Towards Automatic EEG Signal Denoising by Quality Metric Optimization

Arthur Sena Lins Caldas*, Eanes Torres Pereira*, Niago Moreira Nobre Leite†,
Arthur Dimitri Brito Oliveira† and Ellen Ribeiro Lucena†
*Department of Systems and Computing
†Department of Electrical Engineering
Federal University of Campina Grande, Campina Grande, Brazil
Email: arthur.caldas@ccc.ufcg.edu.br
eanes@computacao.ufcg.edu.br
niago.leite@ee.ufcg.edu.br
arthur.dimitri@ee.ufcg.edu.br
ellen.lucena@ee.ufcg.edu.br

***Abstract*—Electroencephalography (EEG) signals are widely used in areas such as: mental disease research, psychological evaluation and Affective Computing. One of the obstacles faced by researchers is related to EEG noise filtering. Involuntary muscle activity, such as eye-blinks and mandibular movements, insert noise into the signal which has a negative impact on its quality thus may result in misleading conclusions. Consequently, this study proposes an approach to remove noise in EEG signals based on a deep learning strategy that optimizes quality assessment algorithms. Furthermore, our methodology trains a model that learns how to optimize algorithms of quality assessment. In such manner, EEG signal users will not need human interference to extract noise which saves time and resources. To evaluate the robustness of our approach, we compare it with a baseline band-pass filter in terms of Peak Signal-to-Noise Ratio (PSNR), Structural Similarity index (SSIM) and quality scores. Our approach has demonstrated superior performance over the baseline technique in terms of PSNR and considering the second quality score applied. These are still preliminary results, yet they show great potential for continued development. Furthermore, this approach provides a new perspective on how to build deep learning methodologies in order to remove noise in EEG signals.**

***Index Terms*—Electroencephalography, Deep Convolutional Neural Networks, Denoising, Quality Metric, Signal Filtering**

I. INTRODUCTION

Emotion recognition which uses Electroencephalography (EEG) signals is one of the most investigated methods in the field of Affective Computing [1] [2]. The EEG signal represents the electrical record captured from one's cerebral activity and it is collected through electrodes attached to one's scalp. Using EEG to correctly identify the emotional state of humans can support psychiatric diagnosis, automate the evaluation of products and services and of course, improve our knowledge about the influence that emotions have in our brains. Therefore, it is important to ensure the quality of the signals as they may contain undesired noise that may lead to artifacts that are not originated in the brain. These artifacts may resemble cognitive activity and thus affect EEG results, and can be categorized into non-physiological and physiological. Non-physiological artifacts contaminate the signals through devices in the recording environment while physiological artifacts are generated by the subject itself.

Some well-known physiological artifacts are originated from eye blinking, jaw clenching, and ingestion of saliva. One of the ways to mitigate this interference is by controlling the environment in which the signals are being collected. Unfortunately, that approach can not reduce all external influence, in addition to being unwieldy for real-life applications. Considering these facts, there is a strong need for developing methods of noise identification and removal. Though great improvements have been achieved, most of the techniques offer good performance removing only particular artifacts as stated by Jiang et al. [3]. For that reason, a more general artifact removal approach will have a positive impact on decreasing the required effort for the EEG filtering process.

In light of this, all EEG signal users can benefit from our proposal of developing a method for training a deep neural network to automatically perform EEG denoising without supervision. Furthermore, our methodology trains a model that learns how to optimize algorithms of quality assessment. In such manner, EEG signal users will not need human interference to extract noise which saves time and resources. On that account, the success of this study can deepen our knowledge of how to create DL approaches to extract EEG artifacts and add them to the preprocessing pipeline of emotional classification.

A. Signal Noise Evaluation

The distortion caused by the noise on the signal must be analyzed and a common way to this is the Peak Signal-to-Noise-Ratio (PSNR), which evaluates the rate between maximum power of the signal and the noise's power, to see how this one is corrupting the signal's integrity. The Equation 1 shows how PSNR is defined.

$$PSNR = 10 \times \log_{10} \frac{MAX_i^2}{MSE} \tag{1}$$

where MAX_i indicates the maximum amplitude value for the channel without noise and MSE is the Mean Square Error between the channel without noise and the noisy channel after filter procedure.

Another metric that we can use to evaluate the filtered signals is called Structural Similarity index (SSIM), which was introduced by Wang et al. [4] as an approach to measure similarity between two images. Considering two signals x and y to be compared, SSIM is defined as follows:

$$SSIM(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2\mu_y^2 + C_1} \cdot \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2\sigma_y^2 + C_2} \cdot \frac{\sigma_{x_y} + C_3}{\sigma_x\sigma_y + C_3} \quad (2)$$

where μ_x and σ_x denote the mean and the standard deviation of x, respectively; μ_y and σ_y, similarly, denote mean and standard deviation of y and σ_{x_y} represents the cross correlation between x and y. C_1, C_2, and C_3 are small positive constants values, introduced to guarantee numerical stability of aforementioned statistical parameters.

B. Related work

One of the classical and simple methods to remove noise from EEG signals is the band-pass Butterworth filter, although other techniques that use independent component analysis (ICA), principal component analysis (PCA), and a combination of adaptive filter algorithms [5] [6] [7] have also been exploited. However, these methods are not automated as human supervision is still needed. Thus, machine learning models can aid the process of filter automation. For example, Deep Learning (DL) models have demonstrated being a great asset in other research fields, but it is still in its early stages inside the field of EEG filtering. Alarcão and Fonseca's research [2] reinforces this idea, arguing that support vector machines are employed in 59% of the EEG emotional classification cases. Yet, Yang et al. [8] suggest that the correct application of deep learning to extract EEG noise has significant potential. In their investigation, they propose a method to remove Electrooculography (EOG) artifacts by applying a deep learning model, an autoencoder composed by three hidden layers. The main idea is to build an autoencoder capable of identifying features encountered only in noiseless EEGs. Afterward, contaminated signals were filtered by the autoencoder. To evaluate the robustness of their approach, they compared it to independent component analysis (ICA), kurtosis-ICA (K-ICA), second-order blind identification (SOBI) and a shallow neural network. Results showed that in most cases, autoencoder's accuracy outperformed other methods.

Following a similar approach, Leite et al. [9] implemented an autoencoder model that is composed of convolutional layers. Additionally, the authors investigated two classes of physiological artifacts: eye blinking and jaw clenching. They concluded that the autoencoder method outweighed the band-pass Butterworth filtering in terms of peak signal-to-noise ratio (PSNR).

Similarly, Zheng et al. [10] applied a deep learning model, but in the context of EEG-based emotion recognition. They trained a deep belief network (DBN) which is integrated with a hidden markov model (HMM) to classify two categories of emotions (positive and negative). The proposed approach reached higher accuracy than state-of-the-art methods. These findings suggest that methodologies based on deep learning models are a great alternative to improve affective modeling.

Despite the success of the deep learning models, we can assert that there are still opportunities for further enhancements. For this reason, our study presents a new process, which uses a deep neural network to filter artifacts from EEG signals. This method aims to optimize the algorithms of quality assessment proposed by Mohamed et al. [11] so that they can be used as a tool to automate the process of noise removal. The advantage of this method is that it would be used as an unsupervised approach so that there is no need for a quality ground truth dataset. For evaluation purposes, we use PSNR and SSIM metrics along with the quality assessment algorithms. Moreover, we compare our method to a baseline band-pass filtering technique and the results achieved by Leite et al. [9] as this study is motivated by theirs. We can fragment this idea into the steps below:

- Implement Scores 1 and 2 as proposed by Mohamed et al. [11] as loss functions so they can be bound to a neural network.
- Compare the neural network results with a baseline method for noise filtering along with the results from Leite et al. [9] in terms of PSNR, SSIM and quality assessment algorithms.

II. Proposed Approach

This paper is inspired by the approach proposed by Leite et al. [9]. The main contribution of our approach is the optimization of quality scores as loss functions. This paper is organized as follows: Section II-A presents the used dataset along with the procedure proposed by Leite et al. [9] for creating synthetic EEG samples, Section II-B presents the quality scores which were used for loss optimization, and Section II-C presents details about network training.

A. Dataset

This research used two EEG datasets: one with noisy signals and another with clean signals. The noisy signal dataset was constructed by an experiment in which EEG sensors were disposed in the scalp of volunteers while they were asked to perform specific movements, such as: eye-blinking, looking left and right [12], raising arms, mandibular contraction [13], and swallowing [14]. The EEG equipment used was Neurovirtual's Brain Wave II with 25 channels [1] and the disposition of electrodes followed the 10-20 international standard.

The filtered signals were obtained from the DEAP dataset [15] which contains data from 32 subjects. The DEAP dataset was collected while volunteers watched one-minute video clips, and for each volunteer there were 40 channels, 32 of which contain EEG information.

[1]For more information about the equipment see the catalog available at: https://neurovirtual.com/br/wp-content/uploads/2016/10/Neurovirtual-Catalog-2019_PTG.pdf

In order to perform channel-wise data augmentation the approach proposed by Leite et al. [9] was employed. The approach is described in reduced fashion in the following steps:

1) Compute average and standard-deviation from one EEG channel signal $eeg[n]$ obtained from the real EEG signal dataset (this paper used the DEAP dataset);
2) Generate a white noise uniform distribution, $s[n]$, using statistical data obtained in step 1;
3) Compute the multiplication of power spectral densities of $eeg[n]$ and $s[n]$. The power spectral density for a discrete-time signal was obtained through an implementation of the method proposed by P. Welch [16] ;
4) Compute the amplitude $A[k]$, $k \in [1, m]$ per sample using the results in step 2, and the relative amplitude using Equation 3;
5) Generate random phases in the interval $[0, 2\pi]$ for the amplitudes acquired in step 4;
6) Compute the array $Z[\omega]$, in the discrete frequency domain, using amplitudes and phases from steps 4 and 5;
7) Obtain the synthetic signal applying the inverse Fourier Transform to the array $Z[\omega]$;
8) Correct any dimensionality incoherences by means of interpolation.

$$A[k] = \sqrt{2 \cdot PSD_{eeg}[k] \cdot PSD_s[k]} \tag{3}$$

B. Signal Quality Scores

The main problem in EEG signal filtering is that most of the times, it is not possible to know in advance if the signal is noisy and if it is, which type(s) of noise it has. Therefore, instead of dealing with specific approaches for noise filtering, one should be interested in the EEG signal quality. In order to support the evaluation of EEG signal quality, Mohamed et al. [11] proposed six quality scores for EEG signals based on frequency bands, from which we used the following two for loss function optimization:

- **Score 1**: For each signal channel, a histogram is constructed from the frequency values of the channel amplitudes. Ideally, the format of this histogram should indicate that the values increase towards a maximum value and decrease a single time. If, for a given signal, that pattern is inconsistent or occurs multiple times, then the signal may have low quality.
- **Score 2**: If we assume that the EEG signal is normal and without noise, then the maximum amplitude should occur in channels O1, O2, P3, P4, T5, T6, C3, C4, A1, A2, T3 and T4. The final score is computed from the position of those channels inside a channel array ordered by maximum amplitude.

Each score is plugged in the loss function, whose output is given according to Equation 4.

$$Loss = \frac{1}{n}\Sigma_{i=1}^{n}\Big(Score_{MAX} - Score_i\Big)^2 \tag{4}$$

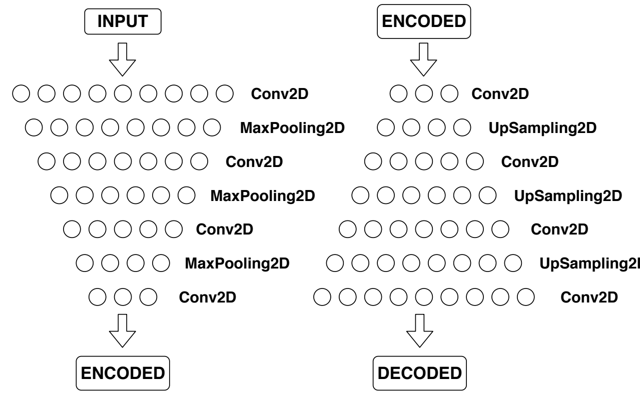

Fig. 1. Deep convolutional autoencoder architecture used in this research.

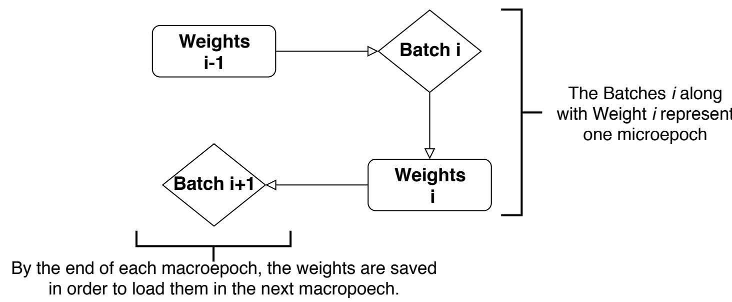

Fig. 2. Workflow structure used in the training process.

where $Score_{MAX}$ is constant and represents the maximum value from the respective score, and $Score_i$ refers to the score calculated from the actual signal.

C. Network Training

The Deep Convolutional Autoencoder (DCAE) architecture used in this research is similar to that proposed by Leite et al. [9], and its structure is presented in Figure 1 [2].

The parameters used for training were:

- Activation Function: tanh (Hyperbolic Tangent)
- Convolution Kernel Size: 8;
- Number of Filters: 100;
- Not using regularization;
- Loss Functions: Score 1 and Score 2.
- Optimizer: SGD (Stochastic Gradient Descent)
- Learning Rate: 0.0001;

III. Experimental Results

Two DCAEs were trained and evaluated using the proposed approach. The first DCAE was trained for filtering blink noise using Score 2 as the loss function. The second DCAE was trained for filtering mandibular contraction noise using Score 1 as the loss function. However, the same hardware limitation described by Leite et al. [9] was experienced in this work as we could not load all the signals at once while training the networks. Therefore we opted to follow the approach proposed

[2]All Deep Learning code was implemented using Keras Deep Learning Framework available at https://keras.io/ for Python 3.x

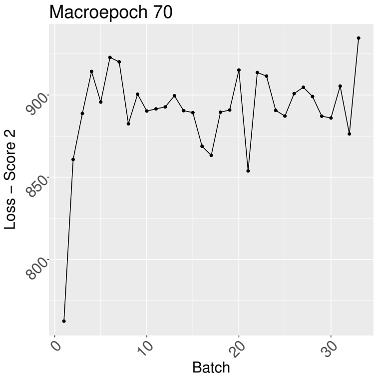

Fig. 3. Loss progression for the last macroepoch of the training for eye blink and using Score 2.

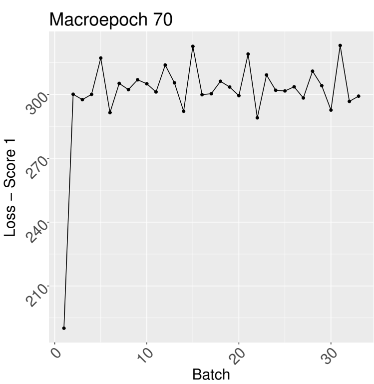

Fig. 4. Loss progression for the last macroepoch of the training for jaw clenching and using Score 1.

by Leite et al. [9] for training with fragments the set of signals into batches and uses each batch to train the network during one epoch. This process is described by the Figure 2. Once all the batches have gone through the network, a "macroepoch" is finished. The last macroepoch loss progression for both DCAEs can be seen in Figures 3 and 4.

Both autoencoders were trained with 70 macroepochs. The pattern of loss over the training process provides evidence of not being adequately optimized for both scores. In other words, there are still many opportunities for further enhancements. After training, the evaluation of the networks was performed by computing the Peak Signal-to-Noise-Ratio (PSNR) and the Structural Similarity index (SSIM), to strengthen our results. The PSNR equation can be seen as Equation 1.

The data used for evaluation was filtered using two approaches: a DCAE and a Butterworth filter in the band [4, 45]Hz. Tables I and II present the mean and standard-deviation per channel of the PSNRs. Each of those tables has a column η which represents the difference between the average PSNR for the DCAE filtered signal and the signal filtered by the baseline approach (Butterworth). A positive η indicates that the DCAE approach achieved higher filtering quality. Moreover, there is the column η_{prior} which represents the average PSNR values found by Leite et al. [9] in their study. Comparing η with η_{prior} we can conclude that the approach proposed by Leite et al. [9] reached higher values which means that although our approach overcomes the baseline, Leite et al. [9] method had even superior performance in comparison with the same baseline technique. Nevertheless, our research does not need a clean set of EEG signals during training, which is an advantage over their study.

TABLE I
AVERAGES AND STANDARD-DEVIATIONS OF PSNR FOR EYE-BLINK NOISE

Channels	**Filtered by DCAE**		**Baseline Filtered**		η	η_{prior}
	$\overline{x}$(dB)	s(dB)	$\overline{x}$(dB)	s(dB)		
1 (Fp1)	15.198	2.280	12.938	1.964	2.260	4.024
2 (Fp2)	16.390	2.917	13.707	2.570	2.683	4.022
3 (F3)	15.881	2.232	13.510	1.942	2.371	5.510
4 (F4)	16.058	2.694	13.536	2.643	2.522	6.228
5 (C3)	16.884	2.910	14.141	2.556	2.743	5.126
6 (C4)	16.422	2.912	13.859	2.532	2.563	5.772
7 (P3)	16.078	2.660	12.964	2.176	3.114	4.523
8 (P4)	16.383	2.343	13.633	2.179	2.750	4.988
9 (O1)	15.382	2.053	12.333	1.606	3.049	3.885
10 (O2)	15.509	2.054	12.448	1.760	3.061	4.394
11 (F7)	15.514	2.338	13.062	2.116	2.452	3.874
12 (F8)	16.125	3.227	13.566	2.788	2.559	4.134
13 (T3)	15.985	2.471	13.184	2.300	2.801	4.052
14 (T4)	16.181	2.577	13.375	2.236	2.806	3.646
15 (T5)	16.045	3.125	12.929	2.530	3.116	4.069
16 (T6)	16.051	2.776	13.080	2.341	2.971	4.874
17 (Fz)	16.386	2.486	13.784	2.419	2.602	5.875
18 (Cz)	16.543	2.961	13.797	2.541	2.746	6.036
19 (Pz)	16.012	3.092	13.044	2.500	2.968	4.573

TABLE II
AVERAGES AND STANDARD-DEVIATIONS OF PSNR FOR MANDIBULAR CONTRACTION NOISE

Channels	**Filtered by DCAE**		**Baseline Filtered**		η	η_{prior}
	$\overline{x}$(dB)	s(dB)	$\overline{x}$(dB)	s(dB)		
1 (Fp1)	15.360	2.079	12.596	1.934	2.764	5.125
2 (Fp2)	15.943	2.742	13.394	2.609	2.549	4.374
3 (F3)	15.888	2.289	13.223	2.074	2.665	6.612
4 (F4)	16.038	2.904	13.411	2.524	2.627	6.814
5 (C3)	16.837	2.926	13.891	2.559	2.946	5.902
6 (C4)	16.343	2.705	13.972	2.624	2.371	7.397
7 (P3)	16.237	2.932	12.300	1.913	3.937	4.947
8 (P4)	16.445	2.447	12.917	1.882	3.528	4.988
9 (O1)	15.372	2.091	11.520	1.462	3.852	3.359
10 (O2)	15.233	1.884	11.857	1.497	3.376	4.081
11 (F7)	15.718	2.298	12.886	1.938	2.832	4.591
12 (F8)	16.122	3.272	13.353	2.752	2.769	5.136
13 (T3)	16.116	2.692	13.230	2.421	2.886	5.163
14 (T4)	16.119	2.472	13.369	2.337	2.750	4.866
15 (T5)	16.235	3.198	12.407	2.476	3.828	4.293
16 (T6)	15.974	2.677	12.483	2.195	3.491	4.243
17 (Fz)	16.519	2.604	13.582	2.467	2.937	7.810
18 (Cz)	16.694	2.899	13.530	2.571	3.164	7.479
19 (Pz)	16.011	2.975	12.594	2.516	3.417	5.497

Additionally, confidence intervals were computed for each filtered channel's PSNR, in order to show the statistical difference between the results presented by DCAE versus the baseline method (Butterworth) and can be seen in Figures 5 and 6. The confidence interval for the original signal's PSNR before the filtering process was also calculated. Inspecting the confidence intervals, it can be seen that our approach has a higher performance than the baseline method in terms of PNSR. However, it should be pointed out that both filtering methods had a lower PSNR than the original signal, meaning that the filtering process also distorts it.

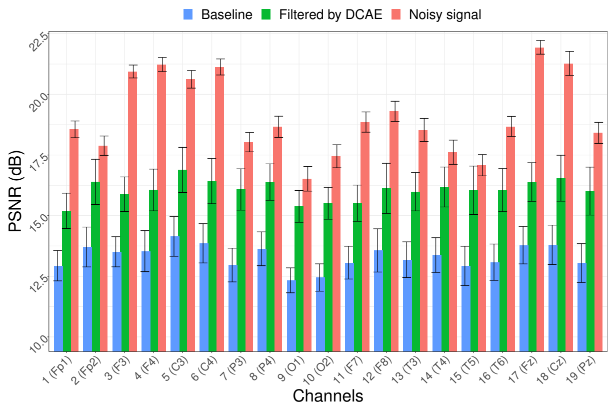

Fig. 5. Confidence intervals for PSNR average per channel, using Score 2 as loss function applied to signals containing eye blink noise.

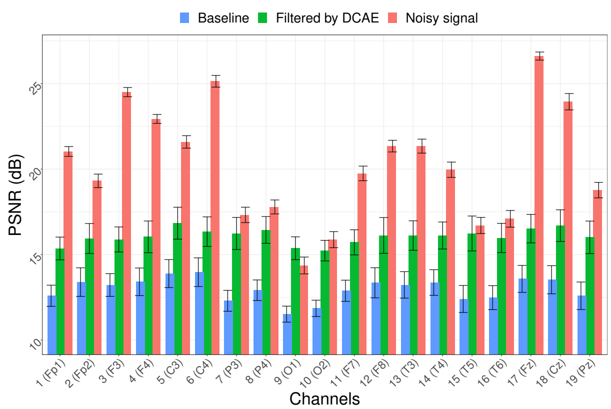

Fig. 6. Confidence intervals for PSNR average per channel, using Score 1 as loss function for filtering mandibular contraction noise from EEG signals.

TABLE III
AVERAGES AND STANDARD-DEVIATIONS OF SSIM FOR EYE-BLINK NOISE

Channels	Filtered by DCAE		Baseline Filtered		η
	$\overline{x}$(dB)	s(dB)	$\overline{x}$(dB)	s(dB)	
1 (Fp1)	0.022	0.021	0.040	0.040	-0.018
2 (Fp2)	0.043	0.051	0.076	0.059	-0.033
3 (F3)	0.019	0.016	0.040	0.031	-0.021
4 (F4)	0.016	0.032	0.062	0.053	-0.046
5 (C3)	0.031	0.033	0.067	0.044	-0.036
6 (C4)	0.027	0.023	0.059	0.045	-0.032
7 (P3)	0.036	0.030	0.046	0.028	-0.010
8 (P4)	0.020	0.016	0.045	0.028	-0.025
9 (O1)	0.016	0.020	0.025	0.020	-0.009
10 (O2)	0.030	0.027	0.031	0.030	-0.001
11 (F7)	0.014	0.014	0.039	0.036	-0.025
12 (F8)	0.037	0.044	0.059	0.051	-0.022
13 (T3)	0.028	0.028	0.049	0.030	-0.021
14 (T4)	0.030	0.029	0.051	0.032	-0.021
15 (T5)	0.027	0.026	0.040	0.029	-0.013
16 (T6)	0.021	0.022	0.034	0.034	-0.013
17 (Fz)	0.017	0.020	0.054	0.048	-0.037
18 (Cz)	0.021	0.020	0.054	0.038	-0.033
19 (Pz)	0.018	0.022	0.052	0.036	-0.030

TABLE IV
AVERAGES AND STANDARD-DEVIATIONS OF SSIM FOR MANDIBULAR CONTRACTION NOISE

Channels	Filtered by DCAE		Baseline Filtered		η
	$\overline{x}$(dB)	s(dB)	$\overline{x}$(dB)	s(dB)	
1 (Fp1)	0.025	0.022	0.037	0.041	-0.012
2 (Fp2)	0.020	0.022	0.065	0.049	-0.045
3 (F3)	0.017	0.018	0.040	0.031	-0.023
4 (F4)	0.030	0.034	0.072	0.060	-0.042
5 (C3)	0.032	0.035	0.066	0.045	-0.034
6 (C4)	0.035	0.021	0.088	0.078	-0.053
7 (P3)	0.047	0.038	0.048	0.031	-0.001
8 (P4)	0.018	0.040	0.047	0.027	-0.029
9 (O1)	0.026	0.033	0.030	0.022	-0.004
10 (O2)	0.024	0.021	0.031	0.024	-0.007
11 (F7)	0.025	0.019	0.050	0.045	-0.025
12 (F8)	0.032	0.048	0.064	0.055	-0.032
13 (T3)	0.032	0.041	0.064	0.048	-0.032
14 (T4)	0.019	0.020	0.061	0.047	-0.042
15 (T5)	0.027	0.045	0.045	0.031	-0.018
16 (T6)	0.021	0.034	0.030	0.029	-0.009
17 (Fz)	0.037	0.040	0.054	0.050	-0.017
18 (Cz)	0.033	0.026	0.053	0.039	-0.020
19 (Pz)	0.037	0.036	0.048	0.035	-0.011

Following the same strategy of comparison, we calculated the values for each channel using the SSIM metric. Tables III and IV present the mean and standard-deviation per channel of the SSIMs.

Regarding the average SSIM difference between baseline and our approach, we see that in most cases this difference was negative which indicates a better result from baseline. However, it should be pointed out that this difference seems to be insignificant as the values are close to zero. This conclusion is also corroborated by the confidence intervals in Figures 7 and 8. The baseline method seems to have a slight advantage over our approach although in most cases the confidence intervals overlap. That being said we can see the importance of using two different metrics of evaluation so we can draw more realistic conclusions from the results of our model.

Furthermore, Scores 1 and 2 were also computed for filtered signals. As DCAEs were trained for optimizing those scores, one may hope the scores for DCAE filtered signals are higher than for baseline filtered signals. Results for each score after signal filtering may be seen in Tables V and VI. Analyzing the results from Score 1, it can be seen that after the process of filtering jaw clenching noise by using DCAEs the set of test signals has a great and much higher improvement than the baseline filtering method as the average scores (α = 95%) are 84.87 and 36.29, respectively. This indicates that although there was not a significant decrease in the pattern of the loss function in the training process, the autoencoder was capable of optimizing the score. On the other hand, the optimization process for Score 2 was not so successful as the average scores for DCAE and baseline are 73.00 e 64.87, respectively. In this case, both methods achieved similar performance. To better understand these results, the Figures 9 and 10 show confidence intervals for each score after and before the filtering processes.

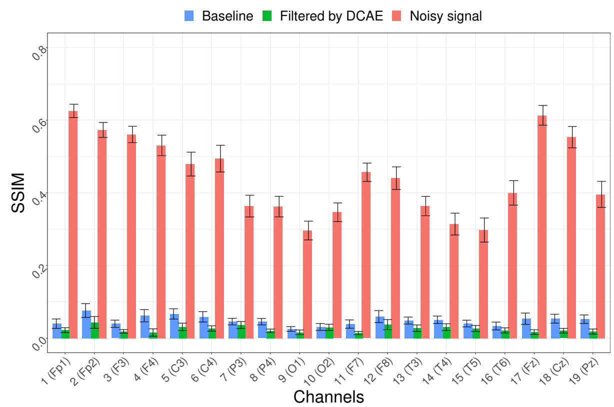

Fig. 7. Confidence intervals for SSIM average per channel, using Score 2 as loss function applied to signals containing eye blink noise.

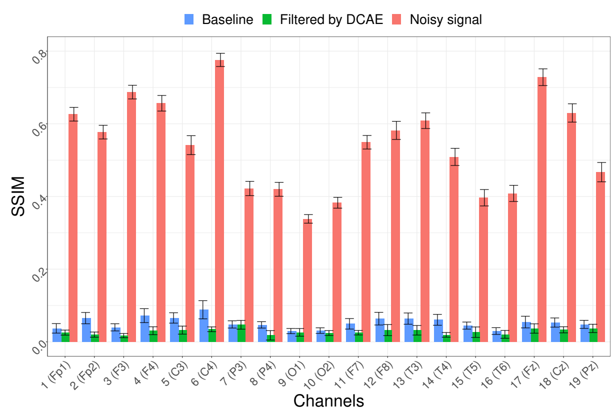

Fig. 8. Confidence intervals for SSIM average per channel, using Score 1 as loss function for filtering mandibular contraction noise from EEG signals.

TABLE V
SCORE 1 RESULTS WHEN APPLIED TO TEST DATA FILTERED BY BASELINE APPROACH AND BY PROPOSED APPROACH FOR MANDIBULAR CONTRACTION NOISE REMOVAL

Score 1					
Test no.	**Proposed**	**Baseline**	**Test no.**	**Proposed**	**Baseline**
1	88.547	31.359	*21*	85.118	31.277
2	83.766	39.541	*22*	85.538	45.299
3	85.424	40.437	*23*	84.534	39.652
4	80.201	31.741	*24*	86.865	31.254
5	82.452	40.995	*25*	85.950	39.394
6	86.542	43.728	*26*	84.092	34.404
7	86.787	42.313	*27*	81.377	37.828
8	85.594	27.814	*28*	86.910	26.865
9	88.080	40.259	*29*	82.179	37.680
10	85.114	25.596	*30*	82.484	38.898
11	88.321	43.053	*31*	83.058	39.374
12	82.156	32.273	*32*	86.695	29.773
13	85.700	37.590	*33*	80.306	35.299
14	82.211	39.915	*34*	81.178	34.550
15	83.195	34.300	*35*	87.355	31.384
16	86.680	41.850	*36*	85.782	39.013
17	88.829	31.781	*37*	81.804	36.923
18	90.898	38.545	*38*	87.859	32.136
19	82.113	35.182	*39*	82.802	39.303
20	84.936	40.141	*40*	85.392	32.772

	Filtered by DCAE	**Baseline Filtered**
μ (α=95%)	84.87	36.29
σ	2.57	4.86

TABLE VI
SCORE 2 RESULTS FOR TEST DATA FILTERED USING EYE BLINK NOISE FILTERING PROPOSED NETWORK AND BASELINE FILTERING

Score 2					
Test no.	**Proposed**	**Baseline**	**Test no.**	**Proposed**	**Baseline**
1	75	65	*21*	75	75
2	85	65	*22*	80	60
3	75	85	*23*	75	60
4	70	55	*24*	70	60
5	75	70	*25*	75	70
6	80	65	*26*	80	45
7	70	70	*27*	55	65
8	70	60	*28*	75	60
9	70	65	*29*	75	65
10	70	65	*30*	70	60
11	85	70	*31*	80	65
12	70	65	*32*	70	60
14	75	50	*33*	65	65
14	70	70	*34*	75	65
15	65	65	*35*	70	70
16	65	65	*36*	85	55
17	70	65	*37*	70	65
18	80	60	*38*	70	75
19	65	80	*39*	70	70
20	75	80	*40*	75	50

	Filtered by DCAE	**Baseline Filtered**
μ (α=95%)	73.00	64.87
σ	6.07	7.96

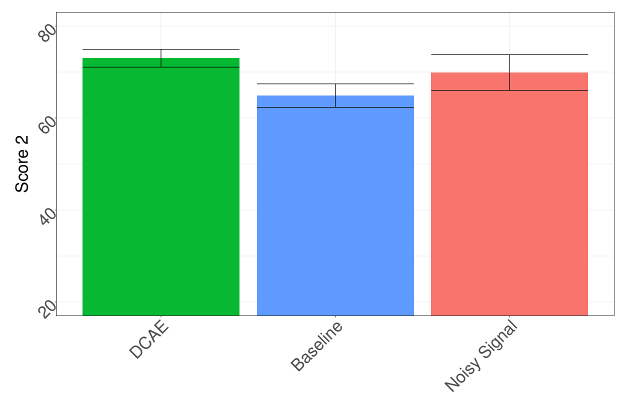

Fig. 9. Confidence intervals for Score 2 calculated before and after the filtering process of eye blink noise.

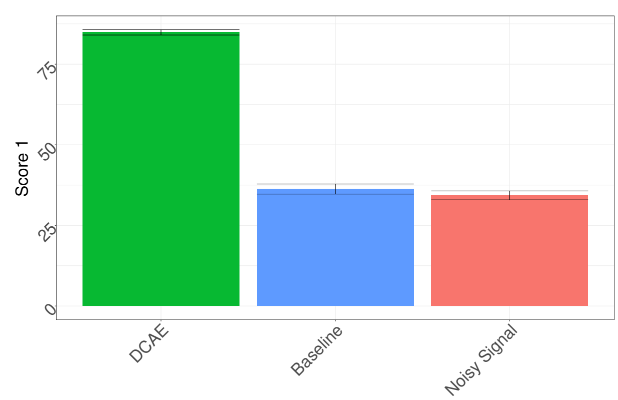

Fig. 10. Confidence intervals for Score 1 calculated before and after the filtering process of mandibular contraction noise.

IV. Conclusion

This paper presented results of two implementations of a deep learning autoencoder which optimize two algorithms of quality assessment in order to denoise EEG signals. The main contribution of the proposed approach is the fact that this method does not require a ground-truth dataset as it is an unsupervised technique. Although the training process has not properly optimized the quality scores, the network filtering had superior performance than the baseline filtering method in terms of PSNR. Likewise, a similar conclusion is found when comparing the values of the first quality score algorithm calculated after the filtering process from both methods. However, it should draw our attention to the fact that in terms of PSNR the autoencoder proposed by Leite et al. [9] achieved better results than ours.

For future work, there are some improvement possibilities. One of them is the experimentation with other network architectures. Leite et al. [9] state that network architectures along with other classes of layers may increase the quality of the filtering process. One of the reasons for which loss function has had a weak decrease over the epochs might be the architecture utilized. Furthermore, there were hardware limitations that prevented us from quickly experimenting with the train and test process. We aim to continue this research by experimenting with other architecture networks along with more computing power. We also plan to investigate goal-oriented DL approaches, such as the improvement of emotion classifiers which use EEG signals as input.

Finally, it may be concluded that the results of this study corroborate the Leite et al. methodology [9] as it was also possible to achieve a deep learning model with better performance than the baseline method evaluated for signal filtering considering the PSNR metric. On top of that, there is great potential for improving the network optimizing process. We believe that encountering the correct set of layers and parameters can surely have a great performance improvement.

V. Acknowledgment

This study was financed in part by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), with resources from Chamada Pública MCTI/CNPQ/Universal 14/2014.

References

[1] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion recognition from eeg using higher order crossings," *IEEE Transactions on Information Technology in Biomedicine*, pp. 186–197, 2010.

[2] S. M. Alarcão and M. J. Fonseca, "Emotions recognition using eeg signals: A survey," *IEEE Transactions on Affective Computing*, pp. 374–393, 2019.

[3] J. Xiao, B. Gui-Bin, and T. Zean, "Removal of artifacts from eeg signals: A review," *Sensors*, p. 987, 2019.

[4] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, pp. 600–612, 2004.

[5] T. P. Jung, S. Makeig, M. J. McKeown, A. J. Bell, T. W. Lee, and T. J. Sejnowski, "Imaging brain dynamics using independent component analysis," *Proceedings of the IEEE*, pp. 1107–1122, 2001.

[6] V. Roy and S. Shukla, "Mth order FIR filtering for eeg denoising using adaptive recursive least squares algorithm," in *2015 International Conference on Computational Intelligence and Communication Networks (CICN)*, 2015, pp. 401–404.

[7] D. Kang and L. Zhizeng, "A method of denoising multi-channel eeg signals fast based on pca and debss algorithm," in *2012 International Conference on Computer Science and Electronics Engineering*, 2012, pp. 322–326.

[8] Y. Banghua, D. Kaiwen, F. Chengcheng, H. Chenxiao, and W. Jinlong, "Automatic ocular artifacts removal in eeg using deep learning," *Biomedical Signal Processing and Control*, pp. 148–158, 2018.

[9] N. M. N. Leite, E. T. Pereira, E. C. Gurjao, and L. R. Veloso, "Deep convolutional autoencoder for eeg noise filtering," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2018, pp. 2605–2612.

[10] W.-L. Zheng, J.-Y. Zhu, Y. Peng, and B.-L. Lu, "EEG-based emotion classification using deep belief networks," in *2014 IEEE International Conference on Multimedia and Expo (ICME)*, 2014, pp. 1–6.

[11] S. Mohamed, S. Haggag, S. Nahavandi, and O. Haggag, "Towards automated quality assessment measure for eeg signals," *Neurocomputing*, 2017.

[12] B. Fisch and R. Spehlmann, *Basic Principles of Digital and Analog EEG*. Elsevier, 1999.

[13] J. W. Britton, L. C. Frey, J. L. Hopp, P. Korb, M. Z. Koubeissi, W. E. Lievens, E. M. Pestana-Knight, and E. K. St. Louis, *Electroencephalography (EEG): An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children, and Infants*. American Epilepsy Society, Chicago, 2016.

[14] K. Matsuo and J. B. Palmer, "Anatomy and physiology of feeding and swallowing: Normal and abnormal," *Physical medicine and rehabilitation clinics of North America*, pp. 691–707, vii, 2008.

[15] S. Koelstra, C. Muhl, M. Soleymani, J. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis; Using physiological signals," *IEEE Transactions on Affective Computing*, pp. 18–31, 2012.

[16] P. D. Welch, "The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, pp. 70–73, 1967.