Multiclass Alpha Integration of Scores from Multiple Classifiers

Gonzalo Safont
gonsaar@upvnet.upv.es
Addisson Salazar
asalazar@dcom.upv.es
Luis Vergara
lvergara@dcom.upv.es
Universitat Politècnica de València, Instituto de Telecomunicaciones y Aplicaciones Multimedia, 46022 Valencia, Spain

Alpha integration methods have been used for integrating stochastic models and fusion in the context of detection (binary classification). Our work proposes separated score integration (SSI), a new method based on alpha integration to perform soft fusion of scores in multiclass classification problems, one of the most common problems in automatic classification. Theoretical derivation is presented to optimize the parameters of this method to achieve the least mean squared error (LMSE) or the minimum probability of error (MPE). The proposed alpha integration method was tested on several sets of simulated and real data. The first set of experiments used synthetic data to replicate a problem of automatic detection and classification of three types of ultrasonic pulses buried in noise (four-class classification). The second set of experiments analyzed two databases (one publicly available and one private) of real polysomnographic records from subjects with sleep disorders. These records were automatically staged in wake, rapid eye movement (REM) sleep, and non-REM sleep (three-class classification). Finally, the third set of experiments was performed on a publicly available database of single-channel real electroencephalographic data that included epileptic patients and healthy controls in five conditions (five-class classification). In all cases, alpha integration performed better than the considered single classifiers and classical fusion techniques.

1 Introduction

The fusion of information derived from different sources has gained much attention because of the potential improvement of automatic detectors and classifiers. Many such sources have been considered in the literature, including different modalities (Lahat, Adali, & Jutten, 2015), different features (Rivet, Wang, Naqvi, & Chambers, 2014), and different classifiers

(Wang et al., 2011; Mohandes, Deriche, & Aliyu, 2018). The combination of multiple experts or classifiers is the most typical of these, and it is being applied in such diverse areas as automatic staging of biomedical data (Wang et al., 2011; Zhang, Wu, Bai, & Chen, 2015), sentiment analysis (Fattah, 2015), risk analysis for financial services (Abellán & Mantas, 2014), and intruder detection in network security applications (Kevric, Jukic, & Subasi, 2017). Among the advantages to the fusion of multiple classifiers are improved classification performance, increased confidence, and greater reliability (Mohandes et al., 2018). The fusion can be performed at three steps of the classification process: (1) early fusion—a combination of the inputs used for each classifier; (2) late soft fusion—a combination of the scores assigned by each classifier to each class; and (3) late hard fusion—a combination of the decisions yielded by each classifier.

Many fusion algorithms have been proposed from different, yet sometimes equivalent, perspectives (Kittler, Hatef, Duin, & Matas, 1998; Yuksel, Wilson, & Gader, 2012; Khaleghi, Khamis, Karray, & Razavi, 2013). One main issue to deal with in this letter is statistical dependence between the outputs of the classifiers. Assuming independence generally leads to rather simple rules (hard fusion) or functions (soft fusion). However, when the classifiers are correlated, more sophisticated and complex methods are required (see, e.g., Poh & Bengio, 2005; Vergara, Soriano, Safont, & Salazar, 2016). One practical solution is to consider relatively simple rules or functions whose parameters are trained to optimize some predefined cost function. Thus, any possible correlation between the outputs of the classifiers is implicitly encompassed by the optimized parameters. One such approach is alpha integration, which was originally conceived as a method to integrate stochastic models (Amari, 2007; Wu, 2009). Least mean squared error (LMSE) methods for estimating the parameters of the model were proposed by Choi, Choi, Katake, and Choe (2010) and Choi, Choi, and Choe (2013).

Alpha integration was recently proposed as a soft fusion method for binary problems in a detection context (Soriano, Vergara, Ahmed, & Salazar, 2015). In that work, the fusion improved classification performance for the automatic detection of microarousals in patients with obstructive sleep apnea. Several optimality criteria were developed for the estimation of the parameters of alpha integration, such as LMSE and minimum probability of error (MPE).

In this letter, we propose a new method based on alpha integration to perform soft fusion of scores in multiclass classification problems. Essentially, alpha integration is performed separately on the scores assigned to each class by the classifiers. Thus, the multiclass problem is split into K binary classification problems. We have called this method separated score integration (SSI). Training algorithms are presented to optimize the parameters of SSI with respect to the LMSE and MPE criteria.

The performance of alpha integration was tested on several sets of simulated and real data. The first set of experiments used synthetic data to

simulate a case of simultaneous detection and classification of three types
of ultrasonic pulses buried in noise (the four-class classification problem).
The second set of experiments analyzed a publicly available database and
a private database containing polysomnographic (PSG) records from sub-
jects with sleep disorders. These records were automatically staged in wake,
rapid eye movement (REM) sleep, and non-REM sleep (the three-class clas-
sification problem). The last set of experiments considered a public database
of single-channel electroencephalographic data from epileptic patients and
control subjects in five conditions (the five-class classification problem). In
all cases, alpha integration was used to fuse the scores from four single clas-
sifiers: linear discriminant analysis (LDA), naive Bayes, classification trees,
and random forests (RDF). These methods were selected because of their
performance and their widespread use in many applications. Furthermore,
two types of classical fusion techniques were also considered: majority vot-
ing (hard fusion) and mean (soft fusion).

The rest of this letter is organized as follows. Section 2 reviews alpha
integration methods for the two-class case, which are extended to soft fu-
sion of K classes in section 3. Section 4 presents the results of the proposed
method on several sets of simulated and real data. Section 5 concludes.

2 Two-Class Alpha Integration

The two-class problem is essentially a test between hypotheses H_1 and H_0.
Given a group of D detectors working on these hypotheses, we would ob-
tain a set of scores s_i, $i = 1, \ldots, D$. We assume that these scores are normal-
ized between 0 and 1, with higher values of s_i indicating that H_1 is more
likely than H_0. The goal of alpha integration is optimally integrating these
scores into a unique score, s_α, given by

$$s_\alpha\left(\mathbf{s} = [s_1 \ldots s_D]^T\right) = \begin{cases} \left\{\displaystyle\sum_{i=1}^{D} w_i s_i^{\frac{1-\alpha}{2}}\right\}^{\frac{2}{1-\alpha}}, & \alpha \neq 1 \\ \exp\left\{\displaystyle\sum_{i=1}^{D} w_i \log(s_i)\right\}, & \alpha = 1 \end{cases}, \tag{2.1}$$

where α and the coefficients $\mathbf{w} = [w_1 \ldots w_D]^T$ are the parameters to be opti-
mized, subject to $w_i \geq 0$, $\sum_{i=1}^{D} w_i = 1$. Due to these constraints, s_α is bound
between 0 and 1.

Most simple soft fusion functions can be obtained as particular selections
of the parameters of alpha integration. For instance, given $w_i = 1/D \;\forall i$, one
can set α to obtain the arithmetic mean ($\alpha = -1$), the geometric mean ($\alpha =$
1), and the harmonic mean ($\alpha = 3$). Similarly, $\alpha = \infty \; (-\infty)$ is equivalent to
computing the minimum (maximum) of the scores. In general, however, the
parameters of alpha integration are optimized to satisfy some criterion.

2.1 LMSE Criterion. Let us assume we have a set of couples $\{\mathbf{s}^j, y^j\}$, $j = 1, \ldots, N$ where $\mathbf{s}^j = [s_1^j, \ldots, s_i^j, \ldots, s_D^j]^T$ is the vector of scores provided by the D detectors when y^j is the corresponding known binary decision ($y^j = 1$ if H_1 is true and $y^j = 0$ if H_0 is true). The mean squared error of alpha integration is defined as

$$\varepsilon = \frac{1}{N} \sum_{j=1}^{N} \left(y^j - s_\alpha(\mathbf{s}^j)\right)^2. \tag{2.2}$$

To learn the parameters of alpha integration, the derivatives of equation 2.2 with respect to α and $\mathbf{w}$ are calculated,

$$\frac{\partial \varepsilon}{\partial \alpha} = -\frac{2}{N} \sum_{j=1}^{N} \left(y^j - s_\alpha(\mathbf{s}^j)\right) \frac{\left(\partial s_\alpha(\mathbf{s}^j)\right)}{\partial \alpha}, \tag{2.3a}$$

$$\frac{\partial s_\alpha(\mathbf{s}^j)}{\partial \alpha} = \frac{2 s_\alpha(\mathbf{s}^j)}{1 - \alpha} \left\{ \frac{\log\left(\sum_{i=1}^{D} w_i f_\alpha\left(s_i^j\right)\right)}{1 - \alpha} + \frac{\sum_{i=1}^{D} w_i \frac{\partial f_\alpha\left(s_i^j\right)}{\partial \alpha}}{\sum_{i=1}^{D} w_i f_\alpha\left(s_i^j\right)} \right\}, \tag{2.3b}$$

$$\frac{\partial f_\alpha\left(s_i^j\right)}{\partial \alpha} = -\frac{1}{2} \log\left(s_i^j\right) \left(s_i^j\right)^{\frac{1-\alpha}{2}}, \tag{2.3c}$$

where, according to equation 2.1, $f_\alpha(\cdot)$ is a differentiable monotone function given by

$$f_\alpha(z) = \begin{cases} z^{\frac{1-\alpha}{2}}, & \alpha \neq 1 \\ \log z, & \alpha = 1 \end{cases}. \tag{2.4}$$

Moreover,

$$\frac{\partial \varepsilon}{\partial w_i} = -\frac{2}{N} \sum_{j=1}^{N} \left(y^j - s_\alpha(\mathbf{s}^j)\right) \frac{\partial s_\alpha(\mathbf{s}^j)}{\partial w_i}, \tag{2.5a}$$

$$\frac{\partial s_\alpha(\mathbf{s}^j)}{\partial w_i} = \begin{cases} \dfrac{2}{1 - \alpha} \left(\dfrac{s_\alpha(\mathbf{s}^j) f_\alpha\left(s_i^j\right)}{\sum_{l=1}^{D} w_l f_\alpha\left(s_l^j\right)} \right), & \alpha \neq 1 \\ \\ s_\alpha(\mathbf{s}^j) \log\left(s_i^j\right), & \alpha = 1 \end{cases}. \tag{2.5b}$$

Equations 2.3 to 2.5 can be used to optimize the parameters of the model—for instance, using gradient descent to update α and $\mathbf{w}$ as

$$\alpha(l+1) = \alpha(l) - \eta_\alpha \frac{\partial \varepsilon}{\partial \alpha}(l), \tag{2.6a}$$

$$w(l+1) = w(l) - \eta_w \frac{\partial \varepsilon}{\partial w}(l), \tag{2.6b}$$

where values η_α and η_w are the learning rate constants that control the speed of convergence.

2.2 MPE Criterion. The minimization of the probability of error, P_e is equivalent to the maximization of the probability of obtaining correct decisions, P_c through the whole set of couples $\{\mathbf{s}^j, y^j\}$, $j = 1, \ldots, N$. The MPE criterion was introduced in Soriano et al. (2015) as

$$-\log P_c = -\sum_{j=1}^{N} \{y_j \log(s_\alpha(\mathbf{s}^j)) + (1 - y_j)\log(1 - s_\alpha(\mathbf{s}^j))\}. \tag{2.7}$$

The derivatives of equation 2.7 with respect to the parameters of alpha integration are

$$\frac{\partial(-\log P_c)}{\partial \alpha} = -\sum_{j=1}^{N} \left(\frac{y^j}{s_\alpha(s^j)} - \frac{1 - y^j}{1 - s_\alpha(s^j)} \right) \frac{\partial \mathbf{s}_\alpha(s^j)}{\partial \alpha}, \tag{2.8a}$$

$$\frac{\partial(-\log P_c)}{\partial w_i} = -\sum_{j=1}^{N} \left(\frac{y^j}{s_\alpha(s^j)} - \frac{1 - y^j}{1 - s_\alpha(s^j)} \right) \frac{\partial \mathbf{s}_\alpha(s^j)}{\partial w_i}, \tag{2.8b}$$

where $\frac{\partial \mathbf{s}_\alpha(s^j)}{\partial \alpha}$ can be calculated using equations 2.3a and 2.3b, and $\frac{\partial \mathbf{s}_\alpha(s^j)}{\partial w_i}$ can be calculated using equation 2.5b. These new derivatives can be used to estimate the parameters that optimize the MPE criterion, for instance, with a gradient algorithm similar to equations 2.6a and 2.6b.

3 Multiclass Alpha Integration by Separated Score Integration

The modification of two-class alpha integration to the general case ($K \geq 2$) can be obtained by performing alpha integration separately on the scores corresponding to each class. Given K classes, indexed by $k = 1, \ldots, K$, and D classifiers, the ith classifier will produce a vector of scores $\mathbf{s}_i = [s_{1i}, \ldots, s_{Ki}]^T$, $i = 1, \ldots, D$. We will assume the scores are normalized to unit sum, $\sum_{k=1}^{K} s_{ki} = 1$. The true class identifier vector is defined as $\mathbf{y} = [y_1, \ldots, y_K]^T$, where

$$y_k = \begin{cases} 1 & \text{if the true class is } k \\ 0 & \text{otherwise} \end{cases}. \tag{3.1}$$

We define α_k and w_{ki}, $i = 1, \ldots, D$ as the parameters to integrate the scores corresponding to class k. Given a set of scores $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_D]$, we can directly apply the integration function, equation 2.1, to every class

$$s_{\alpha_k}(\mathbf{r}_k) = \begin{cases} \left(\displaystyle\sum_{i=1}^{D} w_{ki} \cdot s_{ki}^{\frac{1-\alpha_k}{2}} \right)^{\frac{2}{1-\alpha_k}}, & \alpha_k \neq 1 \\[2em] \exp \left(\displaystyle\sum_{i=1}^{D} w_{ki} \cdot \log (s_{ki}) \right), & \alpha_k = 1 \end{cases}, \quad k = 1, \ldots, K, \tag{3.2}$$

where $\mathbf{r}_k^T$ is the kth row of matrix $\mathbf{S}$. This way, the multiclass problem with K classes is converted in K separate two-class problems. In general, $\sum_{k=1}^{K} s_{\alpha_k} \neq 1$; this can be solved by normalizing the integrated scores in the form

$$s_{\alpha_k}^{norm} = \frac{s_{\alpha_k}}{\sum_{k=1}^{K} s_{\alpha_k}}. \tag{3.3}$$

Once we have fused the scores for all classes, classification is performed by selecting the class with the highest score:

$$\hat{k} = \underset{k}{\operatorname{argmax}} \, s_{\alpha_k}^{norm} = \underset{k}{\operatorname{argmax}} \, s_{\alpha_k}. \tag{3.4}$$

Notice that normalization is not needed to select the class through maximization. SSI is a more general version of the alpha integration of experts proposed in Amari (2016), since the weights w_{ki} depend on both the classifier and the class. The optimality results shown for alpha integration of experts in Amari (2016), however, hold true for SSI. Therefore, SSI is optimal under the alpha risk.

In the following, we present learning algorithms to optimize the parameters of SSI multiclass alpha integration with respect to two criteria, LMSE and MPE. Thus, we have two possible methods: SSI-LMSE and SSI-MPE.

3.1 Multiclass LMSE Criterion. Let us assume we have a set of couples $\{\mathbf{S}^j, \mathbf{y}^j\}$, $j = 1, \ldots, N$, where $\mathbf{S}^j = [\mathbf{s}_1^j, \ldots, \mathbf{s}_D^j]$ represents a matrix formed by the score vectors provided by the classifiers for the jth input sample and $\mathbf{y}^j$ is the corresponding true class identifier vector. The LMSE cost function for class k is

$$\varepsilon_k = \frac{1}{N} \sum_{j=1}^{N} \left(y_k^j - s_{\alpha_k} \left(\mathbf{r}_k^j \right) \right)^2 . \tag{3.5}$$

Hence, the optimum integration parameters for each class can be optimized separately using equations 2.3, 2.5, and 2.6.

3.2 Multiclass MPE Criterion. The MPE cost function for class k is given by

$$-\log P_c = -\sum_{j=1}^{N} \left[y_k^j \log \left(s_{\alpha_k} \left(\mathbf{r}_k^j \right) \right) + \left(1 - y_k^j \right) \log \left(1 - s_{\alpha_k} \left(\mathbf{r}_k^j \right) \right) \right].$$

$$\tag{3.6}$$

Hence, equations 2.3, 2.5, 2.6, and 2.8 can be used to obtain the optimum integration parameters separately for each class under the MPE criterion.

4 Experiments and Results

In order to test the proposed methods, we considered three kinds of experiments: a simulated experiment on ultrasound data, an experiment on real biomedical data from subjects with sleep disorders, and an experiment on real biomedical data from epileptic patients and healthy controls.

4.1 Simulated Nondestructive Testing Experiment by Ultrasound. The simulated experiment replicated a typical pulse-echo nondestructive testing (NDT) application. In these kinds of applications, several pulses are introduced into the test medium by a transducer. These pulses travel through the medium until they encounter defects or discontinuities in the material, at which point they are reflected back to the transducer and captured.

The parameters of the simulation were representative of real NDT applications, for example, the examination of archaeological ceramics for classification purposes (Salazar & Vergara, 2010). The simulated ultrasonic signal was sampled at $f_s = 100$ MHz during 0.75 ms for 75,000 samples. During this time, 12 pulses were received. The pulses had a duration of 10 microseconds, central frequency $f_0 = 10$ MHz, and their time of arrival was random but set so that there were no overlaps between pulses. Three types of pulses were considered, depending on their envelope: (P1) a gaussian envelope, (P2) a Poisson envelope, and (P3) a Mexican hat envelope. These pulses were buried in noise drawn from a K distribution with $v = 10$, which has been shown to describe the statistics of the envelope of the backscattered ultrasonic echo from a scattering medium (Eltoft, 2006). The peak signal-to-noise ratio (SNR) of the pulses was changed from 0 to 20 dB during the
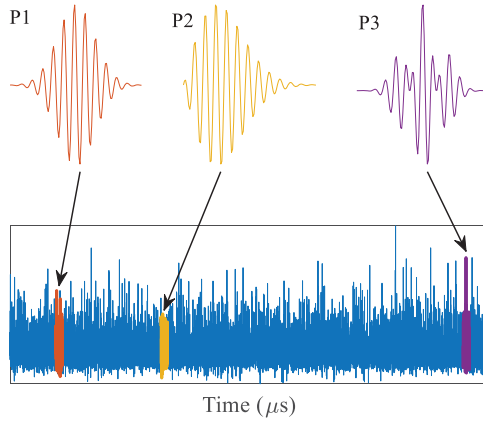
Figure 1: Example of the simulated ultrasound data with three types of pulses buried in background noise: (P1) gaussian envelope, (P2) Poisson envelope, and (P3) Mexican hat envelope.

experiment. Figure 1 shows the simulated signal, including one ultrasonic pulse of each type, for a peak SNR of 0 dB. As shown in the figure, the pulses were buried in noise and hard to find.

The simulated ultrasound signal was split into 10 microsecond epochs, and the following features were extracted: average amplitude, mean absolute value, standard deviation, skewness, kurtosis, maximum frequency, and centroid frequency. These features were then fed to several classifiers to perform classification in four classes: noise (N) and pulse types P1, P2, and P3. Therefore, the methods perform simultaneous detection and classification of the pulses. Four types of single classifiers were considered in this work: linear discriminant analysis (LDA), naive Bayes, classification trees, and random forests (RDF). The scores returned by the single classifiers were considered on their own and then fused using the proposed alpha integration method, optimized according to either the LMSE criterion (SSI-LMSE) or the MPE criterion (SSI-MPE). Two types of classical fusion techniques were included for comparison: majority voting (hard fusion) and mean (soft fusion).

To test the performance of the considered methods, the epochs were split equally into three data sets: training, validation, and testing. The single classifiers were trained using the training data set, and alpha integration methods were trained using the scores obtained by the single classifiers on the validation data set. The parameters of alpha integration were optimized using an interior point method for constrained optimization of the cost function, using the derivatives developed in section 3. Finally, the performance of all methods (single classifiers, classical fusion techniques, and alpha
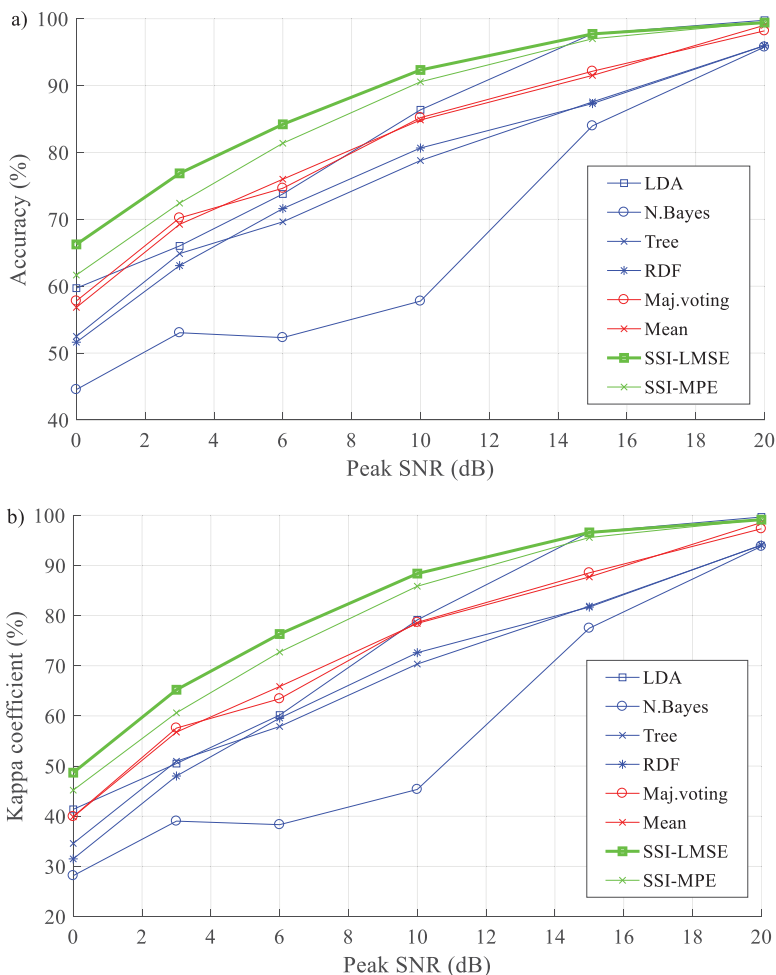
Figure 2: Average classification results of the simulation for increasing peak SNR of the pulses: (a) accuracy; (b) kappa values.

integration) was verified on the testing data set. Classification performance was measured as the average accuracy and Cohen's kappa for 1000 iterations of the experiment. In order to determine the effect of noise on the results, we repeated the experiment for the following peak SNR of the pulses: 0, 3, 6, 10, 15, and 20 dB.

Figure 2 shows the average classification accuracy and kappa values as a function of peak SNR. In all cases, performance increased with SNR, with all methods achieving very high results at 20 dB. LDA yielded the best

Table 1: Average Confusion Matrices for a Peak SNR of 10 dB of the Best Representatives of Each Kind of Method: Single Classifiers, Classical Fusion, and Alpha Integration.

True Class	LDA				Majority Voting				SSI-LMSE			
	N	P1	P2	P3	N	P1	P2	P3	N	P1	P2	P3
N	**98%**	0%	1%	1%	**81%**	3%	8%	8%	**97%**	0%	1%	2%
P1	0%	**81%**	18%	1%	0%	**98%**	1%	1%	0%	**96%**	2%	2%
P2	27%	1%	**67%**	5%	16%	3%	**80%**	1%	19%	2%	**78%**	1%
P3	0%	20%	8%	**72%**	6%	2%	1%	**91%**	2%	4%	2%	**92%**

Note: Bold numbers denote the diagonal of each confusion matrix—that is, the probability of correctly detecting each class with each method.

result out of all the considered single classifiers, and classical fusion techniques were able to improve over LDA only at low levels of SNR (SNR $\leq$ 6 dB). Conversely, alpha integration methods consistently outperformed all the other methods, with SSI-LMSE returning the best performance. The best-performing alpha integration method (SSI-LMSE) achieved an average 5.57% more accuracy and 7.81% more kappa than the best-performing single classifier (LDA), and 6.44% more accuracy and 8.15% more kappa than the best-performing classical fusion technique (majority voting). At very high SNR values (SNR $\geq$ 15 dB), LDA and alpha integration methods both return very similar results. In those instances, the weights w_{ki} estimated by alpha integration favored LDA over the rest of the classifiers.

The performance of alpha integration methods can be more easily shown on the distribution of the error for each class. Table 1 shows the average normalized confusion matrix at peak SNR = 10 dB for the best-performing single classifier (LDA), the best-performing classical fusion technique (majority voting), and the best-performing alpha integration method (SSI-LMSE). LDA was able to detect P1 and P3 pulses correctly but missed 27% of P2 pulses and experienced some confusion between pulse types. Classical fusion techniques were able to reduce pulse confusions and recover some missed P2 pulses, at the cost of an increased number of false alarms. Conversely, the best-performing alpha integration method was able to duplicate the benefits of classical fusion techniques without increasing the number of false alarms. This improvement shows that alpha integration (in this case, SSI-LMSE) was able to adapt its parameters and maximize the contributions of each classifier in the spaces it could be better used.

4.2 Analysis of EEG Data from Subjects with Sleep Disorders. The proposed alpha integration methods were also tested on a publicly available set of real polysomnograms (PSG) from the St. Vincent's University Hospital/University College Dublin Sleep Apnea Database in Physionet
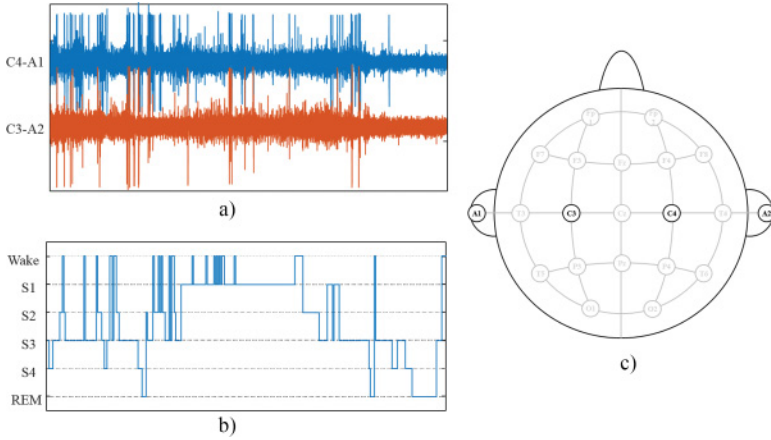
Figure 3: Example of the extracted data from one of the subjects: (a) the two bipolar EEG channels considered, C3-A2 and C4-A1; (b) hypnogram for that time frame; (c) diagram of the used 10–20 electrode system with the considered electrodes in black and ignored electrodes grayed out. In this work, three classes were considered: wake, REM sleep, and non-REM sleep (sleep stages 1 through 4).

(Heneghan, 2011). The database contains PSG from 25 adult subjects (21 male, 4 female) with suspected sleep disorders. The PSG is a multimodal biomedical record that includes many kinds of physiological signals. A full PSG is present for each subject, but in this work, we considered only two bipolar electroencephalographic (EEG) channels, C3-A2 and C4-A1. The EEG signals were sampled at 128 Hz and bandpass-filtered between 0.5 and 30 Hz; scoring was available for every 30 second epoch. An example of such signals is shown in Figure 3.

For the experiment, three classes were considered: wake, rapid eye movement (REM) sleep, and non-REM sleep (sleep stages 1 through 4). In order to perform classification, the following features were extracted from each EEG channel in 30 second epochs: power in frequency bands delta (0–4 Hz), theta (5–7 Hz), alpha (8–12 Hz), sigma (13–15 Hz), and beta (16–30 Hz); and the activity, mobility, and complexity of the signal (Hjorth, 1973). These features are typically used in the literature on sleep staging (Motamedi-Fakhr, Moshrefi-Torbati, Hill, Hill, & White, 2014).

We considered the same methods used for the simulated experiments: four single classifiers (LDA, naive Bayes, classification trees, and RDF), two classical fusion techniques (majority voting and mean), and the proposed alpha integration method (SSI-LMSE and SSI-MPE). For the experiment, each subject was classified independently from the rest. Similar to the simulated experiment, the data were split into three data sets: training,
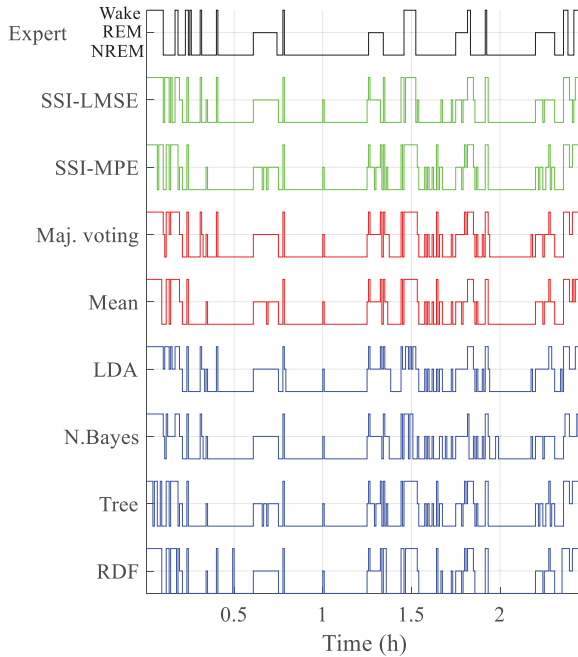
Figure 4: Resulting classifications on the testing data set for one of the subjects.

validation, and testing. In order to preserve the prior probabilities, the observations of each class were distributed as evenly as possible across the three data sets. For some subjects, however, this meant that the single classifiers had to be trained using fewer observations than variables, which led to stability issues. For such subjects, the class containing insufficient observations was simply eliminated from the data. As in the simulated experiment, the considered single classifiers were trained on the training data set, the proposed alpha integration methods were trained on the scores of the single classifiers on the validation data set, and the performance of all methods was estimated on the testing data set. The results for each subject were obtained as the average of 100 iterations.

An example of the classification obtained for one of the patients is shown in Figure 4. It can be seen that alpha integration methods yielded classes that were more in line with the actual labels provided by the expert. This was particularly true for classes after 1.5 hours, where the variability of the labels of the single classifiers and classical fusion techniques was larger than that of alpha integration methods, particularly SSI-LMSE.

The average accuracy and kappa values for all 25 patients are shown in Figure 5. These values are similar to those in the literature for this dataset
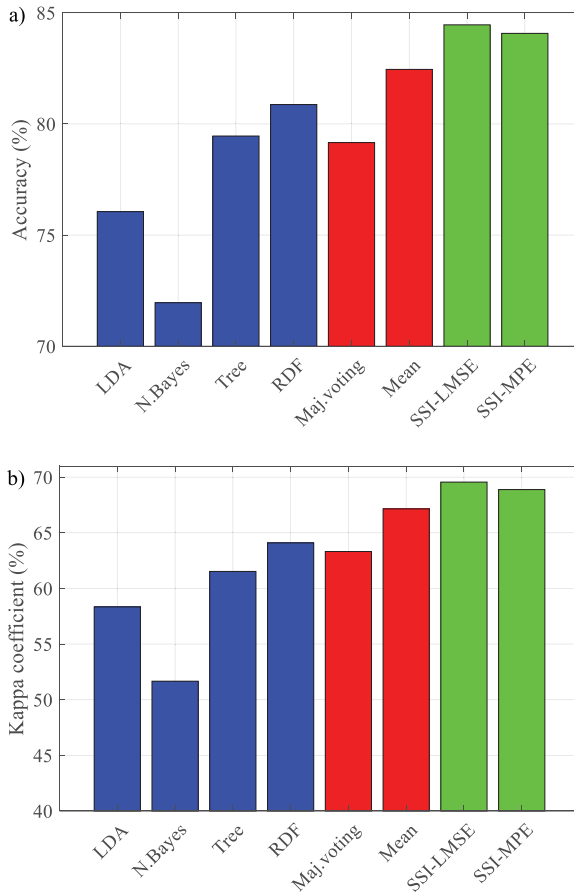
Figure 5: Average classification results on subjects with sleep disorders: (a) accuracy; (b) kappa values.

(Xie & Minn, 2012; Wang, Hua, Hao, & Xie, 2017). Results are largely in accordance with those of the simulated experiment (see Figure 2), albeit with reduced overall performance, as expected from the difficulty of the problem and the use of real data. All the proposed alpha integration methods yielded better results than the considered single classifiers and classical fusion techniques. The best-performing alpha integration method, SSI-LMSE, achieved an average 3.59% more accuracy and 5.47% more kappa than the best-performing single classifier (RDF), and 2.01% more accuracy and 2.41% more kappa than the best-performing classical fusion technique (mean).

The accuracy yielded by each method for each patient is shown in Figure 6. Higher values are indicated by brighter colors; the red squares
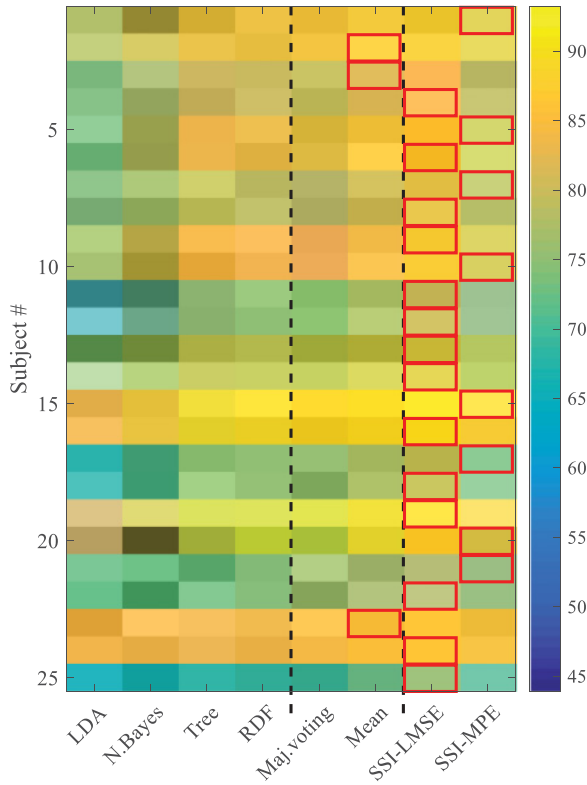
Figure 6: Accuracy obtained for each method for each subject. The red squares mark the best result for each subject.

indicate the best result for each patient. In accordance with the values in Figure 5, the proposed alpha integration methods yielded the best result in 22 of the 25 subjects, with classical fusion techniques yielding the best result for the remaining three subjects. In the subjects where the best performance was yielded by a classical fusion technique (subjects 2, 3, and 23), the difference with the results yielded by alpha integration was small (less than 1% in all cases), and both alpha integration and classical fusion techniques were able to improve on the results of the single classifiers. These results indicate that the improvement in performance yielded by the proposed alpha integration methods was consistent across subjects.

In order to further verify the performance of the proposed alpha integration methods, a second experiment was performed using different data. This second experiment considered six subjects diagnosed with obstructive sleep apnea, and the data were captured with the help of the Neurology and Neurophysiology Units of Hospital Universitari i Politècnic La Fe, Valencia
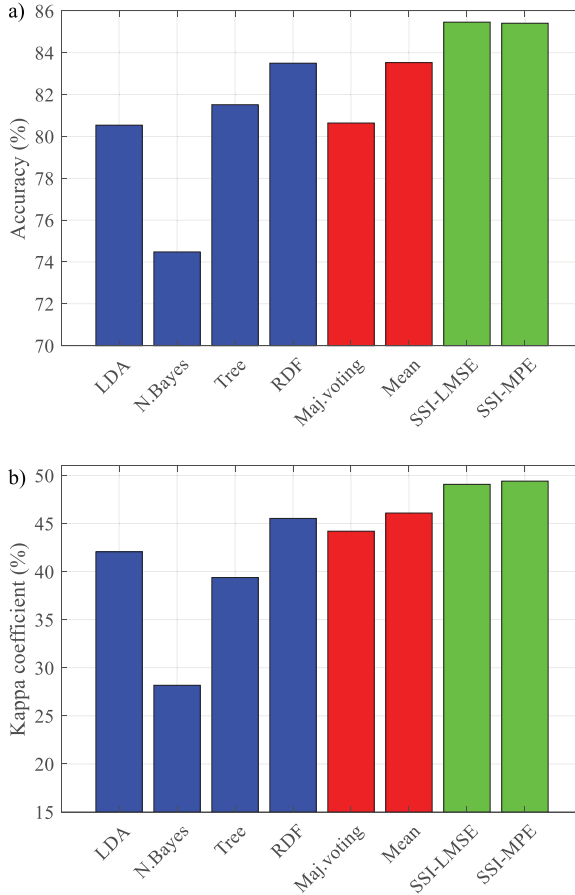
Figure 7: Average classification results on the second data set subjects with sleep disorders: (a) accuracy; (b) kappa values.

(Spain). A full PSG is present for each subject, but in this work, we considered only channels C3-A2 and C4-A1. The EEG signals were sampled at 256 Hz, and scoring was provided by a physician at irregular intervals for every night. These signals were then converted to the same 30 second epochs that were considered for the first experiment on real data. In concordance with the first experiment, any classes containing insufficient observations were removed from the data prior to classification. Performance was evaluated using the same Monte Carlo experiments and performance indicators considered for the previous experiments.

The average classification results on the second data set are shown in Figure 7. Results are consistent with those of the simulated experiment (see

Figure 2) and the first experiment on real data (see Figure 5). The considered classical fusion techniques were generally unable to improve on the single classifiers, whereas alpha integration methods were capable of optimally combining the available results from the different classifiers. Numerically, the best-performing alpha integration method (SSI-LMSE) obtained an average improvement of 1.96% more accuracy and 3.54% more kappa than the best-performing single classifier (RDF) and 1.94% more accuracy and 2.98% more kappa than the best-performing classical fusion technique (mean).

4.3 Analysis of EEG Data from Epileptic Subjects and Healthy Controls. The proposed alpha integration methods were also tested on a publicly available database of 500 single-channel real EEG segments from epileptic subjects and healthy controls (Andrzejak et al., 2001). The EEG were recorded under five conditions: (A) healthy control, eyes open; (B) healthy control, eyes closed; (C) epileptic subject, electrode contralateral to the epileptogenic zone; (D) epileptic subject, electrode in the epileptogenic zone but without seizures; and (E) EEG record with epileptic activity. Each EEG segment was sampled at 173.61 Hz and lasted 23.6 seconds. Figure 8 shows one representative EEG segment from each group.

The feature extraction procedure follows the one presented in Nunes, Coelho, Lima, Papa, and de Albuquerque (2014). The EEG segments were band-filtered between 0.5 and 40 Hz and then decomposed in five levels using the discrete wavelet transform (DWT) with a Coiflet wavelet of order 2. Then the following statistics were applied to each subband: standard deviation, power, absolute mean, and ratio of absolute means between adjacent subbands. This resulted in 23 features per EEG segment.

We considered the same methods used for the simulated experiments: four single classifiers (LDA, naive Bayes, classification trees, and RDF), two classical fusion techniques (majority voting and mean), and the proposed alpha integration methods (SSI-LMSE and SSI-MPE). Similar to the simulated experiment, the data were split into three data sets (training, validation, and testing) in such a way so as to preserve prior probabilities. The considered single classifiers were trained on the training data set, the proposed alpha integration methods were trained on the scores of the single classifiers on the validation data set, and the performance of all methods was estimated on the testing data set. The results for each subject were obtained as the average of 100 iterations. The results of this experiment, shown in Figure 9, are in concordance with those in the previous sections.

Classical fusion techniques were able to improve over the best-performing single classifier, which was LDA. In turn, alpha integration outperformed classical fusion techniques, with SSI-MPE yielding the best performance. Numerically, alpha integration (SSI-MPE) achieved an average 3.30% more accuracy and 4.12% more kappa than the best-performing single classifier (LDA), and 1.77% more accuracy and 2.22% more kappa than the best-performing classical fusion technique (mean). Furthermore,

A) Healthy control, eyes open

B) Healthy control, eyes closed

C) Epileptic patient, hippocampus contralateral to focal point

D) Epileptic patient, hippocampus, close to focal point
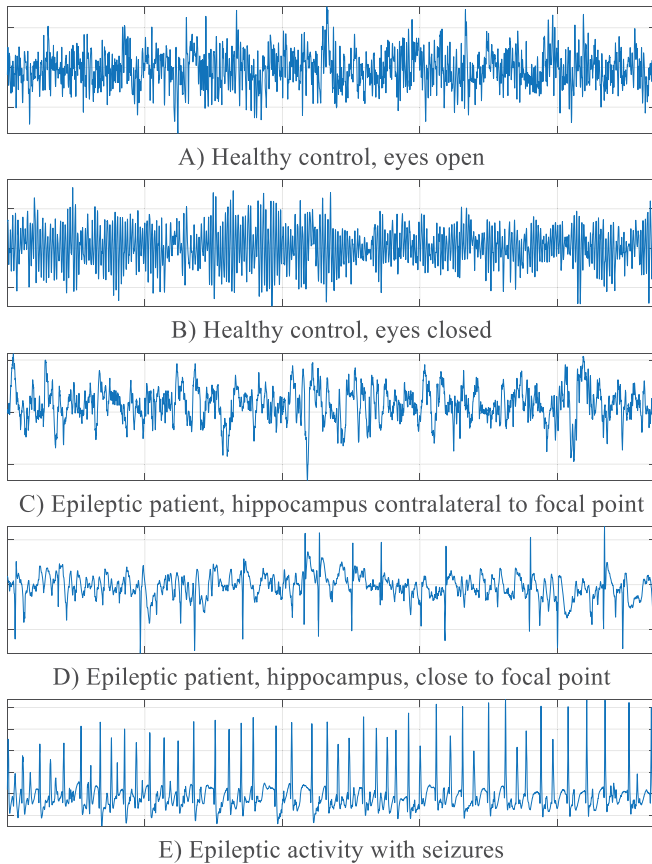
E) Epileptic activity with seizures

Figure 8: Examples of the EEG segments captured for each group.

the values obtained by alpha integration equal or exceed those of state-of-the-art methods as the ones presented in (Nunes et al., 2014).

5 Conclusion

This work has proposed a new method based on alpha integration to perform soft fusion of scores in multiclass classification problems: separated score integration (SSI). Training algorithms have been developed to optimize the parameters of SSI with respect to either of two criteria, least mean squared error (LMSE) or minimum probability of error (MPE). The proposed methods were tested on several sets of simulated and real data. In these experiments, alpha integration was used to fuse the results from four single classifiers: LDA, naive Bayes, classification trees, and RDF. Two
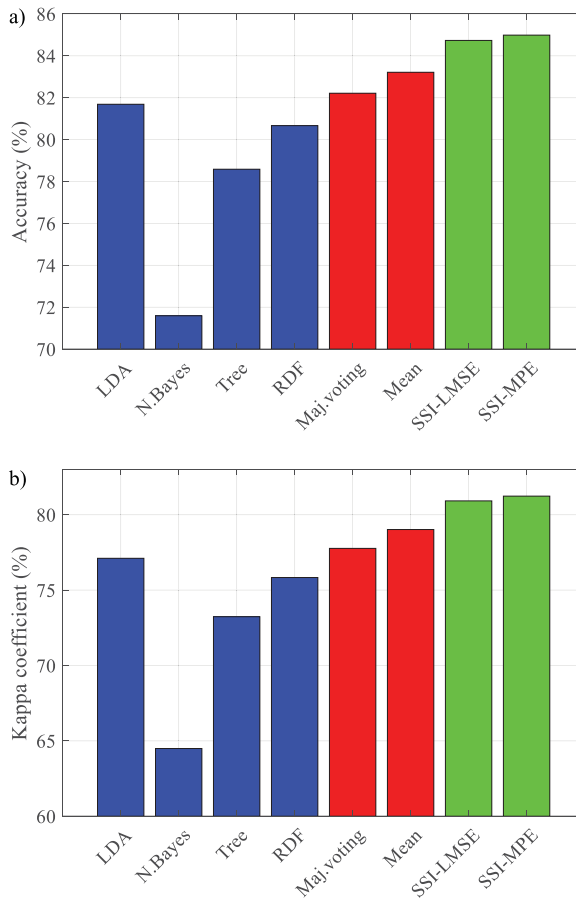
Figure 9: Average classification results on the data set of epileptic subjects and healthy controls: (a) accuracy; (b) kappa values.

classical fusion techniques (majority voting and mean) were also included for comparison.

The simulated experiment modeled the simultaneous detection and classification of three types of ultrasonic pulses buried in K-distributed noise (i.e., a four-class problem), considering the effect of the signal-to-noise ratio on the result. Two experiments were performed on real data from publicly available databases. The first experiment performed automatic staging of polysomnographic records from 25 patients with sleep disorders in three classes: wake, rapid eye movement (REM) sleep, and non-REM sleep. These results were verified on a private database. The second experiment

considered single-channel EEG records captured from epileptic patients and healthy controls and classified them into five different classes.

The proposed alpha integration methods yielded better results than the single classifiers and classical fusion techniques across all three experiments. For the simulated experiment, this improvement was higher for low values of peak SNR, the most difficult cases. For the experiments on real data, the classical fusion techniques were unable to improve significantly on the considered single classifiers; however, the proposed alpha integration methods were still able to exploit information from all single classifiers in order to yield optimally fused scores.

Acknowledgments

References

Abellán, J., & Mantas, C. (2014). Improving experimental studies about ensembles of classifiers for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, *41*, 3825–3830.

Amari, S. (2007). Integration of stochastic models by minimizing α-divergence. *Neural Computation*, *19*, 2796–2780.

Amari, S. (2016). *Information geometry and its applications.* Berlin: Springer.

Andrzejak, R., Lehnertz, K., Rieke, C., Mormann, F., David, P., & Elger, C. (2001). Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E*, *64*, 061907.1–061907.8.

Choi, H., Choi, S., & Choe, Y. (2013). Parameter learning for alpha integration. *Neural Computation*, *25*, 1585–1604.

Choi, H., Choi, S., Katake, A., & Choe, Y. (2010). Learning α-integration with partially-labeled data. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* (pp. 2058–2061). Piscataway, NJ: IEEE.

Eltoft, T. (2006). Modeling the amplitude statistics of ultrasonic images. *IEEE Transactions on Medical Imaging*, *25*(2), 229–240.

Fattah, M. (2015). New term weighting schemes with combination of multiple classifiers for sentiment analysis. *Neurocomputing*, *167*, 434–442.

Heneghan, C. (2011). St. Vincent's University Hospital/University College Dublin Sleep Apnea Database. https://www.physionet.org/pn3/ucddb/

Hjorth, J. (1973). The physical significance of time domain descriptors in EEG analysis. *Electroencephalography and Clinical Neurophysiology*, *34*(3), 321–325.

Kevric, J., Jukic, S., & Subasi, A. (2017). An effective combining classifier approach using tree algorithms for network intrusion detection. *Neural Computation and Applications*, *28*(suppl.), 1051–1058.

Khaleghi, B., Khamis, A., Karray, F., & Razavi, S. (2013). Multisensor data fusion: A review of the state-of-the-art. *Information Fusion*, *14*, 28–44.

Kittler, J., Hatef, M., Duin, R., & Matas, J. (1998). On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*, 226–239.

Lahat, D., Adali, T., & Jutten, C. (2015). Multimodal data fusion: An overview of methods, challenges and prospects. *Proceedings of the IEEE*, *103*, 1449–1477.

Mohandes, M., Deriche, M., & Aliyu, S. (2018). Classifiers combination techniques: A comprehensive review. *IEEE Access*, *6*, 19626–19639.

Motamedi-Fakhr, S., Moshrefi-Torbati, M., Hill, M., Hill, C., & White, P. (2014). Signal processing techniques applied to human sleep EEG signals—A review. *Biomedical Signal Processing and Control*, *10*, 21–33.

Nunes, T., Coelho, A., Lima, C., Papa, J., & de Albuquerque, V. (2014). EEG signal classification for epilepsy diagnosis via optimum path forest: A systemaic assessment. *Neurocomputing*, *123*, 103–123.

Poh, N., & Bengio, S. (2005). How do correlation and variance of base experts affect fusion in biometric authentication tasks? *IEEE Transactions on Signal Processing*, *53*, 4384–4396.

Rivet, B., Wang, W., Naqvi, S., & Chambers, J. (2014). Audiovisual speech source separation: An overview of key methodologies. *IEEE Signal Processing Magazine*, *31*(3), 125–134.

Salazar, A., & Vergara, L. (2010). ICA Mixtures applied to ultrasonic nondestructive classification of archaeological ceramics. *EURASIP Journal on Advances in Signal Processing*, *2010*, 125201.

Soriano, A., Vergara, L., Ahmed, B., & Salazar, A. (2015). Fusion of scores in a detection context based on alpha-integration. *Neural Computation*, *27*, 1983–2010.

Vergara, L., Soriano, A., Safont, G., & Salazar, A. (2016). On the fusion of nonindependent detectors. *Digital Signal Processing*, *50*, 24–33.

Wang, S., Anugu, V., Nguyen, T., Rose, N., Burns, J., McKenna, M., . . . Summers, R. (2011). Fusion of machine intelligence and human intelligence for colonic polyp detection in CT colonography. In *Proceedings of the International Symposium on Biomedical Imaging: From Nano to Macro* (pp. 160–164). Piscataway, NJ: IEEE.

Wang, S., Hua, G., Hao, G., & Xie, C. (2017). A cycle deep belief network model for multivariate time series classification. *Mathematical Problems in Engineering*, 2017, 1–7. doi: 10.1155/2017/9549323

Wu, D. (2009). Parameter estimation for α-GMM Based on maximum likelihood criterion. *Neural Computation*, *21*, 1776–1795.

Xie, B., & Minn, H. (2012). Real-time sleep apnea detection by classifier combination. *IEEE Transactions on Information Technology in Biomedicine*, *16*(3), 469–477.

Yuksel, S., Wilson, J., & Gader, P. (2012). Twenty years of mixture of experts. *IEEE Transactions on Neural Networks and Learning Systems*, *23*, 1177–1193.

Zhang, J., Wu, Y., Bai, J., & Chen, F. (2015). Automatic sleep stage classification based on sparse deep belief net and combination of multiple classifiers. *Transactions of the Institute of Measurement and Control*, *38*(4), 435–451.