WEEGNET: AN WAVELET BASED CONVNET FOR BRAIN-COMPUTER INTERFACES

Anonymous authors
Paper under double-blind review

ABSTRACT

Brain-computer interfaces (BCI) are systems that link the brain with machines using brainwaves as a medium of communication using electroencephalography to explore the brain activity which is an affordable solution, noninvasive, easy setup, and portability. However, the neural signals are noisy, non-stationary, and nonlinear where the processing of those signals in a pattern recognition problem needs a complex pipeline of preprocessing, feature extraction, and classification algorithms that need an apriori knowledge to avoid compatibility issues and a deep understanding of the studied signals. Moreover, some techniques need a huge computational power on the CPU and a huge size of RAM. Therefore, several papers proposed to use Deep Learning to get state of the art performance and visualization of the learned features to have more understanding about the neural signals. But, the convolutional neural network (Convnet) are not used properly and the results are often random when we reproduced the works. Hence, we propose a combination of the discrete wavelet transform (DWT) and a Convnet that processes raw EEG data. The DWT will be used to reduce the size of the data without losing useful information. Also, a modified version of EEGNET will be used to extract the features and classification.

1 INTRODUCTION

Brain-computer interface (BCI) is an instrument that aims to link the brain with machines by translating the brain activities that are recorded with electroencephalography (EEG) into useful information Nicolas-Alonso & Gomez-Gil (2012); Teplan (2002). This comes possibly from several breakthroughs in neurosciences, cognitive sciences, and signal processing that permits to discover the underlying phenomena several cognitive events and their neural responses. To automatize the system, machine learning where used to create fully data-driven systems Lotte et al. (2007). This permits several applications such a neuroprosthesis, sleep monitoring system, or gamepad for videogames Nijholt (2008); Chambon et al. (2018); Zhang et al. (2017); Faust et al. (2018).

Several researchers classified the brain activities into multiple brainwave bands where each one is related to their cognitive states. The Motor Imagery (MI) is the neural response of the imagination of a movement (without physical action) which is the same as the same response as a real movementPfurtscheller & Neuper (2001). MI produces perturbations on the mu (also called sensorimotor rhythm or SMR) band [8,13] Hz and the beta band[13,30] HzNicolas-Alonso & Gomez-Gil (2012). When the amplitude of the band increase it called Event-Related Synchronization (ERS), and the opposed (the decease) is called Event-Related Desynchronization (ERD)Jeon et al. (2011).

The processing of the EEG signal is a complicated task: The EEG signals are a non-linear and non-stationary signal with a low signal-to-noise ratio. It is also affected by several types of artifacts like eyes-blink (EOG), heartbeat (ECG), and muscular contraction (EMG). Fortunately, there are several techniques to remove those noises but are complicated to set up like the EOG noises that need a human expert while using the independent component analysis. Also, there is a compatibility issue between the preprocessing methods, feature extraction techniques, and classifiers that complicated the implementation of such systems Lotte et al. (2007).

As a solution, several researchers proposed deep learning as a solution where they use a light preprocessing (filtering the signal and normalization), and let the deep neural networks to handle the feature extraction and classification Lawhern et al. (2018); Schirrmeister et al. (2017a). Moreover, the interpretability of the learned features can give several clues to understand the brain. And beyond that, deep learning can be a concrete solution to several BCI challenges. The neural response of each person is different overtime (session to session) and over subjects (intersubject) which a universal neural network that can generalize a model to work without specific data from a person could reduce time to train where a pre-entrained model will be trained with a low learning rate as proposed by Schirrmeister et al. (2017b); Farshchian et al. (2019).

In this paper, we introduce a new method based on DWT and Convnet. At the opposite of the existent method, we will feed our Convnet with raw DWT which aims to reduce the dimensionality of the input data where we only use the half without removing the essential frequency band. Then, we use a modified EEGNET that will process the data without the inconvenient of the original one. The new version will use more features to increase the capacity of the network and we use a higher dropout probability reduce any risk of overfitting. Also, this network is designed as it allows us to visualize filter weights to explain which parts of the brain are contributing to the final decision of the classification. We compare our results with filter bank common spatial pattern and EEGNET on the dataset IIa of the BCI competition IV which is and open dataset with the experimental protocol of Lawhern et al. (2018) as we aim to explore the ability of DWT to increase the performances on inter subject problems.

2 Related works

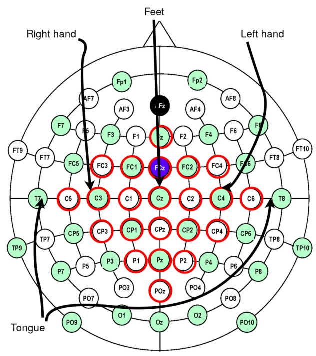
MI classification problems are solved with approaches that rely on the discrimination of frequency and spatial characteristics of the EEG signals. The state of the art technology is the filter bank common spatial pattern (FBCSP) Ang et al. (2012): The first step is a filter bank that filters the data into several non-overlapping subbands scaling mainly from 4 Hz to 40 Hz. The second step is the common spatial pattern that computes specific features that lead to optimal discrimination between the different classes. Some variants include feature selection methods that reduce the size of the feature vector. Finally, a classifier is used mainly linear discriminant analysis.

Then, some works were inspired from the FBCSP to create neural networks that have the same paradigm where the data is shaped into a matrix of multivariate signal $D_{C,T}$ (rows represent the channels, the column represents the timestamp). Cecotti & Graser (2011) proposed to replace the filter bank with a convolution performed with a kernel of size $(1, n_t)$, the CSP is replaced with a kernel with a size $(C, 1)$ where C is the number of the channels. Then a multilayer perceptron classifies the features extracted by the previous layers. Those works inspired Schirrmeister et al. (2017b) to propose several architectures, like a shallow architecture composed with the two convolutional layers then the classification layers, A deep architecture that includes more aggregative layer after the convolutional layers, and a ResNet with 31 layers. Those architectures were working well except that they include a huge number of parameters which Lawhern et al. (2018) reduced by creating EEGNET. EEGNET is a convnet that relies on Depthwise convolutional and separable convolution which permitted to reduce the number of the parameter using 796 parameters only for the EEGNET4,2. Those architectures were studied for several signals and EEGNET got state of the art performance. Those works are the only ones that use the convnet well and compared with the stat of the art techniques according to Lotte et al. (2018). EEGNET was compared on within-subject and inter-subject problem.

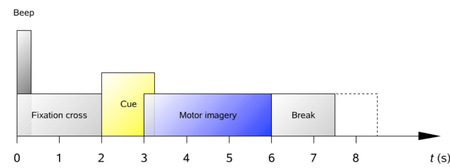
3 Experiment protocol

We use the cross-subject protocol for SMR signals as used by with the codes provided by the authors. The protocol relies on the dataset IIa of the BCI competition IV that consists of a 4 class problem (Left hand, right-hand, tongue, and both feet)Tangermann et al. (2012). The data is recorded with EEG headset composed of 22 electrodes that cover mainly the motor cortex and its surrounding areas. The signals were sampled at 250 Hz then filtered with a bandpass filter between 0.5Hz and 100Hz together with a notch filter to suppress the line noises. Three EOG sensors were used with the same configuration as the EEG. Fig 1a represents the position of every sensor.

In this dataset, nine subjects were comfortably sitting in an armchair and were facing a screen that gives them instructions following a scenario in the Fig 1b that was agreed on:



(a) Position of electrode for the dataset in red, corresponding area of the brain for the tasks of the dataset



(b) Timeline of the sceario

Figure 1: Extra material for the dataset: electrodes and timing

- At $t = 0$, a cross appears in the screen simultaneously with a beep, it signals to the subject to be ready for the MI.
- At $t = 2s$, the subject performs the MI of the body part that the orientation of an arrow that appear in the screen.
- At $t = 6s$, the screen turn blank, the subject interrupt the MI and start resting and waits for the next cross.

This operation is repeated 72 times for every class resulting in 288 trials in two sessions, the first is the training set and the second is the testing set.

As Lawhern et al. (2018) proposed, The data is resampled to 128 Hz and filtered between 4 Hz and 38 Hz, then an electrode-wise exponential moving standardization is performed as described in Schirrmeister et al. (2017b). The segment of 2s after the beginning of the MI is used. Since the protocol is cross-subject, the data is partitioned as follows: for subject i, we discard his original training set and we create a new training set from five other subjects, and a validation set from the three that remain. The testing set stays the same as the competition. In the end, we create 9 combinations of the training set, validation set, and test set composed respectively with 1440/864/288 trials. The same operation is repeated 10 times to create 90 cross-validation sets folds.

4 Proposed method

4.1 The ConvNet

As we changed our data from raw EEG signals into a first detail of the decomposition of the DWT, we add few changes to EEGNET of Lawhern et al. (2018), the architecture, described in Table 1, becomes as follow:

Firstly, we create a block that acts like FBCSP with a sequence of temporal filtering followed by spatial filtering. Essentially, we aim to learn spatial filters that are specific to frequency by the temporal filters: We use convolution with a F kernels of size $(1, 16)$ with padding where the kernels will extract the temporal proprieties of the input. Since the data was originally sampled at 128 Hz and we used the DWT, the data is now sampled at 64 Hz, that's why we used 16 instead of 32. Then, we use the depthwise convolution with kernels of size $(22, 1)$ where 22 represents the number of sensors, and the depth parameter is set to 2 to learn $2 * F$ filters. This layer will act like a regression where each coefficient represents a sensor. This configuration aims to spatially filtrate the data where sensors of high importance will get a high amplitude and the value of the others will be close to 0. To regularize this block, every convolution is followed with batch normalization and after each convolution, a dropout of probability $p = 0.5$ is used at the end (the original used a probability of $p = 0.25$), and the depthwise convolution will have a maximum norm constraint on its weights. No improvement has occurred when there is an activation between the two convolutions, we add one

Table 1: Details of the architecture

Block	Layer	#filter	size	Output	Option
1	Input			(1,C,T)	
	Conv2D	F	(1,16)	(F,C,T)	mode =same
	BatchNorm			(F,C,T)	
	DepthwiseConv2D	D*F	(C,1)	(D*F,C,T)	mode = valid,max norm = 1
	Batchnorm			(D*F,C,T)	
	Activation(ELU)			(D*F,C,T)	
	Maxpool2D		(1,2)	(D*F,C,T//2)	
	Dropout			(D*F,C,T//2)	p = 0.5
2	SeparableConv2D	2*F	(1,3)	(D*F,C,T//2)	mode = same
	BatchNorm			(D*F,C,T//2)	
	Activation (ELU)			(D*F,C,T // 4)	
	Dropout			(D*F,C,T // 4)	p = 0.5
	GlobalMaxPooling2D			D*F	
3	Dense	4		4	
	Softmax			4	max norm = 0.25

after the second batch normalization. A max-pooling is used between the activation and the dropout, average pooling seems to reduce the performance.

Secondly, we set up a block to aggregate the output of the preview layers and reduce the dimension: we use a separable convolution which consists of a depthwise convolution with $2 * F$ kernels with a size of $(1, 3)$ (the original uses $(1, 16)$ which in our case must be $(1, 8)$ due to subsampling, but no significative difference where occurred) followed with a pointwise convolution. Those convolutions aim to learn how to summarize the individual feature maps in the time dimension and then combine them. We use batch normalization, an activation, and dropout with the same probability as before. In opposite of the original version, we use a GlobalMaxpooling to drastically reduce the number of the parameter to $2 * F$.

Finally, we use a classification softmax with 4 units corresponding to the classes of the problem. the weights of the dense layer were regularized with a max norm constraint. Aggregative dense layers were not used to reduce the complexity and overfitting issues.

4.2 Hyperparameter and evaluation

All deep learning methods were trained in a $Nvidia\ Tesla\ T4$, with $CUDA$ 10, $Tensorflow$ 1.14.0, with $tf - keras$.

The hyperparameters are the corner stone of the learning of the training because a wrong choice can gives low result. Some works do not give a justification of the choices as reported in Lotte et al. (2018); in the following, we give some explanation about our choices in the Table

- Batch size as for EEGNET in Lawhern et al. (2018).
- ADAM Optimizer is a generalization of the others and is more efficient and faster. Kingma & Ba (2014).
- Categorical cross entropy is the most appropriate choice for multiclass problems
- 100 epochs was enough to achieve state of the art performance.
- Dropout probability is set to 0.5 as a result of manual tuning and as adviced by Baldi & Sadowski
- Learning rate 5×10^{-4} This value seems to be more efficient based on a manual tuning.
- Bias Only in dense layer As in the work of Lawhern et al. (2018).
- By manual fine tuning , we use $F = 32$, and $D = 2$.

We calculate several metrics to evaluate multiple aspects of the model. The only difference with the original protocol is that we are obligated to use the Wilcoxon sign-rank test instead of ANOVA

Table 2: Details of the results of different methods

metrics	ACC			Log Loss		
statistic	mean	median	var	mean	median	var
EEGNET	0.3921	0.3559	0.0170	1.3799	1.3425	0.1037
FBCSP	0.3578	0.3507	0.0070	1.5073	1.4567	0.0761
bior1.3	0.4147	0.3594	0.0182	1.2951	1.3137	0.0205
bior2.2	0.4168	0.3889	0.0166	1.2944	1.3039	0.0178
coif1	**0.4215**	0.3750	0.0182	1.2904	1.3073	0.0194
coif3	0.4144	0.3889	0.0170	1.2947	1.3206	0.0176
db1	**0.4225**	**0.3976**	0.0175	**1.2844**	**1.2972**	0.0184
db2	0.4180	0.3837	0.0176	1.2855	1.3040	0.0177

Table 3: Result of Wilcoxon test for the comparing

	Metrics	
methods	ACC	Log Loss
EEGNET-FBCSP	1.796139e-03	4.863454e-07
EEGNET-bior1.3	1.736546e-01	5.263838e-03
EEGNET-bior2.2	1.252494e-01	4.416667e-03
EEGNET-coif1	7.880995e-03	2.738729e-03
EEGNET-coif3	1.384940e-01	6.729246e-03
EEGNET-db1	1.892683e-02	1.552796e-03
EEGNET-db2	7.626364e-02	2.183704e-03
FBCSP-bior1.3	3.250105e-08	4.220340e-15
FBCSP-bior2.2	4.307781e-10	5.811730e-15
FBCSP-coif1	6.507993e-10	6.395253e-15
FBCSP-coif3	1.216967e-08	1.029638e-14
FBCSP-db1	4.106375e-11	2.777580e-15
FBCSP-db2	3.616928e-09	2.144278e-15
bior1.3-bior2.2	7.379606e-01	4.726164e-01
bior1.3-coif1	3.194718e-02	5.028582e-02
bior1.3-coif3	8.096274e-01	7.157506e-01
bior1.3-db1	1.303267e-01	5.396168e-03
bior1.3-db2	7.029243e-01	1.174073e-03
bior2.2-coif1	1.217876e-01	1.796284e-01
bior2.2-coif3	9.759437e-01	8.045547e-01
bior2.2-db1	4.567045e-02	1.067901e-02
bior2.2-db2	4.210724e-01	7.963280e-03
coif1-coif3	7.010800e-02	3.880928e-01
coif1-db1	9.900371e-01	1.170617e-01
coif1-db2	2.682423e-01	2.424513e-01
coif3-db1	5.043233e-02	3.647789e-03
coif3-db2	4.640355e-01	1.959992e-03
db1-db2	2.684949e-01	8.358381e-01

because our data do not follow the presumption of the test (normality) for the comparison between the methods. The Filter Bank Common Spatial Pattern (FBCSP) considered as the state-of-the-art technology is one of the most promising methods, we trained as follows: In the first step, a bank filter is used to filter the signals with several bandpass filters with 9 non overlapping filters a frequency range between 4Hz and 40Hz. In the second step, we compute two CSP filters for each band. Feature selection algorithm did not gives any improvement, we omit to use it. For the last step, the Linear Discriminant Analysis (LDA) trained with one-versus-others strategy.

Table 4: Comparison of different methods per subject

	EEGNET	FBCSP	bior1.3	bior2.2	coif1	coif3	db1	db2
S1	0.4743	0.4920	0.6021	0.5927	**0.6045**	0.5865	0.5990	0.5997
S2	**0.2861**	0.2812	0.2757	**0.2861**	0.2858	0.2816	0.2802	0.2778
S3	0.4819	0.3753	0.5726	0.5736	0.5847	0.5878	**0.5934**	0.5837
S4	0.3833	0.3743	0.3517	0.3795	0.3667	0.3740	**0.3972**	0.3819
S5	0.2646	0.2663	0.2594	0.2542	0.2576	0.2580	**0.2649**	0.2597
S6	0.2830	0.2892	0.3160	0.3108	**0.3177**	0.3132	0.3128	0.3122
S7	**0.3406**	0.3003	0.3205	0.3243	0.3326	0.3160	0.3250	0.3198
S8	**0.5694**	0.4573	0.5424	0.5538	0.5656	0.5392	0.5503	0.5372
S9	0.4455	0.3840	**0.4924**	0.4760	0.4778	0.4729	0.4795	0.4899

5 Experiment & Discussion

5.1 Performance comparison

To compare our methods, we used several types of wavelets. We chose the Biorthogonal wavelet 1.2 and 1.3, Daubechies wavelet db2 and db1 (also known as haar), and Coiflets wavelet coif1 and coif3. The notation $WEEGNET_{wav}$ where wav references the type of the used wavelet, we only reference the name of the wavelet in the tables Table 2, 3, and 4.

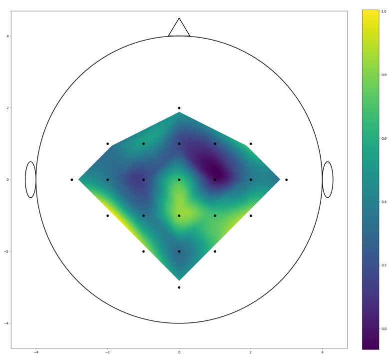
In Table 2, we compared the accuracy and the log loss of each method. We observe that $WEEGNET_{db1}$ provide the highest mean and median for the accuracy, and got the lowest log loss compared with all the other methods. The lowest accuracy and the highest log loss were recorded for $FBCSP$. $WEEGNET_{db2}$, $WEEGNET_{coif1}$, and $WEEGNET_{coif3}$ got close values for accuracy and log loss in the case of the mean and median. EEGnet results are $WEEGNET_{bior1.3}$ and FBCSP for the accuracy and $WEEGNET_{coif3}$ and FBCSP for the log loss.

To conclude on the supremacy of a method over others, we use the statistical test which is shown in Table 3. For the accuracy and over the $WEEGNET$ variants, we observe that the difference between $WEEGNET_{bior1.3}$ and $WEEGNET_{coif1}$ is statistically significant $(p < 0.05)$, the same for $WEEGNET_{bior2.2}$ and $WEEGNET_{coif1}$ advantaging $WEEGNET_{coif1}$. The difference between $WEEGNET_{coif1}$ and $WEEGNET_{coif3}$ and difference between $WEEGNET_{coif1}$ and $WEEGNET_{coif3}$ are marginally significative $(0.05 < p < 0.1)$ disadvantaging $WEEGNET_{coif3}$. For EEGNET, It was significatively outperformed by $WEEGNET_{coif1}$ and $WEEGNET_{db1}$ $(p < 0.05)$ and marginally by $WEEGNET_{db2}$. $FBCSP$ was outperformed by all the other methods with a very significative difference $(p < 0.005)$ even with it state of the art status. For the Logloss, $WEEGNET_{db1}$ and $WEEGNET_{db2}$ ouperform the $WEEGNET$'s variants based on biorthogonal wavelet, $WEEGNET_{coif3}$, EEGNET and FBCSP. FBCSP statisticly performed the worse method $(p < 5.10^{-7})$ compared with all other methods. EEGNET was only outperformed by the $WEEGNET$s.

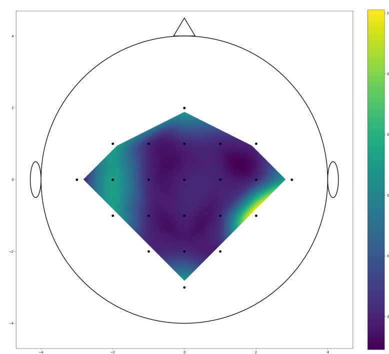
Table 4 exposes the accuracy of each method by subject. For the subject 1, $WEEGNET_{coif1}$ got the highest value followed $WEEGNET_{bior1.3}$. EEGNET and $WEEGNET_{bior2.2}$ were egal and reached the highest performance for subject 2. $WEEGNET_{db1}$ outperforms all methods for subject 2,3,and 4. $WEEGNET_{bior2.2}$ got the maximum value for subject 6 but it is very close to $WEEGNET$'s variant. EEGNET overcome the others for subject 7 and 8. $WEEGNET_{bior1.3}$ outperforms others on subject 9.

5.2 Visualization

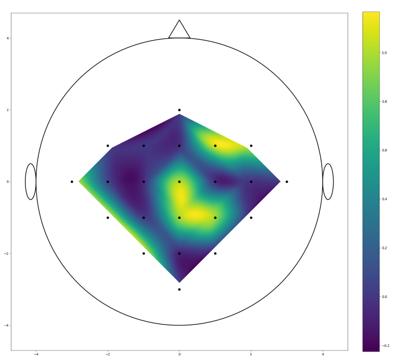
Visualization of the spatial filters can inform us about the spatial distribution of the interest of the network over the brain's area. Fig 2 represent an summarized spatial filter for some methods and with the help of Fig 1a, we can understand. First thing that we remark is that $WEEGNET_{db1}$ and $WEEGNET_{coif1}$ are very similare, but the difference is that the activation on some areas are maximal for $WEEGNET_{coif}$. there is an maximal for the electrodes Cz,CPz and CP2 for both methods. There is a strong interest in C5 CP3 and P1. $WEEGNET_{db1}$ has lower activation in CP4, P2 and around CF4. $WEEGNET_{coif1}$ has a strong activation on FC4. For $WEEGNET_{db2}$ and



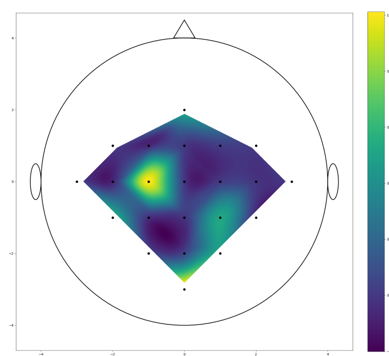
(a) $WEEGNET_{db1}$



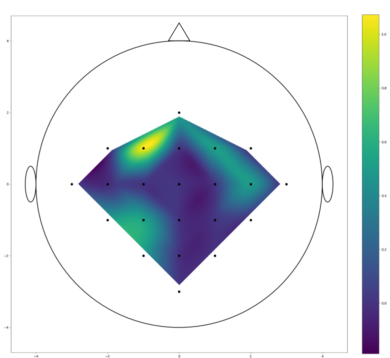
(b) $WEEGNET_{db2}$



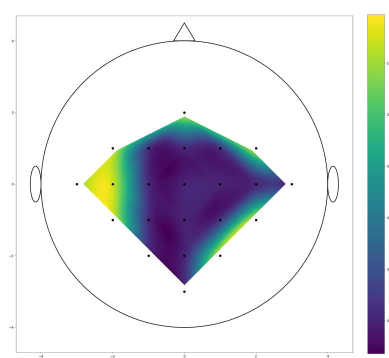
(c) $WEEGNET_{coif1}$



(d) $WEEGNET_{coif3}$



(e) $WEEGNET_{bior2.2}$



(f) $EEGNET$

Figure 2: text

EEGNET, we have similar pattern where have lower activation on the central area compared with the others. Both got a medium activation in the areas surrounding C4. $WEEGNET_{coif3}$ has a strong interest in the electrodes C3 and CP3, and a medium activation on CP4 and P2. $WEEGNET_{bior2.2}$ has a strong interest in the electrode CP1, and a medium activation the electrode surrouding the central area. The most succesfull methods are the ones that extract features from C4,C3, and Cz which are the areas of left hand, right hand, and feet.

6 DISCUSSION

We can explain the difference by comparing the theory of each method, the architecture of the Convnets, ad weights visualization.

If we compare the FBCSP and the convnet, they are based on the same paradigm where they use temporal filtering, then spatial filtering. But, the difference is that the Convnet learns temporal filters in the opposite of the FBCSP that use a filter bank with imposed parameter. This freedom gives the advantage to Convnet to shape temporal filtering that reduces the loss according to the given dataset. Lawhern et al. (2018) shows that the temporal filtering learns in most cases to extract two frequencies around mu band and beta band, which are the most imminent sign of the occurrence of an MI.

EEGNET is the result of the idea of creating a compact Convnet. The problem is that the reduction of the feature maps increases the risk of overfitting. Also, while increasing the number of feature maps, we observe that the network struggled with the learning and the average pooling was behind it. Smooth feature extraction of subject-specific neural response is not compatible with the cross-subject paradigm for EEG application, replacing with max-pooling was the solution where it reduces the noises and increasing the dropout probability to 0.5. While increasing the number of the feature maps, the network started to perform better which lead that the capacity of the convnet must be higher of EEG feature extraction. Also, global pooling reduces the number of features of the final dense input with no improvement on the accuracy. Moreover, our method needs only 100 epochs too train where EEGNET needs 500 epoch.

$WEEGNET_{db1}$ and $WEEGNET_{coif1}$ get the best performance across all the $WEEGNET$'s variants and their result. $WEEGNET$s that are based on Biorthogonal wavelet did not get important results. $WEEGNET_{db2}$ and $WEEGNET_{coif3}$ results were lower than their variants from the same families and $db1$ is the simplest wavelet, we conclude that more the wavelet is complex, more the result is low. The visualization of the spatial filters shows that the filters with several activation and mainly covering the central part increase the chance to perform better where we can explain that db1 wavelet is more generalizing the result as it result were high across all subject and outperform other methods in three cases which means that db1 is more compatible with extract subject-specific feature.

7 CONCLUSION

We introduced a novel method based on the DWT and a modified EEGNET. We exposed the weaknesses of the EEGNET and demonstrate that by increasing the number of the feature map, using max-pooling and increasing the dropout probability have a considerable influence on the result and speed of learning (epoch and size of the input). We used the DWT to reduce the size of the input without alteration the data. We found out that db1 wavelet and coif1 wavelet which are the simplest wavelet of the families, lead to the maximal accuracies and low losses as was found by Uyulan & Erguzel (2016). With the help of weights visualization, we could expose that the db1 and coif1 extract the features better than EEGNET. Even so, the results are low for several subjects due to a restricted dataset. In future studies, we will increase the size of the dataset with generative methods which gave encouraging results in several studiesGoodfellow et al. (2014); Hartmann et al. (2018); Abdelfattah et al. (2018).

REFERENCES

Sherif Abdelfattah, Ghodai Abdelrahman, and Min Wang. Augmenting The Size of EEG datasets Using Generative Adversarial Networks. May 2018. doi: 10.1109/IJCNN.2018.8489727. 00002.

Kai Keng Ang, Zheng Yang Chin, Chuanchu Wang, Cuntai Guan, and Haihong Zhang. Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. *Frontiers in neuroscience*, 6:39, 2012.

Pierre Baldi and Peter J Sadowski. Understanding Dropout. pp. 9.

Hubert Cecotti and Axel Graser. Convolutional neural networks for P300 detection with application to brain-computer interfaces. *IEEE transactions on pattern analysis and machine intelligence*, 33 (3):433–445, 2011.

Stanislas Chambon, Valentin Thorey, Pierrick J. Arnal, Emmanuel Mignot, and Alexandre Gramfort. DOSED: A deep learning approach to detect multiple sleep micro-events in EEG signal. *arXiv:1812.04079 [cs, eess, stat]*, December 2018.

Ali Farshchian, Juan A Gallego, Lee E Miller, Sara A Solla, Joseph P Cohen, and Yoshua Bengio. ADVERSARIAL DOMAIN ADAPTATION FOR STABLE BRAIN-MACHINE INTERFACES. pp. 14, 2019. 00001.

Oliver Faust, Yuki Hagiwara, Tan Jen Hong, Oh Shu Lih, and U Rajendra Acharya. Deep learning for healthcare applications based on physiological signals: A review. *Computer Methods and Programs in Biomedicine*, 161:1–13, July 2018. ISSN 01692607. doi: 10.1016/j.cmpb.2018.04.005.

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Networks. *arXiv:1406.2661 [cs, stat]*, June 2014.

Kay Gregor Hartmann, Robin Tibor Schirrmeister, and Tonio Ball. EEG-GAN: Generative adversarial networks for electroencephalograhic (EEG) brain signals. *arXiv:1806.01875 [cs, eess, q-bio, stat]*, June 2018. 00008.

Yongwoong Jeon, Chang S. Nam, Young-Joo Kim, and Min Cheol Whang. Event-related (De)synchronization (ERD/ERS) during motor imagery tasks: Implications for brain–computer interfaces. *International Journal of Industrial Ergonomics*, 41(5):428–436, September 2011. ISSN 0169-8141. doi: 10.1016/j.ergon.2011.03.005. 00067.

Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]*, December 2014.

Vernon J. Lawhern, Amelia J. Solon, Nicholas R. Waytowich, Stephen M. Gordon, Chou P. Hung, and Brent J. Lance. EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, 15(5):056013, October 2018. ISSN 1741-2552. doi: 10.1088/1741-2552/aace8c.

F Lotte, M Congedo, A Lécuyer, F Lamarche, and B Arnaldi. A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, 4(2):R1–R13, June 2007. ISSN 1741-2560, 1741-2552. doi: 10.1088/1741-2560/4/2/R01. 01903.

F Lotte, L Bougrain, A Cichocki, M Clerc, M Congedo, A Rakotomamonjy, and F Yger. A review of classification algorithms for EEG-based brain–computer interfaces: A 10 year update. *Journal of Neural Engineering*, 15(3):031005, June 2018. ISSN 1741-2560, 1741-2552. doi: 10.1088/1741-2552/aab2f2.

Luis Fernando Nicolas-Alonso and Jaime Gomez-Gil. Brain Computer Interfaces, a Review. *Sensors*, 12(2):1211–1279, January 2012. ISSN 1424-8220. doi: 10.3390/s120201211.

Anton Nijholt. BCI for Games: A 'State of the Art' Survey. In Scott M. Stevens and Shirley J. Saldamarco (eds.), *Entertainment Computing - ICEC 2008*, volume 5309, pp. 225–228. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. ISBN 978-3-540-89221-2 978-3-540-89222-9. doi: 10.1007/978-3-540-89222-9_29. 00003.

G. Pfurtscheller and C. Neuper. Motor imagery and direct brain-computer communication. *Proceedings of the IEEE*, 89(7):1123–1134, July 2001. ISSN 00189219. doi: 10.1109/5.939829.

Robin Tibor Schirrmeister, Lukas Gemein, Katharina Eggensperger, Frank Hutter, and Tonio Ball. Deep learning with convolutional neural networks for decoding and visualization of EEG pathology. *arXiv:1708.08012 [cs, stat]*, August 2017a.

Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. Deep learning with convolutional neural networks for EEG decoding and visualization: Convolutional Neural Networks in EEG Analysis. *Human Brain Mapping*, 38(11):5391–5420, November 2017b. ISSN 10659471. doi: 10.1002/hbm.23730.

Michael Tangermann, Klaus-Robert Müller, Ad Aertsen, Niels Birbaumer, Christoph Braun, Clemens Brunner, Robert Leeb, Carsten Mehring, Kai J. Miller, Gernot Mueller-Putz, Guido Nolte, Gert Pfurtscheller, Hubert Preissl, Gerwin Schalk, Alois Schlögl, Carmen Vidaurre, Stephan Waldert, and Benjamin Blankertz. Review of the BCI Competition IV. *Frontiers in Neuroscience*, 6, 2012. ISSN 1662-453X. doi: 10.3389/fnins.2012.00055.

M Teplan. FUNDAMENTALS OF EEG MEASUREMENT. *MEASUREMENT SCIENCE REVIEW*, 2:11, 2002. 01132.

Caglar Uyulan and Turker Erguzel. Comparison of Wavelet Families for Mental Task Classification. *The Journal of Neurobehavioral Sciences*, 3(2):59, 2016. ISSN 2148-4325. doi: 10.5455/JNBS.1454666348. 00001.

Xiang Zhang, Lina Yao, Quan Z Sheng, Salil S Kanhere, Tao Gu, and Dalin Zhang. Converting Your Thoughts to Texts: Enabling Brain Typing via Deep Feature Learning of EEG Signals. *arXiv preprint arXiv:1709.08820*, 2017.