A Deep Neural Network Classifier for Decoding Human Brain Activity Based on Magnetoencephalography

Abdullah Caliskan[1], Mehmet Emin Yuksel[1], Hasan Badem[2], Alper Basturk[2]
[1]Department of Biomedical Engineering, Erciyes University,
[2]Department of Computer Engineering, Erciyes University,
38039 Kayseri, Turkey
acaliskan@erciyes.edu.tr

Abstract—**Magnetoencephalography(MEG) is an emerging medical signal processing methodology that uses the magnetic field of brain to decode internal brain activity. However, MEG signals are very complicated and usually corrupted with significant amount of noise. Therefore, it is not easy to directly understand how the human brain responds to visual stimulus by analysing the MEG signals without utilizing advanced signal processing techniques such as feature extraction and classification. The feature extraction of MEG signals can be accomplished by applying the Riemannian approach. Moreover, the extracted features can be classified by classification algorithms such as SVM and KNN to complete the decoding process. However, these classification methods don't produce satisfying results as the number of features is very high. In this paper, the classification problem of MEG signals is addressed and a deep neural network based classifier is proposed to classify the MEG signals that were produced as the brain output for two different types of visual stimuli. The visual stimuli comprise a data base of faces and scrambled faces. Our experimental results demonstrate that the proposed classifier exhibits superior classification performance over the other competing methods used in the paper.**

Index Terms—**Autoencoder; brain decoding; deep neural network; magnetoencephalography; Riemannian approach.**

I. INTRODUCTION

The human brain has an excellent visual system, which has the ability to capture the "gist" of a scene within the millisecond time scale [1]. Therefore, several research studies have been made to predict the behaviour of the visual system by decoding brain state from neural response signals such as MEG [2]–[6]. MEG signals with many time series have been extensively utilized for understanding how the human brain works [2]. A strength of MEG is that it is suitable for studying human brain dynamics, and has a perfect temporal resolution (i.e. 1 kHz) [3].

Brain decoding finds a connection between a stimulus [7]–[9], e.g. a visual cue, and a mental activity, which can be predicted by MEG signals. Trials obtained by the signals and the category of the stimulus are collected to construct a classifier during a brain decoding experiment. The classifier can be used to decode the mental content of the human brain which maps the input stimulus to the MEG signals.

However, the construction of the classifier from MEG signals is a challenging process because of the complicated nature of the MEG signals, which are multi-channel time series with noise. Besides, the construction of the classifier not only needs a perfect feature extraction method, but also requires a good classification algorithm.

There exist several feature extraction methods for multi-channel time series such as Electroencephalography and MEG signals [10]–[12] to construct efficient classifiers. Recently, the Riemannian approach employed in the feature extraction process enables direct manipulation of multi-channel signal covariance matrices, which are used as features. Among existing feature extraction methods, Riemannian approach is highly competitive and usually superior to the other feature extraction methods such as Common Spatial Pattern and Linear Discriminant Analysis [12], [13].

In the last few years, due to their superior classification capability, the deep neural networks (DNN) [14]–[17] have been extensively utilized in classification problems. In most cases, the DNN has exhibited surprising classification performance over conventional classification methods [18], [19], including support vector machine (SVM), k-nearest neighbours (KNN), naive Bayes (NB) and decision trees (DT), because of its capability of generating new features from raw features [20], [21]. Therefore, DNNs can be used as a classifier for features extracted from MEG signals.

In this study, we present a deep neural network classifier for decoding human brain based on MEG. Indeed, this MEG decoding technique offers the ability of the Riemannian approach to reduce the dimension of the MEG signals and the capability of DNN to construct an efficient classifier which produces excellent results. DNN classifiers trained with high-level features clearly outperform state-of-the-art methods on a MEG data set.

The remainder of this paper is organized as follows: In Section II, we review the DNN classifier. In Section III, we describe how to apply our approach to MEG signals and we show our experimental results with statistical analyses. Conclusions are presented in Section IV.

II. METHOD

DNNs attempt to model a high-level abstraction in a data

set by using building blocks consisting of multiple nonlinear transformations [15]. The deep neural network classifier presented in this paper consists of an autoencoder and a softmax classifier.

A. Autoencoder

An autoencoder is a simple neural network shown in Fig. 1 which generates its own input. An autoencoder consists of two parts. The first one is an encoder that takes an input $\mathbf{x}$ and maps it to (usually) a lower dimensional representation $\mathbf{c}$. The second one is a decoder that tries to reconstruct the original signals $\hat{\mathbf{x}}$ [22], [23]. The aim of training an autoencoder is to minimize the error between input $\mathbf{x}$ and its output $\hat{\mathbf{x}}$ as much as possible by training the autoencoder, whose weights $\mathbf{W}$ and biases $\mathbf{b}$ are tuned with a convenient optimization algorithm. The objective function of the autoencoder consists of three parts as follows [24]

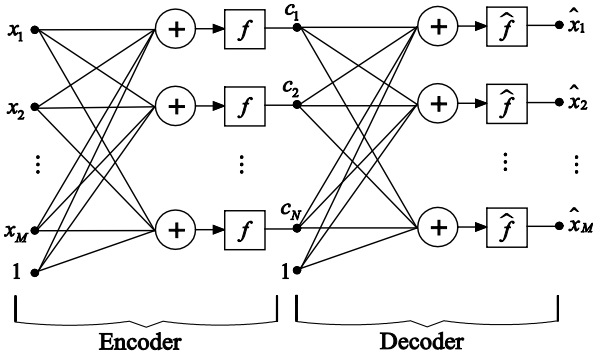$$E_T = E_M + E_R + E_S. \tag{1}$$

Fig. 1. An autoencoder network.

The first part E_M at the objective function is the mean square error which is given as

$$E_M = \frac{1}{S} \sum_{k=1}^{S} e_k^2, \tag{2}$$

where $e_k = \left\| \mathbf{x}^{(k)} - \hat{\mathbf{x}}^{(k)} \right\|$ for $k = 1 \ldots S$ and S is the number of the samples.

The second part is defined as follows

$$E_R = \frac{\lambda}{2} \left\| \mathbf{W} \right\|_2^2. \tag{3}$$

E_R has a regularization term λ used to prevent overfitting of the objective function.

The third part E_S imposes a sparsity constraint, which allows the autoencoder to discover interesting features in the hidden layer

$$E_S = \beta \sum_{j=1}^{N} KL\left(\rho \| \hat{\rho}_j\right), \tag{4}$$

where β is the weight of the sparsity penalty term and $KL\left(\rho \| \hat{\rho}_j\right)$ is the Kullback-Leibler divergence which is defined as follows [23], [24]

$$KL\left(\rho \| \hat{\rho}_j\right) = \rho \log \frac{\rho}{\hat{\rho}_j} + \left(1 - \hat{\rho}_j\right) \log \frac{1-\rho}{1-\hat{\rho}_j}, \tag{5}$$

where ρ is a constant sparsity parameter and $\hat{\rho}_j$ computed below is the mean activation value of jth neuron in the hidden layer of the autoencoder [24]

$$\hat{\rho}_j = \frac{1}{S} \sum_{i=1}^{S} f_j\left(\mathbf{x}^{(i)}\right), \tag{6}$$

where the activation function of the jth neuron of the hidden layer is f_j.

B. Softmax Classifier

A softmax classifier is a supervised layer of the deep classifier which generalizes logistic regression defined as follows

$$f_{\boldsymbol{\theta}}(\mathbf{x}) = \frac{1}{1 + \exp(-\boldsymbol{\theta}^T \mathbf{x})}, \tag{7}$$

where $f_{\boldsymbol{\theta}}(\mathbf{x})$ is a sigmoid function, which is trained with $\left\{\mathbf{c}^{(k)}, \mathbf{y}^{(k)}\right\}$ for $k = 1 \ldots S$; $\mathbf{c}^{(k)}$ is the output of the hidden layer of the autoencoder and $\mathbf{y}^{(k)}$ denotes the labelled data sets. The cost function of the softmax classifier tries to decrease the discrepancy between $\mathbf{y}^{(k)}$ and model output by tuning the $\boldsymbol{\theta}$ which represents the model parameters.

C. Training of the Deep Neural Network Classifier

The input $\left\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)} \ldots \mathbf{x}^{(k)}\right\}$ of the DNN classifier is features of the MEG signals. The features are extracted by using the Riemannian geometry [14], which manipulates the covariance matrices of the MEG signals. The output of the DNN classifier $\left\{\mathbf{y}^{(1)}, \mathbf{y}^{(2)} \ldots \mathbf{y}^{(k)}\right\}$ is the label representing a face or a scrambled face which is represented with 1 or 0 respectively. The training procedure of the DNN classifier is as follows:

1. First, the autoencoder is trained to minimize the error between the input vectors $\left\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)} \ldots \mathbf{x}^{(k)}\right\}$, which are features extracted from the MEG signals, and the output $\left\{\hat{\mathbf{x}}^{(1)}, \hat{\mathbf{x}}^{(2)} \ldots \hat{\mathbf{x}}^{(k)}\right\}$ of the autoencoder. This training process is completely unsupervised.

2. Secondly, the softmax classifier is trained with the data set $\left\{\mathbf{c}^{(k)}, \mathbf{y}^{(k)}\right\}$ where $\left\{\mathbf{c}^{(1)}, \mathbf{c}^{(2)} \ldots \mathbf{c}^{(k)}\right\}$ are the output of the hidden layer of the autoencoder and $\left\{\mathbf{y}^{(1)}, \mathbf{y}^{(2)} \ldots \mathbf{y}^{(k)}\right\}$ are labels representing the face and scrambled face.

3. Thirdly, the encoder part of the trained autoencoder and the trained softmax classifier are combined to construct the DNN classifier. The decoder part of the autoencoder is not used in this process.

4. Finally, the constructed DNN is trained with data sets $\left\{\left\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)} \ldots \mathbf{x}^{(k)}\right\}, \left\{\mathbf{y}^{(1)}, \mathbf{y}^{(2)} \ldots \mathbf{y}^{(k)}\right\}\right\}$ to allow the fine-tuning of the parameters of the autoencoder and softmax classifier.

III. EXPERIMENTAL RESULTS

A. Description of the Data Set

The performance of the presented classifier is investigated on a MEG data set where subjects were presented with visual stimuli, which composed of the faces and the scrambled faces. The human brain decoding with the MEG signals is achieved by using the trained DNN classifier, which can predict whether the subjects see the face or the scrambled face from the MEG signals. The data set is originally created by [25] and modified in the "Dec2Meg2014" competition [5], [26]. The data set contains 16 subjects with 580 trials, therefore the entire data set comprises 9414 trials.

The MEG signals have 306 channels, each with a time duration of 1.5 seconds and sampled at 250 Hz. Before the feature extraction process, the first 0.5 second of the signals, where the stimulus is presented to the subject, is discarded. After that, a band pass filter having a pass band of 1 Hz–20 Hz is applied to the obtained signals. Following this, a spatial filter is applied to signals for the purposes of dimensionality reduction and signal-to-noise ratio enhancement. Finally, a total of 2176 features used for classification are obtained by the Riemannian approach [27] for the signals.

TABLE I. COMPARATION TABLE.

Method	Average	STD	AUC
DNN	**80.85**	**0.13**	**0.81**
SVM	78.01	0.24	0.78
KNN	72.84	0.69	0.72
NB	71.92	0.17	0.72
DT	68.36	0.70	0.69

B. Simulation Results

The objective of the simulation in this application is to develop the DNN classifier that can identify the MEG signal as belonging to either one of the two classes: a face or a scrambled face. The DNN classifier has two parts. The first part is the autoencoder part, which is trained in an unsupervised fashion to obtain hidden features and to decrease the number of features used for classification. The second part is the softmax layer, which is trained in a supervised fashion with features obtained from the autoencoder part and labels to classify MEG signals. For developing an efficient DNN classifier, some tuning parameters such as sparsity, the regularization term and the weight of the sparsity penalty term must be determined. However, there is no analytical strategy to choose the best values for these parameters, which will yield the best experimental results for a given simulation. Therefore, the values of these parameters are heuristically chosen and experimentally validated for the simulation. In the simulations performed in this study, the values proposed in the references [23], [24] are used. These are chosen for the sparsity $\rho = 0.01$, the weight of the sparsity penalty $\beta = 1$

and the regularization term $\lambda = 0.003$. Moreover, setting the hidden layer size of the autoencoder to a small number such as 5 reduces the computational cost of the optimization problem.

9414 examples or trials with 2176 features are randomly sorted for training the DNN and the-state-art-methods such as SVM, KNN, NB, DT are applied to the data. The classifiers are applied with 5-fold cross validation technique, which divides the trials into five parts for each method. Four of the five part trials are provided as training, and a fifth trial is withheld for testing. This is done such that each trial part is analysed as the test trial one time to obtain an average classification across all five repetitions collected. The experiment is repeated for thirty times for each classification technique.

The results of the experiments are shown in Fig. 2 and the average values of the thirty experiments are shown in Table I. As it is seen from this table, the performances of the KNN, NB and DT classification methods are nearly the same, with the KNN being slightly better than the NB and DT. Besides, the SVM exhibits relatively better results for MEG signal's features, which are nonlinear complex data sets with high dimension features. The proposed classifier, however, demonstrates the best classification performance of all with considerable differences. The DNN classifier not only conveniently reduces the huge dimension of the feature space, i.e. from 2176 features to only 5, but it also produces significant classification performances.

The obtained results are supported with statistical analyses. The Mann-Whitney U test is applied to experiments repeated for thirty times because of non-homogeneity of distribution of the results and the results are demonstrated in TABLE II where mean difference, Z-value p-value and significance are shortly denoted as MD, Z, p and sig, respectively. It is seen that the DNN classifier is a very efficient classifier with $p \, value \cong 0$ where statistical significance is accepted at $p \, value < 0.05$.
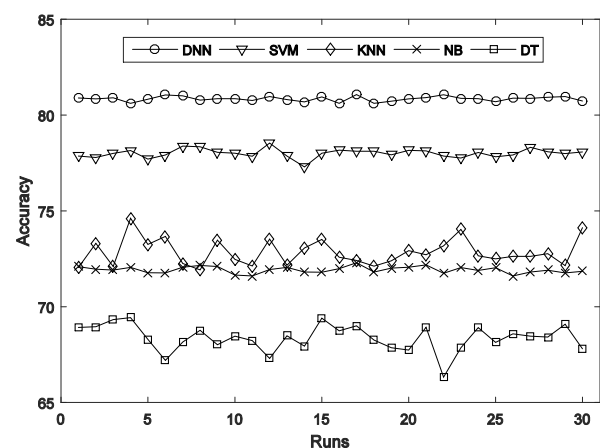

Fig. 2. Accuracy values of the runs.

Both classification accuracy and statistical analyses confirm that the DNN classifier performs better than the other methods. In order to obtain a more objective comparison, receiver operating characteristics (ROC) graphs are drawn to verify the obtained experimental results. The ROC is an important statistical procedure to measure and visualize the performance of a classifier and to compare

classification methods [28]. For each classifier, the ROC is demonstrated in Fig. 3 where the DNN has the best performance with the highest *area-under-curve* (AUC) value, which can be seen from Table I. AUC (i.e. maximum value = 1, minimum value = 0) is the area under the ROC. The higher AUC value means a better performance [28].

TABLE II. THE MANN-WHITNEY U TEST RESULTS.

	MD	Z	p	Sig.
DNN-SVM	2.840	-6.653	0.000	DNN
DNN-KNN	8.006	-6.653	0.000	DNN
DNN-DT	8.925	-6.653	0.000	DNN
DNN-NB	12.482	-6.653	0.000	DNN

IV. CONCLUSIONS

The objective of DNN classifier is to classify the MEG signals for decoding the visual stimuli from the human brain. In the literature, various signal processing algorithms are applied to classify such a complex signal (i.e., the MEG signals) with too much noise, but they usually exhibit poor performance. In this paper, we have presented a very efficient DNN based classifier for decoding the human brain activity from MEG signals.

Our simulations and experimental studies both demonstrate the superior performance of the DNN classifier over other traditional methods such as SVM, NB, KNN and DT. It should be pointed out that the DNN classifier needs relatively more time for training than the other methods used in this paper.

As a future work, other possible applications of the proposed DNN classifier should be further explored in both clinical and nonclinical domains. The DNN classifier may be used to study and compare the functional roles of different cortical regions of brain when performing specific brain tasks. It may, in turn, serve as a powerful tool for many different clinical applications.
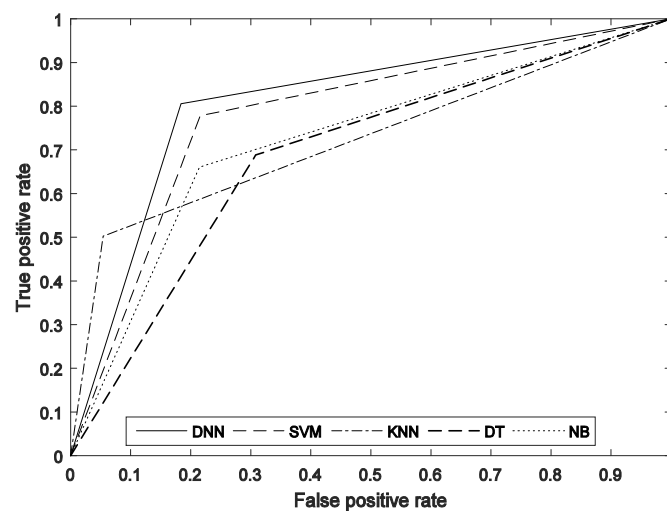

Fig. 3. The ROC curve with respect to the best accuracy of the DNN.

REFERENCES

[1] S. Thorpe, D. Fize, C. Marlot, "Speed of processing in the human visual system", *Nature*, vol. 381, no. 6282, pp. 520–522, 1996. [Online]. Available: http://dx.doi.org/10.1038/381520a0

[2] M. Hamalainen, R. Hari, R. J. Ilmoniemi, J. Knuutila, O. V. Lounasmaa, "Magnetoencephalography theory, instrumentation, and applications to noninvasive studies of the working human brain", *Reviews of modern Physics*, vol. 65, no. 2, pp. 413–497, 1993. [Online]. Available: https://doi.org/10.1103/RevModPhys.65.413

[3] H. Cecotti, "Single-trial detection with magnetoencephalography during a dual-rapid serial visual presentation task", *IEEE Trans. Biomedical Engineering*, vol. 63, no. 1, pp. 220–227, 2016. [Online]. Available: https://doi.org/10.1109/TBME.2015. 2478695

[4] M. K. Abadi, R. Subramanian, S. M. Kia, P. Avesani, I. Patras, N. Sebe, "DECAF: MEG-based multimodal database for decoding affective physiological responses", *IEEE Trans. Affective Computing*, vol. 6, no. 3, pp. 209–222, 2015. [Online]. Available: https://doi.org/10.1109/TAFFC.2015.2392932

[5] E. Olivetti, S. M. Kia, P. Avesani, "MEG decoding across subjects", in *Proc. Int. Workshop on Pattern Recognition in Neuroimaging*, Tubingen, 2014, pp. 1–4. [Online]. Available: https://doi.org/10.1109/prni.2014.6858538

[6] T. J. Bradberry, F. Rong, J. L. Contreras-Vidal, "Decoding center-out hand velocity from MEG signals during visuomotor adaptation", *Neuroimage*, vol. 47, no. 1, pp. 1691–1700, 2009. [Online]. Available: https://dx.doi.org/10.1016/j.neuroimage.2009.06.023

[7] Y. Kamitani, F. Tong, "Decoding the visual and subjective contents of the human brain", *Nature Neuroscience*, vol. 8, no. 5, pp. 679–685, 2005. [Online]. Available: https://doi.org/10.1038/nn1444

[8] K. N. Kay, T. Naselaris, R. J. Prenger, Jack L. Gallant, "Identifying natural images from human brain activity", *Nature*, vol. 452, no. 20, pp. 352–355, 2008. [Online]. Available: https://doi.org/10.1038/nature06713

[9] J. Wang, E. Pohlmeyer, B. Hanna, Y. G. Jiang, P. Sajda, S. F. Chang, "Brain state decoding for rapid image retrieval", in *Proc. 17th ACM Int. Conf. Multimedia*, 2009, pp. 945–954. [Online]. Available: https://doi.org/10.1145/1631272.1631463

[10] V. Raudonis, G. Narvydas, R. Simutis, "A classification of flash evoked potentials based on artificial neural network", *Elektronika ir Elektrotechnika*, vol. 81, no. 1, pp. 31–35, 2008.

[11] R Ursulean, A. M. Lazar, "Detrended cross-correlation analysis of biometric signals used in a new authentication method", *Elektronika ir Elektrotechnika*, vol. 89, no. 1, pp. 55–58, 2009.

[12] A. Barachant, S. Bonnet, M. Congedo, C. Jutten, "Multiclass brain–computer interface classification by Riemannian geometry", *IEEE Trans. Biomedical Engineering*, vol. 59, no. 4, pp. 920–928, 2012. [Online]. Available: https://doi.org/10.1109/TBME.2011. 2172210

[13] F. Yger, M. Berar, F. Lotte, "Riemannian approaches in brain-computer interfaces: a review", *IEEE Trans. Neural Systems and Rehabilitation Engineering*, vol. 99, pp. 1–1, 2016. [Online]. Available: https://doi.org/10.1109/TNSRE.2016.2627016

[14] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, A. Madabhushi, "Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images", *IEEE Trans. Medical Imaging*, vol. 35, no. 1, pp. 119–130, 2016. [Online]. Available: https://doi.org/10.1109/TMI.2015.2458702

[15] Y. LeCun, Y. Bengio, G. Hinton, "Deep learning", *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. [Online]. Available: https://doi.org/10.1038/nature14539

[16] J. Xu, X. Luo, G. Wang, H. Gilmore, A. Madabhushi, "A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images", *Neurocomputing*, vol. 191, pp. 214–223, 2016. [Online]. Available:

http://dx.doi.org/10.1016/j.neucom.2016.01.034

[17] A. Krizhevsky, I. Sutskever, G. E. Hinton, *Imagenet classification with deep convolutional neural networks*. Advances in Neural Information Processing Systems, CA: Curran Associates, Inc, 2012, pp. 1097–1105.

[18] I. Cetiner, A. A. Var, H. Cetiner*, "Classification of knot defect types using wavelets and KNN", *Elektronika ir Elektrotechnika*, vol. 22, no. 6, pp. 67–72, 2016. [Online]. Available: http://dx.doi.org/10.5755/j01.eie.22.6.17227

[19] A. Lorencs, I. Mednieks, M. Pukitis, J. S. Siņavskis, "Fusion of multisensor data based on different multidimensional distributions", *Elektronika ir Elektrotechnika*, vol. 22, no. 4, pp. 68–72, 2016. [Online]. Available: http://dx.doi.org/10.5755/j01.eie.22.4.12600

[20] G. E. Hinton, R. Salakhutdinov, "Reducing the dimensionality of data with neural networks", *Science*, vol. 313, no. 5786, pp. 504–507, 2006. [Online]. Available: http://dx.doi.org/10.1126/science.1127647

[21] Y. Bengio, *Practical recommendations for gradient-based training of deep architectures*. Neural Networks: Tricks of the Trade, CA: Springer, 2012, pp. 437–478. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-35289-8_26

[22] M. Ranzato, C. Poultney, S. Chopra, Y. LeCun, "Efficient learning of sparse representations with an energy-based model", in *Proc. Neural Information and Processing Systems*, Vancouver, 2006, pp. 1137–1144.

[23] J. Ngiam, A. Coates, A. Lahiri, B. Prochnow, Q. V. Le, A. Y. Ng, "On optimization methods for deep learning", *Proc. 28th Int. Conf. Machine Learning (ICML-11)*, Washington, 2011, pp. 265–272.

[24] A. Ng, "Sparse autoencoder", *CS294A Lecture Notes*, 2011.

[25] R. N. Henson, D. G. Wakeman, V. Litvak, K. J. Friston, "A parametric empirical Bayesian framework for the EEG/MEG inverse problem: generative models for multi-subject and multi-modal integration", *Frontiers in Human Neuroscience*, vol. 5, p. 76, 2011. [Online]. Available: http://dx.doi.org/10.3389/fnhum.2011.00076

[26] "DecMeg2014-Decoding the Human Brain". [Online]. Available: https://www.kaggle.com/c/decoding-the-human-brain

[27] A. Barachant, "Covariance toolbox". [Online]. Available: https://github.com/alexandrebarachant/covariancetoolbox

[28] T. Fawcett, "An introduction to ROC analysis", *Pattern Recognition Letters,* vol. 27, no. 8, pp. 861–874, 2006. [Online]. Available: http://dx.doi.org/10.1016/j.patrec.2005.10.010