

Application of functional data analysis in classification and clustering of functional near-infrared spectroscopy signal in response to noxious stimuli

Ahmad Pourshoghi
Issa Zakeri
Kambiz Pourrezaei

Application of functional data analysis in classification and clustering of functional near-infrared spectroscopy signal in response to noxious stimuli

Ahmad Pourshoghi,^a Issa Zakeri,^b and Kambiz Pourrezaei^{a,*}

^aDrexel University, School of Biomedical Engineering, Science and Health Systems, 3141 Chestnut Street, Bossone Research, Suite 718, Philadelphia, Pennsylvania 19104, United States

^bDrexel University, Dornsife School of Public Health, Department of Epidemiology and Biostatistics, 3215 Market Street, Nesbitt Hall, Philadelphia, Pennsylvania 19104, United States

Abstract. We introduce the application of functional data analysis (fDA) on functional near-infrared spectroscopy (fNIRS) signals for the development of an accurate and clinically practical assessment method of pain perception. We used the cold pressor test to induce different levels of pain in healthy subjects while the fNIRS signal was recorded from the frontal regions of the brain. We applied fDA on the collected fNIRS data to convert discrete samples into continuous curves. This method enabled us to represent the curves as a linear combination of basis functions. We utilized bases coefficients as features that represent the shape of the signals (as opposed to extracting defined features from signal) and used them to train a support vector machine to classify the signals based on the level of induced pain. We achieved 94% of accuracy to classify low-pain and high-pain signals. Moreover applying hierarchical clustering on the coefficients, we found three clusters in the data which represented low-pain (one cluster) and high-pain groups (two clusters) with an accuracy of 91.2%. The center of these clusters can represent the prototype fNIRS response of that pain level. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JBO.21.10.101411](https://doi.org/10.1117/1.JBO.21.10.101411)]

Keywords: biomarkers of pain; pain measurement; functional data analysis; functional near-infrared spectroscopy; machine learning; frontal cortex.

Paper 150819SSRR received Dec. 9, 2015; accepted for publication Mar. 24, 2016; published online May 6, 2016.

1 Introduction

During the past two decades, neurophysiological techniques that measure cerebral metabolism and circulation changes have been widely employed to open a window into human cerebral responses to pain with a long-term goal of obtaining a more direct measurement of pain perception. Different functional modalities and brain imaging techniques—especially noninvasive methods such as functional magnetic resonance imaging (fMRI)—have been used to study brain reactivity to pain in both normal subjects and patients with clinical pain conditions. There has been shown a relation between subjects' report of an ongoing pain and blood oxygen level dependent (BOLD) signal acquired by fMRI.¹ Similar relation has been reported between subjects' self-report and functional near-infrared spectroscopy (fNIRS) parameters.^{2–8} For example, Lee et al.⁴ reported that as the intensity of the noxious pressure stimuli increases, the HbO₂ in the frontal cortex increases as well, consistent with an increase in the perceived pain.

Moreover, applying machine-learning techniques on neuroimaging data in the field of pain assessment has shown promising results in recent years. Marquand et al.⁹ showed that using fMRI data from an individual, one could train a support vector machine (SVM) to predict the same individual's pain. Their

SVM model had an accuracy of 68% for distinguishing pain perception level (low pain) from pain tolerance level (high pain) and an accuracy of 91% for distinguishing heat perception level (no pain) from pain tolerance level (high pain).

Furthermore, Brown et al.¹⁰ developed a model that was not individual-based and therefore could be used on different groups of subjects. In this study, whole-brain patterns of activity were used to train a SVM to distinguish painful from nonpainful thermal stimulation. They have reported an accuracy of 81% at distinguishing painful from nonpainful stimuli.

In another important fMRI study, Wager et al.¹¹ used a machine-learning-based regression technique to identify a pattern of fMRI activity across brain regions in response to heat-induced pain. They first identified the brain regions activated by painful thermal stimuli as the dorsal posterior insula, the secondary somatosensory cortex the anterior insula, the ventrolateral and medial thalamus, the hypothalamus, and the dorsal anterior cingulate cortex. Using data from these regions, a model has been developed and then tested on a separate dataset which resulted in an accuracy of 94% to discriminate between painful heat and nonpainful warmth. This study was a strong demonstration of the existence of a universal pain signature in fMRI data.

These studies show that some features from neuroimaging data—such as BOLD signal change in fMRI—are sufficiently consistent between individuals to train a pain classifier that performs accurately when trained on one group of subjects and tested on another. In other words, it might be a universal pattern

*Address all correspondence to: Kambiz Pourrezaei, Email: pourrezk@drexel.edu

of pain activation (neurological signature) across individuals that could be used to detect pain objectively across other subjects.

Despite these advances in medical imaging technology that significantly help basic science, there remains an unmet clinical need for a practical, inexpensive tool for the reliable and objective assessment of human response to pain. fNIRS is a noninvasive technique with a short set-up time which makes it more clinic friendly for applications such as pain measurement or management in the clinic.

Due to the ease of fNIRS measurements on the frontal region, in this paper we pursued feasibility of employing this signal for objective assessment of pain.

fNIRS measure data from a continuous physiological process as discrete samples subject to observational noise. Usually, analysis of fNIRS data has been limited to discrete-time methods and most of the fNIRS studies involve analyses on features extracted from the hemodynamic time series.

However, the measured hemodynamic response by fNIRS is a realization of a naturally continuous physiological phenomenon and it is fair to assume that the true underlying trajectory is a smooth function. Functional data analysis (fDA) is a framework which enables us to convert the nonsmooth fluctuating discrete samples measured by fNIRS into smooth functions using a linear combination of basis functions. Utilizing fDA approach on fNIRS data was first introduced in Ref. 12. Here, we propose the application of fDA method on the fNIRS signal for both classification and clustering purposes.

The rest of the paper is organized as follows: in Sec. 2, we shortly describe the methodology behind our measurement system (fNIRS), the mathematical techniques that we have employed for classification and feature selection, the fDA framework that we have applied on the fNIRS signal and our experiment protocol; in Sec. 3, we present the application of fDA on the fNIRS data collected during a cold pressor test and demonstrate both classification and clustering results. Lastly, in Sec. 4, we summarize and discuss the results.

2 Materials and Methods

2.1 fNIRS

fNIRS is an optical imaging modality for noninvasive, continuous monitoring of tissue oxygenation and regional blood flow. fNIRS works based on the fact that brain activity is associated with changes in optical properties of brain tissue in near-infrared range.

Propagated light can be absorbed or scattered by the tissue. In the near-infrared range (NIR), 600 to 1000 nm optical window, the major absorbing components—chromophores—in the soft tissues are water, oxyhemoglobin (HbO₂) and deoxyhemoglobin (Hb). There are also minor contributions from other tissue chromophores, such as melanin, lipids, and so on. Each one has a different level of absorption at each wavelength. Components such as water, lipids, CFS, and melanin can be assumed to keep a constant concentration during the test period (static absorbers) and have a little contribution to the overall attenuation in the specific window. On the other hand, concentration of dynamic absorbers—oxygenated and deoxygenated hemoglobin—changes during the experiment according to the function and metabolism of the tissue.

When light strikes a blood vessel, some photons are absorbed by oxyhemoglobin and deoxyhemoglobin in the blood. This

absorption changes the intensity of the light which scatters back to the surface. So there is a direct relationship between concentration of oxyhemoglobin and deoxyhemoglobin in the blood and changes in the intensity of light measured on the surface. The equation that governs this relationship is known as modified Beer–Lambert law.¹³

The scattered light reflects back to the tissue surface mostly within a banana-shaped optical path length. By putting a photo-detector on the surface of the skin one can sample the amount of absorption changes within this volume.

2.2 Classification

The classification goal is to train a model, based on the given examples known as training set, that can predict the output (class) of future examples based on their input (features). In other words, in classification we want to find a decision function that assigns the inputs from the feature space into the classes (target space). Different methods use different strategies to find these decision functions, boundaries, and rules.

Assume a linear decision boundary (hyperplane in n -dimension) is defined by $w^T x + b = 0$ in the feature space x where w is referred to as weight vector and b is called the bias term. We can define the binary classification problem as using training data to find w and b such that the hyperplane can separate the data into two groups (classes). In general, many solutions may exist for w and b which can classify the training data exactly. However, it is desirable to find the solution that will give the smallest generalization error (the error estimated for future unknown samples). Therefore, the problem is to find decision rules that generate such a decision boundary that separates the data into two groups and has the best classification performance.

The SVM finds a solution based on the concept of margin, which is defined as the distance between the separating hyperplane to the closest data point from either class. The optimal separating hyperplane (defined by w and b) is chosen to be the one for which margin is maximized (i.e., $1/\|w\|$, where $\|w\|$ is the norm of vector w).^{14–16}

In cases that linear boundaries separation is not possible between classes in the same space, SVM uses kernel method as a measure of similarity between features to map the observations into a higher dimension in which the data can be separated in Refs. 17 and 18.

2.3 Feature Selection

Feature selection is defined as process of selecting a subset out of the feature space which minimizes a predefined criterion, usually classification error in case of supervised learning and cluster detection error in case of unsupervised learning.¹⁹

In this paper we do not define or extract commonly used fNIRS features (such as average, extremum, or slope) from the signal. Instead, we use fDA to represent each signal with a number of coefficients. Although these coefficients may not fit in the classic definition, however, for the sake of convenience we use the term “feature” for them.

Among all the benefits of feature selection techniques two of them are of most interest to us. First of all it improves the prediction performance of our classifier by avoiding overfitting; second, it will result in a faster and more cost-effective model by reducing dimension of predictors.

Here, we used recursive feature elimination-SVM (RFE-SVM) method which is a wrapper feature selection technique.²⁰ Wrapper methods (also known as classification-guided feature selection) use the classifier as a black box to rank subsets of features based on their predictive performance. In the most general formulation, wrapper methods use the prediction performance of a given classifier to assess the relative usefulness of subsets of features (as opposed to filter methods, which select subsets of features as a preprocessing step and independent of the classifier performance; in filter methods, features are ranked based on a predefined criterion such as correlation coefficients).

The RFE-SVM method is based on a backward elimination selection meaning that the algorithm starts with all features and repeatedly removes the least promising feature at each step until all variables have been ranked. In order to find the least promising feature at each step RFE-SVM utilizes SVM and finds w and b . The removed feature is the one whose removal minimizes the variation of $\|w^2\|$. In other words, the removed variable is the one that has the least influence on the weight vector norm. The feature selection procedure at each step can be summarized as

1. Train the classifier.
2. For each feature i , compute the ranking criterion w_i^2 where w_i is the i th component of w .
3. Remove the feature with smallest ranking criterion.

2.4 Cross Validation

The leave-one-out cross validation procedure consists of removing one example from the training set, constructing the predictor on the basis only of the remaining training data, then testing on the removed example. In this fashion, one tests all examples of the training data and averages the results. In this study, we used subject-level leave-one-out cross validation meaning that we set aside all trials of a subject prior to feature selection and training and we repeat this process for all the subjects.

2.5 Functional Data Analysis

Medical devices measure data from a continuous physiological process as discrete samples subject to observational noise. We want a mathematical description of the curve or data distributed over time (in general form it can be any other types of continuum, e.g., space). Considering only one recorded sample path we can write

$$y_i = x(t_i) + e_i \quad \text{for } i = 1, 2, \dots, N,$$

where y_i is the observed sample at time point t_i , $x(t)$ is the underlying continuous process that we want to determine and e_i is the noise or measurement error of the sample at time point t_i . fDA tries to find a function—or better to say, a linear combination of basis functions—that best describes the data recorded at discrete times as a smooth function, in the sense of possessing a certain number of derivatives. In other words, we assume the existence of a smooth function x given rise to the observed data vector $y = (y_1, \dots, y_N)'$. This falls under the general class of approximation theory. The philosophy behind fDA is “to think of observed data functions as single

entities, rather than merely as a sequence of individual observations.”²¹ The advantages of using such representation is twofold: (1) it provides a computational platform to reduce dimension of data especially in cases where there are a huge number of measurements on a small number of subjects, known as large p small n problem and (2) smooth functions allow study of the dynamics of the underlying processes through their derivatives.¹²

2.5.1 Basis functions

In a functional domain, we study functional objects rather than sample points; therefore the discrete data need to be converted to a smooth functional object. However, before we can convert raw discrete samples into a functional data object, we must specify a system of basis functions that is a system of simple smooth functions that are combined linearly to approximate actual functions with an arbitrary level of accuracy. Here, we replace observations y_1, y_2, \dots, y_N (N = number of datapoints) with $x(t_1), x(t_2), \dots, x(t_N)$ where $x(t)$ is a smooth function formed by a linear combination of basis functions ϕ_j .

$$x(t) \approx \sum_{j=1}^K c_j \phi_j(t).$$

When these basis functions ϕ_j are specified, then the conversion of the data into a functional data object involves computing the coefficient of the expansion c_j .

Based on the choice of basis functions, different methods have been developed: Fourier series (suitable for periodic phenomenon), wavelet (suitable for sharp local features phenomenon like heart rate), and B-spline (suitable for smoother and slower phenomenon like hemodynamic response or body temperature).²¹

As explained above, the choice of basis functions depends on the nature of the signal, and for a physiological signal like hemodynamic response B-spline is the best option.

Spline functions are formed by joining polynomials together at fixed points called knots. The order of the polynomial (m) is defined as the number of coefficients defining the polynomial, and degree of polynomial is the highest power of the polynomial which is one less than its order ($m - 1$).

A spline system can be defined by (1) the order m of the polynomial and (2) the location and number of the knots.²² Location of knots can be distributed equally over the data or more knots can be assigned to the parts of data which have more variability.

A much better representation of splines for computation is a linear combination of a set of basis functions called B-splines. These splines have an attractive feature known as compact support (meaning that any given basis function is nonzero over a span of at most five distinct knots), which has been shown to have more stable numerical properties than methods employing other bases.²³

We use penalized least-square estimation procedure, described later, for smoothing based on B-spline bases.

In our model we have

$$y_i = x(t_i) + e_i.$$

And we want to estimate $x(t)$ as

$$x(t) \approx \sum_{j=1}^K c_j \varnothing_j(t).$$

The goodness of fit is measured by the least-squares criterion as

$$SSE = \sum_j [Y_j - x(t_j)]^2.$$

In a regular regression we try to minimize this cost function. However, here we add an extra term to the least-square criterion to measure roughness of $x(t)$. This term is defined as

$$PEN_2 = \int \left[\frac{d^2x(t)}{dt^2} \right]^2,$$

where PEN_2 measures the curvature and smoothness of the function. A smaller PEN_2 implies a less variable function, whereas a larger PEN_2 indicates a rougher curve.

The new penalized squared error cost function can be written as

$$PENSSE = \sum_j [Y_j - x(t_j)]^2 + \lambda \int \left[\frac{d^2x(t)}{dt^2} \right]^2,$$

where λ is smoothing parameter that controls the trade-off between the closeness to the observed values and the smoothness of the function. If λ is close to zero, we obtain an estimate too close to the data and if is too large, we obtain an estimate equivalent to the linear regression estimate of the data.

In this study, knots were equally spaced and the number of knots was determined based on trial-and-error methods suggest by Ref. 24 in which we start with 5, 10, 15, 20, and so on until we obtain the right amount of smoothing for the data.

Once the number of knots is determined, other methods can be used to determine the smoothing parameter (λ). An appropriate smoothing parameter may be chosen subjectively by visual judgment and prior knowledge of the process generating the data. An objective, data-driven method is also developed using the generalized cross validation (GCV) measure²⁵

$$GCV(\lambda) = \left[\frac{N}{N - df(\lambda)} \right] \left[\frac{SSE}{N - df(\lambda)} \right],$$

where N is the number of observations, SSE is the mean square error, and $df(\lambda)$ is the trace of the smoothing matrix.²⁶ In this study, we used GCV method for choosing smoothing parameter (λ).

2.6 Participants

Nineteen healthy, right-handed individuals (10 females) from the Drexel University community participated in this study after giving the informed consent form approved by the Institutional Review Board. Subjects were instructed to avoid smoking and drinking any caffeinated or alcoholic beverages for at least 3 hours prior to the experiment.

2.7 Protocol

The collected data⁸ includes 76 dataset of hand immersion in cold water in four different temperatures (1 deg, 5 deg, 10 deg, and 15 deg; 19 dataset for each temperature). These temperatures have been used to generate low-, moderate-, and high-pain levels. After a baseline recording for 30 s, subjects were asked to immerse their hand in the tepid water for 2 min for adaptation. Then, an experimenter asked them to put their hand in a constant temperature bath which is kept at a certain temperature for as long as they can tolerate the stimulated pain but no more than 5 min (Fig. 1). First 90 s of hand immersion in cold water is used in this study. During each experiment numeric pain rating scores from a 0 to 10 scale—where zero means no pain and ten means an intolerable pain—was recorded every 15 s.

2.8 Measurements

A continuous wave fNIRS system designed and developed at Drexel University was used in this study. The principle and instrumentation of fNIRS are described elsewhere.²⁷⁻²⁹ Throughout the procedure, two fNIRS sensors (Fig. 2) were positioned symmetrically on the left and right sides of a subject’s forehead proximate to the anterior median line and were secured using a medical bandage and a Velcro strap. The fNIRS sensors consisted of one light source with two light-emitting diodes (LEDs) at 730- and 850-nm wavelengths and three photodetectors. Two detectors were placed at 2.8 cm from the LED making the “far” channels to investigate the hemodynamic response at intracranial layers and one detector was located at 1 cm from the LED making the “near” channel to measure the hemodynamic changes within the superficial extracranial tissues (Fig. 2).

The sampling rate of raw optical intensity measurements was 2 Hz. The optical density parameters for 730- and 850-nm wavelengths were calculated by taking the logarithm of the ratio of the detected light intensity during baseline to the detected light intensity during the task. The optical density time series were converted to changes in oxyhemoglobin (HbO₂) and deoxyhemoglobin (Hb) concentrations using the modified Beer-Lambert law.

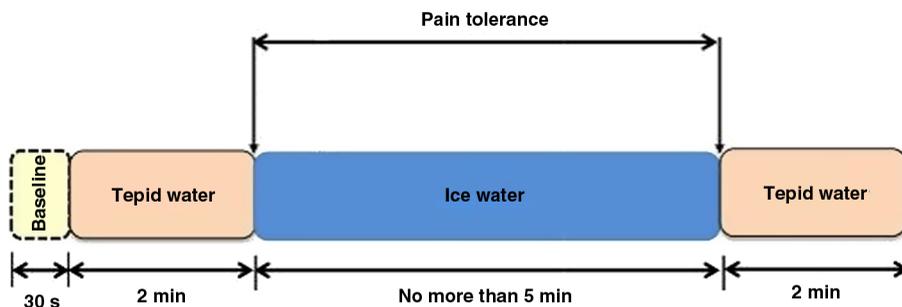


Fig. 1 The block diagram of tolerance test protocol.

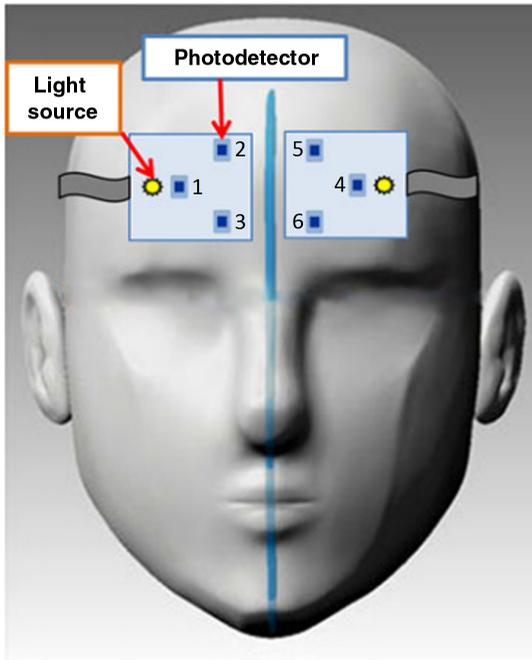


Fig. 2 Placement of NIRS probes on a subject's forehead. Six photodetectors collect the data for 6 channels. Channels 1 (right) and 4 (left) are in near distance from the sources and channels 2, 3 (right) and 5, 6 (left) are in far distance from the sources.

2.8.1 Reported pain scores

Experiment was designed in four different temperatures (0°C, 5°C, 10°C, and 15°C) to generate different levels of pain. Subjects were asked to report their pain in a numeric 0 to 10 scale (NRS-11³⁰) every 15 s. In this study, reported pain scores in the first 90 s of the experiment are averaged for each subject and assigned as subject's self-reported pain. Figure 3 shows the histogram and boxplots of these self-reported pain scores.

In order to reduce the subjective nature of self-reported pain scores and minimize ambiguity between painful and nonpainful data we decided to group pain scores equal and higher than 8 as high pain and equal and lower 6.5 as low pain. Using these criteria, we had an almost even number of trials in each category (55% painful, 45% nonpainful) and trials that were in the gray area in between high pain and low pain were excluded from the dataset. The final dataset that has been used includes 61 trials.

3 Results

3.1 Functional Data Analysis Results

In this study, we use fDA methodology to convert Hb and HbO₂ data collected by fNIRS during the cold pressor test into smooth functions estimated by cubic B-spline basis functions (order 4). The effect of using different orders for B-spline basis is shown in Fig. 4. The most popular choice is the order 4 (cubic) B-spline basis in the literature. Since in cubic splines the segments join with matching derivatives up to order 2 the curvature appears to change smoothly. This is because the second derivative measures the curvature of a curve, and the curvatures match at the breakpoints. Moreover, a cubic spline is the unique minimizer of the penalized squared error cost function explained in Sec. 2.5 (see Ref. 23 for more details).

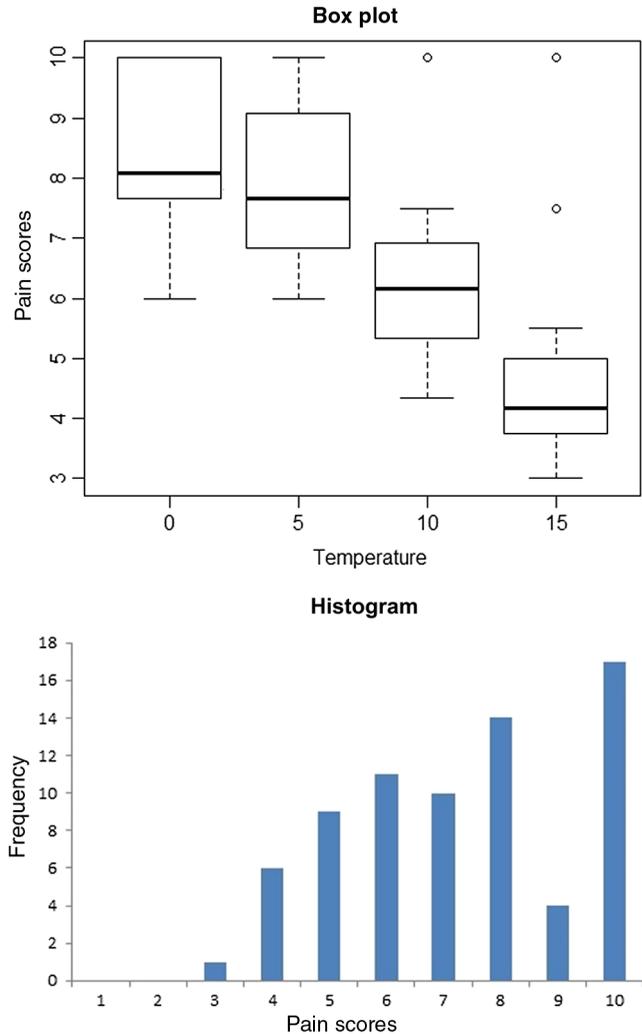


Fig. 3 Boxplots and histogram of subjects self-reported pain scores: (a) boxplots in four different temperatures, and (b) histogram. We used the thresholds 6.5 and lower for low-pain (44%) and 8 and higher for high-pain (56%) groups.

The advantage of our method is that we use the shape of the signal itself (as opposed to extracted features from the signal) to classify and cluster the data. In other words, here, we do not need to define some features from the signal and we use the basis coefficients of fDA curves instead. We do both classification and clustering on the fDA data. In the classification part, our main goal is to explore the possibility of classifying high-pain and low-pain signals based on their shapes and curvatures using fDA coefficients. fDA provides us a statistical tool to answer questions like, "What are the main ways in which the curves vary from one signal to another?"

In the clustering part we specifically would like to answer: (1) How many different types of curves exists in the data? (2) Is there a relation between reported pain levels and the clusters in any of Hb or HbO₂ curves? (3) Is it possible to find Hb or HbO₂ responses as prototype curves that represent corresponding pain levels?

3.1.1 Data processing

Raw fNIRS data for 730- and 850-nm wavelengths were first filtered using a finite impulse response low-pass filter with a

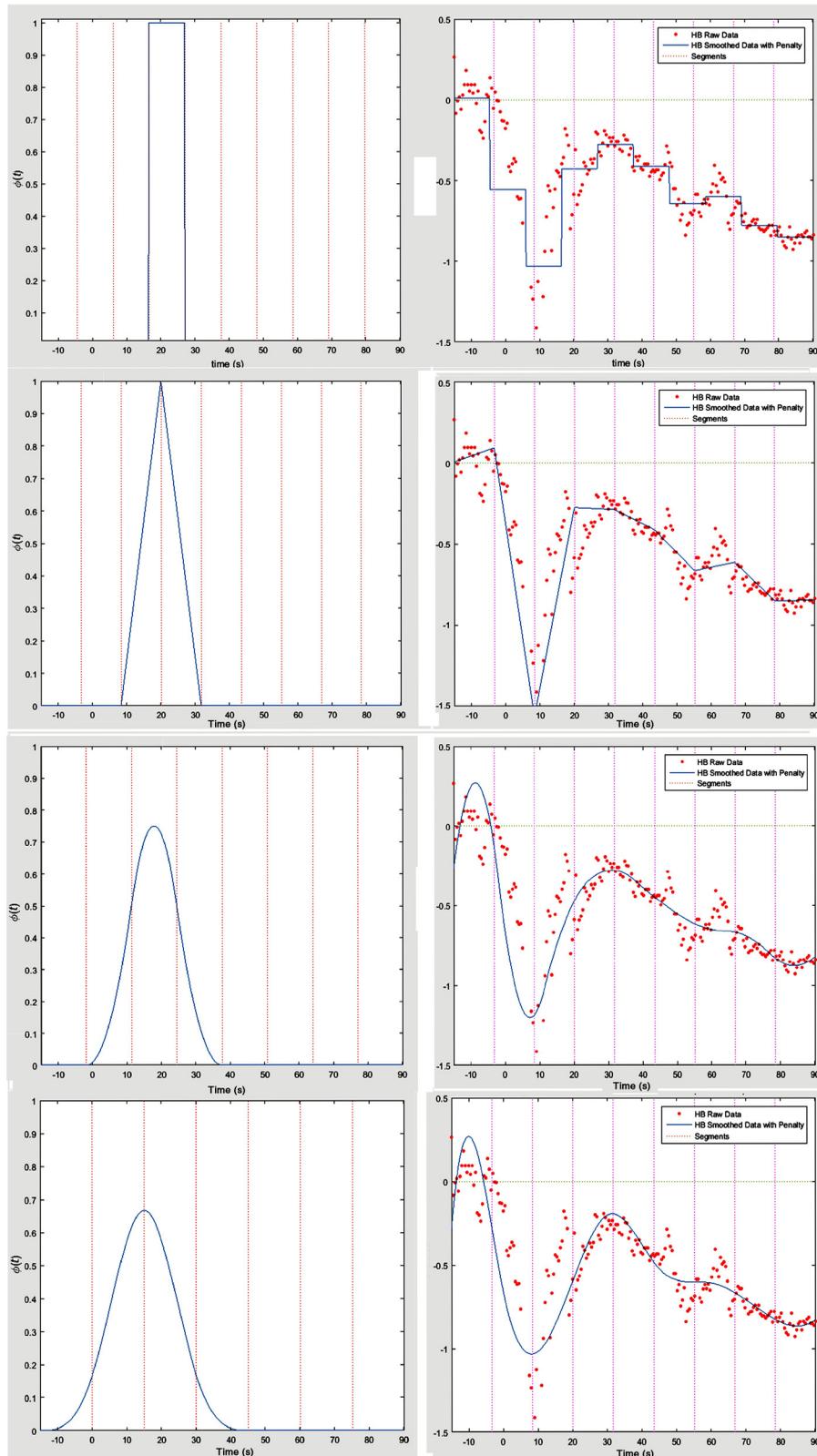


Fig. 4. Effect of using different orders for B-spline bases. B-spline basis (left), a sample HB signal estimated by 10 bases (right): (a) B-spline order 1, (b) B-spline order 2, (c) B-spline order 3, and (d) B-spline order 4.

cut-off frequency of 0.14 Hz to remove high frequency noise, respiration, and heart pulsation artifacts. Motion artifact effects were minimized in the data by securing NIRS probes with a medical bandage and a Velcro strap. The filtered raw data

were converted to changes in Hb and HbO₂ concentrations relative to the baseline using modified Beer–Lambert law.

All fDAs were conducted in MATLAB (R2015a, MathWorks) using the fDA package for MATLAB.²² Hb and

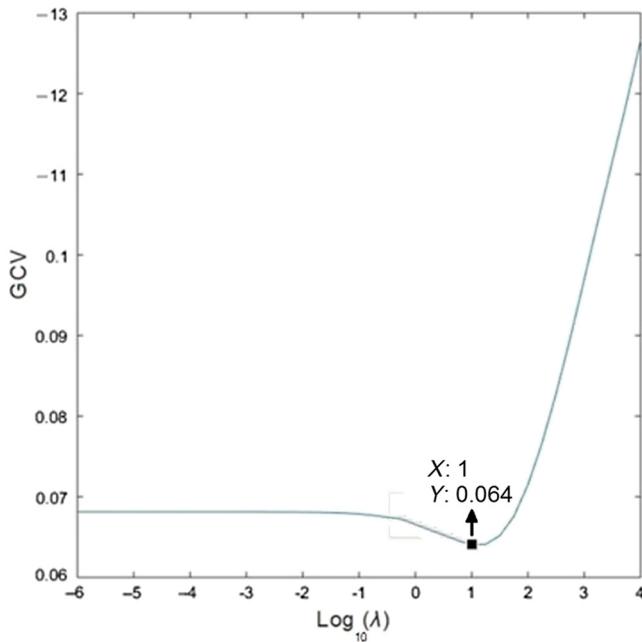


Fig. 5 GCV method was used to find the optimum smoothing parameter for each trial.

HbO₂ were smoothed by imposing a penalty on the roughness of the second derivative of the data (PEN₂) as described in the previous section. The smoothing parameter was chosen using the GCV method for each dataset (Fig. 5). Having 30 bases (clustering part), the lambda varies in the order of 0.1 and for 10 bases (classification part) it changes in the order of 10 to 100.

3.2 Classification of fNIRS Data Using Functional Data Analysis

Since there are six channels of measurements, by converting Hb and HbO₂ into functional objects, each dataset will have a coefficient matrix of size ($n \times 12$) in which n is the number of bases. We decided to use a fixed number of bases for all the trials. Figure 6 shows a sample Hb data which is estimated with a different number of bases. We found that 10 bases are enough to represent the data correctly.

The goal is to classify the data into high pain and low pain using these coefficients. We utilize SVM with Gaussian kernel as the classifier here Ref. 17.

Using 10 bases, we will have 120 coefficients for each dataset which is a large number relative to number of observations (65). Therefore, we need to apply dimension reduction techniques first. We used the recursive feature elimination (RFE-SVM) method explained in Ref. 20 to sort the coefficients based on their classification capability. Figures 7(a)–7(c) show the classifier accuracy using only near channels measuring skin responses (2 channels, 40 bases), only far channels measuring mostly cortex and a little bit of skin responses (4 channels, 80 bases), and both near and far channels together (6 channels, 120 bases), respectively.

These results show that the classifier needs both skin and brain responses to achieve the best performance. The maximum accuracy achieved by using only near channels—skin responses—is 80% while using only far channels reaches an accuracy of 88%. The best classification accuracy (94%) is achieved by using both near and far channels. This performance happens

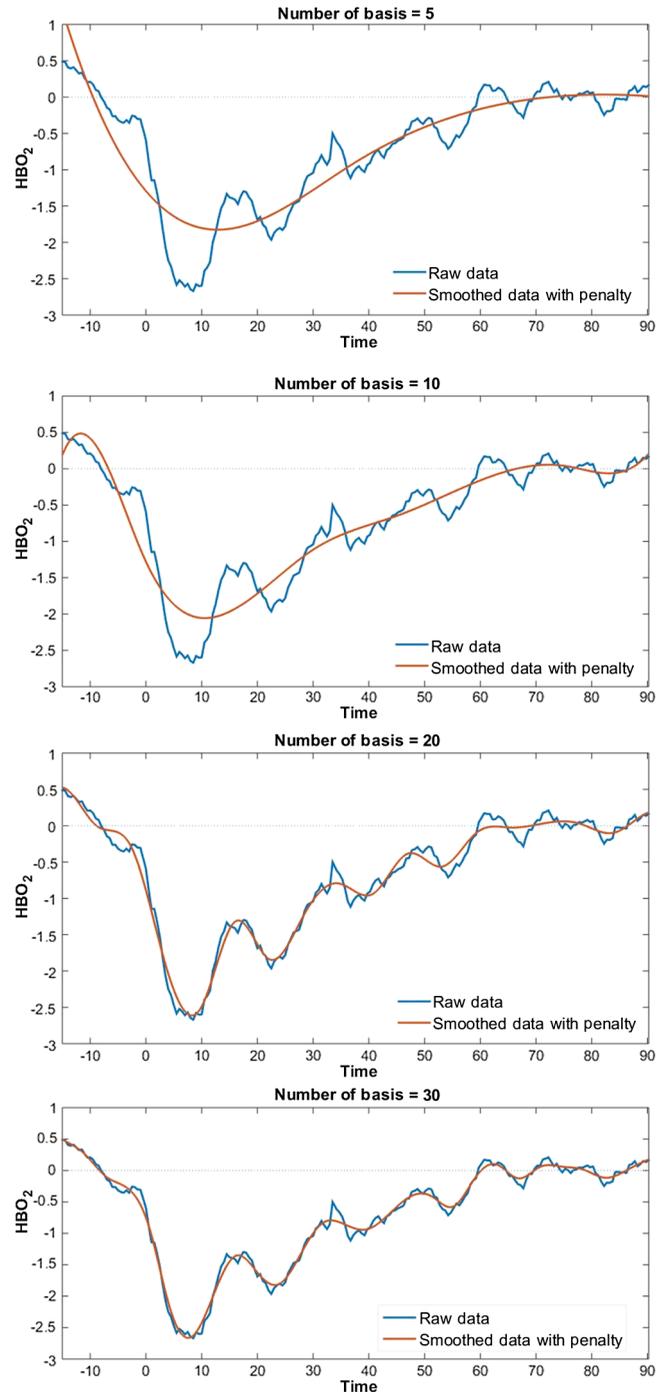
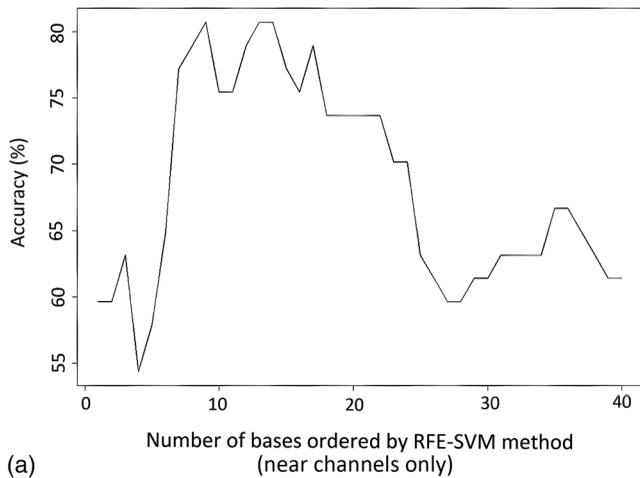


Fig. 6 The effect of choosing a different number of basis (B-spline order 4): (a) 5 bases, (b) 10 bases, (c) 20 bases, and (d) 30 bases.

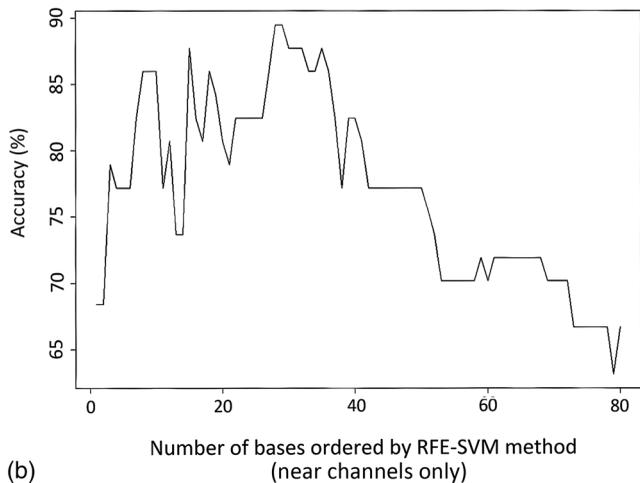
by using the first 15 bases ordered by RFE-SVM method (8 bases from far channels and 7 bases from near channels).

3.3 Clustering of NIRS Curves Using Functional Data Analysis

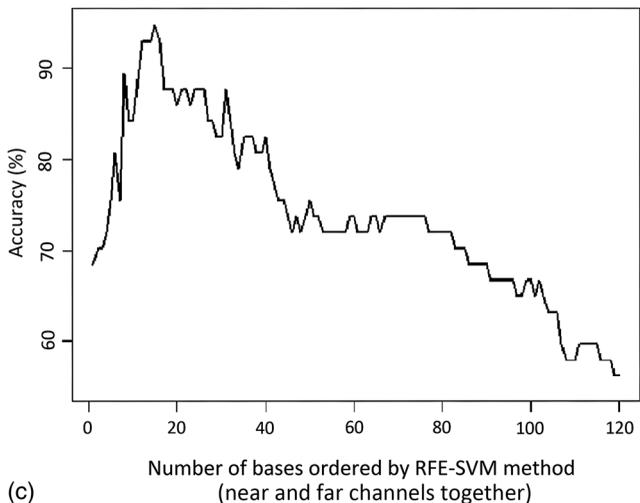
As opposed to classification which is a supervised learning method, here we use clustering which is an unsupervised technique. The clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to



(a)



(b)



(c)

Fig. 7 RFE_SVM feature selection classification accuracy with FDA data. Basis = 10, (a) using only near channels, (b) using only far channels, and (c) using all channels: First 15 coefficients which resulted in 94% accuracy are: (numbers in the parentheses show bases numbers which can be between 1 and 10, base 1 represents first 1/10 of the data, base 2 second 1/10 and so on): Far_left_Hb(10), NearleftHbO₂(10), Far_right_Hb(1), FarleftHbO₂(5), Far_left_Hb(1), NearleftHbO₂(4), Near_left_Hb(3), Far_right_Hb(7), NearleftHbO₂(2), FarrightHbO₂(4), NearrightHbO₂(7), Far_left_Hb(9), FarrightHbO₂(5), NearleftHbO₂(1), Near_left_Hb(2).

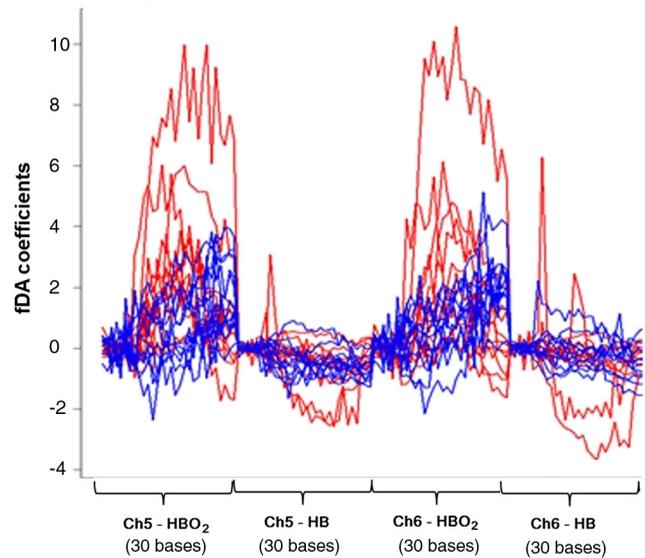


Fig. 8 HBO₂ and HB curves for channel 5 and channel 6 (far channels on the left side) using 30 bases. First 30 numbers are coefficients for Channel 5 HbO₂ signal, 31 to 60 are coefficients for Channel 5 Hb signal, 61 to 90 are coefficients for Channel 6 HbO₂ signal, and 91 to 120 are coefficients for Channel 6 Hb signal. Low-pain and high-pain signals are shown in blue and red, respectively.

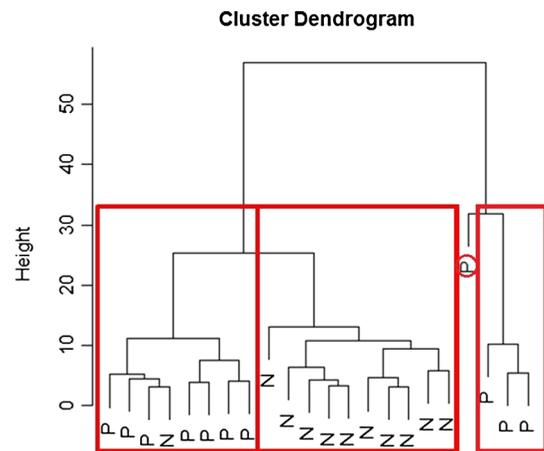


Fig. 9 Clustering of painful (P) and not-painful signals (N) based on their HBO₂ fDA curves.

those in other groups (clusters). Here, we want to cluster the data based on their curves similarity estimated by fDA coefficients.

We selected the subjects from two ends of the spectrum. The ones who reported pain scores less than five (real low pain) and the ones who could not tolerate the test for the whole 5 min (real high pain). We found that 10 bases are not enough to cluster the trials accurately and more details from signal curvatures are needed in the clustering part. Therefore, we used 30 bases to estimate the data.

Figure 8 shows a sample of fDA curves for channel 5 and channel 6 (far channels on left side) using 30 bases. High-pain and low-pain signals are shown in red and blue, respectively.

We performed a hierarchical agglomerative clustering (bottom-up clustering) analysis using a set of dissimilarities, for the observations being clustered. Initially, each observation is assigned to its own cluster and then the algorithm proceeds

iteratively, at each stage joining the two most similar clusters (according to the Ward's minimum variance method), continuing until there is just a single cluster.

The created dendrogram for channel 6 HbO_2 curves is shown in Fig. 9. Dendrogram is a graph with clusters on x-axis and height (distance) on y-axis. The height of each

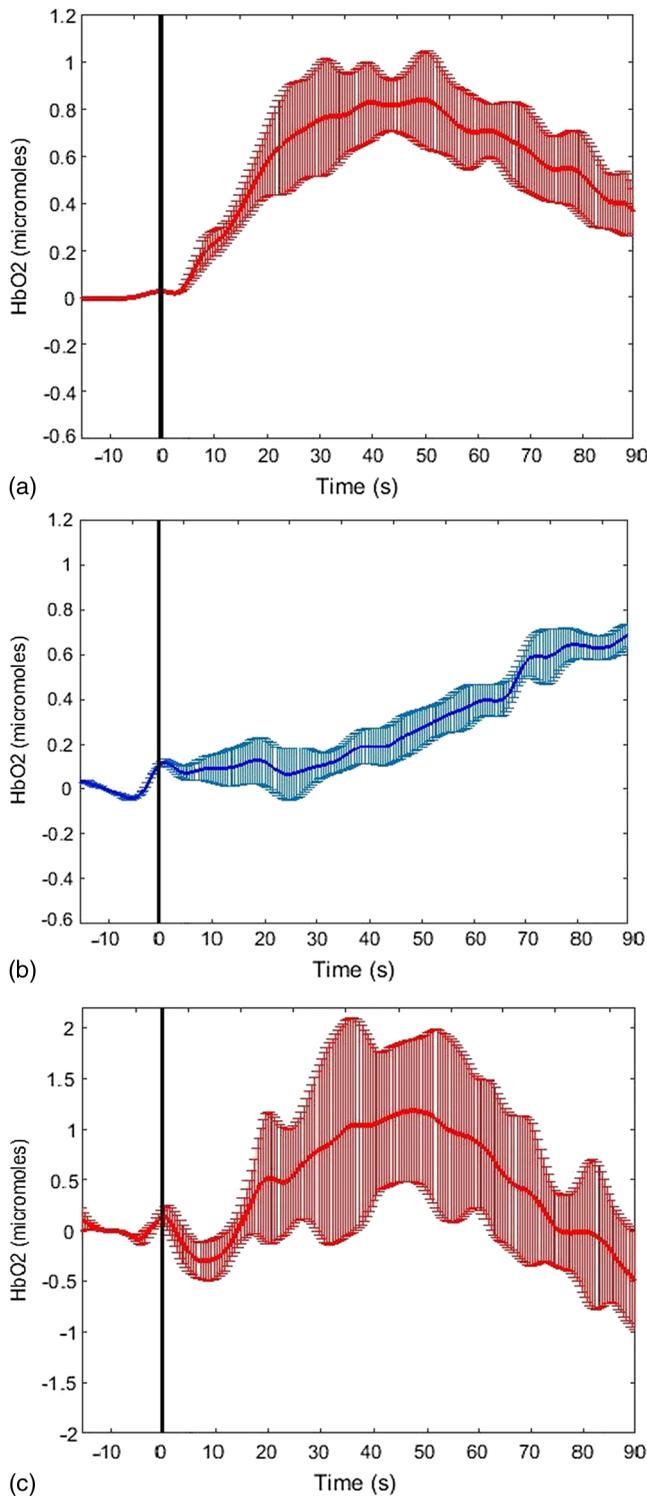


Fig. 10 Mean and error bars for channel 6 (far left) HbO_2 curves: (a) painful stimuli cluster 1, (b) nonpainful stimuli, and (c) painful stimuli cluster 2.

node (parent) is proportional to the value of the intergroup dissimilarity between its two subnodes (daughters).³¹ Then by cutting the tree at different heights (distance threshold) we can have a different number of classes. It is up to the user to decide which level (if any) actually represents a “natural” clustering in the sense that observations within each of its groups are sufficiently more similar to each other than to observations assigned to different groups at that level.

It can be seen in Fig. 9 that one observation (shown by a red circle) appears high in the tree structure that can be a cluster by itself or play the role of an outliers. One needs further study and more subjects to verify that. Other than that there are three major clusters in the data. One cluster overlaps with not-painful data completely while two other clusters cover the painful signals with an accuracy of 91.6%. In other words, it seems that there are two types of painful curves (63% of painful responses fall into pain cluster 1 and 37% into pain cluster 2) and one type of not-painful curve in the HbO_2 data.

Figure 10 shows the mean and variance of each cluster. The cluster mean can be seen as a prototype curve that represents the cluster. In other words, if HbO_2 curve of a new subject is more similar to Figs. 10(a) or 10(c) it can be considered as painful and if it is more similar to Fig. 10(b) it can be considered as nonpainful.

4 Conclusion

Despite the advances in medical imaging technology that significantly help basic science, there remains an unmet clinical need for a practical, inexpensive tool for the reliable and objective assessment of human response to pain. Advanced functional imaging modalities such as fMRI and PET scans deliver superior spatial information which comes at a high equipment and maintenance price. Therefore, they are not readily accessible for routine clinical use.

On the other hand, techniques such as fNIRS are not only noninvasive and safe but also portable, affordable and with a short set up time which makes it more clinic friendly for applications such as pain measurement or management in the clinic. Due to the ease of fNIRS measurements on the frontal region, in this paper we pursued feasibility of employing this signal for objective assessment of pain. Using fDA we found parts of the fNIRS signals that correlate to the pain and could be used to train a machine-learning system to classify signals based on different level of pains with a high accuracy. Moreover, using hierarchical clustering techniques we found that three different types of curves existed in the fNIRS HbO_2 responses in which two of them related to the high-pain stimuli and the other one related to the low-pain stimuli. This produced two prototype curves of fNIRS response to high-pain stimuli and one prototype response to low-pain stimuli.

Feature-based approaches need some features to be defined and extracted from the data prior to classification. Obviously, the accuracy and results depend on how well the features are defined and extracted. As opposed to them, we aimed another approach which tried to classify and cluster the data based on the shape of the smooth curves obtained from the data. In other words, there is no need for an expert opinion to define features prior to the classification. We used fDA procedure to convert discrete measured samples of data into continuous functions explained by a linear combination of some basis functions. This method,

combined with an RFE-SVM model selection algorithm, reaches an accuracy of 94% using an SVM classifier.

In addition to classification, we were able to cluster high-pain and low-pain signals correctly based only on the shape of their HBO₂ signals. We observed two different clusters of high-pain signal and one cluster of low-pain signal.

We used subject-level leave-one-out cross validation, meaning that we set aside all trials of a subject prior to feature selection and training and we repeat this process for all the subjects. Although the leave-one-out cross validation is known to have the least bias in estimating the true error of the model, however its effect on the generalizability of the results should be considered for new datasets.

We believe that our proposed approaches here are one step toward the goal of establishing a clinical method for objective assessment of pain because:

1. Our approach provided trial-by-trial predictions of pain level from fNIRS measurement meaning that each data was collected only once from the subject as opposed to averaging trials over repeated measurements. Moreover, all the subjects' data are used individually to evaluate the classifier accuracy through leave-one-out cross validation.
2. In addition to the inherent advantages of fNIRS for clinical applications, we used fNIRS measurements only from the forehead which makes the process of data collection very fast, easy, and clinical friendly.

Further refinement of proposed methods, including incorporating more datasets and employing other noxious stimuli (e.g., electrical and mechanical), is required to make the fNIRS technique a powerful clinical tool for pain assessment.

Acknowledgments

We would like to thank Dr. Zeinab Barati for her key role in experiment design and collection of the data presented in this paper.

References

1. C. A. Porro et al., "Temporal and intensity coding of pain in human cortex," *J. Neurophysiol.* **80**(6), 3312–3320 (1998).
2. L. Becerra et al., "Diffuse optical tomography activation in the somatosensory cortex: specific activation by painful vs. non-painful thermal stimuli," *PLoS One* **4**(11), e8016 (2009).
3. L. Becerra et al., "Diffuse optical tomography of pain and tactile stimulation: activation in cortical sensory and emotional systems," *NeuroImage* **41**(2), 252–259 (2008).
4. C. H. Lee et al., "Analysis for distinctive activation patterns of pain and itchy in the human brain cortex measured using near infrared spectroscopy (NIRS)," *PLoS One* **8**(10), e75360 (2013).
5. A. Yennu et al., "A preliminary investigation of human frontal cortex under noxious thermal stimulation over the temporomandibular joint using functional near infrared spectroscopy," *J. Appl. Biobehav. Res.* **18**, 134–155 (2013).
6. R. Re et al., "Multichannel time domain fNIRS mapping of cortical activation and superficial systemic responses during neuromuscular electrical stimulation," *Proc. SPIE* **8804**, 880404 (2013).
7. Z. Barati et al., "Hemodynamic response to repeated noxious cold pressor tests measured by functional near infrared spectroscopy on forehead," *Ann. Biomed. Eng.* **41**(2), 223–237 (2013).
8. Z. Barati, *Monitoring and Analysis of Hemodynamic Response to Cold Noxious Stimuli Using Functional Near Infrared Spectroscopy*, Drexel University, Philadelphia, Pennsylvania (2013).
9. A. Marquand et al., "Quantitative prediction of subjective pain intensity from whole-brain fMRI data using Gaussian processes," *NeuroImage* **49**(3), 2178–2189, (2010).
10. J. E. Brown et al., "Towards a physiology-based measure of pain: patterns of human brain activity distinguish painful from non-painful thermal stimulation," *PLoS One* **6**(9), e24124 (2011).
11. T. D. Wager et al., "An fMRI-based neurologic signature of physical pain," *N. Engl. J. Med.* **368**(15), 1388–1397 (2013).
12. Z. Barati, I. Zakeri, and K. Pourrezaei, "Functional data analysis view of functional near infrared spectroscopy data," *J. Biomed. Opt.* **18**(11), 117007 (2013).
13. M. Cope et al., "Methods of quantitating cerebral near infrared spectroscopy data," *Adv. Exp. Med. Biol.* **222**, 183–189 (1988).
14. V. N. Vapnik, *Statistical Learning Theory*, A Wiley-Interscience Publication, John Wiley & Sons, Inc., New York (1998).
15. C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.* **21**(3), 273–297 (1995).
16. B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Fifth Annual Workshop on Computational Learning Theory (COLT '92)*, p. 144 (1992).
17. J. P. Vert, K. Tsuda, and B. Schölkopf, "A primer on kernel methods," in *Kernel Methods in Computational Biology*, MIT Press, Cambridge, Massachusetts (2004).
18. F. Jäkel, B. Schölkopf, and F. A. Wichmann, "A tutorial on kernel methods for categorization," *J. Math. Psychol.* **51**(6), 343–358 (2007).
19. A. E. Isabelle Guyon, "An introduction to variable and feature selection," *J. Mach. Learn. Res.* **3**, 1157–1182 (2003).
20. I. Guyon et al., "Gene selection for cancer classification using support vector machines," *Mach. Learn.* **46**, 389–422 (2002).
21. J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*, Springer, New York (2005).
22. S. Graves, G. Hooker, and J. Ramsay, *Functional Data Analysis with R and MATLAB*, Springer, New York (2009).
23. C. D. Boor et al., *A Practical Guide to Splines*, Applied Mathematical Sciences (Book 27), Springer-Verlag, New York (2001).
24. D. Ruppert, "Selecting the number of knots for penalized splines," *J. Comput. Graph. Stat.* **11**(4), 735–757 (2002).
25. G. W. Peter Craven, "Smoothing noisy data with spline functions," *Numerische Mathematik* **31**(4), 377–403 (1978).
26. H. Ung et al., "Multivariate classification of structural MRI data detects chronic low back pain," *Cerebral Cortex* **24**(4), 1037–1044 (2014).
27. A. Bozkurt and B. Onaral, "Safety assessment of near infrared light emitting diodes for diffuse optical measurements," *Biomed. Eng. Online.* **3**(1), 9–9 (2004).
28. A. Bozkurt et al., "A portable near infrared spectroscopy system for bedside monitoring of newborn brain," *Biomed. Eng. Online* **4**(1), 29–29 (2005).
29. S. C. Bunce et al., "Functional near-infrared spectroscopy—an emerging neuroimaging modality," *IEEE Eng. Med. Biol. Mag.* **25**(4), 54–62 (2006).
30. W. W. Downie et al., "Studies with pain rating scales," *Ann. Rheum.* **37**(4), 378–381 (1978).
31. R. T. Trevor Hastie and J. Friedman, *The Elements of Statistical Learning*, Springer Series in Statistics, Springer, New York (2013).

Ahmad Pourshoghi received his PhD in biomedical engineering from the School of Biomedical Engineering, Sciences, and Health Systems at Drexel University in 2015. His research interests include brain imaging and biooptics, biostatistics, applications of statistical machine learning to health sciences, and medical device development.

Issa Zakeri received his PhD in statistics from the University of Illinois at Urbana-Champaign. He is a professor in the Department of Epidemiology and Biostatistics at Dornsife School of Public Health at Drexel University. His research interests include biostatistics, longitudinal data analysis, functional data analysis, applications of statistical machine learning to health sciences, and time series analysis.

Kambiz Pourrezaei received his PhD in electrical engineering from the Rensselaer Polytechnic Institute in 1982. Since then he has been a faculty member at Drexel University. Presently he is a professor in the School of Biomedical Engineering, Sciences, and Health Systems at Drexel University. His research interest is in the areas of biooptics and medical device development.