Artificial Intelligence In Medicine 110 (2020) 101963

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

Predicting defibrillation success in out-of-hospital cardiac arrested patients: Moving beyond feature design

Marija D. Ivanović [a,*], Julius Hannink [b], Matthias Ring [b], Fabio Baronio [c], Vladan Vukčević [d], Ljupco Hadžievski [a,e], Bjoern Eskofier [b]

[a] Vinca Institute of Nuclear Scientists, University of Belgrade, Belgrade, Serbia
[b] Machine Learning and Data Analytics Lab, Department of Computer Science, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany
[c] CNR and Department of Information Engineering, University of Brescia, Brescia, Italy
[d] School of Medicine, University of Belgrade, Belgrade, Serbia
[e] Diasens, Belgrade, Serbia

ARTICLE INFO

Keywords:
Ventricular fibrillation (VF)
Defibrillation
Shock outcome
Deep learning
Convolutional neural networks (CNN)

ABSTRACT

Objective: Optimizing timing of defibrillation by evaluating the likelihood of a successful outcome could significantly enhance resuscitation. Previous studies employed conventional machine learning approaches and hand-crafted features to address this issue, but none have achieved superior performance to be widely accepted. This study proposes a novel approach in which predictive features are automatically learned.

Methods: A raw 4s VF episode immediately prior to first defibrillation shock was feed to a 3-stage CNN feature extractor. Each stage was composed of 4 components: convolution, rectified linear unit activation, dropout and max-pooling. At the end of feature extractor, the feature map was flattened and connected to a fully connected multi-layer perceptron for classification. For model evaluation, a 10 fold cross-validation was employed. To balance classes, SMOTE oversampling method has been applied to minority class.

Results: The obtained results show that the proposed model is highly accurate in predicting defibrillation outcome (Acc = 93.6 %). Since recommendations on classifiers suggest at least 50 % specificity and 95 % sensitivity as safe and useful predictors for defibrillation decision, the reported sensitivity of 98.8 % and specificity of 88.2 %, with the analysis speed of 3 ms/input signal, indicate that the proposed model possesses a good prospective to be implemented in automated external defibrillators.

Conclusions: The learned features demonstrate superiority over hand-crafted ones when performed on the same dataset. This approach benefits from being fully automatic by fusing feature extraction, selection and classification into a single learning model. It provides a superior strategy that can be used as a tool to guide treatment of OHCA patients in bringing optimal decision of precedence treatment. Furthermore, for encouraging replicability, the dataset has been made publicly available to the research community.

1. Introduction

Ventricular fibrillation (VF) represents the most frequent initial rhythm associated with sudden cardiac death. Restoration of organized electrical activity after cardiac arrest caused by VF can be re-established using electrical defibrillation in conjunction with cardiopulmonary resuscitation (CPR) [1,2]. However, despite major efforts to improve outcome from cardiac arrest, survival rate after out-of-hospital cardiac arrest (OHCA) remains at 9 % [3] whereas after in-hospital cardiac arrest (IHCA) survival is 22.3 % [4].

In short duration VF, immediate defibrillation has proven its benefit for increasing the survival rate [5,6]. However, the precedence of immediate defibrillation over CPR in case of a prolonged duration of untreated cardiac arrest has changed over the years [7–9]. In the European Resuscitation Council Guidelines from 2005 the routine delivery of a specific period of CPR before shock delivery was recommended if the first shock should be delivered after 5 min of the collapse. But, in 2010 European Resuscitation Council Guidelines asserted the insufficient evidence to support or refute the survival benefit of a pre-specified period of CPR before shock delivery wherefore immediate

* Corresponding author at: Vinca Institute of Nuclear Scientists, Mike Petrovica Alasa 12-14, 11001 Belgrade, Serbia.
E-mail address: marijap@vin.bg.ac.rs (M.D. Ivanović).

https://doi.org/10.1016/j.artmed.2020.101963
Received 11 February 2019; Received in revised form 23 August 2020; Accepted 22 September 2020
Available online 7 October 2020
0933-3657/

defibrillation was called again. This led to the application of the same treatment protocol for every VF patient – immediate defibrillation as soon as a defibrillator became available. A prolonged duration of untreated VF can cause a state of acidosis [10]. Upon VF rhythm detection, immediate defibrillation of a heart in that state reduces the probability of successful defibrillation, whereas some studies suggest that, in that state, a period of CPR prior to defibrillation may optimize myocardial perfusion and improve survival rate [1,11,12]. Moreover, repetitive unsuccessful defibrillation attempts are injurious to an already ischemic myocardium and decreases post resuscitation myocardial function and survival [13]. Therefore, optimizing timing of defibrillation by evaluating the likelihood of a successful outcome using VF waveform analysis, and thus determining the priority of receiving immediate defibrillation or alternative therapy, such as CPR or drug administration, could potentially lead to more benefit than applying the same treatment protocol.

Over the past few decades, a plethora of classification strategies have been applied to ECG signal classification and arrhythmia analysis (see for example [[14] and references therein] [15,16],). In particular, a description of the efforts in the prediction of the defibrillation outcome can be found in [[17] and references therein], [18–38]. The strategies available in the existing works along with the validation of their results are given in Table A1 in Appendix. The common characteristic of all of previously reported strategies is to follow conventional machine learning (ML) approaches which implies feature engineering and classification. In the feature engineering stage, hand-crafted" features were designed using time domain, frequency domain, time-frequency domain and/or non-linear dynamical analysis of the pre-shock VF signal. In the classification stage a single VF feature or combination of features was input into various classifiers for defibrillation outcome prediction. Even though none of the previously reported strategies have achieved superior performance to be widely accepted, the amplitude spectrum area (AMSA) has been demonstrated as one of the most accurate predictors for successful defibrillation [17–29,31].

Unlike conventional approaches, deep learning has the capacity to learn multiple levels of representation for a given data [39]. The main advantage over conventional methods is that a deep learning approach does not need to be trained using predesigned features. The input layers of a hierarchical architecture automatically discover useful features from raw signals, based on which the output layers perform classification. Therefore, deep learning methods typically lead to higher performance. There are two common types of deep learning methods. The first type includes deep discriminative models, such as deep neural networks (DNNs) [40], convolutional neural networks (CNNs) [39], recurrent neural networks (RNNs) [39], whereas the second type incorporates generative models such as restricted Boltzmann machines (RBMs) [41], deep belief networks (DBNs) [42] and deep Boltzmann machines (DBMs) [43].

CNNs are typical deep learning architectures that have revolutionized computer vision. Inspired by the feature learning capacity for image classification, in this paper we proposed adaptation of the CNNs applied in computer vision to the time series data for automatic feature learning in predicting defibrillation success. Several publications have appeared in recent years documenting utilization of CNNs on ECG signals. Most of these publications were related to two important issues of ECG classification: ECG beat classification [44–47] and ECG arrhythmia detection [48–52]. Additionally, CNN were also used for biometric human identification [53] and myocardial infraction detection [54].

This study proposes a novel employment of CNNs on ECG signals – defibrillation outcome prediction, and provides a superior strategy that can be used as a tool to guide treatment of OHCA patients in bringing optimal decision of precedence treatment (immediate defibrillation vs. CPR). The main purpose of the applied strategy is to determine the optimal timing of shock delivery by evaluating the probability of successful defibrillation, so that if the shock has a high likelihood of defibrillation success, a defibrillation shock should be delivered. Otherwise, unnecessary shocks should be avoided and instead CPR or drug administration should be employed. It is important to note that the survival rate of VF patients is decreasing with the duration of untreated VF. Therefore, the high sensitivity of the successful defibrillation attempts is the most important characteristic of the defibrillation outcome prediction algorithm. In this direction this study is focused to evaluate CNNs in learning useful data representations from the raw VF waveforms to improve prediction of successful shocks beyond conventional ML algorithms with hand-crafted features. To the best of our knowledge this is the first study to address defibrillation outcome prediction using features learned from raw data. Furthermore, for encouraging replicability, the dataset has been made publicly available to the research community [55].

2. Methods

2.1. The study data

This study is a retrospective analysis of ECG recordings immediately prior to the first countershock in 260 adult patients with sudden out-of-hospital cardiac arrest in Brescia, Italy. The database is publicly available at [55] for the purpose of replicability. It contains only the first defibrillation attempts in order to exclude the effects of administration of any drugs and/or CPR procedures. The data was acquired between 2006 and 2009 following the 2005 European CPR guidelines [56]. The data and all relevant demographic information were recorded according to the Utstein guidelines [57] and by using a semiautomatic Heartstart 3000 defibrillator (Laerdal Medical, Stavanger, Norway). Details of data collection have been described in our previous paper [18]. Ethical approval of this study was obtained through the ethical committee of Brescia (application number NP2753).

Three experienced cardiologists were independently examining 1 min post-shock ECGs and annotated each as successful, unsuccessful or indeterminable. The decision was made based on whether the shock resulted in return of an organized electrical activity (ROEA) as defined in [18], VF persisted after shock or it was not possible to ascertain. Based on the cardiologists' annotations, 9 signals were considered indeterminable and discarded from the analysis. The other 251 valid first shocks were categorized as successful (ROEA) or unsuccessful (no-ROEA) in that the majority of doctors' decisions was taken.

2.2. Convolutional neural networks

There are 4 major ideas that lay the foundation of CNN: local connections, weight sharing, pooling and the use of many layers [39]. The architecture of a typical CNN consists of a series of convolutional layers, pooling layers and fully connected layers.

The convolution is one of the fundamental operations of a convolutional layer. The role of the convolutional layer is to detect local links of features from the preceding layer. Each neuron in convolution layer is connected to a small local region (receptive field) of the preceding layer in order to extract features. The length of a receptive field is called the kernel size. To go through the entire region of the preceding layer, the striding operation is performed. So, neurons at different locations actually share the same weights, and therefore are trained to detect the same feature but on different locations. The result of convolution operation is then passed through a non-linear activation. Nowadays, the most commonly used non-linearity is rectified linear unit (ReLU) activation. The output represents a feature map [58]. To be able to detect more than just a single feature in a convolutional layer, additional filters needs to be learned. Hence, a complete convolutional layer comprise several feature maps, where different feature maps use different filters.

By applying strided convolution the information contained near the edges can be neglected along with output size shrinking. To avoid loss of information on the edges, it is possible to perform padding before applying convolutions.

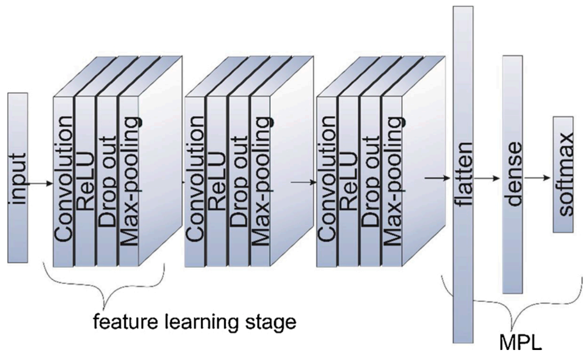

Fig. 1. CNN architecture.

Table 1
The details of the CNN structure.

layer	number of parameters	output shape	kernel size	stride
input	0	(b,1000,1)	–	–
feature learning stage I - convolution	256	(b,1000,16)	15	1
feature learning stage I - max pooling	0	(b,500,16)	2	2
feature learning stage II - convolution	7712	(b,500,32)	15	1
feature learning stage II - max pooling	0	(b,250,32)	2	2
feature learning stage III - convolution	30784	(b,250,64)	15	1
feature learning stage III - max pooling	0	(b,125,64)	2	2
flatten	0	(b,8000)	–	–
dense	8193024	(b,1024)	–	–
softmax	2050	(b,2)	–	–

*b – batch size.

The role of a pooling layer is to emphasize the most significant features. A pooling layer is commonly inserted between successive convolutional layers [58]. This layer simplifies the information in the output from the convolutional layer by computing the maximum or the average of the features among the adjacent neurons in the preceding layer. The stride step in the pooling layer is usually the same as the length of the pooling window. Since, padding is usually omitted in the pooling layer, this layer decrease the dimension of the convolved extracted features and increase the speed of the learning mechanism.

In general, the earlier convolutional layers of the CNN learns lower-level features. The higher-level features are obtained by composing the lower-ones in a manner that the higher the number of the convolutional layers, the more complex features can be extracted.

After a set of convolutional and pooling layers, one or more fully connected layers are used. In the fully connected layer the neurons are connected to all the neurons in the preceding layer.

2.3. Research methodology and network architecture

A 4 s episode during hands off time and immediately prior to first defibrillation shock on each patient was selected for prediction analysis. For the purpose of suppression of residual baseline drift, power line interference and high frequency noise, every signal was band-pass filtered with a lower cutoff frequency of 0.5 Hz and upper cutoff frequency of 48 Hz.

In our original dataset, 195 VF signals were labeled as no-ROEA and only 56 as ROEA. In order to address the problem of imbalanced training data, the SMOTE oversampling method has been applied [59]. In this manner, the ROEA class, as minority class, has been oversampled by creating new synthetic" examples interpolated between original minority class examples. Therefore, a total of 390 signals, 195 instances in each class, were included in the analysis. Additionally, each instance is scaled to have values between 0 and 1 with min-max normalization to address the problem of amplitude scaling before feeding it into the 1D CNN.

The feature extraction, feature selection and classification, steps that have to be explicitly defined in conventional approach, are embedded in one model and learnt jointly in deep learning. For this purpose, the traditional 2D CNN was modified and applied to the 1D time series classification task at hand.

A raw signal of 1000 samples length (4 s signal sampled with sampling rate 250 Hz) was fed into a 3-stage CNN feature extractor. Each feature learning stage (convolutional block) was composed of 4 components: convolutional, ReLU activation, dropout and max-pooling (Fig. 1). In each convolutional block the signals were convolved with 1-D kernels having filter length of 15 samples with stride of 1. The number of feature maps used in blocks were 16, 32 and 64, respectively (Table 1). After convolution, the non-linearity was introduced into the neural network by applying ReLU activation. In that manner, the network was allowed to learn more complex structures with preventing saturations of the gradients [60]. Additionally, during the training procedure, dropout with probability 0.6 for keeping individual nodes was applied. The dropout technique randomly drops" some nodes in each training iteration. It has been shown that this technique improves the generalization capability by preventing nodes from co-adapting too much [61]. The purpose of max-pooling was to subsample the feature map by using a pool kernel with size of 2 samples and a stride of 2, therefore making it more robust to small variations of previously learned feature maps [39]. At the end of the feature extractor, the output was flattened and fed into a subsequent fully connected multi-layer perceptron (MLP). The output layer utilized a softmax activation function for classification.

2.4. Model building

For model evaluation, a 10 fold cross-validation was employed [62]. Since we are examining only the first shocks (one ECG signal per patient), the cross-validation was performed over patients. The CNN model was trained by optimizing categorical cross-entropy objective function utilizing Adam as gradient based optimization method. The hyper-parameters (learning rate, $\beta1$, $\beta2$ and ε) were set as suggested in the original paper [63]. The data in the training set were randomly shuffled and divided into mini-batches with size 100 for faster convergence. Training was performed from scratch in 150 epochs by initializing the weights of the convolutional layers using a Xavier normal initializer [64]. After training the network, the whole validation set was propagated through the network for evaluation of performance metrics. Eventually, the reported performances were obtained by averaging the performance metrics recorded in each fold of the cross-validation.

The proposed CNN algorithm was developed in Python using the Keras open source neural network library. It was trained on a PC computer with Intel(R) Core(TM) i5-4590 CPU 3.3 GHz processor and 16 GB of RAM. The average time needed to complete one epoch of training was approximately 1.4 s. During evaluation, a single VF signal was analysed in 3 ms.

2.5. Evaluation metrics

Classification performance is evaluated using the 5 standard metrics: classification accuracy (Acc), sensitivity (Sen), specificity (Spec), precision (P), and negative predictive value (NPV). Acc provides a simple way of measuring classifier's overall performance. However, in certain situations, such as having imbalanced classification problem [65] or having unequally important classes (as in this study), the model with the highest accuracy may not be the optimal model. To provide comprehensive assessments of classifier's performance, we have included other

Table 2
Performance of the proposed model.

metrics	Acc [%]	Sen [%]	Spec [%]	P [%]	NPV [%]
mean	93.6	98.8	88.2	89.0	99.1
std	2.6	3.8	2.9	3.8	2.7
95 % CI	91.7–95.5	96.1–100	86.1–90.3	86.2–91.7	97.1–100

*CI – confidence interval.

metrics that are specific to each class. Sen and Spec are measures of completeness, which tell how many examples of the positive or negative class were classified correctly. Whereas P and NPV are measures of exactness, which state of all examples classified as positive or negative how many are actually classified correctly. The metrics are defined as:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Sen = \frac{TP}{TP + FN} \tag{2}$$

$$Spec = \frac{TN}{TN + FP} \tag{3}$$

$$P = \frac{TP}{TP + FP} \tag{4}$$

$$NPV = \frac{TN}{TN + FN} \tag{5}$$

in which TP, TN, FP and FN denote the true positives, true negatives, false positives and false negatives, respectively. The shocks that resulted in ROEA were considered as positives.

3. Results

Table 2 summarizes the performance of the proposed model. The reported performances were obtained as a mean value from the 10 fold cross-validation. We could be 95 % confident that Acc of our model falls between 91.7 % and 95.5 %, Sen between 96.1 % and 100 %, Spec between 86.1 % and 90.3 %, P between 86.2 % and 91.7 % and NPV between 97.1 % and 100 %. The results show that the proposed model is highly sensitive in predicting successful defibrillations. Additionally, with the reported Spec our results would lead to avoiding 88.2 % of unsuccessful shocks and allowing an alternate therapy before defibrillation. Therefore, the proposed model could be considered as a safe and useful predictor for defibrillation decision, since it has a sensitivity higher than 95 % and a specificity much higher than 50 %, the level that was in 2005 postulated as the threshold for feature prediction implementation in defibrillators [24,30].

Deep-learning approaches are known to overfit easily due to the large number of parameters to be learned. By using 10 fold cross-validation and obtaining narrow 95 % confidence intervals, we demonstrated that our proposed model generalizes well and produces stable results on different training set distributions.

4. Discussion

In order to evaluate effectiveness of learned features for predicting defibrillation success, we compared our CNN model against traditional ML algorithms with the hand-engineered features on the same dataset. The results obtained by employing seven conventional ML algorithms (logistic regression, Naïve-Bayes, decision tree C4.5, AdaBoost M1, support vector machine (SVM), k nearest neighbor (kNN) and random forest (RF)) on this dataset were reported in our previous paper [18]. In both studies, 4 s episodes of the pre-shock signals were used and pre-processed in the same manner (band-pass filtering, instance normalization and SMOTE oversampling of minority class). We showed

Table 3
Models performance comparison on the reported dataset.

Model	Acc [%]	Sen [%]	Spec [%]	P [%]	NPV[%]
CNN	93.6	98.8	88.2	89.0	99.1
RF	82.8	82.8	82.8	82.8	82.8
kNN	81.8	81.8	81.8	81.8	81.8
SVM	81.5	81.5	81.5	81.5	81.5
AMSA	72.6	90.3	54.9	66.7	84.9

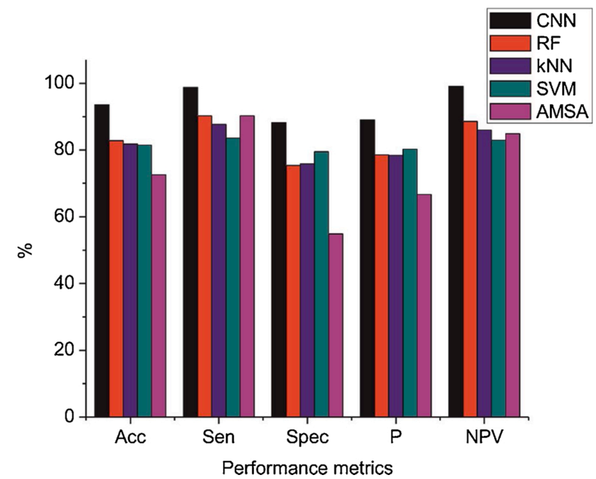

Fig. 2. Models performance comparison on the reported dataset.

that RF, kNN and SVM, with the best performing feature combination selected from 28 previously reported predictive features via a wrapper feature selection method, outperformed other algorithms using single feature or best performing combination. Therefore, the proposed CNN model has been compared to those three conventional ML algorithms (obtained from their best performing feature combination). This comparison is shown in Table 3 and Fig. 2. Additionally, since the medical community considers AMSA feature as a standard highly accurate predictor for successful defibrillation, we have also included this feature in comparison.

These results demonstrate the excellent feature learning capability of the CNN model for predicting defibrillation outcome. Additionally, since the learned features obtained superior results compared to the hand-crafted ones, it can be an indication that previously reported features are not sufficient to distinguish between successful and unsuccessful defibrillation. In order to use conventional ML approach for predicting defibrillation outcome with state-of-the-art performance, further feature engineering is needed to arrive at sufficiently descriptive data characteristics.

In the related literature on predicting the shock outcome, a variety of classification approaches using single feature and/or feature combinations have been proposed [[17] and references therein], [18–38]. However, to the authors' best knowledge there is no publicly available database that can be used for predicting defibrillation outcomes from pre-shock VF rhythms. Each study dealing with this issue in the literature uses its own dataset (Table A1). This imposes limitations to direct comparison among reported studies. For instance, a different pre-shock episode is selected, but the effect of window length on the performance has not been systematically investigated. The waveform design significantly differs among manufactures, but the reliability of VF waveform analysis with the use of different defibrillation waveforms still needs to be researched. The interpretation of the results may be biased due to different patient number, definition of defibrillation success, ratio between successful and unsuccessful defibrillations, (in)consideration of imbalanced learning problem and whether only first shocks or also subsequent shocks were considered.

A pre-shock episode of 1–12 seconds is usually selected, but the effect of window length on the performance has not been systematically investigated. The study of Jekova et al. [34] showed the decrease in performance with increasing the duration of ECG segment used to calculate the features (window length), while Strohmenger et al. [37] demonstrated a significant improvement in performance when a segment of 12 s were used instead of 3 s (Table A1). When selecting only the studies that use 4 s before delivering defibrillation shock Howe et al. [22] reported Acc of 81 %, Sen 100 %, Spec 95 % when the termination of VF was consider as successful and Acc of 75.9 %, Sen 86.2 %, Spec 100 % if the return of organized rhythm was consider as successful. Further, Wu et al. [23] obtained Sen of 90 % and Spec 86 %, Firoozabadi et al. [25] AUC of 83 %, Sen 90 % and Spec 63 %, Nakagawa [26] Sen 100 % and Spec 64 % for biphasic defibrillators, Gundersen et al. [29] and Eftestol el al [32]. reported only AUC of 87.7 %, and 80 %, respectively, Jekova et al. [34] achieved Acc of 69.9 %, Sen 58.8 % and Spec 77.6 %, Monsieurs et al. [36] Acc of 77 %, Spec 86 % and Spec 73.2 %, while Brown et al. [38] reported Sen of 100 % along with Spec of 47.1 %.

The waveform design significantly differs among manufactures, but the reliability of VF waveform analysis with the use of different defibrillation waveforms still needs to be researched. The aim of the study by Nakagawa et al. [26] was to evaluate the prognostic performance of AMSA in relation to waveforms of defibrillators; however their dataset contained only 3 successful cases when monophasic defibrillator was used (Table A1).

The interpretation of the results may be biased due to different patient number, ratio between successful and unsuccessful defibrillations as well as definition of defibrillation success. The largest set reported in the literature contained 3828 defibrillation attempts (1086 successful and 2742 unsuccessful) [20], while the smallest dataset contained only 83 shocks (17 successful, 64 unsuccessful and 2 excluded) [26] (Table A1). In the literature, there is no consistent definition of defibrillation success. The three most commonly used are return of organized electrical activity (ROEA), sometimes referred to as return to organized rhythm (ROR), return of spontaneous circulation (ROSC) and survival, being the most rigorous definition [17]. The effect of the different definitions for successful shocks can be explicitly seen in the study by Howe et al. [22] in which the authors reported Acc of 81 % along with Sen of 100 % and Spec of 95 % if the termination of the VF was considered a success, and Acc of 75.9 % along with Sen of 86.2 % and Spec of 100 % if the ROR was considered a success (Table A1). The studies that use ROEA (or ROR) as a definition of the successful defibrillation reported the Acc of 75.9 %, Sen 86.2 % and Spec 100 % [22] and Sen of 90 % along with Spec 86 % [23]. Both studies do not reach a threshold of 95 % Sen and 50 % Spec for implementation in defibrillators.

The amount of successful shock attempts in the database of the existing works has a wide range, from around 7% in [30] and [38] to 63 % in [22] (Table A1). In most cases, the amount of successful shocks does not exceed 30 %. The possible reason for such a high amount of successful shocks in [22] could be the inclusion of IHCA patients. The small amount of successful shocks rose an additional issue – dealing with the imbalanced dataset. However, to the authors' best knowledge only two studies addressed the issue of class imbalance", our previous work [18] by performing SMOTE method and [27] by using cost sensitive classification. Additionally, there are some studies like [33] where the authors used cost sensitive classification, but in order to achieve high sensitivity.

Whether only the first shocks or also the subsequent shocks were considered could have an influence on the reported performance as well. Thus, Watson et al. [31] reported an increase in sensitivity and decrease in specificity with the increase of maximum number of shocks per patients (Table A1). By considering the studies that investigate first shocks, the study by Gong et al. [21] reported AUC of 82.6 %, Wu et al. [23] Sen of 90 % and Spec of 86 %, Ristagno et al. [24] Acc, Sen and Spec of 73 % and Nakagawa et al. [26] Sen 100 % with Spec 64 %.

When creating a machine learning model one should be extremely careful not to accidentally share information between the training and test/validation sets. This phenomenon is known as data leakage, and often results in overly optimistic performance, since the model is evaluated on data "already seen" by the model. A common cause for data leakage is the selection of hyperparameters and/or features based on the same data that are later used for model evaluation. In the case of small datasets, which need to use cross-validation procedure since there is not enough data to be left for testing, a nested cross-validation should be applied, as reported in [18] and [27]. However, there are some studies like [22], where the authors use a cross-validation, without reporting nesting, for choosing the optimal σ parameter of SVM and evaluating the model performance. Having signals from a single patient, in both training and test sets, could also lead to overly optimistic results. For instance, in [20], the authors separated patients among training and validation set. On the other hand, in [35] the authors did not state that they pay attention not to have signals from the same patients in both training and test sets, leaving the suspicious of their reported performance (Sen 100 %, Spec 97.2 %).

Beside the proposed approach only 5 more approaches reported performance with at least 50 % specificity and 95 % sensitivity, required for their implementation in defibrillators [22,26,30,31] and [35]. Since, the Spec in [26,30] and [31] are below 70 %, the CNN model shows superiority over them. The model proposed by Howe et al. [22] would be considered as safe if the decision was the termination of VF. However, by choosing the decision as in this study, this model would not be considered safe. The only model that outperforms the CNN model was suggested by Podbregar et al. [35]. Nonetheless, as we stated above there is a possibility of data leaking related to this study, and possibly unrealistically high performance.

In the study reported by Acharya et al. [49], the authors designed CNN model to automatically identify shockable (ventricular flutter, tachycardia, fibrillation) and non-shockable ventricular arrhythmias. They achieved a high performance with Acc 93.18 %, Sen 91.04, Spec 95.32 % and P 95.11 %. From these results, they concluded that their proposed CNN model has a great potential to save life and reduce damage inflicted to the heart muscle by unsuccessful defibrillation. Even though the classification of shockable vs. non-shockable arrhythmias is important issue, there are evidences showing that not all patients in VF (or other shockable arrhythmias) might benefit from being treated in the same manner with immediate defibrillation. The studies [1] and [12] indicate that after 4−5 min of VF initialization performing the CPR with chest compression before delivery of a defibrillation can improve the likelihood of restoring an organized electrical activity and increase survival rate. Therefore, to save lives and reduce damage inflicted to the heart muscle by unsuccessful defibrillation, it is not sufficient to correctly distinguish between shockable and non-shockable rhythm, but to be able to successfully predict the probability of shock outcome.

According to the European Resuscitation Council Guidelines for Resuscitation 2010 and 2015 [8,9], the defibrillation shock should not be delayed longer than necessary to establish the need for defibrillation. Following these guidelines it is preferable not to miss to deliver a shock that will lead to return of an organized electrical activity, even at the cost of delivering an unnecessary shock that could be avoided (having Sen 100 % and Spec 0%). Our results demonstrated that by applying the proposed model 98.8 % of the successful shocks will be immediately delivered while 88.2 % of unsuccessful shocks will be avoided allowing CPR or drug administration to be performed first. This indicates a potential of the proposed model to be a tool to guide resuscitation protocols with respect to the condition of the patient. To the best of our knowledge, none of the existing works have addressed the issue of predicting defibrillation outcome using features learned from raw data.

Few limitations of the proposed study should be pointed out. These limitations are mostly related to the small size of database. The total number of defibrillation shocks resulting in ROEA was relatively low and approximately four times lower in comparison with the number of

Table A1
Summary of previous works on predicting defibrillation outcome.

Ref/ Author/ year	Database: number of defibrillations	Signal duration [s]	Classification strategy	Performance [%]	Balance classes
Ivanovic et al. 2019 [18]	260 from 260 patients: ROEA: 56, no-ROEA: 195, ex: 9	4	Logistic regression Naïve Bayes Decision tree C4.5 Ada Boost SVM with GRBF Nearest neighbor Random Forest	**Nested 10 fold cross-validation**: A:72.6, Se:73.3, Sp: 71.8, P: 72.2, NPV: 72.9 A:72.1, Se:69.7, Sp: 74.4, P: 73.1, NPV: 71.1 A: 73.6, Se: 82.6, Sp: 64.6, P: 70, NPV: 78.8 A:73.3, Se:81, Sp: 65.6, P:70.2, NPV: 77.6 A:81.5, Se:83.6, Sp: 79.5, P:80.3, NPV: 82.9 A:81.8, Se:87.7, Sp: 75.9, P: 78.4, NPV: 86 A:82.8, Se:90.3, Sp: 75.4, P:78.6, NPV: 88.6	SMOTE
Chicote et al. 2016 [19]	419 from 163 patients: suc: 107, unsuc: 312	5	feature threshold selection based on ROC curve SVM with LOPCV	Se: 90, Sp: 55.9, AUC: 81.6 Se: 80.4, Sp: 76.9, P: 54.4, NPV: 92	NO
He et al. 2015 [20]	3828 from 1617 patients: suc: 1086, unsuc: 2742 Training set: 2447 from 1050 patients suc: 641, unsuc: 1806 Validation set:1381 from 567 patients suc: 428, unsuc: 953	2.048	feature threshold selection based on ROC curve Logistic regression Neural network SVM with GRBF	**Validation set:** A: 80.8, Se: 79.6, Sp: 81.4, P: 67, NPV: 89.3 A: 79.6, Se: 79.6, Sp: 79.6, P: 65, NPV: 89.1 A: 80.9, Se: 80.9, Sp: 80.9, P: 66.8, NPV:89.9 A: 78, Se: 71.3, Sp: 80.1, P: 53, NPV: 89.9	NO
Gong et al. 2015 [21]	159 from 159 patients: suc: 61, unsuc: 98	2.05	feature threshold selection based on ROC curve	AUC: 82.6, 95% CI: 76–89.1	NO
Howe et al. 2014 [22]	115 from 41 patients: OHCA: 78%, IHCA: 22% VF_T:73, no- VF_T: 42 ROR: 55, no-ROR: 60	4.1	SVM with GRBF	**2 fold cross-validation** VFT vs. no-VFT:A:81, Se:100, Sp:95, P:94.4, NPV:100 ROR vs. no-ROR: A: 75.9, Se: 86.2, Sp: 100, P: 86.7, NPV: 81.8	NO
Wu et al. 2013 [23]	350 from 350 patients: ROR: 134, no-ROR: 216	4.1	feature threshold selection based on ROC curve	Se: 90, Sp: 86, P: 80, NPV: 93	NO
Ristagno et al. 2013 [24]	1260 from 609 patients: suc: 316, unsuc: 944 578 first shocks: suc: 156, unsuc: 422	2.05	feature threshold selection based on ROC curve	All DF attempts:A: 78, Se: 78, Sp: 78, P: 54, NPV: 91 First shocks: A: 73, Se: 73, Sp: 73, P: 50, NPV: 88	NO
Firoozabadi et al. 2013 [25]	469 from 116 patients: ROSC: 49, no-ROSC: 420	4	feature threshold selection based on ROC curve	Se: 90, Sp: 63, AUC: 83	NO
Nakagawa et al. 2013 [26]	83 from 83 patients biphasic:ROSC: 14, no-ROSC:43, ex: 2 monophasic:ROSC: 3, no-ROSC: 21	4.096	feature threshold selection based on ROC curve	biphasic: Se: 100, Sp: 64, P: 42, NPV: 100 monophasic:Se: 100	NO
Shandilya et al 2012 [27]	90 from 57 patients: ROSC: 34, no-ROSC: 56 test set: 8	7.8	SVM Decision stump	**Nested 10 fold cross-validation**: A: 82.2, ROC AUC: 85 Se: 44.1, Sp: 77.2, AUC: 60.9	Cost sensitive
Endoh et al. 2010 [28]	415 from 152 patients: suc: 69, unsuc: 164, ex:182	1.0 and 5.12	Stepwise multiple logistic regression	Se: 76.8, Sp: 62.8 AUC: 77, 95% CI: 70-83	NO
Gundersen et al. 2008 [29]	530 form 86 patients: ROSC: 64, no-ROSC: 466	4	Mixed effects logistic regression	AUC: 87.7	NO
Neurauter et al. 2007 [30]	1077 from 197 patients: ROSC: 60, no-ROSC: 710, ex: 307 Training + validation: 483 ROSC:, 42, no-ROSC: 441 Test: 287: ROSC:18, no-ROSC: 269	2.5	feature threshold selection based on ROC curve Neural networks for combination of features	**test set**: single feature: Se: 95.2, Sp: 55.3, ROC AUC: 84.2 Combination of features: ROC AUC: 61-86	NO
Watson et al. 2006 [31]	878 from 110 patients: 17 patients had more than 15 shocks, 2 patients had over 30 shocks First shocks: ROSC: 8, no-ROSC: 102	at least 10	feature threshold selection based on ROC curve	**cross-validation:** Max 3 shocks/patient: Se: 95±4, Sp: 66±4 Max 6 shocks/patient: Se: 97±2, Sp: 61±4 Max 9 shocks/patient: Se: 98±2, Sp: 56±1	NO
Eftestol et al. 2005 [32]	589 from 136 patients: ROSC: 82, no-ROSC: 507	4	feature threshold selection based on ROC curve	**Test set**: AUC ROC: 80	NO
Watson et al. 2004 [33]	868 ROSC: 87, no-ROSC: 781	10	Bayesian non-parametric techniques	**Cross-validation:** Se: 90±4, Sp: 64±4	NO
	>700, 84 selected	3, 4 and 5	Stepwise discriminant analysis	**Leave-one-out**	NO

(continued on next page)

Table A1 (*continued*)

Ref/ Author/ year	Database: number of defibrillations	Signal duration [s]	Classification strategy	Performance [%]	Balance classes
Jekova et al. 2004 [34]				For 3 s: A: 72.3, Se: 61.8 , Sp: 79.6 For 4 s: A: 69.9, Se: 58.8, Sp: 77.6 For 5 s: A: 66.3, Se: 58.8, Sp: 71.4	
Podbregar et al. 2003 [35]	203 from 47 patients: suc: 79, unsuc: 124 Training set: 100; suc: 47, unsuc: 53 Test set: 103; suc: 32 unsuc: 71	3	Genetic programming	**Test set** Se: 100, Sp: 97.2, P: 94.1, NPV: 100	NO
Monsieurs et al. 1998 [36]	100 patients (83 witnessed, 10 not witnessed): sur: 29, no-sur: 71	4	Fisher's linear discriminant analysis	A: 77, Se: 86, Sp: 73.2, P: 56.8, NPV: 92.9	NO
Strohmenger et al. 1997 [37]	154 from 26 patients: suc: 20, unsuc: 134	3 and 12	feature threshold selection based on ROC curve	for 3s: Se: 100, Sp: 25.37 for 12s: Se: 100, Sp 46.27	NO
Brown et al. 1996 [38]	128 from 55 patients: suc: 9 (in 7 patients), unsuc: 119	4	feature threshold selection based on ROC curve	Se: 100, Sp: 47.1	NO

*suc – successful, ex – excluded, sur – survivor, ROC – receiver operating characteristic, ROEA - restoration of organized electrical activity, ROSC – return of spontaneous circulation, OHCA – out-of-hospital cardiac arrest, IHCA – in-hospital cardiac arrest, VF_T – termination of ventricular fibrillation, ROR – return of organized rhythm, SVM – support vector machine, LOPCV – leave one out patient cross-validation, GRBF – Gaussian radial basis function, A - accuracy, Se - sensitivity, Sp - specificity, P – precision, NPV – negative predictive values, AUC – area under the curve

unsuccessful shocks. This imposes the necessity for applying oversampling method and omission of the independent test set. SMOTE method could lead to over-generalization and artificially induced variance. Therefore, even though we performed 10 fold cross-validation synthetically generated samples of ROEA class were part of holdout folds of the 10 fold cross-validation, and thus could affect performance. Nevertheless, since our previous study [18], which employs traditional ML algorithms, suffers from the same limitations the implication of superiority of learned features over engineered features is still valid. To the authors' best knowledge most of the previously reported studies do not address the issue of class imbalance, even though it is a very essential issue to address. Moreover, the CNN model shows better performance when compared to the studies that consider only first shocks, use 4 s ECG segment for calculating the features, and contemplate ROEA (ROR) as a definition of the successful shock outcome, that correspond to the conditions used in this study. This further confirms superiority of learned features over engineered ones.

5. Conclusion

In this study, a novel approach for predicting defibrillation outcome was developed. A proposed CNN architecture was capable of learning useful data representations from the raw VF signals that are superior to hand-crafted features. This approach benefits from being fully automatic by fusing feature extraction, selection and classification into a single model. The robustness of the proposed model was demonstrated through a 10 fold cross-validation. We achieved Acc of 93.6 %, Sen of 98.8 % and Spec of 88.2 %, which indicate that the proposed model could be considered a safe and useful predictor for defibrillation decision and a beneficial tool to guide treatment of OHCA patients. The proposed strategy provides superiority over the conventional ML algorithms with hand-crafted features.

Even though the proposed model has many advantages, training of the CNNs are usually computationally expensive. On the other hand, once the training is completed, the classification is fast. In the future work, the authors would like to further investigate and validate this approach on larger amounts of data. Additionally, we would like to evaluate the possibility of integrating it in automated external defibrillators for real-time usage.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgments

This work was financially supported in part by European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 691051 and the Ministry of Education, Science and Technological Development of Serbia (III45010). B.E. was supported by the German Research Foundation within the framework of the Heisenberg professorship programme (grant No. ES 434/8-1). M.R. was supported by the Federal Ministry of Education and Research of Germany (grant No 01IS17070). The authors' thanks are also extended to Dr Aleksandra Maluckov for her useful comments and discussions.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.artmed.2020.101963.

Appendix

References

[1] Wik L, Hansen TB, Fylling F, Steen T, Vaagenes P, Auestad BH, et al. Delaying defibrillation to give basic cardiopulmonary resuscitation to patients with out-of-hospital ventricular fibrillation: a randomize trial. JAMA 2003;289(11):1389–95.
[2] Valenzuela TD, Roe DJ, Nichol G, Clark LL, Spaite DW, Hardman RG. Outcomes of rapid defibrillation by security officers after cardiac arrest in casinos. N Engl J Med 2000;343(17):1206–9.
[3] Abrams HC, McNally B, Ong M, Moyer PH, Dyer KS. A composite model of survival from out-of-hospital cardiac arrest using the Cardiac Arrest Registry to Enhance Survival (CARES). Resuscitation 2013;84(8):1093–8.
[4] Girotra S, Nallamothu BK, Spertus JA, Li Y, Krumholz HM, Chan PS. Trends in survival after in-hospital cardiac arrest. N Engl J Med 2012;367(20):1912–20.
[5] White RD, Asplin BR, Bugliosi TF, Hankins DG. High discharge survival rate after our-of-hospital ventricular fibrillation with rapid defibrillation by police and paramedics. Ann Emerg Med 1996;28(5):480–5.
[6] Capucci A, Aschieri D, Piepoli MF, Bardy GH, Iconomus E, Arvedi M. Tripling survival from sudden cardiac arrest via early defibrillation without traditional education in cardiopulmonary resuscitation. Circulation 2002;106(9):1065–70.
[7] Handley AJ, Koster R, Monsieurs K, Perkins GD, Davies S, Bossaert L. European Resuscitation Council Guidelines for Resuscitation 2005. Section 2. Adult basic life support and use of automated external defibrillators. Resuscitation 2005;67:S7–23.
[8] Koster RW, Baubin MA, Bossaert LL, Caballero A, Cassan P, Castren M, et al. European Resuscitation Council Guidelines for Resuscitation 2010. Section 2. Adult basic life support and use of automated external defibrillators. Resuscitation 2010; 81(10):1277–92.
[9] Perkins GD, Handley AJ, Koster RW, Catren M, Smyth MA, Olasveengen T, et al. European Resuscitation Council Guidelines for Resuscitation 2015. Section 2. Adult basic life support and automated external defibrillation. Resuscitation 2015;95: 81–99.
[10] Johnson BA, Weil MH, Tang W, Noc M, McKee D, McCandless D. Mechanisms of myocardial hypercarbic acidosis during cardiac arrest. J Appl Physiol 1995;78(4): 1579–84.
[11] Link MS, Atkins DL, Passman RS, Halperin HR, Samson RA, White RD, et al. Part 6: Electrical therapies: automated external defibrillators, defibrillation, cardioversion, and pacing: 2010 American Heart Association guidelines for

cardiopulmonary resuscitation and emergency cardiovascular care. Circulation 2010;122:706–19.
[12] Cobb LA, Fahrenbruch CE, Walsh TR, Copass MK, Olsufka M, Breskin M, et al. Influence of cardiopulmonary resuscitation prior to defibrillation in patients with out-of-hospital ventricular fibrillation. JAMA 1999;281(13):1182–8.
[13] Strohmenger HU. Predicting defibrillation success. Curr Opin Crit Care 2008;14(3): 311–6.
[14] da Luz EJS, Schwartz WR, Camera-Chavez G, Menotti D. ECG-based heartbeat classification for arrhythmia detection: a survey. Comput Meth Prog Bio 2016;127: 144–64.
[15] Sangaiah AK, Arumugam M, Bian G-B. An intelligent learning approach for improving ECG signal classification and arrhythmia analysis. Artif Intell Med 2020; 103:101788.
[16] Atanasoski V, Ivanovic MD, Marinkovic M, Gligoric G, Bojovic B, Shvilkin AV, et al. Unsupervised classification of premature ventricular contraction based on RR interval and heartbeat morphology. In: Proc. 14th Symposium on Neural Networks and Applications (NEUREL); 2018. p. 1–6.
[17] He M, Chen B, Gong Y, Wang K, Li Y. Prediction of defibrillation outcome by ventricular fibrillation waveform analysis: a clinical review. J Clin Exp Cardiol 2013;S10:1–8.
[18] Ivanovic MD, Ring M, Baronio F, Calza S, Vukcevic V, Hadzievski Lj, et al. ECG derived feature combination versus single features in predicting defibrillation success in out-of-hospital cardiac arrested patients. Biomed Phys Eng Express 2019; 5:015012.
[19] Chicote B, Irusta U, Alcaraz R, Rieta JJ, Aramendi E, Isasi I, et al. Application of entropy-based features to predict defibrillation outcome in cardiac arrest. Entropy 2016;18(9):313.
[20] He M, Gong Y, Li Y, Mauri T, Fumagalli F, Bozzola M, et al. Combining multiple ECG features does not improve prediction of defibrillation outcome compared to single features in a large population of out-of-hospital cardiac arrests. Crit Care 2015;19:425.
[21] Gong Y, Lu Y, Zhang L, Zhang H, Li Y. Predict defibrillation outcome using stepping increment of Poincare plot for out-of-hospital ventricular fibrillation cardiac arrest. Biomed Res Int 2015;2015:493472.
[22] Howe A, Escalona OJ, Di Maio R, Massot B, Cromie NA, Darragh KM, et al. A support vector machine for predicting defibrillation outcome from waveform metrics. Resuscitation 2014;85(3):343–9.
[23] Wu X, Bisera J, Tang W. Signal integral for optimizing the timing of defibrillation. Resuscitation 2013;84(12):1704–7.
[24] Ristagno G, Li Y, Fumagalli F, Finzi A, Quan W. Amplitude spectrum area to guide resuscitation – a retrospective analysis during out-of-hospital cardiopulmonary resuscitation in 609 patients with ventricular fibrillation cardiac arrest. Resuscitation 2013;84(12):1697–703.
[25] Firoozabadi R, Nakagawa M, Helfenbein ED, Babaeizadeh S. Predicting defibrillation success in sudden cardiac arrest patients. J Electrocardiol 2013;46 (6):473–9.
[26] Nakagawa Y, Sato Y, Kojima T, Wakabayashi T, Morita S, Amino M, et al. Amplitude spectral area: predicting the success of electric shock delivered by defibrillators with different waveforms. Tokai J Exp Clin Med 2013;38(2):71–6.
[27] Shandilya S, Ward K, Kurz M, Najarian K. Non-linear dynamical signal characterization for prediction of defibrillation success through machine learning. BMC Med Inform Decis Mak 2012;12:116.
[28] Endoh H, Hida S, Oohashi S, Hayashi Y, Kinoshita H, Honda T. Prompt prediction of successful defibrillation from 1-s ventricular fibrillation waveform in patients with out-of-hospital sudden cardiac arrest. J Anesth 2010;25(1):34–41.
[29] Gundersen K, Kvaløyb JT, Kramer-Johansen J, Eftestøl T. Identifying approaches to improve the accuracy of shock outcome prediction for out-of-hospital cardiac arrest. Resuscitation 2008;76(2):279–84.
[30] Neurauter A, Eftestøl T, Kramer-Johansen J, Abella BS, Sunde K, Wenzel V, et al. Prediction of countershock success using single features from multiple ventricular fibrillation frequency bands and feature combinations using neural networks. Resuscitation 2007;73(2):253–63.
[31] Watson JN, Addison PS, Clegg GR, Steen PA, Robertson CE. Practical issues in the evaluation of methods for the prediction of shock outcome success in out-of-hospital cardiac arrest patients. Resuscitation 2006;68(1):51–9.
[32] Eftestøl T, Losert H, Kramer-Johansen J, Wik L, Sterz F, Steen PA. Independent evaluation of a defibrillation outcome predictor for out-of-hospital cardiac arrested patients. Resuscitation 2005;67(1):55–61.
[33] Watson JN, Uchaipichat N, Addison PS, Clegg GR, Robertson CE, Eftestol T, et al. Improved prediction of defibrillation success for out-of-hospital VF cardiac arrest using wavelet transform methods. Resuscitation 2004;63(3):269–75.
[34] Jekova I, Mougeolle F, Valance A. Defibrillation shock success estimation by a set of six parameters derived from the electrocardiogram. Physiol Meas 2004;25(5): 1179–88.
[35] Podbregar M, Kovačić M, Podbregar-Marš A, Brezocnik M. Predicting defibrillation success by 'genetic' programming in patients with out-of-hospital cardiac arrest. Resuscitation 2003;57(2):153–9.
[36] Monsieurs KG, Cauwer HD, Wuyts FL, Bossaert LL. A rule for early outcome classification of out-of-hospital cardiac arrest patients presenting with ventricular fibrillation. Resuscitation 1998;36(1):37–44.
[37] Strohmenger HU, Lindner KH, Brown CG. Analysis of the ventricular fibrillation ECG signal amplitude and frequency parameters as predictors of countershock success in humans. Chest 1997;111(3):584–9.
[38] Brown CG, Dzwonczyk R. Signal analysis of the human electrocardiogram during ventricular fibrillation: frequency and amplitude parameters as predictors of successful countershock. Ann Emerg Med 1996;27(2):184–8.
[39] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44.
[40] Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw 2015; 61:85–117.
[41] Hinton GE. A practical guide to training restricted boltzmann machines. Neural networks: tricks of the trade. 2nd ed. Berlin Heidelberg: Springer; 2012. p. 599–619.
[42] Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. Neural Comput 2006;18(7):1527–54.
[43] Salakhutdinov R, Hinton G. Deep Boltzmann machines. In: Proc. 12th International Conference on Artificial Intelligence and Statistics (AISTATS); 2009. p. 448–55.
[44] Zubair M, Kim J, Yoon C. An automated ECG beat classification system using convolutional neural networks. In: Proc. 6th International Conference on IT Convergence and Security (ICITCS); 2016. 16450587.
[45] Al Rahhal MM, Bazi Y, AlHichri H, Alajlan N, Melgani F, Yager RR. Deep learning approach for active classification of electrocardiogram signals. Inf Sci 2016;345(1): 340–54.
[46] Kiranyaz S, Ince T, Gabbouj M. Real-time patient-specific ECG classification by 1D convolutional neural networks. IEEE Trans Biomed Eng 2015;63(3):664–75.
[47] Acharya UR, Oh SL, Hagiwara Y, Tan JH, Adam M, Gertych A, et al. A deep convolutional neural network model to classify heartbeats. Comput Biol Med 2017; 89:389–96.
[48] Ivanovic MD, Atanasoski V, Shilkin A, Hadzievski L, Maluckov A. Deep learning approach for highly specific atrial fibrillation and flutter detection based on RR intervals. In: Proc. 41st Annual International Conference of the IEEE EMBC; 2019. p. 1780–3.
[49] Acharya UR, Fujita H, Oh SL, Raghavendra U, Tan JH, Adam M, et al. Automated identification of shockable and non-shockable life-threatening ventricular arrhythmias using convolutional neural network. Future Gener Comput Syst 2018; 79(3):952–9.
[50] Acharya UR, Fujita H, Oh SL, Hagiwara Y, Tan JH, Adam M. Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network. Inf Sci 2017;405:81–90.
[51] Pourbabaee B, Roshtkhari MJ, Khorasani K. Deep convolutional neural networks and learning ECG features for screening paroxysmal atrial fibrillation patients. IEEE Trans Syst Man Cybern: Syst 2017;48(12):2095–104.
[52] Rajpurkar P, Hannun AY, Haghpanahi M, Courn C, Ng AY. Cardiologist-level arrhythmia detection with convolutional neural networks. ArXiv preprint, arXiv 2017. 1707.01836.
[53] Zhang Q, Zhou D, Zeng X. HeartID: a multiresolution convolutional neural network for ECG-based biometric human identification in smart health applications. IEEE Access 2017;5:11805–16.
[54] Liu W, Zhang M, Zhang Y, Liao Y, Huang Q, Chang S, et al. Real-time multilead convolutional neural network for myocardial infraction detection. IEEE J Biomed Health Inform 2017;22(5):1434–44.
[55] Benini S, Ivanovic MD, Savardi M, Krsic J, Hadžievski L, Baronio F. Cardially - ECG waveform dataset for predicting defibrillation outcome in out-of-hospital cardiac arrested patients. Mendeley Data 2020;1. https://doi.org/10.17632/wpr5nzyn2z.1.
[56] Deakin CD, Nolan JP. European Resuscitation Council guidelines for resuscitation 2005. Section 3. Electrical therapies: automated external defibrillators, defibrillation, cardioversion and pacing. Resuscitation 2005;67:25–37.
[57] Recommended guidelines for uniform reporting of data from out-of-hospital cardiac arrest: the 'Utstein, style'. Prepared by a task force of representatives from the European resuscitation council, American heart association, heart and stroke foundation of Canada, Australian resuscitation council. Resuscitation 1991;22(1): 1–26.
[58] Nielsen MA. Neural networks and deep learning. Determination 2015. press.
[59] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 2002;16(1):321–57.
[60] Nair V, Hinton GE. Rectified linear units improve restricted Boltzmann machines. In: Proc. 27th International Conference on Machine Learning (ICML-10); 2010. p. 807–14.
[61] 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 (1):1929–58.
[62] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. Data mining, inference, and prediction. 2nd ed. New York: Springer-Verlag; 2008.
[63] Kingma D, Ba J. Adam: a method for stochastic optimization. ArXiv preprint, arXiv 2020;1412(69).
[64] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: Proc. 13th International Conference on Artificial Intelligence and Statistics (AISTATS); 2010. p. 249–56.
[65] He H, Garcia EA. Learning from imbalanced data. IEEE Trans Knowl Data Eng 2009;21(9):1263–84.