IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 7 November 2023, accepted 13 December 2023, date of publication 19 December 2023,
date of current version 27 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3344776

RESEARCH ARTICLE

Graph Regularization Methods in Soft Detector Fusion

ADDISSON SALAZAR[1], (Member, IEEE), GONZALO SAFONT[1], LUIS VERGARA[1], AND ENRIQUE VIDAL[2], (Member, IEEE)
[1]Institute of Telecommunications and Multimedia Applications, Universitat Politècnica de València, 46022 Valencia, Spain
[2]Pattern Recognition and Human Language Technology (PRHLT), Universitat Politècnica de València, 46022 Valencia, Spain

Corresponding author: Addisson Salazar (asalazar@dcom.upv.es)

This work was supported in part by the European Commission under Grant HORIZON-MSCA-2021-DN, in part by Generalitat Valenciana under Grant CIPROM/2022/20, and in part by the Universitat Politècnica de València.

ABSTRACT This paper presents a theoretical derivation of two new graph-based regularization methods for fusing the individual results of multiple detectors (two-class classifiers). The proposed approach considers linear combination of the individual detector statistics and its extension to a general nonlinear fusion method known as α-integration. A cost function that includes a mean-square error and a regularization term is minimized. The inclusion of the regularization term, which is based on graph signal processing, reduces the dispersion of the fused statistics, and thus improves the separation between the fused statistics corresponding to every detection hypothesis. The proposed methods (linear and non-linear regularized α-integration) are experimentally compared with commonly used classification methods (random forest, linear and quadratic discriminant analysis, and naive Bayes) and competitive fusion methods (Dempster-Shafer, copulas, behavior knowledge space, independent component analysis mixture modeling, majority voting, the mean, and α-integration). Two challenging problems were approached using simulated and electroencephalographic data, respectively: (i) detection of ultrasound pulses buried in high noise, and (ii) detection of changes in electroencephalographic signals for neuropsychological test staging. An experimental convergence analysis of the proposed regularized method for these two applications is included. Besides, the proposed methods were tested using several benchmark datasets. Results on the basis of classification accuracy, kappa index, F1 score, and receiver operating characteristic curve analysis demonstrate the superiority of the proposed regularized fusion methods.

INDEX TERMS Graph regularization, detector fusion, alpha integration, graph signal processing, electroencephalographic signal processing, ultrasounds, two-class classification.

I. INTRODUCTION

The fusion of information from several sources is a booming research area that has been increasingly studied as a suitable method to solve many complex problems. The great development of sensor capabilities and facilities to acquire data from several modalities as well as the availability of hardware of low-cost high-computational performance have paving the way to the advance of fusion methods. This broad research area has been named in different ways, for instance, sensor data fusion, decision fusion, multimodal fusion, heterogeneous sensor fusion, mixture of experts, classifier combination, and multiway signal processing. The applications of fusion methods cover a large number of interesting problems including emotion recognition, fall detection, daily activity recognition, and hand movement recognition [1]; social networks [2]; Alzheimer's disease [3]; automatic sleep staging [4]; image object classification [5]; multi-media [6]; brain computer interfacing [7]; archaeological ceramic provenance [8]; object detection fusing infrared and visible images [9]; banking customer classification [10]; stock movement prediction [11]; maritime tracking [12]; and several benchmark datasets [13], [14].

Detector fusion refers to the combination of different detectors, all of them deciding about the same two hypotheses H_1 and H_0 with the aim of improving the individual

The associate editor coordinating the review of this manuscript and approving it for publication was Yiming Tang.

performance. Detector fusion can be broadly classified into hard and soft fusion. In hard fusion, the individual binary decisions are combined to obtain a final decision. In soft fusion, some continuous statistics generated by the detectors are combined to obtain one fused statistics, which are considered to obtain the final decision. Detector fusion has been implemented in several areas using different terminologies [15], [16], [17]. All of them share the need for optimum methods. In principle, the definition of optimality regards to minimizing the probability of error or maximizing the probability of detection. Hence, the joint probability masses of the individual decisions (hard fusion) or the joint probability density of the individual statistics conditioned to H_1 and H_0, must be estimated. This is in general a complex problem; an alternative is to predefine a reasonable fusion rule (hard) or function (soft) which normally incorporates some parameters which are to be optimized from training samples [18], [19].

The design of optimum fusion rules uses to be simpler than the design of an optimum fusion functions, although possible detector statistical dependence must be taken into account [20], [21]. However, whenever possible, soft fusion is preferable as hard fusion produces a loss of information due to the previous thresholding. When the statistics generated by the individual detectors are normalized between 0 and 1, soft detector fusion may be considered a fusion of scores, a terminology largely used, e.g., in biometrics [17]. However, considering the normalized statistics as posterior probabilities, the soft detector fusion also fall into the machine learning and classification areas [22], [23], for the two-class classification problem.

In this paper, we will consider soft detector fusion. First, the linear combination of the individual statistics will be considered and then extended to a general nonlinear fusion function known as α-integration [24], [25], [26], [27]. The optimization of those functions considers minimizing a properly defined cost, which includes a mean-square error term plus a regularization term. This later, will correspond to the so called ''smoothness'' in graph theory, a quadratic form of the Laplacian matrix [28], [29], [30], [31], [32]. Smoothness constraints have been used in different problems of semi-supervised learning [33], [34], it is of increasing interest in the emerging field of signal processing on graphs [35], [36].

The rest of the paper is organized as follows. In Section II, we propose a graph regularized linear combiner that includes a regularization term derived from a graph model. Theoretical derivation of the cost function is included. Section III extends the method of Section II to a non-linear combination of the statistics considering the function named α-integration. Section IV includes practical experiments with competitive comparison of the proposed methods to solve the following detection problems: ultrasound pulse detection; neuropsychological test staging from electroencephalographic (EEG) signals; and classification of six benchmark datasets (University of California, Irvine, UCI - Machine Learning Repository). Finally, Section V contains the conclusions and future work derived from this work.

II. GRAPH REGULARIZED LINEAR COMBINER

Let us assume that we have d different detectors (two-class classifiers) working on the same hypotheses H_1 and H_0, everyone contributes with a statistic s_i (score), which is a normalized value between 0 and 1. The individual statistics are linearly combined to obtain a fused statistic x.

$$x = \sum_{k=1}^{K} s_k \cdot w_k = \boldsymbol{s}^T \boldsymbol{w}. \tag{1}$$

where $\boldsymbol{s} = [s_1 \ldots s_K]^T$ and $= [w_1 \ldots w_K]^T$, K is the number of classifiers and T means transposition. This statistic will have conditioned probability densities $p\left(x/H_1\right)$ and $p\left(x/H_0\right)$, that can be used to implement a likelihood ratio test (LRT) to yield a final decision. Considering that the scores s_k are posterior probabilities generated by each of the individual detectors, the statistic x could be explained as a final posterior probability from the fusion.

In addition, we assume the availability of a set of labeled samples $\left\{\boldsymbol{s}^{(n)}, y^{(n)}\right\} n = 1 \ldots N$ where $\boldsymbol{s}^{(n)} = \left[s_1^{(n)} \ldots s_K^{(n)}\right]^{\boldsymbol{T}}$ is the vector formed by the statistics provided by the detectors, and y^n is the corresponding known binary decision ($y^{(n)} = 1$ if H_1 is true and $y^{(n)} = 0$ if H_0 is true). Given some coefficient vector, the fused statistics corresponding to the labeled samples will be $x^{(n)} = \boldsymbol{s}^{(n)T}\boldsymbol{w}n = 1 \ldots N$. Let us define the vectors $\boldsymbol{x} = \left[x^{(1)} \ldots x^{(N)}\right]^T$ and $\boldsymbol{y} = \left[y^{(1)} \ldots y^{(N)}\right]^T$. The optimum coefficients $\boldsymbol{w} = [w_1 \ldots w_K]^T$ will be obtained by minimizing a cost function

$$\boldsymbol{w}_{\text{opt}}^{\text{lin}} = \min_{\boldsymbol{w}} \left(\|\boldsymbol{y} - \boldsymbol{x}\|^2 + \beta \boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x} \right). \tag{2}$$

The first term of (2) is proportional to the mean-square error (MSE) between the fused statistic $x^{(n)}$ and the true label $y^{(n)}$. Notice that, ultimately, the performance of the final detector will depend on $p\left(x/H_1\right)$ and $p\left(x/H_0\right)$. Thus, by minimizing the MSE, $p\left(x/H_1\right)$ is shifted to 1, while $p\left(x/H_0\right)$ is shifted to 0. The second term is a regularization of the MSE derived from a graph model with Laplacian matrix $\boldsymbol{L}$ as detailed below. As we will see, minimizing this term will reduce the dispersion of the fused statistics corresponding to the same hypothesis, and thus improving the separation between $p\left(x/H_1\right)$ and $p\left(x/H_0\right)$. The real and positive constant β defines a trade-off between both terms.

Let us consider the undirected and weighted graph, $G\{V, \boldsymbol{A}\}$ where V represents the set of N vertices of the graph; and $\boldsymbol{A}$ is the adjacency matrix. The element a_{nm} of $\boldsymbol{A}$ is the weight corresponding to the edge connecting vertex k and vertex j. We assign $x^{(n)}$ to vertex $n, n = 1 \ldots N$, thus forming a graph signal [35], [36]. On the other hand, let us connect with a weight $= 1$ those vertices of G corresponding to the same true hypothesis, while keeping disconnected (weight $= 0$) those vertices corresponding to different true hypothesis, i.e.,

$$a_{nm} = a_{mn} = \begin{cases} 1 \; if \; y^{(n)} = y^{(m)} \\ 0 \; if \; y^{(n)} \neq y^{(m)}. \end{cases} \tag{3}$$

The Laplacian matrix is defined as $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$, where $\boldsymbol{D}$ is a diagonal matrix with diagonal elements $d_{nn} = \sum_{m=1}^{N} a_{nm}$. It is straightforward to demonstrate that

$$\boldsymbol{x}^T \boldsymbol{L}\boldsymbol{x} = \sum_{n=1}^{N} \sum_{m=1}^{N} a_{nm} \left(x^{(n)} - x^{(m)}\right)^2, \tag{4}$$

so the Laplacian quadratic form is normally considered a measure of the smoothness of the signal $\boldsymbol{x}$ on graph G. In fact, smoothness 0 is obtained if and only if $\boldsymbol{x}$ is a constant signal. Considering the definition in (3), the equation (4) can be expressed in the form:

$$\boldsymbol{x}^T \boldsymbol{L}\boldsymbol{x} = \sum_{n=1}^{N} \sum_{m|y^m=y^n} \left(x^{(n)} - x^{(m)}\right)^2. \tag{5}$$

Hence, by minimizing the second term in (2), the statistics corresponding to the same true hypothesis will reduce its dispersion, and thus reducing the overlapping between $p\left(x/H_1\right)$ and $p\left(x/H_0\right)$.

Let us compute the optimum coefficient vector $\boldsymbol{w}_{opt}^{lin}$. We define the matrix $\boldsymbol{S} = \left[\boldsymbol{s}^1 \ldots \boldsymbol{s}^N\right]^T$ so that we can express $\boldsymbol{x} = \boldsymbol{S}^T \boldsymbol{w}$. Hence, the objective cost function for minimization is

$$\begin{aligned} J &= \left\|\boldsymbol{y} - \boldsymbol{S}^T \boldsymbol{w}\right\|^2 + \beta \boldsymbol{w}^T \boldsymbol{S}\boldsymbol{L}\boldsymbol{S}^T \\ \boldsymbol{w} &= \boldsymbol{y}^T \boldsymbol{y} - 2\boldsymbol{w}^T \boldsymbol{S}\boldsymbol{y} + \boldsymbol{w}^T \boldsymbol{S}\left(\boldsymbol{I} + \beta \boldsymbol{L}\right)\boldsymbol{S}^T \boldsymbol{w}, \end{aligned} \tag{6}$$

and the corresponding derivative

$$\frac{\delta J}{\delta \boldsymbol{w}} = -2\boldsymbol{S}\boldsymbol{y} + 2\boldsymbol{S}\left(\boldsymbol{I} + \beta \boldsymbol{L}\right)\boldsymbol{S}^T \boldsymbol{w}. \tag{7}$$

We can solve it by equating (7) to 0

$$\boldsymbol{w}_{opt}^{lin} = \left(\boldsymbol{S}\left(\boldsymbol{I} + \beta \boldsymbol{L}\right)\boldsymbol{S}^T\right)^{-1} \boldsymbol{S}\boldsymbol{y}. \tag{8}$$

The parameter β defines the degree of importance given to the smoothness of the fused statistics with respect to the MSE. To establish an effective balance between both figures, we need some reference values to fit a reasonable interval for β. We consider the [K × K] matrix to be inverted in (8): $\boldsymbol{M} = \boldsymbol{S}\left(\boldsymbol{I} + \beta \boldsymbol{L}\right)\boldsymbol{S}^T = \boldsymbol{M}^I + \beta \boldsymbol{M}^L$, where $\boldsymbol{M}^I = \boldsymbol{S}\boldsymbol{S}^T$ and $\boldsymbol{M}^L = \boldsymbol{S}\boldsymbol{L}\boldsymbol{S}^T$. The Laplacian matrix is real and symmetric, so it can be diagonalized by a unitary transformation $\boldsymbol{L} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$, the columns of $\boldsymbol{U}$ are the (orthonormal) eigenvectors $\boldsymbol{u}^{(n)} n = 1 \ldots N$ and the main diagonal elements of the diagonal matrix $\boldsymbol{\Lambda}$ are the corresponding (real and nonnegative) eigenvalues $\lambda^{(n)}, n = 1 \ldots N$, which we assume sorted in increasing order. Let us call $\boldsymbol{s}_k = \left[s_k^{(1)} \ldots s_k^{(N)}\right]^T$ to the vector formed by the elements of the k-th row of matrix $\boldsymbol{S}$, i.e., a vector formed by all the statistics afforded by the k-th detector in the training set. We also define $\boldsymbol{r}_k = \boldsymbol{U}\boldsymbol{s}_k = \left[r_k^{(1)} \ldots r_k^{(N)}\right]^T$, the corresponding graph Fourier transform [35]. Thus, the generic elements of the two matrices contributing to $\boldsymbol{M}$ can be expressed in the form (notice that $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}$)

$$\begin{aligned} m_{kl}^I &= \boldsymbol{s}_k^T \boldsymbol{s}_l = \boldsymbol{s}_k^T \boldsymbol{U}^T \boldsymbol{U}\boldsymbol{s}_l = \boldsymbol{r}_k^T \boldsymbol{r}_l = \sum_{n=1}^{N} r_k^{(n)} r_l^{(n)} \\ m_{kl}^L &= \boldsymbol{s}_k^T \boldsymbol{U}^T \boldsymbol{\Lambda}\boldsymbol{U}\boldsymbol{s}_l = \boldsymbol{r}_k^T \boldsymbol{\Lambda}\boldsymbol{r}_l = \sum_{n=1}^{N} \lambda^{(n)} r_k^{(n)} r_l^{(n)}. \end{aligned} \tag{9}$$

Let us first concentrate on the main diagonal elements m_{kk}^L and m_{kk}^I. In that case, the contributions in the summations are all positive. Moreover, in any Laplacian matrix $\lambda^{(1)} = 0$ [30]. Actually, in the particular type of Laplacian matrix here defined, also $\lambda^{(2)} = 0$, this is because we have a disconnected graph: there is no path available to go from one vertex n labeled with $y^{(n)} = 1$ to a vertex m labeled with $y^{(m)} = 0$ (we have two disconnected subgraphs). Hence it is evident that

$$\text{If } 0 \le \beta \le \frac{1}{\lambda^{(N)}} \Rightarrow \beta m_{kk}^L < m_{kk}^I, \tag{10}$$

i.e., the contribution of the main diagonal elements introduced by the regularization term in matrix $\boldsymbol{M}$ are bounded by the contributions of the main diagonal elements of the MSE term. Regarding the out diagonal elements we may resort to the Schwartz inequality, and the result of (10), thus

$$\begin{aligned} &\text{If } 0 \le \beta \le \frac{1}{\lambda^{(N)}} \Rightarrow \\ &\qquad \Rightarrow \beta \left|m_{kl}^L\right| \le \sqrt{\beta m_{kk}^L \beta m_{ll}^L} < \sqrt{m_{kk}^I m_{ll}^I}. \end{aligned} \tag{11}$$

Notice that (11) is a general result which includes (10) as a particular case for $k = l$. The lower limit $\beta = 0$ implies that no regularization term is considered and $\boldsymbol{w}_{\text{opt}}^{\text{lin}}$ will be the classical least MSE solution. For any other value of β inside the proposed interval, the contribution of the regularization term to the main diagonal of $\boldsymbol{M}$ is properly bounded by the MSE term contribution. Certainly, we could use higher values of β, but $1/\lambda^{(N)}$ may be considered a reference about the importance given to every term of the cost function.

III. GRAPH REGULARIZED ALPHA INTEGRATION

Regularization can be extended to nonlinear combination of the statistics given by the individual detectors. We consider here a general nonlinear function named α-integration. It was originally proposed to integrate stochastic models [37], [38]. Then it was applied to the combination of positive numbers in regression problems [24], to the fusion of scores in a detection context [25], and to multi-class classification [26], [27]. The general expression of the α-integration function is given by

$$x = \begin{cases} \left(\sum_{k=1}^{K} w_k \cdot s_k^{\frac{1-\alpha}{2}}\right)^{\frac{2}{1-\alpha}}, & \alpha \neq 1 \\ \exp\left(\sum_{k=1}^{K} w_k \cdot \log\left(s_k\right)\right), & \alpha = 1 \end{cases}. \tag{12}$$

Notice that if $\alpha = -1$, we have the linear combiner. To allow any other possible value, i.e., $-\infty \le \alpha \le \infty$, it is required that $w_k \ge 0$ and $s_k \ge 0 \forall k$. It is also convenient to impose $\sum_{k=1}^{K} w_k = 1$ to avoid scaling of the fused statistic. With these constraints, other simple fusion functions are

included for specific values of α and $w_k = 1/k \forall k$.

$$\begin{aligned}
&\alpha = -1 \rightarrow x = \frac{1}{K}\sum_{k=1}^{K} s_k \\
&\alpha = 1 \rightarrow x = \prod_{k=1}^{K} s_k^{\frac{1}{K}} \\
&\alpha = 3 \rightarrow x = \left(\sum_{k=1}^{K} \frac{1}{s_k}\right)^{-1} \\
&\alpha = \infty \rightarrow x = \min\left(s_1 \ldots s_K\right) \\
&\alpha = -\infty \rightarrow x = \max\left(s_1 \ldots s_K\right)
\end{aligned} \tag{13}$$

In addition to this generalization property, (12) notice that the individual statistics are being observed through a nonlinear, static and invertible function, namely $f\left(s_k\right) = s_k^{2/1-\alpha}$ if $\alpha \neq 1, f\left(s_k\right) = e^{s_k}$ if $\alpha = 1$, which should be compensated before a linear combiner could be effective. As the nonlinear function is the same for all k, (otherwise, we will be introducing too many parameters) it seems convenient to normalize the statistics before integration [39].

Optimum α and $\boldsymbol{w}$ values may be estimated from a set of labeled samples $\left\{\boldsymbol{s}^{(n)}, y^{(n)}\right\} n = 1 \ldots N$ by minimizing the cost function (2).

$$J\left(\alpha, \boldsymbol{w}\right) = \left\|\boldsymbol{y} - \boldsymbol{x}\right\|^2 + \beta \boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x} \tag{14}$$

Unfortunately, the presence of the nonlinearity in the fusion function precludes obtaining a close form solution to this optimization problem, but we can resort to iterative gradients algorithms

$$\alpha\left(i+1\right) = \alpha\left(i\right) - \eta_\alpha \frac{\partial J}{\partial a}\left(i\right), \tag{15a}$$

$$\boldsymbol{w}\left(i+1\right) = \boldsymbol{w}\left(i\right) - \eta_w \frac{\partial J}{\partial \boldsymbol{w}}\left(i\right), \tag{15b}$$

Let us call $\varepsilon\left(\alpha, \boldsymbol{w}\right) = \left\|\boldsymbol{y} - \boldsymbol{x}\left(\alpha, \boldsymbol{w}\right)\right\|^2$ and $R\left(\alpha, \boldsymbol{w}\right) = \beta \boldsymbol{x}^T\left(\alpha, \boldsymbol{w}\right) \boldsymbol{L} \boldsymbol{x}\left(\alpha, \boldsymbol{w}\right)$. Clearly, $\frac{\partial J}{\partial \alpha} = \frac{\partial \varepsilon}{\partial \alpha} + \frac{\partial R}{\partial \alpha}$ and $\frac{\partial J}{\partial \boldsymbol{w}} = \frac{\partial \varepsilon}{\partial \boldsymbol{w}} + \frac{\partial R}{\partial \boldsymbol{w}}$. The derivatives $\frac{\partial \varepsilon}{\partial \alpha}$ and $\frac{\partial J}{\partial \boldsymbol{w}}$ have been calculated in [25],

$$\frac{\partial \varepsilon}{\partial \alpha} = -\sum_{n=1}^{N}\left(y^{(n)} - x^{(n)}\right)\frac{\partial x^{(n)}}{\partial \alpha} \tag{16a}$$

$$\frac{\partial x^{(n)}}{\partial \alpha} = \frac{2x^{(n)}}{1-\alpha}\left(\frac{\log\left(\sum_k w_k \cdot f_\alpha\left(s_k^{(n)}\right)\right)}{1-\alpha} + \frac{\sum_k w_k \cdot \frac{\partial f_\alpha\left(s_k^{(n)}\right)}{\partial \alpha}}{\sum_k w_k \cdot f_\alpha\left(s_k^{(n)}\right)}\right) \tag{16b}$$

and

$$\frac{\partial \varepsilon}{\partial w_k} = -2\sum_{n=1}^{N}\left(y^{(n)} - x^{(n)}\right)\frac{\partial x^{(n)}}{\partial w_k} \tag{17a}$$

$$\frac{\partial x^{(n)}}{\partial w_k} = \begin{cases} \frac{2}{1-\alpha}\left(\frac{s_\alpha \cdot f_\alpha\left(s_k^{(n)}\right)}{\sum_k w_k \cdot f_\alpha\left(s_k^{(n)}\right)}\right), & \alpha \neq 1 \\ x^{(n)} \cdot \log\left(s_k^{(n)}\right), & \alpha = 1. \end{cases} \tag{17b}$$

where we have defined

$$f_\alpha\left(s_k\right) = \begin{cases} s_k^{\frac{1-\alpha}{2}}, & \alpha \neq 1 \\ \log\left(s_k\right), & \alpha = 1 \end{cases}$$

$$\frac{\partial f_\alpha\left(s_k\right)}{\partial \alpha} = -\frac{1}{2}\log\left(s_k\right) \cdot s_k^{\frac{1-\alpha}{2}}. \tag{18}$$

It remains to compute $\frac{\partial R}{\partial \alpha}$ and $\frac{\partial R}{\partial \boldsymbol{w}}$. Considering (4) it follows that

$$\frac{\partial R}{\partial \alpha} = \beta\sum_{n=1}^{N}\sum_{m=1}^{N} a_{nm} \cdot 2\left(x^{(n)} - x^{(m)}\right)\left(\frac{\partial x^{(n)}}{\partial \alpha} - \frac{\partial x^{(m)}}{\partial \alpha}\right). \tag{19}$$

And we can make use of (16b) and (18) to compute $\frac{\partial x^{(n)}}{\partial \alpha}$ and $\frac{\partial x^{(m)}}{\partial \alpha}$. Equivalently

$$\frac{\partial R}{\partial w_k} = \beta\sum_{n=1}^{N}\sum_{m=1}^{N} a_{nm} \cdot 2\left(x^{(n)} - x^{(m)}\right)\left(\frac{\partial x^{(n)}}{\partial w_k} - \frac{\partial x^{(m)}}{\partial w_k}\right). \tag{20}$$

And we can use (17b) and (18) to compute $\left(\frac{\partial x^{(n)}}{\partial w_k} - \frac{\partial x^{(m)}}{\partial w_k}\right)$.

IV. EXPERIMENTS

A. SYNTHETIC DATA – ULTRASOUND PULSE DETECTION

The simulation experiment consisted of flaw detection in materials using ultrasounds. We have assumed that the reflectivity of the material tested by ultrasounds returns a measured signal that contains information on the microstructure and defects inside the material. Thus, the ultrasound signal is a summation of echoes from backscattering noise of the homogeneous part and echoes from reflections of sufficiently large internal inhomogeneities of the material. Reflections induce changes during signal traveling that should be measured in the features extracted from the ultrasound signals. This is a challenging problem considering the low signal to noise ratio (SNR) of the measured signal, i.e., the strong contribution of the homogeneous part of the material masks the relatively weak contribution of the defects. Fig. 1 shows an example of the simulated ultrasound signals where the locations of the defects are highlighted in the noisy signal.

In this experiment, the data consisted of ultrasound targets buried into background noise. The targets were modeled using Gaussian-modulated tones with random initial phase, i.e., $x\left(t\right) = A \cdot \sin(2\pi f_c t + \emptyset_0) \cdot \exp[-(2(t-\tau)/T)^\gamma]$, where: A is the peak amplitude; f_c, τ and T are respectively the central frequency, time center, and duration of the tone; and γ is an even number that determines the shape of the envelope of the pulse. We used $f_c = 20kHz$, $T = 1ms$, $\gamma = 4$, τ =20, 100, 150 ms, $\emptyset_0$ was randomly drawn from a uniform distribution in the range $[0, 2\pi)$, and A was calculated to obtain a specific peak signal-to-noise ratio (PSNR). The background noise was modeled by a K-distribution as in (21), which can describe

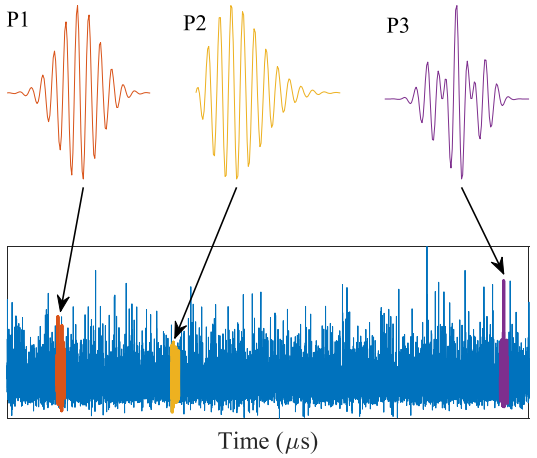

FIGURE 1. Example of the simulated ultrasound data with ultrasonic pulses (marked in color) buried in background noise for a 6 dB PSNR.

realistically the statistics of the envelope of the backscattered ultrasonic echo from a scattering medium [40]. The shape parameters of the K-distribution were set to $\mu = L = 1$ and $v = 10$, thus

$$p(X) = \frac{2}{x}\left(\frac{LvX}{\mu}\right)^{\frac{L+v}{2}} \frac{1}{\Gamma(L)\,\Gamma(v)} K_{v-L}\left(2\sqrt{\frac{LvX}{\mu}}\right) \quad (21)$$

Finally, it was assumed that the data were sampled at 50 kHz. For each iteration of the experiment, we generated a signal filled with background noise during 200 ms to obtain a total of 1000 samples. Then, three ultrasound targets were buried into the noise at 20, 100 and 150 ms from the start of the simulated signals. The targets were obtained by mixing four Gaussian-modulated tones with random initial phases. Thus, the areas of the simulation with background noise corresponded to a different model than the areas with both noise and ultrasound targets. Note that the targets in Fig. 1 (P1, P2, and P3, marked in red, yellow, and purple, respectively) are hardly distinguishable from the background noise, which shows the difficulty of the problem. The ultrasound experiments were repeated four times with different values of peak SNR (PSNR): 3, 6, 10, and 15 dB. These values were selected in order to simulate interesting detection cases, where targets are difficult to distinguish from background noise.

After data generation, the simulated ultrasonic signals were split into overlapped epochs that were labeled into two classes: 0 (noise background) and 1 (ultrasound pulse), i.e., a two-class classification problem. Hereinafter, we will use detector or classifier (two-class classifier) interchangeably. We estimated the epoch size and the overlap between consecutive epochs using Monte Carlo (MC) experiments. The value of the parameters were tuned in order to obtain the maximum accuracy in classification. The tuned parameters were epoch size = 256 samples (5 ms) and overlap = 50%.The features extracted from every signal epoch were the following: mean, mean absolute value, standard deviation, skewness, kurtosis, maximum frequency, centroid frequency, and average negative envelope.

A total of 13 classification methods were implemented: 4 individual classifiers; 7 competitive fusion methods; and the two proposed regularized fusion methods. The individual classifiers were the following: random forest (RDF); linear discriminant analysis (LDA); quadratic discriminant analysis (QDA), and naive Bayes (NB). All these methods allow to obtain a normalized score value, between 0 and 1, (s_k) of equation (1). The competitive fusion methods were the following: Dempster-Shafer (DS) [41], copulas [42], behavior knowledge space (BKS) [43], independent component analysis mixture modeling (ICAMM) [44], majority voting (MV) [45], the mean (average fusion [14]), and standard α-integration (α-LMSE) [27]. Note that these fusion methods are state-of-the-art and are currently being employed in various challenging applications. Finally, the proposed implemented methods were the following: regularized linear α-integration (Linear LMSE (REG)); and regularized α-integration (α-LMSE (REG)). Table 1 shows details of the implemented methods, which parameters were tuned in a rather experimental manner for the particular experiments in Section IV.

TABLE 1. Details of the implemented methods.

Implemented method	Parameters
LDA	'Prior' uniform.
NB	Ensure non-zero group variances.
QDA	'Prior' uniform.
RDF	'Ntrees' 30 'prior' uniform.
DS	Focal elements propagated through find minimum of constrained nonlinear multivariable function 'fmincon'. Normal Bayesian approximation to estimate cumulative density functions (cdfs).
Copulas	'Type' Gaussian 'Copula fitting' Maximum Likelihodd 'Confidence interval' 0.5
BKS	Final decision based on the weighted decisions of all classifiers according to their F1-scores.
ICAMM	'ICA number' 2 'Source number' 2 'Density estimation' Non-parametric.
Mean	Estimate the average of the single classifier posteriors.
MV	Selection of the most voted category using the decisions of all classifiers.
α -LMSE, Linear LMSE REG), α -LMSE (REG)	Initial value of α = -1. Initial values of $\boldsymbol{w}$ should be non-negative and add up to one, $w_k = 1/K$. Optimization method 'fmincon'.

In order to verify the performance of the considered classification methods, a series of MC experiments were run. For each experiment, the available data samples were separated equally into three sets: training, validation, and testing. The training pieces were used to train the single classifiers, the α integration-based and competitive fusion methods were trained using the scores from the trained single classifiers on the validation dataset. Finally, the performance was estimated on the testing dataset. Three performance indicators were

calculated: the classification accuracy (Acc); the kappa index (Kap); and the F1 score (F1), i.e., the harmonic average of precision and recall [46]. The results were obtained as the average of 1000 MC experiments. We considered several statistical tests on the results of the MC experiments for all the experiments in Section IV. First, the Kolmogorov-Smirnov test has been considered to verify that the results were Normally-distributed ($p \ll 0.05$). Then, given the Normal distribution and the number of available samples, we considered one-way analyses of variance (ANOVA) and nonparametric Kruskal-Wallis ANOVAs to determine whether the differences in outcome were statistically significant.

Table 2 shows the average results of the classification, which were estimated using a posterior probability threshold of 0.5. The error variance in the classification results measured by standard deviation (std.) was less than 2%.The proposed α-LMSE (REG) method obtained the best results for every PSNR value, yielding the smallest std. (0.4) and therefore their results being the most stable.

TABLE 2. Classification results for the simulated experiment, ultrasound pulse detection (values in %).

Method	3 dB PSNR			6 dB PSNR			10 dB PSNR			15 dB PSNR		
	Acc	Kap	F1	Acc	Kap	F1	Acc	Kap	F1	Acc	Kap	F1
LDA	90.6	54.1	91.1	92.9	60.1	93.6	97.6	84.7	97.9	97.4	84.7	97.5
NB	86.4	40.1	83.2	90.3	61.8	90.6	94.3	67	94.5	94.3	67	94.4
QDA	91.7	54.3	92.3	94	70	94.5	97.8	85.1	98.2	97.9	85	98.1
RDF	91.9	54.2	92.8	95.5	72	95.9	96.7	78.8	97.3	98.4	86.2	98.8
DS	93.4	66	94.3	95.4	76.3	96.2	96.4	85.5	97	96.8	88	97.1
Copulas	93	65.2	93.8	95	75.7	95.6	96.2	84.9	96.6	96.7	87.2	96.9
BKS	92.6	64.9	93.3	94.5	75.1	95	95.6	84.2	95.9	96.2	86.5	96.3
ICAMM	92.5	64.7	93	94.7	75	95.1	95.8	84.1	96.1	96.6	86.8	96.8
Mean	93.7	62	93.9	95.6	74	95.8	97.2	84.9	97.3	98.3	86	98.4
MV	92.1	61	92.5	94	71.7	94.3	95.1	81	95.3	95.8	84	95.9
α-LMSE	94	64.6	95.4	96	76	97.3	98	87.4	99	98.7	90	99.4
Linear LMSE (REG)	**94.6**	**68**	**96.2**	**96.6**	**78.2**	**98**	**98.2**	**87.5**	**99.4**	**98.7**	**90.2**	**99.4**
α-LMSE (REG)	**96**	**70**	**97.8**	**97.8**	**80.1**	**99.4**	**98.6**	**89**	**99.4**	**98.8**	**91**	**99.5**

The result tables for all the experiments in Section IV has been divided for readability in four zones: single classification methods (LDA, NB, QDA, and RDF); standard fusion methods (DS, copulas, BKS, ICAMM, MV, and Mean); standard α-LMSE; and the proposed regularized methods α-LMSE (REG), and Linear LMSE (REG). The best results for standard fusion methods and single classification methods are highlighted by shadowing and the best results overall are in bold.

The improvements are larger for kappa index than for Acc and F1. It is because kappa index take into account the possible bias in classification results due to the imbalance that exists between the number of samples of the two classes, i.e., a priori probability of the classes (ultrasound pulse and noise background). The results for every classifier improve when the PSNR increases, except for NB which accuracy decreases or does not improve for 10 and 15 dB. This is because the statistics of the data move away from the statistical independence among the extracted features assumed by NB.

In general, the classification results of the fusion methods were better than the ones of the individual classifiers and the results of the fusion methods based on α integration overcame the ones of the competitive fusion methods (DS, copulas, BKS, ICAMM, MV, Mean, and α-LMSE). Besides, the proposed regularized α integration methods perform better than the classic α integration. The improvement of the results by the proposed methods were greater for lowest values of PSNR, i.e., the most difficult classification cases. For instance, the differences between α-LMSE (REG) and the best of the competitive fusion methods that are not based on alpha integration were 2.3% accuracy (Mean); 4% kappa (DS); and 3.5% F1 (DS) for 3dB PSNR and were 2.2% accuracy (Mean); 3.8% kappa (DS); and 3.2% F1 (DS) for 6dB PSNR. Regarding to the differences between α-LMSE (REG) and the best of the single classification methods were 4.1% accuracy (RDF); 15.7% kappa (QDA); and 5% F1 (RDF) for 3dB PSNR and were 2.3% accuracy (RDF); 8.1% kappa (RDF); and 3.5% F1 (RDF) for 6dB PSNR. With the respect to the standard alpha integration (α-LMSE) method, the improvement obtained by α-LMSE (REG) were the following: 2%, 5.4%, and 2.4% (Acc, kap, F1) and 1.8%, 4.1%, and 2.1% (Acc, Kap, F1) for 3 dB PSNR and 6 dB PSNR, respectively. These differences were statistically significant.

Furthermore, the average receiver operating characteristic (ROC) curve was calculated for each value of PSNR and for every method implemented. This metric has been extendedly used in many applications, see for instance [47], [48]. The ROC curves for each of the PSNR values were similar, thus, we show in Fig. 2, as an example, the results for 3dB PSNR. Figures 2a and 2b show the ROC curves for the full range of false positive rate (FPR) values and for low and very low false alarm operating regimens (FPR $\leq$ 10%), respectively. Note that the axes, true positive rate (TPR) and FPR are in logarithmic scale.

The results of Fig. 2 are consistent with the ones of Table 2. The difference between α-LMSE (REG) and the other methods were more important in the region of low FPR (less than 10%). This region of FPR is particularly important in many real-world applications where operative working with relative high FPR is costly and unfeasible (e.g., credit card fraud detection [49]). The ROC curve of α-LMSE (REG) rose rapidly with FPR, yielding over 73% TPR with only 2% FPR, even in the case of PSNR = 3 dB. All the fusion methods obtained better results than the ones obtained by single classifiers. The best results were obtained for the proposed regularized alpha integration methods, in order, α-LMSE (REG) and Linear LMSE (REG).

Table 3 shows the values of area under the ROC curve (AUC) corresponding to Fig. 2. The difference of AUC between α-LMSE (REG) and the best (α-LMSE) and the second best (DS) competitive fusion methods were 1.09 and 1.29 in the full range of FPR and 0.88 and 0.92 for

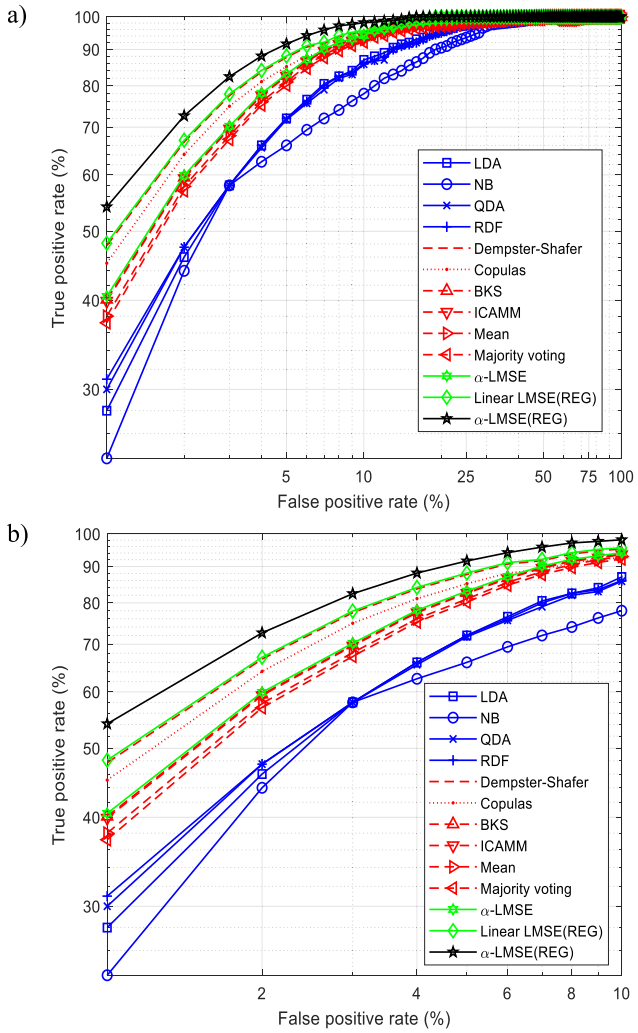

FIGURE 2. ROC curves for the simulated experiment at PSNR = 3dB. (a)FPR full range; (b) Low FPR ≤ 10%.

TABLE 3. AUC from ROC curves of Fig. 2.

Method	FPR ≤ 100%	FPR ≤ 10%
LDA	95.2630	6.5500
NB	93.2480	6.0520
QDA	95.2010	6.5020
RDF	95.3090	6.5900
DS	97.1461	7.5760
Copulas	97.0831	7.5560
BKS	97.0924	7.0565
ICAMM	96.4464	7.3965
Mean	96.9120	7.4290
MV	96.3490	7.3490
α -LMSE	97.3381	7.6160
Linear LMSE (REG)	**97.9185**	**8.0965**
α -LMSE (REG)	**98.4321**	**8.4960**

FPR ≤ 10%, respectively. The best result of the single classification methods was obtained by RDF with difference of 3.12 and 1.91 AUC with respect to α-LMSE (REG) in the full range of FPR and FPR ≤ 10%, respectively. These differences were statistically significant.

Fig. 3 shows a convergence analysis of the proposed regularized method (α-LMSE (REG) versus the non-regularized α-LMSE for an ultrasound pulse detection experiment. It was the normal behavior of the convergence for all the experiments. This analysis allows to investigate the robustness and stability of the methods across a range of parameter settings to ensure their consistent performance. The convergence was steady and achieved in a low number of iterations, at most 45 iterations. The considered regularization did not preclude convergence, although it did increase the number of required iterations on average, and its effect depended on the value of the regularization term, βmax defined as $1/\lambda^{(N)}$, see equations (10) and (11).

In addition, the regularized alpha integration tended to allocate a higher weight to some classifiers over the others. For this experiment, the higher weights were assigned to RDF and LDA. It is because those classifiers yield scores that are more concentrated for each of the classes in values close to 0 and 1, and thus, the histograms for each class will be less overlapping. This is the main factor considered by the regularized classifier fusion proposed here. In summary, the results of the fusion demonstrated the capabilities of the proposed method to find complementarities among the classifiers to improve classification accuracy.

B. EEG DATA – NEUROPSYCHOLOGICAL TEST STAGING FROM EEG SIGNALS

We implemented the proposed methods also for a challenging real-data application, the automatic staging of EEG signals from subjects performing a neuropsychological test. The data from six epileptic patients were analyzed. The neuropsychological tests are essential to evaluate cognitive functions such as learning and memory, which is a critical part of the patient neuropsychological assessment.

The processing of information becomes impossible if the brain is unable to store a sufficient amount in short-term (working) memory or retrieves past experiences, events, and strategies from long-term memory. Conversely, information stored in short- or long-term memory is useless without the means to properly access and activate it. The neuropsychological test was the visual paired associates subtest from the Wechsler Memory Scale (WMS-R, [50]). This test considers visual stimuli, and it consists of two stages, stimulus display (D) and subject response (R) that are repeated in several trials changing the stimuli.

The EEG was composed of M = 19 EEG channels, using the 10-20 electrode system, that were sampled at 500 Hz. Fig. 4 shows an example of the EEG data measured for one of the subjects; the profile of the classes changing from "Display" to "Response" and vice versa; and the 10-20 electrode system employed. The EEG signals were split into epochs, which were labeled into two classes: 0 (display) and 1 (response). As in the previous experiment (Section IV-A),

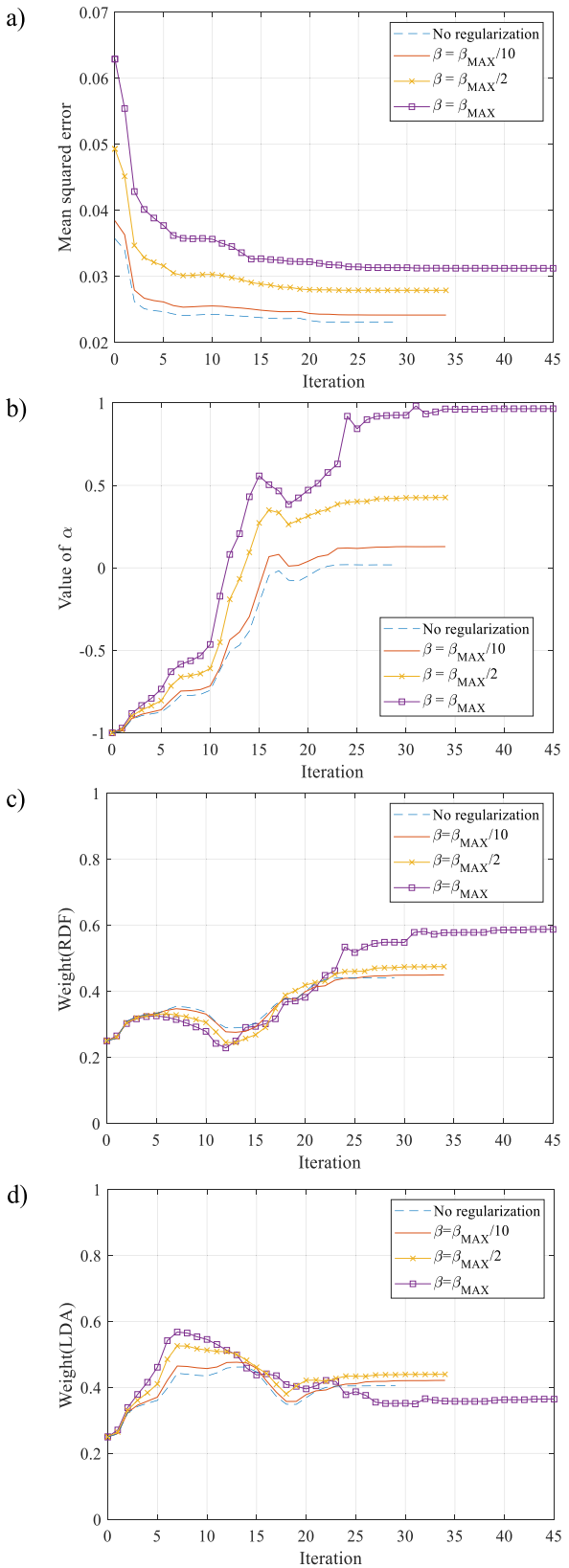

FIGURE 3. Convergence of alpha integration parameters with respect to the number of iterations for the simulation experiment: a) convergence error (LMSE); b) alpha value; c) weight of RDF; d) weight of LDA. The weights of the other two classifiers (QDA and NB) were much lower than those of RDF and LDA.

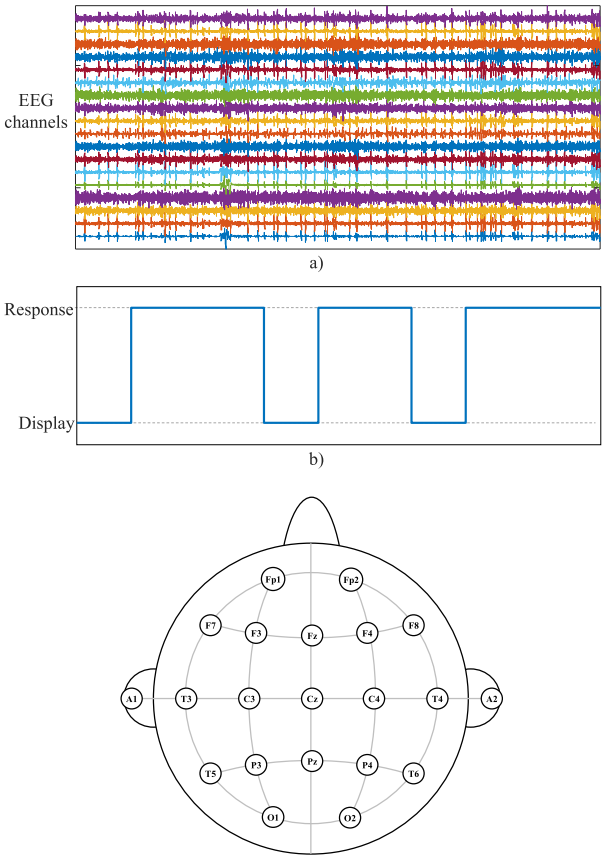

FIGURE 4. Example of the extracted data from one of the subjects: a) the EEG signal channels considered at a time frame; b) real classes for that time frame; c) diagram of the used 10-20 electrode system, with the considered electrodes.

we estimated the epoch size and overlap between consecutive epochs using MC experiments. The experimentally tuned parameters were epoch size = 125 samples (0.25 s) not overlapping. The features extracted from the EEG signals were the following: frequency band power (delta, theta, alpha, sigma, and beta); and Hjorth parameters (activity, mobility, and complexity) [51]. The set of classification and fusion methods; training-validation-testing methodology; and statistical significance evaluation of the results were the same implemented in simulation experiments (Section IV-A).

Table 4 presents the average classification results estimated for a posterior probability threshold of 0.5 in terms of accuracy, kappa index, and F1 score. The error variance in the classification results measured by standard deviation (std.) was less than 3.2%. The proposed α-LMSE (REG) method obtained the best results for the three performance indicators, yielding the smallest std. (0.5) and therefore their results being the most stable. The improvements are larger for kappa index, which considers unbalancing of class sample size, than for accuracy and F1. In general, the classification results of the fusion methods were better than the ones of the individual classifiers (RDF, LDA, QDA, and NB) and the regularized fusion methods (Linear LMSE (REG) and

TABLE 4. Classification results for the experiment on EEG data from the visual associates subtest of WMS-R (values in %).

Method	Acc	Kap	F1
LDA	83	66.1	83.2
NB	76.1	52	76.4
QDA	84.9	68.4	85.4
RDF	84	66.2	84.5
DS	84.9	68.5	85.1
Copulas	83.8	68	84
BKS	83	67.7	83.2
ICAMM	82.7	67.3	82.7
Mean	84.1	67.5	84.7
MV	82.4	65.7	82.5
α-LMSE	85.2	69.4	86.6
Linear LMSE (REG)	**86.3**	**71.1**	**87.9**
α- LMSE (REG)	**88.1**	**74.7**	**89.9**

α-LMSE (REG)) outperformed the standard fusion methods (DS, copulas, BKS, ICAMM, MV, Mean, and α-LMSE).

The proposed method, α-LMSE (REG), exhibited an improvement in the three indices (Acc, Kap, F1) of 2.9%, 5.3%, and 3.3%, respectively, compared to the best standard fusion method (α-LMSE). The differences with the second standard fusion method with the best results (DS) were 3.2%, 6.2%, and 4.8%, (Acc, Kap, F1), respectively. The best result of the single classifiers was obtained by QDA with differences of 3.2%, 6.3%, and 4.5%, (Acc, Kap, F1) with respect to α-LMSE (REG). These differences were statistically significant.

Thus, the results of the two regularized fusion methods were better than the ones of standard α integration. This shows the regularization methods were able to take advantage from the complementarities between all the single classifiers. Finally, NB obtained the worst results giving the statistical independence assumed in this method.

Fig. 5 shows the estimated ROC curves for the experiment with EEG data. Figures 5a and 5b show the ROC curves for the full range of false positive rate (FPR) values and for low and very low (FPR $\leq$ 10%) false alarm operating regimens, respectively. Note that the axes, (TPR and FPR) are in logarithmic scale. The results are consistent with the ones of Table 4 and the simulated data results of Fig. 2. The difference between α-LMSE (REG) and the other methods were more important in the region of low FPR (less than 10%). The ROC curve of α-LMSE (REG) rose rapidly with FPR, yielding above 77% TPR with only 2% FPR. The best results were obtained for the proposed regularized alpha integration methods, in order, α-LMSE (REG) and Linear LMSE (REG).

Table 5 shows the AUC values corresponding to the ROC curves in Fig. 5. The difference of AUC between α-LMSE (REG) and the best (α-LMSE) and the second best (DS) competitive fusion methods were 1.12 and 1.18 in the full range of FPR and 0.38 and 0.42 for FPR $\leq$ 10%, respectively. The best result of the single classification methods was obtained by RDF with difference of 1.66 and 0.81 AUC with respect to α-LMSE (REG) in the full range of FPR and FPR $\leq$ 10%, respectively. These differences were statistically significant.

TABLE 5. AUC from ROC curves of Fig.5.

Method	FPR $\leq$ 100%	FPR $\leq$ 10%
LDA	95.4740	7.1050
NB	91.3990	4.8300
QDA	95.2835	6.5785
RDF	96.2790	7.3975
DS	96.7624	7.7850
Copulas	96.6870	7.7650
BKS	96.6210	7.7775
ICAMM	95.9298	7.7175
Mean	94.0200	7.6800
MV	94.3780	7.6000
α -LMSE	96.8220	7.8250
Linear LMSE (REG)	**96.8504**	**8.0175**
α -LMSE (REG)	**97.9380**	**8.2050**

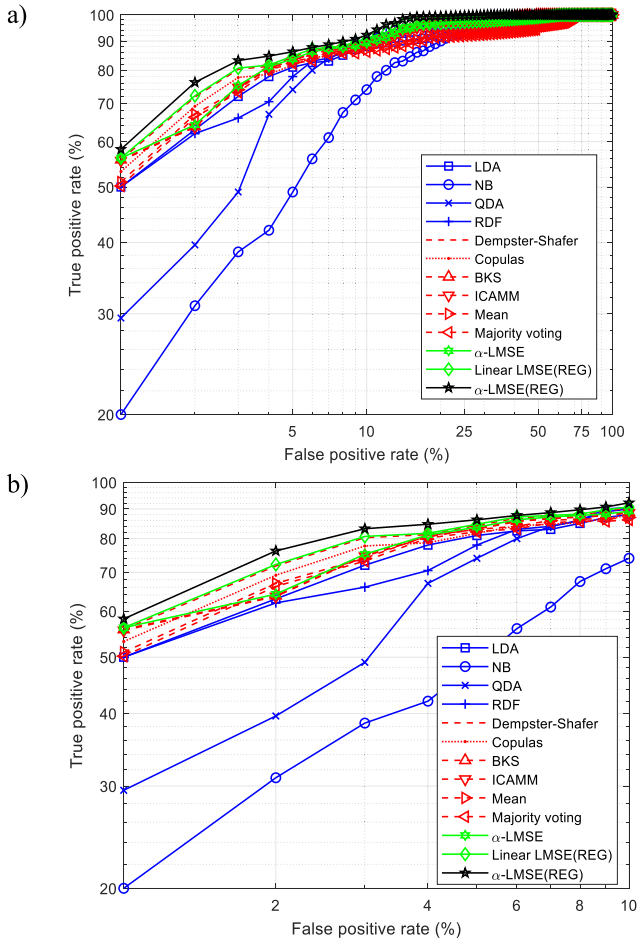

FIGURE 5. ROC curves for the experiment on EEG data. (a) FPR full range; (b) Low FPR $\leq$ 10%.

Fig. 6 shows a convergence analysis through iterations of the proposed regularized algorithm, α-LMSE (REG), versus the non-regularized α-LMSE for an EEG experiment. It was the normal behavior of the convergence for all the experiments. This analysis allows to investigate the robustness and stability of the methods across a range of parameter settings to ensure their consistent performance.

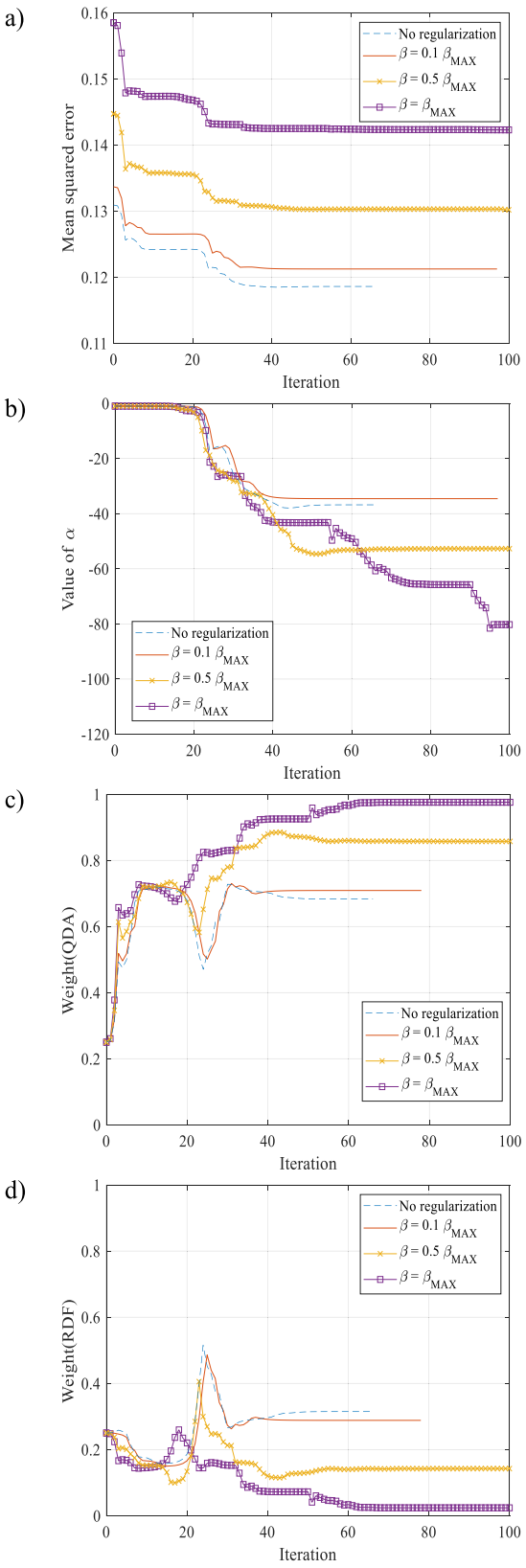

FIGURE 6. Convergence of alpha integration parameters with respect to the number of iterations for the experiment on real EEG data: a) convergence error (LMSE); b) alpha value; c) weight of QDA; d) weight of RDF. The weights of the other two classifiers (LDA and NB) were much lower than those of QDA and RDF.

The convergence was steady and achieved in at most 100 iterations. The considered regularization did not preclude convergence, although it did increase the number of required iterations on average, and its effect depended on the value of the regularization term, β.

For this experiment, the higher weights of the fusion were assigned to QDA and RDF. As for the previous experiment with simulations, fused results have also demonstrated for EEG data the capabilities of the proposed method to find complementarities among the results of the individual classifiers. Besides, the regularization term has allowed a higher concentration of the fused scores in values close to 0 and 1.

C. CLASSICATION OF UCI DATASETS

In this section, we include several results from the application of the implemented methods to six publicly available datasets of the UCI repository. Thus, the generalizability and validation of the effectiveness of the proposed methods across different applications is assessed.

The datasets were the following: (i) SPECTF: data on cardiac single proton emission computed tomography (SPECT) images. Each patient classified into two categories: normal and abnormal. (ii) Bupa: blood tests of patients sensitive to liver disorders that might arise from excessive alcohol consumption. (iii) Wbdc: Breast Cancer Wisconsin (Diagnostic). Features computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. (iv) Ionosphere: classification of radar returns from the ionosphere. ''Good'' radar returns are those showing evidence of some type of structure in the ionosphere. ''Bad'' returns are those that do not; their signals pass through the ionosphere. (v) Sonar: contains several patterns obtained by bouncing sonar signals off a metal cylinder at various angles and under various conditions. Discriminate between sonar signals bounced off a metal cylinder and those bounced off a roughly cylindrical rock. (vi) Magic: data are MC generated to simulate registration of high energy gamma particles in an atmospheric Cherenkov telescope. The events have to be classified background or signal.

Table 6 shows the acronyms, number of attributes, and number of instances for each of the six UCI repository datasets.

TABLE 6. Datasets from the uci machine learning repository.

Dataset	Acronym	Attributes	Instances
SPECTF Heart	SH	44	267
Bupa	Bu	6	345
Wbdc	Wb	30	569
Ionsphere	Io	33	351
Sonar	So	60	208
Magic	Ma	10	19020

The methodology applied for classification and evaluation of results was the same as explained in Sections IV-A and IV-B. Tables 7 and 8 show the results for each of the six UCI datasets in terms of Acc, Kap, and F1 score.

TABLE 7. Classification results for UCI datasets - A (values in %).

Method	SH			Bu			Wb		
	Acc	Kap	F1	Acc	Kap	F1	Acc	Kap	F1
LDA	70.4	65.4	70.6	60.4	54.9	60.6	88.7	83.9	88.8
NB	65.4	54.2	65.5	52.6	41.1	52.7	72.7	61.7	72.8
QDA	76.1	70.9	76.4	58.4	52.7	58.7	88.2	83.2	88.4
RDF	78.5	73.8	79.0	62.1	57.3	62.6	91.6	87.4	91.6
DS	77.2	73.2	77.6	60.6	56.5	61.0	93.3	92.3	93.5
Copulas	78.2	73.2	78.5	58.3	53.1	58.6	90.3	85.8	90.5
BKS	79.6	73.3	79.7	56.9	50.4	57.0	91.9	85.8	91.9
ICAMM	75.7	68.8	75.8	55.1	48.1	55.2	91.7	85.3	91.8
Mean	73.4	68.4	73.8	60.3	54.9	60.7	91.7	87.1	91.9
MV	71.2	67.2	71.5	56.2	52.0	56.5	92.6	89.0	92.7
α-LMSE	83.1	79.0	83.6	63.2	59.0	63.7	94.3	90.3	94.7
Linear LMSE (REG)	**83.7**	**79.9**	**84.1**	**63.2**	**59.6**	**63.6**	**94.7**	**91.4**	**95.0**
α- LMSE (REG)	**84.2**	**81.4**	**84.8**	**70.3**	**67.7**	**70.9**	**96.6**	**94.2**	**97.0**

TABLE 8. Classification results for UCI datasets - B (values in %).

Method	Io			So			Ma		
	Acc	Kap	F1	Acc	Kap	F1	Acc	Kap	F1
LDA	80.1	75.3	80.2	69.4	63.9	69.6	71.9	67.4	72.0
NB	65.0	54.0	65.1	63.9	52.4	64.0	65.5	55.4	65.6
QDA	82.8	77.8	83.0	71.7	66.0	72.0	70.3	66.3	70.4
RDF	86.8	82.6	86.8	72.4	67.6	72.9	74.0	69.9	74.1
DS	82.6	81.6	82.8	71.2	67.1	71.6	73.4	71.8	73.5
Copulas	85.8	81.3	86.0	75.9	70.7	76.2	72.6	68.3	72.9
BKS	90.4	84.3	90.4	64.2	57.7	64.3	73.5	67.2	73.6
ICAMM	90.0	83.6	90.1	60.1	53.1	60.2	72.1	66.1	72.3
Mean	75.5	70.9	75.7	73.1	67.7	73.5	73.4	69.2	73.7
MV	74.1	70.5	74.2	74.0	69.8	74.3	73.3	70.2	73.4
α-LMSE	91.2	87.2	91.6	80.3	76.1	80.8	74.1	70.9	74.4
Linear LMSE (REG)	**91.8**	**88.5**	**92.1**	**81.3**	**77.7**	**81.7**	**75.9**	**72.8**	**76.3**
α- LMSE (REG)	**91.9**	**89.5**	**92.3**	**82.5**	**79.9**	**83.1**	**76.7**	**74.5**	**77.0**

The results in Tables 7 and 8 show that, in general, fusion methods outperform the single classification methods and the proposed regularized methods, in order, α-LMSE (REG), and Linear LMSE (REG) yield the best results for all the indicators and datasets. Therefore, the results obtained for the UCI datasets were consistent with those explained in detail in sections IV-A and IV-B.

D. DISCUSSION

The proposed method has demonstrated to yield better results than the state-of-the-art methods with which it has was compared in various applications in previous sections. The improvement of the regularized alpha integration method for late fusion of classifiers has been both in accuracy and in the stability (standard deviation) of the results. Recently, the differences between early and late fusion has have been studied [18]. Theoretically, we can conclude that early fusion is the best option assuming ideal data availability. However, late fusion is the workable option considering real setups where the number of samples is limited. This is because, in late fusion, the number of parameters to be learned from training data is small, in the case of α-integration only $K+1$ parameters are to be estimated. Thus, overfitting is avoided and results are more stable.

Therefore, we can consider late fusion as an effort to approximate ideal early fusion under the finite training set size constraint. From that perspective, in general terms, any reasonable modification of the mean-square error cost function could lead to a better approximation. In our case, the inclusion of the regularization term should reduce the dispersion of the scores corresponding to the same class (intra-class separation). This should have a significant effect in the reduction of the probability of error, even if some reduction of the interclass separation would appear. The theoretical demonstration of this later is not straightforward (depending on the specific statistical data models) and is out the scope of this paper, however, we have demonstrated it, thoroughly, in practice.

The proposed iterative gradient algorithm showed steady convergence for minimizing the cost function. Gradient algorithms are simple to implement and have reasonably good convergence properties, particularly in combination with ad hoc techniques to avoid blocking in local minima. To this end, we used an annealing method in the implementation of the algorithm. The step size or learning rate was annealing during the adaptation process in order to provide faster and proper convergence. However, in order to improve the convergence to global minimum, there are other numerical approaches that could be implemented [52].

In the optimization process, the starting values for the parameters alpha and weights corresponded to the fusion by mean, i.e. $\alpha = -1$, $w_k = 1/k \forall k$. The fusion by mean usually gives an improvement over the single classifier results, thus, it was a good starting point for alpha integration. From this approach, the regularized alpha integration method was able to learn non-linear relationships between the random variables of the single classifier scores to improve the results. Non-linearities depend on the particular geometry of the score distributions, and their learning can lead to differences in classification, which, even if they are small, can be important from the point of view of the application.

V. CONCLUSION

Two new graph-based regularization methods for soft detector fusion have been theoretically derived and experimentally tested. The methods consider the linear combination of the individual detector statistics and their extension to a general non-linear fusion function called α-integration. The cost function formulated considers the mean-square error and a regularization term based on graph signal processing. The minimization of the regularization term reduces the dispersion of the fused statistics, and thus improves the separation between the statistics corresponding to the different hypotheses. The convergence of the proposed regularized methods was experimentally studied showing the robustness and stability of the proposed methods across a range of parameter settings, and thus, ensuring their consistent performance.

The performance of the proposed method was evaluated in a comparative analysis with competitive methods in two challenging problems (detection of ultrasound pulses and staging of neuropsychological tests) and six different applications from publicly available datasets. The superiority of the proposed methods over several state-of-the-art methods in terms of classification accuracy, kappa index, F1 score, and receiver operating characteristic analysis was demonstrated. Thus, the generalizability and effectiveness of the proposed methods were validated in an extensive range of datasets and domains across different applications.

From the results of this work, several objectives of research might be pursued such as the extension of the methods to multi-class classification. Of course, other applications could also be addressed.

REFERENCES

[1] H. Shi, H. Zhao, and W. Yao, "A transfer fusion framework for body sensor networks (BSNs): Dynamic domain adaptation from distribution evaluation to domain evaluation," *Inf. Fusion*, vol. 91, pp. 338–351, Mar. 2023.
[2] Y. Du, Y. Wang, J. Hu, X. Li, and X. Chen, "An emotion role mining approach based on multiview ensemble learning in social networks," *Inf. Fusion*, vol. 88, pp. 100–114, Dec. 2022.
[3] S. El-Sappagh, F. Ali, T. Abuhmed, J. Singh, and J. M. Alonso, "Automatic detection of Alzheimer's disease progression: An efficient information fusion approach with heterogeneous ensemble classifiers," *Neurocomputing*, vol. 512, pp. 203–224, Nov. 2022.
[4] H. Kim, S. M. Lee, and S. Choi, "Automatic sleep stages classification using multi-level fusion," *Biomed. Eng. Lett.*, vol. 12, no. 4, pp. 413–420, Nov. 2022.
[5] M. Ren, P. He, and J. Zhou, "Decision fusion of two sensors object classification based on the evidential reasoning rule," *Expert Syst. Appl.*, vol. 210, Dec. 2022, Art. no. 118620.
[6] S. Alyahyan and W. Wang, "Decision level ensemble method for classifying multi-media data," *Wireless Netw.*, vol. 28, no. 3, pp. 1219–1227, Apr. 2022.
[7] N. Padfield, J. Ren, C. Qing, P. Murray, H. Zhao, and J. Zheng, "Multi-segment majority voting decision fusion for MI EEG brain-computer interfacing," *Cognit. Comput.*, vol. 13, no. 6, pp. 1484–1495, Nov. 2021.
[8] A. Salazar, G. Safont, L. Vergara, and E. Vidal, "Pattern recognition techniques for provenance classification of archaeological ceramics using ultrasounds," *Pattern Recognit. Lett.*, vol. 135, pp. 441–450, Jul. 2020.
[9] Y. Zhang and H. J. Lee, "A target-aware fusion framework for infrared and visible images," *IEEE Access*, vol. 11, pp. 33666–33681, 2023.
[10] M. Alizadeh, D. S. Zadeh, B. Moshiri, and A. Montazeri, "Development of a customer churn model for banking industry based on hard and soft data fusion," *IEEE Access*, vol. 11, pp. 29759–29768, 2023.
[11] J. Choi, S. Yoo, X. Zhou, and Y. Kim, "Hybrid information mixing module for stock movement prediction," *IEEE Access*, vol. 11, pp. 28781–28790, 2023.
[12] Ø. K. Helgesen, A. Stahl, and E. F. Brekke, "Maritime tracking with georeferenced multi-camera fusion," *IEEE Access*, vol. 11, pp. 30340–30359, 2023.
[13] K. Zhao, L. Li, Z. Chen, R. Sun, G. Yuan, and J. Li, "A new multi-classifier ensemble algorithm based on D-S evidence theory," *Neural Process. Lett.*, vol. 54, no. 6, pp. 5005–5021, Dec. 2022.
[14] Z. Liu, Q. Pan, J. Dezert, J.-W. Han, and Y. He, "Classifier fusion with contextual reliability evaluation," *IEEE Trans. Cybern.*, vol. 48, no. 5, pp. 1605–1618, May 2018.
[15] P. K. Atrey, M. A. Hossain, A. El Saddik, and M. S. Kankanhalli, "Multimodal fusion for multimedia analysis: A survey," *Multimedia Syst.*, vol. 16, no. 6, pp. 345–379, Nov. 2010.
[16] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, "Multisensor data fusion: A review of the state-of-the-art," *Inf. Fusion*, vol. 14, no. 1, pp. 28–44, Jan. 2013.
[17] A. Ross and K. Nandakumar, "Fusion, score-level," in *Encyclopedia of Biometrics*, S. Z. Li, A. Jain, Eds. Boston, MA, USA: Springer, 2009, pp. 611–616, doi: 10.1007/978-0-387-73003-5_158.
[18] L. M. Pereira, A. Salazar, and L. Vergara, "A comparative analysis of early and late fusion for the multimodal two-class problem," *IEEE Access*, vol. 11, pp. 84283–84300, 2023.
[19] A. Salazar, L. Vergara, and E. Vidal, "A proxy learning curve for the Bayes classifier," *Pattern Recognit.*, vol. 136, Apr. 2023, Art. no. 109240.
[20] L. Vergara, A. Soriano, G. Safont, and A. Salazar, "On the fusion of non-independent detectors," *Digit. Signal Process.*, vol. 50, pp. 24–33, Mar. 2016.
[21] L. Vergara, "On the equivalence between likelihood ratio tests and counting rules in distributed detection with correlated sensors," *Signal Process.*, vol. 87, no. 7, pp. 1808–1815, Jul. 2007.
[22] S. E. Yuksel, J. N. Wilson, and P. D. Gader, "Twenty years of mixture of experts," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 8, pp. 1177–1193, Aug. 2012.
[23] M. Wozniak, M. Graña, and E. Corchado, "A survey of multiple classifier systems as hybrid systems," *Inf. Fusion*, vol. 16, pp. 3–17, Mar. 2014.
[24] H. Choi, S. Choi, and Y. Choe, "Parameter learning for alpha integration," *Neural Comput.*, vol. 25, no. 6, pp. 1585–1604, Jun. 2013.
[25] A. Soriano, L. Vergara, B. Ahmed, and A. Salazar, "Fusion of scores in a detection context based on alpha integration," *Neural Comput.*, vol. 27, no. 9, pp. 1983–2010, Sep. 2015.
[26] G. Safont, A. Salazar, and L. Vergara, "Multiclass alpha integration of scores from multiple classifiers," *Neural Comput.*, vol. 31, no. 4, pp. 806–825, Apr. 2019.
[27] G. Safont, A. Salazar, and L. Vergara, "Vector score alpha integration for classifier late fusion," *Pattern Recognit. Lett.*, vol. 136, pp. 48–55, Aug. 2020.
[28] D. Spielman, "Spectral graph theory," in *Combinatorial Scientific Computing*, U. Naumann and O. Schnek, Eds. Boca Raton, FL, USA: CRC Press, 2012, ch 16, pp. 1–23.
[29] R. Merris, "Laplacian matrices of a graph: A survey," *Linear Algebra Appl.*, vol. 197, pp. 143–176, Jan. 1994.
[30] X. D. Zhang, "The Laplacian eigenvalues of graphs: A survey," in *Linear Algebra Research Advances*, G. D. Ling, Ed. Hauppauge, NY, USA: Nova Science Publishers, 2007, pp. 201–228.
[31] J. Belda, L. Vergara, G. Safont, and A. Salazar, "Computing the partial correlation of ICA models for non-Gaussian graph signal processing," *Entropy*, vol. 21, no. 1, p. 22, Dec. 2018.
[32] J. Belda, L. Vergara, G. Safont, A. Salazar, and Z. Parcheta, "A new surrogating algorithm by the complex graph Fourier transform (CGFT)," *Entropy*, vol. 21, no. 8, p. 759, Aug. 2019.
[33] D. Zhou and B. Schölkopf, "A regularization framework for learning from graph data," in *Proc. ICML Workshop Stat. Relational Learn. Connections Other Fields*, Banff, AB, Canada, 2004, pp. 132–137.
[34] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *J. Mach. Learn. Res.*, vol. 7, pp. 2399–2434, Nov. 2006.
[35] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.
[36] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3042–3054, Jun. 2014.
[37] S.-I. Amari, "Integration of stochastic models by minimizing α-divergence," *Neural Comput.*, vol. 19, no. 10, pp. 2780–2796, Oct. 2007.
[38] D. Wu, "Parameter estimation for α-GMM based on maximum likelihood criterion," *Neural Comput.*, vol. 21, no. 6, pp. 1776–1795, Jun. 2009.
[39] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognit.*, vol. 38, no. 12, pp. 2270–2285, Dec. 2005.
[40] T. Eltoft, "Modeling the amplitude statistics of ultrasonic images," *IEEE Trans. Med. Imag.*, vol. 25, no. 2, pp. 229–240, Feb. 2006.
[41] X. Li, X. Zheng, T. Zhang, W. Guo, and Z. Wu, "Robust fault diagnosis of a high-voltage circuit breaker via an ensemble echo state network with evidence fusion," *Complex Intell. Syst.*, vol. 9, no. 5, pp. 5991–6007, Oct. 2023.
[42] C. Li, Y. Huang, and Y. Xue, "Dependence structure of Gabor wavelets based on copula for face recognition," *Expert Syst. Appl.*, vol. 137, pp. 453–470, Dec. 2019.
[43] A. Ragab, H. Ghezzaz, and M. Amazouz, "Decision fusion for reliable fault classification in energy-intensive process industries," *Comput. Ind.*, vol. 138, Jun. 2022, Art. no. 103640.

[44] G. Safont, A. Salazar, L. Vergara, E. Gómez, and V. Villanueva, "Multichannel dynamic modeling of non-Gaussian mixtures," *Pattern Recognit.*, vol. 93, pp. 312–323, Sep. 2019.
[45] S. Deepak and P. M. Ameer, "Brain tumor categorization from imbalanced MRI dataset using weighted loss and deep feature fusion," *Neurocomputing*, vol. 520, pp. 94–102, Feb. 2023.
[46] H. C. Kraemer, "Kappa coefficient," in *Wiley StatsRef: Statistics Reference Online*. New York, NY, USA: Wiley, 2014, pp. 1–4.
[47] N. Q. K. Le, D. T. Do, T.-T.-D. Nguyen, and Q. A. Le, "A sequence-based prediction of kruppel-like factors proteins using XGBoost and optimized features," *Gene*, vol. 787, Jun. 2021, Art. no. 145643.
[48] Q. Yuan, K. Chen, Y. Yu, N. Q. K. Le, and M. C. H. Chua, "Prediction of anticancer peptides based on an ensemble model of deep learning and machine learning using ordinal positional encoding," *Briefings Bioinf.*, vol. 24, no. 1, pp. 1–10, Jan. 2023.
[49] L. Vergara, A. Salazar, J. Belda, G. Safont, S. Moral, and S. Iglesias, "Signal processing on graphs for improving automatic credit card fraud detection," in *Proc. Annu. IEEE Int. Carnahan Conf. Secur. Technol. (ICCST)*, Madrid, Spain, 2017, pp. 1–6.
[50] E. Strauss, *A Compendium of Neuropsychological Tests*. London, U.K.: Oxford Univ. Press, 2006.
[51] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalogr. Clin. Neurophysiol.*, vol. 29, no. 3, pp. 306–310, Sep. 1970.
[52] J. Nocedal and S. J. Wright, *Numerical Optimization* (Springer Series in Operations Research), 2nd ed., USA: Springer, 2006, doi: 10.1007/978-0-387-40065-5.

ADDISSON SALAZAR (Member, IEEE) received the Ph.D. degree in electrical engineering from Universitat Politècnica de València (UPV), in 2011. He is currently a Senior Researcher with the Institute of Telecommunications and Multimedia Applications, UPV. He has published more than 100 papers in statistical signal processing, machine learning, decision fusion, and pattern recognition.

GONZALO SAFONT received the B.Eng. and Ph.D. degrees in telecommunications engineering from Universitat Politècnica de València (UPV), Valencia, Spain, in 2008 and 2015, respectively. He has involved in different applied problems in telecommunications and multimedia applications with UPV, including financial data mining, non-destructive testing, and biomedical diagnosis. He has authored 50 papers in JCR-indexed journals, book chapters, and international conference papers.

LUIS VERGARA received the Ph.D. degree in electrical engineering from Universidad Politécnica de Madrid, in 1983. He is currently a Full Professor in telecommunications, signal and data processing with Universitat Politècnica de València. He has more than 250 publications in theoretical and applied problems of signal and data processing and has led many important projects in these fields.

ENRIQUE VIDAL (Member, IEEE) is currently an emeritus Professor in computer science with Universitat Politècnica de València, Spain. He has published more than 250 research papers in the fields of pattern recognition, multimodal interaction and applications to language, and speech and image processing. He has led many important projects in these fields. He is a fellow of the International Association for Pattern Recognition (IAPR).