Research Article | Vol. 12, No. 3 / 1 March 2021 / Biomedical Optics Express 1635

Biomedical Optics EXPRESS

Multimodal fusion of EEG-fNIRS: a mutual information-based hybrid classification framework

ROOHOLLAH JAFARI DELIGANI,[1] SEYYED BAHRAM BORGHEAI,[1] JOHN MCLINDEN,[1] AND YALDA SHAHRIARI[1,2,*]

[1] Department of Electrical, Computer and Biomedical Engineering; University of Rhode Island, Kingston, RI 02881, USA
[2] Interdisciplinary Neuroscience Program; University of Rhode Island, Kingston, RI 02881, USA
[*] yalda_shahriari@uri.edu

Abstract: Multimodal data fusion is one of the current primary neuroimaging research directions to overcome the fundamental limitations of individual modalities by exploiting complementary information from different modalities. Electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRS) are especially compelling modalities due to their potentially complementary features reflecting the electro-hemodynamic characteristics of neural responses. However, the current multimodal studies lack a comprehensive systematic approach to properly merge the complementary features from their multimodal data. Identifying a systematic approach to properly fuse EEG-fNIRS data and exploit their complementary potential is crucial in improving performance. This paper proposes a framework for classifying fused EEG-fNIRS data at the feature level, relying on a mutual information-based feature selection approach with respect to the complementarity between features. The goal is to optimize the complementarity, redundancy and relevance between multimodal features with respect to the class labels as belonging to a pathological condition or healthy control. Nine amyotrophic lateral sclerosis (ALS) patients and nine controls underwent multimodal data recording during a visuo-mental task. Multiple spectral and temporal features were extracted and fed to a feature selection algorithm followed by a classifier, which selected the optimized subset of features through a cross-validation process. The results demonstrated considerably improved hybrid classification performance compared to the individual modalities and compared to conventional classification without feature selection, suggesting a potential efficacy of our proposed framework for wider neuro-clinical applications.

© 2021 Optical Society of America under the terms of the OSA Open Access Publishing Agreement

1. Introduction

Numerous mathematical tools and computational methods have been utilized to combine data from different modalities efficiently and obtain a criterion that optimally selects the best fused features from these different modalities. These fusion methods are especially useful in neuro-clinical studies to support more accurate decoding of neural information, and thus, improve the performance of relevant applications. These algorithms have shown promising applications in various fields, including brain-computer interfaces (BCIs) [1,2], neuro-pathological diagnosis [3,4], and neural source localization [5].

So far, many fusion frameworks have exploited the common and complementary properties of different types of neuroimaging data, including electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRS). These modalities are both portable scalp located devices that can be easily employed for data acquisition in multiple populations of patients with neurological impairments. Considering the first modality, EEG captures macroscopic cortical dynamics with relatively fine temporal resolution (~5 msec). Although EEG classification has been widely investigated to detect and extract underlying pathological neural signatures, outcomes remain

#413666 https://doi.org/10.1364/BOE.413666
Journal © 2021 Received 28 Oct 2020; revised 15 Feb 2021; accepted 16 Feb 2021; published 26 Feb 2021

Biomedical Optics EXPRESS

poor for multiple reasons, including low signal-to-noise ratio (SNR), poor spatial resolution, and insufficient classifiable measurements which cannot be addressed easily by existing computational algorithms [6]. One way to overcome these drawbacks is combining EEG with other modalities in an integrated framework that can provide a complimentary basis for the more accurate and robust detection of neural signatures to improve classification performance. For this purpose, fNIRS has shown promising capacity in improving classification performance [7–10] as a modality for measuring the underlying hemodynamic properties with higher spatial resolution (~ 1 cm) than EEG. The integration of EEG and fNIRS provides us with two different types of neural data associated with the same regional neural activities, each one reflecting the underlying changes as potentially different sources of information. Exploiting the complementary features of the two data modalities with proper fusion algorithms to achieve a higher classification accuracy for hybrid EEG-fNIRS measures than for single modality approaches can provide a basis for improving performance in many existing neuro-assisted applications ranging from BCI to improving diagnostic methods for neurological impairments.

Fusion frameworks for EEG-fNIRS classification can be broadly classified into two categories based on the level at which the combination takes place. The first category is decision-level, in which the features are separately fed to a classifier, and the outcome is used in a feedback loop to optimize accuracy. For example, in a motor imagery study conducted by Fazli et al. [11], three groups of features, specifically EEG band-power, oxy-, and deoxy-hemoglobin (HbO2 and HbR respectively) were separately classified, and then a meta classifier optimally combined the three classifier outputs in a feedback loop based on the global peak cross-validation accuracy of each classifier. Putze et al. [12] used a similar framework to classify auditory and visual perception using hybrid EEG-fNIRS spectral and temporal features. Both studies achieved an average of 5% improved classification accuracy over single modality classification. In another study [13], the authors used decision level fusion to combine the outputs of two local support vector machine (SVM) classifiers, one for EEG signals and the other for fNIRS signals in which each classifier was calibrated based on the optimal operating points of the EEG and fNIRS receiver operating characteristic (ROC) curves. At the end, both outputs were fed to a global classifier, which improved classification accuracy by 7.76% compared to single modal approach. A similar study for classifying mental work achieved 6% improvement compared to single modal data [14]. Another decision-level hybrid classification criterion is the fuzzy fusion-based approach, as was done in [15] to integrate the temporal and spectral features of EEG for motor imagery classification. After employing traditional classification methods, the authors adopted Choquet and Sugeno integrals to consider possible interactions between the obtained outputs from the different classifiers by fusing their posterior probabilities. They achieved ~7% improvement compared to conventional classifiers including linear discriminant analysis (LDA).

The second category is feature-level fusion in which features are concatenated, transformed, or optimally selected before training the classifier. Work on the simple concatenation of EEG-fNIRS features has shown a modest improvement compared to that obtained with a single modality, which is likely caused by the lack of comprehensive computational approaches for a proper feature integration that exploit the complementarity between each modality's unique properties as a preferred alternative over feature concatenation [6]. For example, in another study conducted by Buccino et al. [16], EEG-fNIRS features were integrated through concatenation without any feature fusion strategy. In this study, the authors reported that the feature set was small, had no imposed computational load on the classification, and reached a 2% accuracy improvement compared to features from a single modality. In a study by Nguyen et al. [17], driver drowsiness during long-term simulated driving classified using concatenated EEG-fNIRS features yielded an average 5.5% accuracy improvement using combined classification compared with single modal features. Another modest improvement of 1% using hybrid classification was achieved by concatenating EEG and fNIRS features for distinguishing Parkinson's disease [18]. Feature-level

Biomedical Optics EXPRESS

EEG-fNIRS fusion has also been done by projecting the original feature set to a new feature space to provide better separability than the original feature set. These projection methods are known as feature extraction methods, and their main disadvantage is that the newly created feature space is difficult to interpret and may not have a clear physical meaning [19]. In a study conducted by Saadati et al. [20], the authors extracted temporal and spectral features from EEG-fNIRS data and then used a convolutional neural network (CNN) to pass the features through different layers of the network and change the dimensions in a deep learning process for classifying mental workload from EEG-fNIRS data, which improved classification by 7%. Other transformation approaches have used a specific criterion for projecting the feature set into a new space. For example, in a study of mental stress assessment [21], the temporal properties of EEG have been combined with the spatial properties of fNIRS by transforming their signals to a mixed model, respectively using temporal and spatial independent component analysis (ICA), achieving a 3.4% accuracy improvement. In another study [22], the authors used a joint sparse canonical correlation analysis (CCA) to jointly estimate multiple pairs of canonical vectors to fuse EEG-fNIRS features and then fed these features to a SVM classifier, which significantly improved the hybrid classification accuracy by 5%. In a similar study on mental stress assessment [23], a CCA was used, to project two different feature sets into a space with maximum correlation across two sets. The authors reported that by using this criterion, the redundant information has been reduced, and they obtained a 7.9% accuracy improvement. As the last category of feature-level fusion frameworks, feature selection algorithms have been used to optimally select a subset of features from the original combined feature set based on a criterion that maximizes classification performance. Depending on whether the classifier is included in the selection process, feature selection methods can be grouped into wrapper and filter methods [24]. While wrapper methods generally consider classification performance as the feature selection criterion, filter methods select an optimized feature set independent of the classification algorithm. Thus, in filter methods, the biases of the feature selection procedure do not interfere with the learning algorithm–this results in improved generalization capability for the classifier. One example of feature selection is the method used by Lin et al. [25], who conducted correlation analysis as the selection criterion between EEG and fNIRS channels (features in their study) and selected the most correlated channels, which yielded a 9% sensitivity improvement compared to single modalities. In the use of conventional classification algorithms, fused feature selection is a fundamental difficulty given a large number of possible features and the often small amount of available data. Furthermore, as the number of samples in real-world EEG-fNIRS recordings is relatively small, avoiding underfitting or overfitting is a primary challenge [20]. The existence of redundant information in the original feature space can also hinder classification performance [26] since a system that memorizes training data involving redundancy can achieve perfect training performance while completely failing to generalize to new data.

The mutual information criterion is a powerful mathematical tool for feature selection, which can minimize the redundancy between features (i.e. the joint entropy of features subtracted from the individual entropies of the features). Yin et al. [26] used this criterion to decode the force and speed of hand clenching. In this study, the authors used band-power, amplitude, phase, and frequency to construct time-phase-frequency EEG features, and the differences between HbO2 and HbR were extracted as fNIRS features. They used a feature optimization method based on joint mutual information to remove redundant information that may reduce classification accuracy. This combination of EEG-fNIRS features resulted in improved performance (up to a 5% increase). In addition to minimizing redundancy, maximizing the relevance of a feature set to the discrete output of the classifier can significantly increase classification performance [27]. Another important contributing factor to improving classification performance is maximizing the complementarity between features obtained from multimodal data. This property has been defined as a combination of features that can return more information on the output class than the

Biomedical Optics EXPRESS

sum of the information returned by each of the features taken individually [27]. This advantage has special importance while fusing two different modalities with unique complementary properties, which can be efficiently exploited to improve classification performance. The mutual information criterion has also been adopted for combining other modalities in the literature. The authors in [28] minimized the conditional entropy between EEG and magnetoencephalography (MEG) features to reduce the degree of redundancy or similarity between the two signals for optimal estimation of the parameters to model localized sources. In another study [29], the authors used EEG and functional magnetic resonance imaging (fMRI) data in a hybrid source activation model by minimizing the mutual information to maximize the independence for joint ICA analysis. In another study [30], the authors used EEG and electrocardiography (ECG) data to classify mental workload. In this study, the authors first extracted features from both modalities and then used a criterion called co-information to maximize the mutual information between the output labels and the integrated feature subset. The authors reported that their proposed fusion method could increase the classification accuracy indicating their multi-modal fusion approach is promising to identify mental workload.

To date, EEG-fNIRS multimodal approaches have shown a considerable capacity to improve classification performance by measuring two different brain functions. However, they suffer from a lack of strong computational methods to systematically and optimally integrate the features. Computational integration methods should be developed that consider the differential characteristics of features from multimodal EEG-fNIRS signals. It is anticipated that efforts towards optimizing multimodal integration of EEG and fNIRS can make substantial advancement to the existing brain measurement packages with improved performance compared to EEG or fNIRS modalities alone.

In this paper, a mutual information-based feature selection algorithm was adopted to propose a classification framework for multimodal EEG-fNIRS data. This study is the first that systematically exploits the complementarity aspect of such multimodal fused features through a feature selection algorithm that quantifies the complementarity between features and selects the optimal fused subset towards improving the classification performance. In this algorithm, the optimal features from a fused set of EEG-fNIRS features were determined with respect to minimized redundancy between features, maximized relevance, and maximized complementarity between features and class labels. EEG and fNIRS data were recorded from healthy participants and participants with ALS during a visuo-mental paradigm and were used to distinguish between the two aforementioned groups as a two-class problem. Features were first extracted from each modality and then the optimized subset of features was selected from the original combined set of EEG and fNIRS features through the aforementioned mutual information-based algorithm. This process was repeated for each modality (i.e., EEG and fNIRS) separately to evaluate the classification performance's improvement due to the integration of features compared to those obtained from each single modality. Finally, the selected optimal feature sets from each individual modality and from the two modalities combined were fed into a support vector machine (SVM) classifier in which the hyper-parameter was the adequate number of features that was chosen according to the best classification results.

2. Materials and methods

2.1. Participants

A total of 18 subjects were recruited and assigned to two groups: Nine individuals with ALS (ALS: 7 males, average age 56.8 years old) with ALS revised Functional Rating Scale (ALSFRS-R) scores of 0, 4, 4, 23, 22, 39, 41, 33, 26, respectively for subjects 1 to 9 (mean: 21.3 ± 15.5) on a 48-point scale and nine age-matched healthy controls (HC: 4 males, average age 60.7 years old). All the protocols in this study were approved by the Institutional Review Board (IRB) of the University of Rhode Island (URI) and written informed consent was provided directly by each

Biomedical Optics EXPRESS

subject or by each patient's caregiver. Age-matched control subjects had no reported history of visual, mental, or substance-related disorders that could potentially affect the results or their performance during data recording.

2.2. Experimental protocol

Subjects participated in two sessions, each consisting of one run with 14 trials. The participants were asked to perform a visuo-mental paradigm based on the conventional visual oddball paradigm followed by a mathematical task, as fully described in our previous work [31] and in the supplementary materials (Fig. S1 in Supplement 1). The dual nature of our visuo-mental paradigm provokes both electrical and hemodynamic responses associated with visual oddball stimulations and mental arithmetic operations.

2.3. Data acquisition

Both signals were recorded simultaneously using a single cap mounted with both EEG electrodes and fNIRS optodes. fNIRS data were recorded using NIRScout (NIRx Inc.) with two NIR lights (760 nm and 850 nm wavelengths) and digitized at 7.81 Hz. EEG data were recorded using the g.USBamp amplifier (g.tec Medical Tech., Schiedlberg, Austria) and digitized at 256 Hz. Figure 1 shows a schematic head montage model of the fNIRS-EEG sensors. EEG was recorded from 16 channels: AF3*, AF4*, F1*, Fz*, F2*, T7, Cz, T8, P7, P3, Pz, P4, P8, PO7, PO8, and Oz covering all of the prefrontal, frontal, central, parietal, temporal and occipital areas, which are investigated commonly in whole head surface ALS studies [32–34] (note: AF3*, AF4*, F1*, Fz*, and F2* respectively, were the nearest electrode placements to fNIRS-occupied AF3, AF4, F1, Fz, and F2 according to the 128-channel montage). As depicted in Fig. 1, we used eight emitters and seven detectors to create a total of 16 fNIRS channels. Most of the fNIRS channels were mounted on the frontal and prefrontal areas that cover the regions in which extra-motor alterations and cognitive impairments are most often reported in people with ALS [32], along with two parietal channels. Following the modified combinatorial nomenclature (MCN) montage, the emitters were located at Fpz, AF3, AF4, F3, Fz, F4, CP5, and CP6, while the detectors were placed at Fp1, Fp2, AFz, F1, F2, P5, and P6. Each fNIRS channel used an emitter-detector pair with the optimal 3-cm distance recommended by Yamamoto et al. [35]. This multimodal montage follows standards closely and is convenient to mount, making it an appropriate candidate for future multimodal applications. All experimental protocols and data acquisition for EEG and fNIRS were controlled using BCI2000 and NIRStar software.

2.4. Data analysis

EEG data were band-pass filtered at 0.3–35 Hz and detrended to remove baseline drift and out of band artifacts. Then, the data were checked for extreme values and outliers. Participants from both the ALS and HC groups had the same total number of $9 \times 2 \times 14 = 252$ (number of participants $\times$ number of runs $\times$ number of trials) observation points (i.e., samples) for both modalities (i.e., EEG and fNIRS). For EEG spectral features, the data were decomposed into spectrograms using a set of 30 complex Morlet wavelets ranging from 1–30 Hz and 3–10 cycles. The baseline-corrected spectrograms were obtained by dividing each frequency bin and time point by the baseline (-3 to -1 sec pre-stimulus window) average and calculating the percentage changes. The spectrograms from [0–5 sec] post-stimulus were then averaged across four traditional frequency bands: delta (1–3 Hz), theta (4–7 Hz), alpha (8–12 Hz), and beta (13–30 Hz) to generate four different features. In total, there were $16 \times 4 = 64$ (channels $\times$ frequency bands) spectral features extracted from EEG data. For EEG temporal features, we used event-related potentials (ERPs), the averaged EEG waveforms of time-locked to stimulus or response events, in which the data were segmented to [0–800 ms] post-stimulus and the ERPs were then obtained. Five ERP features corresponding to peak amplitudes of P200, P300,

Biomedical Optics EXPRESS

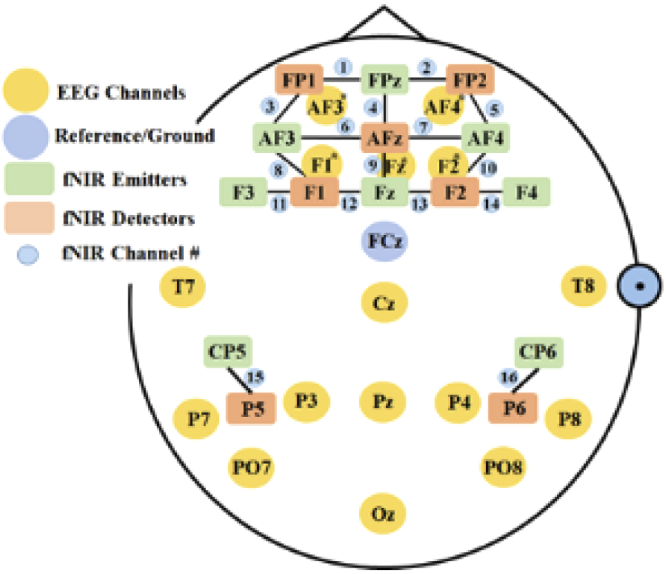


Fig. 1. Schematic head model of the fNIRS-EEG sensors.

P600, N200, and N400 components were then extracted in which the P200, P300, and P600 components were defined as the maximum peaks between 100–250, 250–400, and 650–800 ms post-stimulus, respectively, while the N200 and N400 components were defined as the minimum peaks between 150–280 and 360–560 ms post-stimulus, respectively. Following our previous work [31], these features have previously reflected significant differences between ALS patients and healthy controls, and thus have been considered as proper features with high separability for the classification procedure. In total $16 \times 5 = 80$ (channels $\times$ ERP components) temporal features were extracted from the EEG data.

fNIRS data were band-pass filtered at 0.01–0.2 Hz to mitigate physiological noises caused by respiratory and cardiac activities [36]. Then, oxy-hemoglobin (HbO2) concentration changes were extracted from the raw optical intensity data as features using the modified Beer-Lambert Law [37]. The average baseline (-2 to -1 sec pre-stimulus window) was then subtracted from the following post-stimulus signal for each epoch, and then, the peak and the area under the curve (AUC) of HbO2 were extracted using [0–6 sec] post-stimulus window for each of the 16 fNIRS channels, providing a total of $16 \times 2 = 32$ (channels $\times$ feature types) features extracted from fNIRS data.

All features were then normalized by subtracting the mean and dividing by the standard deviation of each feature vector (z-score). Outliers were clipped by setting all the values that were more than three feature standard deviations from the feature mean to only three standard deviations from the mean [38]. This was done to eliminate any degradational effect of the feature value range on the feature selection process. All the EEG and fNIRS vectors of features were then concatenated and the whole dataset was shuffled and partitioned into two main (equal size) folds with five sub-folds in each main fold for cross-validation testing to optimize the features.

To improve the discriminative performance of our classification procedure, we used an optimization framework following that proposed by Meyer et al. [27]. This framework consists of three steps: 1) maximizing the relevance of a selected feature set to the class labels, 2) minimizing the redundancy between features within a selected subset of the original features, and 3) maximizing the complementarity between features with respect to the class labels. The

Biomedical Optics EXPRESS

optimization formulation in which the features were selected is defined in the equation below.

$$X_S^{Opt} = \arg\max_{X_s \subseteq X} \left\{ \sum_{X_i \in X_S} \sum_{X_j \in X_S} I(X_{i,j}; Y) \right\} \tag{1}$$

In this formulation, Y represents the vector of output labels (HC = 1, ALS = −1), X, X_S and $X_{i,j}$ represent the original set of n features [n is the number of features in Eq. (2)], a subset of original features, and a subset of original features consisting of two single features (X_i and X_j) respectively defined in the equations below. The term under optimization inside the objective function represents the mutual information $I(.)$ between $X_{i,j}$ and Y. The term "arg max" states that the objective function is supposed to be maximized by searching for the $X_S \subseteq X$ to find the optimized feature set (i.e., X_S^{Opt}).

$$X = \{X_i : i \in A = \{1, \ldots, n\}\} \tag{2}$$

$$X_S = \{X_i : i \in S \subseteq A\} \tag{3}$$

$$X_{i,j} = \{X_i, X_j\} \tag{4}$$

Equation (1) is an optimization formulation for finding a subset of features that can maximize the joint mutual information of class labels with each pair of features inside the selected subset of original features. The joint mutual information of two random variables with another variable can be defined by the equation below.

$$I(X_{i,j}; Y) = I(X_i; Y) + I(X_j; Y) - C(X_i; X_j; Y) \tag{5}$$

The first two terms in this equation are the mutual information between single features and the class labels. These terms represent the relevance of each feature to the class labels, which means maximizing the term in Eq. (1) will optimize the relevance of each feature alone. The last term, denoted as $C(.)$ represents the interaction among the whole set of both features and the class labels. The lower the interaction term, the less redundant the variables are, and the higher their complementarity is (if the interaction term is negative). The interaction term in Eq. (5) for three variables can be obtained using the entropies and joint entropies of the set of variables according to the equation below.

$$\begin{aligned} C(X_i; X_j; Y) &= H(X_i) + H(X_j) + H(Y) \\ &-H(X_i, X_j) - H(X_i, Y) - H(X_j, Y) + H(X_i, X_j, Y) \end{aligned} \tag{6}$$

The entropy of variable(s) is denoted with $H(.)$ in this formulation. If the interaction term becomes negative, it can be inferred from Eq. (5) that $I(X_{i,j}; Y) > I(X_i; Y) + I(X_j; Y)$. Therefore, the gain resulting from using the joint mutual information of the two features will be more than the sum of the individual features' information. This property is caused by the existence of complementarity between two features.

As finding the optimized subset of features according to Eq. (1) is a non-deterministic polynomial-time hardness (NP-hard) problem [39], a semi-optimized strategy based on forward selection search was used to solve this equation [40]. This approach consists of updating a set of selected features X_s with the feature X_i from the set of remaining features that have not been selected yet. This new feature has been paired with all the members of the pre-selected set of features and should maximize the summation of joint mutual information between all paired sets of features and class labels. In other words, instead of attempting to find an optimized solution

Biomedical Optics EXPRESS

for Eq. (1), a semi-optimized solution will be substituted based on the equation below using a procedural updating approach.

$$X_S^{Opt} = \arg \max_{\substack{X_s \subseteq X \\ X_i \in X_{-S}}} \left\{ \sum_{X_j \in X_S} I(X_{i,j}; Y) \right\} \tag{7}$$

In this formulation, X_{-S} represents the whole set of original features with those in X_S removed. This can be defined as the equation below.

$$X_{-S} = \{X_i : i \in A - S\} \tag{8}$$

This strategy starts with an empty set of variables and progressively updates the solution by adding the variable that maximizes the objective function in Eq. (7) until an adequate number of features is reached. The pseudo-code for the sequential feature selection algorithm is shown in Fig. 2.

1. Initialize feature set
   $X_1 = \arg \max_{X_k \in X} [I(X_k; Y)]$
2. Update original and selected feature set
   $X = X - X_1;\ X_S = X_1$
3. Select next best feature
   $X^+ = \arg \max_{X_i \in X} [\sum_{X_j \in X_S} I(X_{i,j}; Y)]$
4. Update original and selected feature set
   $X = X - X_i;\ X_S = X_S + X_i$
5. While size $(X_S) \leq$ minimum number of features per modality
   Go to step 3

Fig. 2. Sequential feature selection pseudo-code.

A support vector machine (SVM) classifier, which has been widely used for brain signal classification was used to classify data points corresponding to two classes of HC and ALS denoted as $Y \in \{\text{HC} = 1, \text{ALS} = -1\}$. A non-linear polynomial kernel was used for SVM in this study to maximize discrimination between data points, as it allows complex separation surfaces requiring optimization of a reduced number of hyper-parameters. In order to reduce the bias associated with training and test data and to improve the generalizability of the proposed framework, a cross-validation technique was employed in which the generalization error was estimated based on resampling. A 2-fold cross-validation strategy was then used to partition each dataset into separate datasets for feature selection and validation as follows: the dataset was first split into two equal parts. Each half-dataset was separately used as training data to conduct the learning process and optimize the parameters. The results were then applied on the other half (i.e., testing dataset) to produce the classification accuracy for that corresponding fold. The final accuracy was the average of both folds' accuracies. Within the inner level of the aforementioned cross-validation, each half-dataset was split into five sub-fold to select and validate the best number of features (i.e., our only hyperparameter under optimization at the classification level). In a leave-one-out strategy for the aforementioned 5-fold cross-validation, the feature selection and classifier training was done for each 80% of the half-dataset and was repeated five times to cover all the sub-folds. Each training process was done for a number of

Biomedical Optics EXPRESS

optimally selected features ranging from 1 to 32 (32 is the minimum number of features per modality). The classification accuracies of the five validation sets were then averaged for each number of features, and the best number of features was then selected. This whole process was done in a similar way for each single modality and for the multimodal data. To evaluate the classifier, three metrics of accuracy, sensitivity, and specificity were used as follows:

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \tag{9}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{10}$$

$$Specificity = \frac{TN}{TN + FP} \tag{11}$$

where TP denotes the correct classifications of positive cases, TN denotes the correct classifications of negative cases, FP denotes the incorrect classifications of negative cases into class positive, and FN denotes the incorrect classifications of positive cases into class negative.

3. Results

The classification accuracy of the validation dataset for different numbers of selected features using the three modality options (i.e., EEG, fNIRS, EEG + fNIRS) are shown in Fig. 3. The

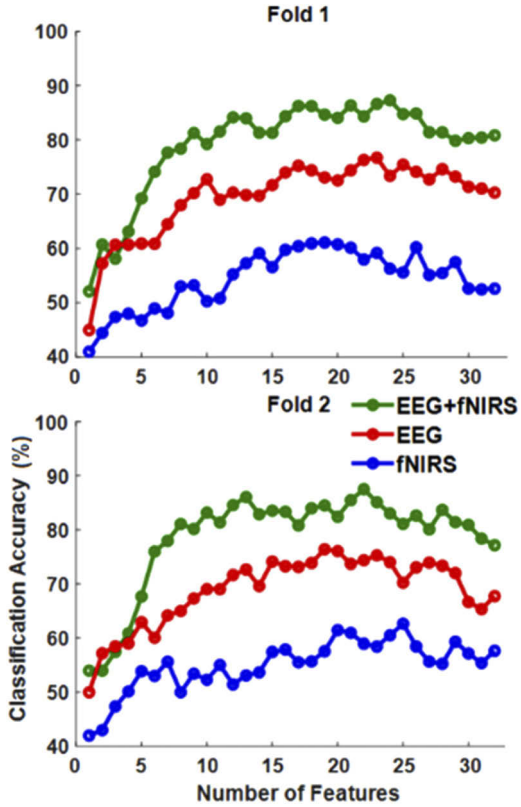


Fig. 3. Classification accuracy of single and hybrid modalities for variable sizes of the selected optimal feature subset (averaged across sub-folds of the validation dataset for fold 1 (top) and fold 2 (bottom)).

Biomedical Optics EXPRESS

averaged accuracy across the five validation sub-folds of the first main fold (fold 1) is shown in the top plot, and the bottom plot shows the averaged accuracy across the five validation sub-folds of the second main fold (fold 2). In both plots, at first, the curve (classification accuracy) ascends as the size of the optimally selected feature subset increases. It then remains around the range of maximum accuracy after increasing the number of features, reaches its maximum classification accuracy at a certain point, and finally descends. In general, the hybrid EEG-fNIRS modality performs considerably better than other single modalities in terms of the classification accuracy. In the first fold, the optimal number of features with the maximum accuracies for different modalities were: EEG + fNIRS: 87.32% accuracy with 24 features, EEG: 76.71% accuracy with 23 features, and fNIRS: 60.19% accuracy with 26 features. In the second fold, the maximum accuracies for different modalities were: EEG + fNIRS: 87.51% accuracy with 22 features, EEG: 76.39% accuracy with 19 features, and fNIRS: 62.64% accuracy with 25 features.

Figure 4. shows the relative portions of included features from each feature category/subcategory when averaged over optimal selected feature sets from all sub-folds. This figure highlights the relative discriminatory importance of each feature in the final classification procedure. As it is seen, EEG spectral features were the most selected features with 49% presence, followed by fNIRS features with 27% and EEG temporal features with 24% presence. The most selected three feature types were beta-band power with 22% presence, theta-band power with 18% presence, and P300 peak with 16% presence.

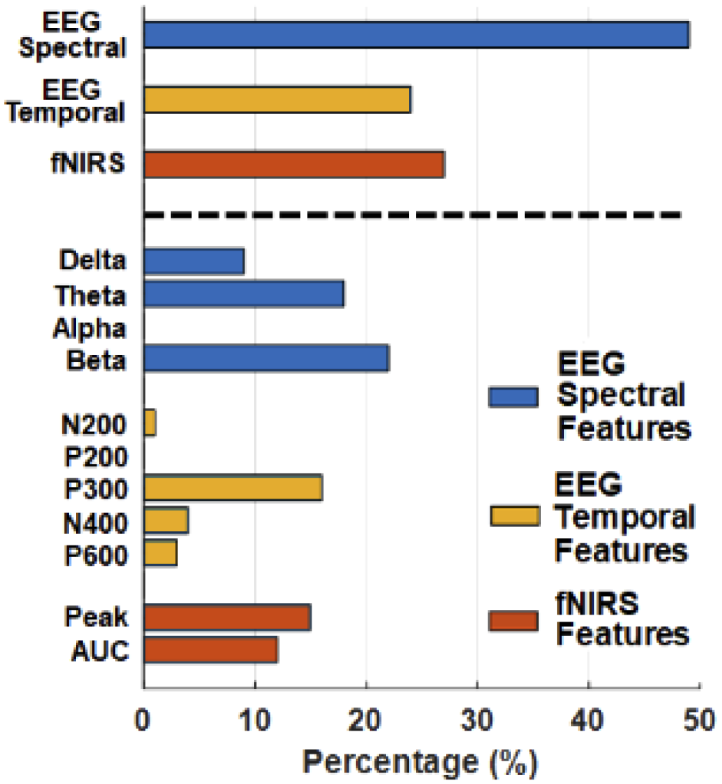
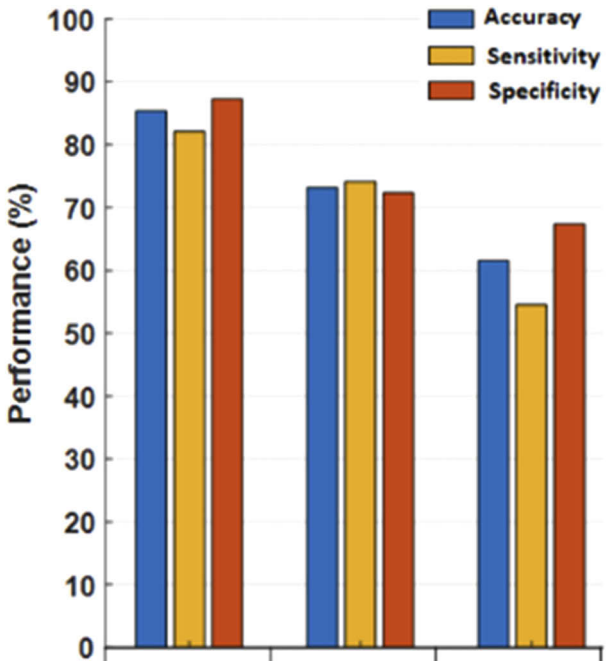


Fig. 4. Relative portions of included features from each feature category/subcategory averaged over optimal selected feature sets from all sub-folds.

Figure 5. shows classification performance characteristics based on the optimal selected subset of features which was obtained from sub-folds for single and hybrid modalities, averaged across

Biomedical Optics EXPRESS

both test folds. The hybrid classification achieved the best test accuracy of 85.38%, outperforming EEG with its best accuracy of 73.23%, and fNIRS with its best accuracy of 61.56%.

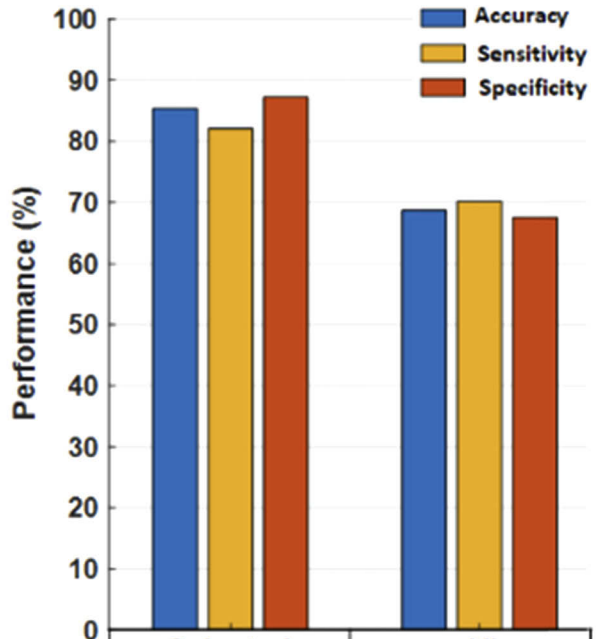


	EEG+fNIRS	EEG	fNIRS
Accuracy	85.38 %	73.23 %	61.56 %
Sensitivity	82.13 %	74.19 %	54.53 %
Specificity	87.26 %	72.38 %	67.42 %

Fig. 5. Classification performance characteristics for single and hybrid modalities.

Figure 6 shows the performance characteristics of the hybrid classification for the optimally selected set of features compared to hybrid classification using all features without any feature selection procedure. The feature selection procedure improved accuracy by 16.67% over the test set.

Biomedical Optics EXPRESS



	Selected Features	All Features
Accuracy	85.38 %	68.71 %
Sensitivity	82.13 %	70.14 %
Specificity	87.26 %	67.54 %

Fig. 6. Classification performance characteristics for the selected optimal feature subset and the original set of features without any feature selection procedure.

4. Discussion

In this paper, we used an information theory-based method to optimize feature selection and thereby classify between a healthy group and a pathological one, people with ALS in this case, during a visuo-mental task using multimodal EEG and fNIRS data. The proposed technique takes the first steps to systematically exploit the complementarity aspect of the fused features extracted from electrical and hemodynamic neural activities through a feature selection algorithm that quantifies the complementarity between features and selects the optimal fused subset to improve classification performance. Although the feature selection algorithm was adopted from "Meyer et al. [27]" in which the authors used the algorithm for a single-modality dataset, to the best of our knowledge, it has not been applied to any hybrid dual modality dataset in which both modalities have complementary information. Considering complementary information in multimodal data can make a remarkable increase in the classification performance compared to the simple concatenation of the features if only certain features from each modality that can increase the complementarity function are selected for the classification. Thus, it can be inferred that applying this algorithm to a dual modality dataset can exploit the full potential of such an algorithm which was presented in our results. Our results showed that when an integrated set of features from both modalities was used, classification performance was considerably improved compared to when EEG or fNIRS alone was used. Moreover, classification performance was

Biomedical Optics EXPRESS

substantially improved for the integrated subset of optimally selected features compared to when no feature selection was done.

Our overall classification results revealed that considerable improvements in all three performance metrics are achievable with the proposed fusion approach. This supports our central hypothesis that the systematic selection of fused complementary EEG and fNIRS features of can improve classification performance. The fused feature selection model enabled us to take advantage of the strengths of both modalities in unified analytics. Although it is impossible to make fair quantitative comparisons with other similar studies as the algorithms were run on different datasets the improvement in hybrid classification accuracy achieved in this study relative to single modality accuracies was competitive with previous EEG and fNIRS fusion studies, including those reported by Fazli et al. [11] and Putze et al. [12]. Our improved fusion results may be due to the level of fusion being adopted, as both of their studies applied fusion at the decision level, i.e., using a meta classifier to integrate the outputs from one EEG classifier and one fNIRS classifier. Indeed, the cross-modality inconsistencies which negatively affect the efficiency of modality fusions [41] cannot be avoided in decision level fusions, while such inconsistencies between modalities and their features are removed by the feature selection algorithm used in our study. Moreover, it is likely that the outputs from the EEG classifier and fNIRS classifier in these studies are highly correlated with less complementary information, and thus a systematic fusion of the features to properly maximize the complementary benefits from both modalities has been lacking. In contrast to studies done by Fazli et al. [11] and Putze et al. [12], Yin et al. [26] considered the feature level fusion of bimodal EEG and fNIRS and were able to improve the decoding of motor imagery tasks using a feature selection algorithm based on removing redundancy between the integrated EEG and fNIRS features. However, Yin et al. achieved a modest improvement, which may be due to not systematically exploiting the potential of complementarity and focusing only on removing redundancy between their hybrid modalities in their feature selection method, although the authors mentioned that EEG and fNIRS complement each other in presenting cortex activation.

The technique used for feature selection in our study selects an optimal subset of features that have maximum pairwise mutual information with the specified classes of interest (two classes in our case). Although the most complete method would consider all possible feature subsets, even with a small number of features, this procedure is computationally impossible and cannot be used in practice [42]. Given the fact that most feature sets used to represent EEG and fNIRS signals are sets of different types of features with redundancies and complementarities, this technique considers a trade-off between computational cost and the number of chosen features. This contrasts with other techniques that select features individually without considering interactions between features. The classification accuracy using features obtained by applying our technique outperforms those obtained by applying individual feature selection methods when applied to EEG and fNIRS signals. Moreover, mutual information measures non-linear dependencies between a set of random variables, taking into account higher-order statistical structures existing in the data, as opposed to linear and second-order statistical measures such as correlation and covariance. This makes mutual information-based techniques especially beneficial for a combination of features from different modalities that are likely to have non-linear relationships with each other.

This study considered complementarities between features only up to order two to avoid the additional computational complexity required by higher orders of feature fusion. Future work might consider higher levels of feature fusion with more complexity, requiring greedy search algorithms but potentially providing more advanced solutions. The small sample size and the heterogeneous characteristics of our patient group was another limitation of this study. If a larger number of patients are recruited in future studies, it will be possible to classify them into subgroups based on the onset of clinical symptoms and cognitive deficits to better discriminate between different patterns rather than considering putative patterns of altered brain functions for

Biomedical Optics EXPRESS

all ALS patients. In addition, we did not analyze differences in gender and education, which might affect the obtained neuro-markers measures. Future research with larger patient samples should be conducted to further consider demographic information in smaller sub-groups. Applying the proposed framework in this study to other datasets of integrated EEG and fNIRS in future work will further validate the efficiency of the adopted feature selection algorithm for neuro-clinical studies. Furthermore, in the future, applying other state of the art algorithms that are designed for dual-modality data classification on the same dataset will provide a more robust ground to make fair quantitative comparisons between the proposed framework and other approaches.

5. Conclusion

Overall, in this study, we adopted a mutual information-based feature selection algorithm to propose a classification framework for hybrid EEG-fNIRS data which was used to classify between a healthy and a pathological group, patients with ALS in this application, during a visuo-mental task. The optimized process of selecting features to increase classification performance was based on exploring three properties of the fused features, including decreasing redundancy, increasing relevance and increasing complementarity. The multimodal results revealed a considerable improvement of classification performance characteristics, including 16% accuracy improvement over hybrid classification with no feature selection, 12% accuracy improvement over single modal classification using EEG, and 23% accuracy improvement over single modal classification using fNIRS. These results support the idea of using complementary features from fused EEG-fNIRS in neuro-clinical studies for optimized decoding of neural information, and thus, improve the performance of relevant applications, including BCI and neuro-pathological diagnosis.

Funding. National Science Foundation (NSF-1913492); Institutional Development Award (IDeA) Network for Q3 Biomedical Research Excellence (P20GM103430).

Acknowledgments. The authors would like to thank the participants who took part in this study, without whom this study would not have been possible. We would also like to thank the ALS Association Rhode Island Chapter and the National Center for Adaptive Neurotechnologies for their continuous support, and Alyssa Zisk for proofreading this manuscript.

Disclosures. The authors declare no conflicts of interest.

Supplemental document. See Supplement 1 for supporting content.

References

1. Z. Ma, Z. H. Tan, and J. Guo, "Feature selection for neutral vector in EEG signal classification," Neurocomputing **174**, 937–945 (2016).
2. M. A. Li, Y. F. Wang, S. M. Jia, Y. J. Sun, and J. F. Yang, "Decoding of motor imagery EEG based on brain source estimation," Neurocomputing **339**(1), 182–193 (2019).
3. M. Yang, J. Li, Z. Li, D. Yao, W. Liao, and H. Chen, "Whole-brain functional connectome-based multivariate classification of post-stroke aphasia," Neurocomputing **269**, 199–205 (2017).
4. H. Peng, C. Li, J. Chao, T. Wang, C. Zhao, X. Huo, and B. Hu, "A novel automatic classification detection for epileptic seizure based on dictionary learning and sparse representation," Neurocomputing **424**, 179–192 (2021).
5. K. Liu, Z. L. Yu, W. Wu, Z. Gu, and Y. Li, "Imaging brain extended sources from EEG/MEG based on variation sparsity using automatic relevance determination," Neurocomputing **389**, 132–145 (2020).
6. S. Ahn and S. C. Jun, "Multi-modal integration of EEG-fNIRS for brain-computer interfaces – Current limitations and future directions," Front. Hum. Neurosci. (2017).
7. S. B. Borgheai, J. Mclinden, A. H. Zisk, S. I. Hosni, R. J. Deligani, M. Abtahi, K. Mankodiya, and Y. Shahriari, "Enhancing Communication for People in Late-Stage ALS Using an fNIRS-Based BCI System," IEEE Trans. Neural Syst. Rehabil. Eng. **28**(5), 1198–1207 (2020).
8. J. Hennrich, C. Herff, D. Heger, and T. Schultz, "Investigating deep learning for fNIRS based BCI," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, EMBS (2015).
9. K. S. Hong, N. Naseer, and Y. H. Kim, "Classification of prefrontal and motor cortex signals for three-class fNIRS-BCI," Neurosci. Lett. **587**, 87–92 (2015).
10. S. B. Erdoğan, E. Özsarfati, B. Dilek, K. S. Kadak, L. Hanoğlu, and A. Akin, "Classification of motor imagery and execution signals with population-level feature sets: Implications for probe design in fNIRS based BCI," J. Neural Eng. **16**(2), 026029 (2019).

Biomedical Optics EXPRESS

11. S. Fazli, J. Mehnert, J. Steinbrink, G. Curio, A. Villringer, K. R. Müller, and B. Blankertz, "Enhanced performance by a hybrid NIRS-EEG brain computer interface," Neuroimage **59**(1), 519–529 (2012).
12. F. Putze, S. Hesslinger, C. Y. Tse, Y. Y. Huang, C. Herff, C. Guan, and T. Schultz, "Hybrid fNIRS-EEG based classification of auditory and visual perception processes," Front. Neurosci. **8**, 373 (2014).
13. F. Al-Shargie, T. B. Tang, and M. Kiguchi, "Stress Assessment Based on Decision Fusion of EEG and fNIRS Signals," IEEE Access **5**, 19889–19896 (2017).
14. Y. Liu, H. Ayaz, and P. A. Shewokis, "Mental workload classification with concurrent electroencephalography and functional near-infrared spectroscopy," Brain-Computer Interfaces **4**(3), 175–185 (2017).
15. L. W. Ko, Y. C. Lu, H. Bustince, Y. C. Chang, Y. Chang, J. Ferandez, Y. K. Wang, J. A. Sanz, G. Pereira Dimuro, and C. T. Lin, "Multimodal fuzzy fusion for enhancing the motor-imagery-based brain computer interface," IEEE Comput. Intell. Mag. **14**(1), 96–106 (2019).
16. A. P. Buccino, H. O. Keles, and A. Omurtag, "Hybrid EEG-fNIRS asynchronous brain-computer interface for multiple motor tasks," PLoS One **11**(1), e0146610 (2016).
17. T. Nguyen, S. Ahn, H. Jang, S. C. Jun, and J. G. Kim, "Utilization of a combined EEG/NIRS system to predict driver drowsiness," Sci. Rep. **7**(1), 43933 (2017).
18. M. Abtahi, S. Bahram Borgheai, R. Jafari, N. Constant, R. Diouf, Y. Shahriari, and K. Mankodiya, "Merging fNIRS-EEG brain monitoring and body motion capture to distinguish parkinsons disease," IEEE Trans. Neural Syst. Rehabil. Eng. (2020).
19. A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," IEEE Trans. Pattern Anal. Mach. Intell. **22**(1), 4–37 (2000).
20. M. Saadati, J. Nelson, and H. Ayaz, "Convolutional neural network for hybrid fNIRS-EEG mental workload classification," in *Advances in Intelligent Systems and Computing* (2020).
21. F. Al-Shargie, M. Kiguchi, N. Badruddin, S. C. Dass, A. F. M. Hani, and T. B. Tang, "Mental stress assessment using simultaneous measurement of EEG and fNIRS," Biomed. Opt. Express **7**(10), 3882–3898 (2016).
22. T. B. T. Al-Shargie, Hasan Fares, and Al-Nashash, "Assessment of mental stress among undergraduate students using novel fusion method on EEG and fNIRS features," Frontiersin 1 (2019).
23. F. Al-Shargie, T. B. Tang, and M. Kiguchi, "Assessment of mental stress effects on prefrontal cortical activities using canonical correlation analysis: an fNIRS-EEG study," Biomed. Opt. Express **8**(5), 2583–2598 (2017).
24. G. Brown, A. Pocock, M. J. Zhao, and M. Luján, "Conditional likelihood maximisation: A unifying framework for information theoretic feature selection," J. Mach. Learn. Res. **13**, 27–66 (2012).
25. X. Lin, L. Sai, and Z. Yuan, "Detecting concealed information with fused electroencephalography and functional near-infrared spectroscopy," Neuroscience **386**, 284–294 (2018).
26. X. Yin, B. Xu, C. Jiang, Y. Fu, Z. Wang, H. Li, and G. Shi, "A hybrid BCI based on EEG and fNIRS signals improves the performance of decoding motor imagery of both force and speed of hand clenching," J. Neural Eng. **12**(3), 036004 (2015).
27. P. E. Meyer, C. Schretter, and G. Bontempi, "Information-theoretic feature selection in microarray data using variable complementarity," IEEE J. Sel. Top. Signal Process. **2**(3), 261–274 (2008).
28. S. Baillet, L. Garnero, G. Marin, and J. P. Hugonin, "Combined MEG and EEG source imaging by minimization of mutual information," IEEE Trans. Biomed. Eng. **46**(5), 522–534 (1999).
29. M. A. B. S. Akhonda, Y. Levin-Schwartz, S. Bhinge, V. D. Calhoun, and T. Adali, "Consecutive independence and correlation transform for multimodal fusion: application to EEG and FMRIdata," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings* (2018).
30. P. Zhang, X. Wang, J. Chen, and W. You, "Feature weight driven interactive mutual information modeling for heterogeneous bio-signal fusion to estimate mental workload," Sensors **17**(10), 2315 (2017).
31. S. B. Borgheai, R. J. Deligani, J. McLinden, A. Zisk, S. I. Hosni, M. Abtahi, K. Mankodiya, and Y. Shahriari, "Multimodal exploration of non-motor neural functions in ALS patients using simultaneous EEG-fNIRS recording," J. Neural Eng. **16**(6), 066036 (2019).
32. F. Christidi, E. Karavasilis, M. Rentzos, N. Kelekis, I. Evdokimidis, and P. Bede, "Clinical and radiological markers of extra-motor deficits in amyotrophic lateral sclerosis," Front. Neurol. **9**, 1005 (2018).
33. P. M. Iyer, C. Egan, M. Pinto-Grau, T. Burke, M. Elamin, B. Nasseroleslami, N. Pender, E. C. Lalor, and O. Hardiman, "Functional connectivity changes in resting-state EEG as potential biomarker for amyotrophic lateral sclerosis," PLoS One **10**(6), e0128682 (2015).
34. K. Kopitzki, A. Oldag, C. M. Sweeney-Reed, J. Machts, M. Veit, J. Kaufmann, H. Hinrichs, H. J. Heinze, K. Kollewe, S. Petri, B. Mohammadi, R. Dengler, A. R. Kupsch, and S. Vielhaber, "Interhemispheric connectivity in amyotrophic lateral sclerosis: A near-infrared spectroscopy and diffusion tensor imaging study," NeuroImage Clin. **29**(12), 666–672 (2016).
35. T. Yamamoto, A. Maki, T. Kadoya, Y. Tanikawa, Y. Yamada, E. Okada, and H. Koizumi, "Arranging optical fibres for the spatial resolution improvement of topographical images," Phys. Med. Biol. **47**(18), 3429–3440 (2002).
36. F. Scarpa, S. Cutini, P. Scatturin, R. Dell'Acqua, and G. Sparacino, "Bayesian filtering of human brain hemodynamic activity elicited by visual short-term maintenance recorded through functional near-infrared spectroscopy (fNIRS)," Opt. Express **18**(25), 26550–68 (2010).
37. L. Kocsis, P. Herman, and A. Eke, "The modified Beer-Lambert law revisited," Phys. Med. Biol. **51**(5), N91–N98 (2006).

Biomedical Optics EXPRESS

38. E. Cernadas, M. Fernández-Delgado, E. González-Rufino, and P. Carrión, "Influence of normalization and color space to color texture classification," Pattern Recognit. (2017).
39. D. Pisinger, "Upper bounds and exact algorithms for p-dispersion problems," Comput. Oper. Res. **33**(5), 1380–1398 (2006).
40. A. Billionnet and F. Calmels, "Linear programming for the 0-1 quadratic knapsack problem," Eur. J. Oper. Res. **92**(2), 310–325 (1996).
41. C. W. Wu, P. J. Tsai, S. C. J. Chen, C. W. Li, A. L. Hsu, H. Y. Wu, Y. T. Ko, P. C. Hung, C. Y. Chang, C. P. Lin, T. J. Lane, and C. Y. Chen, "Indication of dynamic neurovascular coupling from inconsistency between EEG and fMRI indices across sleep–wake states," Sleep Biol. Rhythms **17**(4), 423–431 (2019).
42. M. Deriche and A. Al-Ani, "A new algorithm for EEG feature selection using mutual information," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221)* (Vol. 2) (2001), p. 4.