Neurophotonics

Investigation of data-driven optical neuromonitoring approach during general anesthesia with sevoflurane

Gabriela Hernandez-Meza
Meltem Izzetoglu
Ahmet Sacan
Michael Green
Kurtulus Izzetoglu

Investigation of data-driven optical neuromonitoring approach during general anesthesia with sevoflurane

Gabriela Hernandez-Meza,[a,*] **Meltem Izzetoglu,**[a] **Ahmet Sacan,**[a] **Michael Green,**[b] **and Kurtulus Izzetoglu**[a]
[a]Drexel University, School of Biomedical Engineering, Science and Health Systems, Philadelphia, Pennsylvania, United States
[b]Drexel University College of Medicine, Hahnemann University Hospital, Department of Anesthesiology, Philadelphia, Pennsylvania, United States

Abstract. Anesthesia monitoring currently needs a reliable method to evaluate the effects of the anesthetics on its primary target, the brain. This study focuses on investigating the clinical usability of a functional near-infrared spectroscopy (fNIRS)-derived machine learning classifier to perform automated and real-time classification of maintenance and emergence states during sevoflurane anesthesia. For 19 surgical procedures, we examine the entire continuum of the maintenance–transition–emergence phases and evaluate the predictive capability of a support vector machine (SVM) classifier during these phases. We demonstrate the robustness of the predictions made by the SVM classifier and compare its performance with that of minimum alveolar concentration (MAC) and bispectral (BIS) index-based predictions. The fNIRS-SVM investigated in this study provides evidence to the usability of the fNIRS signal for anesthesia monitoring. The method presented enables classification of the signal as maintenance or emergence automatically as well as in real-time with high accuracy, sensitivity, and specificity. The features local mean HbTotal, std HbO$_2$, local min Hb and HbO$_2$, and range Hb and HbO$_2$ were found to be robust biomarkers of this binary classification task. Furthermore, fNIRS-SVM was capable of identifying emergence before movement in a larger number of patients than BIS and MAC. © 2017 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.NPh.4.4.041408]

Keywords: anesthesia monitoring; functional near-infrared spectroscopy; machine learning; cerebral hemodynamics; depth of anesthesia.

Paper 17038SSR received Mar. 18, 2017; accepted for publication Jul. 24, 2017; published online Aug. 19, 2017.

1 Introduction

The core of the work of anesthesiologists consists of integrating a variety of signals from the human body during anesthetic use. In order to do this, physicians undergo rigorous training to obtain in-depth understanding of how drugs affect the patient's awareness, responsiveness, and pain. Currently, the tool box that anesthesiologists have for evaluation of the anesthetized state of patients contains primarily peripheral measurements, such as heart rate (HR), respiration rate, oxygen saturation, and end-tidal anesthetic concentration. The clinician must then fuse the values obtained from the peripheral monitors with the clinical picture to deliver anesthetics that are best suited for each case. One of the guiding parameters that can be used when delivering volatile anesthetics is the minimum alveolar concentration (MAC). MAC is estimated from the patient's age and the end-tidal volatile anesthetic concentration. Even though MAC is a widely used method, it is known that at MAC 1.00 only 50% of the population will not respond to painful stimuli and during nonsteady-state conditions the end-tidal estimates may lag.[1] Furthermore, the MAC value does not account for nonvolatile agents administered during surgery that could have an effect on pain and consciousness.[1] Since the anesthesiologist tool box does not contain information on the effects of the anesthetics on the brain, a method that directly measures brain activity would provide important data to help clinicians deliver individualized anesthetic care.

Early studies using positron emission tomography visualized decrease in the cerebral metabolic rate of glucose and the cerebral metabolic rate of oxygen during the use of general anesthesia.[2–6] This decrease was observed in the cortical areas of the brain including the prefrontal cortex (PFC), which formed the basis for the development of anesthesia monitoring technology.[5–10] Furthermore, in the PFC specifically, anesthetics can alter the hemodynamics by affecting the regional cerebral blood flow and neurovascular coupling.[6–9] Today, the most widely used anesthesia monitor is the bispectral index (BIS™ Medtronic-Covidien, Dublin, Ireland). However, with routine use estimated at only 1.8%, it has not gained widespread acceptance due to variability of its performance with opioid use and its inability to predict emergence with sufficient lead-time to take a preventive action.[11–14] The low-rate use of these devices may be influenced by the results of studies that have shown that BIS has a reduced performance in the presence of opioids, did not improve the quality of postoperative recovery, did not improve early- and intermediate-term survival, or decrease the incidence of intraoperative awareness.[12,15–22] In order to address these deficiencies, as well as enhance the existing capabilities, recent studies in anesthesia neuromonitoring have investigated the use of alternative modalities, particularly functional near-infrared spectroscopy (fNIRS), to evaluate the effects of anesthetics on the PFC.[23] fNIRS is a safe, noninvasive, and portable optical method that can be used to monitor activity within the cortical areas of the human brain. By deploying specific wavelengths of light, continuous wave fNIRS provides relative measurements of oxygenated (HbO$_2$) and deoxygenated (Hb) hemoglobin that are associated with the human brain activity changes.[24–26]

*Address all correspondence to: Gabriela Hernandez-Meza, E-mail: gabriela. hernandezmeza@icahn.mssm.edu.edu

Recent studies following patients during the transition from anesthesia maintenance to emergence have shown that fNIRS can measure significant changes in the concentrations of Hb and HbO$_2$ from the PFC associated with the transition between anesthetized states.[23] To date, four independent studies examining the effects of the anesthetic agent sevoflurane on the PFC have shown that during the transition from maintenance to emergence, a decrease in global mean Hb can be observed.[27–30] Furthermore, significant decreases in global mean HbO$_2$ and HbTotal, derived as Hb + HbO$_2$, have also been reported.[29,30] The global mean changes in Hb, HbO$_2$, and HbTotal found in these studies were relative to a baseline obtained prior to the induction of anesthesia. However, changes relative to a pre-induction baseline period can be influenced by events that are unrelated to the anesthetic regime such as changes in position and others described in the literature.[23] In order to avoid such confounders, local features describing the changes in the local Hb, HbO$_2$, and HbTotal signals can be quantified, instead, as features of the transition from maintenance to emergence during sevoflurane washout.[30] In this local analysis approach, each minute of data is normalized relative to a 5-s baseline at the beginning of the minute. The following features describing the hemodynamic response were then found to be associated with the transition from maintenance to emergence: the standard deviations, minimum and range of local Hb and HbO$_2$.[30] Using these features, it was found that the emergence state is associated with higher signal variability. These hemodynamic changes in the PFC measured by fNIRS features are associated with the competing effects of vasoconstriction from decreasing sevoflurane concentration and increased oxygen demand from increasing brain activity during emergence.[30] Furthermore, machine learning algorithms were able to make use of these features to develop a model for automated classification of the maintenance and emergence phases with a high accuracy even with a limited number of sample points.[30] Quantification of the usability and robustness of an fNIRS-based machine learning model for anesthesia evaluation still needs further investigation in a larger number of samples.

This study focuses on investigating the clinical usability of an fNIRS-derived machine learning classifier to perform automated and real-time classification of maintenance and emergence states. While previous studies represented the maintenance and emergence phases of a surgical operation as single data points,[27–30] here we consider the entire continuum of the maintenance–transition–emergence phases and evaluate the predictive capability of a support vector machine (SVM) classifier during these phases. We demonstrate the robustness of the predictions made by the SVM classifier and compare its performance with that of MAC and BIS index-based predictions.

2 Materials and Methods

Patients undergoing elective limb or abdominal surgery at the Hahnemann University Hospital in Philadelphia, Pennsylvania, were enrolled in this observational study. The study protocol and statements of informed consent were approved by the Institutional Review Board of Drexel University. The study was conducted with the subjects' understanding and written informed consent. A total of 50 patients were enrolled in the study, out of which a group of 19 patients receiving the anesthetic agent sevoflurane exclusively from maintenance until emergence were selected for this study, to be consistent with the anesthetic used during the procedure. The criteria for exclusion included: incomplete records of physiologic data, technical issues during data recording, excessive number of artifacts in the fNIRS recording, use of desflurane for anesthesia maintenance, use of a combination of sevoflurane and desflurane during maintenance, and use of nitrous oxide. In our 19 patient group all patients were American Society of Anesthesiologist (ASA) class I (normal health patient) and class II (patient with mild systemic disease), 5 male (26%) and 14 female (74%), had a mean age of 42.49($\pm$11.42) and a mean weight of 198.96($\pm$56.17) pounds. This study follows the experimental design, data collection, and fNIRS description method published by Hernandez-Meza et al. 2017.[30]

2.1 Experimental Design

For the purpose of identifying the times of interest in this study, the perioperative time course was divided into preinduction, induction, maintenance, and light anesthesia/emergence phases, as shown in Fig. 1. During the preinduction phase, 13 patients received midazolam to decrease anxiety. Oxygenation with 100% oxygen was then administered. All patients were induced intravenously with a bolus of lidocaine (87.0 $\pm$ 21.5 mg), fentanyl (2 to 3 mcg/kg), propofol (1.5 mg/kg) and, for those requiring intubation, rocuronium (50 mg). Maintenance was defined as the time following intubation until the end of wound closure. During this time, anesthesia was maintained with sevoflurane. Drug dosing was left to the discretion of the anesthesiologist. Light anesthesia/emergence followed the end of wound closure with the start of sevoflurane washout until the first movement of the patient (Fig. 1).

2.2 Data Collection

Standard physiological monitoring of HR, blood pressure (BP), and peripheral oxygen saturation (SpO$_2$) began during the preinduction phase. At this time, the fNIRS sensor pad was placed on the right side of the forehead while the BIS sensor was placed on the left side as shown in Fig. 2. The right-side preference for the fNIRS sensor was studied and validated by previous findings where greater differences in the deoxygenated hemoglobin levels were found on the right side of the forehead.[27,28] After obtaining a preinduction baseline, the fNIRS signal was recorded throughout the procedure with a sampling rate of 2 Hz. Simultaneous data acquisition from continuous wave fNIRS and BIS monitor is not expected to affect either signal as multiple studies have used both modalities [e.g., electroencephalogram (EEG) and fNIRS, BIS and fNIRS] in combination

General anesthesia phases					
Preinduction	**Induction**		**Maintenance**	**Light anesthesia/ emergence**	
Data collection starts	Intravenous induction	Sevoflurane anesthesia starts	Surgical procedure	Sevoflurane washout	Observation of purposeful movement

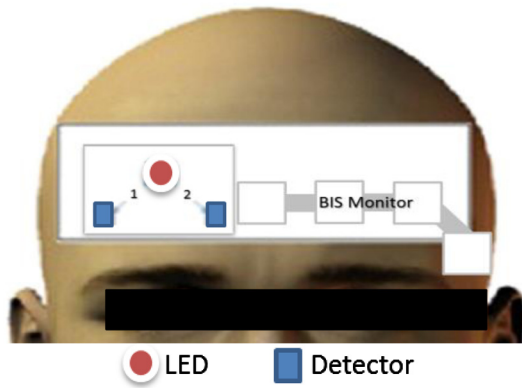**Fig. 1** Timeline of the study in relation to phases of general anesthesia.

Fig. 2 fNIRS and BIS sensor data were simultaneously collected in this study.

light and the optical sensitivity of the hemoglobin molecules. The noncommercial continuous wave system used in this study was developed by the Optical Brain Imaging Lab at Drexel University based on the technique demonstrated by Chance et al. 1993.[24,37] The sensor pad contained one LED light source and two photodetectors spaced 2.5 cm apart. The pad is composed of a flexible circuit board encased in foam (Fig. 2). Attachment to the forehead, at the Fp2 position of the international 10–20 system,[38] was achieved using medical grade double-sided adhesive. The fNIRS system captures measurements at the photodetectors at a rate of 2 Hz. During each 500-ms cycle, the LED emits light at wavelengths of 730 and 850 nm for 33 ms sequentially. The optical power of the LEDs is dependent on the wavelength of light and the current. The power emitted at either wavelength is ~12 mW, at a current of 20 mA, and has average irradiance of 41 mW/cm^2.

without interference.[27,28,31–34] Furthermore, after starting the fNIRS data acquisition during the preanesthesia phase, we confirmed that the signal quality index (SQI) output of the BIS monitor indicated high EEG signal reliability. Events of interest during the procedure were time-stamped in the fNIRS signal for accuracy in analyses.

Vital signs and gas concentrations were recorded automatically from the anesthesia machine at a rate of 0.1 Hz. The vital signs recorded included the BP, HR, pulse rate, SpO$_2$, BIS, end-tidal carbon dioxide concentration (EtCO$_2$), and temperature. The gas analyzer data included inspired and end-tidal concentrations of oxygen (O$_2$) and sevoflurane. The data from the anesthesia machine and fNIRS were recorded simultaneously with synced clocks. The data from the anesthesia machine were extracted using custom optical character recognition software and downsampled to 2 Hz during offline data analysis. The end-tidal concentration of sevoflurane (EtSev) was converted to MAC and adjusted for the subject's age.[35,36] All data were processed with MATLAB® (The Mathworks, Inc., Natick, Massachusetts).

2.3 fNIRS Description

fNIRS is a noninvasive optical neuroimaging modality that measures the relative concentrations of Hb and HbO$_2$ molecules in a given tissue using the transparency of tissues to near-infrared

2.4 fNIRS Data Processing

fNIRS data were obtained continuously from preinduction until emergence. Although the fNIRS sensor pad contained two photodetectors, the data from channel-2 suffered from ambient light interference due to failure of the adhesive related to the curvature of the forehead and the need for space for the BIS sensor. Therefore, only the data from channel-1 was used for the analysis presented here. The first step in the processing of the signal was to apply a low pass filter with 0.1 Hz cutoff frequency to extract the hemodynamic signal changes without confounding artifacts from the respiration.

Since our aim was to evaluate the robustness of classification of the anesthetized state as maintenance versus emergence using the fNIRS signal in a machine learning model, we divided the anesthetic time course into three periods of interest: maintenance (M), transition (T), and emergence (E), see Fig. 3. The maintenance phase starts at the beginning of wound closure, when the patient has been returned to supine position, no electrocautery is being used, the abdomen has been deflated after laparoscope removal, and EtSev alone was adequate for surgical anesthesia. The selection of this time period allowed us to reduce the number of possible confounders. The transition period starts when the inspired concentration of sevoflurane is reduced in preparation for emergence. Emergence was defined as the 60-s period preceding the first purposeful movement.

Time periods of interest		
Maintenance (M)	**Transition (T)**	**Light Anesthesia/ Emergence (E)**
Starts when the following conditions are met: *Patient is in supine position for the remainder of the procedure *Sevoflurane concentration is adequate for surgical procedure *No electrocautery is used *Laparoscope is removed and pneumoperitonium deflated Ends with the beginning of sevoflurane washout	Starts with the beginning of sevoflurane washout Ends 60 s before observation of movement	Starts 60 s before observation of movement Ends when movement is observed

Fig. 3 Description of the time periods studied: maintenance (M), transition (T), and light anesthesia/ emergence (E).

One-minute long epochs were extracted from the three periods of interest M, T, and E. A window of 60 s with a step size of 60 s was used to extract the epochs. The number of epochs for each patient varied based on the type of procedure and the surgeon's timing.

After each epoch was extracted, the changes in Hb and HbO_2 were calculated relative to a 5-s local baseline (calculated at the beginning of each epoch) using the modified Beer–Lambert law.[39,40] Our previous study showed significant changes in features from the fNIRS signal [standard deviation (std) of Hb and HbO_2, the local minimum (min) of Hb and HbO_2, and the range of Hb and HbO_2] during the transition from maintenance to emergence.[30] Furthermore, these features, along with the local mean HbTotal, were previously found to be good predictors for the classification of single maintenance and emergence samples.[30]

Since the aim of this study is to evaluate the robustness of an fNIRS-based classifier for the differentiation of maintenance and emergence during the anesthetic time course, and to develop evidence for the clinical usability of this signal, the following features extracted from each epoch were used to form a feature vector: std of HbO_2, local mean HbTotal, local min Hb and HbO_2, and range of Hb and HbO_2. The extracted features were assessed in an SVM classifier in order to investigate their ability to serve as robust and reliable biomarkers for depth of anesthesia and were compared to BIS and MAC classification. All data were processed with MATLAB$^®$ (The Mathworks, Inc. Natick, Massachusetts). Feature plots were prepared with SPSS v.24 (IBM Corporation, Armonk, New York).

2.5 Classification Method

Classification using machine learning allows us to build a model from a set of training examples with known categories, in order to predict the unknown category of a new observation.[41] In this study, the two categories of interest are maintenance and emergence, forming a binary classification problem. The machine learning analysis was implemented using MATLAB 2015 Statistics and Machine Learning Toolbox (The Mathworks, Inc., Natick, Massachusetts).

To develop our model, we use SVM classifiers trained with features from the fNIRS signal (fNIRS-SVM) in a supervised learning paradigm. To train the model, the training set is $P = \{(X_i, y_i), i = 1, \ldots, N\}$, which contains the feature vectors $X_i = x_{i,1}, \ldots, x_{i,n}$ and the target categories $y_i \in \{-1, 1\}$ for $i = 1, \ldots, N$; where N is the number of samples and n is the number of features. The training set used to develop the model included $n = 6$ features (local mean HbTotal, local min Hb and HbO_2, std HbO_2, range Hb and HbO_2) and $N = 38$ samples (two from each patient, representing 19 samples from the first epoch of maintenance and 19 from emergence). By having an equal number of samples from each category and each patient, we prevent classifier bias toward a given class or patient. Using the training set, the SVM algorithm maps the features into a multidimensional space where a hyperplane that achieves category division with the largest margin can be computed as the target function $f(X_i)$. This function represents the decision boundary between the categories and can be used to predict the class of a feature vector input of unknown category. Due to the nonlinear nature of the data, a radial basis function kernel was used to map the features to a higher dimensional feature space in order to determine $f(X_i)$. The parameters of the model are selected to maximize crossvalidated prediction accuracy of the test data. More information about the SVM classification method can be found in the literature.[41–44]

In order to test the generalization performance of the classifier, a leave-one-patient-out crossvalidation method was employed. The data set is divided so that the training set excludes one patient's data. The classifier is trained on the data from the remaining 18 patients and the classification performance is evaluated using the test set of the left-out patient. This is repeated 19 times, leaving one of the patients out for testing at each time. This allows us to obtain a measure of the generalization capacity of the fNIRS-SVM to correctly classify the data from a patient that was not used to train the algorithm.

The test set included all samples from the maintenance class that were extracted from the epochs in period M, the transition period extracted from T, and the emergence class extracted from period E. The true label of the transition period that occurs during sevoflurane washout is unknown. Therefore, it was classified as either maintenance or emergence in order to obtain a measure of classification performance during this dynamic time period and to quantify the number of epochs before movement that would fall under the emergence category for each patient. Because our data set is composed of real surgical cases, the number of epochs of maintenance and transition varied between patients. Using leave-one-patient-out crossvalidation, classification of maintenance and emergence was tested on 229 maintenance samples, 246 transition samples, and 19 emergence samples from 19 patients.

Using the method outlined above, SVM models were trained to evaluate the classification of maintenance and emergence using the six features outlined in Sec. 2.4. We evaluated overall accuracy, sensitivity, specificity, positive predictive value or precision (PPV), and negative predictive value (NPV) for periods M and E, because these are the only time periods where there is some certainty of the anesthetized state of the patient. The transition period T was classified during testing and the results are also presented as the number of continuous minutes before movement that were classified as emergence, in order to obtain a measure of how far back in time emergence can be predicted using the fNIRS features in our sample set. However, the true category of this time period is difficult to assign due to the dynamic nature of this phase.

The classification of maintenance and emergence using BIS was performed in accordance to the manufacturer's indications (Table 1).[27] For each epoch of BIS data, the mean was calculated and maintenance was assigned when the mean BIS < 60 and emergence was assigned when the mean BIS ≥ 60. For MAC, classification was performed by calculating the mean MAC over each epoch. An MAC ≤ 0.34 was classified as emergence, since this number has been stated as the MAC awake for sevoflurane in the literature.[35]

To evaluate the classification performance of fNIRS, BIS, and MAC, we computed the overall accuracy of classification, sensitivity, specificity, PPV, and NPV during maintenance and emergence. Furthermore, sensitivity versus 1-specificity (false-positive rate) was plotted to obtain the receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC). AUC was used to evaluate the discriminative power of the classifier, as this number approaches one the classifier performs without error. When AUC ≤ 0.5, the performance of the classifier is worse than random classification. The ROC curves of fNIRS-SVM, BIS, and MAC were compared by

Table 1 BIS range guidelines.

BIS index range	Anesthetized state
100 to 80	Awake
80 to 60	Responds to loud commends or mild prodding/shaking
60 to 45	General anesthesia
60 to 46	Deep hypnotic state
20 to 0	Burst suppression

Table 2 Demographic and procedural information ($N = 19$).

Description	Mean (±std. or percent of cases)
Age (yrs.)	42.49 (±11.47)
Weight (lbs.)	198.96 (±56.17)
Height (in.)	65.63 (±4.33)
Hemoglobin (g/dL)	12.86 (±1.54)
Hematocrit (%)	38.36 (±4.25)
Anesthetic risk (cases)	7 ASA I and 12 ASA II
Sex female	14 (73.7%)
Race Caucasian	11 (57.9%)
Duration (min)	115.6 (±56.8)
Propofol dose (mg)	182.5 (±25.8)
Midazolam	13 cases (68.4%)
Fentanyl	18 cases (94.7%)
Sevoflurane (MAC)	1.15 (±0.09)
Rocuronium	9 cases (47.4%)
Ephedrine	1 case (5.3%)
Morphine	2 cases (10.5%)
MAP < 90 mm Hg	4 cases (21.1%)
HR < 50 bpm	2 cases (10.5%)
$\Delta EtCO_2 > 50\%$	1 case (5.3%)
Procedures	8 minimally invasive (42%)
	11 open (58%)

Friedman test for repeated measures. For *post-hoc* analysis, we employed the Wilcoxon signed-ranks test for repeated measures in order to establish significant differences. The significance level was set at $p < 0.05$. The statistical evaluation was carried out with SPSS v.24 (IBM Corporation, Armonk, New York).

3 Results

A total of 19 patients receiving sevoflurane for anesthesia maintenance were evaluated in this study. Demographic and procedural statistics are presented in Table 2 as the mean and standard deviation or as the number of cases and percent of all cases. No procedures required aggressive interventions and all the patients recovered consciousness upon emergence.

The training set was composed of the first epoch of maintenance and the emergence epoch of each patient. The local mean HbTotal, std HbO$_2$, local min Hb and HbO$_2$, and range of Hb and HbO$_2$ were calculated for these two epochs forming a 6×38 training set, as explained in Sec. 2.5. Over the same time points, a mean BIS and MAC data set were also derived. The training set values per feature and MAC are presented as mean and 95% confidence interval of 19 patients in Fig. 4. These specific features were identified from a set of 11 statistical features using a backward selection strategy for the classification of maintenance and emergence during sevoflurane anesthesia.[30] Figure 5 shows the mean HR, SpO$_2$, EtCO$_2$, and mean arterial BP (MAP) for the same epochs that compose the training set. The measurements of MAP were recorded by BP cuff noninvasively every 5 min and, therefore, represent only estimates for actual values during the presented epochs. For the test set, the same six features were calculated for each epoch of the maintenance, transition, and emergence phases, forming a 6×494 matrix. The number of epochs per patient depended on the type of procedure.

Using a leave-one-patient-out crossvalidation method, fNIRS-SVM was trained and tested using six fNIRS features. Performance was assessed as explained in Sec. 2.5. The confusion matrix of the SVM classifier presented in Fig. 6 shows that it achieves an overall accuracy of 94.8% in the classification of the 248 M and E epochs. This fNIRS-SVM classifier was found to have a sensitivity of 94.8%, specificity of 94.7%, PPV of 99.5%, and NPV of 60% (Table 3).

By examining the classification results of the transition phase, we can gain some information on the predictive capacity of fNIRS to detect changes associated with light anesthesia/emergence before the patient's first movement. Table 4 presents the number of continuous minutes before the first movement that were classified as emergence. When we look further into
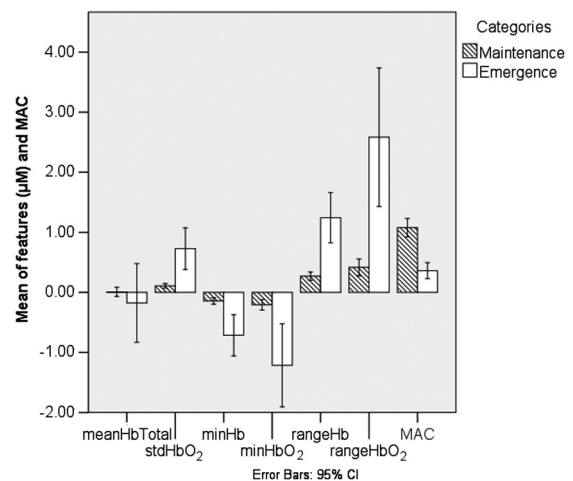

Fig. 4 Mean and 95% confidence interval of six features used in the training set (mean HbTotal, std HbO$_2$, min Hb, min HbO$_2$, range Hb, and range HbO$_2$) and MAC for 19 patients represented in the training set with categories maintenance and emergence.
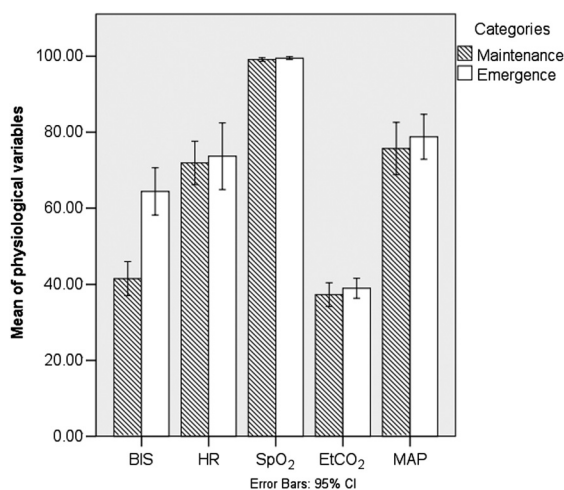
Fig. 5 Mean and 95% confidence interval of MAC, BIS index, mean HR (bpm), mean SpO_2 (%), mean $EtCO_2$ (%), and mean MAP (mmHg) for 19 patients represented in the training set with categories of maintenance and emergence.

		Target class	
		Maintenance (positive)	Emergence (negative)
Output class	Maintenance (positive)	True positive = 217	False positive = 1
	Emergence (negative)	False negative = 12	True negative = 18

Fig. 6 fNIRS-SVM confusion matrix.

the classification of the transition period, however, we find that several cases contained intermittent epochs classified as emergence. Figures 7(a)–7(c) show examples of such cases. Figure 7(d) shows continuous classification of emergence over the 3 min that precedes emergence. The true-positive rate versus False-positive rate is presented as the ROC curves in Fig. 8.

Classification of maintenance and emergence using the data from the BIS sensor was performed following the manufacturer's guidelines.[12] Each epoch of BIS data was averaged and assigned a maintenance label when mean BIS < 60; the results are presented as a confusion matrix in Fig. 9. Using this method,

Table 3 Classification results for fNIRS-SVM, BIS, and MAC.

Parameter	Classifier type		
	fNIRS-SVM	BIS	MAC
Specificity (%)	94.7	68.4	52.6
Sensitivity (%)	94.8	97.4	100.0
Accuracy (%)	94.8	95.6	96.4
PPV (%)	99.5	97.4	96.2
NPV (%)	60.0	72.2	100.0
AUC	0.965	0.842	0.761

Table 4 Number of continuous epochs/minutes before the observation of movement that were classified as emergence with fNIRS-SVM, BIS, and MAC.

Timing of emergence detection	Number of patients per classifier type		
	fNIRS-SVM	BIS	MAC
Not detected	1 (5%)	6 (32%)	9 (47%)
1 min before emergence	10 (53%)	1 (5%)	2 (11%)
2 min before emergence	4 (21%)	1 (5%)	1 (5%)
3+ min before emergence	4 (21%)	11 (57%)	7 (37%)

BIS obtained an overall classification accuracy of 95.6%, sensitivity of 97.8%, specificity of 68.3%, PPV of 97.4%, and NPV of 72.2% (Table 3). MAC was classified as emergence when mean $MAC \leq 0.34$. The overall accuracy of MAC was 96.4%, sensitivity 100%, specificity 52.6%, PPV 96.2%, and NPV 100% (Table 3). Table 4 presents the number of continuous minutes before the first movement that were classified as emergence for MAC and BIS. The ROC curves for the classification of BIS and MAC are shown in Fig. 8. Figure 10 shows the confusion matrix for the classification of MAC. The fNIRS-SVM AUC was found to be 0.965, superior to that of BIS and MAC. The Friedman Omnibus test comfirmed significant differences among the performances of fNIRS-SVM, MAC, and BIS $[X^2(2) = 122.435, p < 0.001]$. *Post-hoc* analysis by Wilcoxon signed-ranks test showed differences in performance between fNIRS-SVM and MAC ($Z = -7.074$, $p < 0.001$), fNIRS-SVM, and BIS ($Z = -6.060$, $p < 0.001$), and but not for MAC and BIS ($Z = -2.047$, $p = 0.41$).

4 Discussion

Delivery of anesthetics remains without a gold standard method for quantifying its effects directly on the brain. Previous fNIRS studies during general anesthesia with sevoflurane concluded that the global mean Hb and HbO_2 decreased as the subjects transitioned from anesthesia maintenance to light anesthesia and finally to emergence.[27–30] The global means of the signal, however, can be easily confounded by changes in the position of the body during surgery.[23] The search for fNIRS-derived features that describe the hemodynamic effects of the transition from maintenance to emergence led to the realization that emergence after sevoflurane washout was associated with variability in the local fNIRS signal, which could be modeled using the local mean HbTotal, std of HbO_2, local min Hb and HbO_2, and the range of Hb and HbO_2 calculated over 1-min long epochs.[30] Physiologically, the increased variability in these hemodynamic parameters during emergence reflects the competing effects of increased vasoconstriction and increased cerebral metabolic rate that occur during sevoflurane washout.[30] On the other hand, the physiological signals routinely measured in the operating room, such as HR, SpO_2, and $EtCO_2$, have been found to be unreliable as a measure of depth of anesthesia;[45,46] hence, they have not been included in this analysis.

In this study, we expand on previous efforts to perform gross classification of maintenance and emergence states during general anesthesia with sevoflurane in relatively healthy and young individuals by examining the performance of
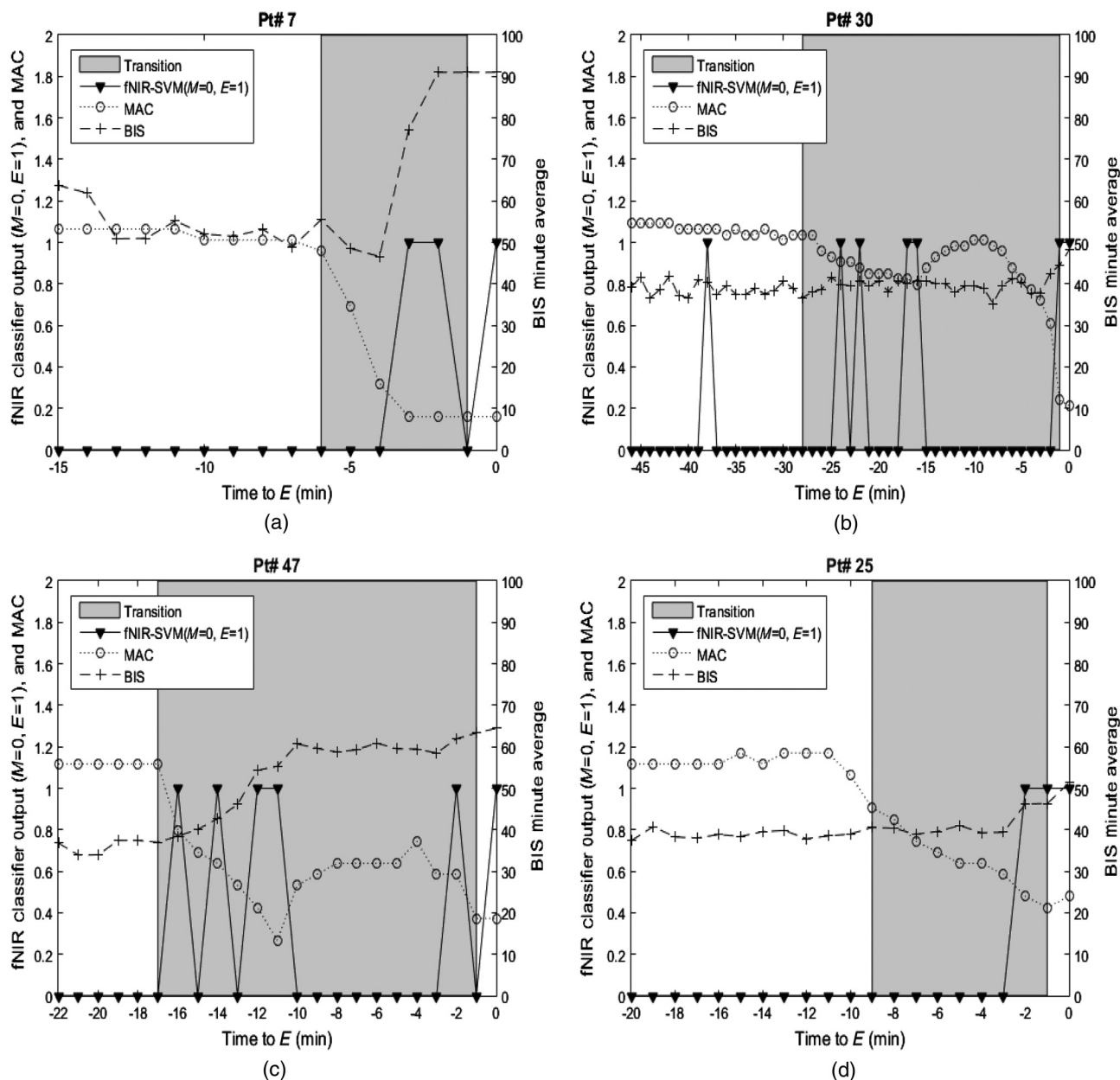
Fig. 7 (a–d) Four examples of classification results with fNIRS-SVM along with mean BIS and mean MAC versus time to emergence. The transition phase is shown in light gray. Examples "a," "b," and "c" show intermittent classification of emergence during maintenance. Example "d" shows continuous classification of emergence using fNIRS-SVM during the 3-min preceding movement/emergence.

fNIRS-SVM classification over the entire continuum of the maintenance–transition–emergence phases in a 19 patient cohort. Furthermore, we compare the classification using fNIRS features to BIS and MAC. In this study, we did not expect to find interference between fNIRS and EEG. Simultaneous, data acquisition of fNIRS and EEG has been performed successfully in previous applications including in the operating room using the BIS monitor.[27,28,31–34] Furthermore, we used the SQI indicator to confirm the reliability of the EEG signal acquired by the BIS monitor and did not observe indications of a low quality signal. However, we did encounter space constraints when placing the fNIRS alongside the BIS sensor in patients with smaller forehead space and adhesion issues due to the curvature of the forehead, which

limited our data acquisition to one channel at the Fp2 position. Future studies would benefit from a more compact and flexible fNIRS sensor pad.

For the classification task, we trained an SVM classifier with a supervised learning paradigm using the local mean HbTotal, std HbO_2, local min Hb and HbO_2, and range Hb and HbO_2 as features and using a leave-one-patient-out crossvalidation method. The training set was composed of one sample of maintenance and one sample of emergence for each patient. Classification was then tested on a total of 229 maintenance and 19 emergence samples. The fNIRS-based SVM model was found to perform well, with AUC = 0.965, and to have high accuracy (94.8%), sensitivity (94.8%), specificity (94.7%), and PPV (99.5%). In 18 out of 19 patients, emergence was detected
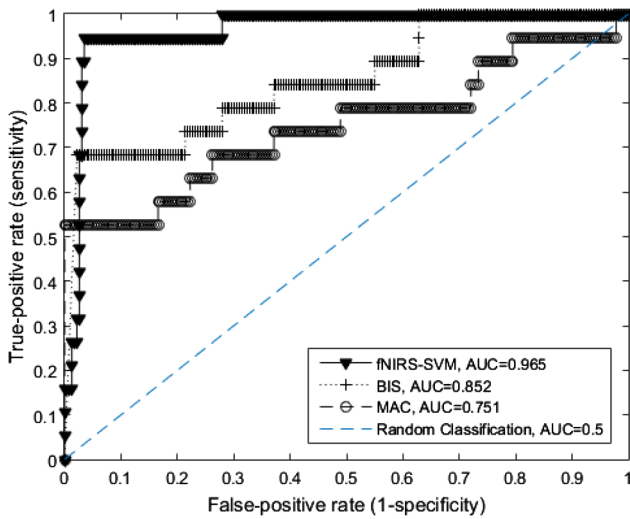
Fig. 8 ROC curves for fNIRS-SVM, BIS, and MAC.

	Target class-BIS	
	Maintenance (positive)	Emergence (negative)
Maintenance (positive)	True positive = 224	False positive = 6
Emergence (negative)	False negative = 5	True negative = 13

Fig. 9 Confusion matrix for classification of BIS.

	Target class-MAC	
	Maintenance (positive)	Emergence (negative)
Maintenance (positive)	True Positive = 229	False positive = 9
Emergence (negative)	False Negative = 0	True negative = 10

Fig. 10 Confusion matrix for classification of MAC.

successfully at least 1 min before purposeful movement. However, the NPV was low (60%). These results on 248 samples validate the performance observed by Hernandez-Meza et al. 2017 on 38 samples.[21]

Classification of the transition period allowed us to examine the number of continuous minutes that were classified as emergence in the 19 patient samples. For most patients (10/19), classification of emergence with fNIRS-SVM occurred 1 min before movement. In eight patients, classification of emergence occurred at least 2 min before movement. We hypothesize that the number of continuous minutes classified as emergence before the first purposeful movement is dependent on the rate of sevoflurane washout, the anesthetic concentration during maintenance and the rate at which the patient metabolizes sevoflurane. During analysis, however, it was observed that, in some cases, emergence was assigned intermittently to some epochs during the transition phase [Figs. 7(a)–7(c)]. Example Fig. 7(a) shows a possible false positive in the second to last minute before emergence, which could indicate that additional biomarkers for this classification task may be needed to increase the reliability of the results or that the binary labels offer insufficient resolution

for some patients. In example Fig. 7(b), we observe intermittent classification of emergence during transition, then with an increase in MAC we observe maintenance classification that continues until the 2-min preceding emergence. Figure 7(c) shows intermittent classification of emergence at the beginning of the transition phase, when MAC is decreasing. Then MAC increases again and we obtain maintenance classification until the 3 min before emergence where we again see intermittent emergence classification accompanied by decreasing MAC. Example Fig. 7(d) shows, in contrast, continuous classification of emergence during the 3 min that preceded movement. In these examples, it can be seen that fNIRS-SVM is sensitive to changes in the end-tidal anesthetic concentration, but other features or classification methods may be needed for higher reliability during the transition phase. The ideal number of minutes for emergence detection is not an agreed upon metric in the anesthesiology community, although we estimate that a warning at least 1 min before emergence will enable the clinician to take preventive action. However, the goal of our study was to determine if fNIRS was capable of better performance than the currently available methods. Given that the specificities of BIS and MAC (68.4% and 52.6%, respectively) are lower than fNIRS-SVM (94.7%), we conclude that fNIRS monitoring could offer a significant improvement in the detection of emergence over the currently available methods.

Classification with BIS and MAC was performed using cut-off values established by the manufacturing company and the literature, respectively.[27,35] The results of classification with BIS and MAC can be found in Tables 3 and 4. From these results, it was observed that fNIRS was more prone to false negatives during maintenance 12 of 229 time points evaluated, whereas BIS had five and MAC had none. On the other hand, emergence was not detected in 1 out of 19 cases 1 min before movement with fNIRS-SVM, in contrast to 6 out of 19 cases for BIS and 9 out of 19 cases for MAC. BIS was found to have high sensitivity but low specificity compared to the fNIRS-SVM. Similarly, MAC was found to have lower specificity than fNIRS.

The BIS index is a proprietary algorithm that utilizes the power spectrum of the EEG signal, and perhaps also values from an electromyogram, to output an index that estimates the awareness of the patient.[47] While the BIS index has high accuracy and sensitivity, its inability to detect emergence in 6/19 patients in the minute before movement indicates that this may be an unreliable metric to deliver individualized anesthetic care. However, since the electrical activity measured by EEG could be an indicator of the activity of the brain, future work could assess features derived from the EEG signal in combination with fNIRS. Similarly, the low specificity of MAC (52.6%) indicates that on its own it may be insufficient to predict emergence with enough lead-time for the anesthesiologist to take action. During the anesthetic procedures in our sample set, MAC was calculated from the end-tidal sevoflurane concentration and the patient's age. MAC can be a good indicator of the effect of anesthetic drugs during steady-state conditions, but during dynamic times the values may have significant lag.[1] Furthermore, the effect of other drugs such as opioids is not accounted in this number and each combination of drugs will have specific effects on analgesia.[1] In recent years, however, predictive modeling of drug concentrations in anesthesia monitors have been implemented to account for the effect of multiple drugs and these algorithms could be used to improve anesthesia monitoring in combination with fNIRS and EEG.[1] BIS and

MAC also have the advantage of providing a continuous index instead of binary classification provided by fNIRS-SVM. In this study, we provide the first evidence of the ability of fNIRS to provide gross classification of maintenance and emergence robustly in terms of a binary output. We recognize this as a limitation of this study and suggest that future work with additional data could make use of multiple SVM classifiers, in sequence, to provide finer classification by increasing the number of classes. This task, however, will require additional data in order to have sufficient examples to train the algorithm to recognize each individual class. Higher resolution for the classification task could also be achieved with other learning algorithms, such as SVM regression, given sufficient data.

In this study, we focused on evaluating the anesthetic time course after the patient had been returned to supine position and the pneuperitoneum deflated, as those events have been identified as confounders of the fNIRS signal.[23] These confounding factors along with others described in the literature[23] may reduce the performance of the fNIRS-derived features. Given these limitations, we suggest that improvements in sensitivity and specificity could be achieved by a hybrid models that combine fNIRS, EEG, and predictive modeling of drug concentrations. Such a model might more accurately follow the effects of anesthesia and prevent the detection of false positives and negatives in a given patient leading to personalized, controlled, and safe delivery of anesthetics. This level of personalization might especially benefit the pediatric and elderly populations that are particularly sensitive to the effects of anesthetics.

For this study, we limited the feature space to a previously studied and validated feature set that showed high accuracy of classification in a small number of samples.[30] During this study, we observed that maintenance can be classified accurately throughout the continuum of the maintenance–emergence phases in patients undergoing different types of procedures (abdominal and limb surgery, open and laparoscopic procedures). Furthermore, classification of emergence was reliable at least 1 min before movement was observed. However, during the transition-phase additional features, classification with multiple classifiers or regression methods may offer the resolution needed to obtain reliable results. Analysis in a larger set of patients is needed in order to increase the number of emergence samples.

5 Conclusion

The goal of this study was to evaluate the robustness of an fNIRS-SVM-based model in the classification of the continuum of the maintenance–transition–emergence phases and to compare the results to classification with BIS and MAC. The fNIRS-SVM investigated in this study provides further evidence to the usability of the fNIRS signal for anesthesia monitoring. The method presented in this study enables classification of the signal as maintenance or emergence automatically as well as in real-time with high accuracy, sensitivity, and specificity. The features local mean HbTotal, std HbO_2, local min Hb and HbO_2, and range Hb and HbO_2 are found to be robust biomarkers of this binary classification task. Furthermore, fNIRS-SVM was capable of identifying emergence before movement in a larger number of patients than BIS and MAC. The predictive capacity of the algorithm during transition periods, however, still needs further investigation. In addition, this study was conducted in a relatively small cohort of patients and these results need to be confirmed in a much larger scale. Fusion of fNIRS with other sources of data related to anesthetic depth, such as EEG, peripheral physiologic measurements, and predictive modeling of drug concentrations may improve anesthesia monitoring and lead to a clinically viable real-time anesthesia monitoring tool.

Disclosures

The authors declare that there is no conflict of interest.

References

1. J. DeCou and K. Johnson, "An introduction to predictive modelling of drug concentration in anaesthesia monitors," *Anaesthesia* **72**(Suppl. 1), 58–69 (2017).
2. M. T. Alkire et al., "Cerebral metabolism during propofol anesthesia in humans studied with positron emission tomography," *Anesthesiology* **82**(2), 393–403 (1995).
3. M. T. Alkire et al., "Positron emission tomography study of regional cerebral metabolism in humans during isoflurane anesthesia," *Anesthesiology* **86**, 549–557 (1997).
4. M. T. Alkire et al., "Functional brain imaging during anesthesia in humans: effects of halothane on global and regional cerebral glucose metabolism," *Anesthesiology* **90**(3), 701–709 (1999).
5. P. M. Patel and J. C. Drummond, "Cerebral physiology and the effects of anesthetic drugs," in *Miller's Anesthesia*, R. D. Miller, Ed., Vol. **1**, pp. 594–674, Churchill-Livingstone, Philadelphia (2010).
6. U. León-Domínguez et al., "A chronometric functional sub-network in the thalamo-cortical system regulates the flow of neural information necessary for conscious cognitive processes," *Neuropsychologia* **51**(7), 1336–1349 (2013).
7. R. A. Veselis et al., "A neuroanatomical construct for the amnesic effects of propofol," *Anesthesiology* **97**(2), 329–337 (2002).
8. E. R. John and L. S. Prichep, "The anesthetic cascade: a theory of how anesthesia suppresses consciousness," *Anesthesiology* **102**(2), 447–471 (2005).
9. W. Heinke et al., "Sequential effects of propofol on functional brain activation induced by auditory language processing: an event-related functional magnetic resonance imaging study," *Br. J. Anaesth.* **92**(5), 641–650 (2004).
10. X. Cui et al., "A quantitative comparison of NIRS and fMRI across multiple cognitive tasks," *NeuroImage* **54**(4), 2808–2821 (2011).
11. M. S. Avidan and G. A. Mashour, "II. The incidence of intraoperative awareness in the UK: under the rate or under the radar?" *Br. J. Anaesth.* **110**(4), 494–497 (2013).
12. S. C. Manyam et al., "When is a bispectral index of 60 too low? Rational processed electroencephalographic targets are dependent on the sedative-opioid ratio," *Anesthesiology* **106**(3), 472–483 (2007).
13. R. J. Gajraj et al., "Analysis of the EEG bispectrum, auditory evoked potentials and the EEG power spectrum during repeated transitions from consciousness to unconsciousness," *Br. J. Anaesth.* **80**(1), 46–52 (1998).
14. S. von Delius et al., "Auditory evoked potentials compared with bispectral index for monitoring of midazolam and propofol sedation during colonoscopy," *Am. J. Gastroenterol.* **104**(2), 318–325 (2009).
15. M. S. Avidan et al., "Prevention of intraoperative awareness in a high-risk surgical population," *N. Engl. J. Med.* **365**(7), 591–600 (2011).
16. M. S. Avidan et al., "Anesthesia awareness and the bispectral index," *N. Engl. J. Med.* **358**(11), 1097–1108 (2008).
17. G. A. Mashour et al., "Prevention of intraoperative awareness with explicit recall in an unselected surgical population a randomized comparative effectiveness trial," *Anesthesiology* **117**(4), 717–725 (2012).
18. P. S. Myles et al., "Bispectral index monitoring to prevent awareness during anaesthesia: the B-aware randomised controlled trial," *Lancet* **363**(9423), 1757–1763 (2004).
19. K. Leslie et al., "Recovery from bispectral index-guided anaesthesia in a large randomized controlled trial of patients at high risk of awareness," *Anaesth. Intensive Care* **33**(4), 443 (2005).
20. K. Leslie et al., "The effect of bispectral index monitoring on long-term survival in the B-aware trial," *Anesth. Analg.* **110**(3), 816–822 (2010).
21. M. D. Kertai et al., "Association of perioperative risk factors and cumulative duration of low bispectral index with intermediate-term mortality

after cardiac surgery in the B-unaware trial," *Anesthesiology* **112**(5), 1116–1127 (2010).

22. M. D. Kertai et al., "Bispectral index monitoring, duration of bispectral index below 45, patient risk factors, and intermediate-term mortality after noncardiac surgery in the B-unaware trial," *Anesthesiology* **114**(3), 545–556 (2011).

23. G. Hernandez-Meza et al., "Near-infrared spectroscopy for the evaluation of anesthetic depth," *BioMed. Res. Int.* **2015**, 939418 (2015).

24. A. Villringer and B. Chance, "Non-invasive optical spectroscopy and imaging of human brain function," *Trends Neurosci.* **20**(10), 435–442 (1997).

25. A. Villringer et al., "Near infrared spectroscopy (NIRS): a new tool to study hemodynamic changes during activation of brain function in human adults," *Neurosci. Lett.* **154**(1–2), 101–104 (1993).

26. P. W. McCormick et al., "Noninvasive cerebral optical spectroscopy for monitoring cerebral oxygen delivery and hemodynamics," *Critical Care Med.* **19**(1), 89–97 (1991).

27. U. Leon-Dominguez et al., "Molecular concentration of deoxyHb in human prefrontal cortex predicts the emergence and suppression of consciousness," *NeuroImage* **85**, 616–625 (2014).

28. K. Izzetoglu et al., "The evolution of field deployable fNIR spectroscopy from bench to clinical settings," *J. Innov. Opt. Health Sci.* **4**(3), 239–250 (2011).

29. Z. Liang et al., "Design of multichannel functional near-infrared spectroscopy system with application to propofol and sevoflurane anesthesia monitoring," *Neurophotonics* **3**(4), 045001 (2016).

30. G. Hernandez-Meza et al., "Investigation of optical neuro-monitoring technique for detection of maintenance and emergence states during general anesthesia," *J. Clin. Monit. Comput.* **31**, 1–17 (2017).

31. Y. Liu et al., "Neural adaptation to a working memory task: a concurrent EEG-fNIRS Study," in *Int. Conf. on Augmented Cognition*, pp. 268–280, Springer International Publishing (2015).

32. Y. Liu et al., "Towards a hybrid P300-based BCI using simultaneous fNIR and EEG," in *Int. Conf. on Augmented Cognition*, pp. 335–344, Springer Berlin Heidelberg (2013).

33. A. C. Merzagora et al., "Functional near-infrared spectroscopy and electroencephalography: a multimodal imaging approach," in *Int. Conf. on Foundations of Augmented Cognition*, pp. 417–426, Springer Berlin Heidelberg (2009).

34. M. Butti et al., "Multimodal analysis of a sustained attention protocol: continuous performance test assessed with near infrared spectroscopy and EEG," in *28th Annual Int. Conf. of the IEEE on Engineering in Medicine and Biology Society*, pp. 1040–1043, IEEE (2006).

35. E. Eger, "Age, minimum alveolar anesthetic concentration, and minimum alveolar anesthetic concentration-awake," *Anesth. Analg.* **93**(4), 947–953 (2001).

36. R. Nickalls and W. Mapleson, "Age-related iso-MAC charts for isoflurane, sevoflurane and desflurane in man," *Br. J. Anaesth.* **91**(2), 170–174 (2003).

37. B. Chance et al., "Cognition-activated low-frequency modulation of light absorption in human brain," *Proc. Natl. Acad. Sci. U. S. A.* **90**(8), 3770–3774 (1993).

38. H. Jasper, "Report of the committee on methods of clinical examination in electroencephalography," *Electroencephalogr. Clin. Neurophysiol.* **10**, 370–375 (1958).

39. M. Cope and D. Delpy, "System for long-term measurement of cerebral blood and tissue oxygenation on newborn infants by near infra-red transillumination," *Med. Biol. Eng. Comput.* **26**(3), 289–294 (1988).

40. D. Delpy et al., "Estimation of optical pathlength through tissue from direct time of flight measurement," *Phys. Med. Biol.* **33**(12), 1433–1442 (1988).

41. V. Vapnik, *Estimation of Dependences Based on Empirical Data*, Springer Science and Business Media Inc., New York (2006).

42. N. Aronszajn, "Theory of reproducing kernels," *Trans. Am. Math. Soc.* **68**(3), 337–337 (1950).

43. N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, Cambridge (2000).

44. S. Theodoridis and K. Koutroumbas, "Pattern recognition and neural networks," in *Machine Learning and Its Applications*, G. Paliouras, V. Karkaletsis, and C. Spyropoulos, Eds., pp. 169–195, Springer, Berlin (2001).

45. K. B. Domino et al., "Awareness during anesthesia: a closed claims analysis," *Anesthesiology* **90**(4), 1053–1061 (1999).

46. N. Moerman, B. Bonke, and J. Oosting, "Awareness and recall during general anesthesia: facts and feelings," *Anesthesiology* **79**, 454–464 (1993).

47. Z. Hajat, N. Ahmad, and J. Andrzejowski, "The role and limitations of EEG-based depth of anaesthesia monitoring in theatres and intensive care," *Anaesthesia* **72**(S1), 38–47 (2017).

Gabriela Hernandez-Meza is a biomedical engineering PhD candidate at Drexel University, Philadelphia, Pennsylvania, USA. She is interested in exploring the use of data obtained from noninvasive sensors to improve human performance and clinical monitoring. Her focus is on the use of machine learning algorithms to derive meaningful outcomes that can guide decision making.

Biographies for the other authors are not available.