



Article

Does Score Bias Correction Improve the Fusion of Classifiers?

Luis Vergara and Addisson Salazar *

Institute of Telecommunications and Multimedia Applications, Universitat Politècnica de València,
Camino de Vera s/n, 46022 València, Spain; lvergara@dcom.upv.es

* Correspondence: asalazar@dcom.upv.es

Abstract

We demonstrate that the potential bias in the scores generated by individual classifiers negatively affects their fusion. Consequently, we present an algorithm to improve the effectiveness of score fusion in classification. The algorithm corrects the score class conditional bias before fusion. The interest of the procedure is demonstrated theoretically, first in general terms and then considering exponential models for the score class conditional distributions. The case of beta distributions is also addressed using Monte Carlo simulations. Finally, a real-life application of fusion of two modalities (EEG, ECG) and two classifiers (Gaussian Bayes and Logistic Regression) is included, showing significant improvement with respect to conventional fusion without bias correction.

Keywords: score fusion; bias correction; classification; fusion statistic; classifier fusion; multimodal fusion; unimodal fusion

1. Introduction

Fusion of scores is of interest whenever different entities collaborate to arrive at a common opinion on the occurrence of a given phenomenon. This is particularly relevant in the fusion of automatic classifiers where a score is to be assigned to each possible class [1–7]. Fused classifiers can correspond to one or several modalities. In the latter case, the term “multimodal fusion” is often used [8–15]. In the Bayesian approach, the scores can be considered posterior probabilities of each class, so they are restricted within the range 0–1 and must add up to 1. This imposes a relevant limitation on the score fusion framework, as explained below.

Let us consider the two-class problem $c = 0, 1$. Given a feature vector x , a score s within $(0, 1)$ is assigned to class $c = 1$, and a score $1 - s$ is assigned to class $c = 0$. Therefore s can be considered an estimate of a binary random variable from a given feature vector x . A good estimator should provide scores close to 1 for feature vectors belonging to class $c = 1$ and close to 0 for feature vectors belonging to class $c = 0$. From that perspective, notice that s is inevitably a conditionally biased estimate. This is immediately evident. Let us call $p(s/c = i)$ to the probability density function (PDF) of s conditional to class $c = i$. Moreover, conditional to class $c = i$, the true value of the binary random variable will be i . Then, let us compute the conditional biases b_0 and b_1 (for simplicity in the rest of the analysis, we define the bias so that both are positive).



Academic Editor: Guoqing Chao

Received: 17 October 2025

Revised: 15 November 2025

Accepted: 21 November 2025

Published: 24 November 2025

Citation: Vergara, L.; Salazar, A. Does Score Bias Correction Improve the Fusion of Classifiers? *Mach. Learn. Knowl. Extr.* **2025**, *7*, 151. <https://doi.org/10.3390/make7040151>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

$$\begin{aligned}
 b_0 &= E(s/c = 0) - 0 = \underbrace{\int_0^1 s \cdot p(s/c = 0) ds}_{\geq 0} - 0 \Rightarrow 0 \leq b_0 \leq 1 \\
 b_1 &= 1 - E(s/c = 1) = 1 - \underbrace{\int_0^1 s \cdot p(s/c = 1) ds}_{\leq 1} \geq 0 \Rightarrow 0 \leq b_1 \leq 1,
 \end{aligned} \tag{1}$$

where the lower bounds only hold for perfect (unrealistic) estimators, i.e., $p(s/c = i) = \delta(s - i)$. These (inevitable) biases impose a limitation on classifier performance. Let us demonstrate it from a rather inverse perspective: if we could compensate the bias, the detector performance would improve. Before that, let us assume that we use a maximum a posteriori (MAP) criterion, where the most probable class is selected, i.e., as follows:

$$\begin{aligned}
 &\begin{matrix} c=1 \\ > \\ s & 1-s \\ < \\ c=0 \end{matrix} \Leftrightarrow \begin{matrix} c=1 \\ > \\ s & 0.5 \\ < \\ c=0 \end{matrix}.
 \end{aligned} \tag{2}$$

This implicitly assumes that the scores are calibrated, that is, that $P(c = 1/s) = s$. However, there is no conceptual problem in implementing the bias correction method described in the following section with uncalibrated scores. In fact, the goal of the correction method is to improve accuracy (a “post-thresholding indicator”), whereas calibration refers to using the scores as probabilities of class membership (a “pre-thresholding indicator”). Therefore, calibration and correction can be complementary; for example, it will always be possible to calibrate the corrected scores.

Equation (2) defines the two-class classification problem as a detection problem, where class $c = 1$ is detected when the statistic s is greater than the threshold 0.5 (the term detector will be frequently used throughout the article). As we see below, if we could correct the conditional bias b_0 , the probability of false alarm $PFA = \int_{0.5}^1 p(s/c = 0) ds$ will decrease and if we could correct the conditional bias b_1 , the probability of detection $PD = \int_{0.5}^1 p(s/c = 1) ds$ will increase. Correcting b_0 involves a shift to the left (towards 0) of $p(s/c = 0)$, while correcting b_1 involves a shift to the right (towards 1) of $p(s/c = 1)$. Let us call PFA_b and PD_b to the corresponding probabilities after bias correction, then

$$\begin{aligned}
 PFA_b &= \int_{0.5}^{1-b_0} p(s + b_0/c = 0) ds \underbrace{=}_{s'=s+b_0} \int_{0.5+b_0}^1 p(s'/c = 0) ds' < \int_{0.5}^1 p(s'/c = 0) ds' = PFA \\
 PD_b &= \int_{0.5}^{1+b_1} p(s - b_1/c = 1) ds \underbrace{=}_{s'=s-b_1} \int_{0.5-b_1}^1 p(s'/c = 1) ds' > \int_{0.5}^1 p(s'/c = 1) ds' = PD.
 \end{aligned} \tag{3}$$

Therefore, correcting the bias reduces PFA and increases PD , thus improving the detector performance. Unfortunately, direct correction of the conditional bias is an ill-posed problem. Certainly, we can compute sample estimates $\hat{b}_0$ and $\hat{b}_1$ from labeled training data. However, compensating for the bias of the sample under test requires knowledge of the true class because we have to subtract $\hat{b}_0$ from the computed score if the true class is $c = 0$, but we must add $\hat{b}_1$ if the true class is $c = 1$. But determining the true class of the sample under test is exactly the problem we are trying to solve from the very beginning. Furthermore, averaging individual scores from different detectors to obtain a score with reduced conditional bias is not an option because the individual biases are all positive. For example, it is well-known that by averaging N independent and identically distributed random variables (i.i.d.r.v.), we obtain a random variable where variance has been reduced by a factor N . However, the mean (and so the possible bias in an estimation context)

remains the same. A more formal analysis of the conditional bias influence on the optimum fusion of classifiers is given in [16] for some specific models.

Despite the above discouraging arguments, we propose in the following section an algorithm for conditional bias correction in the framework of score fusion. We first provide an intuitive (rather heuristic) rationale of the proposal. Then, a general formal analysis is carried out. This later indicates that the expected improvement depends on the integration of the bivariate class conditional PDFs in a given area. So, in Section 3, the analytically tractable case of exponential PDFs is analyzed. Then in Section 4, we consider the more general case of beta-PDFs. Beta-distribution is the more common model for random variables in $(0, 1)$. Unfortunately, this is analytically infeasible, so we resorted to Monte Carlo simulations. Finally, in Section 5, we evaluated the usefulness of the proposed algorithm in a real-world experiment, combining two biosignal modalities, electroencephalograms (EEGs) and electrocardiograms (ECGs), and two lightweight classifiers, Gaussian Bayes (GB) and Logistic Regression (LR), to improve sleep arousal detection.

2. The Algorithm for Conditional Bias Correction

Let us consider two individual classifiers defined, respectively, by the variable $m = 1, 2$. Given a sample under test x_m , each classifier provides a score $0 \leq s_m \leq 1$, which, according to the previous section, represents the posterior probability of class $c = 1$, i.e., $P(c = 1/x_m) = s_m$, and hence $P(c = 0/x_m) = 1 - s_m$. Every classifier implements the MAP test (2) with s_m ; hence, the corresponding probabilities of false alarm and detection will be given by

$$\begin{aligned} PFA_m &= Pr(s_m > 0.5/c = 0) = \int_{0.5}^1 (s_m/c = 0) ds_m \\ PD_m &= Pr(s_m > 0.5/c = 1) = \int_{0.5}^1 (s_m/c = 1) ds_m . \end{aligned} \quad (4)$$

We will assume that s_1 and s_2 are i.i.d.r.v. to facilitate the analysis, so that we can write

$$\begin{aligned} (s_1 s_2 / c = i) &= p(s_1 / c = i) \cdot p(s_2 / c = i) = p_i(s_1) p_i(s_2) \\ PFA_1 &= PFA_2 = PFA \\ PD_1 &= PD_2 = PD . \end{aligned} \quad (5)$$

Let us start by the case of no bias correction. Scores are fused by computing the mean $\frac{1}{2}(s_1 + s_2)$, which is to be compared with the 0.5 threshold as in (2). For convenience, we prefer to consider the score $= s_1 + s_2$ $0 \leq z \leq 2$, and compare z with a threshold 1, which is an equivalent test. It is evident that if both original scores exceed the threshold of 0.5, their sum will exceed the threshold of 1. And vice versa if both individual scores are below the threshold of 0.5. Therefore, the possible improvement through fusion will occur when both classifiers make different decisions, such that the score of the one who makes a mistake is offset by the score of the one who makes a correct decision.

Now consider that before adding the scores we apply a conditional bias correction. Let us define the biases of every classifier as in (1): $b_{m0} = E[s_m/c = 0] - 0 > 0$ $b_{m1} = 1 - E[s_m/c = 1] > 0$. From assumption (5), we may simplify to $b_{10} = b_{20} = b_0$ and $b_{11} = b_{21} = b_1$. These values can be estimated during training from labeled samples, but for simplicity, we keep the notation b_0 and b_1 for the computed estimates. The bias correction is implemented as indicated by the following equations (Figure 1 depicts the general scheme):

$$s_m^b = \begin{cases} s_m - b_0 & \text{if } s_m < 0.5 \\ s_m + b_1 & \text{if } s_m > 0.5 \end{cases} . \quad (6)$$

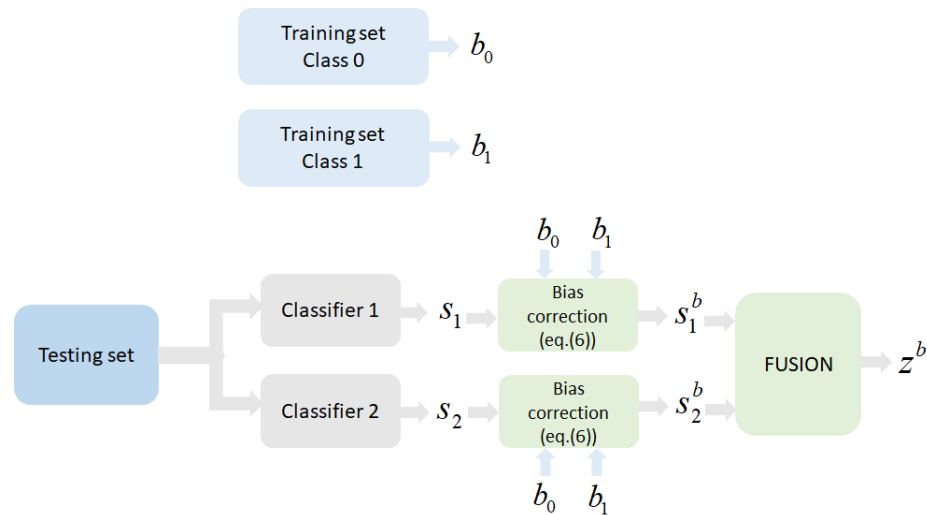


Figure 1. General scheme for conditional bias correction.

This can be interpreted as accepting the class selected by the individual classifier as “true” to perform the corresponding correction of the conditional bias. Certainly, the individual classifier will provide both correct and incorrect decisions, depending on its own performance (essentially its operating point given by PD_m and PFA_m). The following analysis deduces the expected performance of the proposed method, implicitly incorporating the performances of the individual detectors.

First notice that working with the corrected scores by no means modifies the performance of the individual classifiers. This is because the bias is always positive, hence if $s_m < 0.5 \Rightarrow s_m^b = s_m - b_0 < 0.5$ and if $s_m > 0.5 \Rightarrow s_m^b = s_m + b_1 > 0.5$. Consider now that the fusion consists of $z^b = s_1^b + s_2^b - 2b_0 \leq z^b \leq 2 + 2b_1$, and compare z^b with a threshold of 1. Again, if both original scores exceed the threshold of 0.5, then $z^b = s_1 + s_2 + 2b_1 > 1$. But if both original scores are below the threshold of 0.5, then $z^b = s_1 + s_2 - 2b_0 < 1$. Therefore, as with z , the potential improvement using z^b will be obtained in cases where individual classifiers make different decisions, i.e., the following:

$$z^b = s_1 + s_2 + b_1 - b_0 \begin{cases} \text{where } s_1 > 0.5 \text{ } s_2 < 0.5 \\ \text{where } s_1 < 0.5 \text{ } s_2 > 0.5 \end{cases} \quad (7)$$

Obviously, if $b_1 = b_0 \Rightarrow z^b = z$, i.e., the correction of biases does not contribute anything with respect to conventional fusion. However, let us assume that the class to be detected, as defined in (2), is the one with the most biased scores, i.e., $b_1 > b_0 \Rightarrow z^b > z$. Therefore, if the correct decision in (7) is $c = 1$, bias correction will contribute to increasing the probability of detection. Certainly, if the correct decision is $c = 0$, bias correction will contribute to increasing the probability of false alarm. But, considering that $b_1 > b_0$ (the scores corresponding to $c = 1$ are “worst” than the scores corresponding to $c = 0$), disagreement events of the type (7), should be more likely when the true class is $c = 1$. Then, it is expected that the increasing probability of detection will be more “significant” than the increasing probability of false alarm. This rather heuristic argument encourages us to undertake a formal analysis of the procedure, as we do below.

Let us call, respectively, PD_z and PFA_z the probability of detection and false alarm of $z = s_1 + s_2$, and PD_{z^b} and PFA_{z^b} the probability of detection and false alarm of $z^b = s_1^b + s_2^b$. Our aim is to deduce how PD_z and PFA_z change when bias correction is implemented, i.e., to deduce the relation between PD_{z^b} and PD_z , as well as PFA_{z^b} and PFA_z . But, as already stated above $z^b > z$, which allows us to ensure that $PD_{z^b} > PD_z$ and $PFA_{z^b} > PFA_z$. The increase in both probabilities will occur due to those cases in which both detectors make a

different decision, being $z < 1$ and $z^b > 1$. Let us express this conclusion in terms of s_1 and s_2 (we define $\Delta = b_1 - b_0 \Rightarrow 0 \leq \Delta \leq 1$).

$$\begin{aligned} PD_{z^b} &= PD_z + 2PD_{z^b} = PD_z + 2Pr(s_1 + s_2 + \Delta > 1/s_1 + s_2 < 1, s_1 < 0.5, s_2 > 0.5, c = 1) \\ PFA_{z^b} &= PFA_z + 2PFA_{z^b} = PFA_z + 2Pr(s_1 + s_2 + \Delta > 1/s_1 + s_2 < 1, s_1 < 0.5, s_2 > 0.5, c = 0) \end{aligned} \quad (8)$$

The factor 2 appears because the two events, $s_1 < 0.5, s_2 > 0.5$, and $s_1 > 0.5, s_2 < 0.5$, are equivalent as far as we have assumed in (5) that s_1 and s_2 have the same class conditional PDFs. Figure 2 helps us to compute the required probabilities in (8). The shadowed area corresponds to the values s_1 and s_2 that satisfy the conditioning in (8); therefore, we have to integrate the corresponding bivariate PDF in that area. From (5) and Figure 2, we can write

$$\begin{aligned} D_{z^b} &= PD_z + 2 \left(\int_0^{0.5-\Delta} \int_{1-s_1-\Delta}^{1-s_1} p_1(s_1)p_1(s_2)ds_1ds_2 + \int_{0.5-\Delta}^{0.5} \int_{0.5}^{1-s_1} p_1(s_1)p_1(s_2)ds_1ds_2 \right) 0 \leq \Delta \leq 0.5 \\ PD_{z^b} &= PD_z + 2 \int_0^{0.5} \int_{0.5}^{1-s_1} p_1(s_1)p_1(s_2)ds_1ds_2 0.5 \leq \Delta \leq 1 \\ PFA_{z^b} &= PFA_z + 2 \left(\int_0^{0.5-\Delta} \int_{1-s_1-\Delta}^{1-s_1} p_0(s_1)p_0(s_2)ds_1ds_2 + \int_{0.5-\Delta}^{0.5} \int_{0.5}^{1-s_1} p_0(s_1)p_0(s_2)ds_1ds_2 \right) 0 \leq \Delta \leq 0.5 \\ PFA_{z^b} &= PFA_z + 2 \int_0^{0.5} \int_{0.5}^{1-s_1} p_0(s_1)p_0(s_2)ds_1ds_2 0.5 \leq \Delta \leq 1. \end{aligned} \quad (9)$$

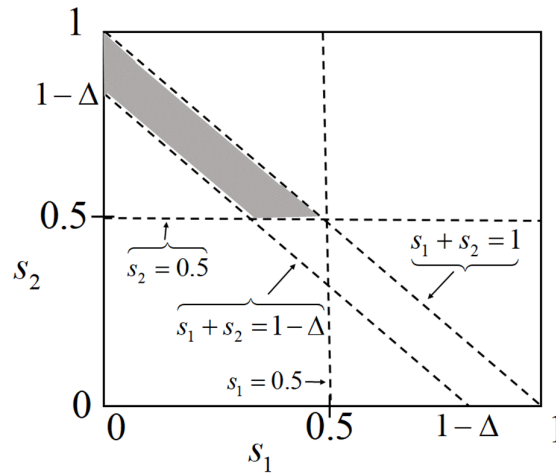


Figure 2. Area (shadowed) where $p_1(s_1)p_1(s_2)$ and $p_0(s_1)p_0(s_2)$ have to be integrated to calculate, respectively, the increase in probability of detection and probability of false alarm.

Thus, the potential improvement of fusion with bias correction depends on how the individual PDFs integrate inside the shadowed area of Figure 2. In principle, it is expected that the bivariate PDF for class $c = 0$ would be concentrated in the lower left corner of the square in Figure 2, while the corresponding for class $c = 1$ should be concentrated in the upper right corner. On the other hand, the greater the bias, the greater the occupancy of the bivariate PDF along the square, so the integral of $p_1(s_1)p_1(s_2)$ in the shadowed area should be larger than that of $p_0(s_1)p_0(s_2)$. A more specific conclusion requires knowledge of $p_1(\cdot)$ and $p_0(\cdot)$. Therefore, in the next section, we will consider conditional PDFs to be truncated exponentials. While this is reasonably realistic, the exponential assumption makes the analysis tractable.

Furthermore, in Section 4, we will assume the beta distribution. This is the most common distribution for random variables in the finite domain $(0, 1)$. It covers a wide variety of distributions by adjusting two parameters. Unfortunately, a theoretical analysis for the beta distribution is not feasible, so we resort to Monte Carlo simulations.

3. The Exponential Distribution Case

Let us consider that $p_0(s_m)$ and $p_1(s_m)$ are, respectively, a truncated exponential and a transformed truncated exponential PDF in the interval $(0, 1)$, namely the following:

$$\begin{aligned} p_0(s_m) &= \begin{cases} \frac{\lambda_0}{1-e^{-\lambda_0}} e^{-\lambda_0 s_m} & 0 \leq s_m \leq 1 \\ 0 & \text{rest} \end{cases} \\ p_1(s_m) &= \begin{cases} \frac{\lambda_1}{1-e^{-\lambda_1}} e^{-\lambda_1(1-s_m)} & 0 \leq s_m \leq 1 \\ 0 & \text{rest} \end{cases} \end{aligned} \quad (10)$$

These two exponential models, respectively, concentrate more probability as we approach $s_m = 0$ from the right side or as we approach $s_m = 1$ from the left side, which seems an expected property of reasonable classifiers. Moreover, according to (1)

$$\begin{aligned} b_0 &= E(s_m / c = 0) = \frac{1-e^{-\lambda_0}}{\lambda_0} \\ b_1 &= 1 - E(s_m / c = 1) = \frac{1-e^{-\lambda_1}}{\lambda_1} \end{aligned} \quad (11)$$

We have represented in Figure 3a two exponential distributions, where the blue one corresponds to a truncated exponential $p_0(s_m)$ with $\lambda_0 = 10 \rightarrow b_0 = 0.1$ and the red one corresponds to a transformed truncated exponential $p_1(s_m)$ with $\lambda_1 = 3.2 \rightarrow b_1 = 0.3$. Note that the greater the bias, the slower the decay of the exponentials. Then in Figure 3b, we have represented the contour plots of $p_0(s_1)p_0(s_2)$ (blue) and $p_1(s_1)p_1(s_2)$ (red) considering the conditional PDFs of Figure 3a. We have also marked in Figure 3b the shadowed area from Figure 2 where both bivariate PDFs have to be integrated to calculate the increase in PD_z and PFA_z due to bias correction. Clearly, the distribution corresponding to the class with more bias expands more along the square, leading to a greater overlap with the shaded area. Therefore, greater increase in detection probability than in false alarm probability is to be expected. In order to obtain a more precise and quantitative assessment, we have calculated the integrals in the shaded area for this exponential case. Thus, we first calculate in Appendix A the probabilities corresponding to the individual detectors and to the fusion without bias correction. The expressions obtained are

$$\begin{aligned} PD &= \frac{1-e^{-\lambda_1 0.5}}{1-e^{-\lambda_1}}; PFA = e^{-\lambda_0 0.5} \frac{1-e^{-\lambda_0 0.5}}{1-e^{-\lambda_0}} \\ PD_z &= \frac{(1-e^{-\lambda_1}(1+\lambda_1))}{(1-e^{-\lambda_1})^2}; PFA_z = \frac{e^{-\lambda_0}(1-e^{-\lambda_0}-\lambda_0)}{(1-e^{-\lambda_0})^2} \end{aligned} \quad (12)$$

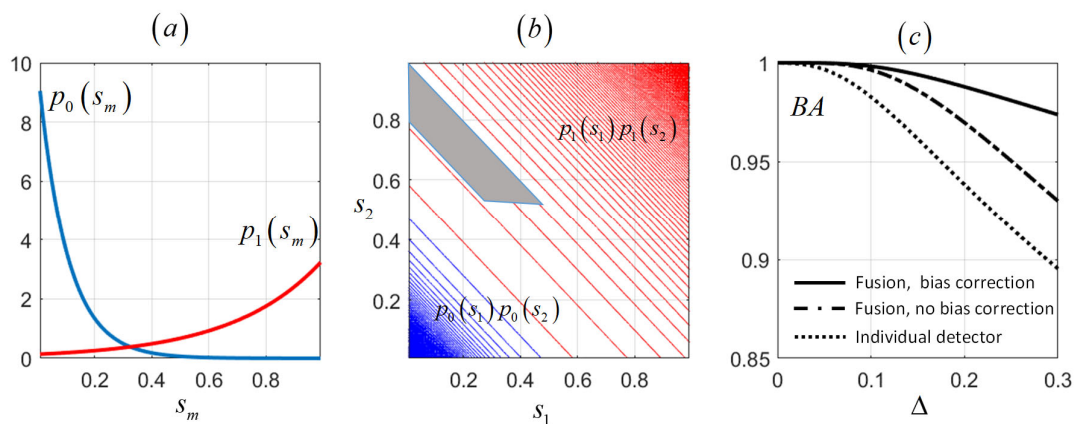


Figure 3. Exponential PDF cases. (a) $p_0(s_m)$ with $b_0 = 0.1$, and $p_1(s_m)$ with $b_1 = 0.3$. (b) Contour plots of the corresponding bivariate PDFs and area (shadowed) to be integrated. (c) Balanced accuracy for $b_0 = 0.05$ and $0.05 \leq b_1 \leq 0.35 \Leftrightarrow 0 \leq \Delta \leq 0.30$.

Then, we also calculate in Appendix A the integrals required in (9) for the computation of PD_{zb} and PFA_{zb} .

$$\begin{aligned}
 PD_{zb} &= PD_z + 2 \left(\frac{\lambda_1 e^{-\lambda_1}}{(1-e^{-\lambda_1})^2} (1 - e^{-\lambda_1 \Delta}) (0.5 - \Delta) + \frac{\lambda_1 e^{-\lambda_1}}{(1-e^{-\lambda_1})^2} \left(\Delta - \frac{1}{\lambda_1} (1 - e^{-\lambda_1 \Delta}) \right) \right) \quad 0 \leq \Delta \leq 0.5 \\
 PD_{zb} &= PD_z + 2 \left(\frac{\lambda_1 e^{-\lambda_1}}{(1-e^{-\lambda_1})^2} \left(0.5 - \frac{1}{\lambda_1} (1 - e^{-\lambda_1 0.5}) \right) \right) \quad 0.5 \leq \Delta \leq 1 \\
 PFA_{zb} &= PFA_z + 2 \left(\frac{\lambda_0 e^{-\lambda_0}}{(1-e^{-\lambda_0})^2} (e^{\lambda_0 \Delta} - 1) (0.5 - \Delta) + \frac{\lambda_0 e^{-\lambda_0}}{(1-e^{-\lambda_0})^2} \left(\frac{1}{\lambda_0} (e^{\lambda_0 \Delta} - 1) - \Delta \right) \right) \quad 0 \leq \Delta \leq 0.5 \\
 PFA_{zb} &= PFA_z + 2 \left(\frac{\lambda_0 e^{-\lambda_0}}{(1-e^{-\lambda_0})^2} \left(\frac{1}{\lambda_0} (e^{\lambda_0 0.5} - 1) - 0.5 \right) \right) \quad 0.5 \leq \Delta \leq 1.
 \end{aligned} \tag{13}$$

Using the previous expressions, we can calculate the balanced accuracy (BA), defined as the mean of the “sensitivity” (equivalent to probability of detection) and the “specificity” (equivalent to 1 minus the probability of false alarm), namely the following:

$$\begin{aligned}
 BA &= \frac{1}{2} (PD + 1 - PFA) = 1 - \left(\frac{1}{2} (1 - PD) + \frac{1}{2} PFA \right) = 1 - PE \\
 BA_z &= \frac{1}{2} (PD_z + 1 - PFA_z) = 1 - \left(\frac{1}{2} (1 - PD_z) + \frac{1}{2} PFA_z \right) = 1 - PE_z \\
 BA_{zb} &= \frac{1}{2} (PD_{zb} + 1 - PFA_{zb}) = 1 - \left(\frac{1}{2} (1 - PD_{zb}) + \frac{1}{2} PFA_{zb} \right) = 1 - PE_{zb},
 \end{aligned} \tag{14}$$

where PE , PE_z , and PE_{zb} are the probabilities of error assuming equal priors $P_0 = P_1 = \frac{1}{2}$. Then, we show in Figure 3c the balanced accuracies for $b_0 = 0.05$ and $0.05 \leq b_1 \leq 0.35 \Leftrightarrow 0 \leq \Delta \leq 0.30$. As expected, the balanced accuracy decreases with increasing Δ , but is always higher when bias correction is applied. In fact, the improvement achieved compared to individual classifiers, as well as conventional fusion, is greater the larger Δ is.

4. The Beta Distribution Case

In the previous section, we carried out a theoretical study, made possible by the analytical tractability of the exponential distribution. However, the most common distribution for random variables in the finite domain $(0, 1)$ is the beta distribution. This distribution is defined by two parameters, $\alpha > 0$ and $\beta > 0$, which allow for a wide variety of PDFs to be considered. Unfortunately, its analytical expression prevents a theoretical study like the previous one, so we will resort to Monte Carlo simulations to evaluate the effectiveness of the bias correction. Let us consider that $p_0(s_m)$ is a beta-PDF: $p_0(s_m) = \frac{s_m^{\alpha_0-1} (1-s_m)^{\beta_0-1}}{B(\alpha_0, \beta_0)}$ where $B(\alpha_0, \beta_0)$ is the beta-function, and that $p_1(s_m) = \frac{(1-s_m)^{\alpha_1-1} s_m^{\beta_1-1}}{B(\alpha_1, \beta_1)}$, which also corresponds to a beta-PDF but with interchanged parameters. Among the wide variety of beta-PDFs, we have selected some that seem appropriate for the case at hand. This was carried out by first setting a value $\alpha_0 = \alpha_1 = \alpha$ in $p_0(s_m)$ and $p_1(s_m)$, and then calculating the values β_0 and β_1 which, respectively, give biases b_0 and b_1 . Notice that $\beta_i = \alpha_i \frac{1-b_i}{b_i}$. Thus, in Figure 4a(i–iii), we have represented beta-PDFs, respectively, corresponding to $\alpha = 0.4$, $\alpha = 0.7$, and $\alpha = 1.0$. In all three cases, the values β_0 and β_1 have been calculated to give biases $b_0 = 0.1$ and $b_1 = 0.3$. Note that each α parameter defines a different type of PDF, in any case appropriate to concentrate more probability to the left ($p_0(s_m)$) or to the right ($p_1(s_m)$) of the threshold of 0.5. As in Figure 3b, we have represented in Figure 4b(i–iii) the contour plots of $p_0(s_1)p_0(s_2)$ (blue) and $p_1(s_1)p_1(s_2)$ (red) considering, respectively, the conditional PDFs of Figure 4a(i–iii). We have marked in Figure 4b(i–iii) the shadowed area from Figure 2 where both bivariate PDFs have to be integrated to calculate the increase in PD_z and PFA_z due to bias correction. Once again, the distribution corresponding to the class $c = 1$ expands more along the square, leading to a greater overlap with the shaded

area. Then, we performed Monte Carlo simulations for the three α values, a unique value $b_0 = 0.05$ for class $c = 0$ and a range of values $0.05 \leq b_1 \leq 0.35$ for class $c = 1$, so that $0 \leq \Delta \leq 0.30$. We have generated 10,000 samples from every required beta-PDF. Then, we have computed the balanced accuracies using (14), considering the MAP detector (2) for the individual scores, for the sum of the original scores and for the sum of the bias-corrected scores. Correction was implemented as indicated in (6) with sample estimates of the conditional biases b_0 and b_1 . The results are shown in Figure 4c(i–iii). Similarly to the exponential case, an increase in bias reduces the balanced accuracy, while correcting for bias always produces the best result, with the relative improvement increasing with Δ .

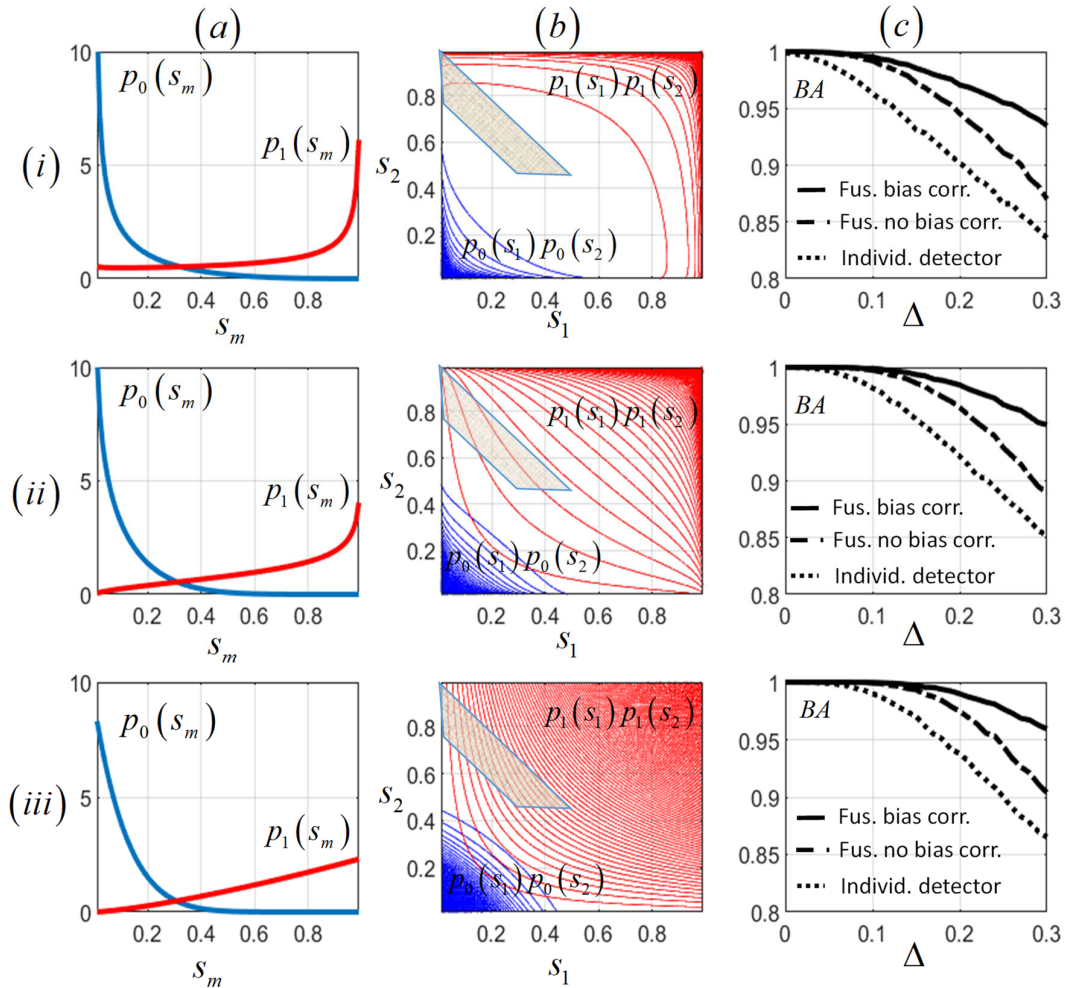


Figure 4. (a) Beta-PDFs for $\alpha = 0.4$ (i), $\alpha = 0.7$ (ii), and $\alpha = 1.0$ (iii), in all three cases $b_0 = 0.1$ and $b_1 = 0.3$. (b) Contour plots of the corresponding bivariate PDFs and area (shadowed) to be integrated. (c) Balanced accuracies computed by Monte Carlo simulations for $\alpha = 0.4$ (i), $\alpha = 0.7$ (ii), and $\alpha = 1.0$ (iii) with $b_0 = 0.05$ and $0.05 \leq b_1 \leq 0.35 \Leftrightarrow 0 \leq \Delta \leq 0.30$.

5. A Real Data Case

This section is devoted to evaluating the effectiveness of conditional bias correction in an automatic biomedical signal analysis problem to facilitate diagnosis. Two signal modalities and two classification methods will be considered. As we will see, this will open up the opportunity for 10 possible options, including unimodal classifiers and fusion of scores from two different modalities and/or two different classification methods.

The goal is to develop an automatic classifier for arousals [17,18] during sleep, as their frequency is correlated with the presence of apnea and epilepsy. An expert physician, through a thorough analysis of so-called polysomnograms, usually performs this

classification manually: a set of different types of biomedical signals recorded during monitoring of the patient's sleep. The record of arousal detections is called hypnogram. It is of obvious interest to use automatic hypnograms to alleviate the tedious and subjective task of the physician.

The polysomnograms were obtained from the public database Physionet [19]. In this experiment, we will consider the EEG and ECG signals recorded synchronously during patient sleep. Then, 8 EEG features and 12 ECG features were obtained, respectively, from the EEG and ECG signals in non-overlapped intervals of 30 sec. A total number of 10 patients were considered. The number of intervals varies for every patient, ranging from 749 (6 h and 14 min) to 925 (7 h and 42 min). The eight EEG features were powers in the frequency bands delta (0–4 Hz), theta (5–7 Hz), alpha (8–12 Hz), sigma (13–15 Hz), and beta (16–30 Hz), and three Hjorth parameters: activity, mobility, and complexity. These features are routinely used in the analysis of EEG signals [20–22]. The 12 ECG features were the following: autoregressive coefficients (4), Shannon's entropy maximal overlap discrete wavelet packet transform at level four (4), and multiscale wavelet variance estimates up to fourth order using a Daubechies wavelet (4). More details on how to compute these features may be found in [23–26].

In this problem, class 1 corresponds to arousal and class 0 corresponds to any of the other possible sleep stages. We have considered two different classifiers. The first one is the Gaussian Bayes (GB), a generative classifier which assumes general Gaussian models for both $p(x/c = 1)$ and $p(x/c = 0)$, then the scores are calculated using

$$s = \frac{e^{-\frac{1}{2}((x-\mu_1)^T C_1^{-1}(x-\mu_1))}}{e^{-\frac{1}{2}((x-\mu_1)^T C_1^{-1}(x-\mu_1))} + e^{-\frac{1}{2}((x-\mu_0)^T C_0^{-1}(x-\mu_0))}}, \quad (15)$$

where μ_0 , μ_1 , C_0 , C_1 are the mean vectors and covariance matrices of the feature vectors, which have been estimated from the training data using maximum likelihood (ML) estimates. The second one is Logistic Regression, a discriminative method which computes the score using

$$s = \frac{1}{1 + e^{-w^T x}}. \quad (16)$$

There are different options for fitting the coefficient vector w ; in our case, we opted for a closed form solution by solving an overdetermined system of linear equations. There exists an equation of the form $w^T x = \nabla$ for each instance x of the training set, where ∇ is a large positive number if x belongs to $c = 1$ or a large (magnitude) negative number if x belongs to $c = 0$. The classifiers were trained separately for every patient using the first half of the respective EEG or ECG recordings. Training includes the calculation of the conditional bias estimates $\hat{b}_0$ and $\hat{b}_1$ required by the proposed correction algorithm (5). The second half was then used for testing. The scores of all patients were grouped by class. Thus, each classifier generated a total of 6706 class 0 scores and 1708 class 1 scores.

We have considered ten options. The first four correspond to one modality—one classifier. The next two also consider one modality but the two classifiers are fused. Finally, the last four correspond to the fusion of the two modalities, either with the same classifier or a different one. Fusion is always implemented by simply computing the mean score.

Thus, Table 1 shows the results of the 10 options (numbered from 1 to 10) without applying bias correction. The conditional bias is shown as well as the balanced accuracy. Notice the bias is significantly higher for class 1 and that fusion does not reduce it (in fact, the bias after fusion is the mean bias). Also notice that fusion of GB + LR for one modality does not improve over using a single classifier for one modality. This is because of the high correlation between the scores of the same modality, as discussed later. The best balanced

accuracy in Table 1 is obtained by option 9, which is the fusion of EEG GB + ECG LR, although the 83.02% obtained is only slightly greater than the 82.55% of the single modality EEG + GB.

Table 1. Bias, probability of detection, probability of false alarm, and balanced accuracy of different options without bias correction. The best result without fusion and the best result with fusion are highlighted.

	−1 Modality −1 Method				−1 Modality −2 Methods (Fusion)		−2 Modalities −1 Method (Fusion)		−2 Modalities −2 Methods (Fusion)	
OPTION	1	2	3	4	5	6	7	8	9	10
Modality and Method	EEG GB	EEG LR	ECG GB	ECG LR	EEG GB + EEG LR	ECG GB + ECG LR	EEG GB + ECG GB	EEG LR + ECG LR	EEG GB + ECG LR	EEG LR + ECG GB
Bias class 0	0.067	0.044	0.070	0.028	0.055	0.049	0.068	0.036	0.047	0.057
Bias class 1	0.291	0.355	0.354	0.439	0.323	0.396	0.322	0.397	0.365	0.354
Prob. of Detection (%)	70.86	64.01	64.60	55.94	68.05	59.92	70.10	64.31	69.69	65.24
Prob. of f. alarm (%)	5.76	4.03	5.67	1.95	4.40	2.82	4.92	1.83	3.65	2.80
Balanced Accuracy (%)	82.55	79.99	79.47	76.99	81.83	78.55	82.59	81.24	83.02	81.22

Then, Table 2 shows the results corresponding to conditional bias correction. Notice that bias is clearly reduced in all options. The mean factor of reduction is 3 in class 0 and 2.5 in class 1. On the other hand, the balanced accuracy stays the same for the first four options as bias correction does not affect it when operating with individual detectors. However, the balanced accuracy increases in each of the six fusion options when compared with the corresponding option without bias correction. The best balanced accuracy is provided by option 10, which is very close to option 7. An improvement of 4.63% is achieved compared to the best case of Table 2.

To achieve statistical validation, we have repeated the previous experiments 100 times. In each repetition, a partition of the data set into two distinct halves was considered. In this way, we have obtained the mean and standard deviation of all the values shown in Tables 1 and 2. The mean values are similar to the data in Tables 1 and 2. The standard deviations never exceed 3% of the mean values.

We also performed significance tests on the balance accuracies obtained by bias correction in options 5 to 10, characterizing the distribution of the null hypothesis with the 100 balance accuracy values obtained with the corresponding option without bias correction. In all cases, the p -value was less than 0.05, so the null hypothesis is rejected, and we can consider that the balance accuracies from the bias correction are statistically significant.

Table 2. Bias, probability of detection, probability of false alarm, and balanced accuracy of different options with bias correction. The best result without fusion and the best result with fusion are highlighted.

	–1 Modality –1 Method				–1 Modality –2 Methods (Fusion)		–2 Modalities –1 Method (Fusion)		–2 Modalities –2 Methods (Fusion)	
OPTION	1	2	3	4	5	6	7	8	9	10
Modality and Method	EEGb GB	EEGb LR	ECGb GB	ECGb LR	EEGb GB + EEGb LR	ECGb GB + ECGb LR	EEGb GB + ECGb GB	EEGb LR + ECGb LR	EEGb GB + ECGb LR	EEGb LR + ECGb GB
Bias class 0	0.020	0.016	0.024	0.009	0.018	0.016	0.022	0.012	0.014	0.020
Bias class 1	0.104	0.143	0.150	0.205	0.124	0.178	0.127	0.174	0.155	0.146
Prob. of Detection (%)	70.86	64.01	64.60	55.94	73.84	71.85	82.27	75.95	78.41	81.10
Prob. of f. alarm (%)	5.76	4.03	5.67	1.95	6.17	5.44	7.90	4.44	5.77	6.84
Balanced Accuracy (%)	82.55	79.99	79.47	76.99	83.84	83.21	87.13	85.75	86.32	87.18

For better visualization, we have represented all the balanced accuracies from Tables 1 and 2 together in Figure 5. It is clear that the fusion, in general, improves with respect to the use of a single detector; furthermore, the fusion of modalities improve with respect to the fusion of methods and fusion with bias correction improves with respect to fusion without bias correction.

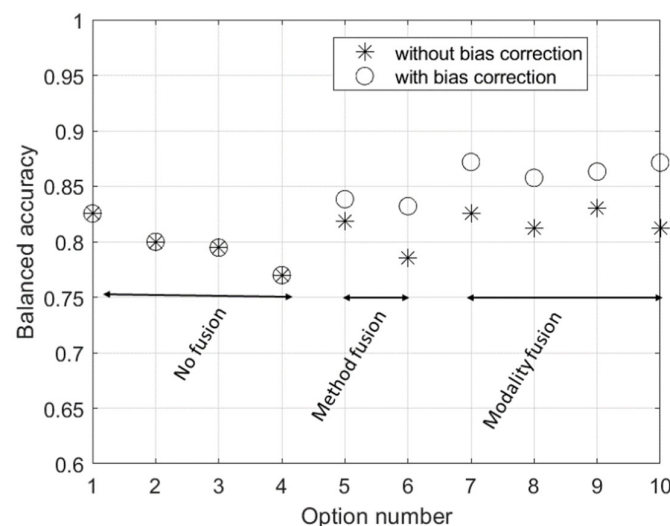


Figure 5. Balanced accuracy values for the 10 options from Table 1 (*) and Table 2 (O).

The improved results of modality fusion can be explained in terms of the correlation between the fused scores. Note that in the previous theoretical developments, as well as in the simulations, we assumed uncorrelated scores. However, in real-world application, a certain degree of correlation can be expected, which will result in some deterioration in performance. It is expected that the correlation level between scores from different

modalities should be lower than that from different methods applied to the same modality. This is because, if the modality is the same, the feature vector at the detector input will be the same regardless of the method used, resulting in a high correlation between the output scores. Changing the modality will reduce the output score correlation. This is shown in Table 3, where the correlation coefficients between scores for class 0 and class 1 are indicated. The minimum correlation coefficient in both classes corresponds to EEG-LR and ECG-GB, which coincides with the maximum balanced accuracy in Table 2.

Table 3. Correlation coefficients between the different sets of scores. The smallest correlation values are highlighted.

Class 0	EEG-GB	EEG-LR	ECG-GB	ECG-LR	Class 1	EEG-GB	EEG-LR	ECG-GB	ECG-LR
EEG-GB	1.00	0.61	0.32	0.26	EEG-GB	1.00	0.66	0.28	0.31
EEG-LR	0.61	1.00	0.20	0.30	EEG-LR	0.66	1.00	0.21	0.36
ECG-GB	0.32	0.20	1.00	0.44	ECG-GB	0.28	0.21	1.00	0.51
ECG-LR	0.26	0.30	0.44	1.00	ECG-LR	0.31	0.36	0.51	1.00

In the experiments with real data in this section, as well as in the analyses and simulations in Sections 3 and 4, we have always considered the MAP detector (threshold 0.5). This is convenient for potential extensions to the multi-class case, but it is not essential in the two-class case, where we could vary the threshold to adjust different operating points. In Table 4, we reproduce the balanced accuracy results obtained with different thresholds for the 10 options from Tables 1 and 2. A practical range of thresholds between 0.4 and 0.7 was selected, avoiding excessive false alarms for lower thresholds and low detectability for higher thresholds. We observe that the results maintain the improvement from bias correction and are consistent with what was observed in the previous results.

Table 4. Balanced accuracy of the different options for different thresholds, with bias correction (BC) and with no bias correction (NBC). The best result for each threshold is highlighted.

OPTION		−1 Modality −1 Method				−1 Modality −2 Methods (Fusion)		−2 Modalities −1 Method (Fusion)		−2 Modalities −2 Methods (Fusion)	
		1	2	3	4	5	6	7	8	9	10
Modality and Method		EEG GB	EEG LR	ECG GB	ECG LR	EEG GB + EEG LR	ECG GB + ECG LR	EEG GB + ECG GB	EEG LR + ECG LR	EEG GB + ECG LR	EEG LR + ECG GB
BA (%) threshold 0.40	BC	83.23	81.28	79.64	78.06	85.24	84.20	87.38	86.59	86.44	87.39
	NBC	83.23	81.28	79.64	78.06	83.56	81.99	86.96	84.32	85.91	86.59
BA (%) threshold 0.45	BC	82.92	80.50	79.68	77.28	84.62	83.64	87.26	86.93	86.44	87.35
	NBC	82.92	80.50	79.68	77.28	82.83	80.78	86.37	83.50	84.94	85.74
BA (%) threshold 0.50	BC	82.55	79.99	79.47	76.69	83.84	83.21	87.13	85.75	86.32	87.18
	NBC	82.55	79.99	79.47	76.69	81.83	78.55	82.59	81.24	83.02	81.22
BA (%) threshold 0.55	BC	82.26	79.71	79.26	76.46	83.20	82.34	86.67	84.67	85.82	86.49
	NBC	82.26	79.71	79.26	76.46	80.39	75.74	77.78	77.64	78.12	77.28
BA (%) threshold 0.60	BC	82.04	79.38	78.98	76.67	82.61	81.24	85.51	83.91	85.20	85.81
	NBC	82.04	79.38	78.98	75.67	79.56	74.86	76.55	75.35	76.03	75.40

Thus, bias corrections do not affect individual detectors, while fusion with bias correction always provides improvement compared to fusion without correction. It is also observed that the relative improvement between correction and no correction increases as

the threshold rises. This is explainable since the class with the highest bias was chosen as the class to detect, which means that corrections tend to raise the score values, with the impact of the correction becoming greater as the threshold increases.

6. Discussion

Let us discuss in terms of first principles. The optimum detector given s_1 and s_2 imply computation of the likelihood ratio $\frac{p_1(s_1, s_2)}{p_0(s_1, s_2)}$, which is to be compared with a given threshold. Unfortunately, most of the time, this only has theoretical interest because knowledge of the likelihood ratio is far from being realistic. Simple rules like computing the mean score (the one considered in this work) are much more useful yet reasonably grounded if the scores are considered as posterior probability of class $c = 1$. However, it seems tentative to incorporate in the fusion rule any knowledge about the conditional PDFs which could be learned during training. Notice that assuming independence it is $p_i(s_1, s_2) = p_i(s_1)p_i(s_2)$, and biases b_0 and b_1 are respective statistics of $p_0(s_1)$ and $p_1(s_1)$ that are estimated during training and incorporated in the proposed method to improve the mean rule. Essentially, the method helps to resolve the situation in which both detectors make different decisions and the mean is below the threshold. This is achieved by increasing the value of the fusion statistic (the sum in our case) by the difference between the bias of the class to be detected and the bias of the other class. By defining the class to be detected as the one with the greatest bias, the difference will always be positive, which will increase both the probability of detection and the probability of false alarm. However, it is expected that the increase in detection probability will be more significant than the increase in false alarm probability, due to the fact that class $c = 1$ has a greater bias, which implies uncorrected scores closer to the 0.5 threshold than those of class $c = 0$.

An interesting point to discuss is that the analyses and simulations in Sections 2–4 assumed independence between the scores to be fused. To overcome this limitation, we should incorporate a score joint probability density model. Unfortunately, this could make the necessary analysis and/or simulations intractable. Notice that the chosen dependence model would still be an approximation, not necessarily appropriate for a real-world problem. Recently, it has been formally shown that the existence of correlation is one of the factors that can reduce the effectiveness of the fusion [16], so, in this sense, we can consider that the conclusions of Sections 2–4 set an upper limit on the performance achievable with the proposed method. Notice that in the real-world application of Section 5, the scores to be fused exhibit relatively high correlation values (Table 3), yet the bias correction proved effective, especially when the correlation is lower, as expected. Clearly, it would be valuable to investigate methods that reduce the correlation between the scores to be fused. For example, if different classifiers are involved in the fusion, different training subsets could be used for each classifier. Similarly, subsets recorded at different times for different modalities could be used for training. In any case, it would be advisable to analyze whether the improvement due to decorrelation outweighs the worsening due to the reduction in the effective size because of the partition of the whole training set into smaller subsets. Another approximation to the dependence problem could be devising more complex fusion methods that account for the possible correlation to obtain optimum solutions [16,27].

A line of research that also deserves further investigation concerns extending the method to the multi-class case. This extension is not always obvious in the field of machine learning. Frequently, the extension is carried out by decomposing the multi-class problem into a number of two-class problems. This approach can also be applied here, with everything considered in Sections 2–4 being directly applicable. Obviously, a global multi-class approach is always desirable. To this end, note that in the multi-class problem, each classifier generates as many scores as there are classes, which must satisfy the constraint of

summing to 1. From training data, it is possible to determine the conditional biases of each of these scores, keeping in mind that the value to be estimated is 1 for the correct class and 0 for the rest of the classes. Corrections would be made, as in the two-class case, the class with the highest score is considered to be the “correct” one. One point to consider is that the corrected scores should sum to 1. We could, for example, correct all scores except 1, and adjust the uncorrected score so that they all sum to 1. In any case, extending the analyses in Sections 3 and 4 to the multi-class case is not straightforward and requires significant additional effort.

Other possible lines of interest involve making a “soft” correction, for example, proportional to the score, thus taking into account the reliability of the decision made by the individual detector on the correct class.

Finally, other real-world data applications could be topics of future research.

7. Conclusions

Regarding the conclusions, we have shown that conditional bias is always greater than zero and that bias correction improves accuracy. However, we face an ill-posed problem since, strictly speaking, conditional bias correction requires prior knowledge of the true class. Nevertheless, we have shown that corrections can be made to the scores provided by each individual detector before fusion, using its own detections as true. Conditions for improvements have been derived theoretically, first in general and then for the exponential distribution case. Interest has also been demonstrated for the versatile beta distribution case, resorting to Monte Carlo simulations. Beta distribution is particularly well-suited for normalized random variables and allows for very different shapes of the class conditional PDF. Finally, a real data case was considered. We observed that the proposed method provides the best results, despite the different (more realistic) conditions of the real data with respect to the simplifying assumptions of the prior analysis.

Author Contributions: Conceptualization, L.V. and A.S.; methodology, L.V. and A.S.; software, L.V. and A.S.; validation, A.S. and L.V.; formal analysis, L.V.; investigation, L.V. and A.S.; resources, A.S.; data curation, A.S.; writing—original draft preparation, L.V.; writing—review and editing, A.S. and L.V.; supervision, L.V.; project administration, L.V.; funding acquisition, L.V. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Generalitat Valenciana, Spain under Grant CIPROM/2022/20 and Agencia Estatal de Investigación, Spain under Grant PID2024-161353OB-I00.

Data Availability Statement: The data presented in this study are available on request from the corresponding authors.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Let us first calculate the *PD* and *PFA* corresponding to the individual classifiers

$$\begin{aligned}
 PD = Pr(s_m > 0.5/c = 1) &= \int_{0.5}^1 \frac{\lambda_1}{1 - e^{-\lambda_1}} e^{-\lambda_1(1-s_m)} ds_m = \frac{\lambda_1 e^{-\lambda_1}}{1 - e^{-\lambda_1}} \left(\frac{e^{\lambda_1 s_m}}{\lambda_1} \Big|_{0.5}^1 \right) = \\
 &= \frac{e^{-\lambda_1}}{1 - e^{-\lambda_1}} (e^{\lambda_1} - e^{\lambda_1 0.5}) = \frac{1 - e^{-\lambda_1 0.5}}{1 - e^{-\lambda_1}}
 \end{aligned} \tag{A1}$$

$$\begin{aligned}
 PFA &= Pr(s_m > 0.5/c = 0) = \int_{0.5}^1 \frac{\lambda_0}{1 - e^{-\lambda_0}} e^{-\lambda_1 s_m} ds_m = \frac{\lambda_0}{1 - e^{-\lambda_1}} \left(\frac{e^{-\lambda_0 s_m}}{-\lambda_0} \Big|_{0.5}^1 \right) = \\
 &= \frac{1}{1 - e^{-\lambda_0}} (e^{-\lambda_0 0.5} - e^{-\lambda_0}) = e^{-\lambda_0 0.5} \frac{1 - e^{-\lambda_0 0.5}}{1 - e^{-\lambda_0}}
 \end{aligned} \tag{A2}$$

Now, let us calculate PD_z and PFA_z . We have to compute the integral of the corresponding bivariate PDF in the upper triangle of Figure 2, delimited below by the straight line $s_2 = 1 - s_1$, hence

$$\begin{aligned}
 PD_z &= Pr(s_1 + s_2 > 1/c = 1) = \int_0^1 \int_{1-s_1}^1 \frac{\lambda_1}{1 - e^{-\lambda_1}} e^{-\lambda_1(1-s_1)} \frac{\lambda_1}{1 - e^{-\lambda_1}} e^{-\lambda_1(1-s_2)} ds_1 ds_2 = \left(\frac{\lambda_1 e^{-\lambda_1}}{1 - e^{-\lambda_1}} \right)^2 \int_0^1 e^{\lambda_1 s_1} \left(\int_{1-s_1}^1 e^{\lambda_1 s_2} ds_2 \right) ds_1 = \\
 &= \left(\frac{\lambda_1 e^{-\lambda_1}}{1 - e^{-\lambda_1}} \right)^2 \int_0^1 e^{\lambda_1 s_1} \left(\frac{e^{\lambda_1 s_2}}{\lambda_1} \Big|_{1-s_1}^1 \right) ds_1 = \left(\frac{\lambda_1 e^{-\lambda_1}}{1 - e^{-\lambda_1}} \right)^2 \frac{1}{\lambda_1} \int_0^1 e^{\lambda_1 s_1} (e^{\lambda_1} - e^{\lambda_1(1-s_1)}) ds_1 = \\
 &= \left(\frac{\lambda_1 e^{-\lambda_1}}{1 - e^{-\lambda_1}} \right)^2 \frac{1}{\lambda_1} e^{\lambda_1} \int_0^1 (e^{\lambda_1 s_1} - 1) ds_1 = \frac{\lambda_1 e^{-\lambda_1}}{(1 - e^{-\lambda_1})^2} \left(\left(\frac{e^{\lambda_1 s_1}}{\lambda_1} \Big|_0^1 \right) - 1 \right) = \frac{\lambda_1 e^{-\lambda_1}}{(1 - e^{-\lambda_1})^2} \left(\frac{e^{\lambda_1} - 1}{\lambda_1} - 1 \right) = \frac{(1 - e^{-\lambda_1}(1 + \lambda_1))}{(1 - e^{-\lambda_1})^2},
 \end{aligned} \tag{A3}$$

and

$$\begin{aligned}
 PFA_z &= Pr(s_1 + s_2 > 1/c = 0) = \int_0^1 \int_{1-s_1}^1 \frac{\lambda_0}{1 - e^{-\lambda_0}} e^{-\lambda_0 s_1} \frac{\lambda_0}{1 - e^{-\lambda_0}} e^{-\lambda_0 s_2} ds_1 ds_2 = \left(\frac{\lambda_0}{1 - e^{-\lambda_0}} \right)^2 \int_0^1 e^{-\lambda_0 s_1} \left(\int_{1-s_1}^1 e^{-\lambda_0 s_2} ds_2 \right) ds_1 = \\
 &= \left(\frac{\lambda_0}{1 - e^{-\lambda_0}} \right)^2 \int_0^1 e^{-\lambda_0 s_1} \left(\frac{e^{-\lambda_0 s_2}}{-\lambda_0} \Big|_{1-s_1}^1 \right) ds_1 = \frac{\lambda_0}{(1 - e^{-\lambda_0})^2} \int_0^1 e^{-\lambda_0 s_1} (e^{-\lambda_0} - e^{-\lambda_0(1-s_1)}) ds_1 = \\
 &= \frac{\lambda_0 e^{-\lambda_0}}{(1 - e^{-\lambda_0})^2} \int_0^1 (e^{-\lambda_0 s_1} - 1) ds_1 = \frac{\lambda_0 e^{-\lambda_0}}{(1 - e^{-\lambda_0})^2} \left(\left(\frac{e^{-\lambda_0 s_1}}{-\lambda_0} \Big|_0^1 \right) - 1 \right) = \frac{\lambda_0 e^{-\lambda_0}}{(1 - e^{-\lambda_0})^2} \left(\frac{1 - e^{-\lambda_0}}{\lambda_0} - 1 \right) = \frac{e^{-\lambda_0} (1 - e^{-\lambda_0} - \lambda_0)}{(1 - e^{-\lambda_0})^2}
 \end{aligned} \tag{A4}$$

Finally, from (9), we can calculate PD_{z^b} and PFA_{z^b} . Assuming for the moment that $\Delta < 0.5$, we can write

$$PD_{z^b} = PD_z + 2 \left(\underbrace{\int_0^{0.5-\Delta} \int_{1-s_1-\Delta}^{1-s_1} p_1(s_1) p_1(s_2) ds_1 ds_2}_{v_{PDz}} + \underbrace{\int_{0.5-\Delta}^{0.5} \int_{0.5}^{1-s_1} p_1(s_1) p_1(s_2) ds_1 ds_2}_{\eta_{PDz}} \right), \tag{A5}$$

where

$$\begin{aligned}
 v_{PDz} &= \int_0^{0.5-\Delta} \int_{1-s_1-\Delta}^{1-s_1} \frac{\lambda_1}{1 - e^{-\lambda_1}} e^{-\lambda_1(1-s_1)} \frac{\lambda_1}{1 - e^{-\lambda_1}} e^{-\lambda_1(1-s_2)} ds_1 ds_2 = \left(\frac{\lambda_1 e^{-\lambda_1}}{1 - e^{-\lambda_1}} \right)^2 \int_0^{0.5-\Delta} e^{\lambda_1 s_1} \left(\int_{1-s_1-\Delta}^{1-s_1} e^{\lambda_1 s_2} ds_2 \right) ds_1 = \\
 &= \left(\frac{\lambda_1 e^{-\lambda_1}}{1 - e^{-\lambda_1}} \right)^2 \int_0^{0.5-\Delta} e^{\lambda_1 s_1} \left(\frac{e^{\lambda_1 s_2}}{\lambda_1} \Big|_{1-s_1-\Delta}^{1-s_1} \right) ds_1 = \left(\frac{\lambda_1 e^{-\lambda_1}}{1 - e^{-\lambda_1}} \right)^2 \frac{1}{\lambda_1} \int_0^{0.5-\Delta} e^{\lambda_1 s_1} (e^{\lambda_1(1-s_1)} - e^{\lambda_1(1-s_1-\Delta)}) ds_1 = \\
 &= \left(\frac{\lambda_1 e^{-\lambda_1}}{1 - e^{-\lambda_1}} \right)^2 \frac{1}{\lambda_1} e^{\lambda_1} \int_0^{0.5-\Delta} (1 - e^{-\lambda_1 \Delta}) ds_1 = \frac{\lambda_1 e^{-\lambda_1}}{(1 - e^{-\lambda_1})^2} (1 - e^{-\lambda_1 \Delta}) (0.5 - \Delta),
 \end{aligned} \tag{A6}$$

$$\begin{aligned}
 \eta_{PDz} &= \int_{0.5-\Delta}^{0.5} \int_{0.5}^{1-s_1} \frac{\lambda_1}{1 - e^{-\lambda_1}} e^{-\lambda_1(1-s_1)} \frac{\lambda_1}{1 - e^{-\lambda_1}} e^{-\lambda_1(1-s_2)} ds_1 ds_2 = \left(\frac{\lambda_1 e^{-\lambda_1}}{1 - e^{-\lambda_1}} \right)^2 \int_{0.5-\Delta}^{0.5} e^{\lambda_1 s_1} \left(\int_{0.5}^{1-s_1} e^{\lambda_1 s_2} ds_2 \right) ds_1 = \\
 &= \left(\frac{\lambda_1 e^{-\lambda_1}}{1 - e^{-\lambda_1}} \right)^2 \int_{0.5-\Delta}^{0.5} e^{\lambda_1 s_1} \left(\frac{e^{\lambda_1 s_2}}{\lambda_1} \Big|_{0.5}^{1-s_1} \right) ds_1 = \left(\frac{\lambda_1 e^{-\lambda_1}}{1 - e^{-\lambda_1}} \right)^2 \frac{1}{\lambda_1} \int_{0.5-\Delta}^{0.5} e^{\lambda_1 s_1} (e^{\lambda_1(1-s_1)} - e^{\lambda_1 0.5}) ds_1 = \\
 &= \left(\frac{\lambda_1 e^{-\lambda_1}}{1 - e^{-\lambda_1}} \right)^2 \frac{e^{\lambda_1}}{\lambda_1} \int_{0.5-\Delta}^{0.5} (1 - e^{\lambda_1(s_1-0.5)}) ds_1 = \frac{\lambda_1 e^{-\lambda_1}}{(1 - e^{-\lambda_1})^2} \left(\Delta - \left(\frac{e^{\lambda_1(s_1-0.5)}}{\lambda_1} \Big|_{0.5-\Delta}^{0.5} \right) \right) = \\
 &= \frac{\lambda_1 e^{-\lambda_1}}{(1 - e^{-\lambda_1})^2} \left(\Delta - \frac{1}{\lambda_1} (1 - e^{-\lambda_1 \Delta}) \right)
 \end{aligned} \tag{A7}$$

and

$$PFA_{zb} = PFA_z + 2 \left(\underbrace{\int_0^{0.5-\Delta} \int_{1-s_1-\Delta}^{1-s_1} p_0(s_1) p_0(s_2) ds_1 ds_2}_{v_{PFAz}} + \underbrace{\int_{0.5-\Delta}^{0.5} \int_{0.5}^{1-s_1} p_0(s_1) p_0(s_2) ds_1 ds_2}_{\eta_{PFAz}} \right). \quad (A8)$$

where

$$\begin{aligned} v_{PFAz} &= \int_0^{0.5-\Delta} \int_{1-s_1-\Delta}^{1-s_1} \frac{\lambda_0}{1-e^{-\lambda_0}} e^{-\lambda_0 s_1} \frac{\lambda_0}{1-e^{-\lambda_0}} e^{-\lambda_0 s_2} ds_1 ds_2 = \left(\frac{\lambda_0}{1-e^{-\lambda_0}} \right)^2 \int_0^{0.5-\Delta} e^{-\lambda_0 s_1} \left(\int_{1-s_1-\Delta}^{1-s_1} e^{-\lambda_0 s_2} ds_2 \right) ds_1 = \\ &= \left(\frac{\lambda_0}{1-e^{-\lambda_0}} \right)^2 \int_0^{0.5-\Delta} e^{-\lambda_0 s_1} \left(\frac{e^{-\lambda_0 s_2}}{-\lambda_0} \Big|_{1-s_1-\Delta}^{1-s_1} \right) ds_1 = \left(\frac{\lambda_0}{1-e^{-\lambda_0}} \right)^2 \frac{1}{\lambda_0} \int_0^{0.5-\Delta} e^{-\lambda_0 s_1} \left(e^{-\lambda_0(1-s_1-\Delta)} - e^{-\lambda_0(1-s_1)} \right) ds_1 = \\ &= \left(\frac{\lambda_0}{1-e^{-\lambda_0}} \right)^2 \frac{e^{-\lambda_0}}{\lambda_0} \int_0^{0.5-\Delta} (e^{\lambda_0 \Delta} - 1) ds_1 = \frac{\lambda_0 e^{-\lambda_0}}{(1-e^{-\lambda_0})^2} (e^{\lambda_0 \Delta} - 1) (0.5 - \Delta). \end{aligned} \quad (A9)$$

$$\begin{aligned} \eta_{PFAz} &= \int_{0.5-\Delta}^{0.5} \int_{0.5}^{1-s_1} \frac{\lambda_0}{1-e^{-\lambda_0}} e^{-\lambda_0 s_1} \frac{\lambda_0}{1-e^{-\lambda_0}} e^{-\lambda_0 s_2} ds_1 ds_2 = \left(\frac{\lambda_0}{1-e^{-\lambda_0}} \right)^2 \int_{0.5-\Delta}^{0.5} e^{-\lambda_0 s_1} \left(\int_{0.5}^{1-s_1} e^{-\lambda_0 s_2} ds_2 \right) ds_1 = \\ &= \left(\frac{\lambda_0}{1-e^{-\lambda_0}} \right)^2 \int_{0.5-\Delta}^{0.5} e^{-\lambda_0 s_1} \left(\frac{e^{-\lambda_0 s_2}}{-\lambda_0} \Big|_{0.5}^{1-s_1} \right) ds_1 = \left(\frac{\lambda_0}{1-e^{-\lambda_0}} \right)^2 \frac{1}{\lambda_0} \int_{0.5-\Delta}^{0.5} e^{-\lambda_0 s_1} (e^{-\lambda_0 0.5} - e^{-\lambda_0(1-s_1)}) ds_1 = \\ &= \left(\frac{\lambda_0}{1-e^{-\lambda_0}} \right)^2 \frac{e^{-\lambda_0}}{\lambda_0} \int_{0.5-\Delta}^{0.5} (e^{-\lambda_0(s_1-0.5)} - 1) ds_1 = \frac{\lambda_0 e^{-\lambda_0}}{(1-e^{-\lambda_0})^2} \left(\left(\frac{e^{-\lambda_0(s_1-0.5)}}{-\lambda_0} \Big|_{0.5-\Delta}^{0.5} \right) - \Delta \right) = \\ &= \frac{\lambda_0 e^{-\lambda_0}}{(1-e^{-\lambda_0})^2} \left(\frac{1}{\lambda_0} (e^{\lambda_0 \Delta} - 1) - \Delta \right). \end{aligned} \quad (A10)$$

Finally, for $\Delta \geq 0.5$, v_{PDz} and v_{PFAz} vanish and η_{PDz} and η_{PFAz} saturates to

$$\begin{aligned} \eta_{PDz} &= \int_0^{0.5} \int_{0.5}^{1-s_1} \frac{\lambda_1}{1-e^{-\lambda_1}} e^{-\lambda_1(1-s_1)} \frac{\lambda_1}{1-e^{-\lambda_1}} e^{-\lambda_1(1-s_2)} ds_1 ds_2 = \frac{\lambda_1 e^{-\lambda_1}}{(1-e^{-\lambda_1})^2} \left(0.5 - \frac{1}{\lambda_1} (1 - e^{-\lambda_1 0.5}) \right) \\ \eta_{PFAz} &= \int_0^{0.5} \int_{0.5}^{1-s_1} \frac{\lambda_0}{1-e^{-\lambda_0}} e^{-\lambda_0 s_1} \frac{\lambda_0}{1-e^{-\lambda_0}} e^{-\lambda_0 s_2} ds_1 ds_2 = \frac{\lambda_0 e^{-\lambda_0}}{(1-e^{-\lambda_0})^2} \left(\frac{1}{\lambda_0} (e^{\lambda_0 0.5} - 1) - 0.5 \right). \end{aligned} \quad (A11)$$

References

1. Ruta, D.; Gabrys, B. An overview of classifier fusion methods. *Comput. Inf. Syst.* **2000**, *7*, 1–10.
2. Tulyakov, S.; Jaeger, S.; Govindaraju, V.; Doermann, D. Review of classifier combination methods. *Stud. Comput. Intell.* **2008**, *90*, 361–386.
3. Ross, A.; Nandakumar, K. *Fusion, Score-Level, Encyclopedia of Biometrics*; Springer: New York, NY, USA, 2009; pp. 611–616.
4. Mohandes, M.; Deriche, M.; Aliyu, S. Classifiers Combination Techniques: A Comprehensive Review. *IEEE Access* **2018**, *6*, 19626–19639. [\[CrossRef\]](#)
5. Heydarian, H.; Adam, M.T.P.; Burrows, T.L.; Rollo, M.E. Exploring Score-Level and Decision-Level Fusion of Inertial and Video Data for Intake Gesture Detection. *IEEE Access* **2015**, *13*, 643–655. [\[CrossRef\]](#)
6. Mi, A.; Wang, L.; Qi, J. A Multiple Classifier Fusion Algorithm Using Weighted Decision Templates. *Sci. Program.* **2016**, *3*, 3943859. [\[CrossRef\]](#)
7. Pereira, L.M.; Salazar, A.; Vergara, L. A comparative analysis of early and late fusion for the multimodal two-class problem. *IEEE Access* **2023**, *11*, 84283–84300. [\[CrossRef\]](#)
8. Atrey, P.K.; Hossain, M.A.; El Saddik, A.; Kankanhalli, M.S. Multimodal fusion for multimedia analysis: A survey. *Multimed. Syst.* **2010**, *16*, 345–379. [\[CrossRef\]](#)
9. Lahat, D.; Adali, T.; Jutten, C. Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects. *Proc. IEEE* **2015**, *103*, 1449–1477. [\[CrossRef\]](#)
10. Pawlowski, M.; Wróblewska, A.; Sysko-Romaczuk, S. Effective Techniques for Multimodal Data Fusion: A Comparative Analysis. *Sensors* **2023**, *23*, 2381. [\[CrossRef\]](#)

11. Barua, A.; Ahmed, M.U.; Begum, S. A Systematic Literature Review on Multimodal Machine Learning: Applications, Challenges, Gaps and Future Directions. *IEEE Access* **2023**, *11*, 14804–14831. [\[CrossRef\]](#)
12. Hanzelik, P.P.; Kummer, A.; Abonyi, J. Data Reconciliation-Based Hierarchical Fusion of Machine Learning Models. *Mach. Learn. Knowl. Extr.* **2024**, *6*, 2601–2617. [\[CrossRef\]](#)
13. Tiwari, A.; Shukla, R.; Tiwari, S. Alzheimer’s Disease Detection from Fused PET and MRI Modalities Using an Ensemble Classifier. *Mach. Learn. Knowl. Extr.* **2023**, *5*, 512–538. [\[CrossRef\]](#)
14. Zhang, L.; Li, T.; Cui, H.; Zhang, Q.; Jiang, Z.; Li, J.; Welsch, R.E.; Jia, Z. A Novel Prediction Model for Multimodal Medical Data Based on Graph Neural Networks. *Mach. Learn. Knowl. Extr.* **2025**, *7*, 92. [\[CrossRef\]](#)
15. Arrowsmith, J.; Susnjak, T.; Jang-Jaccard, J. Multimodal Deep Learning for Android Malware Classification. *Mach. Learn. Knowl. Extr.* **2025**, *7*, 23. [\[CrossRef\]](#)
16. Vergara, L.; Salazar, A. On the Optimum Linear Soft Fusion of Classifiers. *Appl. Sci.* **2025**, *15*, 5038. [\[CrossRef\]](#)
17. Jobert, M.; Shulz, H.; Jahng, P.; Tismer, C.; Bes, F.; Escola, H. A computerized method for detecting episodes of wakefulness during sleep based on the Alpha slow-wave index (ASI). *Sleep* **1994**, *17*, 37–46. [\[CrossRef\]](#)
18. Yazdi, M.; Samaee, M.; Massicotte, D. A Review on Automated Sleep Study. *Ann. Biomed. Eng.* **2024**, *52*, 1463–1491. [\[CrossRef\]](#)
19. Heneghan, C. St. Vincent’s University Hospital/University College Dublin, Sleep Apnea Database. 2011. Available online: <https://physionet.org/content/ucddb/1.0.0/> (accessed on 20 November 2025).
20. Hjorth, B. EEG analysis based on time domain properties. *Electroencephalogr. Clin. Neurophysiol.* **1970**, *29*, 306–310. [\[CrossRef\]](#)
21. Salazar, A.; Vergara, L.; Miralles, R. On including sequential dependence in ICA mixture models. *Signal Process.* **2010**, *90*, 2314–2318. [\[CrossRef\]](#)
22. Motamedi-Fakhr, S.; Moshrefi-Torbati, M.; Hill, M.; Hill, C.; White, P. Signal processing techniques applied to human sleep EEG signals—A review. *Biomedical. Signal Process. Control* **2014**, *10*, 21–33. [\[CrossRef\]](#)
23. Li, T.; Zhou, M. ECG Classification Using Wavelet Packet Entropy and Random Forests. *Entropy* **2016**, *18*, 285. [\[CrossRef\]](#)
24. Sun, H.; Jia, J.; Goparaju, B.; Huang, G.; Sourina, O.; Matt Travis, M.; Westover, B. Large-Scale Automated Sleep Staging. *Sleep* **2017**, *40*, zsx139. [\[CrossRef\]](#)
25. Franz Ehrlich, J.B. Automatic Sleep Arousal Detection Using Heart Rate from a Single-Lead. In Proceedings of the 2022 Computing in Cardiology (CinC), Tampere, Finland, 4–7 September 2022.
26. Clifford, G.; Azuaje, F.; McSharry, P. *Advanced Methods and Tools for ECG Data Analysis*; Artech House: Norwood, MA, USA, 2006.
27. Salazar, A.; Safont, G.; Vergara, L.; Vidal, E. Graph Regularization Methods in Soft Detector Fusion. *IEEE Access* **2023**, *11*, 144747–144759. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.