches need to be combined with a shock advisory algorithm.

The second category of dealing with CPR artifact involves direct analysis of the corrupted ECG during CPR.[19,20] However, their accuracies did not meet the American Heart Association's requirements,[21] as the specificity was lower than required.

Traditional machine learning approaches have been also used in some recent works with the aim of making shock versus no-shock decisions in AEDs.[14,22] Although these methods do not require certain threshold values to make rhythm determinations, they still require a reference signal to suppress the impact of CPR artifacts. In addition, their performance was not satisfactory, albeit better than some of the filtering approaches.

Although deep learning approaches have engendered interest for arrhythmia detection,[23,24] they have not been explored for use in AEDs. Hence, the aim of this study was to use a deep learning approach to make a shock versus no-shock rhythm classification with no prior filtering requirements, while CPR is being performed. The proposed approach determines whether or not shocks are needed on the basis of analyzing only 8 seconds of ECG signal in either the presence or the absence of CPR. No additional reference signal is required.

METHODS

Data Preparation

ECG signals were derived from the following databases to ensure that we had a diverse set of arrhythmias:

1. ECG recordings from Creighton University tachyarrhythmia database (CUDB),[25] Massachusetts institute of technology- Beth Israel hospital (MIT-BIH) malignant ventricular arrhythmia database (VFDB),[26] and the sudden cardiac death Holter database

(SDDB)[26] were used. Institutional review board requirements were waived because these databases are all available online at the PhysioNet Physio bank archive.[27] All recordings have a sampling frequency of 250 Hz with 12-bit resolution over a 10-mV range. The recordings with short duration were discarded. The annotation files provided by PhysioNet and visual confirmation by trained ECG experts were used to label both shockable and nonshockable samples from the CUDB, VFDB, and SDDB databases. Consequently, recordings of 20 subjects from CUDB, 10 subjects from VFDB, and 10 subjects from SDDB were used for the analysis. Thirty subjects had both shockable and nonshockable rhythms. The remaining 10 subjects had only nonshockable data. The bad quality portions, such as large variations in the baseline (caused by electrode misplacement, electrode poor contact, and motion artifact) and railed data, were discarded on the basis of the annotation files.

2. The CPR artifact data, without association to patient identification, were acquired from Defibtech's postmarket AED log files. The CPR data were collected with DDU-2200, DDU-2300, DDU-2450, and DDU-2400 devices as well as the DDU-100 and DDU-120. We selected the recordings that contain CPR during asystole. Because the amplitude of ECG during asystole is negligible, the signal should largely represent only CPR. The AED's sampling frequency was 125 Hz. The CPR artifacts are diverse as they were obtained from 43 different rescuers (Figure 1). These samples were visually confirmed by an experienced cardiology fellow and a trained ECG expert. When there was disagreement between the 2 reviewers, which was a rare event, another trained ECG expert broke the tie.

We used the following equation (1) to create a CPR-contaminated ECG data set[2,28] with a signal/noise ratio (SNR) of –3 dB.

$$ECG_{Corrupted} = ECG_{Clean} + \left[std\left(ECG_{Clean}\right) \times 10^{\left(\frac{-SNR}{20}\right)} \times \frac{Noise}{std\left(Noise\right)} \right] \quad (1)$$

In this equation, the terms "ECG_{Clean}," "$ECG_{Corrupted}$," and "CPR" stand for clean ECG, CPR-contaminated ECG, and CPR artifact, respectively. Figure 2 describes the overall procedure for creation of CPR-contaminated ECG signals with a sampling frequency of 125 Hz. In total, 1131 shockable (285 from CUDB, 576 from VFDB, and 270 from SDDB) and 2741 nonshockable (877 from CUDB, 973 from VFDB, and 891 from SDDB) segments were used to produce our database. The preprocessing steps consisted of bandpass filtering (0.430 Hz) and removing the mean value of the data segment. Figure 3 shows representative combined CPR artifacts with ECG rhythms.

We partitioned each signal into 8-second segments, because this time length has been reported to be the best choice for classification accuracy.[29–31] Moreover, most AEDs analyze and display ECG signals in 6- to 12-second data segments.

For each 8-second ECG segment, randomly chosen CPR data from the total of 43 different types were added to it so that 5 nonshockable and 10 shockable

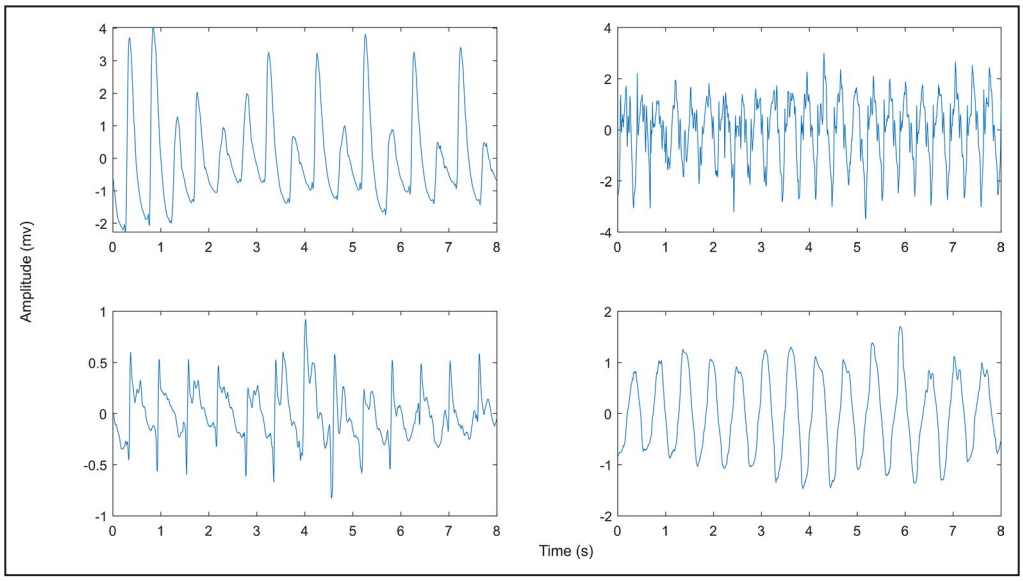


Figure 1. Examples of cardiopulmonary resuscitation (CPR) performed by 4 different rescuers during asystole.
Note the high variability of CPR artifacts in terms of their amplitudes and frequencies.

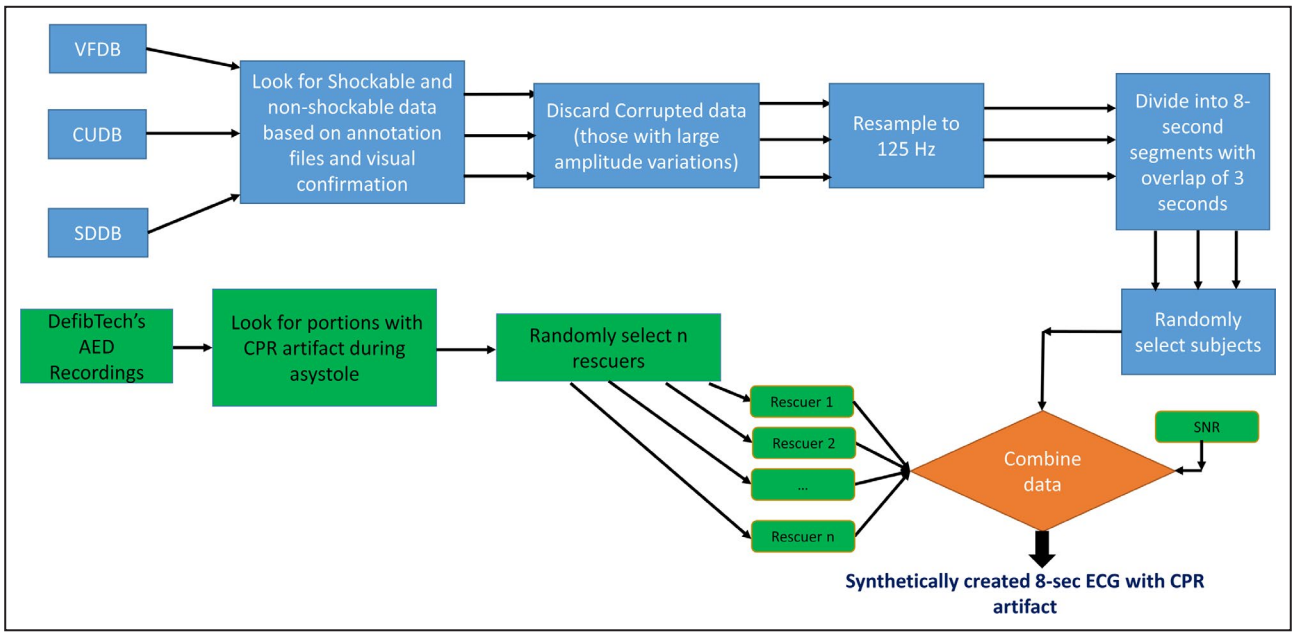


Figure 2. Block diagram of the data preparation.
Samples were from Creighton University tachyarrhythmia database (CUDB), Massachusetts institute of technology- Beth Israel hospital (MIT-BIH) malignant ventricular arrhythmia database (VFDB), and sudden cardiac death Holter database (SDDB). AED indicates automated external defibrillator; CPR, cardiopulmonary resuscitation; DefibTech, defibrillation technician; and SNR, signal/noise ratio.

CPR-contaminated ECG data segments were created. We chose a greater number of CPR types for shockable rhythms than nonshockable rhythms because there were more available nonshockable data. In this manner, we obtained more balanced data sets from both shockable and nonshockable rhythms. Hence, we generated 11 samples (including 10 CPR-contaminated samples and 1 clean sample) for each shockable ECG, and 6 samples (including 5 CPR-contaminated samples and 1 clean sample) for every nonshockable ECG. The reason for preparing a data set containing both CPR-contaminated ECG and clean ECG samples was that our aim is to have an AED able to automatically make shock/no-shock decisions regardless of whether the data contain only the ECG itself or ECG with CPR.

Deep Learning Algorithm Development

Our analytic decision approach is based on the use of the deep neural network (DNN). To train this model, we assume that the rhythm type is consistent throughout each 8-second sample, and no assumption is made about the segments before or after each 8-second segment. Because the model involves the use of the convolutional neural network (CNN), which has been shown to be most effective with digital images, we formulated 2-dimensional images by combining the ECG signal with the amplitude and phase information derived from the short-time Fourier transform. The Hamming window with 50% overlap was used for the short-time Fourier transform. To create the 2-dimensional images, each sample's amplitude (size of 257×8) and phase (size of 257×8) components along with its replicated time series (size of 2000×1) were combined to convert the 1-dimensional signal into a 2-dimensional image representation by simply reshaping elements from a vector of 1×6112 into a matrix of 32×191×1 (Figure 4). The reformulated data images are used as an input to the deep learning model.

Distinguishing between shockable and nonshockable arrhythmia during CPR is a highly challenging task, because of the diversity of ECG arrhythmias, the variability of ECG waveforms, and, most important, variations in CPR artifacts attributable to different CPR delivery among performers. To compensate for these variations, a powerful network is required. The proposed deep learning model, shown in Figure 5, was composed of multiple processing layers consisting sequentially of the CNN, bidirectional long short-term memory (BiLSTM) architecture,[32] and the residual connection layer.[33] As shown in the network architecture (Figure 5), the features are extracted from CNN layers. We used the sequence unfolding layer and the flatten layer to convert images to feature vectors. Subsequently, extracted features were then fed into the BiLSTM layer for classification. Although CNN layers are capable of extracting and learning complex

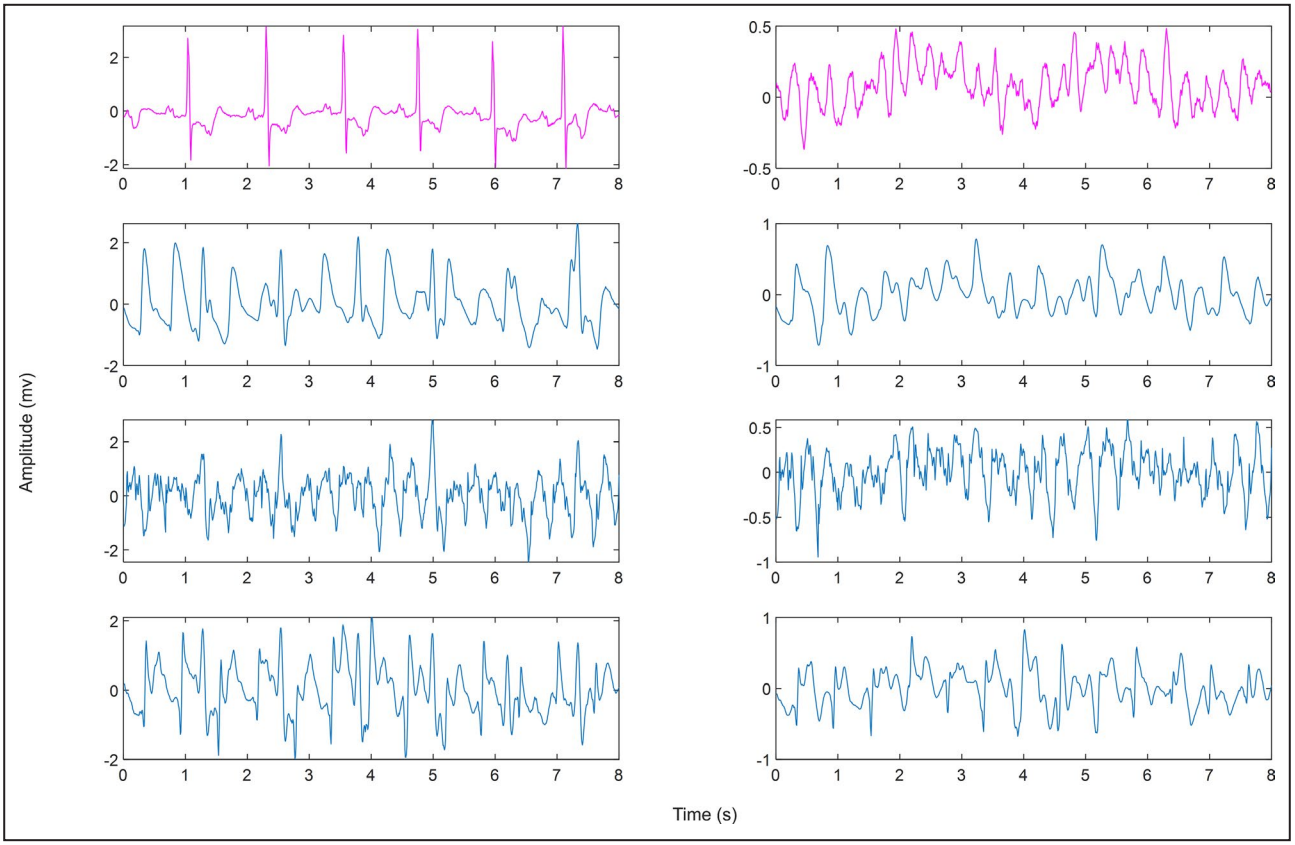


Figure 3. Representative illustrations of the combined various forms of cardiopulmonary resuscitation (CPR) artifacts with either normal sinus rhythm (NSR) or ventricular fibrillation (VF).
In the top row, the left and right panels represent typical NSR and VF, respectively. The bottom 3 rows represent 3 different forms of CPR artifacts that were combined with either NSR (left column) or VF (right column).

features from the input data, the BiLSTM layer is especially well suited for sequences and time-series analysis. Moreover, it can examine the data in both forward and backward directions to look for temporal correlation features. A batch normalization layer after each convolutional layer was used to avoid overfitting during training. We also used residual blocks with 2 CNN architectures in each block, as described in the study by Hannun et al.[23] The residual blocks help in avoiding vanishing gradients in deeper layers and improve the discrimination capability of the features. We examined how different numbers of residual blocks affected the performance, starting from 1, and began to see decreasing performance after 4 blocks. Our proposed network consisted of a far smaller number of residual blocks than the 16 in the study by Hannun et al,[23] plus the BiLSTM layer[34] with the hyperbolic tangent activation function (for hidden layers), fully connected layers, and the sigmoid activation function (for gates) with 128 hidden units. The default initialization, consisting of random weights, was applied. The CNN layers have filter dimensions of 16×32×1, 16×32×2, and 16×32×4, for 1st to 4th, 5th to 8th, and 9th to 12th convolutional layers, respectively. There are dropout layers with a probability of 0.4 before each CNN layer. The adaptive moment estimation (Adam) optimizer with default parameters of β_1=0.9, β_2=0.999, and a mini batch size of 128 was used. The initial learning rate was 0.001, and it decreased every 4 epochs by the drop factor of 0.1. To make a better shock decision based on 8-second samples of ECG data, manual tuning of hyperparameters was performed. Finally, the set of parameters that best fit the objective of the study was chosen.

Statistical Analysis

To avoid overfitting and to provide unbiased evaluations, database 1 was split into training and testing sets using the 4-fold cross-validation procedure. This ensured testing was based on nontrained subjects' ECG segments (Figure 6). To consider nontrained CPR artifacts on each fold, we have separate training (33 rescuers) and testing (10 rescuers) CPR data. The classification results for the entire data set are shown in Table 1. Furthermore, we evaluated the output of the optimally trained DNN model on the clean ECG (Table 2) and CPR-contaminated ECG (Table 3) samples, separately.

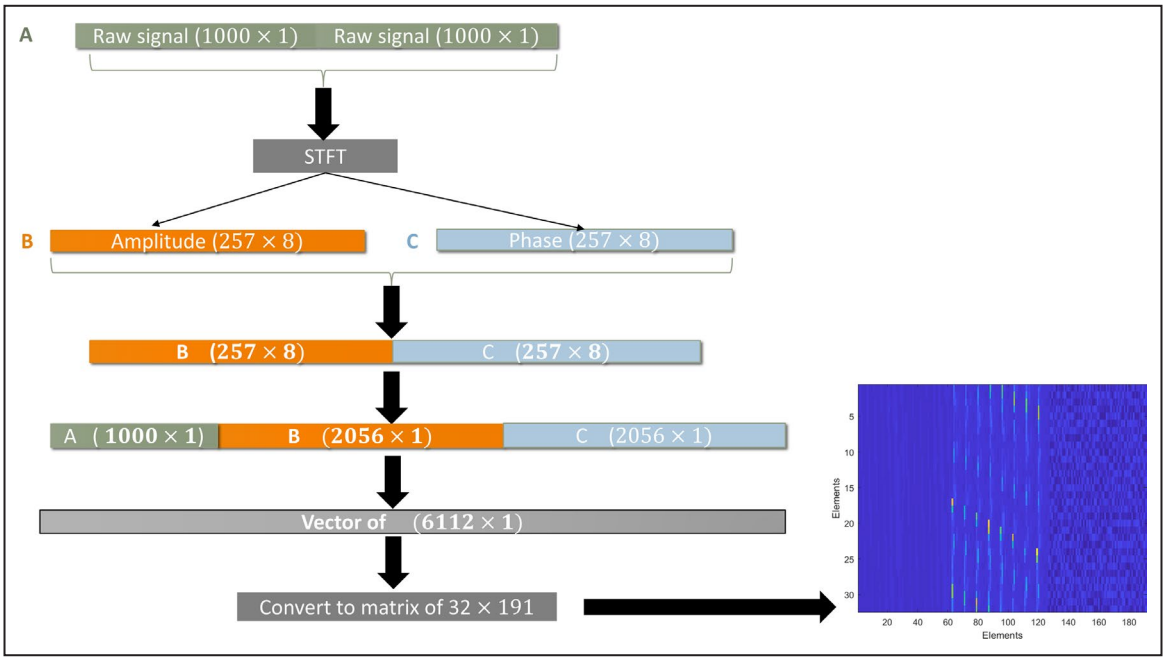
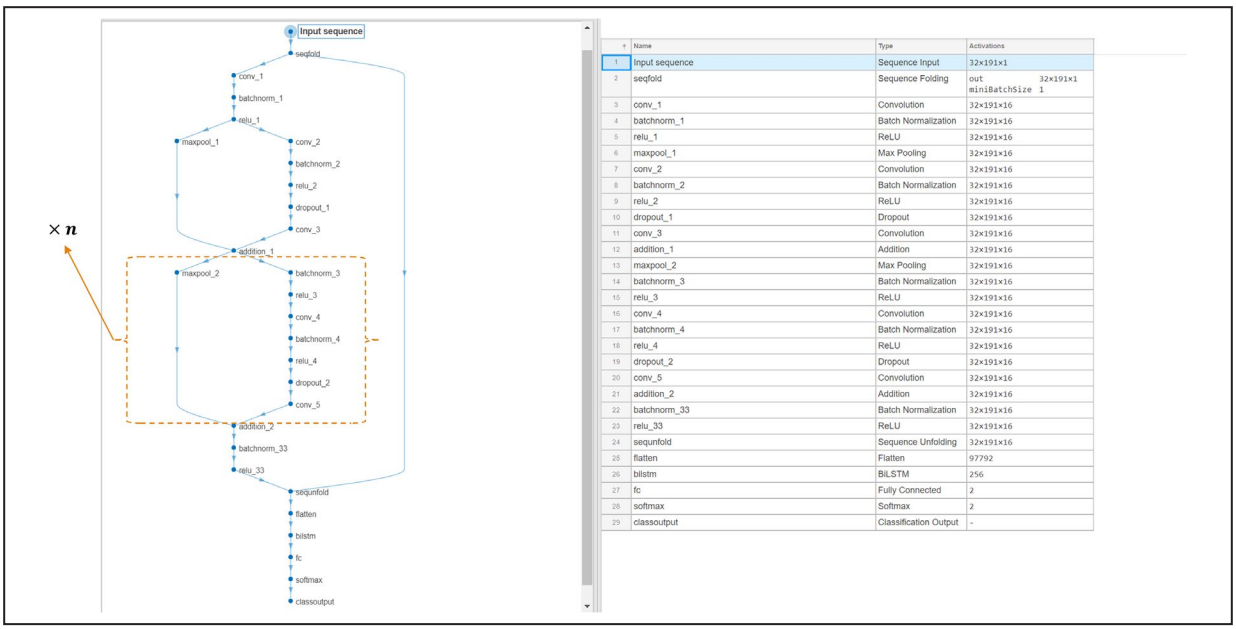


Figure 4. Procedures for converting time series into an image that was then used as an input to a deep neural network model.
First, the 8-second sample is replicated to create time series "A." Then, amplitude "B" and phase "C" of vector A's short-time Fourier transform (STFT) is determined and concatenated. To produce the combined vector of time and frequency information, B and C are reshaped in a single column array and concatenated with A. Finally, this 1-dimensional vector is converted to 2-dimensional matrix, which is appropriate for convolutional layers.

In addition to 4-fold cross-validation, we also examined leave-one-subject-out validation. For this, the DNN model was trained using ECG recordings from 39 subjects and CPR artifacts from 20 different rescuers. Data from the 1 remaining subject and 23 CPR samples (those other than training) were used to create the



	Name	Type	Activations
1	Input sequence	Sequence Input	32×191×1
2	seqfold	Sequence Folding	out 32×191×1 miniBatchSize 1
3	conv_1	Convolution	32×191×16
4	batchnorm_1	Batch Normalization	32×191×16
5	relu_1	ReLU	32×191×16
6	maxpool_1	Max Pooling	32×191×16
7	conv_2	Convolution	32×191×16
8	batchnorm_2	Batch Normalization	32×191×16
9	relu_2	ReLU	32×191×16
10	dropout_1	Dropout	32×191×16
11	conv_3	Convolution	32×191×16
12	addition_1	Addition	32×191×16
13	maxpool_2	Max Pooling	32×191×16
14	batchnorm_3	Batch Normalization	32×191×16
15	relu_3	ReLU	32×191×16
16	conv_4	Convolution	32×191×16
17	batchnorm_4	Batch Normalization	32×191×16
18	relu_4	ReLU	32×191×16
19	dropout_2	Dropout	32×191×16
20	conv_5	Convolution	32×191×16
21	addition_2	Addition	32×191×16
22	batchnorm_33	Batch Normalization	32×191×16
23	relu_33	ReLU	32×191×16
24	sequnfold	Sequence Unfolding	32×191×16
25	flatten	Flatten	97792
26	bilstm	BiLSTM	256
27	fc	Fully Connected	2
28	softmax	Softmax	2
29	classoutput	Classification Output	-

Figure 5. The proposed deep learning neural network architecture.
Left: A block diagram of the model's backbone. As shown by dashed lines, the second block is repeated n times in the models with different numbers of layers. In this study, n varies from 0 to 5. **Right**: Detailed description of each layer and the activations. batchnorm indicates batch normalization layer; bilstm, bidirectional long short-term memory layer; classoutput, classification output layer; conv, convolution layer; fc, fully connected layer; maxpool, maximum pooling layer; sequnfold, sequence unfolding layer; and relu, rectified linear unit layer.

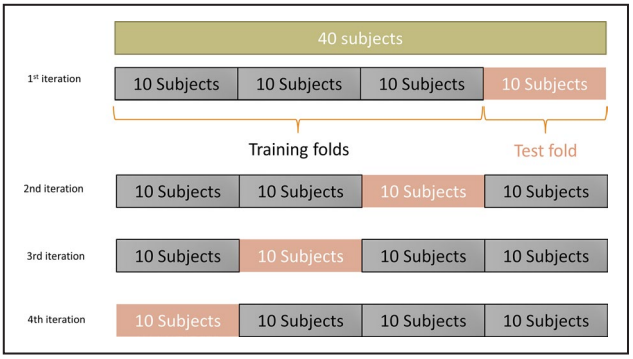


Figure 6. Performance evaluation using 4-fold cross-validation procedure.

testing set. For each training subject's shockable or nonshockable data, 10 CPR samples were randomly selected of 20 CPR samples to produce 10 CPR-contaminated data sets. Repeating this procedure, we created 27 049 nonshockable (2459 clean ECGs and 24 590 CPR-contaminated ECGs) and 10 967 shockable (997 clean ECGs and 9970 CPR-contaminated ECGs) data sets. Each class of data was shuffled. Applying down sample balancing, 10 967 shockable and 10 967 nonshockable samples were produced as the training data. The remaining subject was used for testing, and the choice of this subject was based on the one who had the highest number of the shockable or nonshockable rhythms. For each either shockable or nonshockable ECG sample, CPR artifacts from 23 rescuers (those other than used for training) were combined to produce 3216 shockable rhythms (134 clean ECGs and 3082 CPR-contaminated ECG samples) and 6768 nonshockable rhythms (282 clean ECGs and 6486 CPR-contaminated ECG samples) as the test data.

RESULTS

We determined the optimal number of residual blocks for our data. To examine this, we trained and tested the DNN model using both shockable and nonshockable ECG data with and without CPR artifact using the 4-fold cross-validation method, by varying the number of residual blocks starting from 1 to 6 at an increment of 1. Note that each block contained 2 CNN layers. In addition, there was a single CNN layer at the beginning of the DNN model; thus, in Table 1, we show the total number of CNN layers for our choice of 1 to 6 residual blocks. The results shown in Table 1 were based on the summation of all true positives, true negatives, false positives, and false negatives from the 4-fold cross-validation. These parameters were validated using the information from the annotation files provided by ECG experts who were blinded to initial underlying rhythms. As mentioned in section 2.1, all rhythms (both shockable and nonshockable) without CPR were adjudicated using the information from the associated annotation files, which were created by ECG experts. To these rhythms we added CPR data collected during asystole. Hence, we knew a priori the adjudication of CPR-contaminated rhythms. As shown in Table 1, the network with 4 residual blocks (9 CNN layers) provided the best classification results. Figure 7 better reflects the performance differences between various numbers of residual blocks. Although 5 residual blocks had slightly higher sensitivity (0.31%) than did 4 residual blocks, the specificity was lower (2.66%). In addition, 4 residual blocks had the highest F1-score among all 6 different numbers of residual blocks. Hence, we selected 4 residual blocks containing 9 CNN layers as the final optimal DNN model for subsequent data analysis, to be described henceforth. Increasing the number of residual blocks up to 6 (13 CNN layers) did not further improve the performances, and it came at the expense of higher computational costs.

Table 2 represents the results of the trained DNN model when the validation set contained the ECG samples without CPR artifact. The sensitivity of 99.04% and specificity of 95.2% over the 4-fold cross-validation show that the chosen DNN model is accurate in making shock versus no-shock decisions based on the ECG in absence of CPR. The results for the data set containing ECG with CPR artifact are shown in Table 3.

Table 1. Comparison Results for DNN With Different Numbers of Residual Blocks and CNN Layers for 4-Fold Cross-Validation Sets

No. of Residual Blocks	No. of CNN Layers	Sensitivity, %	Specificity, %	Accuracy, %	F1-Score, %
1	2	90.72	88.23	88.96	82.76
2	5	92.40	86.20	87.30	80.04
3	7	93.50	87.22	88.10	82.20
4	9	95.21	86.03	88.13	83.52
5	11	95.52	83.37	86.22	80.39
6	13	94.50	84.11	86.00	80.10

CNN indicates convolutional neural network; and DNN, deep neural network.

Table 2. Classification Results of Trained DNN Model for ECG in Absence of CPR

Method	Sensitivity, %	Specificity, %	Accuracy, %	F1-Score, %
DNN algorithm (7 CNN layers)	99.04	95.2	96.17	93.50

CNN indicates convolutional neural network; CPR, cardiopulmonary resuscitation; and DNN, deep neural network.

These results are based on the use of the 4-fold cross-validation data, which were iterated 4 times. Evaluation results of the leave-one-subject-out analysis for DNN model are shown in Table 4. The results are for the data set containing ECG with and without CPR artifact. The achieved sensitivity of 92.71% meets the American Heart Association's requirement (>90%). The specificity of 97.6%, which is better than accuracy reported in the literature, is comparable with American Heart Association's requirement (>99% for normal sinus rhythm and >95% for other nonshockables).

Table 5 provides comparison of the result of our proposed DNN model with support vector machine, long short-term memory, and CNN–long short-term memory. For these methods, we show 2 results using 2 different feature vectors. The first feature (feature vector 1) includes both the instantaneous frequency and spectral entropy, whereas the second feature (feature vector 2) includes feature vector 1 as well as the image information converted by taking short-time Fourier transform of the ECG time series. Our proposed method's results are the same as shown in Table 1, and the feature we have used is the image information converted by taking short-time Fourier transform of the ECG time series (following the method in Figure 4). As shown in Table 5, our proposed model provides the best sensitivity, accuracy, and F1-score to classify shockable and nonshockable data with and without CPR artifact when compared with all other methods. For classifiers other than the proposed model, the second feature (feature vector 2) provided better results across the board than the first feature.

DISCUSSION

To the best of our knowledge, this is one of the first studies that has considered the use of deep learning to classify shockable versus nonshockable rhythm using ECG waveforms with and without CPR artifact. Hence, the fact that our DNN approach is able to discriminate this decision only using ECG data with high accuracy is the novel finding of this work. This is a challenging scenario, as CPR artifact often severely overwhelms the ECG morphological features. Our proposed approach does not require any reference signal or additional filtering approaches to remove CPR artifact. Specifically, our proposed deep learning approach is based on the combined deployment of CNN with the residual network and BiLSTM.

Some of the recent notable studies involve the use of data sets extracted from a large prospective study of out-of-hospital cardiac arrest conducted at 3 European sites.[35–38] In these data sets, the ECGs were recorded along with multiple reference signals. The studies first applied adaptive filtering methods using ≥1 reference signal to suppress the CPR artifact. Then, they made shock versus no-shock decision using features extracted from time and frequency domains.[37,38] Recently, the support vector machine classifier[14] and the CNN classifier[39] were applied to make shock versus no-shock decision following the adaptive filtering methods based on the reference signal. The accuracy of the DNN method applied in our work (sensitivity of 95.21% and specificity of 86.03% for the 4-fold cross-validation, and sensitivity of 92.71% and specificity of 97.6% for leave-one-subject-out analysis) is comparable with the accuracy of most of the previous methods using reference signal (sensitivity of ≈90%–96% and specificity of ≈79%–96%). However, the main disadvantage of using a reference signal is that most AEDs do not have the capability to capture it.[14]

Only a few studies have made an attempt to make shock versus no-shock decisions during CPR using the ECG data but without a reference signal. For instance, in the study by Didon et al,[16] the authors proposed a method for shock advisory decisions based on extracting 3 frequency bands derived from the power spectral density of the ECG combined with CPR artifact. This method requires setting threshold values to distinguish between different rhythms. They did not have separate testing and training subjects. The reported sensitivity and specificity values were 94.2% and 87%, respectively.

Table 3. Classification Results of Trained DNN Model for ECG in Presence of CPR

Method	Sensitivity, %	Specificity, %	Accuracy, %	F1-Score, %
DNN algorithm (7 CNN layers)	94.21	86.14	90.1	89.33

The entire process of creating testing samples using 4-fold cross-validation was repeated 4 times for each validation set. Therefore, the results are based on considering results from 16 separate iterations. CNN indicates convolutional neural network; CPR, cardiopulmonary resuscitation; and DNN, deep neural network.

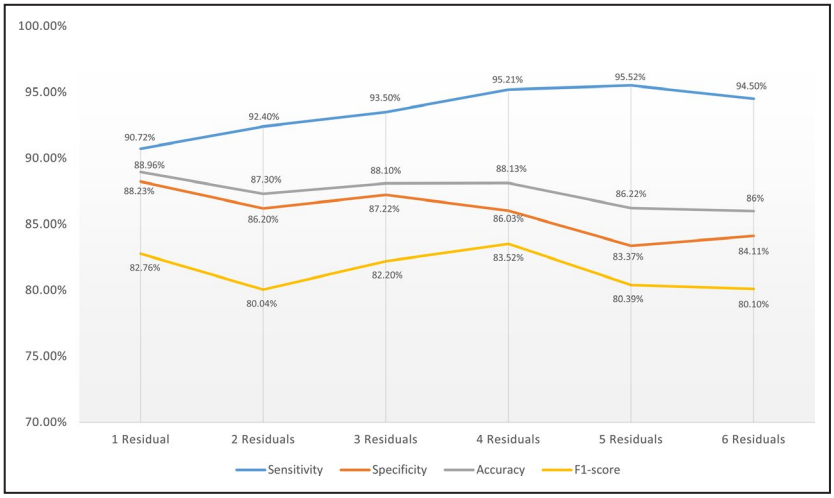


Figure 7. Comparison of model with different numbers of residual blocks.

Although the objective of most recent studies is to remove CPR artifacts so that the AED can make a judicious shockable versus nonshockable decision, the DNN model used in this work automatically makes the decision using only the ECG data, which may or may not be corrupted with CPR artifact. The processing time for the trained DNN model with 9 CNN layers for an 8-second data segment is ≈150 milliseconds on a DELL Intel Core i7 CPU processor using MATLAB 2019a.

Limitations

There are several potential limitations that need to be considered for future studies. First, the results were based on a limited number of subjects. Deep learning requires a vast amount of training data to be effective and to produce generalizable results. Hence, the DNN model used in this work needs to be further tested with larger data sets. The larger data sets should consist of different arrhythmias, various CPR artifacts, and data from different AED devices, to name a few considerations. Second, although we were able to create ECG recordings with CPR artifact, the performance of the proposed DNN model needs to be further tested on ECG data from AED devices during CPR. Third, we have not tested our DNN model during asystole. Finally, determining the optimum number of training iterations with different numbers of layers is challenging. In this study, the number of iterations used for training was fixed for all training data. The iteration number was chosen on the basis of manual tuning. However, this may not have been optimal because different numbers of residual layers may need different numbers of iterations to arrive at their best performance.

CONCLUSIONS

Early and accurate detection of shockable rhythms without interrupting CPR is one of the most important goals for increasing out-of-hospital cardiac arrest survival rate. The current AEDs require stopping CPR while ECG data are analyzed, which deprives critical oxygen supply to the brain. To overcome this limitation in current AED capabilities, a new method based on DNN was used. We found that DNN used in this work was successful in making accurate shock decision even in the presence of CPR, as the sensitivity and specificity were 94.26% and 86.14%, respectively, over 4-fold cross-validation procedure. We also achieved sensitivity of 92.71% and specificity of 97.6%, using test data that were not seen by the deep learning method. Although the results are based on a limited number of subjects, both training and testing were based on several thousand 8-second data segments to minimize overfitting of the model. The sensitivity meets the American Heart Association's requirement for AEDs. However, further validation is needed, including using more training and testing data sets from different AED devices. In addition, the DNN architecture may also need to be optimized so that the most

Table 4. Classification Results of Trained DNN Model for One-Subject-Out Validation Scenario

Method	Sensitivity, %	Specificity, %	Accuracy, %	F1-Score, %
DNN algorithm (7 CNN layers)	92.71	97.6	96.33	93.06

CNN indicates convolutional neural network; and DNN, deep neural network.

Table 5. Comparing the Proposed Method With Some Other Popular Classifiers

Classifier	Features	Sensitivity, %	Specificity, %	Accuracy, %	F1-Score, %
SVM	F vector 1	66.81	53.82	57.61	47.94
SVM	F vector 2	70.57	85.00	80.79	68.21
LSTM	F vector 1	67.73	82.93	78.49	64.78
LSTM	F vector 2	86.78	87.14	87.04	79.64
CNN-LSTM	F vector 1	70.48	81.19	78.07	65.25
CNN-LSTM	F vector 2	82.11	89.08*	87.05	78.74
Proposed model	Raw information	95.21*	86.03	88.13*	83.52*

CNN indicates convolutional neural network; F vector 1, feature vector 1; F vector 2, feature vector 2; LSTM, long short-term memory; and SVM, support vector machine.

*indicates the highest value of each metric.

compact and computationally efficient model can be developed that can be embedded into AEDs for real-time implementation.

ARTICLE INFORMATION

Received August 24, 2020; accepted January 4, 2021.

Affiliations

From the Biomedical Engineering Department, University of Connecticut, Storrs, CT (S.H., K.H.C.); and Defibtech, LLC, Guilford, CT (A.C., M.V.).

Sources of Funding

This work was supported by a grant from Defibtech, LLC.

Disclosures

None.

REFERENCES

1. Ristagno G, Mauri T, Cesana G, Li Y, Finzi A, Fumagalli F, Rossi G, Grieco N, Migliori M, Andreassi A, et al. Amplitude spectrum area to guide defibrillation: a validation on 1617 patients with ventricular fibrillation. *Circulation*. 2015;131:478–487. DOI: 10.1161/CIRCULATIONAHA.114.010989.
2. Ruiz de Gauna S, Irusta U, Ruiz J, Ayala U, Aramendi E, Eftestøl T. Rhythm analysis during cardiopulmonary resuscitation: past, present, and future. DOI: 10.1155/2014/386010. Available at: https://www.hindawi.com/journals/bmri/2014/386010/. Accessed January 9, 2020.
3. Hu Y, Tang H, Liu C, Jing D, Zhu H, Zhang Y, Yu X, Zhang G, Xu J. The performance of a new shock advisory algorithm to reduce interruptions during CPR. *Resuscitation*. 2019;143:1–9. DOI: 10.1016/j.resuscitation.2019.07.026.
4. Waalewijn RA, de Vos R, Tijssen JG, Koster RW. Survival models for out-of-hospital cardiopulmonary resuscitation from the perspectives of the bystander, the first responder, and the paramedic. *Resuscitation*. 2001;51:113–122. DOI: 10.1016/s0300-9572(01)00407-5.
5. Waalewijn RA, Nijpels MA, Tijssen JG, Koster RW. Prevention of deterioration of ventricular fibrillation by basic life support during out-of-hospital cardiac arrest. *Resuscitation*. 2002;54:31–36. DOI: 10.1016/s0300-9572(02)00047-3.
6. Eilevstjønn J, Eftestøl T, Aase SO, Myklebust H, Husøy JH, Steen PA. Feasibility of shock advice analysis during CPR through removal of CPR artefacts from the human ECG. *Resuscitation*. 2004;61:131–141. DOI: 10.1016/j.resuscitation.2003.12.019.
7. Tan Q, Freeman GA, Geheb F, Bisera J. Electrocardiographic analysis during uninterrupted cardiopulmonary resuscitation. *Crit Care Med*. 2008;36:S409–S412. DOI: 10.1097/ccm.0b013e31818a7fbf.
8. 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*. 2008;55:130–137. DOI: 10.1109/TBME.2007.902235.
9. Ruiz J, Irusta U, Ruiz de Gauna S, 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. DOI: 10.1016/j.resuscitation.2010.02.031.
10. Gong Y, Yu T, Chen B, He M, Li Y. Removal of cardiopulmonary resuscitation artifacts with an enhanced adaptive filtering method: an experimental trial. DOI: 10.1155/2014/140438. Available at: https://www.hindawi.com/journals/bmri/2014/140438/. Accessed January 9, 2020.
11. Gong Y, Gao P, Wei L, Dai C, Zhang L, Li Y. An enhanced adaptive filtering method for suppressing cardiopulmonary resuscitation artifact. *IEEE Trans Biomed Eng*. 2017;64:471–478. DOI: 10.1109/TBME.2016.2564642.
12. Irusta U, Ruiz J, de Gauna SR, Eftestøl T, Kramer-Johansen J. A least mean-square filter for the estimation of the cardiopulmonary resuscitation artifact based on the frequency of the compressions. *IEEE Trans Biomed Eng*. 2009;56:1052–1062. DOI: 10.1109/TBME.2008.2010329.
13. Werther T, Klotz A, Kracher G, Baubin M, Feichtinger HG, Gilly H, Amann A. CPR artifact removal in ventricular fibrillation ECG signals using Gabor multipliers. *IEEE Trans Biomed Eng*. 2009;56:320–327. DOI: 10.1109/TBME.2008.2003107.
14. Isasi I, Irusta U, Elola A, Aramendi E, Ayala U, Alonso E, Kramer-Johansen J, Eftestol T. A machine learning shock decision algorithm for use during piston-driven chest compressions. *IEEE Trans Biomed Eng*. 2019;66:1752–1760. DOI: 10.1109/TBME.2018.2878910.
15. Ruiz de Gauna S, Ruiz J, Irusta U, Aramendi E, Eftestøl T, Kramer-Johansen J. A method to remove CPR artefacts from human ECG using only the recorded ECG. *Resuscitation*. 2008;76:271–278. DOI: 10.1016/j.resuscitation.2007.08.002.
16. Didon J, Dotsinsky I, Jekova I, Krasteva V. Detection of shockable and non-shockable rhythms in presence of CPR artifacts by time-frequency ECG analysis. In 2009 36th Annual Computers in Cardiology Conference; 2009:817–820. Available at: https://ieeexplore.ieee.org/document/5445257/citations. Accessed June 15, 2020.
17. Li Y, Tang W. Techniques for artefact filtering from chest compression corrupted ECG signals: good, but not enough. *Resuscitation*. 2009;80:1219–1220. DOI: 10.1016/j.resuscitation.2009.09.003.
18. Aramendi E, de Gauna SR, Irusta U, Ruiz J, Arcocha MF, Ormaetxe JM. Detection of ventricular fibrillation in the presence of cardiopulmonary resuscitation artefacts. *Resuscitation*. 2007;72:115–123. DOI: 10.1016/j.resuscitation.2006.05.017.
19. Li Y, Bisera J, Geheb F, Tang W, Weil MH. Identifying potentially shockable rhythms without interrupting cardiopulmonary resuscitation. *Crit Care Med*. 2008;36:198–203. DOI: 10.1097/01.CCM.0000295589.64729.6B.
20. Krasteva V, Jekova I, Dotsinsky I, Didon J-P. Shock advisory system for heart rhythm analysis during cardiopulmonary resuscitation using a single ECG input of automated external defibrillators. *Ann Biomed Eng*. 2010;38:1326–1336. DOI: 10.1007/s10439-009-9885-9.
21. Kerber RE, Becker LB, Bourland JD, Cummins RO, Hallstrom AP, Michos MB, Nichol G, Ornato JP, Thies WH, White RD, 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. DOI: 10.1161/01.cir.95.6.1677.
22. Isasi I, Rad AB, Irusta U, Zabihi M, Aramendi E, Eftestøl T, Kramer-Johansen J, Wik L. ECG rhythm analysis during manual chest compressions using an artefact removal filter and random forest classifiers. In 2018 Computing in Cardiology Conference; 2018;45:1–4. DOI: 10.22489/CinC.2018.202.
23. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, Ng AY. Cardiologist-level arrhythmia detection in ECGs using a deep neural network. *Nat Med*. 2019;25:65–69. DOI: 10.1038/s41591-018-0268-3.
24. Parvaneh S, Rubin J, Babaeizadeh S, Xu-Wilson M. Cardiac arrhythmia detection using deep learning: a review. *J Electrocardiol*. 2019;57:S70–S74. DOI: 10.1016/j.jelectrocard.2019.08.004.
25. CREI-GARD, a new concept in computerized arrhythmia monitoring systems. | Scholars@Duke. Available at: https://scholars.duke.edu/display/pub754604. Accessed January 10, 2020.
26. Greenwald SD. *The Development and Analysis of a Ventricular Fibrillation Detector*. Cambridge, MA: Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science; 1986. Available at: http://hdl.handle.net/1721.1/92988. Accessed October 25, 2019.
27. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*. 2000;101:E215–E220. DOI: 10.1161/01.cir.101.23.e215.
28. Yu M, Zhang G, Wu T, Li C, Wan Z, Li L, Wang C, Wang Y, Lu H, Chen F. A new method without reference channels used for ventricular fibrillation detection during cardiopulmonary resuscitation. *Australas Phys Eng Sci Med*. 2016;39:391–401. DOI: 10.1007/s13246-016-0425-2.
29. Hajeb-M SH, Ahmadi M, Shahghadami R, Chon KH. Automated method for discrimination of arrhythmias using time, frequency, and nonlinear features of electrocardiogram signals. *Sensors*. 2018;18:2090. DOI: 10.3390/s18072090.
30. Verma A, Dong X. Detection of ventricular fibrillation using random forest classifier. *J Biomed Sci Eng*. 2016;9:259–268. DOI: 10.4236/jbise.2016.95019.
31. Amann A, Tratnig R, Unterkofler K. Reliability of old and new ventricular fibrillation detection algorithms for automated external defibrillators. *Biomed Eng Online*. 2005;4:60. DOI: 10.1186/1475-925X-4-60.
32. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;9:1735–1780. DOI: 10.1162/neco.1997.9.8.1735.
33. He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. In: *European conference on computer vision*. (pp. 630–645) Cham: Springer; 2016.
34. Thireou T, Reczko M. Bidirectional long short-term memory networks for predicting the subcellular localization of eukaryotic proteins. *IEEE/ACM Trans Comput Biol Bioinform*. 2007;4:441–446. DOI: 10.1109/tcbb.2007.1015.
35. Wik L, Kramer-Johansen J, Myklebust H, Sørebø H, Svensson L, Fellows B, Steen PA. Quality of cardiopulmonary resuscitation during out-of-hospital cardiac arrest. *JAMA*. 2005;293:299–304. DOI: 10.1001/jama.293.3.299.
36. Kramer-Johansen JO, Myklebust H, Wik L, Fellows B, Svensson L, Sørebø H, Steen PA. Quality of out-of-hospital cardiopulmonary resuscitation with real time automated feedback: a prospective interventional study. *Resuscitation*. 2006;71:283–292. DOI: 10.1016/j.resuscitation.2006.05.011.
37. Ayala U, Irusta U, Ruiz J, Eftestøl T, Kramer-Johansen J, Alonso-Atienza F, González-Otero D. A reliable method for rhythm analysis during cardiopulmonary resuscitation. *Biomed Res Int*. 2014;2014:872470. DOI: 10.1155/2014/872470. Available at: https://www.hindawi.com/journals/bmri/2014/872470/. Accessed January 13, 2020.
38. Isasi I, Irusta U, Bahrami Rad A, 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. DOI: 10.1109/ACCESS.2019.2935096.
39. Isasi I, Irusta U, Aramendi E, Eftestøl T, Kramer-Johansen J, Wik L. Rhythm analysis during cardiopulmonary resuscitation using convolutional neural networks. *Entropy*. 2020;22:595. DOI: 10.3390/e22060595.