ECG Activity recognition: analysis on a recurrent neural network classifier approach

Andrea Mascari, Davide Pivato

***Abstract*—Activity recognition carried out with a collection of ECG (electrocardiogram) signals, has been significant over the years in several health evaluation tasks, and in relation to the specific act of motion, even with the objective of physical activity assessment differentiation. The application of machine learning approaches in this field has been explored by the use of deep neural network architectures, to individuate local dependencies among adjacent input samples. To extend their application we proposed in addition to the already known models, the usage of recurrent neural networks with gated recurrent units and the possibility to initialize the network's weights with preliminary auto-encoders representations. What we observed as results are that on the classification of previously unseen signals, the already existing models based on a convolution approach are outperformed by recurrent neural networks. This kind of architectures show impressive results in the time needed to achieve robust performance on brand new subjects, of which we don't possess any type of information rather than their heart activity.**

***Index Terms*—Recurrent Neural Networks, Supervised Learning, Neural Networks, ECG Activity recognition, Autoencoders, Convolutional Neural Networks.**

I. Introduction

The studies of electrocardiogram signals are more and more used to identify several physiological tasks, varying from rest state to more active motions like walking as running or doing sports as well, giving an essential contribution in the evaluation of the physical state of a person. The simplicity in the acquisition of the data, increasingly facilitated over the years by the improvement in the sensors field with wearable wireless sensors, has permitted to conduct different works in the monitoring of the correlation between an activity and the human body responses. Lately this prospective and care about this type of data has been adopted in professional environment in sports and in researches in the medicals field. The most significant results has been achieved in this topic by the employment of machine learning techniques and algorithms, able to analyze underlying patterns that characterize the different possible activities. However, often this procedures demands to already know the specific features of the subject to be able to classify new data from her, . What we proposed, really aim to generalize the activity recognition task on subjects that weren't seen during training phase and still achieve a desiderable performance by the mean of a recurrent neural network architecture the exploits the power of gated recurrent units. The main phases of our procedure that will be detailed covered as salience points are:

- The pre-processing and data manipulation approaches in a small dataset scenario.
- The realization of an basic convolutional neural network and the analysis of its limited performance in poor data contexts.
- The final proposed model as an extension technique and its enhanced performance applicability.

This report will be presented in the following structure. In Section III will be described the baseline description of the original model in order to highlight the limitations and the aspects we wanted to improve, Section IV will focus on the proposed data manipulation and pre-processing. The detailed procedure we performed and the steps we followed will be explained in Section V and its final compared results in Section VI. Concluding remarks are provided in Section VII.

II. Related Work

The growing refinement of wearable ECG sensors allowed to easily gather accurate data, this helped to develop researching on machine learning techniques that can perform different classification tasks from heart's signal, and one of them is precisely activity recognition.

Taking as example the paper *Body Movement Activity Recognition for Ambulatory Cardiac Monitoring* [1], classification on multiple activities (even similar between them, such as moving right arm against moving left arm), has been conducted by applying a mathematical model based on *PCA* reconstruction of signal's components, a remarkable accuracy has been obtained in distinguish the most separable activities, (99% recall of climbing up stairs).

At the same time we have examples of recognition tasks even better carried out by neural networks, it's the case presented on the paper *Activity Recognition for Cognitive Assistance Using Body Sensors Data and Deep Convolutional Neural Network* [2], were the multiple class activity recognition has been performed with a 94% of accuracy using deep convolutional networks. An another paper such *Interpretable Parallel Recurrent Neural Networks with Convolutional Attentions for Multi-Modality Activity Modeling* [3], even uses data from multiple sensors all together, opted for an attention mechanism beside a recurrent neural network to capture spatial significant regions.

Even though the approaches presented in the literature are especially good talking about performances they lack on generalization side, indeed the solutions presented by those papers is limited on classifying new data which, however, are gathered from "old" subjects, i.e. that already appear in the training set.

In our paper we want instead to propose a new approach, which requires less prior knowledge but is still able to get close to the results obtained in the state of the arts. What we are going to discuss below, is an architecture that make the most of gated recurrent units, being able to recognize activities not only from new signals, but also from new subjects that have not been observed in the training phase. As results we obtained a model that gives an optimal trade off between complexity and stability in the accuracy, which can be extended to scenarios where we couldn't have recorded data of the subject from which we want to predict the activity.

III. Processing Pipeline

The architecture defined has been thought to perform a binary classification between Walking and Rest activity based on segmented ECG signal, in this context, the first block of our pipeline aimed to detect and subsequently extract every single beat in the signal recorded from subjects. The signals recorded at hand, have been properly scaled and pre-processed, and then the train and test sets are built from the fixed sized segments containing a single beat returned by the detection operation. The data generated, are given as input to the main structure of our model that we are going to introduce in this section: the Recurrent Classifier.

The first developed structure that we built, in order to process the collected data, is a recurrent neural network composed by a Dense layer and a layer exploiting the use of Gated Recurrent Units. The Dense layer has been used as starting point of our design, it is a fully connected layer of neurons that we implemented to carry out the function of 'bridge' between the input and the main part of the network, consisting of GRU cells. This specific kind of memory cells, allows the architecture to maintain a certain amount of information about the past units of the general design, helping to tackle the main problem of "vanishing gradient". In the scenario taken in exam by us, dealing with ECG signals time series, GRU cells play a very critical role, making it possible to taking into account in the learning process, the temporal correlation of the heart activity data, and extend this knowledge through the whole network.

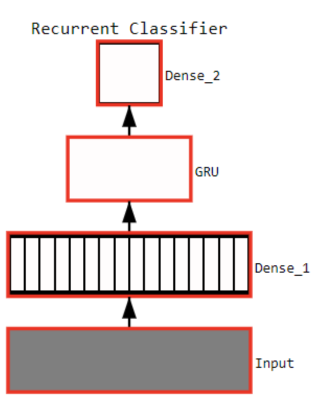

Fig. 1: Design of the recurrent classifier architecture.

IV. Signals and Features

The ECG signals are hand collected, by the mean of the PolarH10 chest strap heart rate sensor. The subjects taken in exam are 7, all of them with no bias about age or gender, that after the whole pre-processing procedure formed a dataset containing 1712 samples. Among this 7 subjects 5 of them were used in the training set, and 2 of them in a brand new test set, doing so we added an ulterior challenge to the task at hand, removing any previous information of the test subjects in the learning process. The samples belong to two classes of observations: the Rest state and the Walking state. Each of the two classes had a recording time of about one minute, an uncertainty that produced a generation of signals of slightly different lengths.

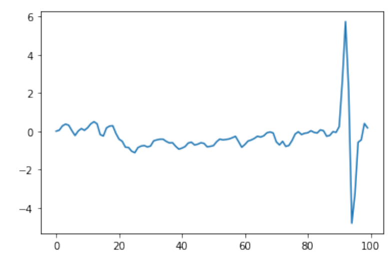

Fig. 2: Example of a single heartbeat detected.

The pre-processing procedure starts with a standard scaling normalization of the whole set of input signals, here each signal has been normalized subtracting its mean and dividing it by its standard deviation, to transform the features of the signals to have a mean of zero and a standard deviation of one, doing so we worked with a smaller range of values and reduced the variance of the machine learning techniques results. After that, we applied a segmentation function that used a QRS complex detection, in the form of two average detector implemented by Elgendi Mohamed & Jonkman Mirjam & De Boer Friso in the paper *Frequency Bands Effects on QRS Detection* [4], that returned ECG beats segments of a fixed feature-size of 100 elements, this allowed us to cope with the problem of the slightly different lengths of the signals and to work with a relatively 'bigger' dataset (each application of the segmentation function used an ad hoc value for the frequency of the segmentation due to the difference of lengths). However this problem still persisted having an effect on the number of samples per class, resulting in a imbalanced dataset; to tackle this collateral effect we implemented an oversampling technique with the SMOTE algorithm augmenting the dataset with synthetic generated samples of the minority class, that in our case was the rest class, interpolating a random instance of the class taken in exam with its closest determined neighbours. Doing so we managed to overcome this situation obtaining a more balanced dataset to work with, with a reduced risk of incoming in overfitting.

V. Learning Framework

The visual representation of our main architecture in detail is shown in the figure 2.

A. Recurrent classifier architecture

Now we dive in the main part of our work, which has the role of compute the actual classification task. First of all, the network of the recurrent classifier begins with the previous mentioned Dense Layer, which doesn't contain a huge number of features in order to keep sustainable the amount of the complexity of the network, already high for the presence of the section that will come next. Then we implemented a *GRU* layer to incorporate the memory in our network; as mentioned in the previous section, the Gated recurrent units are critical to model sequential data, they control the flow of information from the previous layers thanks to two hidden states: the update gate that select the amount of past information to maintain and to delete and the reset gate the amount of past information to forget, finally the information of the current hidden state is computed as shown in the figure 3 . At the end,

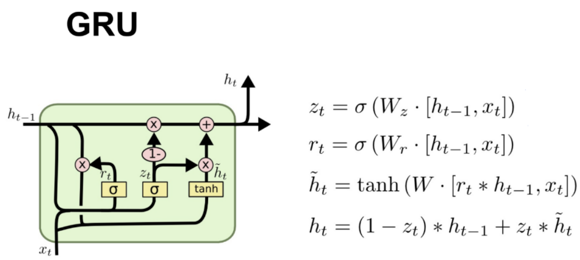

Fig. 3: Architecture of a GRU cell, z_t and r_t represent respectively the update and the reset gate, h_t and h_t-1 the current and the previous hidden state.

we concluded our network with a final output Dense layer, composed of 1 neuron with a *sigmoid* activation function, to perform the actual classification; more specifically the loss function of the model, considering that we tackled a binary classification problem, is 'binary cross-entropy'. We decreased the learning rate of our model from its standard value of '0.01' to '0.001' and opted for a very large batch size of '1024'. All of those choices combined acted as a counter measure to the huge 'power' of the GRU layer and were aimed to improve the generalization and prevent the network to overfit.

B. Experimented approaches

1) Undercomplete Autoencoder: The Autoencoder technique is a unique neural network design that plays an essential role in the field of unsupervised representation learning, being employed for several different tasks such as data denoising, generation, compression. This architecture follows an hourglass shape, reducing, going forward, the dimensionality of the original input samples, permitting to extract the most meaningful information, for then reconstructing them back to their starting form. What we were interested about, is its property to obtain the cited above compressed representation, as conduct in the paper *Autoencoders as Weight Initialization of Deep Classification Networks Applied to Papillary Thyroid Carcinoma* [5], using the weights that the architecture has learned, and feed them as initialization to a first fully connected layer, to accomplish the goal of accelerating the learning phase. Unfortunately this approach didn't returned the desired effect, showing a very similar, if not sometimes worse performance both in the speed of the process and in the actual metrics score as well.

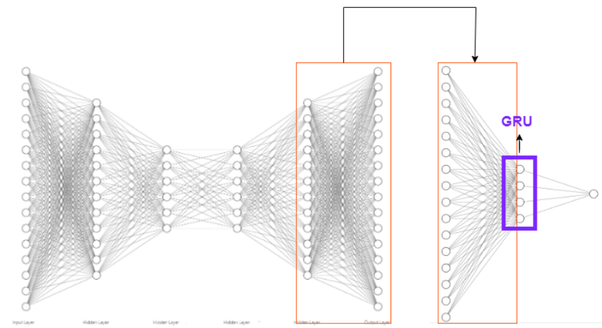

Fig. 4: Whole architecture scheme.

Our autoencoder was composed of 6 layer, 3 for the encoder section and 3 for the decoder one; the order of the number of neurons for each layer is 100, 16, 8 (and reversed) to obtain and reconstruct from a very compressed and dense representation. For the same reason of reconstruction, this technique network has to be fed with the train samples in both input and target, the activation function selected for all the layers was the 'Scaled Exponential Linear Unit'(Selu), preferred to the standard 'Rectified Linear Unit'(Relu) (Figure 6),

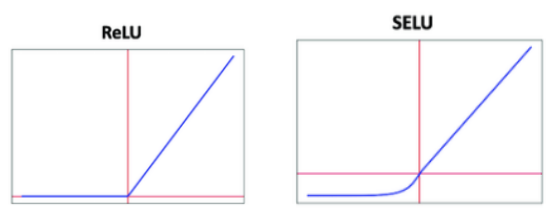

Fig. 5: Comparision of Relu and Selu activaction funtions: Selu is derivable in zero and has positive derivative for negative values.

for its property of self normalization, to continue the idea of mean zero and variance one in the layers, and moreover for the fact that it gives positive derivative to negative values, that are meaningful in the reconstruction task, doing so to prevent the neurons from 'dying', phenomenon that occurs with the application of the 'Relu' function instead. Exception was made for the last layer of the decoder, that having to rebuild the original values of the input signal, which can take both positive and negative values, has been assigned with a 'linear' activation function. To monitor and measure the goodness of our autoencoder we measured the reconstruction error by the mean of a loss function computed with the mean squared error, giving again the nature and the range of possible values for the data given as output.

2) 1-d Convolutional Neural Network: To compare the recurrent neural network with a different topological approach, we built a convolutional neural network to apply it to a 1D representation of the input ECG segments, in which each feature represent a different subsequent time step in the signal. Convolutional neural networks distinguish themselves from other architectures for the usage of the convolution operation by which the network applies a sliding window filter across all the input in order to capture local dependencies in it. The model that we proposed has two one dimensional convolutional layers, since the signal is a time series, that learn respectively 16 and 8 filters, all of them with a window size of 9. Between the layers has been inserted an average pooling layer in order to reduce dimensionality of the input while keeping information of both positive and negative values. We used two different techniques for normalization, a batch normalization layer after the first convolutional layer and two dropout layers after each convolutional layer, the dropout makes the network learn a more distributed information from the data since they randomly inhibit units in each epoch with the goal to don't make the model rely too much on a specific node. The convolutional neural network was mostly develop in order to have an already explored architecture that could be a reliable baseline for the analysis of the recurrent models.

VI. Results

The architecture and the learning method that we propose brought remarkable results both in terms of metrics score and generalization if compared against other approaches that we tested, with the only exchange of an higher training time which by the way could be restrained by the computational expensiveness of the GRU units.
For a first general overview of the perfomance we show below the plot showing the trend of the loss both in the training and in the test set:

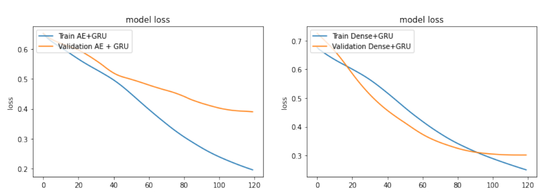

Fig. 6: Plot showing the loss behaviour in AE + RNN and Dense + RNN

It comes up that the starting point injected from the auto-encoder didn't help to reach a faster and more stable convergence, the recurrent classifier randomly initialiazed shows a nice behaviour of the loss, that avoids overfitting.
To verify the results carried out by the neural networks we show the table below containing the scores obtained on the most significant metrics, in the table we inserted also the model made of two one dimensional convolutional layers, since the architecture has been described as successful in the paper cited on section II, we wanted to test our model against convolutional based neural network.

	Accuracy	Std	Precision	Recall	F1 score
Dense + RNN	90.0%	2.1%	97.2%	83.0%	89.5
AE + RNN	87.4%	4.2%	97.2%	84.3%	90.3
DeepCNN	74.2%	0.7%	82.1%	77.0%	79.5

TABLE 1: Table showing the metrics for different models

The gated recurrent unit confirmed the hypothesis to be the best learning unit to adopt in this situation that regards time series, its ability to remember information from the past and to forget it when not significant anymore resulted in remarkable results, indeed the model making use of them were able to achieve the highest accuracies, we can't say the same for the convolutional neural network, were the accuracy score is way worse, evidently the lack of ability in catching temporal dynamics played an important role on making the model perform badly.
To analyze more specifically the kind of predictions made by the best models we showed also the precision and recall metrics. Precision score tells the percentage of walking labelled data among all the data labelled as walking, we can see how the developed model almost never fails on identify the ECG segments related to Rest class.
Another speech needs to be made for the recall metric, where the model still performs well with a score of 83%, but significantly decrease its performance, it means that the task to recognize the walking segments is slightly harder for it, and we should accept this error, since it is more probable, for instance in the situation of a well trained subject, that its heartbeats while walking are closer to rest's heartbeats of an average person, while it is more rare to see the opposite, i.e. a person that while resting has heartbeats similar to an average person that's walking.
To directly see the number of error occurred in a prediction task we show the confusion matrix below, from which the precision and recall metrics are computed.

	Rest	Walking
Predicted Rest	173	35
Predicted Walking	2	300

TABLE 2: Confusion Matrix of Dense + RNN

We can tell our self satisfied with the results achieved by the recurrent neural networks, the accuracy score in the prediction was acceptable since we decided to dive in a difficult prediction task with only a restricted available dataset for train the models. The results overcome our beginning expectations, even though we must admit that we expected an improvement on the metric's scores and convergence when using the RNN initialized with the auto-encoder's weights, so we had to conclude that our starting hypothesis was wrong since strongly disproven by the empirical results.

VII. Concluding Remarks

The solution proposed by this paper made use of gated recurrent units that perfectly fit fit for the processing of long time series data such as ECG signals. The activity recognition task was performed in an innovative point of view, where the approach was to spot differences in the shape of heartbeats in order to recognize the activity, in a way that the knowledge learned by the model could be generalized into unseen subjects. The results obtained by the model, especially if we focus on the precision metric with which the model can assign a segment's label, can possibly open real world scenarios where the application of activity recognition tasks could be well carried out instantaneously, or better, in just prediction time, either for medical purposes or in less precision demanding fields such as sports analytics.

Even though the results obtained in our project are reliable, the selfmade dataset is still a fundamental aspect that should be improved and extended in order to better analyze whether the model would still acceptably generalize in a situation where data are gathered in a wider and less biased sample of person, to compose a training set that contains subjects with different physical shapes, pathological diseases or ethnic groups. Another possible improvement of our work is to include the possibility to predict the activity of a whole ECG signal instead of just one heartbeat, as we explained in the other experimental approaches in chapter V. A possible future implementation could be to predict a portion of signal containing more heartbeats using an ensemble method that exploits the labels returned for each heartbeat by our model in order to make a more robust classification performance.

To conclude the reading we wanted to linger on what we learned during the development of the project. The first contribution regarded the field of ECG signals: to reach good results we firstly had to well understand how the data describes the heart behaviour. In a second phase we had the opportunity to explore different architectures and to get in deep in their understanding, also for the ones that we didn't eventually used, such as the attention mechanisms. We can't hide the fact that we had to get through many different experiments in order to obtain the results that we wanted, also because the most of the time, probably due to the small amount of samples of the dataset, restricted size architectures were the ones that performed better. In the end, we can say that we overcame this problem by keeping a simple, in terms of layers and units, but also not trivial architecture, obtaining good performances in a new case scenario.

References

[1] T. Pawar, S. Chaudhuri, and S. Duttagupta, "Body movement activity recognition for ambulatory cardiac monitoring," *IEEE Trans Biomed Eng.*, May 2007.

[2] Z. Uddin and M. M. Hassan, "Activity recognition for cognitive assistance using body sensors data and deep convolutional neural network," *IEEE Sensors Journal*, Oct. 2019.

[3] K. Chen, L. Yao, X. Wang, D. Zhang, T. Gu, Z. Yu, and Z. Yang, "Interpretable parallel recurrent neural networks with convolutional attentions for multi-modality activity modeling," *IEEE* , Oct. 2018.

[4] E. Mohamed, J. Mirjam, and D. Boer, "Frequency bands effects on qrs detection," *DBLP* , Jan. 2010.

[5] M. F. Ferreira, R. Camacho, and L. F. Teixeira, "Autoencoders as weight initialization of deep classification networks applied to papillary thyroid carcinoma," *IEEE* , Jan. 2019.