Semi-supervised Learning for Dynamic Modeling of Brain Signals During Visual and Auditory Tests

Gonzalo Safont, Addisson Salazar, Jordi Belda, Luis Vergara
Institute of Telecommunications and Multimedia Applications
Universitat Politècnica de València
Valencia, Spain
gonsaar@upvnet.upv.es

Abstract— **Requirements of costly data labeling for data classification are relaxed with semi-supervised learning. This is particularly useful considering monitoring of a physiological process that continuously produces data and can be observed for a long time. We propose a new expectation-maximization (EM) procedure that implements semi-supervised learning and it is based on sequential independent component analysis modeling (SICAMM), that we have called EM-SICAMM. This procedure is applied for dynamic modeling of EEG signals measured from epileptic patients during visual and auditory neuropsychological tests. Those tests are done to evaluate the learning and memory cognitive function of the patients. Classification results demonstrate that EM-SICAMM outperforms, in terms of balanced error rate (BER) and kappa index, the following competitive methods: ICAMM, SICAMM, Gaussian mixture model (GMM), and hidden Markov model (HMM).**

Keywords— semi-supervised learning, dynamic modeling, ICA, EEG, neuropsychological tests

I. Introduction

The amount of knowledge about the labels for grouping data in significant populations mostly determines the success in pattern recognition for several applications. However, labeling could be costly due to several issues, e.g., human time and expertise requirements; economic conditions for deployment of specific experiments; or simply labeling is difficult or unapproachable due to constantly changing in the streaming source of the data. This problematic motivates the so-called semi-supervised learning approach, assuming that is relatively easy to obtain a large amount of unlabeled data [1, 2].

A field of application where capabilities of semi-supervised learning can be exploited is the monitoring of a physiological process that continuously produces data and can be observed for a long time. An application of particular interest is the dynamic modeling of brain waves using electroencephalographic (EEG) signals. This might be useful for several clinical diagnosis and research such as evaluation of cognitive functions [3]; sleep disorder diagnosis [4]; and brain connectivity analysis [5]. EEG signals are non-stationary and non-linear in nature, requiring advanced signal processing methods.

In this paper, we propose an expectation-maximization (EM) procedure that implements semi-supervised learning and it is based on sequential independent component analysis modeling (SICAMM) [6-8]. We have called EM-SICAMM to this proposed method. Independent component analysis (ICA)-based models have been shown to be suitable for biosignal analysis in EEG and electrocardiography, see [9, 10] and the references within. These methods will be explained in more detail in Section II.

We approach the dynamic modeling of EEG signals measured from epileptic patients during visual and auditory neuropsychological tests. Those tests are done as a part of clinical procedures to evaluate the learning and memory cognitive function of the patients. The performance of EM-SICAMM was compared using the balanced error rate (BER) and Cohen's kappa coefficient with that of the following competitive methods: ICAMM, SICAMM, Gaussian mixture models (GMM), and hidden Markov models (HMM). The results demonstrate the proposed EM algorithm was much more capable of learning the underlying data generation model of the EEG signals than the other methods.

II. Sequential Independent Component Analysis Mixture Model (SICAMM)

Sequential ICAMM is an extension of independent component analysis that considers time dependence in the data [6]. Essentially, SICAMM is a HMM whose hidden states correspond to classes of an ICAMM, and whose state emissions are modeled as a non-Gaussian mixture using ICA. Those models have been used in different applications [6-12].

Independent component analysis assumes that the set of observations $\mathbf{x}(1)...\mathbf{x}(N)$ can be expressed as linear combinations of a set of statistically independent sources, $\mathbf{s}(1)...\mathbf{s}(N)$, such that $\mathbf{x}(n)=\mathbf{A}\cdot\mathbf{s}(n),\ n=1...N$. ICA methods obtain the sources and the mixing matrix, $\mathbf{A}$, at the same time [13]. For convenience, we will assume that $\mathbf{A}$ is square and invertible, with demixing matrix $\mathbf{W}=\mathbf{A}^{-1}$. Thus, given the independence of the sources, we can express the joint probability of the data as the product of the marginal distributions:

$$p(\mathbf{x}(n))=|\det\mathbf{W}|\cdot p(\mathbf{s}(n))=|\det\mathbf{W}|\cdot\prod_{m=1}^{M}p(s_m(n)) \quad (1)$$

where M is the number of sources and $s_m(n)$ is the mth source at time n. In practice, $s_m(n)$ is unknown and it is estimated instead from the data as $\hat{s}_m(n)=\mathbf{w}_m{}^T\mathbf{x}(n)$, where $\mathbf{w}_m{}^T$ is the mth row of the demixing matrix.

SICAMM considers that there are K mutually-exclusive classes. Assuming that the data at time n belong to class k, $c(n)=k$, the data can be expressed as $\mathbf{x}(n)=\mathbf{A}_k\cdot\mathbf{s}_k(n)+\mathbf{b}_k$, where the k sub-index denotes the parameters of class $k=1...K$, and $\mathbf{b}_k$ is the bias term of class k [14]. Thus, the probability of the data conditional to class k is

$$p\left(\mathbf{x}(n)\mid c(n)=k\right)=\left|\det\mathbf{W}_k\right|\cdot\prod_{m=1}^{M}p\left(s_{k,m}(n)\mid c(n)=k\right) \quad (2)$$

where $\mathbf{W}_k=\mathbf{A}_k{}^{-1}$, and $s_{k,m}(n)$ is the mth source of class k at time n. The sources $s_{k,m}(n)$ are estimated from the data as $\hat{s}_{k,m}(n)=\mathbf{w}_{k,m}{}^T\left(\mathbf{x}(n)-\mathbf{b}_k\right)$. Given the HMM and the emission probabilities in (2), the likelihood of the data can be expressed using classical methods for HMM such as the forward-backward method. Classification could be performed, for instance, using Viterbi decoding, which can obtain the optimal (in the sense of maximum likelihood) solution to the sequence of hidden states [15].

III. Semi-supervised SICAMM

The parameters of SICAMM comprise the ICA parameters for each class ($\mathbf{W}_k$, $\mathbf{b}_k$, $p\left(s_{k,m}(n)\right)$, $k=1...K$), the prior probabilities of each class ($P(k)$, $k=1...K$), and the transition probabilities between each pair of classes ($\pi_{lk}=p\left(c(n)=k\mid c(n-1)=l\right)$, $1\le l\le K$, $1\le k\le K$). In practice, these parameters are unknown and they have to be estimated from labeled training data. In many applications, however, data is labeled by hand and is therefore more expensive and harder to obtain than unlabeled data. One of the most popular statistical tools to deal with this kind of incomplete problems is expectation-maximization (EM, [16]). Compared to gradient methods, it has been shown that the likelihood of EM is guaranteed to increase after each iteration, and EM avoids the need to calculate the derivatives (and the Hessian) of the cost function.

This work proposes an expectation-maximization semi-supervised learning algorithm for SICAMM that is able to learn the parameters of the model using labeled and unlabeled data, which we denote by EM-SICAMM. In particular, EM-SICAMM is based on the forward-backward method for learning HMM [17]. Furthermore, the method is able to classify unlabeled data simultaneously with the learning.

Let us assume we have a set of N observations $\mathbf{x}(1)...\mathbf{x}(N)$, and each observation belongs to one of K mutually-exclusive classes. The label of the nth observation will be denoted by $c(n)$. Given the semi-supervised learning problem, at least some of these labels are unknown during parameter estimation. Initially, EM-SICAMM proceeds as if no labels were known, and known labels are introduced at a later stage of the process.

Each iteration of EM-SICAMM performs an expectation stage and a maximization stage. The expectation stage builds the cost function for the ith iteration,

$$Q\left(\Theta,\Theta^{(i-1)}\right)=\sum_{c(1)...c(N)}\log p\left(\mathbf{x}(1)...\mathbf{x}(N),c(1)...c(N)\mid\Theta\right)\cdot \cdot p\left(\mathbf{x}(1)...\mathbf{x}(N),c(1)...c(N)\mid\Theta^{(i-1)}\right) \quad (3)$$

which is used during the maximization stage in order to update the values of the parameters, $\Theta^{(i)}$.

As in the forward-backward method, the cost function is not built explicitly, but rather through two intermediate variables ([17]): the forward variable, $\alpha_k(n)=p\left(\mathbf{x}(1)...\mathbf{x}(n),c(n)=k\right)$; and the backward variable, $\beta_k(n)=p\left(\mathbf{x}(n+1)...\mathbf{x}(N)\mid c(n)=k\right)$. These variables can be calculated iteratively

$$\alpha_k(n)=p\left(\mathbf{x}(n)\mid c(n)=k\right)\cdot\sum_{l=1}^{K}\pi_{lk}\cdot\alpha_l(n-1) \quad (4.a)$$

$$\beta_l(n-1)=\sum_{k=1}^{K}\pi_{lk}\cdot p\left(\mathbf{x}(n)\mid c(n)=k\right)\cdot\beta_k(n) \quad (4.b)$$

where π_{lk} are the transition probabilities from class l to class k; $p\left(\mathbf{x}(n)\mid c(n)=k\right)$ is calculated using (2), with $\hat{s}_{k,m}(n)=\mathbf{w}_{k,m}{}^T\left(\mathbf{x}(n)-\mathbf{b}_k\right)$; and the probability density of the sources $p\left(s_{k,m}(n)\mid c(n)=k\right)$ is calculated using a non-parametric kernel density estimator with a Normal kernel. The initial values of (4) are $\alpha_k(1)=p\left(\mathbf{x}(1)\mid c(n)=k\right)\cdot P(k)$ and $\beta_l(N)=1\forall l$. Semi-supervision is introduced at this stage by setting $p\left(\mathbf{x}(n)\mid c(n)=k\right)=0\ \forall k\neq k_n$, where k_n is the known label for the nth observation. The forward and backward variables are used to obtain two auxiliary variables:

$$\gamma_k(n)=p\left(c(n)=k\mid\mathbf{x}(1)...\mathbf{x}(N)\right)=\frac{\alpha_k(n)\cdot\beta_k(n)}{\sum_{l=1}^{K}\alpha_l(n)\cdot\beta_l(n)} \quad (5.a)$$

$$\varepsilon_{l,k}(n)=p\left(\mathbf{x}(1),...,\mathbf{x}(N),c(n+1)=k,c(n)=l\right)= =\frac{\pi_{lk}\cdot\alpha_l(n)\cdot p\left(\mathbf{x}(n+1)\mid c(n+1)=k\right)\cdot\beta_k(n+1)}{\sum_{k=1}^{K}\sum_{l=1}^{K}\pi_{lk}\cdot\alpha_l(n)\cdot p\left(\mathbf{x}(n+1)\mid c(n+1)=k\right)\cdot\beta_k(n+1)} \quad (5.b)$$

The maximization stage updates the parameters of the model. Given the model and the EM method, most of the parameters can be estimated in a closed form using the values in (5). The class priors are estimated as $P(k)=\gamma_k(1)$, whereas the bias terms and transition probabilities are estimated as:

$$\pi_{lk}=\frac{\sum_{n=1}^{N-1}\varepsilon_{l,k}(n)}{\sum_{n=1}^{N-1}\gamma_l(n)} \quad (6.a)$$

$$\mathbf{b}_k=\frac{\sum_{n=1}^{N}\gamma_k(n)\cdot\mathbf{x}(n)}{\sum_{n=1}^{N}\gamma_k(n)} \quad (6.b)$$

Due to the kernel density estimation of the sources, the bias terms in (6.b) do not affect the estimation of $p(s_{k,m}(n)\,|\,c(n)=k)$. These terms affect the estimation of the demixing matrices and the average value of the extracted sources. The ICA parameters of each class are estimated using an embedded ICA method, in a fashion similar to that of MIXCA [6]. Briefly, the demixing matrix for each class *k* is calculated using a weighted version ICA method. A demixing matrix is obtained using all the data, with each observation being weighted by the corresponding $\gamma_k(n)$. In this work, we considered FastICA as an embedded ICA method [13], which is adapted to consider weighted estimators. Once the SICAMM parameters have been updated, the EM procedure is repeated from the forward-backward variables until the values converge.

The EM-SICAMM algorithm is summarized in Table I. The initial values were selected using the labeled samples. The initial ICA demixing matrices were estimated using MIXCA and FastICA on the labeled samples [6]. The initial priors and transition probabilities were estimated by counting, and the initial bias terms were estimated using the sample.

IV. Results

The performance of the semi-supervised SICAMM method was tested on a set of real EEG signals from six epileptic subjects performing two neuropsychological tests to evaluate the learning and memory cognitive function. The first test used visual stimuli and the second test used auditory stimuli. The visual neuropsychological test was the figural memory subtest of the Barcelona test (BT) [18]. During each trial of the BT, the subject is shown a probe figure for ten seconds. The figure is then removed and, after a short retention interval, the subject must recognize the probe among a set of four similar figures. The figures become more difficult to recognize with each trial. The auditory test was the short-term memory subtest of the TAVEC [19, 20]. During each trial of the TAVEC, the subject is read a "shopping list" of 16 items. After a retention interval, the subject must recall as many words from the list as they can. The test comprises five trials: the first three trials use the same list, the fourth trial contains a second list, and the fifth trials uses the initial list again.

TABLE I. EM ALGORITHM FOR LEARNING SICAMM.

Input

Input data $\mathbf{x}(1)\ldots\mathbf{x}(N)$; known labels $c(n)$, $n\in\{1\ldots N\}$, $c(n)\in\{1\ldots K\}$; maximum number of iterations I

Initialization

Estimate priors $P(k)$ and transition probabilities π_{lk} by counting

Estimate bias terms $\mathbf{b}_k$ by averaging labeled observations of class k

Estimate demixing matrices $\mathbf{W}_k$ using labeled observations and an embedded ICA method

For each iteration $i=1\ldots I$

Expectation step

Calculate auxiliary variables, $\gamma_k(n)$ and $\varepsilon_{l,k}(n)$, using (4.a) to (5.b)

Maximization step

Update the values of the priors as $P(k)=\gamma_k(1)$

Update the values of the transition probabilities using (6.a)

Update the values of the bias terms using (6.b)

Re-estimate the demixing matrices using weighted embedded ICA

Repeat until the maximum number of iterations is reached or the parameters no longer change

Eighteen bipolar electroencephalographic channels were acquired concurrently with the tests. Synchronization was maintained using a graphic user interface designed by the authors, and the data were captured by the Neurophysiology and Neurology Units at Hospital La Fe, Valencia. A picture of one of the tests is shown in Fig. 1. Data were captured at 500 Hz and then filtered between 1.6 and 35. The stages of each test were sampled at the same rate. The stages correspond to two classes ($K=2$): "display" and "response." For analysis, the data were split into non overlapping epochs of one fourth of a second (roughly 125 samples). Seven features were extracted for each channel at each epoch: average amplitude, maximum amplitude, average power, centroid frequency, peak frequency, and the relative amount of power in the delta, alpha and sigma frequency bands. These features have been considered in previous works on EEG processing and automatic staging (e.g., [7, 21, 22]). The optimal feature for each subject was obtained using cross-validation and SICAMM.

The performance of the proposed semi-supervised EM-SICAMM was tested using a series of Monte Carlo experiments. For these experiments, we removed a certain amount of labels and the proposed method was used to the now-missing labels. The percentage of remaining labels is defined as the "supervision rate (SR)." The performance of EM-SICAMM was compared with that of ICAMM, SICAMM, GMM, and HMM. Semi-supervised methods were trained using the known labels and all the observations (including unlabeled ones). For these methods, classification was obtained as part of the training algorithm; for EM-SICAMM, by maximization of $\gamma_k(n)$. Supervised methods were trained using only labeled observations. For these methods, classification was obtained using the maximum a posteriori criterion (for ICAMM) or Viterbi decoding (for SICAMM). Classification performance was measured using the balanced error rate (BER) and Cohen's kappa coefficient. SR ranged from 5 to 95% in steps of 5%, and the end results were obtained as the average of 100 iterations per value of the SR.

ISBN 978-90-827970-1-5 © EURASIP 2018

Fig. 1. Subject taking the Barcelona test while the EEG is being captured.

Fig. 2 shows the average performance for all subjects of the considered methods for the automatic staging of the visual test. The overall performance of the semi-supervised SICAMM was better than that of the supervised SICAMM, owing to the exploitation of unlabeled data. Conversely, both versions of SICAMM yielded the best overall performance. The performance of the HMM increased with the amount of known data, but it was always worse than that of SICAMM. Also, dynamic methods improved faster than non-dynamic methods (ICAMM, GMM), owing to the exploitation of known values and their temporal dependences during classification.

Fig. 3 shows the average performance of the considered methods for the automatic staging of the auditory test. Unlike what happened with the visual test (see Fig. 2), non-dynamic methods (GMM, ICAMM) did not improve in performance with rising supervision rate. Conversely, HMM and EM-SICAMM experienced a noticeable improvement in performance, and EM-SICAMM yielded the best performance. The differences with the results of the visual test owed to timing differences between tests. The auditory test takes 10 times longer than the visual test to complete, which results in a larger number of available samples for the auditory test. Therefore, for the same supervision rate, the methods were trained using 10 times more data for the auditory test. The number of available samples for training at the 5% supervision rate for the auditory test was roughly equivalent to that available for the visual test at 50% supervision rate; at which point, as seen in Fig. 2, non-dynamic methods had already stabilized. The increase in time samples, however, did not translate into a larger amount of transitions between classes, since the number of trials is similar for both tests. Therefore, the estimation of the transition probabilities (necessary for dynamic methods) required a similar amount of supervision rate. This was also the reason EM-SICAMM reached such a low performance for very low amounts of supervision rate, since EM-SICAMM considered all observations (labeled and unlabeled) regardless of the supervision rate.

Tables II to V show the average classification performance for each of the subjects in both visual and auditory tests. As shown in Figures 2 and 3, EM-SICAMM yielded the best average performance (lowest BER and highest kappa). In the small amount of cases where EM-SICAMM did not yield the best result, the best performing method was SICAMM. These results confirm that the proposed EM algorithm was much more capable of learning the underlying data generation model than supervised SICAMM and Gaussian-based methods.

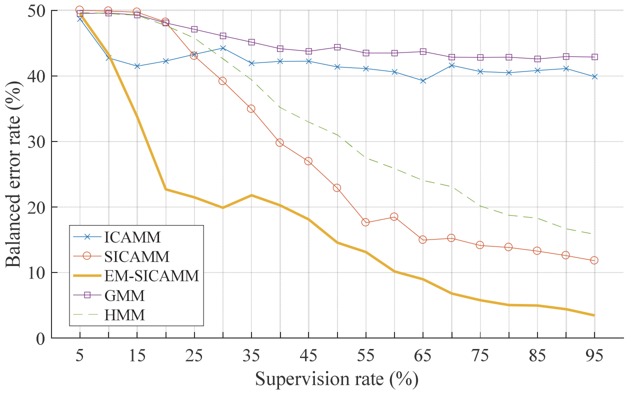

Fig. 2. Classification performance vs supervision rate for the visual test.

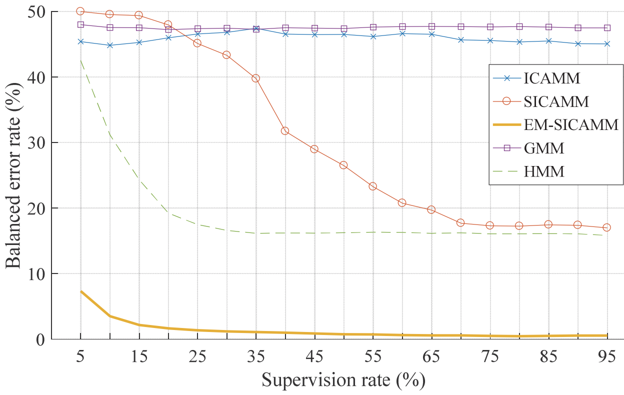

Fig. 3. Classification performance vs. supervision rate for the auditive test.

TABLE II. AVERAGE BER FOR BT.

Subject	1	2	3	4	5	6
ICAMM	42.5%	48.9%	49.0%	39.4%	34.4%	42.8%
SICAMM	20.2%	35.9%	43.5%	24.7%	28.2%	21.1%
EM-SICAMM	11.0%	31.5%	14.0%	16.5%	31.7%	28.3%
GMM	45.6%	48.9%	48.8%	39.8%	49.7%	43.6%
HMM	31.5%	34.4%	39.8%	23.2%	47.4%	28.7%

TABLE III. AVERAGE COHEN'S KAPPA BT.

Subject	1	2	3	4	5	6
ICAMM	0.16	0.02	0.02	0.22	0.27	0.14
SICAMM	0.59	0.31	0.14	0.51	0.35	0.58
EM-SICAMM	0.70	0.38	0.74	0.67	0.27	0.44
GMM	0.10	0.02	0.03	0.21	0.01	0.13
HMM	0.41	0.33	0.23	0.55	0.07	0.45

ISBN 978-90-827970-1-5 © EURASIP 2018

TABLE IV. AVERAGE BER FOR TAVEC.

Subject	1	2	3	4	5	6
ICAMM	45.9%	49.7%	47.8%	45.3%	48.9%	42.8%
SICAMM	29.7%	36.0%	28.8%	31.3%	32.0%	30.3%
EM-SICAMM	1.3%	2.2%	2.9%	2.6%	2.3%	1.8%
GMM	47.6%	50.0%	48.3%	47.7%	49.0%	44.7%
HMM	18.9%	25.2%	17.9%	22.4%	19.8%	15.5%

TABLE V. AVERAGE COHEN'S KAPPA FOR TAVEC.

Subject	1	2	3	4	5	6
ICAMM	0.08	0.01	0.05	0.08	0.02	0.13
SICAMM	0.41	0.29	0.43	0.35	0.34	0.38
EM-SICAMM	0.97	0.95	0.93	0.94	0.94	0.96
GMM	0.06	0.00	0.04	0.05	0.02	0.12
HMM	0.63	0.51	0.65	0.58	0.60	0.68

V. CONCLUSION

This paper has presented a semi-supervised expectation maximization (EM) method to learn sequential independent component analysis mixture models (SICAMM) that we have named EM-SICAMM. The classification performance of the method was tested on several sets of EEG data captured from epileptic patients performing two neuropsychological tasks. EM-SICAMM was compared with (supervised) ICAMM, SICAMM, GMM, and a HMM with Gaussian emissions. The proposed method yielded the best performance. The semi-supervised method was able to exploit both labeled and unlabeled data, being able to equal or exceed the classification performance of SICAMM with a lower amount of labeled data. EM-SICAMM yields a structured representation of the data as well as classification. The parameters of this representation include the ICA mixing matrix and sources, which have been shown to be related with physiological processes underlying the EEG. Therefore, future works will explore the parameters learned by EM-SICAMM to analyze the level of cortical activation of the subjects and define brain functional regions during the different stages of the neuropsychological tests.

Future works will also extend the proposed EM-SICAMM method to coupled hidden Markov models (rather than HMM). This extension could be used, for instance, to consider multimodal data, partition the brain into specific regions, or perform fusion [23, 24]. Furthermore, an online version of the method would classify the data during the test.

ACKNOWLEDGMENT

This work was supported by Spanish Administration (*Ministerio de Economía y Competitividad*) and European Union (FEDER) under grants TEC2014-58438-R and TEC2017-84743-P; and Generalitat Valenciana under grant PROMETEO II/2014/032.

REFERENCES

[1] X. Zhu and A. Goldberg, Introduction to Semi-Supervised Learning (Synthesis Lectures on Artificial Intelligence and Machine Learning), Morgan and Claypool Publishers, 2009.

[2] O. Chapelle, B. Schölkopf, and A. Zien, Semi-supervised Learning, MIT Press, 2006.

[3] A. Megías, M.J. Gutiérrez-Cobo, R. Gómez-Leal, R. Cabello, and P. Fernández-Berrocal, "Performance on emotional tasks engaging cognitive control depends on emotional intelligence abilities: An ERP study," *Scientific Reports*, vol. 7(1), Article no. 16446, 2017.

[4] S. Rossi et al., "Sleep patterns associated with the severity of impairment in a large cohort of patients with chronic disorders of consciousness," *Clin. Neurophysiology*, vol. 129(3), pp. 687-693, 2018.

[5] Y. Yau et al., "Network connectivity determines cortical thinning in early Parkinson's disease progression," *Nature Communications*, vol. 9(1), Article no. 2416, 2018.

[6] A. Salazar, On Statistical Pattern Recognition in Independent Component Analysis Mixture Modelling. Springer, 2013.

[7] G. Safont, A. Salazar, L. Vergara, and A. Rodriguez, "New applications of sequential ICA mixtures models compared with dynamic Bayesian networks for EEG signal processing," in *proceedings of the 5th International Conference on Computational Intelligence, Communication Systems and Networks (CICSyN)*, pp. 397-402, 2013.

[8] G. Safont, A. Salazar, L. Vergara, E. Gomez, and V. Villanueva, "Mixtures of independent component analyzers for microarousal detection," in proceedings of IEEE-EMBS Internat. Conf. on Biomed. and Health Informatics, BHI, Article no. 6864473, pp. 752-755, 2014.

[9] T.P. Jung and T.W. Lee, "Applications of Independent Component Analysis to Electroencephalography," in Statistical and Process Models for Cognitive Neuroscience and Aging. USA: Psychology Press, 2012.

[10] R. Llinares, J. Igual, A. Salazar, and A. Camacho, "Semi-blind source extraction of atrial activity by combining statistical and spectral features," *Digital Signal Processing*, vol. 21(2), pp. 391-403, 2011.

[11] G. Safont, A. Salazar, A. Rodriguez, and L. Vergara, "On recovering missing ground penetrating radar traces by statistical interpolation methods," Remote Sensing, vol. 6(8), pp. 7546-7565, 2014.

[12] A. Salazar, J. Igual, L. Vergara, and A. Serrano, "Learning hierarchies from ICA mixtures", Proceedings of IEEE Internat. Conf. on Neural Networks, IJCNN 2007, Article no. 4371312, pp. 2271-2276, 2007.

[13] P. Common and C. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. AP, 2010.

[14] T.W. Lee, M.S. Lewicki, and T.J. Sejnowski, "ICA mixture models for unsupervised classification of non-gaussian classes and automatic context switching in blind signal separation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22(10), pp. 1078-1089, 2000.

[15] A.J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Transactions on Information Theory*, vol. 13(2), pp. 260-269, 1967.

[16] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 39, pp. 1-38, 1977.

[17] O. Ibe, *Markov Processes for Stochastic Modeling*. Elsevier, 2013.

[18] M. Quintana et al., "Spanish multicenter normative studies (neuronorma project): norms for the abbreviated barcelona test," *Archives of Clinical Neuropsychology*, vol. 26(2), pp. 144-157, 2011.

[19] M. Benedet and M. Alejandre, Test de Aprendizaje Verbal España-Complutense, Spain: TEA Ediciones, 1998.

[20] J.S. Kreutzer, J. DeLuca, and B. Caplan, *Encyclopedia of Clinical Neuropsychology.* Springer, 2011.

[21] S. Sanei, J.A. Chambers, *EEG Signal Processing*. Wiley&Sons, 2013.

[22] G. Safont, A. Salazar, L. Vergara, E. Gomez, and V. Villanueva, "Probabilistic Distance for Mixtures of Independent Component Analyzers," I*EEE Transactions on Neural Networks and Learning Systems*, vol 29(4), pp. 1161-1173, 2017.

[23] A. Soriano, L. Vergara, B. Ahmed, and A. Salazar, "Fusion of Scores in a Detection Context Based on Alpha Integration," Neural Computation, vol. 27, pp. 1983-2010, 2015.

[24] L. Vergara, A. Soriano, G. Safont, and A. Salazar, "On the fusion of non-independent detectors," Digital Signal Processing, vol. 50, pp. 24-33, 2016.