HHS Public Access

Author manuscript

IEEE Trans Biomed Eng. Author manuscript; available in PMC 2018 December 01.

Published in final edited form as:

IEEE Trans Biomed Eng. 2017 December ; 64(12): 2988–2996. doi:10.1109/TBME.2017.2756870.

Quantifying and Characterizing Tonic Thermal Pain across Subjects from EEG Data using Random Forest Models

Vishal Vijayakumar,
Department of Electrical Engineering, University of Minnesota, MN, USA

Michelle Case,
Department of Biomedical Engineering, University of Minnesota, MN, USA

Sina Shirinpour, and
Department of Biomedical Engineering, University of Minnesota, MN, USA

Bin He, IEEE [Fellow]
Department of Biomedical Engineering and the Institute for Engineering in Medicine, University of Minnesota, Minneapolis, MN-55455 USA

Abstract

Objective—Effective pain assessment and management strategies are needed to better manage pain. In addition to self-report, an objective pain assessment system can provide a more complete picture of the neurophysiological basis for pain. In this study, a robust and accurate machine learning approach is developed to quantify tonic thermal pain across healthy subjects into a maximum of ten distinct classes.

Methods—A random forest model was trained to predict pain scores using time-frequency wavelet representations of independent components obtained from electroencephalography (EEG) data, and the relative importance of each frequency band to pain quantification is assessed.

Results—The mean classification accuracy for predicting pain on an independent test subject for a range of 1–10 is 89.45%, highest among existing state of the art quantification algorithms for EEG. The gamma band is the most important to both inter-subject and intra-subject classification accuracy.

Conclusion—The robustness and generalizability of the classifier is demonstrated.

Significance—Our results demonstrate the potential of this tool to be used clinically to help improve chronic pain treatment, and establish spectral biomarkers for future pain-related studies using EEG.

Index Terms

Electroencephalography (EEG); gamma oscillations; pain quantification; machine learning; cingulate cortex

Correspondence to: Bin He.

Pain is considered to be the product of a conscious, multidimensional response to noxious stimuli [1], [2]. The neural correlates of such a response can either reflect sensory information, or can be modulated by various psychological and social factors [3]. As pain is considered a subjective experience, the most prevalent means of measuring it has been through self-report. While there are established and effective treatments for acute pain [4], such an approach to rating pain provides an incomplete picture of the various neurological processes pertinent to both acute and chronic pain. Moreover, the capacity to effectively rate pain is limited in cases where the subject is unable to adequately express himself or herself (due to age or infirmity). Hence, developing an accurate objective pain assessment system which can generalize across a population can increase the efficacy of clinical treatments [5] by serving as an additional unbiased tool to evaluate pain, and drastically reduce the current costs of managing and treating chronic pain [6].

Many recent studies have demonstrated the involvement of several spatially distinct cortical regions in pain processing. The most consistent regions are the anterior cingulate cortex (ACC), the primary (S1) and secondary (S2) somatosensory cortices, the prefrontal cortex (PFC), and the insular cortex [7]–[9]. Due to such spatially distinct regions showing concurrent activation during pain, traditional univariate approaches [10] are not suitable for capturing the spatio-temporal dynamics of the pain response. Hence, a branch of multi-variate approaches, collectively called Multivariate Pattern Analysis (MVPA) techniques [11], [12] has been increasing in popularity over recent years [13]. Machine learning based approaches are a form of MVPA which focus on identifying patterns in data, and making predictions about future behavior of similar data [14]. Such approaches can potentially be used to solve problems associated with real-time brain decoding [15] and for selecting the best biomarkers for various acute and chronic medical conditions [16].

Electroencephalography (EEG) is a useful noninvasive tool for assessing pain responses due to its high temporal resolution, clinical convenience, and low cost of setup and maintenance [17]. Many studies have shown observable pain-related cortical activity using EEG [18]–[20]. Studies have also been carried out which effectively demonstrate the activity of certain frequency bands in the pain response, mainly the alpha band [21], [22], and more recently, the gamma band [23]–[25]. However, most of these studies have investigated the effects of phasic pain stimuli. To mimic responses triggered during chronic pain [26], there are a growing number of studies using tonic thermal stimuli [27]–[31].

The aim of the present study was to build a robust and generalizable classification model for predicting various levels of tonic thermal pain from EEG data recorded in a group of subjects. This approach was used to train and test the performance of classifier models both across subjects and within a subject. Using the constructed model, the frequency bands that consistently contribute most to the accuracy of the model were analyzed.

II. METHODS

A. Subjects

Twenty-five healthy subjects participated in this study with a median age of twenty-four years old, eleven of whom were female. The study was approved by the Institutional Review

Board at the University of Minnesota and conducted in conformity with the Declaration of Helsinki. Subjects were screened so that none of the healthy volunteers had any history of mental illness. All subjects provided written informed consent before participating in the study.

B. Experimental Procedure

The experimental procedure consisted of four phases: pain threshold and tolerance determination, the tonic pain phase, and a control phase, performed on each subject in that order. A tonic thermal stimulus was delivered to subjects using a thermal stimulator (CHEPs, Medoc, Ramat Yishai, Israel). A custom-built analog pain-rating device was used to record pain scores. Subjects could rate pain by indicating the level of pain intensity using the device and aided by a color bar with a gradient from green (indicating no pain and non-painful heat) to red (indicating high pain). A line was placed on the color bar to indicate pain threshold, and subjects were asked to rate any painful thermal stimuli above the bar. Before beginning the experiment, subjects were given time to practice with the device by rating thermal stimuli delivered to their left forearm. In this training session, subjects experienced different levels of the thermal stimulus ranging from non-painful heat to moderately high pain temperatures. This practice allowed subjects to accurately determine how to rate their pain on the provided color bar.

The phases of the experimental procedure are illustrated in Fig. 1. To determine pain threshold, the heat delivered by the thermode was increased at a constant rate of 0.5°C/second with a safety limit of 50°C. Seven trials were conducted with an inter-trial interval of 10 seconds. During this phase, the subjects were asked to pay attention to the stimulus and use a response unit to indicate the instant the stimulus changed from heat to pain, after which the temperature was brought back to baseline at a rate of 5°C/second. The rate of increase of temperature for tolerance testing was the same as that for the threshold test, with the same number of trials. However, for the tolerance test, subjects were instructed to continuously rate their pain with the analog rating device using their left hand, and to use the response unit to indicate the temperature at which the painful stimulus became intolerable. The threshold and tolerance tests were used to determine the minimum and maximum temperatures to be administered to the subject during the tonic pain condition to ensure the desired range of pain scores was obtained for each subject and to make sure no one was administered an unbearable pain stimulus. The thermode was placed on the ventral left forearm for both the tests, while the right arm remained motionless.

The tonic pain phase was split into three conditions: low pain, medium pain, and high pain. The thermode was shifted to the right ventral forearm for tonic pain trials, which subjects were asked to keep motionless while they rated their pain with the device using their left hand. The order of the condition was decided at random, and the intensity of the stimulus for each condition was decided based on the subject's threshold and tolerance trials. The subjects rated their pain responses using the analog device and the color bar. For the low pain condition, the subject was administered heat a degree higher than the median threshold response of the seven trials. During high pain, the administered temperature was half a degree lower than the median tolerance response of the seven trials. The temperature

administered was constant to eliminate the possibility of saliency or awareness of temperature changes being reflected in the EEG recording. Subjects were not aware of the stimulus paradigm so that there was no inherent bias on how they would rate their pain. The only instruction given to the subjects during this phase was to try to accurately rate their pain intensity based on the stimulus received. The low and medium pain conditions consisted of two trials of a constant temperature lasting four minutes each, with a resting period of three minutes between each trial. Due to the intensity of the high pain condition, the trials were shortened to four trials of one minute each, with a resting period of one minute between trials. During rest, subjects were instructed to either stare at a fixed point on the screen, or observe a silent video clip of an aquarium to constrain their streams of consciousness.

The visual control phase was used to account for the visual, attentional and motor components of all the trials to ensure that EEG components from these sources during the tonic pain phase do not adversely influence the machine learning algorithm [30], [32], [33]. During this phase, subjects were asked to use the pain rating system to rate the height of a continuously shifting point displayed on the screen, while the temperature of the thermode was maintained at body temperature (37°C). The height of the point was determined by the temporally inverted pain rating of one of the previous low/medium trials chosen at random, resulting in the trial lasting four minutes.

C. EEG Recording and Pre-processing

During all phases in the procedure, EEG data were recorded at 1 kHz using 64 channel Neuroscan system (Compumedics, NC). The impedance was kept below 5KΩ for all conditions. The pain rating was collected at a sampling rate of 10Hz.

During the analysis of each pain trial, the initial and final 10 seconds of EEG data were discounted to eliminate portions of the EEG signal with high variability and noise due to inconsistencies in initial and final sensations caused by the stimulus. These variations were more likely to reflect awareness of a drastic change in the stimulus rather than actual pain rating scores. The analysis was performed in MATLAB and Python using native toolboxes and those developed by EEGLAB [34].

The EEG signal was down sampled to 256 Hz and band-pass filtered between 1 Hz and 100 Hz to remove any muscle contraction artifacts and voltage drifts. Line noise at 60 Hz was filtered out using a notch filter. EEG was re-referenced to average reference. Artifact Subspace Reconstruction [35] was used to clean eye blink, muscle movement artifacts and reject bad time windows. This method works using sliding window Principal Component Analysis (PCA), which estimates the signal values of high-variance components exceeding a threshold relative to a covariance matrix calculated using 1 minute of empirically determined clean data. This mixing matrix is then used to reconstruct noisy EEG time segments and in extreme cases, reject EEG segments more than five standard deviations away from the statistics of the reference clean data. Independent component analysis (ICA) [36] is then used to reject other noisy components, primarily eye movement, cardio-ballistic artifacts, and other bad electrode sources.

D. EEG Analysis

The EEG data corresponding to the pain and the control conditions were then concatenated along with the pain ratings, with the data acquired from control conditions assigned a pain score of one. Two minutes of resting state data were also appended to enable the model to discriminate between rest and pain states. The data were then concatenated across subjects, and subjected to full rank group ICA [37]. This was done to enable multivariate analysis by focusing on the fraction of source information available at each scalp electrode. Subsequent steps were carried out separately on the time course of each independent component (IC), which was warranted since the dependency between ICs is reduced to a minimum.

The ground truth obtained was labeled as follows: all the rest conditions, including the control conditions and non-painful heat, were assigned a pain score of one, while the continuous pain responses were discretized into 'n' equal interval classes. The variable 'n' is iterated from 2 to 10, 2 corresponding to a common 'pain' class, and 10 corresponding to 9 ordinal "pain" classes, and one rest class.

A continuous wavelet transform using the Gabor wavelet was used to extract the Time-Frequency Representations (TFR) [38], owing to its excellent tradeoff between temporal and frequency resolution among existing signal analysis methods [40]. Wavelet coefficients were computed for 60 scales, corresponding to a frequency range of 2–80 Hz. Choosing a smaller scale would result in more high frequency noise than actual signal, so a maximum of 80 Hz was chosen. This resulted in computing a feature vector of 60 scales per sample, while the number of samples skipped was made equal to half the width of the smallest wavelet scale used to analyze the EEG data. This was done to minimize redundancy in the feature space, while also analyzing all available data. All such time points computed were then used as data points to train and test the classifier model.

E. Pattern Analysis using Random Forests

Random Forest (RF) models are an ensemble method of classification, where the fundamental classifier is a decision tree. A decision tree splits a dataset into several leaf nodes in an attempt to "purify" the dataset at each node, where purification implies finding the best feature in the dataset to split a node to obtain leaf nodes which have a majority class [40]. In short, RF models are created by growing many unpruned decision trees trained on a random subset of the data, and return a class for each data point that is the mode of all classes decided by all the trees. They are known for their high accuracies, efficiency on large datasets, and resistance to overtraining in the presence of sparsity as the number of trees increases by undersampling majority class labels [41]. This characteristic is relevant for the problem of pain quantification since the data corresponding to higher pain classes is relatively sparse. As the primary unit is a decision tree, the splitting rules for identifying the best feature used to split the training set forces the classifier to address all classes in an imbalanced dataset, making it a suitable choice for this study. The ensemble bootstrapping approach to quantification minimizes the trade-off between bias and variance. They also do not make any assumption about the distribution of the data or interactions between the features.

F. Performance Metrics for Quantification

Due to the non-uniform distribution of the discrete pain scores collected for this study, metrics other than classification accuracy were used to judge classifier performance during training and testing. The confusion matrices obtained during the training and testing phases were used to calculate statistics such as Balanced Classification Accuracy (BCA), F-measure (F) and the Matthews Correlation Coefficient (MCC). For a given class i in a multi-class setting consisting of N possible classes, each of the metrics are defined below for a confusion matrix C_{NxN} as:

$$\mathrm{BCA}=\frac{1}{N}\sum_{i=1}^{N}\frac{C_{\mathrm{ii}}}{\sum_j C_{\mathrm{ji}}} \quad (1)$$

$$\mathrm{Precision},\ P_i=\frac{C_{\mathrm{ii}}}{\sum_j C_{\mathrm{ji}}} \quad (2)$$

$$\mathrm{Recall},\ R_i=\frac{C_{\mathrm{ii}}}{\sum_j C_{\mathrm{ij}}} \quad (3)$$

$$F_i=\frac{2\cdot P_i\cdot R_i}{P_i+R_i} \quad (4)$$

$$\mathrm{MCC}=\frac{\sum_{k,l,m} C_{\mathrm{kk}}\,C_{\mathrm{ml}}-C_{\mathrm{lk}}\,C_{\mathrm{km}}}{\sqrt{\sum_k[(\sum_l C_{\mathrm{lk}})(\sum_{f\neq k,g} C_{\mathrm{gf}})]}\,\sqrt{\sum_k[(\sum_l C_{\mathrm{kl}})(\sum_{f\neq k,g} C_{\mathrm{fg}})]}} \quad (5)$$

These measures were computed from confusion matrices calculated for each test subject to ascertain the robustness of the classifier. The trend of the BCA, which is the mean of the precision across all classes, is analyzed across all subjects for various resolutions of the continuous pain score. The F-score is the harmonic mean of the precision (positive predictive value) and the recall (sensitivity). The MCC (5) is a robust measure of the quality of the classifier when it is trained/tested on imbalanced classes. It can be interpreted as a measure of correlation between observed and predicted classes, and has a range of [−1, 1] where a score of 0.4 (70%) or higher indicates good agreement between the observed and predicted class labels.

G. Training and testing paradigms for quantification

The efficacy of the quantification approach was tested both across subjects and within subjects. To test performance across subjects, Leave-One-Out Classification (LOOC) was

used to train the classifier across 24 subjects and test on one subject. An illustration of the training and cross-validation pipeline for LOOC is shown in Fig. 2 (I). The wavelet power spectrum of each time instant of a group IC, along with the corresponding discrete pain scores was used as data for a RF model. The training data for the LOOC case is over a million data points, while the test set is about 100,000 time points. The goal was to find the model and the corresponding group IC that contained the best representation of the pain response. Each model was trained on two-thirds of the training data using various resolutions of the continuous pain score, and cross validation was performed on the remaining one-third of the training data where the accuracy of each model was returned. The model returning the best cross-validation performance metrics, along with the spatial IC weights corresponding to the most discriminative group IC were used in the testing phase shown in Fig. 2(II).

In this phase, the EEG data collected from the remaining test subject were projected into a new subspace using the weights of the most informative group IC obtained during cross-validation. The corresponding RF model was then used to estimate the discrete pain score for each time instant of the wavelet power spectrum calculated from the projected EEG data. Statistics of model performance were then computed using the ground truth pain score for that test subject. For evaluating within-subject performance metrics, the pipeline in Fig. 2(I) was used to train each subject, where the two-thirds of the wavelet scalogram of an IC was used to train a specific model. The model with the best training performance was then tested with the remaining one-third of the corresponding IC time course.

To investigate within-subject classifier performance, two-thirds of the subject's EEG data were used to compute the ICs and individual RF models were trained on the corresponding wavelet time courses. The statistics of classifier performance (BCA, F-score, and MCC) were used to select the most informative IC using ten-fold cross validation. The corresponding spatial map was then used to project the remaining one-third of the EEG data to the component subspace, and the wavelet time course of this test data was used to test classifier performance. This process was repeated thrice to utilize all the EEG data for that subject.

In addition to predicting the pain score, analysis was also performed on the frequency spectrum on the most accurate test ICs for each subject to determine the frequency band most relevant to classification. Consider the feature set $F = \{f_1, f_2, \ldots, f_M\}$ where there are M features per data point. Let the decision set across all trees in the RF classifier be denoted as $D = \{d_1, d_2, \ldots, d_N\}$, where d_i is the pain score predicted by the classifier, and N is the number of decision trees in the forest. To determine the importance of a feature j within the testing set, the values of this feature were then randomly permuted within its range across all data points in the test set. The modified test set is then projected down the unchanged decision tree. Let the new decisions be denoted as $D^j = \{d_1^j, d_2^j, \ldots, d_N^j\}$. The number of votes cast for the correct class is subtracted from that of the unperturbed decision tree. The magnitude of this deviation is the raw importance score, denoted by $R^j = \{r_1^j, r_2^j, \ldots, r_N^j\}$, where

$$r_i^j = d_i^j \oplus d_i, j = 1 \ldots M, i = 1 \ldots N \quad (6)$$

The final score is then computed by dividing the mean of the set R^j by its standard deviation. The feature importance score C for a feature j is thus given as

$$C^j = \frac{E(R^j)}{\sigma(R^j)} \quad (7)$$

Hence, the final set $C = \{C^1, C^2, \ldots, C^M\}$ provides the feature importance scores for every feature independent of the other in the test dataset. The features with the highest scores are found to contribute most to the classification accuracy.

III. RESULTS

A. Results of Classification Analysis

The statistics for the LOOC paradigm performed across subjects are presented in Fig. 3. The horizontal axis shows the number of intervals that the continuous pain score is discretized into, while the vertical axis shows the classification accuracy. The BCA for 2-way classification is 95.33±0.6%, and is 89.45±1.05% for 10-way classification, where chance BCA is 10%. While the slope indicates expected deterioration in the performance of the classifier as the ground truth intervals are made finer, the low rate of decrease demonstrates the scalability of the classifier to finer resolutions of continuous pain scores.

In addition to BCA calculated for different resolutions of the pain score across subjects, the confusion matrix computed for worst case (10-way) classification for one test subject is shown in Fig. 4. Most misclassifications occurred within the neighborhood of the true class label, and on average, 73% of the total misclassifications occurred within the lower half of the pain scale and 62% of all errors occurred within a 2-neighborhood distance from the true class. Hence, it is seen that misclassifications are more prevalent in the lower half of the pain scale than the upper half, and most of the wrongly predicted scores are about the true classes. From a clinical perspective, this is less of a risk compared to a larger misclassification rate on higher pain scores, and across larger pain intervals.

For a test subject, the BCA of each component is shown in Fig. 5. The components are sorted in descending order of BCA performance from left to right for 10-way classification, where the dotted red line corresponds to chance BCA. The spatial topographies of the three best and worst performing components are also depicted. As described in [42], muscle artifacts show the strongest activity in pre-frontal and temporal electrodes, with spikes in higher frequency bands. It is seen from Fig 5 that ICA succeeds in isolating such type of activity. These ICs also perform consistently poorly across subjects for various resolutions of the pain score, showing that classification accuracy is not informed by any muscle activity. Alternatively, these ICs can be removed manually prior to classification to speed up computation time.

The performance metrics computed for the confusion matrices in the 10-way classification for both the cross-subjects and within-subject paradigms are shown in Fig. 6. All performance metrics are higher when the classifier is trained and tested within subjects, as expected. The F-score for each subject is averaged across all classes. The F-score for both the inter-subject and the intra-subject paradigms (91.3±3.4% and 94.2±2.6%, respectively) is slightly higher than the BCA (89.45±1.07% and 93.26±1.49%, respectively). The MCC is 0.466 (73.3%) for the inter-subject paradigm, while it is 0.609 (80.46%) for the intra-subject paradigm, indicating strong agreement between the ground truth and predicted pain scores in the presence of class imbalance.

B. Contributions of Frequency Bands to Pain Quantification

Estimation of frequency importance is performed on all subjects for ten-way classification. The frequency spectrum was divided into the delta (1.5 Hz – 4 Hz), theta (4 Hz – 7 Hz), alpha (7 Hz – 15 Hz), beta (15 Hz – 30 Hz), and the gamma (30 Hz – 80 Hz) bands. Fig. 7 shows a boxplot depicting the normalized contribution scores for all subjects. It is seen that the gamma band contributes the most to classification accuracy, and that the alpha band has the highest variance across all subjects. The alpha band also showed a reduction in relative importance as the resolution of the pain score was made finer (Fig. S4). The effect of each frequency band was also investigated in a paradigm where, during classification, the frequency information was ignored from one band at a time during training to observe its effect during testing. The results illustrated in Fig. 8 show that BCA drop is the highest when the gamma band is ignored, with a 10% decrease in performance compared to the original approach. However, it is interesting to note that accuracy suffers regardless of which frequency band is ignored, showing that all frequency bands seem to be important to pain quantification.

IV. DISCUSSION

A. Improvements over Previous Methods

Table 1 compares the current quantification method to existing state of the art methods. While such studies have had a fair amount of success, most [43]–[48] examined the efficacy of a two-class version of this problem (presence vs absence, or low pain vs high pain), while others [49], [50] pursued a regression-based approach. Those studies that pursued a multi-class approach [25], have broken it down to various two-class sub-problems with a vote-based approach to arrive at a pain score. In addition to the high computational time for training these models, such voting procedures would need testing data from both low-pain and high-pain trials to arrive at a single discrete score for a test subject, which may not be always clinically available. The current model substantially improves on the state of the art, and is trained to accurately predict a maximum pain range of 1–10 based on any 2–3 second snippet of the time course of independent components derived from scalp-recorded EEG, which is clinically more convenient.

In addition, most of the past approaches using EEG have worked towards finding the most informative electrode i.e. the electrode where the quantification accuracy is the highest. While fairly accurate, such approaches may, in fact, be interpreted as univariate rather than

multivariate [13] as a decision is made by processing data from one electrode at a time. In the end, this may not be the most informative due to the distributed nature of pain processing. By projecting EEG data into another subspace in which each component contains phase-locked information from multiple electrodes, group ICA isolates patterns recurring throughout the training data, which can generalize better for larger datasets as it favors selecting an IC with a more consistent temporal pain signature. In addition, ICA also isolates EEG/muscle artifacts, which can be removed either before or after the classification pipeline upon performance analysis. Even though Support Vector Machines (SVMs) are one of the most commonly used classifiers in this field, issues with multi-class implementation and optimal parameter selection [51], [52] make RF decision trees a better alternative due to their scalability to large datasets and resistance to overfitting.

B. Spectral and Spatial Activity Most Important to Pain

The most informative IC found during testing for each subject was used to analyze inter-individual variability of the spatio-temporal pain signature in both the frequency and the topographical domains. It was found that the gamma band (31–80 Hz) was the most important in discriminating between different pain scores, and the degree of importance was consistently high from two-way classification up to (and including) 10-way classification.

The gamma band has been previously found to be important in the subjective perception of both acute and tonic pain. The sensory and attentional aspects of brain activity in response to acute pain were found to be most prevalent in the gamma band [23], [24], [30], [53], Moreover, [53] showed that the predictive power of gamma band activity is independent of the rate of repetition of the stimulus. In this study, the high degree of importance of the gamma rhythm for pain quantification at both coarse and fine resolutions of the continuous pain score was observed across subjects for the first time in tonic thermal pain stimuli. Hence, it is likely that the gamma rhythm primarily encodes reactionary aspects of nociception for the entire duration of the stimulus. As previous studies investigating acute and tonic thermal pain have highlighted the gamma rhythm as well, this would point to the gamma rhythm being the most important biomarker for thermal pain.

C. Salience

Salience, in the context of sensation and pain perception, is defined [55] to be the awareness of how much the stimulus intensity contrasts with the surrounding. It has been suggested in [56], [57] that the regions constituting the pain neuromatrix have been implicated in the processing of other auditory, visual, and tactile stimuli as well. Since non-nociceptive neurons have been observed to outnumber the nociceptive kind [58], it can be argued subjective pain ratings are based on an awareness of change of stimulus intensity [59], and may not provide information about pain processing. It has also been shown that the Default Mode Network (DMN) has significant deactivation during performance of attention-related tasks [60], [61]. This inverse relationship between the DMN and attention could also reflect attention to the task of rating pain, rather than pain processing, contributes to pain prediction scores.

In this study, attempts are made to minimize effects of salience by keeping the stimulus intensity constant. To test if attention to the task was a factor in predicting pain scores, we compared the uncertainty of our classifier during the control task (which accounted for visual and attention-related stimuli) to rest, where no task was performed. As shown in Fig. 9, there was greater uncertainty during the control task, indicating attention does play a role. However, the reduced uncertainty during pain conditions shows the classifier is confident during the pain stimulus and attention to the task is not the only factor contributing to pain prediction. However, further studies need to be conducted as to the extent of the role of non-nociceptive stimuli in pain processing.

D. Study Limitations

Pain is a sensation that is experienced and processed differently by everyone. While the initial shock of sensation is likely the same, the spatio-temporal correlates of affective and emotional processing are uncommon across individuals. While the ICs selected for the inter-subject case seem consistently accurate, there are significant spatial differences between those and the ICs chosen by the within-subject training paradigm. This shows that the uniqueness of the pain response topology diminishes as more subjects are considered for analysis, and this may account for a drop in the accuracy when more subjects are considered for training. With a high enough number of training subjects, the problem of pain prediction could potentially devolve into one of sensory discrimination.

Using mutual information in the RF model is an indirect measure of frequency importance, as no monotonic relationship between any of the frequency bands with the pain score was identified. To determine the extent of similarity between different frequency bands and the pain score at the electrode level, the best performing components were projected back to the scalp domain and the Welch power spectrum was computed for every second of the data to dynamically observe the changes in the frequency spectrum with time. Correlation analysis between any combination of frequency bands and the pain score did not yield results above significance, suggesting that the relationship between EEG data and the pain score is subtler than previously studied.

V. CONCLUSIONS

We have developed a novel noninvasive approach to the problem of pain quantification in the presence of tonic thermal pain and evaluated the excellent performance of our method in 25 human subjects with 89.45% accuracy in 10-way pain quantification. The temporal wavelet characteristics of independent components is used to characterize the EEG information corresponding to the pain score, and the most informative component is evaluated, along with observing its characteristics in the spatial and the frequency domains. The present approach outperforms existing state of the art methods for pain quantification by reliably predicting the pain score and demonstrating scalability of the approach for finer resolutions of a continuous pain scale. The present promising results suggest the generalizability of a biomarker for pain across a population. The gamma band in EEG is found to be the most important predictor of pain for all resolutions of the pain score across inter-subject and intra-

subject prediction paradigms. Changes in other frequency bands can be further explored by extending these approaches; unraveling a potential pain response mechanism.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by the U.S. National Institutes of Health under Grants HL117664, AT009263, NS096761, EB021027, OD021721, and by the U.S. National Science Foundation under Grant DGE-1069104.

References

1. Borsook D, Sava S, Becerra L. The Pain Imaging Revolution: Advancing Pain Into the 21st Century. The Neuroscientist. Apr.2010 16(2):171–185. [PubMed: 20400714]
2. Lee MC, Tracey I. Imaging pain: a potent means for investigating pain mechanisms in patients. Br. J. Anaesth. Jul.2013 111(1):64–72. [PubMed: 23794647]
3. Aslaksen PM, Myrbakk IN, Høifødt RS, Flaten MA. The effect of experimenter gender on autonomic and subjective responses to pain stimuli. PAIN. Jun.2007 129(3):260–268. [PubMed: 17134832]
4. Tracey I, Mantyh PW. The Cerebral Signature for Pain Perception and Its Modulation. Neuron. Aug. 2007 55(3):377–391. [PubMed: 17678852]
5. Gagliese L, Melzack R. Chronic pain in elderly people. Pain. Mar.1997 70(1):3–14. [PubMed: 9106804]
6. Gaskin DJ, Richard P. The Economic Costs of Pain in the United States. J. Pain. Aug.2012 13(8): 715–724. [PubMed: 22607834]
7. Apkarian AV, Bushnell MC, Treede R-D, Zubieta J-K. Human brain mechanisms of pain perception and regulation in health and disease. Eur. J. Pain. Aug.2005 9(4):463–463. [PubMed: 15979027]
8. Duerden EG, Albanese M-C. Localization of pain-related brain activation: A meta-analysis of neuroimaging data. Hum. Brain Mapp. Jan.2013 34(1):109–149. [PubMed: 22131304]
9. Case M, et al. Characterization of functional brain activity and connectivity using EEG and fMRI in patients with sickle cell disease. NeuroImage Clin. 2017; 14:1–17. [PubMed: 28116239]
10. Friston KJ, Holmes AP, Worsley KJ, Poline J-P, Frith CD, Frackowiak RSJ. Statistical parametric maps in functional imaging: A general linear approach. Hum. Brain Mapp. Jan.1994 2(4):189–210.
11. Haynes J-D, Rees G. Decoding mental states from brain activity in humans. Nat. Rev. Neurosci. Jul.2006 7(7):523–534. [PubMed: 16791142]
12. Norman KA, Polyn SM, Detre GJ, Haxby JV. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn. Sci. Sep.2006 10(9):424–430. [PubMed: 16899397]
13. Rosa MJ, Seymour B. Decoding the matrix: Benefits and limitations of applying machine learning algorithms to pain neuroimaging. Pain. May; 2014 155(5):864–867. [PubMed: 24569148]
14. Bishop, CM., Nasrabadi, NM. Pattern recognition and machine learning. Vol. 1. Springer; New York: 2006.
15. van Gerven MAJ, Kok P, de Lange FP, Heskes T. Dynamic decoding of ongoing perception. NeuroImage. Aug.2011 57(3):950–957. [PubMed: 21609771]
16. Klöppel S, Abdulkadir A, Jack CR Jr, Koutsouleris N, Mourão-Miranda J, Vemuri P. Diagnostic neuroimaging across diseases. NeuroImage. Jun.2012 61(2):457–463. [PubMed: 22094642]
17. Giacino JT, Fins JJ, Laureys S, Schiff ND. Disorders of consciousness after acquired brain injury: the state of the science. Nat. Rev. Neurol. Feb.2014 10(2):99–114. [PubMed: 24468878]
18. Bromm B, Chen ACN. Brain electrical source analysis of laser evoked potentials in response to painful trigeminal nerve stimulation. Electroencephalogr. Clin. Neurophysiol. Jul.1995 95(1):14–26. [PubMed: 7621766]

19. Valeriani M, Le Pera D, Niddam D, Chen ACN, Arendt-Nielsen L. Dipolar modelling of the scalp evoked potentials to painful contact heat stimulation of the human skin. Neurosci. Lett. Jan.2002 318(1):44–48. [PubMed: 11786221]
20. Arendt-Nielsen L, Chen ACN. Lasers and other thermal stimulators for activation of skin nociceptors in humans. Neurophysiol. Clin. Clin. Neurophysiol. Dec.2003 33(6):259–268.
21. Domnick, C., Hauck, M., Lorenz, J., Engel, A., Casey, K. [[Accessed: 10-Jan-2017]] C-fiber-related EEG-oscillations induced by laser radiant heat stimulation of capsaicin-treated skin. Journal of Pain Research. Mar.. 2009 [Online]. Available: https://www.dovepress.com/c-fiber-relatedeeg-oscillations-induced-by-laser-radiant-heat-stimulapeer-reviewed-article-JPR
22. Nir R-R, Sinai A, Raz E, Sprecher E, Yarnitsky D. Pain assessment by continuous EEG: Association between subjective perception of tonic pain and peak frequency of alpha oscillations during stimulation and at rest. Brain Res. Jul.2010 1344:77–86. [PubMed: 20460116]
23. Gross J, Schnitzler A, Timmermann L, Ploner M. Gamma Oscillations in Human Primary Somatosensory Cortex Reflect Pain Perception. PLOS Biol. Apr.2007 5(5):e133. [PubMed: 17456008]
24. Tiemann L, Schulz E, Gross J, Ploner M. Gamma oscillations as a neuronal correlate of the attentional effects of pain. Pain. Aug.2010 150(2):302–308. [PubMed: 20558000]
25. Schulz E, Zherdin A, Tiemann L, Plant C, Ploner M. Decoding an Individual's Sensitivity to Pain from the Multivariate Analysis of EEG Data. Cereb. Cortex. Jul.2011 :bhr186.
26. Huber MT, Bartling J, Pachur D, Woikowsky-Biedau Sv, Lautenbacher S. EEG responses to tonic heat pain. Exp. Brain Res. Aug.2006 173(1):14–24. [PubMed: 16552561]
27. Backonja M, Howland EW, Wang J, Smith J, Salinsky M, Cleeland CS. Tonic changes in alpha power during immersion of the hand in cold water. Electroencephalogr. Clin. Neurophysiol. Sep. 1991 79(3):192–203. [PubMed: 1714810]
28. Chen AC, Rappelsberger P. Brain and human pain: topographic EEG amplitude and coherence mapping. Brain Topogr. 1994; 7(2):129–140. [PubMed: 7696090]
29. Nir R-R, Sinai A, Moont R, Harari E, Yarnitsky D. Tonic pain and continuous EEG: Prediction of subjective pain perception by alpha-1 power during stimulation and at rest. Clin. Neurophysiol. Mar.2012 123(3):605–612. [PubMed: 21889398]
30. Schulz E, et al. Prefrontal Gamma Oscillations Encode Tonic Pain in Humans. Cereb. Cortex. Mar. 2015 :bhv043.
31. Huishi Zhang C, Sohrabpour A, Lu Y, He B. Spectral and spatial changes of brain rhythmic activity in response to the sustained thermal pain stimulation. Hum. Brain Mapp. Aug.2016 37(8): 2976–2991. [PubMed: 27167709]
32. Baliki MN, et al. Chronic Pain and the Emotional Brain: Specific Brain Activity Associated with Spontaneous Fluctuations of Intensity of Chronic Back Pain. J. Neurosci. Nov.2006 26(47):12165–12173. [PubMed: 17122041]
33. Hashmi JA, et al. Shape shifting pain: chronification of back pain shifts brain representation from nociceptive to emotional circuits. Brain. Sep.2013 136(9):2751–2768. [PubMed: 23983029]
34. Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods. Mar.2004 134(1):9–21. [PubMed: 15102499]
35. Mullen T, et al. Real-Time Modeling and 3D Visualization of Source Dynamics and Connectivity Using Wearable EEG. Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Conf. 2013; 2013:2184–2187.
36. Hyvärinen A. The Fixed-Point Algorithm and Maximum Likelihood Estimation for Independent Component Analysis. Neural Process. Lett. Aug.1999 10(1):1–5.
37. Calhoun VD, Liu J, Adalı T. A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. NeuroImage. Mar.2009 45(1, Supplement 1):S163–S172. [PubMed: 19059344]
38. Qin L, He B. A wavelet-based time–frequency analysis approach for classification of motor imagery for brain–computer interface applications. J. Neural Eng. 2005; 2(4):65. [PubMed: 16317229]

39. Sifuzzaman M, Islam MR, Ali MZ. Application of Wavelet Transform and its Advantages Compared to Fourier Transform. 2009
40. Quinlan JR. Simplifying decision trees. Int. J. Man-Mach. Stud. Sep.1987 27(3):221–234.
41. Breiman L. Random Forests. Mach. Learn. Oct.2001 45(1):5–32.
42. Goncharova II, McFarland DJ, Vaughan TM, Wolpaw JR. EMG contamination of EEG: spectral and topographical characteristics. Clin. Neurophysiol. Sep.2003 114(9):1580–1593. [PubMed: 12948787]
43. Marquand A, Howard M, Brammer M, Chu C, Coen S, Mourão-Miranda J. Quantitative prediction of subjective pain intensity from whole-brain fMRI data using Gaussian processes. NeuroImage. Feb.2010 49(3):2178–2189. [PubMed: 19879364]
44. Brown JE, Chatterjee N, Younger J, Mackey S. Towards a Physiology-Based Measure of Pain: Patterns of Human Brain Activity Distinguish Painful from Non-Painful Thermal Stimulation. PLOS ONE. Sep.2011 6(9):e24124. [PubMed: 21931652]
45. Brodersen KH, et al. Decoding the perception of pain from fMRI using multivariate pattern analysis. NeuroImage. Nov.2012 63(3):1162–1170. [PubMed: 22922369]
46. Huang G, Xiao P, Hung YS, Iannetti GD, Zhang ZG, Hu L. A novel approach to predict subjective pain perception from single-trial laser-evoked potentials. NeuroImage. Nov.2013 81:283–293. [PubMed: 23684861]
47. Kuo P-C, Chen Y-T, Chen Y-S, Chen L-F. Decoding the perception of endogenous pain from resting-state MEG. NeuroImage. Jan.2017 144:1–11. [PubMed: 27746387]
48. Misra G, Wang W, Archer DB, Roy A, Coombes SA. Automated classification of pain perception using high-density electroencephalography data. J. Neurophysiol. Feb.2017 117(2):786–795. [PubMed: 27903639]
49. Prato M, Favilla S, Zanni L, Porro CA, Baraldi P. A regularization algorithm for decoding perceptual temporal profiles from fMRI data. NeuroImage. May; 2011 56(1):258–267. [PubMed: 21296171]
50. Wager TD, Atlas LY, Lindquist MA, Roy M, Woo C-W, Kross E. An fMRI-Based Neurologic Signature of Physical Pain. N. Engl. J. Med. Apr.2013 368(15):1388–1397. [PubMed: 23574118]
51. Cawley GC, Talbot NLC. On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. J. Mach. Learn. Res. Jul.2010 11:2079–2107.
52. Sharan RV, Moir TJ. Comparison of multiclass SVM classification techniques in an audio surveillance application under mismatched conditions. 2014 19th International Conference on Digital Signal Processing. 2014:83–88.
53. Zhang ZG, Hu L, Hung YS, Mouraux A, Iannetti GD. Gamma-Band Oscillations in the Primary Somatosensory Cortex—A Direct and Obligatory Correlate of Subjective Pain Intensity. J. Neurosci. May; 2012 32(22):7429–7438. [PubMed: 22649223]
54. Rossiter HE, Worthen SF, Witton C, Hall SD, Furlong PL. Gamma oscillatory amplitude encodes stimulus intensity in primary somatosensory cortex. Front. Hum. Neurosci. Jul.2013 7
55. Näätänen R, Picton T. The N1 Wave of the Human Electric and Magnetic Response to Sound: A Review and an Analysis of the Component Structure. Psychophysiology. Jul.1987 24(4):375–425. [PubMed: 3615753]
56. Iannetti GD, Mouraux A. From the neuromatrix to the pain matrix (and back). Exp. Brain Res. Aug.2010 205(1):1–12. [PubMed: 20607220]
57. Liang M, Mouraux A, Hu L, Iannetti GD. Primary sensory cortices contain distinguishable spatial patterns of activity for each sense. Nat. Commun. Jun.2013 4:1979. [PubMed: 23752667]
58. Davis KD, Racine E, Collett B. Neuroethical issues related to the use of brain imaging: Can we and should we use brain imaging as a biomarker to diagnose chronic pain? Pain. Aug.2012 153(8): 1555–1559. [PubMed: 22464695]
59. Iannetti GD, Hughes NP, Lee MC, Mouraux A. Determinants of Laser-Evoked EEG Responses: Pain Perception or Stimulus Saliency? J. Neurophysiol. Aug.2008 100(2):815–828. [PubMed: 18525021]
60. Heine L, et al. Resting State Networks and Consciousness. Front. Psychol. Aug.2012 3

61. Lee MH, Smyser CD, Shimony JS. Resting-State fMRI: A Review of Methods and Clinical Applications. Am. J. Neuroradiol. Oct.2013 34(10):1866–1872. [PubMed: 22936095]

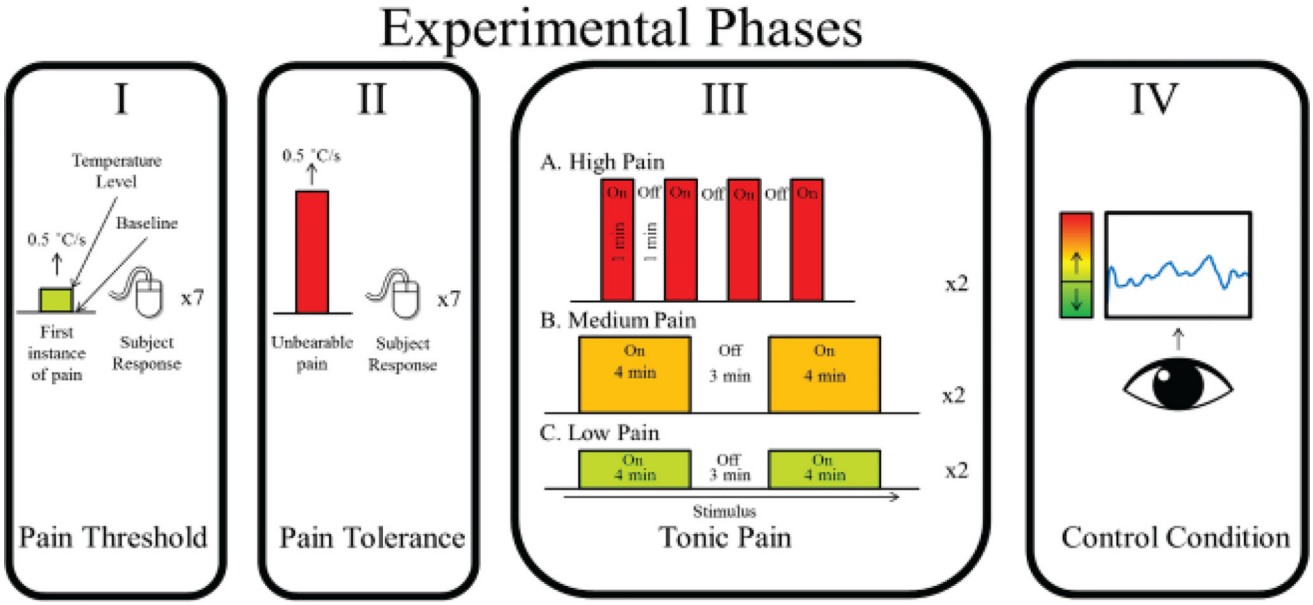

Fig. 1.

Stages of the experimental protocol: (I) Threshold determination. (II) Tolerance determination. (III) Illustration of the stimulus time course for all pain conditions. (IV) Control condition to account for various sensory and attentional stimuli during the experiment.

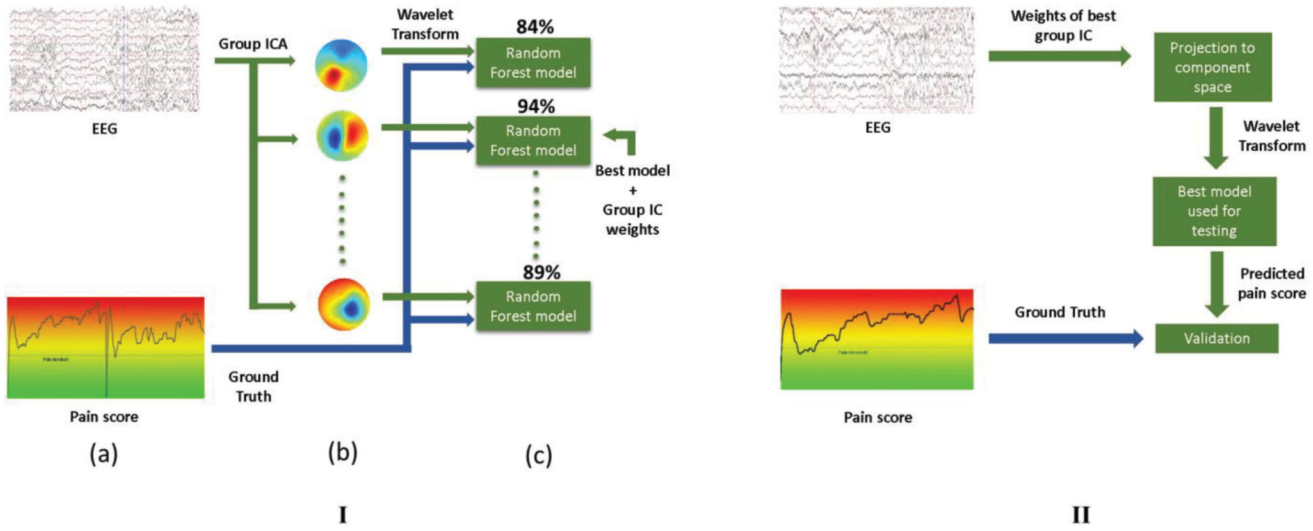

Fig. 2.

Algorithm workflow. (I) Training paradigm. (a) Training EEG data, along with the ground truth is collected from all but one subjects. (b) Group ICA is used on concatenated pre-processed EEG data. (c) Random Forest models are trained on two-thirds of the training data for each component, and the best performing model along with the corresponding group IC weights are picked for testing. (II) Testing paradigm, where the spatial group IC weights of the most discriminative IC are used to project the test data to the component subspace. The best model selected in I(c) is used to test performance of the prediction method on the unseen test subject.

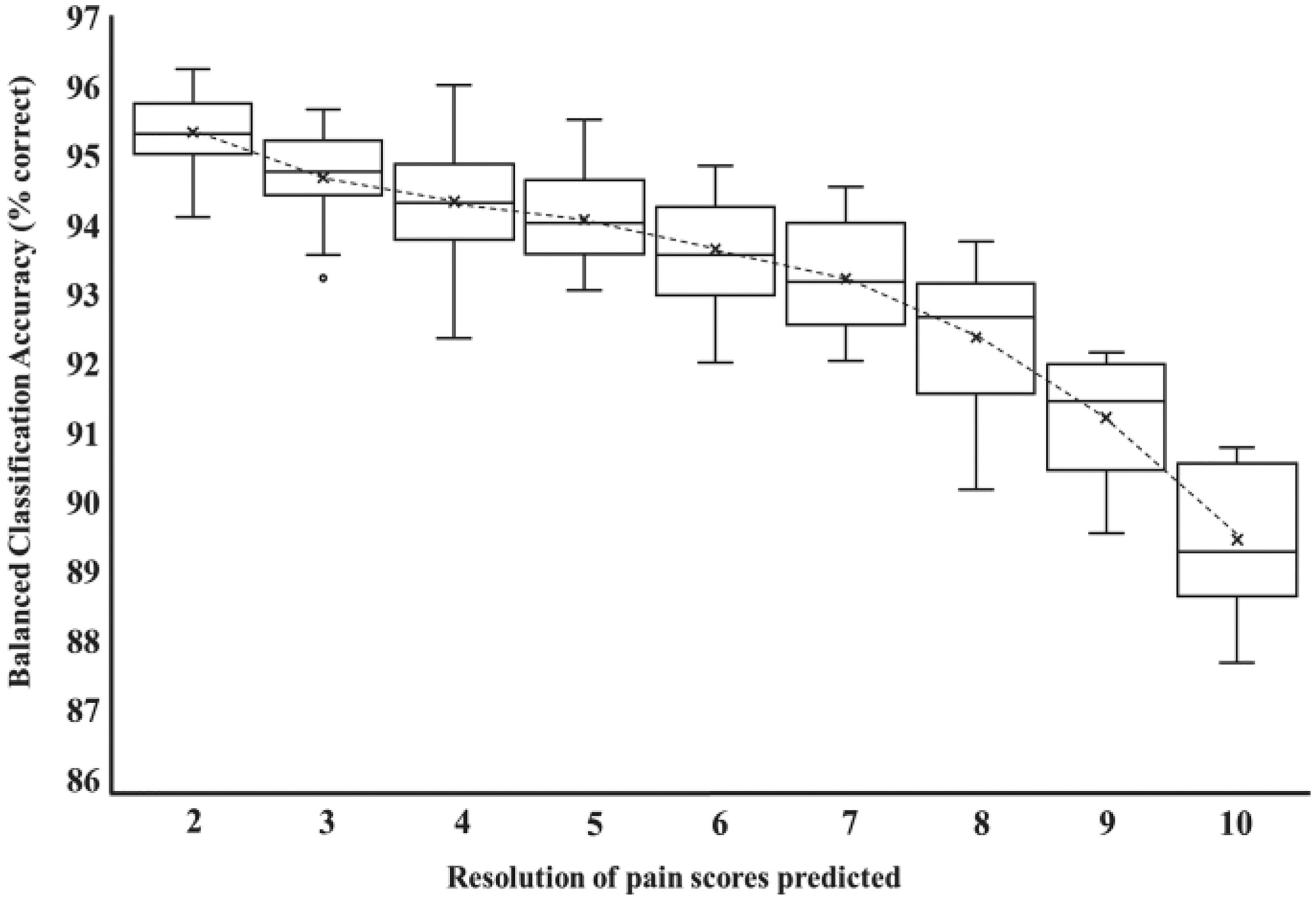

Fig. 3.

Boxplot of balanced classification accuracies (BCA) at different resolutions of the continuous pain score for classification across subjects. The X-axis depicts the results starting from 2-way classification (rest vs pain), up to and including those for 10-way classification (1–10 pain score). The low drop in BCA between 2-way and 10-way classification paradigms (6%) demonstrates the scalability of the classifier.

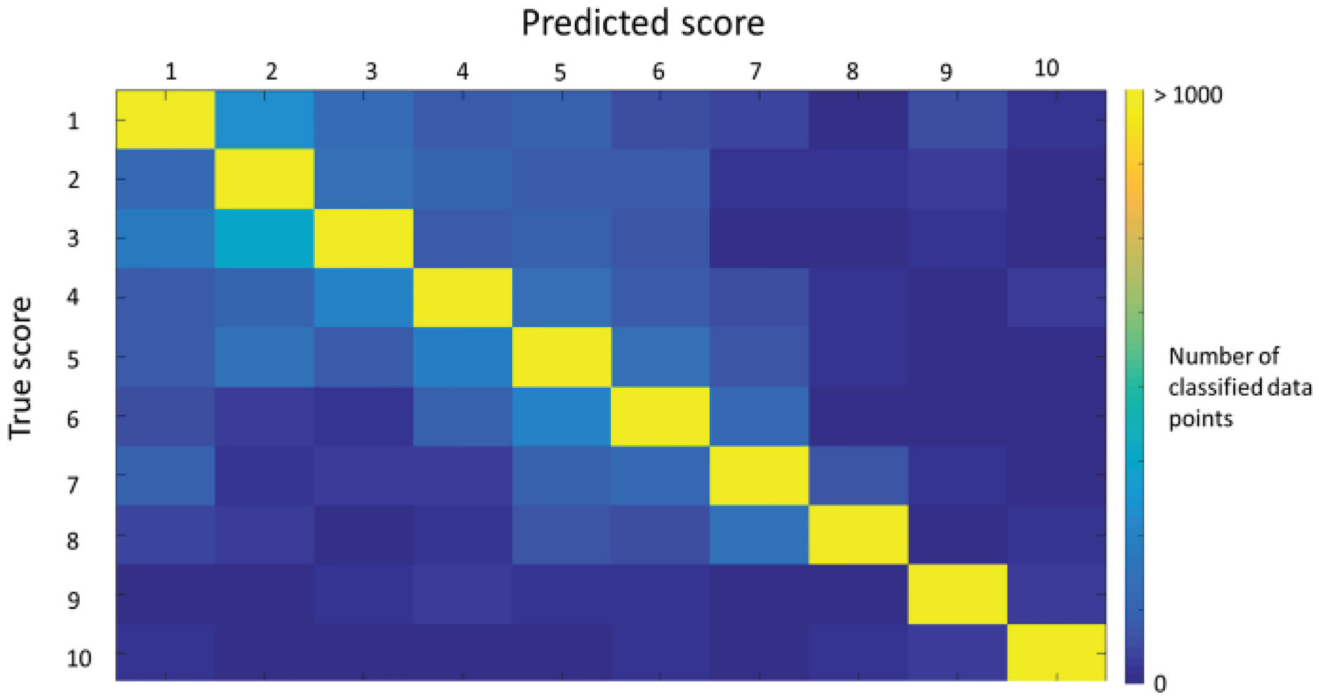

Fig. 4.

Confusion matrix for 10-way quantification using a classifier trained on 24 subjects, and tested on the remaining subject. The color bar indicates the number of classified data points for the test subject. The diagonal nature of the matrix and the skew of errors towards lower pain scores demonstrate the feasibility of such a classifier in a clinical setting.

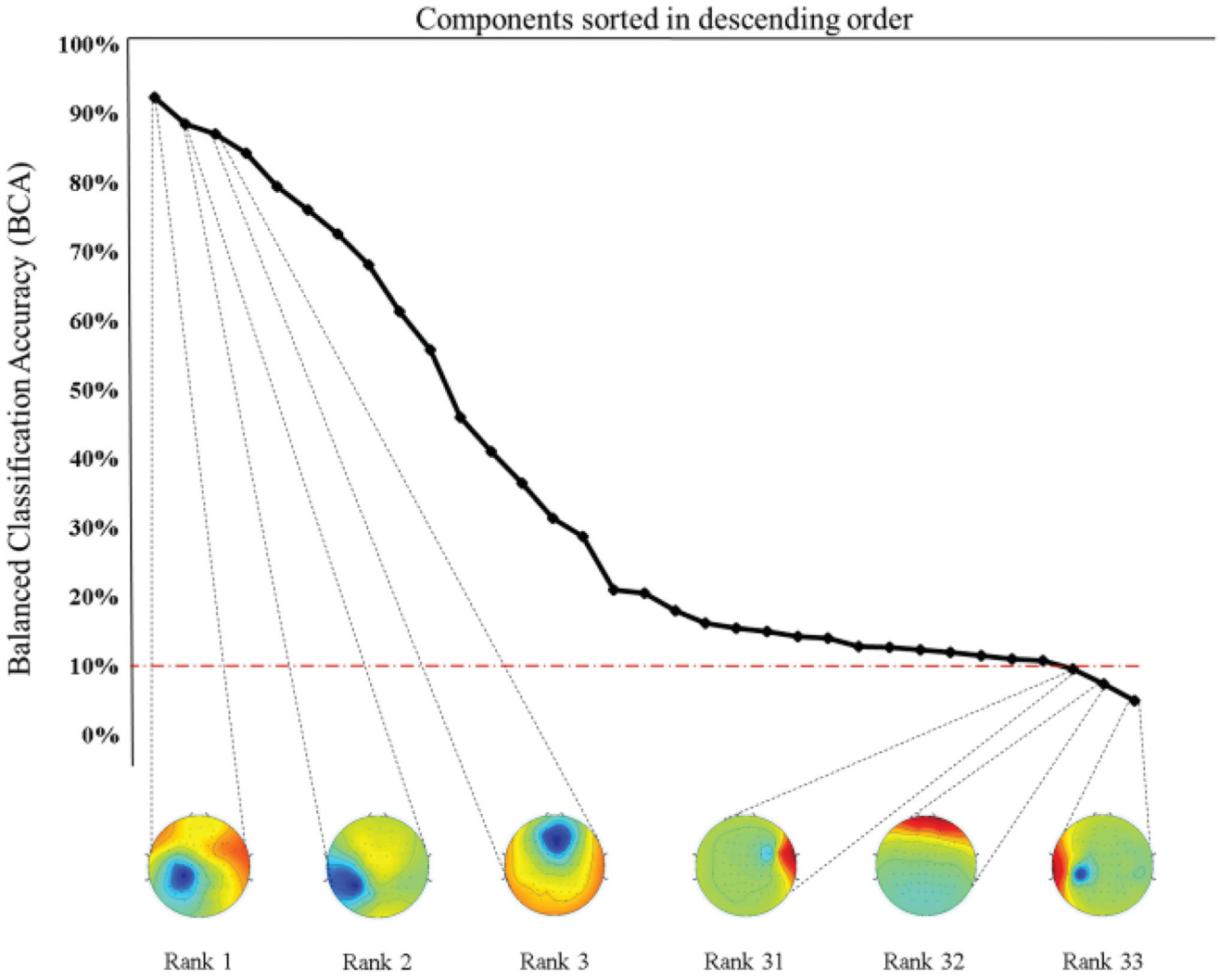

Fig. 5.

Balanced Classification Accuracy (BCA) of each component when tested on a subject, sorted in descending order of performance. The scalp maps of the best and worst three ICs are presented. The red line indicates chance BCA for the classifier. Scalp topographies corresponding to muscle activity or EEG artifacts consistently perform the worst during classification.

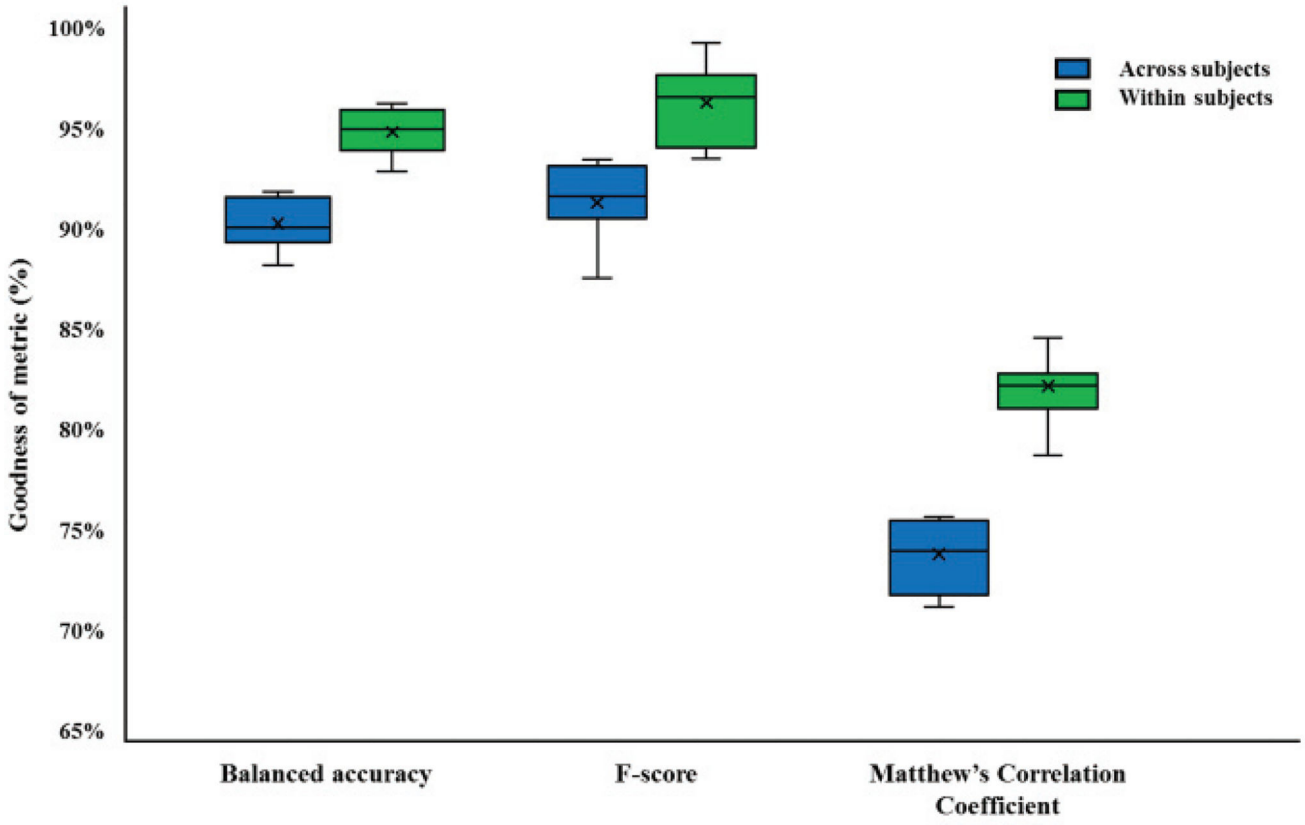

Fig. 6.

Calculating goodness of performance metrics for assessing classifier performance when trained and tested both within and across 25 subjects. All metrics are higher for the intra-subject case compared to the inter-subject case in the presence of class imbalance, as expected.

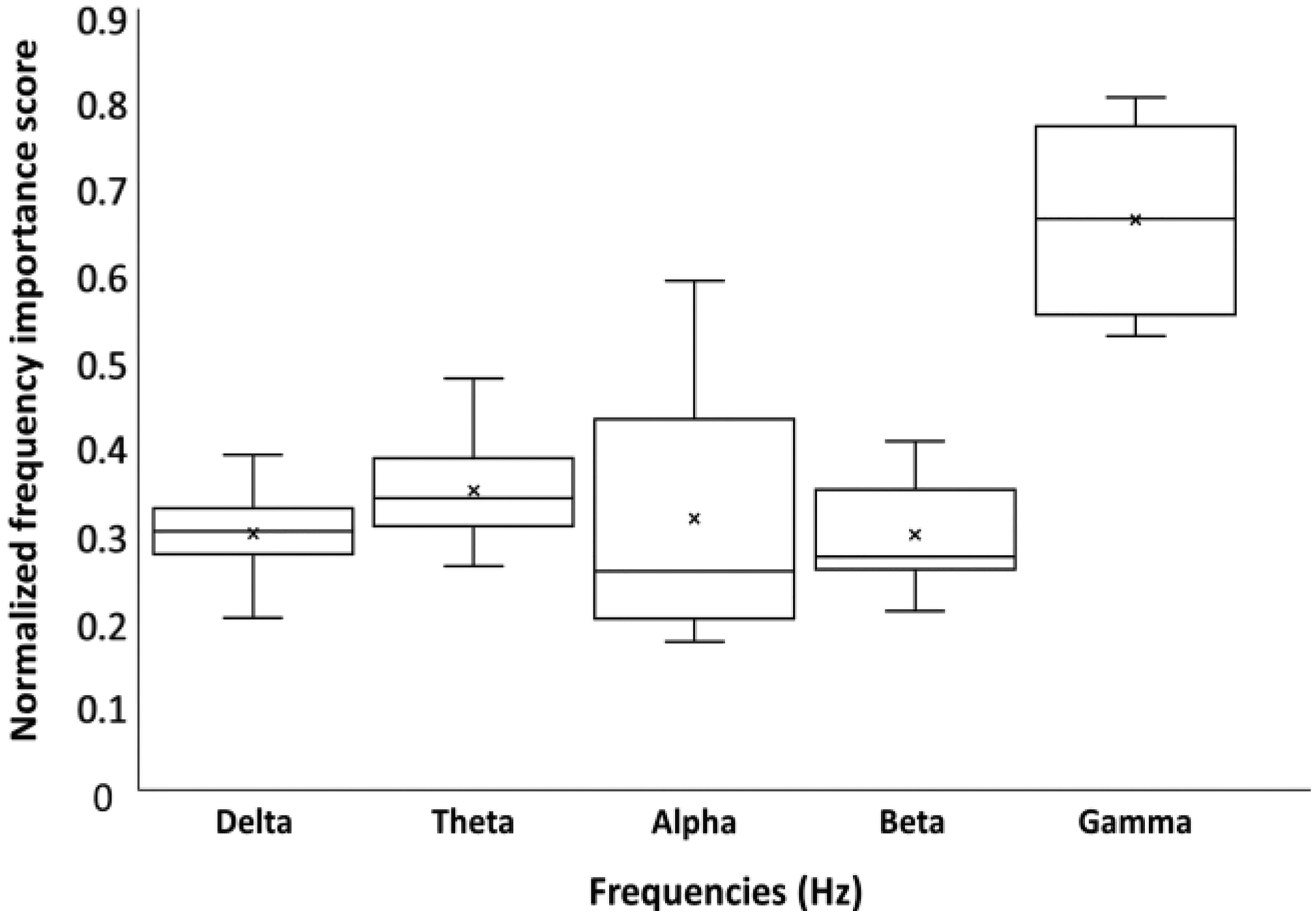

Fig. 7.

Box plot of normalized frequency contribution scores across all subjects for 10-way classification. The gamma band is seen to be consistently important across all subjects, while the alpha band shows the most variability in importance.

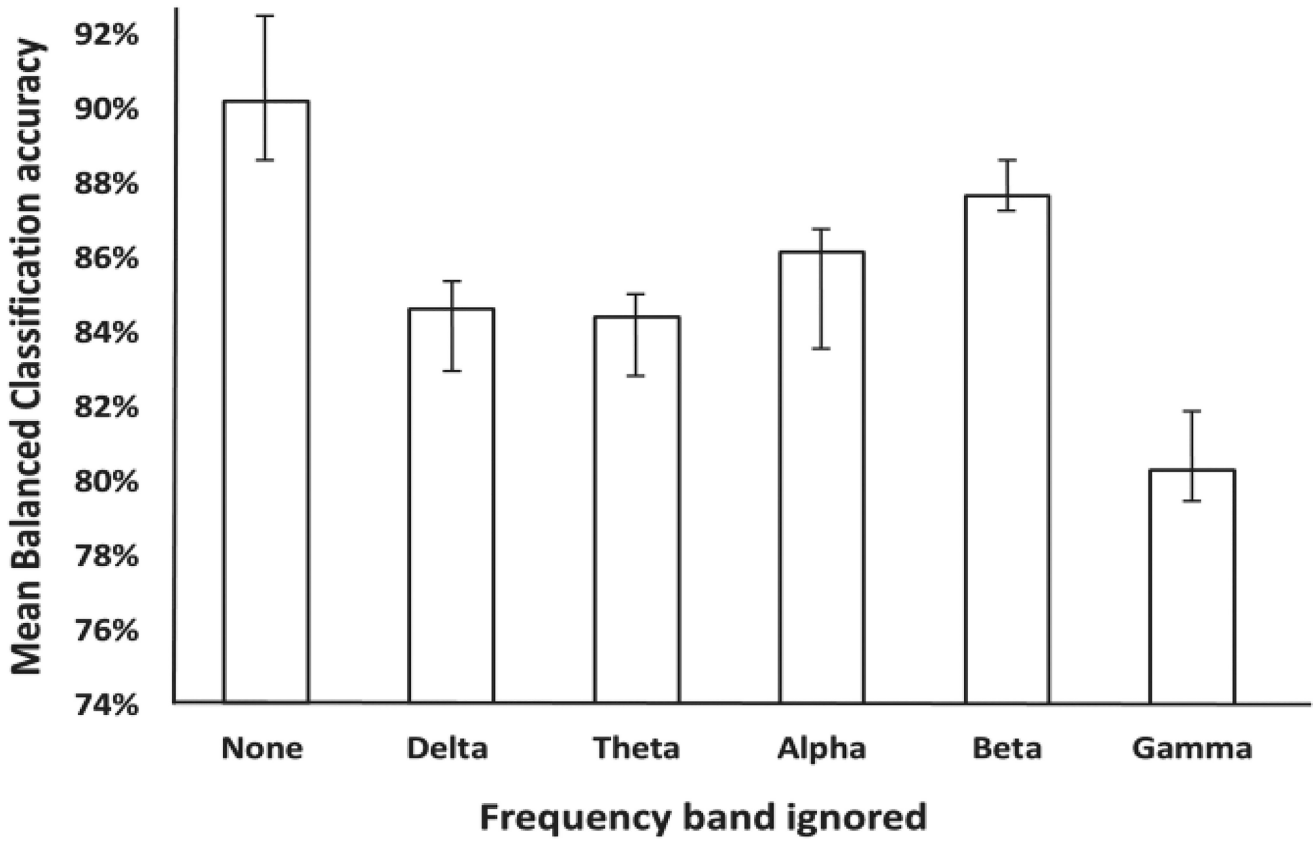

Fig. 8.

Effect of ignoring each frequency band in turn on training classification accuracy. Largest accuracy drop is seen when the gamma band is ignored, suggesting it is the most important.

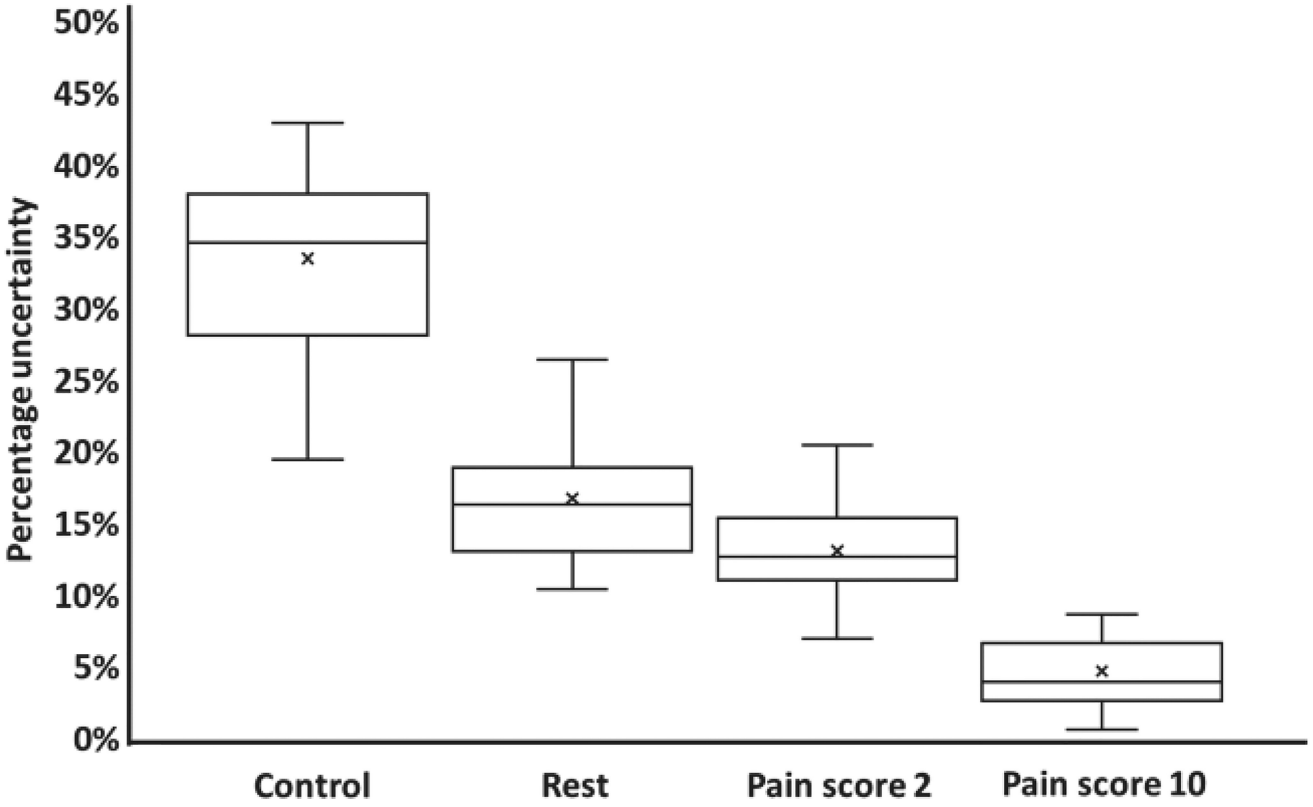

Fig. 9.

Classifier uncertainty for rest, control, and two pain scores. The uncertainty is a probability measure of how unsure the algorithm is of classifying a test point into the correct class. Greater uncertainty while classifying points belonging to the control condition shows that attention to the task plays a role during pain prediction, but low uncertainty while classifying pain scores shows that attention is not the dominant factor.

TABLE I

Comparison of Existing Classification Approaches

Source	Modality	Average Accuracy (%)	
		Intra-subject	Inter-subject
Broderson et al., 2012 [2-class]	fMRI	66.50	
Marquand et al., 2010 [3-Class]	fMRI	72.67	
Schulz et al., 2013 [10-Class]	EEG		83.00 *
Brown et al., 2011 [2-Class]	fMRI		86.60
Misra et al., 2017 [2-class]	EEG		89.58 *
Kuo et al., 2017 [2-class]	MEG		83.00
Current method [10-Class]	**EEG**	**93.26**	**89.45**
Wager et al., 2013 [10-Class]	fMRI	94.00	92.00
Huang et al., 2013[2-class]	EEG	86.30	80.30

* Listed maximum accuracy