A Convolutional Neural Network Approach for Interpreting Cardiac Rhythms from Resuscitation of Cardiac Arrest Patients

Trygve Eftestøl[1], Mari A Hognestad[1], Sander A Søndeland[1], Ali Bahrami Rad[2], Elisabete Aramendi[3], Lars Wik[4,5], Jo Kramer-Johansen[4,5,6]

[1] University of Stavanger, Stavanger, Norway
[2] Emory University, Atlanta, Georgia, USA
[3] University of the Basque Country, Bilbao, Spain
[4] Norwegian National Advisory Unit for Prehospital Emergency Care (NAKOS), Oslo University Hospital, Oslo, Norway
[5] Air ambulance dept., Division of Prehospital Care, Oslo University Hospital, Oslo, Norway
[6] Division of Prehospital Care, Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway

Abstract

Patients undergoing cardiopulmonary ressuscitation (CPR) may respond through rhythm transitions between different rhythms ventricular fibrillation (VF), ventricular tachycardia (VT), asystole (AS), pulseless electrical activity (PEA) and pulse generating rhythm (PR). Rhythm recognition is crucial to address adequate resuscitation efforts, and in this study we applied a deep neural network to classify ECG rhythms during cardiac arrest.

Artifact-free four second segments were extracted from 100 patients in out-of-hospital cardiac arrest. A convolutional neural network (CNN) was trained to discriminate between five cardiac arrest rhytm types. Experiments were conducted with increasing number of layers. For each model, training was repeated 10 times to explore variations in the results.

A five layer network provided the best performance with an accuracy of 80.3 (78.1,81.3)% (median(25th,75th quartiles)).

We have proposed a deep learning approach to automatically recognise five cardiac arrest rhythms common during resuscitation.

1. Introduction

During resuscitation of out-of-hospital cardiopulmonary arrest (OHCA), the patient is treated with chest compressions, ventilations and electroshock. Efficient data analysis of data from OHCA can help understanding the relationship between therapy and patient response. This knowledge can help us to improve the quality of the therapy, thus increasing the likelihood of survival [1].

Among the subsidary goals to improve efficiency our group and collaborators are working on, is the development of systems for automatic chest compression and ventilation detection, shock outcome prediction, compression artifact removal and rhythm interpretation. In our previous work on rhythm classification an approach with handcrafted features and various machine learning approaches were studied[2–4].

Patients undergoing CPR will respond through rhythm transitions between ventricular fibrillation (VF), ventricular tachycardia (VT), asystole (AS), pulseless electrical activity (PEA) and pulse generating rhythm (PR) as illustrated in figure 1.

In this work we address rhythm interpretation with the aim to recognise the five rhythm types applying deep neural networks on the electrocardiogram (ECG) recorded during resuscitation.

2. Materials and Methods

The data originates from a study on quality in cardiopulmonary resuscitation (CPR) from the period between March 2002 and September 2004). A subset of the data collected from the original study[1, 5] are used in this project. Only the ECG signals, annotatated for rhythm, is used. The signal was downsampled by a factor of two to 250 Hz. Four second segments of ECG without artifacts from CPR were collected from 100 different patients. This totals to 2833 segments containing 423 AS, 912 PEA, 689 PR, 643 VF and 166 VT [6].

The dataset was split into training, test and validation

 ISSN: 2325-887X DOI: 10.22489/CinC.2023.304

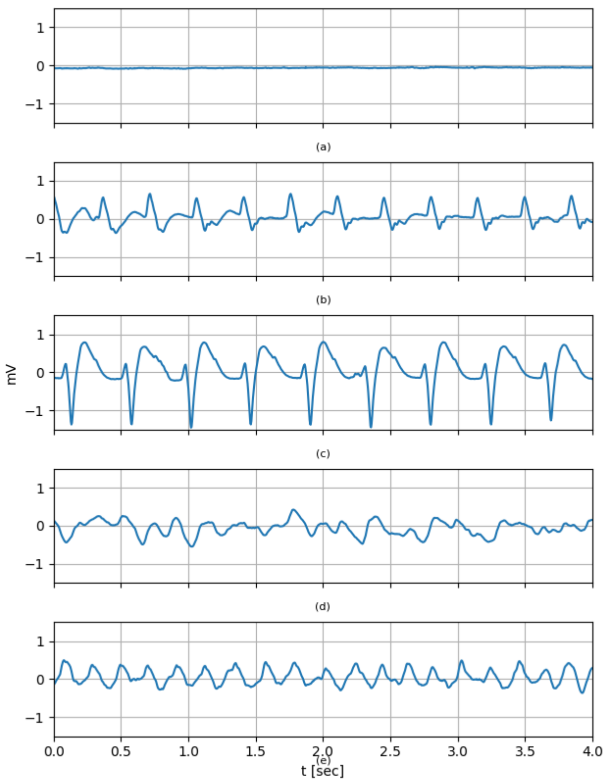


Figure 1: Cardiac rhythms from resuscitation: (a) ASY, (b) PEA, (c) PR, (d) VF, (e) VT.

sets in a stratified manner with 2281, 256 and 296 cuts respectively.

In this study, experiments are conducted to design a one-dimensional convolutional neural network (CNN) to extract features to classify the heart rhythms. A densely connected neural network classifier is attached to the CNN based feature extractor.

The filter kernel is used in the convolution of the input signals. The input signal will be longer than the output signal, so zero padding can be used to get equal lenghth on input and output. This can maintain the information at the boundaries and possibly increase the performance[7].

Pooling reduces the size of the output signal so we get a feature representation that is less sensitive to localisation of events in the input signal.. Pooling can be done in several manners - average and max pooling. Average pooling calculates the average value of the feature representation. Max pooling, on the other hand, calculates the maximum value[8].

A preliminary study was conducted, running with only one repetition for each training. The results from this study concluded that zero padding to the input should be used, and pooling at the end of each CNN-layer. It was also found that it was beneficial to gradually increase the number of filters for layers deeper into the model.

We have used this setup, repeating each model, for one to six layers, training and testing 10 times. For each model performance is measured in terms of total accuracy (tAcc), sensitivity (Se), specificity (Sp) and positive predictive value (PPV). The results are presented with median value and the lower and upper quartiles, 50 (25, 75) percentiles.

3. Results

Figure 2 shows the validation accuracy and loss for both the one-layer and the five-layer models.

Comparing the curves for the two types of models one can see that the validation accuracies has increased and the loss decreased with the increase in number of layers. The training accuracies and loss are more slimilar, with the one layer model slightly higher accuracies and lower losses.

The results from the experiments are shown in table 1, where Se, Sp, PPV and tAcc are provided.

The five layer model clearly shows the best performance.

Figure 3 shows the confusion matrices with mean values over the 10 repeated experiments for both the one layer and the five layer models.

The figure shows that the five-layer models are able to predict ASY, PGR and VF fairly well. PEA is problematic, as it is confused with PR. There are few cases of VT which the algorithm confused with VF. The five-layer model also has smaller off-diagonal values in ths confusion matrix.

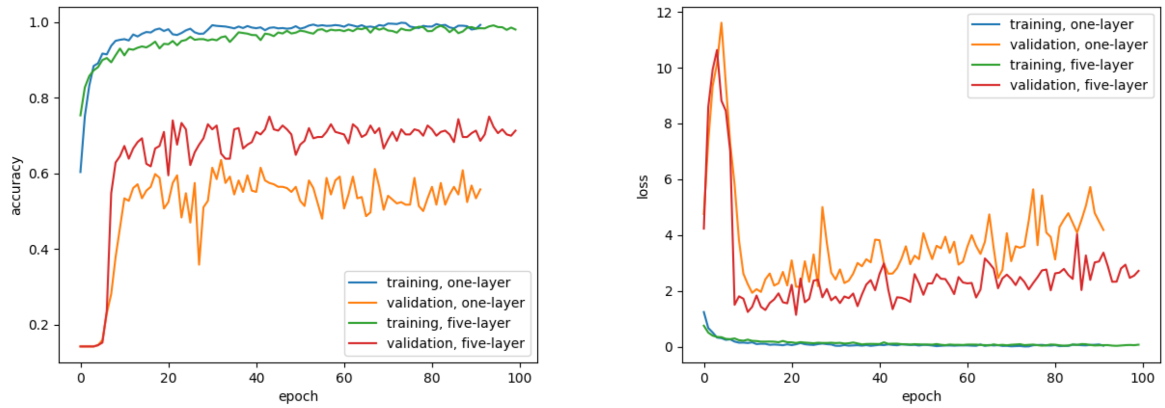
4. Discussion

As expected, the performance increased as new layers were added to the models. The total accuracies increased from 71.1% via 72.3%, 72.3%, and 75.2% for one to four layers until it peaked at 80.3% for five layers. A six layer model was also trained for which the total accuracy dropped to 72.3 (71.5,73.1)%. These results indicates that a five-layer model seems to be the optimal choice deemed from the current experiments. In the thouroghly conducted study by Krasteva et al where one dimensional CNNs were trained for cardiac rhythm recognition, a five-layer model was found to have the highest performance[9].

The confusion matrices in figure 3 also shows that the degree of confusion has decreased as the off diagonal numbers representing misclassification has lower values when comparing the one layer and the five-layer models.

In the repetition of the training of each model, the results vary for each training. This is due to the random initialisation of the neural network weights. The current experiments were run without setting any seed to control the randomisation.

From the accuracy and loss curves one gets the impression that the models are learning, and that the one layer



(a)

(b)

Figure 2: Training and validation data accuracies (a) and loss (b) for the one- and five-layer CNN models

		One-layer	Two-layer	Three-layers	Four-layer	Five-layer
AS	Se	100 (100,100)	100 (98.0,100)	100 (97.4,100)	100 (100,100)	100 (100,100)
	Sp	97.5 (96.8,98.1)	97.2 (97.2,97.7)	98.2 (97.4,98.6)	98.2 (97.4,98.6)	**98.2 (97.7,98.6)**
	PPV	87.4 (84.4,90.0)	86.4 (86.4,87.8)	90.4 (86.9,92.5)	90.5 (86.9,92.6)	**90.5 (88.4,92.7)**
PEA	Se	56.3 (50.6,59.8)	56.9 (52.3,59.5)	54.6 (53.2,58.3)	55.2 (52.0,58.9)	**64.9 (56.0,68.7)**
	Sp	94.1 (91.9,94.7)	93.2 (92.9,94.1)	93.8 (91.9,95.3)	**95.3 (94.7,95.7)**	94.7 (91.4,97.5)
	PPV	81.4 (79.1,83.7)	80.7 (79.0,83.7)	81.6 (78.7,84.8)	85.8 (84.4,86.8)	**86.1 (80.4,92.2)**
PR	Se	71.0 (65.3,75.4)	73.4 (62.9,75.4)	74.2 (67.3,76.6)	86.3 (81.5,88.7)	**92.7 (84.3,96.4)**
	Sp	87.1 (84.5,88.4)	86.3 (85.7,87.5)	86.3 (84.1,87.5)	86.9 (85.3,88.0)	**87.4 (83.8,90.5)**
	PPV	62.4 (58.9,64.4)	61.3 (60.8,63.6)	61.2 (59.1,63.6)	66.7 (65.5,67.8)	**68.8 (64.9,74.0)**
VF	Se	85.6 (75.4,92.4)	83.1 (73.3,86.0)	87.3 (83.9,90.7)	89.0 (73.3,93.2)	**87.3 (83.5,92.4)**
	Sp	90.6 (88.3,94.5)	93.1 (90.5,94.4)	90.4 (89.5,92.0)	95.4 (92.1,97.0)	**97.0 (97.0,98.0)**
	PPV	73.6 (69.3,79.8)	79.4 (71.2,81.2)	72.6 (69.5,77.0)	80.9 (76.8,89.7)	**90.2 (89.4,92.2)**
VT	Se	60.0 (52.5,60.0)	60.0 (60.0,60.0)	60.0 (52.5,60.0)	**60.0 (60.0,70.0)**	60.0 (60.0,60.0)
	Sp	96.1 (92.8,98.2)	93.9 (92.5,97.0)	97.0 (96.0,98.0)	95.9 (94.4,98.2)	**97.6 (96.1,98.3)**
	PPV	38.5 (27.7,54.2)	28.6 (23.0,45.6)	45.6 (38.1,50.0)	38.9 (30.8,57.5)	**48.1 (41.6,58.6)**
tAcc		71.1 (70.5,73.3)	72.3 (69.9,74.0)	72.3 (71.3,75.0)	75.2 (72.3,78.9)	**80.3 (78.1,81.3)**

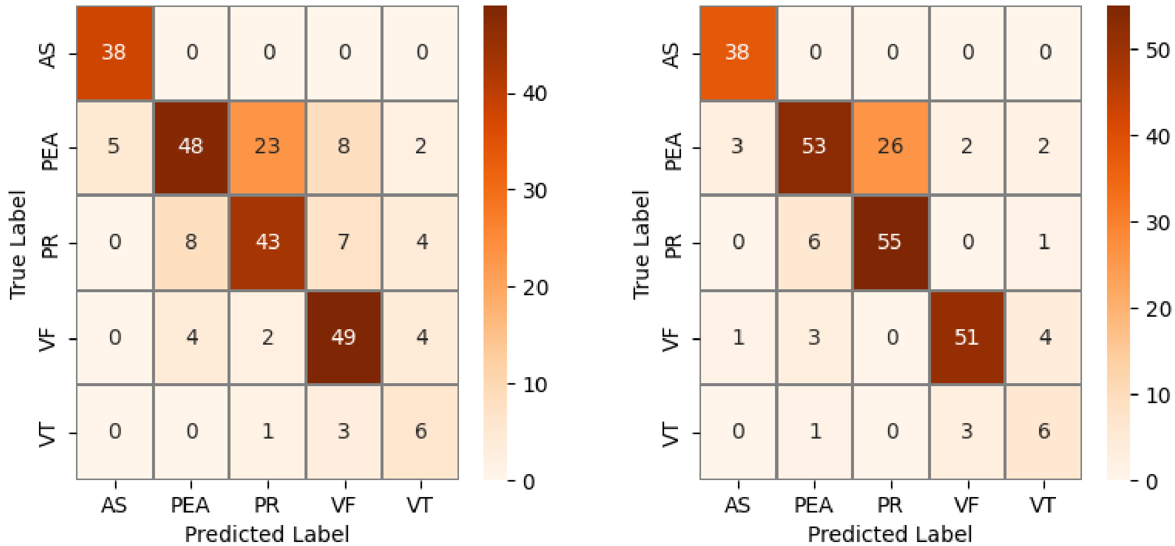
Table 1: Results for CNNs with varying architectures.

model is overtrained as the deviation to the validation accuracies and losses are much larger than for the five-layer models

The results in this study is also comparable to those achieved by B. Rad et al who reported a tAcc of 78.5% and Se (PPV) for AS, PEA, PR, VF, and VT of 88.7% (91.0%), 68.9% (70.4%), 65.9% (69.0%), 86.2% (83.8%), and 78.8% (72.9%), respectively[3].

It must be noted that the current study was conducted on a subset of (100 episodes) of the data set of 298 episodes used by B. Rad et al. In the future we plan to prepare a larger data set for experimentation with deep learning models.

It is also worth noting that only artifact free ECG cuts were used in this study. At some time we will have to include cuts with artifacts from chest compressions to see if deep learning models can learn to classify such ECG cuts. Both approaches with combination of artifact removal by adaptive filters[4, 10] or without such filtering can be considered.



(a)

(b)

Figure 3: Confusion matrix for (a) the one-layer CNN and (b) the five-layer CNN

5. Conclusion

A CNN based model has been demonstratet to be able to classify the rhythms typically seen during out-of-hospital resuscitation. The dataset was limited to 100 episodes, so this must be considered a preliminary study and the results should be confirmed in an expanded data set.

References

[1] Kramer-Johansen J, 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(3):283–292.

[2] Rad AB, Engan K, Katsaggelos AK, Kvaløy JT, Wik L, Kramer-Johansen J, Irusta U, Eftestøl T. Automatic cardiac rhythm interpretation during resuscitation. Resuscitation 2016;102:44–50.

[3] Rad AB, Eftestøl T, Engan K, Irusta U, Kvaløy JT, Kramer-Johansen J, Wik L, Katsaggelos AK. ECG-based classification of resuscitation cardiac rhythms for retrospective data analysis. IEEE Transactions on Biomedical Engineering 2017;64(10):2411–2418.

[4] Rad AB, Eftestøl T, Irusta U, Kvaløy JT, Wik L, Kramer-Johansen J, Katsaggelos AK, Engan K. An automatic system for the comprehensive retrospective analysis of cardiac rhythms in resuscitation episodes. Resuscitation 2018; 122:6–12.

[5] 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(3):299–304.

[6] Lasa H, Irusta U, Eftestøl T, Aramendi E, Rad AB, Kramer-Johansen J, Wik L. Multimodal biosignal analysis algorithm for the classification of cardiac rhythms during resuscitation. In 2020 Computing in Cardiology. IEEE, 2020; 1–4.

[7] Brownlee J. How do convolutional layers work in deep learning neural networks. Machine Learning Mastery 2020; 17.

[8] Brownlee J. A gentle introduction to pooling layers for convolutional neural networks. Machine Learning Mastery 2019;22.

[9] Krasteva V, Ménétré S, Didon JP, Jekova I. Fully convolutional deep neural networks with optimized hyperparameters for detection of shockable and non-shockable rhythms. Sensors 2020;20(10):2875.

[10] Irusta U, Ruiz J, Ruiz de Gauna S, Eftestøl T, Kramer-Johansen J. A least mean square filter for the estimation of the cardiopulmonary resuscitation artifact based on the frequency of the compressions. IEEE Transactions on Biomedical Engineering 2009;56:1052–1062.

Address for correspondence:

Trygve Eftestøl
University of Stavanger, N-4036 Stavanger, Norway
tel./fax: +47-5183-2035/1750
trygve.eftestol@uis.no