# POLITECNICO DI TORINO
## Repository ISTITUZIONALE

Towards Uncovering Feature Extraction from Temporal Signals in Deep CNN: The ECG Case Study

(Article begins on next page)

19 April 2024

# Towards Uncovering Feature Extraction From Temporal Signals in Deep CNN: the ECG Case Study

Jacopo Ferretti
*Dipartimento di Scienze Chirurgiche*
*Università degli Studi di Torino*
Turin, Italy
jacopo.ferretti@unito.it

Pietro Barbiero
*Depertment of Mathematical Sciences*
*Politecnico di Torino*
Turin, Italy
pietro.barbiero@studenti.polito.it
0000-0003-3155-2564

Vincenzo Randazzo
*DET*
*Politecnico di Torino*
Turin, Italy
vincenzo.randazzo@polito.it

Giansalvo Cirrincione
*Lab. LTI*
*University of Picardie Jules Verne*
Amiens, France
exin@u-picardie.fr

Eros Pasero
*DET*
*Politecnico di Torino*
Turin, Italy
eros.pasero@polito.it

*Abstract*—Despite all the progress made in biomedical field, the Electrocardiogram (ECG) is still one of the most commonly used signal used in medical examinations. The problem of ECG classification has been approached in many different ways. Most of them rely on the extraction of features from the signal in the form of temporal or morphological characteristics. Although feature engineering can led to adequately good results, it mostly relies on human ability and experience in selecting the correct feature set. In the last decade, a growing class of techniques based on Convolutional Neural Network (CNN) has been proposed in opposition to feature engineering. The efficiency and accuracy of CNN-based approaches is indisputable, however their ability in extracting and using temporal features from raw signal is poorly understood. The main objective of this work was to uncover the differences and the relationships between CNN feature maps and human-curated temporal features, towards a deeper understanding of neural-based approaches for ECG.

*Index Terms*—1-D CNN, cross-correlation, deep learning, ECG classification, MLP, temporal feature analysis

## I. INTRODUCTION

With the advancement in technology in the last decade the widespread of Internet-Of-Things (IOT) and wearable devices has been constantly increasing. The ubiquity of this kind of devices make them particularly useful for medical purposes, especially for monitoring and prevention. However, the tremendous amount of medical data that a wearable can produce has no use unless is analyzed by a physician. There is still the necessity of an automated system capable of analyzing this data.

Despite all the progress made in biomedical field, the Electrocardiogram (ECG) is still one of the most commonly used signal used in medical examinations. The ECG signal is an electrical signal resulting from the various phases of the heart's muscular activation. Fig. 1 reports an healthy ECG, which presents a single heartbeat with its four important segments (P, QRS, T, U) each correlated to one of the four principal stages of activity of a cardiac cycle: isovolumic relaxation, inflow, isovolumic contraction, ejection. Significant deviation from the rhythmic repetition of this patterns are called *arrhythmia* and are important because they are an indication of an abnormal heart activity. Since cardiac arrhythmia is one of the most common cause of death, normal and abnormal ECG signal automatic classification is raising more and more interest in the scientific community.
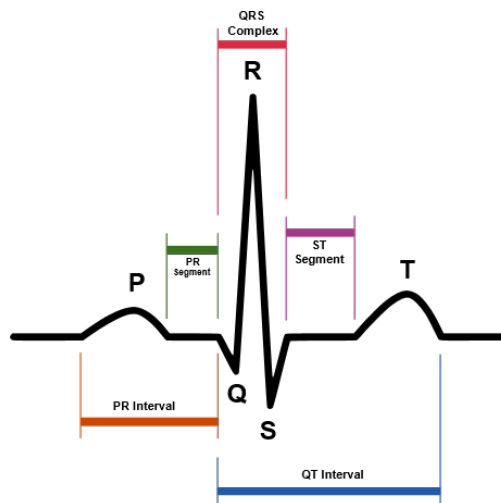


Fig. 1. Example of an healthy ECG [1].

In the years the problem of ECG classification has been approached in many different ways. Most of them rely on the extraction of features from the signal in the form of temporal

or morphological characteristics (feature engineering). The consequential analysis of those extracted features lead to the classification of the signal. For example: in [2] morphological and temporal features are evaluated for R-peak recognition, while in [3] a temporal analysis is used to recognize ventricular fibrillation and ventricular tachycardia. Although feature engineering can led to adequately good results, it relies too much on the correct extraction of the various feature and the assumption that the selected characteristics are the most representative to efficiently classify the original signal. Moreover, selecting the correct feature set requires an in-depth knowledge of the signal and the environment in which it is acquired.

Another commonly used approach, which not rely on hand-engineered features, is machine learning. In machine learning the algorithm is not explicitly designed to perform a specific task, e.g. classify different heart-beat, but instead is *taught* to perform the task by observing data examples. A growing class of techniques based on machine learning is the Convolutional Neural Network (CNN). CNN, as opposed to feature engineering, can automatically select and extract features and hidden patterns from its input data. By mimicking the brain functions, CNN consists of multiple layers, each of which owns a small subset of neurons to process portions of the input data. The results of each layer is composed of multiple features extracted from its input, resulting in increasingly abstracted features extracted with growing layers. Therefore CNN not only can automatically extract features from the signal, but they can also extract more complicated ones by extracting features from the extracted features, and so on. The main downside of this approach is the loss of control on the algorithm as most, if not all, neural networks are considered black boxes since they use too much data with very high levels of abstraction. A particular, quite interesting, class of convolutional neural network is the 1D-CNN, which takes as input data a single stream (i.e signal), e.g. ECG, and slides a kernel along it in search of particular patterns. In literature there are already some heart-beat classification studies based on 1D-CNN; for example in [4] and [5] the networks are trained for the classification of 4 and 5 different arrhythmia respectively; while in [6] there is a direct comparison between 2D-CNN (commonly used for image processing) and 1D-CNN in classifying 5 classes of heart-beat; finally in [7] a very deep 1D-CNN is used to recognize heart-beat amongst 14 different classes.

## II. PROPOSED APPROACH

In this paper we propose a simple 1D-CNN architecture applied to the MIT-BIH database [8] for a complete ECG arrhythmia classification. However, the main goal of this work is the interpretation of the behavior of the neural network by means of comparison with a more traditional feature engineering approach. The features extracted from the two methodologies are compared using cross-correlation analysis to assess similarities. This will help to better understand if the features extracted from the network are somehow related

to those hand-picked, and consequently if those are actually useful for ECG arrhythmia classification.

### A. Input Dataset

The MIT-BIH database from PhysioNet [9] is considered the gold standard when dealing with ECGs. It contains data from 48 different patients in the form of two lead (L2 and V1) ECG recording of 30 minutes. Its approximate 109000 heart-beats are distributed in 16 different classes and each of them has been manually labeled by two professional cardiologists. The first step was to parse each record and split it into smaller segments (1-2 seconds) around the QRS complexes. The original dataset was acquired at 360 samples/s, so the chosen segment corresponds to 500 samples. Data augmentation was performed by means of 10% overlapping among segments. Segments were statistically normalized and labeled appropriately. Finally, a random 90% of the resulting dataset was used for training while the remaining 10% as validation dataset.

### B. 1-D CNN

In Convolutional Neural Networks (CNN) a filter - commonly called kernel - is convoluted (sled) through data in order to learn particular patterns. These pattern extracted grow in complexity along with the depth of the network. Namely, deeper networks extracts more elaborate features.

In Fig. 2 there is an example of a 1D-CNN where a kernel of width 3 is passed across an input vector of length 10, producing an output of dimension 8. The process of convolution starts by superimposing the kernel with part of the input, then, the corresponding elements are multiplied and their results summed with each other forming a new element of the output vector. After that, the kernel is moved and the process is repeated for all the elements of the input matrix, excluding those in the edges where there is no enough data to superimpose the whole kernel. It is important to note that because of how the convolution works, the output matrix is smaller than the input one depending on the size of the kernel.

### C. Temporal features extraction

An alternative approach for ECG automatic classification is based on feature engineering [10]–[12]. As explained in [13], this strategy exploits domain experts, e.g. physicians, to extract the relevant information from data and, then, fed it to a neural network. For example, in [14], several features were extracted from surgeon hand motion recording to determine his skill level during surgical training.

In a previous study [15] of the authors, the proposed dataset manifold have been analyzed and different feature engineering techniques have been used to train a multi-layer perceptron (MLP), and tested on their classification performances. Table I summarizes the fifteen features that yielded the best results based on the temporal evolution of each record of the ECG dataset.

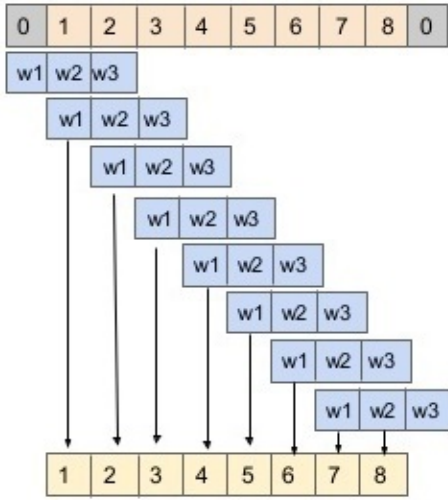Here is how the different features are evaluated:

Fig. 2. Example of 1D Convolution Neural Network.

TABLE I
TEMPORAL FEATURE EXTRACTED FROM THE ECG AND THEIR MEANING.

| Feature # | Feature type |
|-----------|--------------|
| F1 | Mean |
| F2 | Max value |
| F3 | Root Mean Square (RMS) |
| F4 | Square Mean Root (SMR) |
| F5 | Standard deviation |
| F6 | Variance |
| F7 | Shape factor (using RMS) |
| F8 | Shape factor (using SMR) |
| F9 | Crest factor |
| F10 | Latitude factor |
| F11 | Impulse factor |
| F12 | Skewness |
| F13 | Kurtosis |
| F14 | Normalized 5th central moment |
| F15 | Normalized 6th central moment |

$$F1 = \bar{x} = \sum \frac{X_i}{N}$$

$$F2 = max\,(x)$$

$$F3 = \sqrt{\sum \frac{x_i^2}{N}}$$

$$F4 = \left( \sum \frac{\sqrt{|x_i|}}{N} \right)^2$$

$$F5 = \sqrt{\frac{1}{N-1} \sum (x_i - \bar{x})^2}$$

$$F6 = F5^2 = \frac{1}{N-1} \sum (x_i - \bar{x})^2$$

$$F7 = \frac{F3}{\sum \frac{|x_i|}{N}} = \frac{\sqrt{\sum \frac{x_i^2}{N}}}{\sum \frac{|x_i|}{N}}$$

$$F8 = \frac{F4}{\sum \frac{|x_i|}{N}} = \frac{\left( \sum \frac{\sqrt{|x_i|}}{N} \right)^2}{\sum \frac{|x_i|}{N}}$$

$$F9 = \frac{F2}{F3} = \frac{max\,(x)}{\sqrt{\sum \frac{x_i^2}{N}}}$$

$$F10 = \frac{F2}{F4} = \frac{max\,(x)}{\left( \sum \frac{\sqrt{|x_i|}}{N} \right)^2}$$

$$F11 = \frac{F3}{\sum \frac{|x_i|}{N}} = \frac{max\,(x)}{\sum \frac{|x_i|}{N}}$$

$$F12 = \frac{\frac{1}{N} \sum (x_i - \bar{x})^3}{\left[ \frac{1}{N-1} \sum (x_i - \bar{x})^2 \right]^{\frac{3}{2}}}$$

$$F13 = \frac{\frac{1}{N} \sum (x_i - \bar{x})^4}{\left[ \frac{1}{N-1} \sum (x_i - \bar{x})^2 \right]^2}$$

$$F14 = \frac{\frac{1}{N} \sum (x_i - \bar{x})^5}{\left[ \frac{1}{N-1} \sum (x_i - \bar{x})^2 \right]^{\frac{5}{2}}}$$

$$F15 = \frac{\frac{1}{N} \sum (x_i - \bar{x})^6}{\left[ \frac{1}{N-1} \sum (x_i - \bar{x})^2 \right]^3}$$

where $N$ is the total number of the elements of the vector $X$, whilst $x_i$ is the $i^{th}$ element.

### D. Cross-correlation

Given a pair of random variables $X$ and $Y$ with probability density functions $f_X$ and $f_Y$, respectively, the probability density function of their difference $X - Y$ is known as cross-correlation [16] and it is given by:

$$f_{X-Y} = f_X \star f_Y \tag{1}$$

The resulting probability density can be interpreted as a measure of similarity between the two random variables. For two random vectors $\mathbf{X} = (X_1, \ldots, X_m)^T$ and $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$, the cross-correlation matrix is defined by:

$$R_{\mathbf{XY}} := \mathbb{E}[\mathbf{XY}^T] \tag{2}$$

For this reason, the cross-correlation is also known as *sliding sliding dot product*. In order to provide a simple interpretation of the results, the normalized version of the cross-correlation has been used:

$$\rho_{\mathbf{XY}} = \frac{\text{cov}(\mathbf{X}, \mathbf{Y})}{\sigma_{\mathbf{X}} \sigma_{\mathbf{Y}}} \tag{3}$$

ranging in $[-1, 1]$, with 1 indicating perfect correlation, 0 no correlation, and $-1$ perfect anti-correlation.

*E. Feature Maps and Temporal Features Similarity*

Cross-correlation has been used as a similarity measure to compare the feature maps generated by the 1D CNN with the temporal features extracted from the original dataset. For each sample ($i$), for each convolutional filter ($j$), and for each temporal feature ($k$), the cross-correlation has been computed to estimate the similarity between the feature map $x_i^j = (x_{i1}^j, \ldots, x_{im}^j)^T$ with the temporal representation of the sample $x_i^k = (x_{i1}^k, \ldots, x_{in}^k)^T$:

$$\rho_{x_i^j x_i^k} = \frac{\text{cov}(x_i^j, x_i^k)}{\sigma_{x_i^j} \sigma_{x_i^k}} \tag{4}$$

For each convolutional filter, the obtained score has been averaged across all samples in order to estimate the average similarity between feature maps and temporal representations (the higher the score, the higher the similarity):

$$\rho_{j,k} = \frac{1}{N} \sum_i^N \rho_{x_i^j x_i^k} \tag{5}$$

ranging in $[-1, 1]$. The final score $\rho_{j,k}$ has been used to investigate the abstraction level of temporal features with respect to the features extracted by the CNN.

## III. NET ARCHITECTURE

The proposed network architecture follow a bottom-up approach in which deeper layers contain increasing number and dimension of filters. Namely in each consecutive layer dimension and number of filters were doubled alternatively. Therefore starting with a first layer consisted of 4 filters with kernel dimension 8, the second layer had 8 filters of dimension 8; the fourth 8 filters of dimension 16, and so on for 10 total convolutional layers. Furthermore, the results are passed through a fully connected layer and softmax activation that produce a distribution over the 16 output classes. In order to reduce the total number of parameters of the network, the final layers are interleaved with max pooling layer. Finally, to improve the training phase and to avoid over-fitting a dropout layer was added before the softmax. Figure 3 represent the final architecture of the network.

## IV. CROSS-CORRELATION ANALYSIS

The transformed dataset based on temporal features has been analyzed together with the proposed CNN in order to understand *if* and *how* the neural network is able to extract abstract temporal features from raw signals. For each convolutional filter and temporal feature pair, the average normalized cross-correlation score has been computed as described in (5). Table II summarizes the main findings. For each temporal feature, the minimum and maximum cross-correlation with the best matching feature map has been reported alongside with the corresponding filter position. The results clearly show how:
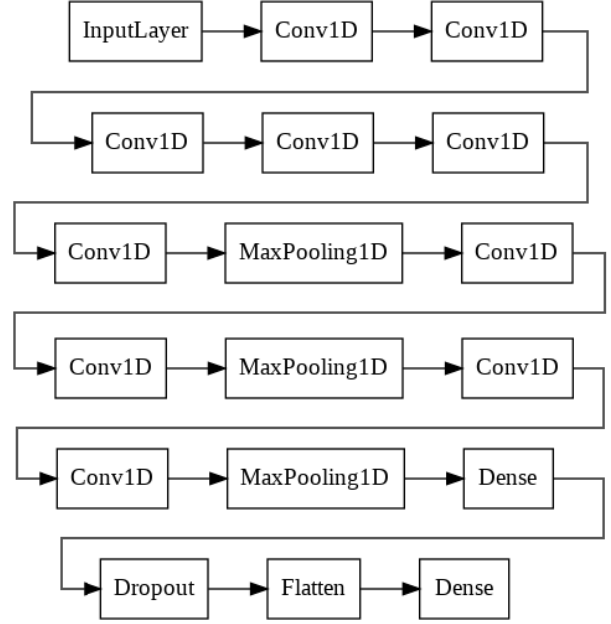


Fig. 3. Net architecture.

- the CNN was able to automatically extract from raw signals *temporal-like* features similar to human-curated ones
- the neural network extracted such features in the earliest layers directly from raw signal

Most likely, in deeper layers the networks exploited its own *temporal-like* features in order to refine the internal representation and improve the classification accuracy. The temporal features displaying the highest cross-correlation were:

- the mean value (F1, $\rho = -0.882$)
- the max value (F2, $\rho = -0.808$)
- the root mean square (F3, $\rho = 0.875$)
- the square mean root (F4, $\rho = 0.882$)
- the crest factor (F9, $\rho = -0.838$)

of the raw signal. All of these temporal features have been extracted in the first layer of the network. Moreover, they can be clustered according to the most similar feature map they have been associated with. In fact, both the mean value (F1), the max value (F2), and the crest factor (F9) are strongly anti-correlated with the feature map generated by the first filter of the first layer. On the other hand, both the root mean square (F3) and the square mean root (F4) are mostly correlated with the second filter of the first layer. Therefore, ***these feature maps can be interpreted as a compact representation of core waveform characteristics such as***:

- ***how extreme the peaks are with respect to the signal mean value*** (first filter of the first layer, associated with mean value, max value, and crest factor)
- ***the peak-to-peak amplitude*** (second filter of the first layer, associated with root mean square and square mean

root)

| TEMPORAL FEATURE | | LAYER | FILTER | $\rho$ |
|---|---|---|---|---|
| **F1** | MAX | 1 | 2 | 0.497 |
| | **MIN** | **1** | **1** | **-0.882** |
| **F2** | MAX | 1 | 2 | 0.685 |
| | **MIN** | **1** | **1** | **-0.808** |
| **F3** | **MAX** | **1** | **2** | **0.875** |
| | MIN | 3 | 7 | -0.236 |
| **F4** | **MAX** | **1** | **2** | **0.882** |
| | MIN | 6 | 4 | -0.252 |
| F5 | MAX | 1 | 2 | 0.748 |
| | MIN | 3 | 7 | -0.236 |
| F6 | MAX | 1 | 2 | 0.699 |
| | MIN | 3 | 7 | -0.246 |
| F7 | MAX | 1 | 2 | 0.553 |
| | MIN | 7 | 3 | -0.180 |
| F8 | MAX | 2 | 6 | 0.658 |
| | MIN | 7 | 6 | -0.138 |
| **F9** | MAX | 1 | 2 | 0.725 |
| | **MIN** | **1** | **1** | **-0.838** |
| F10 | MAX | 1 | 2 | 0.701 |
| | MIN | 1 | 1 | -0.700 |
| F11 | MAX | 1 | 2 | 0.711 |
| | MIN | 1 | 1 | -0.780 |
| F12 | MAX | 1 | 2 | 0.500 |
| | MIN | 1 | 1 | -0.553 |
| F13 | MAX | 1 | 4 | 0.655 |
| | MIN | 2 | 3 | -0.048 |
| F14 | MAX | 2 | 8 | 0.532 |
| | MIN | 1 | 1 | -0.300 |
| F15 | MAX | 1 | 4 | 0.681 |
| | MIN | 3 | 3 | -0.014 |

## V. ECG CLASSIFICATION

The analysis of the previous section has shown how the temporal features were automatically extracted by the proposed deep network in the first layer. In order to further investigate the contribution of the 1D-CNN features to the ECG classification task, a comparison with the MLP used in [15] has been conducted. Because of the under-representation of most of the classes of the original MIT-BIH dataset, only the most significative subset (normal beat, right bundle branch block beat, premature ventricular contraction, atrial premature contraction) was employed, while the remaining were collapsed in a generic label *other anomalies*.
The MLP has a single hidden layer with forty neurons and five output units equipped with the soft-max activation function [17]. Because of the use of the cross-entropy error function [17], they yield the probability of class membership. Fig. 4 shows the associated confusion matrix. The classification reached an overall accuracy of 96%.

The 1D-CNN used to conduct this part of experiments was constructed following the same methodology described previously. However, since the classification problem was
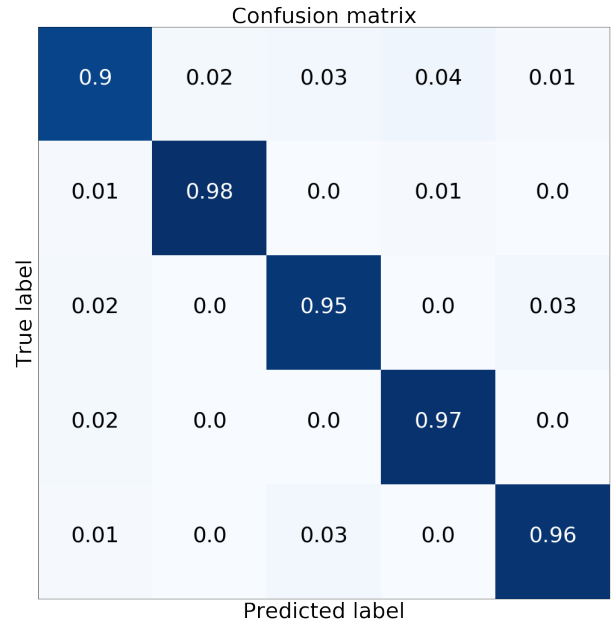


Fig. 4. MLP confusion matrix

intrinsically easier to solve it was used a smaller network consisting of only four CNN layers followed by a softmax classifier. Fig. 5 shows the architecture of the smaller net, while Fig. 6 represents its confusion matrix. An overall accuracy of 99.6% was achieved.
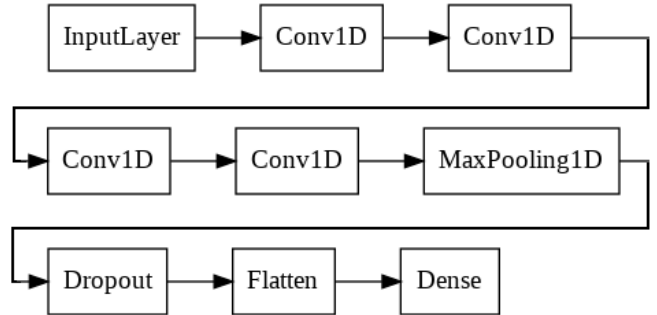


Fig. 5. Second CNN architecture

The comparison between the two architectures on the heartbeat classification has confirmed the quality of the deep learning approach, which was able to achieve better results by extracting the same features engineered with the MLP approach in the first layers and improving the classification (4%) with the others automatic extracted characteristics of the subsequent layers.

Resuming, both networks have been trained and tested on a subset of the database used for the cross-correlation analysis and with a smaller number of classes because the goal of the proposed approach is not the design of the best classifier, but the impact of the more abstract features extracted by the CNN with regard to the previously analyzed temporal features.
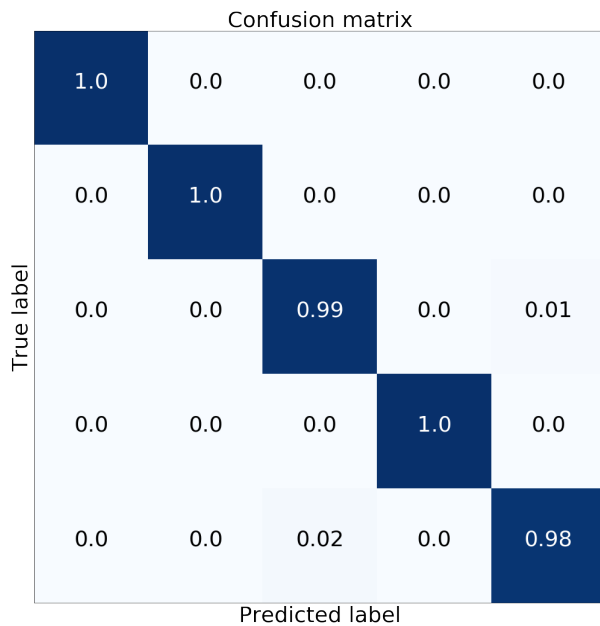
Fig. 6. 1D-CNN confusion matrix

This approach assumes that the analysis given in the previous section can be extended to this particular case, which has been already verified by the deep neural experts that exploit it, for example, for motivating transfer learning.

## VI. CONCLUSIONS

This paper has introduced an original approach, which works both for the evaluation of the impact of feature engineering in a classification problem by means of classical (shallow) neural networks by exploiting the deep convolutional architecture as a kind of non-linear performance evaluation, and, on the contrary, as an interpretation of the deep convolutional layers, by using human-curated features as clues.

In particular, the temporal features have been here taken into account, because of their good performance in ECG classification. To this aim, a cross-correlation analysis has been performed to investigate the similarity between the 1-D CNN feature maps and hand-engineered temporal features. It has been deduced that most temporal features are synthesized in the first convolutional layer, which means that these features are fundamental for the problem at hand. However, there are other more abstract features in the subsequent layers that improve the classification. This analysis has been confirmed by the experiments. Indeed, MLP, based on temporal features, yields good results, but CNN is even more accurate. In conclusion, this work has both justified the importance of a subset of temporal features, thanks to the first CNN layer correlation analysis, and has paved the way of interpreting a convolutional layer by using certain choices of features. The interpretation of the behavior of a deep neural network is an open problem and, probably, their biggest drawback (black box). This paper is not only original for the approach to the ECG classification problem, but suggests a new way to understand deep learning.

Future work will address the analysis of other features and the use of alternative comparison techniques in order to have an idea of the abstract features created by CNN in the deeper layers.

## REFERENCES

[1] Wikipedia electrography page. https://en.wikipedia.org/wiki/Electrocardiography.
[2] J. Pan and W. J. Tompkins. A real-time qrs detection algorithm. *IEEE Transactions on Biomedical Engineering*, BME-32(3):230–236, March 1985.
[3] Muhammad Abdullah Arafat, Abdul Wadud Chowdhury, and Md. Kamrul Hasan. A simple time domain algorithm for the detection of ventricular fibrillation in electrocardiogram. *Signal, Image and Video Processing*, 5(1):1–10, 2011.
[4] Nikhil Gawande and Alka Barhatte. Heart diseases classification using convolutional neural network. In *2017 2nd International Conference on Communication and Electronics Systems (ICCES)*, pages 17–20. IEEE, 2017.
[5] Dan Li, Jianxin Zhang, Qiang Zhang, and Xiaopeng Wei. Classification of ecg signals based on 1d convolution neural network. In *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)*, pages 1–6. IEEE, 2017.
[6] Yunan Wu, Feng Yang, Ying Liu, Xuefan Zha, and Shaofeng Yuan. A comparison of 1-d and 2-d deep convolutional neural networks in ecg classification. *arXiv preprint arXiv:1810.07088*, 2018.
[7] Awni Y Hannun, Pranav Rajpurkar, Masoumeh Haghpanahi, Geoffrey H Tison, Codie Bourn, Mintu P Turakhia, and Andrew Y Ng. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature medicine*, 25(1):65, 2019.
[8] George B Moody and Roger G Mark. The impact of the mit-bih arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine*, 20(3):45–50, 2001.
[9] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *circulation*, 101(23):e215–e220, 2000.
[10] Válber César Cavalcanti Roza, Ana Maria de Almeida, and Octavian Adrian Postolache. Design of an artificial neural network and feature extraction to identify arrhythmias from ecg. In *2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pages 391–396. IEEE, 2017.
[11] Vitoantonio Bevilacqua, Leonarda Carnimeo, Giuseppe Mastronardi, Vito Santarcangelo, and Rocco Scaramuzzi. On the comparison of nn-based architectures for diabetic damage detection in retinal images. *Journal of Circuits, Systems, and Computers*, 18(08):1369–1380, 2009.
[12] David Cuesta-Frau, Juan C Pérez-Cortés, and Gabriela Andreu-García. Clustering of electrocardiograph signals in computer-aided holter analysis. *Computer methods and programs in Biomedicine*, 72(3):179–196, 2003.
[13] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, 33(4):917–963, Mar 2019.
[14] Munenori Uemura, Morimasa Tomikawa, Tiejun Miao, Ryota Souzaki, Satoshi Ieiri, Tomohiko Akahoshi, Alan K. Lefor, and Makoto Hashizume. Feasibility of an AI-based measure of the hand motions of expert and novice surgeons. 2018:9873273.
[15] Giansalvo Cirrincione, Vincenzo Randazzo, and Eros Pasero. *A Neural Based Comparative Analysis for Feature Extraction from ECG Signals*, pages 247–256. Springer Singapore, Singapore, 2020.
[16] Jae-Chern Yoo and Tae Hee Han. Fast normalized cross-correlation. *Circuits, systems and signal processing*, 28(6):819, 2009.
[17] Christopher M Bishop et al. *Neural networks for pattern recognition*. Oxford university press, 1995.