NeuroImage 225 (2021) 117506

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Designing individual-specific and trial-specific models to accurately predict the intensity of nociceptive pain from single-trial fMRI responses

Qianqian Lin[a,b,c], Gan Huang[a,b], Linling Li[a,b], Li Zhang[a,b], Zhen Liang[a,b], Ahmed M. Anter[a,b], Zhiguo Zhang[a,b,d,*]

[a] School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, Guangdong 518060, China
[b] Guangdong Provincial Key Laboratory of Biomedical Measurements and Ultrasound Imaging, Shenzhen, Guangdong 518060, China
[c] Department of Brain Functioning Research, The Seventh Hospital of Hangzhou, 305 Tianmushan Road, Hangzhou, Zhejiang, China
[d] Peng Cheng Laboratory, Shenzhen, Guangdong 518055, China

ARTICLE INFO

Keywords:
Pain prediction
Individualized prediction
fMRI
Machine learning
Individual difference

ABSTRACT

Using machine learning to predict the intensity of pain from fMRI has attracted rapidly increasing interests. However, due to remarkable inter- and intra-individual variabilities in pain responses, the performance of existing fMRI-based pain prediction models is far from satisfactory. The present study proposed a new approach which can design a prediction model specific to each individual or each experimental trial so that the specific model can achieve more accurate prediction of the intensity of nociceptive pain from single-trial fMRI responses. More precisely, the new approach uses a supervised *k*-means method on nociceptive-evoked fMRI responses to cluster individuals or trials into a set of subgroups, each of which has similar and consistent fMRI activation patterns. Then, for a new test individual/trial, the proposed approach chooses one subgroup of individuals/trials, which has the closest fMRI patterns to the test individual/trial, as training samples to train an individual-specific or a trial-specific pain prediction model. The new approach was tested on a nociceptive-evoked fMRI dataset and achieved significantly higher prediction accuracy than conventional non-specific models, which used all available training samples to train a model. The generalizability of the proposed approach is further validated by training specific models on one dataset and testing these models on an independent new dataset. This proposed individual-specific and trial-specific pain prediction approach has the potential to be used for the development of individualized and precise pain assessment tools in clinical practice.

1. Introduction

In basic and clinical studies of pain, the foremost question is how to accurately measure the intensity of pain (Lötsch and Ultsch, 2018; Prato et al., 2011; Lee et al., 2019). Because pain is a highly subjective experience and can hardly be perceived by others, self-report is the golden standard for measuring the presence, absence and intensity of pain in clinical diagnosis and treatment of pain (Cruccu et al., 2010; Haanpää et al., 2011). However, self-report could be biased for people who deliberately exaggerate or conceal pain or could be unavailable for people who are with unconsciousness or limited cognitive abilities (Schnakers and Zasler, 2007; Herr et al., 2004; Buffum et al., 2007). Therefore, much effort has been devoted to developing objective pain assessment tools, and these researches have been focused on the uses of various types of physiological signals, such as electroencephalography (EEG), skin conductance, respiration rate and blood volume pulse, and behaviors, such as facial expression (Bai et al., 2016; Gholami et al., 2010; Gruss et al., 2015; Zautra et al., 2010; Schulz et al., 2012).

Recently, functional magnetic resonance imaging (fMRI) has been increasingly used to measure the intensity of pain perception (Boly et al., 2007; Lindquist et al., 2017; Brown et al., 2011). Because blood-oxygen-level-dependent (BOLD) fMRI can record brain activity with high spatial resolution, it is effective in identifying brain regions that are related to pain experience (Wager et al., 2013; Woo et al., 2017). Furthermore, various machine learning models have been developed to learn the relationship between fMRI responses and the intensity of pain from a set of experimental pain-evoked trials of a cohort of individuals (Brodersen et al., 2012; Marquand et al., 2010; Cecchi et al., 2012). These machine learning models can then be used to predict the intensity of pain from fMRI activities of new experimental trials or new individuals.

However, because of large inter-individual difference (i.e., different people have remarkably different pain-related behavioral and fMRI re-

* Corresponding author at:School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen Guangdong 518060, China.
E-mail address: zgzhang@szu.edu.cn (Z. Zhang).

https://doi.org/10.1016/j.neuroimage.2020.117506
Received 31 August 2019; Received in revised form 6 October 2020; Accepted 21 October 2020
Available online 27 October 2020

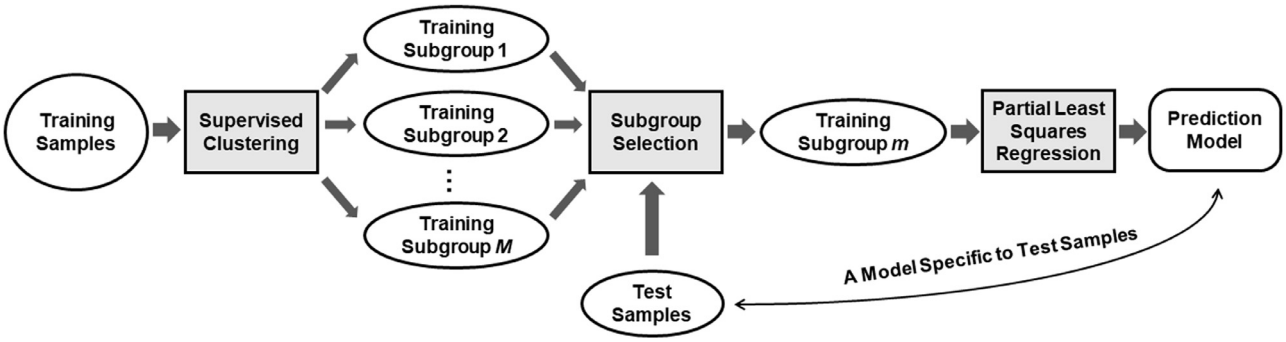

Fig. 1. Illustration of the proposed approach for designing pain prediction models specific to test samples. First, supervised clustering is used to categorize training samples into a set of subgroups. Second, a training subgroup is selected based on its similarity to the test samples. Third, a regression model (here, partial least squares regression) is trained from the selected training subgroup and this model is specific to the test samples.

sponses) (Coghill et al., 2003; Coghill, 2011) and intra-individual difference (i.e., one individual has different pain-related behavioral and fMRI responses at different times) (Tracey and Mantyh, 2007), a pain prediction model trained from certain individuals or trials generally has lower performance when being applied on new individuals or trials (Huang et al., 2013; Hu and Iannetti, 2016; Mun et al., 2019). As summarized in (Vijayakumar et al., 2017), cross-individual pain prediction in general has lower accuracy than within-individual pain prediction. Therefore, it is important to develop a general and effective approach to address the problem of individual variability and to establish a pain prediction model with high accuracy.

To date, there are few studies concerning the development of an fMRI-based pain prediction model that can handle individual variability. Lindquist et al. (2017) proposed to incorporate group-level priors into individual prediction models, but its objective is to address the problem of insufficient samples for each individual. In our previous EEG study (Li et al., 2018), an individual's pain-evoked EEG was normalized by his/her spontaneous EEG to reduce the inter-individual difference in EEG-based pain prediction models. As for the intra-individual variability, our previous work (Tu et al., 2016a) showed that including pre-stimulus brain activities in a pain prediction model could achieve significantly better performance than conventional models, which only used post-stimulus pain-evoked brain activities as features, because pre-stimulus brain activities can largely explain intra-individual variability of pain-evoked brain activities. Recently, our work (Lin et al., 2018) showed that, inter-individual differences in the fMRI responses and pain ratings jointly determine the error of cross-individual pain prediction. However, because pain ratings of a test individual are unknown, it is not feasible to improve the accuracy of pain prediction by minimizing the individual difference of pain ratings. On the other hand, a new individual's fMRI activities can be readily recorded, so it is possible to minimize the difference between fMRI activities of training samples and test samples for a more accurate and individually tailored model. This study inspired us to develop a more accurate model by selecting training sample with similar fMRI activities as test samples.

In the present study, we developed a new pain prediction approach to combat the adverse influence of both inter- and intra-individual differences in fMRI-based pain prediction models. The basic idea of the proposed approach is to select a subset of training samples (instead of all available training samples) with the closest fMRI responses to test samples so that the trained model has minimal difference in fMRI features between training and test samples. The proposed approach is inspired by the dynamic selection technique (Cruz et al., 2018), which is an active research topic in the machine learning community. Dynamic selection techniques are capable of designing a classifier according to each new sample to be classified and one common method is to train the classifier based on a local region of the feature space where the test sample is located (Cruz et al., 2018).

The proposed pain prediction approach has three major steps (Fig. 1). First, a supervised clustering method is adopted to cluster training individuals/trials into a small number of subgroups, in which individuals/trials share similar and consistent pain-related fMRI activation patterns. Second, for a new individual or trial under test, the new approach compares the fMRI patterns of the test individual/trial with the fMRI patterns of those subgroups, and then selects the subgroup with the closest fMRI patterns to the test individual or trial for training of an individual- or trial-specific model. Third, a continuous prediction model (partial least squares regression is used in this study) is trained from the selected training subgroup so that this model is specific to the test samples.

We examined the performance of the proposed approach on a nociceptive-pain-evoked fMRI dataset. To test the robustness of the supervised clustering method, four different features sets (the whole brain, the pain matrix, pain-activated regions, and pain-predictive regions) were independently adopted in the supervised clustering method to define subgroups. A partial least squares regression model was then established from selected training samples and applied on test samples. Results showed that the newly designed individual-specific or trial-specific models achieved significantly higher prediction accuracy than conventional non-specific models using all available training samples. Particularly, the trial-specific model had superior performance over the individual-specific model, though it used much fewer samples to train the model. We also identified a set of brain regions exhibiting significantly different fMRI patterns among those clustered subgroups and these regions account for the inter- or intra-individual differences in the fMRI-based pain prediction models. Lastly, we demonstrated the strong generalizability of the proposed approach by training specific models on one dataset and testing these models on an independent new dataset.

2. Materials and methods

2.1. Participants and experiments

This study has two independent pain-evoked fMRI datasets. Dataset I is the primary data used in this study to develop and test the proposed specific models, and most results are based on this dataset. Dataset II is only used to examine the generalizability of the proposed approach by training models from Dataset I and testing models on this Dataset II. Note that, Dataset I has been used in our previous publications (Tu et al., 2016a, 2016b, 2018; Lin et al., 2018), but Dataset II has not been reported before.

2.1.1. Dataset I

Dataset I included a total of 32 healthy participants (age=22.1±2.0, 20 females). The initial inclusion and exclusion criteria were based on the general health questionnaire, pain safety screening form, and

fMRI safety screening form. Participants reported no history of chronic pain, psychiatric, or neurological disorders. The experimental procedures were approved by the local ethics committee. All participants gave written informed consent, and they were familiarized with the experiment paradigm before the experiment.

Radiant-heat stimulus energies included four levels (E1: 2.5 J; E2: 3.0 J; E3: 3.5 J; E4:4.0 J), and ten laser pulses at each of the four energies were delivered in a random order on the dorsum of the left hand, with a total of 40 pulses per individual. Participants experienced the laser stimulus and then reported the intensity of pain, using a visual analog scale (VAS) ranging from 0 to 10 (0: no pain; 10: pain as bad it could be). Laser stimuli of the four energies elicited graded subjective pain intensities (E1: 2.9 ± 1.5; E2: 3.8 ± 1.7, E3: 5.7 ± 1.6, and E4: 6.9 ± 1.5) (Tu et al., 2016b). Because only using one level of stimulus energy cannot elicit pain ratings in the whole range from 0 to 10, which cannot result in a good prediction model, so we used four different levels of stimulus energies with the aim to produce samples which were suitable for the establishment of a pain prediction model. In addition, according to our previous studies (Zhang et al., 2012; Huang et al., 2013; Tu et al., 2016b, 2018) and literature (Gross et al., 2007; Iannetti et al., 2008; Hu et al., 2014; Hu and Iannetti, 2019), such an experimental design (4 intensities of energy, each having 10 trials for each participant) can generate sufficient samples for developing and validating pain prediction models. It should also be noted that, we have elucidated the relationship between stimulus intensities and pain experience (both pain ratings and pain-evoked fMRI responses were positively correlated with the stimulus intensity) in our previous publication (Tu et al., 2016b). We will not discuss the analysis of nociceptive input (including stimulus intensities) in this study because it is outside the scope of this study, and interested readers can refer to Tu et al. (2016b) for details.

MRI data were collected using a Siemens 3.0 Tesla Trio scanner with a standard head coil. Functional images were acquired with echo planar imaging sequence with the following parameters: 255 mm thick slices and 0.5 mm inter-slice gaps, TR=1500 ms, TE=29 ms, filed of view =192 × 192 mm, 64 × 64 matrix, 3 × 3 × 3 mm^3 voxels, flip angle=90° At the end of the experiment, high-resolution T1-weighted structural images were collected.

2.1.2. Dataset II

Dataset II included 49 healthy participants (aged=22.8 ± 6.0, 32 females). The initial inclusion and exclusion criteria were based on the general health questionnaire, pain safety screening form, and fMRI safety screening form. Participants reported no history of chronic pain, psychiatric, or neurological disorders. The experimental procedures were approved by the local ethics committee. All participants gave written informed consent, and they were familiarized with the experiment paradigm before the experiment.

Participants received brief stimuli of three different sensory modalities: nociceptive somatosensory (pain), auditory, and visual. Nociceptive stimuli were pulses of radiant heat. Auditory stimuli were loud 800 Hz tones. Visual stimuli consisted of a bright white disk displayed on the projection screen. Note that we only used nociceptive laser-evoked fMRI trials in the present study to examine the generalizability of the proposed approach. Nociceptive stimuli included two intensity levels (E1: 3 J; E2: 3.5 J). The experiment consisted of two runs, and within each run, each type of stimulus was delivered 10 times in a pseudo-random order, 5 times for each intensity level. As a result, each participant had 20 pain-activated trials. Participants were asked to rate the intensity of perception for each stimulus using a VAS rating from 0 to 10.

MRI data were collected using a GE 3.0 T MRI scanner with a standard head coil. Functional images were collected using a standard gradient echo planar imaging sequence with following imaging parameters: 43 oblique slices, thickness/gap=3/0 mm, acquisition matrix=64 × 64, TR=2000 ms, TE=29 ms, flip angle=90°, field of view=192 × 192 mm^2, total volume=300. For multi-sensory stimulation task, two sessions with 454 functional volumes each were administered. At the end of the experiment, high-resolution T1-weighted structural images were collected.

2.2. fMRI data preprocessing and feature extraction

For each subject, structural T1-weighted images were co-registered to the mean functional image and were then normalized to the Montreal Neurological Institute (MNI) space using SPM8 (Penny et al., 2007). fMRI data were slice-timing corrected, head motion corrected, normalized to the MNI space (voxel size =3 × 3 × 3) by mapping T1-weighted structural images to the MNI template, and then smoothed with an 8 mm FWHM Gaussian kernel. A high-pass filter was applied (cut-off frequency = 1/128 Hz) to the BOLD time-series to remove low-frequency drifts. Normalization of BOLD responses was performed by subtracting and then dividing the baseline BOLD signals (at the stimulus onset). BOLD responses at the 4th scan (for Dataset I with a TR=1.5) or the 3rd scan (for Dataset II with a TR=2) after stimulus onset were extracted as features for pain prediction, because BOLD responses at these time points are normally strong (or even at peaks) and significantly correlated with pain ratings.

To identify nociceptive-evoked fMRI activations, single-subject fMRI data were analyzed on a voxel-by-voxel basis using the general linear model (GLM) (Friston et al., 1994). BOLD fMRI responses were modeled as a series of events using a stick function and pain ratings were included as a parametric modulator of each stimulus, which were then convolved with a canonical hemodynamic response function. In order to identify nociceptive-evoked fMRI activations at the group level, statistical analyses were carried out using a random effects analysis with one-sample *t*-test as implemented in SPM8. Considering the problem of multiple comparisons, family-wise error (FWE) rate (Hochberg, 1988) was used to correct the significance level (P_{FWE}<0.05).

2.3. Pain prediction model (partial least squares regression)

The trial-by-trial relationship between normalized fMRI features and pain ratings for each individual was modeled using partial least squares regression (PLSR), which is a commonly used machine learning method particularly suitable for multicollinear data (McIntosh et al., 2004; Krishnan et al., 2011; Hu et al., 2014; Tu et al., 2016a; O'Connell et al., 2018). For the *j*-th individual, the PLSR model is formulated as: $Y_i^{(j)} = \boldsymbol{X}_i^{(j)}\boldsymbol{\beta}^{(j)}$, where $Y_i^{(j)}$is the pain rating of the i th trial, $\boldsymbol{X}_i^{(j)}$is the fMRI features of the i th trial, and $\boldsymbol{\beta}^{(j)}$is the PLSR model coefficient vector. Each coefficient in the PLSR coefficient vector $\boldsymbol{\beta}^{(j)}$ represents the predictive capability of the nociceptive-evoked fMRI feature at the corresponding voxel. The SIMPLS algorithm was used to compute the PLSR model coefficients (De Jong, 1993). The Matlab function "plsregress" (https://www.mathworks.com/help/stats/plsregress.html) was used for the implementation of PLSR in MatlabR2018a (MathWorks; Natick, MA, US). To identify pain-predictive brain regions at the group level, we used a one-sample *t*-test against zero to assess the significance of PLSR coefficients across individuals.

In the PLSR model, the number of latent components determines whether the trained model is over-fitted or not. More latent components lead to a better fitting of the model to the data. But, if too many latent components are used, the PLSR model may be over-fitted because it could describe the noise components in the data. On the other hand, if too few latent components are used, the PLSR model could be underfitted for it cannot explain sufficient information of the data. In the present study, the number of latent components in the PLSR analysis was estimated using internal cross validation of the coefficient of determination. The coefficient of determination calculates the percentage of the variance of the values fitted by the latent components and the total variance of the dependent variables, and it is a common metric to measure the fitting performance of a regression model. We used leave-one-individual-out or leave-one-trial-out cross validation within the training set for an individual-specific model or a trial-specific model to calculate

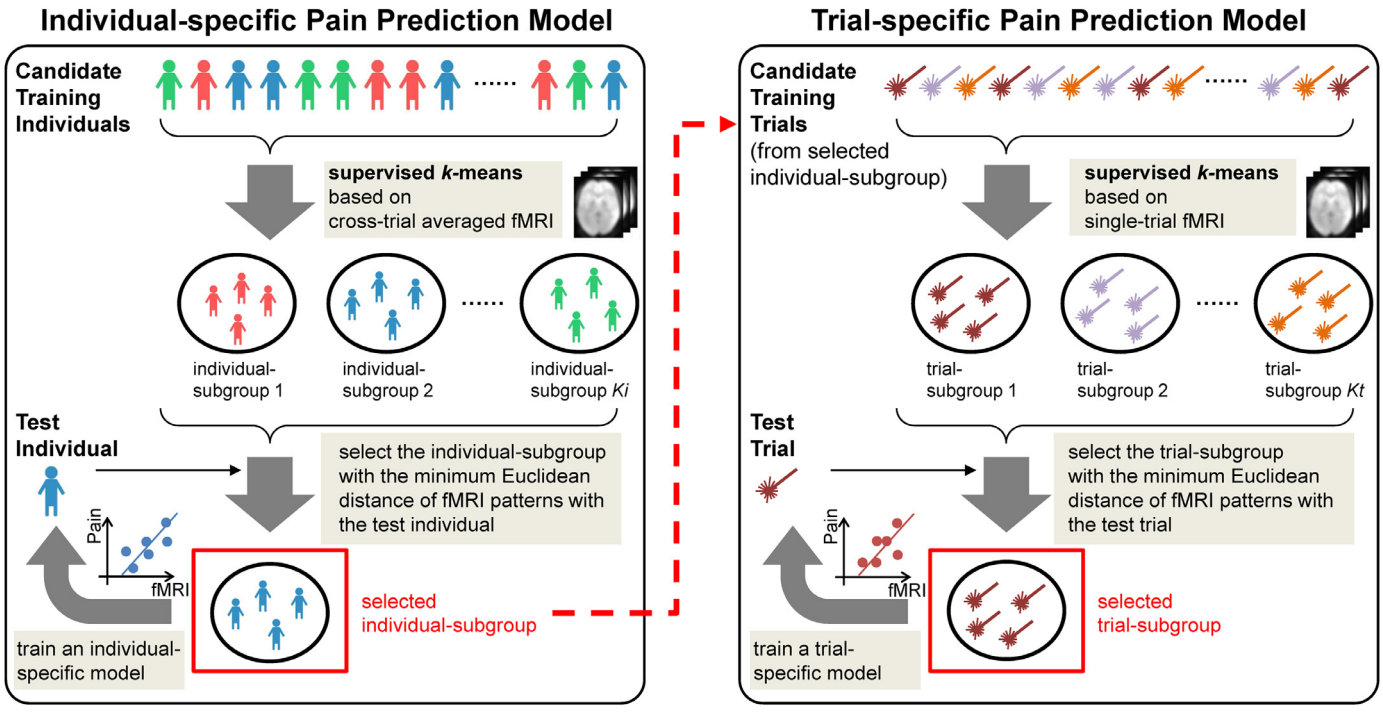

Fig. 2. Schematic representation of the proposed pain prediction approach (individual-specific model and trial-specific model). To build an individual-specific model, candidate training individuals were grouped into several individual-subgroups by using supervised *k*-means. The individual-subgroup having the closest fMRI features to the test individual (as measured by the Euclidean distance between fMRI patterns of the individual-subgroup and the test individual) was selected to train an individual-specific model. To build a trial-specific model, all trials of the selected individual-subgroups were further grouped into several trial-subgroups by using supervised *k*-means. The trial-subgroup having the closest fMRI features to the test trial (as measured by the Euclidean distance between fMRI patterns of the trial-subgroup and the test trial) was selected to train a trial-specific model.

the coefficients of determination with different numbers of latent components (ranging from 10 to 100) and selected the one with the maximum coefficient of determination as the number of latent components used in the PLSR model for the test data. Such a method is popularly used in literature to avoid over-fitting of PLSR models (Faber and Rajko, 2007).

2.4. Proposed individual- and trial-specific model design approach

The basic idea of the proposed individual- and trial-specific fMRI-based pain prediction approach is to only select a subset of training samples with similar fMRI activation patterns as test samples. The procedures of the proposed approach are illustrated in Fig. 2 and detailed in the following.

2.4.1. Clustering of individuals and trials by supervised k-means

The first step of our proposed fMRI-based pain prediction approach was to cluster individuals/trials according to their fMRI patterns, so that one cluster could be selected for training a specific model. More precisely, a "supervised *k*-means" method (Gan et al., 2018) was employed in the proposed approach to cluster individuals/trials into a small number of subgroups based on their fMRI features. Clustering is usually unsupervised, which means no label information is used. However, if label information is available, unsupervised clustering is not the optimal strategy for correctly categorizing samples. Because the label information of training samples is readily for use in building a pain prediction model, the proposed approach uses supervised clustering to group individuals and trials. Compared with unsupervised clustering and semi-supervised clustering, supervised clustering can identify clusters with high probability density with respect to a single class while keeping the total number of clusters small. Supervised clustering has attracted wide attention in recent years and has been effectively used in many applications, such as image segmentation (Gan et al., 2018). Generally, supervised clustering evaluates a cluster based on two criteria: (1) class impurity (which is measured by the percentage of minority samples in different clusters) and (2) the number of clusters M(which should be kept small). Suppose $C_m$was the m-th cluster ($m = 1, 2, \ldots, M$) obtained by unsupervised clustering (i.e., *k*-means used in this study) and $N(C_m)$was the number of minority samples in the m-th cluster (C_m), the objective function of supervised clustering is formulated as:

$$J(X, M, \beta) = P(X) + \beta L(M)$$

where $P(X) = \frac{1}{l}\sum_{m=1}^{M} N(C_m)$is the class impurity and $L(M) = \sqrt{(M-c)/l}$with $l$the size of sample and $c$the number of classes. The value of $P(X)$is in the range between 0 and $1/c$. If samples of one cluster all belong to the same class, then $P(X) = 0$. In this study, the number of classes $c$was set to 2 (pain-sensitive or pain-insensitive for individuals, and low pain or high pain for trials). An individual was assigned to a pain-sensitive class (mean VAS > 5) or a pain-insensitive class (mean VAS ≤ 5). A trial was assigned to a high-pain class (VAS>5) or a low-pain class (VAS ≤5). To obtain the optimal solution for above objective function, we can choose different values of $M$and calculate the corresponding $J(X, M, \beta)$. When $J(X, M, \beta)$gets the smallest value, the corresponding $M$is the optimal.

In this study, we used the popular *k*-means in the framework of supervised clustering. The supervised *k*-means method was applied on individual-level fMRI features (averaged across all trials of one individual) and single-trial fMRI features in order to divide individuals or trials into a small number of subgroups. In the following, we introduced two specifically-designed models, which are respectively established at the individual level and at the trial level. The individual-specific model can tailor make a pain prediction model specific to each individual while the trial-specific model can tailor make a pain prediction model specific to each trial.

2.4.2. . Individual-specific pain prediction model

In the individual-specific pain prediction model, we used supervised *k*-means to group training individuals into several individual-subgroups according to their fMRI features (averaged across trials for each individual). By doing this, the difference of fMRI features of training individuals within each individual-subgroup is minimized and individuals in one individual-subgroup have similar and consistent fMRI features. Next, when predicting the pain intensity of a new individual, we compared the test individual's fMRI features with the fMRI features of each individual-subgroup to find out the most suitable individual-subgroup for training. More precisely, we determined the optimal training subgroup by comparing the Euclidean distance between fMRI features of the test individual and the center of each individual-subgroup. The individual-subgroup having the smallest Euclidean distance with the test individual was selected to train a model specific to this test individual. According to our previous study (Lin et al., 2018), smaller the distance between training data and test data, smaller the prediction error.

2.4.3. Trial-specific pain prediction model

The individual-specific pain prediction model can effectively mitigate the inter-individual variability in pain prediction model. But, the negative influence of intra-individual variability, or inter-trial variability, has not been tackled. Therefore, based on the individual-specific model, we further developed a trial-specific pain prediction model. Suppose one optimal individual-subgroup is selected to train a model for one test individual. Next, for each trial of this test individual, we selected a subgroup of trials (trial-subgroup) from all trials within the selected individual-subgroup to train a pain prediction model specific to this test trial. To achieve this, we used supervised *k*-means again to divide all trials within the selected individual-subgroup into a small number of trial-subgroups. Similarly, we then calculated the Euclidean distance of fMRI features between the test trial and the center of each trial-subgroup, and the trial-subgroup with the smallest difference with the test trial was selected to train a trial-specific model for this test trial.

2.4.4. Feature sets

To perform a complete and rigorous examination of the proposed approach, we tested four different types of features (i.e., BOLD fMRI features within the following four brain regions) separately in the proposed approach. The details of these four feature sets are as follows.

1) Whole Brain: fMRI features of all voxels within the whole brain were used as features.
2) Pain Matrix: The pain matrix included the primary somatosensory cortex (SI), the secondary somatosensory cortex (SII), the supplementary motor area (SMA), the prefrontal cortex (PFC), the anterior cingulate cortex (ACC), insula, and thalamus. SI includes Brodmann areas 1, 2 and 3, SII includes Brodmann areas 40 and 43, and other regions of interest (ROIs) are defined based on the AAL atlas (Tzourio-Mazoyer et al., 2002).
3) Pain-Activated: These pain-activated regions were defined based on the GLM results during leave-one-individual-out cross validation. That is, when building a prediction model for one test individual using all other $(N-1)$ individuals' data, we obtained the pain-activated regions using GLM on $(N-1)$ individuals. Considering the problem of multiple comparisons, FWE was used to correct the significance level.
4) Pain-Predictive: These pain-predictive regions were defined based on the PLSR results during leave-one-individual-out cross validation. That is, when building a prediction model for one test individual using all other $(N-1)$ individuals' data, we obtained the pain-predictive regions using PLSR on data from each of $(N-1)$ training individuals. We further used a one-sample *t*-test against zero to assess the significance of these model coefficients at the group level, i.e., $(N-1)$ training individuals. Considering the problem of multiple comparisons, FWE was used to correct the significance level.

It is worth mentioning that, Pain-Activated and Pain-Predictive feature sets used in the new approach were specific for each test individual in leave-one-individual-out cross validation (see Section 2.4.5): they were not obtained from all N individuals but obtained from $(N-1)$ training individuals. On the other hand, the group-level pain-activated regions (as revealed by GLM) and the group-level pain-predictive regions (as revealed by PLSR) shown in Fig. 3were obtained from all N individuals.

All these 4 feature sets have been popularly used in literature to develop pain prediction models: the whole brain (Brodersen et al., 2012; Tu et al., 2018), the pain matrix (Brodersen et al., 2012), the pain-activated regions (Duff et al., 2012), and the pain-predictive regions (Tu et al., 2016a). The main difference between "pain-activated" regions and "pain-predictive" regions is whether these regions are identified based on subjective pain perception or not. Pain-activated regions were identified by using GLM, which aims to explain the observed fMRI time course of a voxel in terms of a linear combination of several reference functions (in this study, a function representing stimulus presentation convolved with the hemodynamic response function). On the other hand, pain-predictive regions were identified by using PLSR, which describes the perceived pain level (from 0 to 10) as a linear combination of pain-evoked maximum fMRI responses (the 4th scan after the stimulus onset) of the whole brain. We can see from above that the key difference between these two types of regions is that: whether the perceived pain ratings are taken into consideration when identifying these regions. Only the time of stimulus presentation was considered when identifying pain-activated regions, so the identified regions were supposed to be "activated" by the stimulation. On the other hand, subjective pain ratings are needed to identify pain-predictive regions, so the identified regions were "predictive" of pain ratings.

It is also important to note that, these four types of features have two different usages in the proposed model design approach. First, these features can be used in the supervised clustering method to measure the distances among samples so that several subgroups of samples can be defined based on these distances. Second, these features can be used as independent variables in PLSR models to predict pain intensity.

2.4.5. Cross validation and performance evaluation

The performance of the proposed individual-specific and trial-specific models was validated on the nociceptive-evoked fMRI dataset $(N = 32)$ using leave-one-individual-out cross validation. More precisely, a three-step process was used to select the test and the training individuals.

Step 1. A leave-one-individual-out cross validation strategy was used to make one individual as the test individual and $(N-1)$ individuals as the candidate training individuals (but only a subgroup of these candidate training individuals was eventually used for training a model). Such a cross validation was repeated *N*times so that each participant acts as the test individual for once.

Step 2. A supervised *k*-means method was used to divide the $(N-1)$ candidate training individuals into a small number of individual-subgroups, each having a certain number of candidate training individuals, based on these individuals' fMRI patterns.

Step 3. We compared the fMRI patterns of individual-subgroups with the fMRI patterns of the test individual. The individual-subgroup having the closest fMRI feature patterns to the test individual was selected as the training individuals to train an individual-specific model for the test individual in this round of cross validation.

Furthermore, a trial-specific model was designed for each trial of the test individual. That is, all trials in the selected individual-subgroups were used as candidate training data (but only a subgroup of which were eventually used for training a trial-specific model). The supervised *k*-means method was used to divide all these candidate training trials into several trial-subgroups, and the trial-subgroup with the closest fMRI

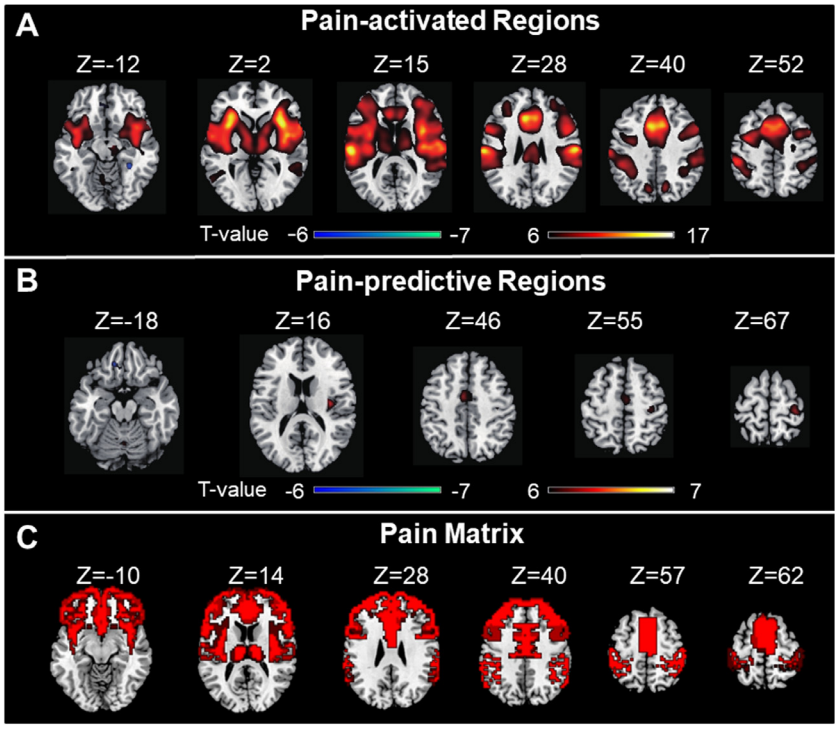

Fig. 3. (A) Pain-activated brain regions defined by GLM at the group level. Activated and deactivated regions are shown in red and blue, respectively (P_{FWE}<0.05). (B) Pain-predictive regions defined by PLSR at the group level. Positively and negatively predictive regions are shown in red and blue, respectively (P_{FWE}<0.05). (C) The pain matrix includes SI, SII, SMA, PFC, ACC, MCC, insula, and thalamus.

features patterns to the test trial was selected to train a trial-specific model. Because one individual had 40 trials, the above procedure for trial-specific prediction was performed 40 times for each test individual. It is also worth noting that subgroup formation was independent of the test samples. Although test samples were indeed used to select one subgroup for training, but only the samples from the training subgroup were used for training and the test samples were not involved in the training.

The conventional cross-individual prediction model, which is not specific to any individual or any trial, was used as control for the comparison with the proposed individual- and trial-specific models. The conventional non-specific pain prediction model was trained and tested based on leave-one-individual-out cross validation. That is, 32 individuals were into 31 training individuals and 1 test individual. The same procedure was repeatedly performed 32 times to make sure that each individual was used as the test individual once. All trials from training individuals were used as training trials to build the model, while all trials from the remaining test individual were used as test trials to validate the model. We can clearly see the main difference between the conventional non-specific model and the proposed specific models is whether training samples are selected to build a model.

The performance of a pain prediction model was evaluated by using mean absolute error (MAE), which is the across-trial average of the absolute values of the estimation error between actual pain ratings and their estimates. We used one-way repeated measures ANOVA and post-hoc paired *t*-test to examine whether there was significant difference in MAE among different models under comparison (the conventional non-specific model, the individual-specific model, and the trial-specific model) for each type of fMRI features (Whole Brain, Pain Matrix, Pain-Activated, and Pain-Predictive). As mentioned earlier, these four types of features can be used in supervised *k*-means as well as in the PLSR model. Because the main novelty in the proposed approach is the supervised clustering method to define subgroups of training samples, we used four different types of fMRI features separately in supervised *k*-means to examine whether the supervised clustering method is robust to features used. On the other hand, we consistently used the Whole Brain features to build the PLSR model for pain prediction in the main results, because we wanted to make the prediction results comparable (the number of features varied largely among four feature sets and had a significant influence on the results). We also did an additional analysis (Additional Analysis 4, as shown below) to examine the prediction performance of proposed approach by using each type of feature in both supervised clustering and PLSR modeling.

2.4.6. Additional analyses

We further performed seven additional analyses to examine the performance of the proposed specific models in a more rigorous and complete manner.

- *Additional Analysis 1 (selection of number of classes in supervised k-means):* We aimed to check how the number of classes influence the proposed model design approach. Thus, we tested the performance of the proposed approach using different number of classes (c = 2, 3 or 5) in supervised clustering. When c = 2, the cutoff value was 5; when c = 3, the cutoff values were 4 and 6; when c = 5, the cutoff values were 2, 4, 6 and 8.
- *Additional Analysis 2 (selection of cutoff values in supervised k-means):* We further examined the performance of the proposed approach under different cutoff values when c = 2. Five different cutoff values (2, 4, 5, 6, and 8) were under comparison.
- *Additional Analysis 3 (performance of using other regression models):* We tested the performance of the proposed approach by using other regression methods rather than the PLSR model. Two popular regression models, support vector regression (SVR) and LASSO-principal component regression (LASSO-PCR), were used in the proposed approach.
- *Additional Analysis 4 (performance comparison among different features used for both supervised k-means and PLSR):* We compared the prediction results when each type of feature is used for both clustering and prediction. More precisely, each type of feature was (1) used in supervised clustering to measure distances among samples for selection of samples and (2) used in PLSR models as features to predict pain ratings.
- *Additional Analysis 5 (comparison of fitting performance on training samples):* We examined the model performance on selected train-

ing samples when these selected samples were used to generate a model for specific test samples. More precisely, for each test individual/trial, we used the proposed approach to select a cluster of training samples to train an individual- or trial-specific model. We then checked the specific model's fitting performance on these selected training samples.

- *Additional Analysis 6 (correlation between actual and predicted pain ratings):* Correlation between actual and predicted values is often used to evaluate the performance of a prediction model. Thus, we checked the correlation between actual and predicted values for the proposed specific models, which used the features Whole Brain for both clustering and prediction. Note that, as pointed out in a recent review paper (Poldrack et al., 2020), correlation is not an appropriate metric and should be avoided when evaluating the predictive performance, because correlation is not sensitive to scaling of the data and could be biased in the case of leave-one-out cross validation. So, we just used correlation as an additional metric to evaluate the proposed approach.
- *Additional Analysis 7 (binary classification of high-pain and low-pain):* The proposed approach can also be used to make binary classification by categorizing samples into high-pain (VAS>5, positive) and low-pain (VAS≤5, negative). For the binary classification problem, we used SVM as the classifier, and other procedures were exactly the same as those in continuous prediction. We also examined and compared the performances of non-specific models, individual-specific models and trial-specific models.

2.4.7. Inter-individual/trial-subgroup variability of fMRI features

We further explored inter-subgroup difference of whole-brain fMRI responses at the individual level as well as at the trial level in a voxel-wise manner, with the aim to identify brain regions showing significant inter- or intra -individual differences. When training the individual-specific prediction model, candidate training trials were divided into several individual-subgroups. Exploring the differences in fMRI features among these individual-subgroups (called as inter-individual-subgroup variability) would be useful in understanding the inter-individual variability in pain-related brain responses. Because of the leave-one-individual-out cross validation used, different rounds of cross validation had different individual-subgroups and different number of subgroups (2, 3, or 4). If there were two individual-subgroups, two-sample *t*-test was used to compare the difference of whole-brain fMRI features between two individual-subgroups in a voxel-wise manner; if the number of individual-subgroups is greater than two, fMRI features of different individual-subgroups were compared using one-way ANOVA, followed by post-hoc pairwise comparisons between each pair of individual-subgroups. The multiple comparison problem was corrected by using FWE rate. Then, we calculated the probability of significant inter-individual-subgroup difference for each voxel (i.e., the percentage of individuals who have significant inter-individual-subgroup difference at this voxel). Similarly, we also calculated inter-trial-subgroup variability to explore differences in fMRI features among trial-subgroups. To achieve this, we compared the difference of features between trial-subgroups using two-sample *t*-test (if there were only two trial-subgroups) or one-way ANOVA (if there were only more than two trial-subgroups) for each individual. Subsequently, we calculated the probability of significant inter-trial-subgroup difference for each voxel (i.e., the percentage of trials which have significant inter-trial-subgroup difference at this voxel). Note that, because there were 32 individuals and each individual had 40 trials, we had in total 1280 trials to calculate the percentage of trials having significant inter-trial-subgroup difference. The histograms of the probabilities of significant inter-individual-subgroup and inter-trial-subgroup differences of all voxels were estimated, and the brain regions exhibiting large inter-individual-subgroup or inter-trial-subgroup differences were illustrated.

2.4.8. Validation of the proposed approach on an independent dataset

To check the generalizability of the specific models, we tested the proposed approach on a new independent dataset. More precisely, training samples and test samples were from two independent pain-evoked fMRI datasets, which were acquired at two sites. We evaluated the generalizability of the proposed approach by training the models using Dataset I and testing the models using Dataset II. More precisely, for each individual and each trial in Dataset II, we used the proposed model design approach to select training samples from Dataset I.

3. Results

3.1. Group-level pain-related fMRI patterns

Fig. 3A shows the brain regions activated (or de-activated) by nociceptive pain stimuli, as revealed by GLM at the group level. Positively activated regions include SI, SMA, ACC, MCC, thalamus, insula, while negatively activated region is the ventromedial prefrontal cortex (vmPFC). Fig. 3B shows the brain regions which are predictive of pain intensity, as revealed by PLSR at the group level. Positively predictive regions include S1, SMA and insula, while a negatively predictive region is vmPFC. Supplementary Table 1 lists the peak coordinates and cluster sizes of "Pain-activated" and "Pain-predictive" regions. Fig. 3C shows the "pain matrix", including SI, SII, PFC, ACC, insula, SMA and thalamus, which are segmented using the Brodmann and AAL templates.

3.2. Performance of the proposed approach

By using four different sets of fMRI features (Whole Brain, Pain Matrix, Pain-Activated, and Pain-Predictive) for supervised clustering to select training samples, we compared the prediction accuracy of three models: the proposed individual-specific model, the proposed trial-specific model, and the conventional non-specific model, which was trained on $(N-1)$ individuals and tested on the remaining one individual. Note that, these four different feature sets were only separately used and compared in the supervised clustering, while the same feature set, Whole Brain, was used in the PLSR model for pain prediction. We did this analysis because we aimed to check the robustness of supervised clustering (the core algorithm of the proposed method) by using different feature sets. Actually, when these features were used for supervised clustering only and the Whole Brain features were used in PLSR, there was no significant difference among different feature sets for individual-specific models ($P = 0.309$, ANOVA) or trial-specific models ($P = 0.284$, ANOVA). Prediction errors (MAE) of the conventional non-specific model, individual-specific model and trial-specific model were provided in Table 1and results of ANOVA and post-hoc paired-sample *t*-test are listed in Table 2. Fig. 4 provides the prediction performance in terms of MAE by the three models for each individual. These results show that, almost all the pair-wise comparisons, except the comparison between the individual-specific model and the non-specific model by using Pain Matrix as features ($P = 0.077$), have significant differences. Specifically, the trial-specific model has significantly higher accuracy (lower MAE) than the individual-specific model. Fig. 4 also suggests that the proposed specific models can improve prediction accuracy consistently for most participants.

The results of seven additional analyses (as introduced in Section 2.4.6) are mainly shown in the supplementary materials, and some key results are summarized as below.

- *Additional Analysis 1 (selection of number of classes in supervised k-means):* The results in Supplementary Fig. S1 show that the proposed specific models had better performance than conventional non-specific models. However, we can see from Supplementary Fig. S1 that individual-specific model and trial-specific model had lower prediction performance with the increase of *c*, which is explained in Section 4.5.

Table 1
Prediction errors (MAE, mean±std) of the conventional non-specific model, individual-specific model and trial-specific model. Four types of fMRI features were separately used for supervised clustering and the Whole Brain features were used for prediction.

fMRI feature used for supervised k-means	Whole brain	Pain matrix	Pain-activated	Pain-predictive
Non-specific model[a]	2.718±0.592[a]			
Individual-specific model	2.571±0.525	2.586±0.700	2.542±0.623	2.487±0.609
Trial-specific model	2.445±0.568	2.346±0.567	2.394±0.592	2.357±0.506

[#a] Note: For the non-specific model, there was no any selection of training samples, and the supervised clustering was not used. So, the four feature sets listed were not used and there was only one value for the non-specific model.

Table 2
Comparison among different pain prediction models based on four different fMRI feature sets used for supervised *k*-means.

fMRI feature used for supervised k-means	Whole brain	Pain matrix	Pain-activated	Pain-predictive
ANOVA	$P < 0.001$	$P < 0.001$	$P < 0.001$	$P < 0.001$
Post-hoc test (individual-specific vs. non-specific)	$P = 0.011$	$P = 0.077$	$P = 0.003$	$P < 0.001$
Post-hoc test (trial-specific vs. non-specific)	$P < 0.001$	$P < 0.001$	$P < 0.001$	$P < 0.001$
Post-hoc test (individual-specific vs. trial-specific)	$P = 0.014$	$P = 0.009$	$P = 0.010$	$P = 0.005$

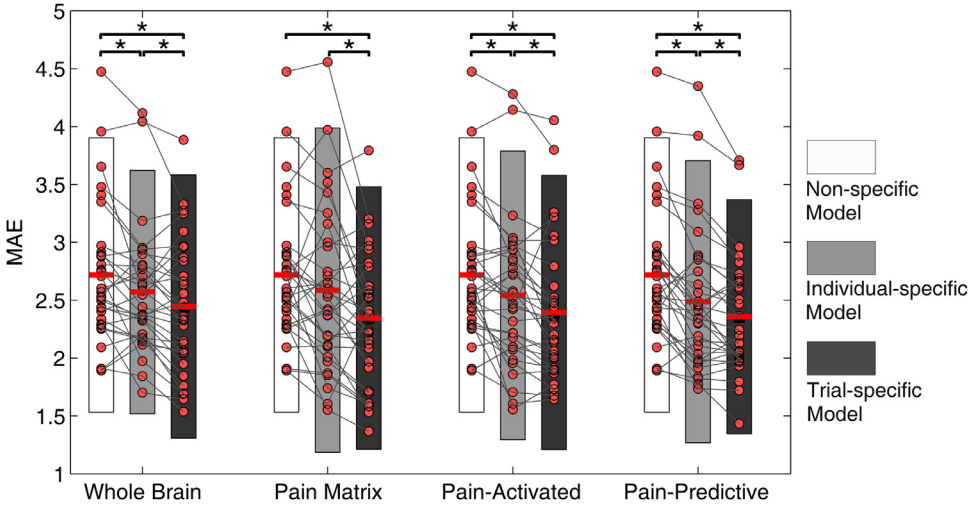

Fig. 4. Comparisons of prediction accuracy (in terms of MAE) among the conventional non-specific model, individual-specific model and trial-specific model. Four types of fMRI features (Whole Brain, Pain Matrix, Pain-Activated, and Pain-Predictive) were used for supervised clustering and the Whole Brain features were separately used for prediction. Each red dot denotes the averaged MAE value of one participant, and the MAE values of the same participants are connected with thin lines. White and gray box shadows indicate 95% confidence intervals (upper and lower boundaries of these box shadows are mean±1.96*std, and the red lines are mean values). Significant differences obtained by post-hoc paired *t*-test are indicated by asterisks ($P < 0.05/3$).

- *Additional Analysis 2 (selection of cutoff values in supervised k-means):* The results in Supplementary Fig. S2 show that different cutoff values did not significantly change our observation: the proposed specific models are more accurate than conventional non-specific models. Of course, some extreme cutoff values (such as 2 and 8) should not be used because they make samples highly unbalanced (i.e., the number of samples in two classes is largely different), which decreases the performance of machine learning models (especially for trial-specific models, as shown in Supplementary Fig. S2-B).
- *Additional Analysis 3 (performance of using other regression models):* Supplementary Fig. S3 shows that LASSO-PCR, SVR and PLSR can achieve similar conclusion that the individual-specific model and the trial-specific model significantly improved prediction accuracy compared with the performance of the non-specific model, and trial-specific model had superior performance over the individual-specific model. However, there is no difference between the results of different ML models (PLSR, SVR, LASSO-PCR) ($P = 0.48$, ANOVA).
- *Additional Analysis 4 (performance comparison among different features used for both supervised k-means and PLSR):* Supplementary Fig. S4 shows that, there was significant difference in prediction accuracy among non-specific, individual-specific, and trial-specific models for Whole Brain ($P = 3.95 \times 10^{-4}$), Pain Matrix ($P = 1.97 \times 10^{-5}$), and Pain-activated ($P = 3.03 \times 10^{-5}$) features, but there was no significant difference among non-specific and specific models for features Pain-predictive ($P = 0.842$). The possible reason for the lower performance of Pain-predictive features is that, the number of Pain-Predictive features is much smaller than others (the average number of features for each feature set is, Whole Brain: 55,017, Pain Matrix: 12,453, Pain-Activated: 17,588, Pain-Predictive: 385). Of course, more features do not guarantee higher accuracy. But in this case, it is more likely that the number of Pain-Predictive features is too small so that this feature set contains less predictive information than other feature sets. In addition, we also compared the prediction accuracies of models with different feature sets using ANOVA, and found no significant difference among four features for non-specific models ($P = 0.117$) and for individual-specific models ($P = 0.475$). There was significant difference among different feature sets for trial-specific models ($P = 0.031$, ANOVA). Post-hoc paired *t*-test showed that, the prediction errors of using Pain-predictive features were marginally higher than those of using Whole Brain ($P = 0.049$), Pain Matrix ($P = 0.035$), and Pain-activated ($P = 0.019$) for trial-specific models.
- *Additional Analysis 5 (comparison of fitting performance on training samples):* Supplementary Fig. S5 shows that, there is significant difference in prediction accuracy among non-specific, individual-specific, and trial-specific models ($P = 2.17 \times 10^{-9}$). Post-hoc paired *t*-test show that, the prediction errors of non-specific model were

higher than those of individual-specific model ($P = 9.54 \times 10^{-7}$), and trial-specific model ($P = 1.14 \times 10^{-18}$), while trial-specific models were marginally better than individual-specific models and ($P = 0.027 > 0.05/3$). Apparently, the fitting errors on training data (selected samples) were much smaller than prediction errors on test data, but they shared a similar pattern (i.e., trial-specific model < individual-specific model < non-specific model). The possible reason is that, individual- and trial-specific models used training samples with similar feature distributions, so that these samples may also share similar model parameters. As a result, the fitting performance on selected training samples is better for specific models.

- *Additional Analysis 6 (correlation between actual and predicted pain ratings):* The correlation results between actual and predicted values for the proposed specific models were shown in Supplementary Fig. S6. The Z-transformed correlation coefficients of three models were 0.46±0.075 (mean±std) for non-specific models, 0.49±0.054 for individual-specific models, and 0.53±0.061 for trial-specific models, respectively. One-sample *t*-test showed that correlation values of non-specific models and specific models are all significantly larger than zero ($P = 1.15 \times 10^{-10}$, 4.57×10^{-13}, 3.14×10^{-13}, for non-specific, individual-specific, and trial-specific models, respectively). Although the correlation coefficients of different models showed a similar pattern as the prediction accuracy (i.e., trial-specific model > individual-specific model > non-specific model), there is no significant difference in Z-transformed correlation coefficients among non-specific, individual-specific, and trial-specific models ($p = 0.248$, ANOVA).
- *Additional Analysis 7 (binary classification of high-pain and low-pain):* The results of binary classification are illustrated in Supplementary Fig. S7. We have three major observations from this figure. First, the proposed individual- and trial-specific models were still better than conventional non-specific models in terms of accuracy, specificity, and AUC, while trial-specific models had the best results. Second, in terms of sensitivity, there was no significant difference among three types of models. Third, the sensitivity of all these models is low, which means high-pain trials are difficult to be correctly detected.

In the cross-site study, the results in Fig. 5 demonstrate that the proposed approach has strong cross-site generalizability. The prediction errors (MAE, mean±std) of non-specific, individual-specific, and trial-specific models were respectively 2.907±0.716, 2.696±0.734, and 2.449±0.844. There was significant difference in prediction accuracy among non-specific, individual-specific, and trial-specific models ($P = 8.16 \times 10^{-5}$). Post-hoc paired *t*-test show that, the prediction errors of non-specific models were higher than those of individual-specific models ($P = 0.0027$) and trial-specific models ($P = 1.32 \times 10^{-4}$), while individual-specific models had higher prediction errors than trial-specific models ($P = 0.0087$). We can see that the proposed approach can achieve better prediction accuracy for most of individuals. In addition, by comparing the results on two datasets (Figs. 4 and 5), we can see that the test performance on Dataset II was relatively lower (as compared with the corresponding results obtained using the Whole Brain features in Dataset I), which was reasonable because these two datasets had different feature distributions. However, when we used two-sample *t*-test to compare the MAE values of non-specific and specific models (all of which used the Whole Brain features for supervised clustering and prediction) between Dataset I and Dataset II, but did not find any significant difference ($P = 0.218$ for non-specific models, $P = 0.407$ for individual-specific models, $P = 0.985$ for trial-specific models).

Collectively, all above results validated that the proposed model design approach is effective, robust, and generalizable.

3.3. Brain regions exhibiting inter-subgroup variability

Fig. 6 shows the histograms of the probabilities of significant inter-individual-subgroup or inter-trial-subgroup differences of all voxels.

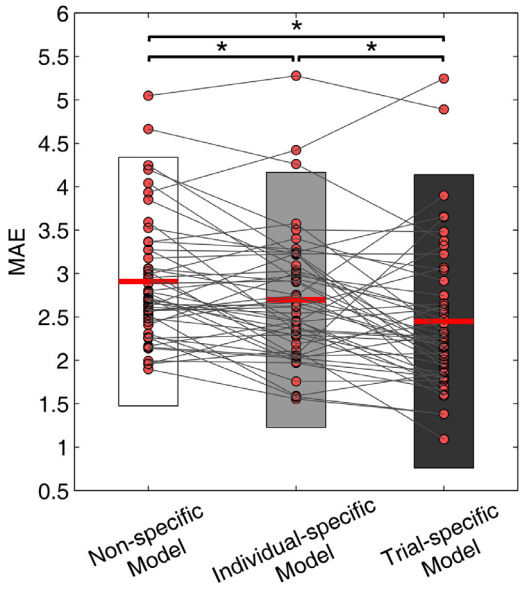

Fig. 5. Comparisons of prediction accuracy (in terms of MAE) among the conventional non-specific model, individual-specific model and trial-specific model, which are trained using Dataset I and tested on an independent Dataset II. The Whole Brain features were used for both clustering and prediction. Each red dot denotes the MAE value of one participant, and the MAE values of the same participants are connected with thin lines. White and gray box shadows indicate 95% confidence intervals (upper and lower boundaries of these box shadows are mean±1.96*std, and the red lines are mean values). Significant differences obtained by post-hoc paired *t*-test are indicated by asterisks (* $P < 0.05/3$).

It can be seen from Fig. 6 that, there are a large number of voxels showing significant difference between individual- or trial-subgroups. A great number of voxels had significant difference between individual-subgroups with a probability around 50%, while many voxels have significant difference between trial-subgroups with a probability ranging from 10% to 60%. Fig. 7 shows the brain regions with large inter-individual-subgroup differences and inter-trial-subgroup difference in the fMRI activation patterns, respectively. It can be seen from Fig. 7 that, the brain regions exhibiting large inter-individual-subgroup differences include the pain matrix and the visual cortex, while the brain regions exhibiting large inter-trial-subgroup differences include S1, MCC, thalamus, precuneus, and visual cortex. More details of brain regions with significant inter-individual-subgroup difference and inter-trial-subgroup difference can be found in Supplementary Tables S2 and S3.

4. Discussion

In the present study, we proposed a novel individual-specific and trial-specific fMRI-based pain prediction approach which can effectively restrain the adverse influence of inter-individual and intra-individual differences on pain prediction. The proposed new approach showed significantly higher prediction accuracy than the conventional method which used a non-specific model for all individuals or trials under test. In the following, we will first summarize the relevance and novelty of this study, and then elucidate why and how the new approach can successfully improve the accuracy of pain prediction. The limitation and future clinical implications of the proposed approach will be also discussed.

4.1. General relevance and methodological novelty

This study is aimed to develop an approach which can produce a pain prediction model specific to each test individual and even each test trial, so it is an important step towards solving the problem of individual variability. Further, because other modalities of data used for

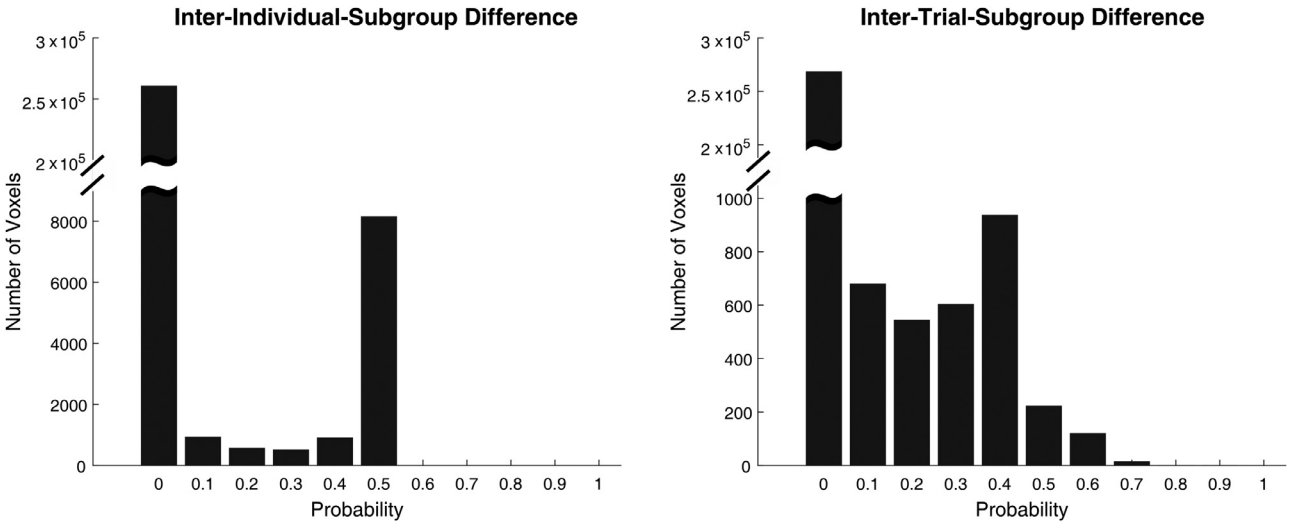

Fig. 6. Histograms of the probabilities (with respect to all individuals or all trials) of significant inter-individual-subgroup or inter-trial-subgroup differences of all voxels in the whole brain. A voxel was counted if it had significantly different values between individual-subgroups or trial-subgroups ($P_{FWE} < 1 \times 10^{-70}$; two-sample *t*-test if the number of subgroups is 2, or one-way ANOVA if the number of subgroups is greater than 2).

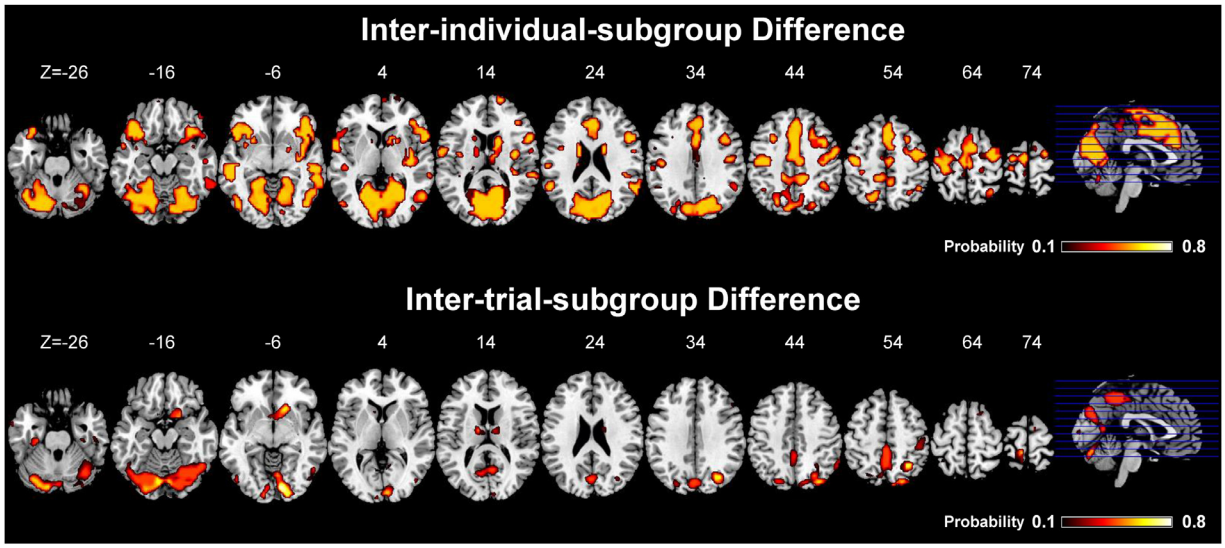

Fig. 7. Brain regions with significant inter-individual-subgroup difference and inter-trial-subgroup difference in their fMRI activation patterns. Colors indicate the probabilities (with respect to all individuals or all trials) of significant inter-individual-subgroup or inter-trial-subgroup differences of each voxel.

pain assessment (such as electrocardiogram, electrodermal activity, surface electromyography, facial expression, etc.) also have large individual variability, this study has a general relevance to the development of automatic pain assessment techniques. From a broader perspective, machine learning has been very popularly used in the neuroimaging community with aims to discover neural signatures and to diagnose brain diseases, but individual difference is a major source of data heterogeneity, which has an adverse effect on the performance of machine learning models. So, the present study is also relevant to other similar applications where neuroimaging-based machine learning models are used.

The novelty of the proposed approach is two-fold. First, there are few studies concerning the development of an fMRI-based pain prediction model that can handle individual variability. The proposed approach can design a specific model for each individual and each trial, and thus it can address both inter-individual variability and intra-individual variability. Second, to the best of our knowledge, this study is the first attempt to use the dynamic selection technique (Cruz et al., 2018), which is an active research topic in the machine learning community, for the design of pain prediction models. The basic idea of dynamic selection is to design a classifier according to each new sample to be classified and one common method is to train the classifier based on a local region of the feature space where the test sample is located (Cruz et al., 2018). The present study adopted the idea of dynamic selection to train an individual-specific or trial-specific model for each new test individual or trial based on their fMRI responses (the feature space). To conclude, the novelty of proposed approach includes, (1) it can well address the problems of two types of individual variability in fMRI-based pain prediction, (2) it uses the idea of dynamic selection to design a model specific to each test sample and can achieve better performance.

It is worth mentioning that, the "specific models" are indeed not generalizable, but the "approach" is generalizable. "Generalizability" means a well-trained machine learning model can be well used for new samples without any updating. But, in this study, the generalizability is realized in a different manner: the model is trained for each new sample. In another word, the proposed approach (not a model) is generalizable by designing a new model specific to each new sample. For a new sample in

a new test dataset, we can use the proposed approach to select training samples from the training dataset. If there are sufficient training samples having similar patterns as the test sample in the feature space, the resultant specific model can achieve good performance on the test sample. In the present study, we used a cross-site study to demonstrate the strong generalizability of the proposed approach. Even if the prediction models are trained from one dataset and applied on another independent dataset, the proposed individual- and trial-specific models can still achieve better performance than conventional non-specific models.

4.2. Principle and interpretability of specific models

From a machine learning point of view, an fMRI-based prediction model elucidates the relationship between fMRI activation patterns (features) and subjective pain intensities (labels), so the model is defined by the probability distributions of fMRI activation patterns and pain intensities. The rationale behind the proposed specific models is, if training samples and test samples have similar feature distributions, then the trained model has good results on test samples. This rationale is based on our previous work (Lin et al., 2018), which has shown that, the difference between two individuals' pain prediction models is determined by the differences between the probability distributions of their fMRI features and pain intensities. Hence, if one wants to develop a more accurate pain prediction model, it is of key importance to minimize the differences in the probability distributions of fMRI activation patterns and pain intensities between training samples and test samples. Because it is not possible to know the pain intensities (i.e., class labels) of test samples in advance, we can only rely on the minimization of the difference in fMRI activation features between training and test samples. This inspires us to only use part of training samples, which have similar fMRI features as the test samples, to design a model.

The new individual-specific and trial-specific prediction models were proposed based on the idea of the dynamic selection technique (Cruz et al., 2018). That is, only a subset of training samples, which have similar fMRI activation patterns as the test sample, should be used to train a model for one individual or one experimental trial. To achieve this, we clustered training samples into several subgroups based on their fMRI activation patterns and then selected one subgroup with the closest fMRI patterns to that of the test sample. The clustering method we used is a new supervised *k*-means method. As compared with the conventional clustering methods, such as *k*-means, the new supervised *k*-means method is able to identify clusters with high probability density with respect to one single class (pain-sensitive vs. pain-insensitive at the individual level, or high-pain vs. low-pain at the trial level), because class information is used in clustering. According to our previous study (Lin et al., 2018), the distance of fMRI features between training samples and test samples is positively correlated with the prediction accuracy. That is, if training samples and test samples are similar in their feature patterns, the model should have a higher accuracy, which was validated in this study.

It is necessary to note that, "individual-specific" and "trial-specific" models do not simply mean the models are different for each individual or trial. Even in the conventional design of models using cross validation, the models are different for individuals/trials because different training samples are assigned by cross validation. Here in the proposed approach, "individual/trial-specific" means these models are specifically designed for each individual/trial by selecting training samples in an individual/trial-specific manner.

Because the prediction model is specific to each sample, it is hard to illustrate and interpret all specific models for all individuals and trials (32 individual-specific models and 1280 trial-specific models for Dataset I) in a straightforward way. However, the designed specific model itself is still simple and has good interpretability because of the following reasons.

First, the PLSR model (no matter whether it is specific or non-specific) itself is not difficult to interpret: the magnitude of the model coefficient indicates the predictability of the BOLD signal at the corresponding voxel. For example, we can rank the PLSR coefficients to find the most predictive regions. PLSR is widely used for fMRI decoding and it has very good interpretability (McIntosh et al., 2004; Krishnan et al., 2011; O'Connell et al., 2018). In Fig. 3, we have shown that, by using the PLSR model, we can find pain-predictive regions. Because the identified pain-predictive regions (including S1, SMA, insula and vmPFC, etc.) are highly consistent with literature, we did not discuss these pain-related regions in this manuscript. Of course, each individual and even each trial may have some specific fMRI response patterns, but it is hard to interpret each specific model. However, we still take a deep look at these models and find brain regions that contribute to the inter-subgroup variability (see Section 4.4).

Second, the flexibility and researchers' degrees-of-freedom in model selection are not high. Subgroups are not arbitrary defined, and they are defined based on similarity of training samples in the feature space (i.e., fMRI responses). Subgroups cannot be defined based on arbitrary variables (such as some demographical and behavioral parameters), especially if they are not related to pain and are not included in the model as features to predict pain. In the present study, fMRI responses are well known to be related to subjective pain intensity and are used as features for pain prediction, so we use fMRI responses to define subgroups. Thus, the specific models are not arbitrarily defined and they are defined based on similarity of pain-related features.

To conclude, the proposed model design approach is not complicated and the designed specific model is not difficult to interpret. The PLSR model is neither nonlinear not complex, and it is just a multivariate linear model with fMRI signals as features. We have a strong rationale behind the proposed model design approach: the approach is based on our observation that prediction performance is positively correlated with the similarity between training samples and test samples and is also based on the common dynamic selection technique. To understand why the pain prediction models have large inter-subject or inter-trial variability, we identified and discussed the brain regions contributing to the individual difference in the prediction models (as discussed in Section 4.4).

4.3. "Less is more": quality of training samples matters

This basic idea of the proposed approach is different from the common viewpoint that more samples should be used to produce a more general and accurate prediction model. In general, it is true that a model trained from more data has better desired properties, which explains why the number of samples in neuroimaging study has gradually increased in recent years (Lindquist et al., 2017). However, the quality of data is as important as the amount of data (if not more so). To build a machine learning model, data should be clean, complete and correctly labeled, so data cleansing is a crucial step. More importantly, training data and test data should be homogeneous so that the trained model can be well applied on test samples. However, in the study of fMRI-based pain prediction, training data and test data may be largely different in their distributions of features and labels, because of the remarkable inter- and intra-individual differences in pain. As a result, we have to carefully screen the training data by only keeping a subset of training data that are similar to test data in their fMRI feature patterns.

Our results clearly showed that, the model trained from a subset of selected samples had better performance than the model trained from all samples. In this study, the number of samples used for training a pain prediction model was ordered as: non-specific model > individual-specific model > trial-specific model, while the prediction accuracy was ordered as: non-specific model < individual-specific model < trial-specific model. These results convincingly proved that a large number of training samples do not guarantee good performance of a machine learning model. Training samples should be carefully selected to match the properties of test samples so that the trained model can obtain satisfactory performance on test samples.

It must be emphasized that, this study did not imply that we should not collect more data when building a machine learning model. Actually, large amount of data is crucial for any machine learning application. We only argue that, the amount of data used to train a model for any specific test individual could be small if the properties of data match the test individual well. Even in our study, the performance of our proposed approach also greatly depends on the amount of candidate training data. Only if the amount of candidate training data is large, we can have more accurately-clustered individual/trial-subgroups and enough data samples from selected individual/trial-subgroups to train a more precise model.

4.4. Brain regions exhibiting inter-subgroup variability

In the study, we identified a set of brain regions showing large difference among individual-subgroups or trial-subgroups. Inter-individual-subgroup difference of fMRI patterns is mainly located in the pain matrix and visual cortex, while inter-trial-subgroup difference of fMRI patterns is mainly located in S1, MCC, precuneus, thalamus and visual cortex.

As for the brain regions showing inter-individual-subgroup difference, it is not surprising to identify that the pain matrix has significantly different fMRI activation patterns among individual subgroups. But, it is interesting to observe the visual cortex showing large inter-individual-subgroup difference, though the visual cortex was not a pain-activated or pain-predicative region (as seen from Fig. 3). This observation suggests that the visual cortex also plays a functional role in pain perception. Conventionally, visual cortex is not considered as a core region of pain perception, but a series of studies have shown that it may also be directly or indirectly involved in the processing of nociceptive stimuli (Sava et al., 2014). For example, visual regions might modulate network of inhibitory interneurons in early somatosensory regions, which is an important region perceiving the pain (Cardini et al., 2011).

As for the brain regions showing inter-trial-subgroup difference, the observed brain regions within the pain matrix include S1, MCC, and thalamus. Similarly to the observations in the inter-individual-subgroup difference, the visual cortex also showed inter-trial-subgroup difference. It is reasonable because visual perception has considerable within-individual moment-to-moment variability. Besides, the precuneus is also identified to exhibit large inter-trial-subgroup difference. Some studies have shown that precuneus is predictive of pain perception (Mouraux et al., 2011; Wager et al., 2013), but, as discussed in (Lin et al., 2018; Goffaux et al., 2014), the modulation effect of the precuneus activity on pain may not be specific to pain, because precuneus plays an important role in salience processing.

4.5. Methodological considerations and alternatives

The proposed approach is based on the idea of dynamic selection, which is a widely-used machine learning technique to design more specific models for test samples. As shown in Fig. 1, two crucial steps in the proposed approach are supervised clustering and the machine learning model. In the following, we will discuss and examine how different parameters in supervised clustering and different machine learning models influence the performance of the proposed approach.

In the operation of supervised clustering, the subgroups (clusters) were primarily determined based on fMRI patterns, while class labels (pain perception) helped make subgroups more homogeneous (i.e., the samples in one cluster belong to the same class). In this work, we set the number of classes c in the objective function of supervised clustering to 2 (pain-sensitive or pain-insensitive for individuals, and low pain or high pain for trials), and M (the number of subgroups) was calculated as 2, 3, or 4, depending on the data. Suppose the number of classes c is larger, the number of clusters M will also be larger, leading to a smaller number of samples per one cluster (subgroup). As a result, the number of samples in the subgroup used to train a model becomes smaller, which could decrease the performance of the model. We tested the performance of the proposed approach using other values of c (3 and 5). The results (see Supplementary Fig. S2) also showed that the proposed specific models had better performance than conventional non-specific models. However, we can see from Supplementary Fig. S2 that individual-specific model and trial-specific model had lower prediction performance with the increase of c, which agreed with our inference. We further found that the performance of the proposed approach was not significantly influenced by different cutoff values (as shown Supplementary Fig. S3). Since it is impossible to exhaust all possible combinations of these parameters to find the optimal settings, we adopted the current setting ($c = 2$, cutoff value = 5) in this study. The proposed approach can also use other regression methods rather than the PLSR model, but, as shown in Supplementary Fig. S4, different regression methods did not have significantly performance, while they can all approve that the individual-specific model and the trial-specific model are more accurate than the non-specific model.

It is necessary to discuss the fMRI features we extracted for pain prediction. We used the fourth scan after the stimulus onset as features, because the fourth scan is normally the peak activation of BOLD signals and its magnitude is strongly correlated with the intensity of pain. BOLD responses at discrete time points (especially at the activation peaks) are often used as features for pain prediction, such as (Marquand et al., 2010; Brown et al., 2011). Our previous papers (Tu et al., 2016a, 2016b, 2018; Lin et al., 2018; Anter et al., 2020) also extracted the peak BOLD responses as features to predict the intensity of pain and achieved satisfactory results. On the other hand, beta-series coefficients (Rissman et al., 2004; Mumford et al., 2012) are also widely used as features in pain prediction (Wager et al., 2013; Lindquist et al., 2017). We checked the correlation between the BOLD magnitudes at the fourth scan and beta-series coefficients at the single-trial level for the pain-activated regions, and found they are very strongly correlated ($R = 0.95$, $P = 2 \times 10^{-5}$). Hence, we believe both types of features could result in reasonable prediction performance.

Further, although the main results in this study are based on continuous prediction, the proposed approach can be used to make binary classification as well. As shown in Supplementary Fig. S5, the specific models still can achieve better results than conventional classifiers in terms of accuracy, specificity, and AUC when classifying high-pain and low-pain trials. However, there was no significant difference among three types of models in terms of sensitivity. It can be seen from confusion matrices that, the major improvement of the proposed specific models stemmed from the increased number of correctly classified low-pain samples. Because the number of low-pain samples was much larger than high-pain samples (665 vs. 535), so the classification performance in classifying true low-pain samples had a larger influence on the overall performance. Also, the sensitivity of all these models is low, which means high-pain trials are difficult to be correctly detected. One possible reason is that, high-pain trials have relatively large variability in their pain-evoked BOLD responses, and some of them are overlapped with low-pain trials in the feature space. As shown in Supplementary Fig. S5-D, features of high-pain trials have fewer samples but larger dispersion. Hence, some high-pain samples are overlapped with low-pain trials in the feature space and it is not easy to achieve high sensitivity in such a circumstance.

In addition, we used the pain ratings and fMRI responses of the first two energy levels (E1 and E2) as baseline to normalize the pain ratings and fMRI responses of the other two energy levels (E3 and E4), and then predict the changes in pain perception observed in E3 and E4 energy levels. Still, we used the PLSR model and the proposed individual-specific and trial-specific models to make prediction. The prediction accuracies (mean±SEM) for non-specific model, individual-specific model, and trial-specific model were respectively 2.05±1.3, 2.04±1.4, and 1.79±1.1. The results showed that, the proposed trial-specific models had significantly better performance than non-specific models ($P = 7.5050 \times 10^{-4}$, paired t-test) and individual-specific models ($P = 0.0033$, paired t-test), but individual-specific modes were only

slightly (not significantly) better than non-specific models ($P = 0.74$, paired t-test). Overall, the results suggest that the proposed approach can still achieve better performance in this scenario (to predict the changes in perception observed in E3 and E4 energy levels using E1 and E2 as baselines).

In summary, the proposed approach has robust and consistent performance. Although different parameters in supervised clustering and different machine learning models could influence the performance of the proposed approach, the resultant individual-specific and trial-specific models normally have higher accuracy than non-specific models in different testing scenarios.

4.6. Limitations and future work

The advantages of the proposed approach have been demonstrated and discussed in previous sections, but the present study still has some limitations to overcome before it can be applied in practice. First, the approach was proposed for and validated on a nociceptive pain dataset, and it could be extended to be used on other sensory (visual, auditory, somatosensory, etc.) event-related fMRI data and resting-state fMRI, which is more popularly used especially in translational study of pain. It is not difficult to apply the basic idea of the proposed approach on resting-state fMRI data for an individual-specific prediction model. We can cluster individuals into a small number of subgroups based on their resting-state fMRI features, such as the amplitude of low frequency fluctuations (Zang et al., 2007) and resting-state networks (Greicius et al., 2003), and then select a training subgroup with the closest fMRI patterns with the test individual to train an individual-specific model. Note that, the trial-specific model is not applicable for resting-state fMRI data, which does not contain any event-related trials.

Second, the number of participants used in this study was still small and only one dataset was tested. It is always necessary to rigorously validate a machine learning model on multiple datasets and large-sample datasets. Also, if more information about the participants (such as demographic variables, behavioral parameters, co-morbidities, and conditions affecting perception) could be recorded and used as independent variables in the pain prediction model, the model's prediction performance could be improved, because such information could also be related to participants' perceived levels of pain. However, we did not record much information during subject recruitment and experiments, and hence could not do further analyses. It will be definitely useful to record this information in future. If more information (such as some demographic and behavioral variables) of participants is available, we can check whether such information is predictive of pain ratings and how such information is related to fMRI responses. Some variables that can provide more predictive capability than fMRI could be considered as new features, and a new type of pain prediction model based on the feature space defined by both fMRI activities and these predictive variables can be established. For example, participants having different co-morbidities may be clustered into different subgroups for the training of individual-specific models.

Third, although the proposed pain prediction model only uses fMRI responses as features, it does not mean the prediction model is only based on pain-related neural signals. fMRI signals generally contain many confounding factors, which may have an important influence on the prediction model. To make clear whether a confounding factor, such as head motion, contributes to a machine learning model, we can examine whether this confounding factor alone (if it can be estimated or measured) is correlated with the variables to be predicted and has any predictive power in the model. For example, to check the potential effects of head motion on the pain prediction results, we computed the frame-wise displacement (FD) of head motion per trial (at the same time as the fMRI features were extracted, i.e., the 4th scan after stimulus onset). Then, we calculated the correlation coefficient between FD values and pain ratings per participant and found that the correlation coefficients (after Fisher Z-transformation) of all participants were not significantly different from zero (correlation coefficients=0.03±0.20, $P = 0.47$, one-sample t-test). This result implied that head motion was not correlated with pain ratings and thus was not predictive of pain ratings. In a recent paper (Tu et al., 2020) which classified migraine patients and healthy controls based on fMRI functional connectivity, we used a similar way to check possible influence of two confounding factors: head motion and drowsiness. We found that, neither head motion nor drowsiness was significantly different between migraine patients and healthy controls and they did not provide significant predictive power for the classifier. Thus, the influence of head motion or drowsiness on the migraine classification model was excluded (Tu et al., 2020). However, because fMRI signals may contain many confounding factors, some of which cannot be measured or estimated, it is almost impossible to examine whether all possible confounding factors have an influence on the fMRI-based prediction model. Therefore, the proposed specific pain prediction model may not depend on neural signals only, even the features used are fMRI signals only. Future studies should investigate the possible influence of other confounding factors on pain prediction, which will be helpful in designing a more accurate and reliable pain prediction model.

Fourth, the computational complexity of the proposed approach could be high because it needs to train a new model for each test individual and even for each trial. However, on the other hand, because the proposed approach selectively used training samples to train a model, it can effectively reduce the sample size of the training set and shorten the training time of the model.

Last but not the least, we still need to explore the applicability of the proposed approach on pain datasets recorded from clinical practices. The proposed individual-specific and trial-specific pain prediction approach could be executed reliably and automatically and could potentially address existing problems of cross-individual pain prediction, so it has the potential to meet the requirements of clinical applications. Successful validation of the proposed method on clinical pain data is a key step towards a precise and individualized pain assessment tool.

5. Conclusions

Remarkable inter- and intra-individual differences in pain experience and pain-related brain greatly decrease the accuracy of fMRI-based pain prediction models. It is desired to design more accurate and individualized pain prediction models, which are not sensitive to individual difference in pain, based on advanced machine learning techniques. To this end, we developed new individual-specific and trial-specific models to estimate the intensity of pain from single-trial fMRI data. Different from the common belief that more data samples should be used to train a more accurate model, the proposed models only selected a small subset of training samples, which have similar features to test samples, to train a specific model for test individuals or even trials. The superior performance of the proposed new models over the conventional non-specific models was validated on a nociceptive-evoked fMRI dataset. The proposed new individual-specific and trial-specific fMRI-based pain prediction models could be potentially used to develop more a precise and objective pain assessment tool.

Credit authorship contribution statement

Qianqian Lin: Conceptualization, Formal analysis, Methodology, Writing - original draft. **Gan Huang:** Software. **Linling Li:** Data curation. **Li Zhang:** Validation. **Zhen Liang:** Writing - review & editing. **Ahmed M. Anter:** Formal analysis. **Zhiguo Zhang:** Supervision, Conceptualization, Methodology, Writing - review & editing.

Data and code availability statement

Data and code are available upon direct request to the corresponding author.

Acknowledgments

We thank Li Hu and Yazhuo Kong (Institute of Psychology, Chinese Academy Sciences), Meng Liang (Tianjin Medical University), and Jixin Liu (Xidian University), for jointly designing, planning, and conducting experiments. We also thank Dr. Yiheng Tu and Xiaoxia Zhang for their help during the revision. This work was supported by National Natural Science Foundation of China (No. 81871443); Science, Technology and Innovation Commission of Shenzhen Municipality Technology Fund (No. JCYJ20170818093322718), and Shenzhen Peacock Plan (No. KQTD20160531120514 97).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2020.117506.

References

Anter, A.M., Huang, G., Li, L.L., Zhang, L., Liang, Z., Zhang, Z.G., 2020. A new type of fuzzy rule-based system with chaotic swarm intelligence for multi-classification of pain perception from fMRI. IEEE Trans. Fuzzy Syst. 28 (6), 1096–1109. doi:10.1109/TFUZZ.2020.2979150.

Bai, Y., Huang, G., Tu, Y.H., Tan, A., Hung, Y.S., Zhang, Z.G., 2016. Normalization of pain-evoked neural responses using spontaneous EEG improves the performance of EEG-based cross-individual pain prediction. Front. Comput. Neurosci. 10, 1–10. doi:10.3389/fncom.2016.00031.

Boly, M., Balteau, E., Schnakers, C., Degueldre, C., Moonen, G., Luxen, A., Phillips, C., Peigneux, P., 2007. Baseline brain activity fluctuations predict somatosensory perception in humans. Proc. Natl. Acad. Sci. 104, 12187–12192. doi:10.1073/pnas.0611404104.

Brodersen, K.H., Wiech, K., Lomakina, E.I., Lin, C.shu, Buhmann, J.M., Bingel, U., Ploner, M., Stephan, K.E., Tracey, I., 2012. Decoding the perception of pain from fMRI using multivariate pattern analysis. Neuroimage 63, 1162–1170. doi:10.1016/j.neuroimage.2012.08.035.

Brown, J.E., Chatterjee, N., Younger, J., Mackey, S., 2011. Towards a physiology-based measure of pain: patterns of human brain activity distinguish painful from non-painful thermal stimulation. PLoS One 6, 2–9. doi:10.1371/journal.pone.0024124.

Buffum, M.D., Hutt, E., Chang, V.T., Craine, M.H., Snow, A.L., 2007. Cognitive impairment and pain management: review of issues and challenges. J. Rehabil. Res. Dev. 44, 315. doi:10.1682/jrrd.2006.06.0064.

Cardini, F., Longo, M.R., Haggard, P., 2011. Vision of the body modulates somatosensory intracortical inhibition. Cereb. Cortex 21, 2014–2022. doi:10.1093/cercor/bhq267.

Cecchi, G.A., Huang, L., Hashmi, J.A., Baliki, M., Centeno, M.V., Rish, I., Apkarian, A.V., 2012. Predictive dynamics of human pain perception. PLoS Comput. Biol. 8. doi:10.1371/journal.pcbi.1002719.

Coghill, R.C., McHaffie, J.G., Yen, Y.F., 2003. Neural correlates of interindividual differences in the subjective experience of pain. Proc. Natl. Acad. Sci. 100, 8538–8542. doi:10.1073/pnas.1430684100.

Coghill, R.C., 2011. Individual differences in the subjective experience of pain: new insights into mechanisms and models. Headache 50, 1531–1535. doi:10.1111/j.1526-4610.2010.01763.x.

Cruccu, G., Sommer, C., Anand, P., Attal, N., Baron, R., Garcia-Larrea, L., Haanpaa, M., Jensen, T.S., Serra, J., Treede, R.D., 2010. EFNS guidelines on neuropathic pain assessment: revised 2009. Eur. J. Neurol. 17, 1010–1018. doi:10.1111/j.1468-1331.2010.02969.x.

Cruz, R.M.O., Sabourin, R., Cavalcanti, G.D.C., 2018. Dynamic classifier selection: recent advances and perspectives. Inf. Fusion 41, 195–216. doi:10.1016/j.inffus.2017.09.010.

De Jong, S., 1993. SIMPLS: an alternative approach squares regression to partial least. Chemom. Intell. Lab. Syst. 18, 2–263. doi:10.1016/0169-7439(93)85002-X.

Duff, E.P., Trachtenberg, A.J., Mackay, C.E., Howard, M.A., Wilson, F., Smith, S.M., Woolrich, M.W., 2012. Task-driven ICA feature generation for accurate and interpretable prediction using fMRI. Neuroimage 60, 189–203. doi:10.1016/j.neuroimage.2011.12.053.

Faber, N.M., Rajko, R., 2007. How to avoid over-fitting in multivariate calibration—the conventional validation approach and an alternative. Anal. Chim. Acta 595, 98–106. doi:10.1016/j.aca.2007.05.030.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.P., Frith, C.D., Frackowiak, R.S., 1994. Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Mapp. 2, 189–210. doi:10.1002/hbm.460020402.

Gan, H., Huang, R., Luo, Z., Xi, X., Gao, Y., 2018. On using supervised clustering analysis to improve classification performance. Inform. Sci. 454-455, 216–228. doi:10.1016/j.ins.2018.04.080.

Gholami, B., Haddad, W.M., Tannenbaum, A.R., 2010. Relevance vector machine learning for neonate pain intensity assessment using digital imaging. IEEE Trans. Biomed. Eng. 57, 1457–1466. doi:10.1109/TBME.2009.2039214.

Goffaux, P., Girard-Tremblay, L., Marchand, S., Daigle, K., Whittingstall, K., 2014. Individual differences in pain sensitivity vary as a function of precuneus reactivity. Brain Topogr. 27, 366–374. doi:10.1007/s10548-013-0291-0.

Greicius, M.D., Krasnow, B., Reiss, A.L., Menon, V., 2003. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. Proc. Natl. Acad. Sci. 100, 253–258. doi:10.1073/pnas.0135058100.

Gross, J., Schnitzler, A., Timmermann, L., Ploner, M., 2007. Gamma oscillations in human primary somatosensory cortex reflect pain perception. PLoS Biol. 5, 1168–1173. doi:10.1371/journal.pbio.0050133.

Gruss, S., Treister, R., Werner, P., Traue, H.C., Crawcour, S., Andrade, A., Walter, S., 2015. Pain intensity recognition rates via biopotential feature patterns with support vector machines. PLoS One 10. doi:10.1371/journal.pone.0140330.

Haanpää, M., Attal, N., Backonja, M., Baron, R., Bennett, M., Bouhassira, D., Cruccu, G., Hansson, P., Haythornthwaite, J.A., Iannetti, G.D., Jensen, T.S., Kauppila, T., Nurmikko, T.J., Rice, A.S.C., Rowbotham, M., Serra, J., Sommer, C., Smith, B.H., Treede, R.D., 2011. NeuPSIG guidelines on neuropathic pain assessment. Pain 152, 14–27. doi:10.1016/j.pain.2010.07.031.

Herr, K.A., Spratt, K., Mobily, P.R., Richardson, G., 2004. Pain intensity assessment in older adults: use of experimental pain to compare psychometric properties and usability of selected pain scales with younger adults. Clin. J. Pain 20, 207–219. doi:10.1097/00002508-200407000-00002.

Hochberg, Y., 1988. A sharper Bonferroni procedure for multiple tests of significance. Biometrika 75, 800–802. doi:10.1093/biomet/75.4.800.

Hu, L., Iannetti, G.D., 2016. Painful issues in pain prediction. Trends Neurosci. 39, 212–220. doi:10.1016/j.tins.2016.01.004.

Hu, L., Iannetti, G.D., 2019. Neural indicators of perceptual variability of pain across species. Proc. Natl. Acad. Sci. 116, 1782–1791. doi:10.1073/pnas.1812499116.

Hu, L., Xiao, P., Zhang, Z.G., Mouraux, A., Iannetti, G.D., 2014. Single-trial time-frequency analysis of electrocortical signals: baseline correction and beyond. Neuroimage 84, 876–887. doi:10.1016/j.neuroimage.2013.09.055.

Huang, G., Xiao, P., Hung, Y.S., Zhang, Z.G., Hu, L., 2013. A novel approach to predict subjective pain perception from single-trial laser-evoked potentials. Neuroimage 81, 283–293. doi:10.1016/j.neuroimage.2013.05.017.

Iannetti, G.D., Hughes, N.P., Lee, M.C., Mouraux, A., 2008. Determinants of laser-evoked EEG responses: pain perception or stimulus saliency? J. Neurophysiol. 100, 815–828. doi:10.1152/jn.00097.2008.

Krishnan, A., Williams, L.J., McIntosh, A.R., Abdi, H., 2011. Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review. Neuroimage 56 (2), 455–475. doi:10.1016/j.neuroimage.2004.07.020.

Lee, J., Mawla, I., Kim, J., Loggia, M.L., Ortiz, A., Jung, C., Chan, S.T., Gerber, J., Schmithorst, V.J., Edwards, R.R., Wasan, A.D., Berna, C., Kong, J., Kaptchuk, T.J., Gollub, R.L., Rosen, B.R., Napadow, V., 2019. Machine learning-based prediction of clinical pain using multimodal neuroimaging and autonomic metrics. Pain 160, 550–560. doi:10.1097/j.pain.0000000000001417.

Li, L., Huang, G., Lin, Q., Liu, J., Zhang, S., Zhang, Z.G., 2018. Magnitude and temporal variability of inter-stimulus EEG modulate the linear relationship between laser-evoked potentials and fast-pain perception. Front. Neurosci. 12, 1–9. doi:10.3389/fnins.2018.00340.

Lin, Q., Li, L., Liu, J., Liu, W., Huang, G., Zhang, Z.G., 2018. Influence of individual differences in fMRI-based pain prediction models on between-individual prediction performance. Front. Neurosci. 12, 1–12. doi:10.3389/fnins.2018.00569.

Lindquist, M.A., Krishnan, A., López-Solà, M., Jepma, M., Woo, C.W., Koban, L., Chang, L.J., Reynolds Losin, E.A., Eisenbarth, H., Ashar, Y.K., Delk, E., Wager, T.D., Roy, M., Atlas, L.Y., Reynolds Losin, E.A., Krishnan, A., Schmidt, L., Schmidt, L., 2017. Group-regularized individual prediction: theory and application to pain. Neuroimage 145, 274–287. doi:10.1016/j.neuroimage.2015.10.074.

Lötsch, J., Ultsch, A., 2018. Machine learning in pain research. Pain 159, 623–630. doi:10.1097/j.pain.0000000000001118.

Marquand, A., Howard, M., Brammer, M., Chu, C., Coen, S., Mourão-Miranda, J., 2010. Quantitative prediction of subjective pain intensity from whole-brain fMRI data using Gaussian processes. Neuroimage 49, 2178–2189. doi:10.1016/j.neuroimage.2009.10.072.

McIntosh, A.R., Lobaugh, N.J., 2004. Partial least squares analysis of neuroimaging data: applications and advances. Neuroimage 23, S250–S263. doi:10.1016/j.neuroimage.2004.07.020.

Mouraux, A., Diukova, A., Lee, M.C., Wise, R.G., Iannetti, G.D., 2011. A multisensory investigation of the functional significance of the “pain matrix”. Neuroimage 54, 2237–2249. doi:10.1016/j.neuroimage.2010.09.084.

Mumford, J.A., Turner, B.O., Gregory Ashby, F., Poldrack, R.A., 2012. Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. Neuroimage 59, 2636–2643. doi:10.1016/j.neuroimage.2011.08.076.

Mun, C.J., Suk, H.W., Davis, M.C., Karoly, P., Finan, P., Tennen, H., Jensen, M.P., 2019. Investigating intra-individual pain variability: methods, applications, issues, and directions. Pain 160, 2415–2429. doi:10.1097/j.pain.0000000000001626.

O’Connell, T.P., Chun, M.M., 2018. Predicting eye movement patterns from fMRI responses to natural scenes. Nat. Commun. 9 (1), 1–15. doi:10.1038/s41467-018-07471-9.

Penny, W.D., Friston, K.J., Ashburner, J.T., Kiebel, S.J., Nichols, T.E., 2007. Statistical Parametric mapping: the Analysis of Functional Brain Images, 1st ed. Elsevier, Academic Press, Boston.

Poldrack, R.A., Huckins, G., Varoquaux, G., 2020. Establishment of best practices for evidence for prediction: a review. JAMA Psychiatry 77 (5), 534–540. doi:10.1001/jamapsychiatry.2019.3671.

Prato, M., Favilla, S., Zanni, L., Porro, C.A., Baraldi, P., 2011. A regularization algorithm for decoding perceptual temporal profiles from fMRI data. Neuroimage 56, 258–267. doi:10.1016/j.neuroimage.2011.01.074.

Rissman, J., Gazzaley, A., D’Esposito, M., 2004. Measuring functional connectivity during distinct stages of a cognitive task. Neuroimage 23, 752–763. doi:10.1016/j.neuroimage.2004.06.035.

Sava, S.L., De Pasqua, V., Magis, D., Schoenen, J., 2014. Effects of visual cortex activation on the nociceptive blink reflex in healthy subjects. PLoS One 9. doi:10.1371/journal.pone.0100198.

Schnakers, C., Zasler, N.D., 2007. Pain assessment and management in disorders of consciousness. Curr. Opin. Neurol. 20, 620–626. doi:10.1097/WCO.0b013e3282f169d9.

Schulz, E., Zherdin, A., Tiemann, L., Plant, C., Ploner, M., 2012. Decoding an individual's sensitivity to pain from the multivariate analysis of EEG data. Cereb. Cortex 22, 1118–1123. doi:10.1093/cercor/bhr186.

Tracey, I., Mantyh, P.W., 2007. The cerebral signature for pain perception and its modulation. Neuron 55, 377–391. doi:10.1016/j.neuron.2007.07.012.

Tu, Y.H., Tan, A., Bai, Y., Hung, Y.S., Zhang, Z.G., 2016a. Decoding subjective intensity of nociceptive pain from pre-stimulus and post-stimulus brain activities. Front. Comput. Neurosci. 10, 32. doi:10.3389/fncom.2016.00032.

Tu, Y.H., Zhang, Z.G., Tan, A., Peng, W., Hung, Y.S., Moayedi, M., Iannetti, G.D., Hu, L., 2016b. Alpha and gamma oscillation amplitudes synergistically predict the perception of forthcoming nociceptive stimuli. Hum. Brain Mapp. 37, 501–514. doi:10.1002/hbm.23048.

Tu, Y.H., Fu, Z., Huang, G., Hu, L., Hung, Y.S., Zhang, Z.G., 2018. A novel and effective fMRI decoding approach based on sliced inverse regression and its application to pain prediction. Neurocomputing 273, 373–384. doi:10.1016/j.neucom.2017.07.045.

Tu, Y.H., Zeng, F., Lan, L., Li, Z.J., Maleki, N., Liu, B., Chen, J., Wang, C.C., Park, J., Lang, C., Gao, Y.J., Liu, M.L., Fu, Z.N., Zhang, Z.G., Liang, F.R., Kong, J., 2020. An fMRI-based neural marker for migraine without aura. Neurology 94, e741–e751. doi:10.1212/wnl.0000000000008962.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15, 273–289. doi:10.1006/nimg.2001.0978.

Vijayakumar, V., Case, M., Shirinpour, S., He, B., 2017. Quantifying and characterizing tonic thermal pain across subjects from EEG data using random forest models. IEEE Trans. Biomed. Eng. 64, 2988–2996. doi:10.1109/TBME.2017.2756870.

Wager, T.D., Atlas, L.Y., Lindquist, M.A., Roy, M., Woo, C.W., Kross, E., 2013. An fMRI-based neurologic signature of physical pain. N. Engl. J. Med. 368, 1388–1397. doi:10.1056/NEJMoa1204471.

Woo, C.W., Chang, L.J., Lindquist, M.A., Wager, T.D., 2017. Building better biomarkers: brain models in translational neuroimaging. Nat. Neurosci. 176, 139–148. doi:10.1038/nn.4478.

Zang, Y.F., He, Y., Zhu, C.Z., Cao, Q.J., Sui, M.Q., Liang, M., Tian, L.X., Jiang, T.Z., Wang, Y.F., 2007. Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI. Brain Dev. 29, 83–91. doi:10.1016/j.braindev.2006.07.002.

Zautra, A.J., Fasman, R., Davis, M.C., (Bud) Craig, A.D., 2010. The effects of slow breathing on affective responses to pain stimuli: an experimental study. Pain 149, 12–18. doi:10.1016/j.pain.2009.10.001.

Zhang, Z.G., Hu, L., Hung, Y.S., Mouraux, A., Iannetti, G.D., 2012. Gamma-band oscillations in the primary somatosensory cortex-A direct and obligatory correlate of subjective pain intensity. J. Neurosci. 32, 7429–7438. doi:10.1523/JNEUROSCI.5877-11.2012.