Recurrent Neural Networks to Predict the Outcome of Subsequent Defibrillation Shocks in Cardiac Arrest

Xabier Jaureguibeitia[1], Gorka Zubia[1], Unai Irusta[1,2], Elisabete Aramendi[1,2], Giuseppe Ristagno[3]

[1] University of the Basque Country (UPV/EHU), Bilbao, Spain
[2] Biocruces Bizkaia Health Research Institute, Cruces University Hospital, Barakaldo, Spain
[3] IRCCS – Istituto di Ricerche Farmacologiche Mario Negri, Milano, Italy

Abstract

Accurate prediction of defibrillation success could help to optimize the treatment of out-of-hospital cardiac arrest (OHCA). On top of the classical predictors derived from the ventricular fibrillation (VF) waveform, the outcome of the preceding shock can also be used to enhance prediction. This work introduces a recurrent neural network (RNN) to predict the outcome of subsequent shocks, using the full history of previous outcomes and VF measures.

Data from 957 OHCA patients were analyzed, comprising 3159 shocks of which 2202 were subsequent shocks. Shocks were labeled as successful/unsuccessful based on their ability to restore an organized rhythm. Each patient data were modeled as a time sequence, one shock per time-step, and fed to a RNN. Individual shocks were characterized using a single classical predictor (computed on a 2s pre-shock VF window) and the outcome of the previous shock. Seven predictors were independently analyzed.

Model performance was assessed through 10-fold cross-validation (CV) over 50 different CV data partitions. Balanced accuracy (BAC) was chosen as target performance metric. Mean Slope was the best predictor, with median (interquartile range, IQR) BAC of 84.2 (84.0 - 84.4)%, 3.6 points above that of simple thresholding with no previous shock information. RNNs could improve subsequent shock-outcome prediction using previous shock history.

1. Introduction

Defibrillation and high-quality cardiopulmonary resuscitation (CPR) are the main therapies in the early treatment of out-of-hospital cardiac arrest (OHCA). However, the optimal timing and balance between both is still unclear. Not every defibrillation attempt succeeds at restoring spontaneous pulse, and these unsuccessful shocks result in CPR interruptions which are, in turn, detrimental to patient survival. Unnecessary and repetitive shocks may also cause damage to the myocardium, making the heart less responsive to further attempts. Conversely, continued CPR may improve myocardial state, increasing the chances of a successful follow-up shock. Consequently, an accurate prediction of shock success could help rescuers select the optimal therapy, and potentially improve the rates of survival.

Ventricular fibrillation (VF) is the main arrhythmia sensible to defibrillation during OHCA. The morphology of the VF waveform may provide information on myocardial state, so many different ECG-based features have been proposed over the years as shock-outcome predictors [1–5]. The type of heart rhythm prior to the onset of VF is also related to shock success [6], with higher success rates being observed for organized rhythms. Since shock success is typically determined by the restoration of an organized rhythm, some studies have proposed to use the outcome of the preceding shock as additional predictor [7, 8]. This may significantly enhance outcome prediction for subsequent shocks.

The goal of this study was to assess if a recurrent neural network (RNN), fed with the complete history of previous shock-outcomes and VF features, could help improve subsequent shock-outcome prediction.

2. Data Materials

The study dataset comprised the electronic recordings from 957 OHCA episodes treated by the Italian emergency services in the period between 2008 and 2010. Only cases with several defibrillation attempts were considered. The dataset included files from various manufacturers and defibrillator models: Philips Heartstart FR2 (Philips Medical Systems, Andover, MA, US), ZOLL M-Series (ZOLL Medical, Chelmsford, MA, US), and Lifepak LP12, LP500 and LP1000 (Stryker, Kalamazoo, MI, US). All files were converted to a common MATLAB (MathWorks Inc., Natick, MA, US) format, and all ECG recordings were resampled to a common sampling frequency $f_s = 250\,\text{Hz}$. Shocks were located in the ECG according to the event logs recorded by the AEDs. Then, the ECG waveform fol-

 Page 1 ISSN: 2325-887X DOI: 10.22489/CinC.2021.183

lowing each shock was analyzed by expert clinicians, and the shock annotated as successful or unsuccessful. Shocks were annotated as successful if an organized heart rhythm, with rate $\geq$ 40 beats per minute, could be located within one minute of the shock [1]. The dataset included 3159 shocks, of which 2202 were subsequent shocks. Among the subsequent shocks, 619 (28.1%) were annotated as successful, and 1583 (71.9%) as unsuccessful.

3. Methods

Each patient data were modeled as a time sequence, one shock per time-step, and fed to a RNN. Input data for each individual shock comprised one out of seven ECG-based state-of-the-art predictors (computed over a 2s pre-shock VF window), and the outcome of the immediately previous shock. Model performance was assessed using 10-fold cross-validation (CV), over 50 different CV data partitions. Performance was independently assessed for each predictor.

3.1. ECG-based Feature Extraction

A 512 sample ECG window (≈ 2 s), 0.5 s prior to shock delivery, was used to analyze the VF waveform at each defibrillation (see Figure 1). Seven shock-outcome predictors were calculated: amplitude spectrum area (AMSA) [1], mean and median slopes (MS, MdS) [2], average peak-to-peak amplitude (PPA) [3], sample and fuzzy entropies (SampEn, FuzzEn) [4], and the mean step increment (MSI) of the Poincarè plot [5]. All predictors were computed using the specific implementation and ECG window preprocessing described by Chicote et al [9], and available online through https://github.com/BChicote/shockOutcome.

3.2. Classification

Each patient's data were modeled as a time series, where each time-step corresponded to a shock. As shown in Figure 1, three variables were used to characterize each shock: a shock-outcome predictor, computed from the VF waveform prior to shock delivery, and two binary variables, PS1 and PS0, indicating whether the immediately previous shock was successful or unsuccessful, respectively. Both PS0 and PS1 took zero value for the first shock in each series.

The shock series were fed to a recurrent neural network, consisting of 20 Long-Short Term Memory (LSTM) units. A sequence-to-sequence configuration was selected, which allows for the individual prediction on each of the shocks within a series. Models were trained for 200 epochs, using a batch size of 32 time-series, and an Adam optimizer with an initial learning rate of 5×10^{-4}. Sample weights were applied to balance the importance of successful (minority class) and unsuccessful shocks. A label-smoothing of 0.1 was also applied.

3.3. Data Partitioning and Evaluation

Patients were divided into training and test sets using 10-fold CV. A quasi-stratified fold assignment was used, so that the overall distributions of the complete dataset

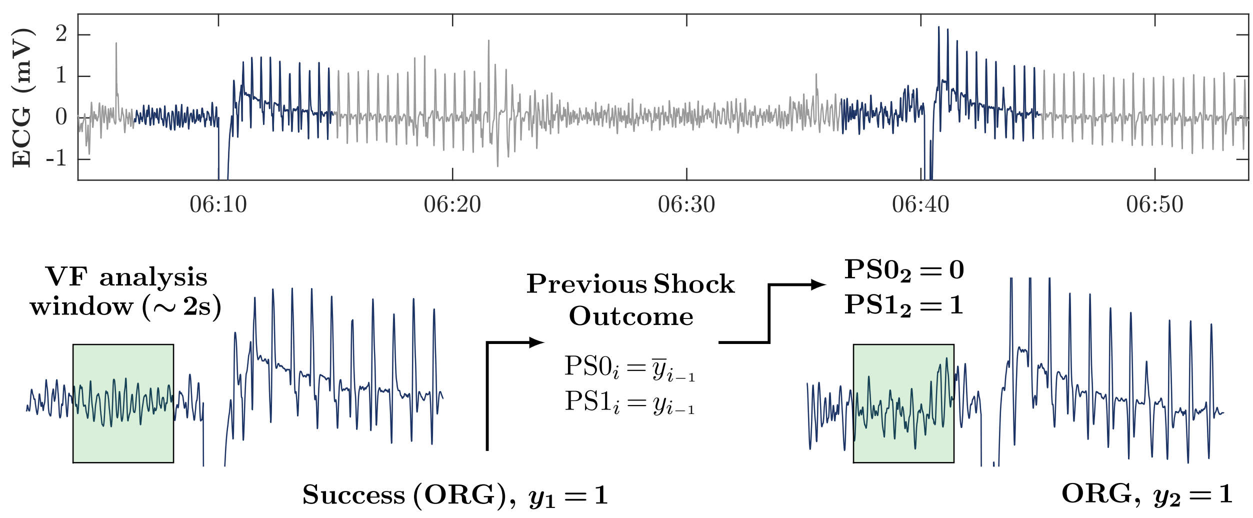

Figure 1. A 50 s ECG trace including two shocks, a first shock and a subsequent one (top image). Bottom image shows a magnified version of both shock surroundings. An ECG window of about 2 s prior to each shock was used to analyze and extract predictors related to the VF waveform. Both shocks were annotated as successful ($y_i = 1$), as an organized rhythm could be identified within 1 min of delivery. The output of the first shock was fed to the second one as additional predictors (PS0 = 0, PS1 = 1). Both PS0 and PS1 were set to zero for the first shock.

were approximately preserved. The number of subsequent shocks, the proportion of first and subsequent successful shocks, and the proportion of shocks for each defibrillator model, all were restricted to a maximum deviation of 10%. This process was repeated 50 times to avoid partition bias and to statistically characterize the results.

Performance was assessed in terms of sensitivity (Se, proportion of correctly predicted successful shocks), specificity (Sp, proportion of correctly predicted unsuccessful shocks), and balance accuracy (BAC, average of Se and Sp). The BAC was chosen as target performance metric, as both successful and unsuccessful shock-outcome prediction are of clinical relevance. The analysis was aimed at subsequent shocks, so first shocks were excluded from performance calculations. Median (interquartile range, IQR) scores, computed over the 50 CV partitions, are reported.

4. Results and Discussion

Table 1 shows the performance metrics obtained for the LSTM models for each predictor. The table also includes the BAC improvement (ΔBAC) over single predictor thresholding, with no previous shock information (the threshold was chosen to maximize the BAC in the training set). The models based on MS scored the highest median performance. AMSA, the best shock-outcome predictor (by simple thresholding) for both all shocks and exclusively first shocks, seemed to benefit less from previous shock information. Still, predictors were found highly correlated (Pearson's correlation coefficients above 0.9 between each pair of predictors) and performance differences were small. The addition of previous shock information significantly improved results compared to simple thresholding ($p < 0.05$ by a cost-sensitive McNemar's test).

Table 2 shows the median performance metrics by state-of-the-art solutions, cross-validated on the study dataset. He et al proposed a two-hidden layer artifical neural network, fed with AMSA, the outcome of the previous shock (akin to PS1), and the variation of AMSA with respect this previous shock [7]. Coult et al proposed a linear regression model using the outcome of the previous shock and the natural logarithm of either AMSA (a) or MdS (b) [8]. ECG-based predictors were computed using the window preprocessing and implementation described in the original studies, although restricted to the 2 s analysis window used in this work. The LSTM models based on MS outperformed the best state-of-the-art solution by about 0.8 BAC points. Even for the same ECG-based features, the results of the RNN were still better, suggesting that the full history of previous outcomes and VF measures could positively contribute to prediction. Differences were small though ($p > 0.05$), meaning that the outcome of the immediately previous shock is most likely the main contributor by a large margin. This is in line with the observations of He et al on the contribution of AMSA variation [7]. Further studies are needed to assess the individual contributions within the previous shock history.

It should also be noticed that none of the state-of-the-art solutions was designed for BAC maximization. Several other performance metrics have usually been targeted on shock-outcome prediction studies, including Se at 90% Sp, Sp at 90% Se, or the area under the ROC curve (AUC). Since no conclusive clinical trial has been yet conducted on shock-outcome prediction assisted therapy, the optimal performance target from a clinical perspective remains unclear. The differences in Se/Sp balance in both Table 1 and Table 2 suggest that different predictors could be better at targeting different performance metrics. It is also worth noting that maximizing an specific metric may prove a difficult task due to the high amount of outliers. As a prediction problem, there's a relevant number of instances, mostly unsuccessful shocks, for which the outcome can not be explained neither through VF measures nor through previous outcomes, and which may introduce a considerable bias into the loss function. For simple models like a logistic regression, this can be corrected to some extent by probability output thresholding, but more complex models, like neural networks, are less responsive to such approach and must rely mostly on training weights. The LSTM models in this study made use of label-smoothing to deal with these outliers, but the effect was minor and further strategies should be explored.

Table 1. Median (IQR) performance metrics per prediction feature

Predictor	BAC (%)	ΔBAC (%)	Se (%)	Sp (%)
MS	84.2 (84.0-84.4)	3.6 (3.4-3.7)	87.4 (87.2-87.5)	81.0 (80.8-81.3)
MSI	84.0 (83.9-84.1)	4.5 (4.4-4.8)	87.1 (86.9-87.2)	80.8 (80.7-81.0)
PPA	83.9 (83.7-84.0)	4.7 (4.4-5.0)	85.8 (85.3-85.9)	82.1 (81.9-82.3)
FuzzEn	83.8 (83.6-84.0)	4.4 (4.2-4.7)	86.1 (85.8-86.4)	81.6 (81.4-81.8)
MdS	83.6 (83.4-83.8)	3.7 (3.6-3.9)	86.4 (86.2-86.7)	80.8 (80.5-80.9)
SampEn	83.6 (83.4-83.7)	3.5 (3.3-3.7)	86.1 (85.8-86.4)	81.2 (80.9-81.3)
AMSA	83.6 (83.4-83.7)	3.0 (2.8-3.2)	85.4 (85.3-85.6)	81.7 (81.5-81.8)

Table 2. Median performance of state-of-the-art solutions

Solution	BAC (%)	Se (%)	Sp (%)
Coult et al [a] [8]	83.3	86.4	80.2
Coult et al [b]	83.1	81.3	85.0
He et al [7]	82.8	84.5	80.9

5. Conclusions

A recurrent neural network, fed with the full history of previous VF measures and shock-outcomes, could potentially improve shock-outcome prediction for subsequent shocks. The outcome of the immediately prior shock is most likely the main contributor for this improvement. Mean slope (MS) was found the best ECG-based predictor at maximizing BAC using previous shock information.

Acknowledgments

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), by the Basque Government through grants IT1229-19 and PRE-2020-2-0182, and by the University of the Basque Country (UPV/EHU) through grant COLAB20/01.

References

[1] G. Ristagno, T. Mauri, G. Cesana *et al.*, “Amplitude spectrum area to guide defibrillation: a validation on 1617 patients with ventricular fibrillation.” *Circulation*, vol. 131, pp. 478–487, Feb. 2015.

[2] A. Neurauter, T. Eftestøl, J. Kramer-Johansen *et al.*, “Prediction of countershock success using single features from multiple ventricular fibrillation frequency bands and feature combinations using neural networks.” *Resuscitation*, vol. 73, pp. 253–263, May 2007.

[3] R. Firoozabadi, M. Nakagawa, E. D. Helfenbein *et al.*, “Predicting defibrillation success in sudden cardiac arrest patients.” *Journal of Electrocardiology*, vol. 46, pp. 473–479, 2013.

[4] B. Chicote, U. Irusta, R. Alcaraz *et al.*, “Application of entropy-based features to predict defibrillation outcome in cardiac arrest,” *Entropy*, vol. 18, no. 9, p. 313, 2016.

[5] Y. Gong, Y. Lu, L. Zhang *et al.*, “Predict defibrillation outcome using stepping increment of poincare plot for out-of-hospital ventricular fibrillation cardiac arrest.” *BioMed research international*, vol. 2015, p. 493472, 2015.

[6] J. Eilevstjønn, J. Kramer-Johansen, and K. Sunde, “Shock outcome is related to prior rhythm and duration of ventricular fibrillation.” *Resuscitation*, vol. 75, pp. 60–67, Oct. 2007.

[7] M. He, Y. Lu, L. Zhang *et al.*, “Combining amplitude spectrum area with previous shock information using neural networks improves prediction performance of defibrillation outcome for subsequent shocks in out-of-hospital cardiac arrest patients.” *PloS One*, vol. 11, p. e0149115, 2016.

[8] J. Coult, H. Kwok, L. Sherman *et al.*, “Ventricular fibrillation waveform measures combined with prior shock outcome predict defibrillation success during cardiopulmonary resuscitation.” *Journal of Electrocardiology*, vol. 51, pp. 99–106, 2018.

[9] B. Chicote, U. Irusta, E. Aramendi *et al.*, “Fuzzy and sample entropies as predictors of patient survival using short ventricular fibrillation recordings during out of hospital cardiac arrest,” *Entropy*, vol. 20, no. 8, p. 591, 2018.

Address for correspondence:

Xabier Jaureguibeitia
Plaza Ingeniero Torres Quevedo 1. 48013 - Bilbao, Spain
xabier.jaureguibeitia@ehu.eus