# Affective Relevance

Inferring Emotional Responses via fNIRS Neuroimaging

Tuukka Ruotsalo
University of Copenhagen and University of Helsinki
tr@di.ku.dk

Kalle Mäkelä
University of Helsinki
kalle.o.makela@helsinki.fi

Michiel M. Spapé
University of Helsinki
michiel.spape@helsinki.fi

Luis A. Leiva
University of Luxembourg
name.surname@uni.lu

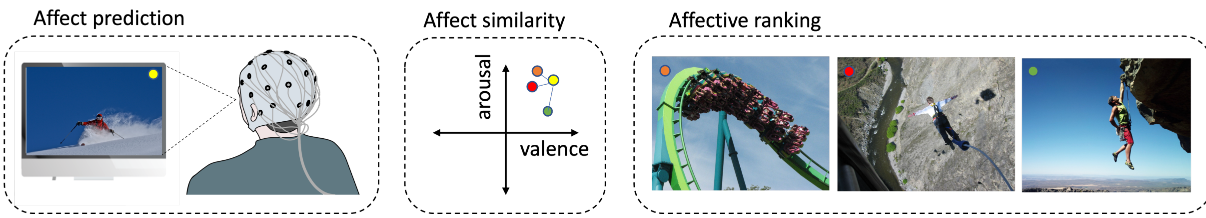


**Figure 1: Affective relevance accounts for subjective measurement of emotional responses. Here, a user is experiencing an image of skiing, which evokes a positive exciting emotion, or a high-arousal high-valence response. This can be measured via fNIRS neuroimaging, then predicted and mapped to the valence/arousal affective space. Images that are similar in the affective space, such as extreme sports and the roller coaster ride in this example, can be retrieved based on their affective similarity despite the fact that they have different visual and topical appearance.**

## ABSTRACT

Information retrieval (IR) relies on a general notion of relevance, which is used as the principal foundation for ranking and evaluation methods. However, IR does not account for more a nuanced affective experience. Here, we consider the emotional response decoded directly from the human brain as an alternative dimension of relevance. We report an experiment covering seven different scenarios in which we measure and predict how users emotionally respond to visual image contents by using functional near-infrared spectroscopy (fNIRS) neuroimaging on two commonly used affective dimensions: *valence* (negativity and positivity) and *arousal* (boredness and excitedness). Our results show that affective states can be successfully decoded using fNIRS, and utilized to complement the present notion of relevance in IR studies. For example, we achieved 0.39 Balanced accuracy and 0.61 AUC in 4-class classification of affective states (vs. 0.25 Balanced accuracy and 0.5 AUC of a random classifier). Likewise, we achieved 0.684 Precision@20 when retrieving high-arousal images. Our work opens new avenues for incorporating emotional states in IR evaluation, affective feedback, and information filtering.



*SIGIR '23, July 23–27, 2023, Taipei, Taiwan*

ACM ISBN 978-1-4503-9408-6/23/07.
https://doi.org/10.1145/3539618.3591946

## CCS CONCEPTS

• **Information systems** → **Users and interactive retrieval**.

## KEYWORDS

Affective computing; Emotion detection; Affective feedback

**ACM Reference Format:**
Tuukka Ruotsalo, Kalle Mäkelä, Michiel M. Spapé, and Luis A. Leiva. 2023. Affective Relevance: Inferring Emotional Responses via fNIRS Neuroimaging. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '23), July 23–27, 2023, Taipei, Taiwan.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3539618.3591946

## 1 INTRODUCTION

The present information retrieval (IR) paradigm relies on matching information with representations reflecting the user's perception of relevance, typically estimated from textual or visual features [47], click-through data [29], or other content-based and behavioral data [1, 10, 32, 48]. However, previous approaches do not account for what we call *affective relevance*. That is, the affective states of users when they experience content. For example, consider a user who receives content from a search engine, media feed, or recommender system. After examining the content, the user might feel positive, negative, excited, or bored about the content. However, the present signals that are used for predicting whether the content is relevant for a user would rely on dwell time or click-through rate, which are not predictive for such affective experiences of users.

The lack of affective features in IR is often referred to as the 'affective gap' [52]. It challenges models that rely only on content-based or behavioral features, as opposed to predictions on how users emotionally experience content. While researchers have proposed feature engineering methods that would account for affective content [34, 49–51], as well as deep learning to learn efficient representations [4, 13, 25, 41], the affective gap has been turned out to be challenging to address.

This observation has fundamental implications for user modeling and information retrieval: we build ranking models, user models, and user profiles by utilizing user signals, but our understanding of whether and how they reflect the users' emotional experiences toward content remain elusive. As a result, the models may predict user interest on information that users click or spend time on, while users may still experience the information negatively or find it appearing outrageous or even false.

Here, we propose a novel methodology using brain signals recorded with fNIRS to predict *arousal* (interesting or boring) and *valence* (positivity or negativity) of information [5, 42]. Arousal indicates a form of excitability and valence reflects the degree an emotion has a pleasant or unpleasant quality [23].

Our work complements previous studies [7, 8, 24] that focused on emotion decoding. For example, previous research has revealed associations between emotional experiences or affective states of users and behavioral and physiological correlates [3, 9, 30, 35–39, 43] and utilized physiological data for various interactive tasks [15, 18, 20, 21, 28, 31]. Nevertheless, the practical utilization of affective states, as mesured from neural activity related to how content is *emotionally or affectively experienced*, has not been previously utilized in IR methodologies.

The use of fNIRS mitigates reliance on any explicit user interaction as a probe for affect and can directly measure the affective processing as it occurs in the brain. Therefore, our methodology only requires users to perceive stimuli information, and their subjective mental processes can be continuously recorded. fNIRS relies on the measurement of blood oxygenation using the dissociable levels of light-absorption for oxygenated and deoxygenated hemoglobin [6, 11]. Thus, by measuring absorption at different wavelengths of light, fNIRS may quantify cortical activity. While fNIRS has previously been used to quantify mainly motor activity [26] and cognitive load [45], fNIRS has recently shown success in measuring emotion-related activity in frontal areas as well [8, 27]. Other apporaches such as electroencephalography (EEG) and peripheral wearable sensors have been extensively studied, but their use in realistic IR applications is still limited [16, 17].

We report initial results using data collected from participants viewing images according to two predictive tasks: affect prediction and affective ranking. For the former, we show that machine learning models can predict the affective class with a reasonably high accuracy in four-class and several two-class setups. For the latter, we show that rankers utilizing the predictions for affective similarity lead to a reasonably high precision rankings when retrieving affectively similar images. This result holds particularly for high-arousal images, that can be understood as high-attention grasping content. In summary, our contributions can be summarized as follows:

(1) We present a novel affective relevance estimation methodology from brain signals using fNIRS brain imaging.
(2) We show that affective states can be predicted from fNIRS in a single-trial scenario.
(3) We show that the predicted affective states can be succesfully used in ranking information according to their affective relevance.

## 2 NEUROIMAGING DATA

### 2.1 Participants, stimuli, and procedure

Data were recorded from thirty-one volunteers (18 female, 12 male, 1 non-binary) aged M=31.2 years (SD=7.4). Participants were fully informed about the nature of the experiment, and their rights as participants, including the right to withdraw at any time without fear of negative consequences. They signed informed consent prior to the experiment's commencement and were compensated with a voucher to the local cinema. The study was approved by Ethical Review Board of the University of Helsinki.

A sample of 120 images from the international affective pictures system (IAPS) [33] were used as stimuli. The images represent the entire range of emotional reactions associated with valence and arousal scores. Each participant was presented a random sample of 40 images from the range of high/low valence and high/low arousal. The experiment took about 45 minutes to complete, excluding participant preparation and device setup.

Images were displayed usinng the E-Prime 3 (Psychology Software Tools, Inc, Sharpsburg, PA) software. Synchronization was conducted via the DCOM interface and fNIRS was recorded using the OxySoft (Artinis Medical Systems, Elst, The Netherlands) software. Raw optical density (OD) data of fNIRS were recorded using the Artinis Brite-24 device in a configuration with 10 diodes transmitting light at two wavelengths (760 and 850 nm) and 8 photodiodes. The diodes were positioned on an elastic cap and placed such that the distance between receivers and optodes was ≈30 mm. Each receiver obtained light from three different transmitters, resulting in OD signals from 12 sources, which weredigitized at 50 Hz.

### 2.2 Data preprocessing

Raw OD data were exported from OxySoft and processed using MNE.[1] To determine artefactual channels, the scalp coupling index (SCI) [40] was calculated for each channel and channels with poor contact (SCI < 0.8) were replaced by interpolating from neighboring channels. Artefacts in the continuous signal were corrected using temporal derivative distribution repair [22]. The OD data were then converted to hemoglobin concentrations using the modified Beer-Lambert law [19] to derive oxy-hemoglobin (HbO) and deoxy-hemoglobin (HbR) levels. A 0.1 Hz low pass filter was applied to remove physiological noise, while a 0.01 Hz high pass filter was applied to remove signals that were unrelated to evoked activity. Then, the continuous data were time-locked to the onset of experimental stimuli and segmented into epochs of 17 s, including 5 s of pre-stimulus baseline activity, which was used to normalize each epoch. The 12-second post-stimulus period of each epoch was

[1] https://mne.tools/stable/auto_tutorials/preprocessing/70_fnirs_processing.html

divided into three 4-second windows, and the mean of each window was extracted for each channel to be used as input features in the prediction experiments. Finally, features from HbR channels were filtered out since HbO and HbR channel pairs are strongly dependent [14].

# 3 AFFECT PREDICTION

The goal of our affective prediction experiments was to use the features extracted from the fNIRS data as input to classify the presented stimuli under seven different scenarios, starting with four-class classification into high/low arousal and high/low valence. Predictions were also conducted for various two-class setups considering different data splits, as in many practical scenarios a downstream task would benefit from identification of valence (whether the information is positive or negative), arousal (whether the information draws attention or not), or some combination, such as detecting e.g. valence of high-arousal content (positivity or negativity of content that draws attention).

## 3.1 Experimental setup

Individual models were trained and validated for each participant with stratified 10-fold cross-validation. All features were normalized to zero mean and unit variance within each training set split. Two standard classifiers were used: Logistic Regression (LReg) and Shrinkage Linear Discriminant Analysis (SLDA). Hyperparameters were optimized using Optuna [2]. We report Balanced Accuracy and AUC ROC score as evaluation metrics.

## 3.2 Classification results

As reported in Table 1, Balanced Accuracy and AUC ROC values show classification performance varying from the worst-case scenario of high/low arousal classification of all images (Acc=0.553 and AUC=0.576 with LReg) to the best-case scenario of high/low valence classification of high-arousal images (Acc=0.654 and AUC=0.686 with SLDA). All results are statistically significant ($p < .05$) over a random classifier using permutation testing. The results suggest that classification of all images into either low/high arousal or low/high valence is a challenging task, whereas classification of different data splits (e.g. high-arousal images only) show higher classification performance.

# 4 AFFECT SIMILARITY AND RANKING

We now focus on retrieving similar images in terms of their affective content, for the same set of scenarios considered before. The data from the participants who achieved a classification accuracy below 50% in our previous experiments were excluded, motivated by the “BCI illiteracy” phenomenon, which states that brain-computer interfacing does not work for a non-negligible portion of users, estimated to be around 15–30% of the population [46]. This resulted in valid data for 29 participants on average.

## 4.1 Experimental setup

We focus on the query-by-example setting: Each image at a time is used as an input query $x$ to rank the rest of the images. An image is considered relevant if it belongs to the same class as the query image, and vice versa for non-relevant images.

**Table 1: fNIRS classification results using Logistic Regression (LReg) and Shrinkage Linear Discriminant Analysis (SLDA). Mean ± Std. Err.**

| Model | No. Classes | Balanced Accuracy | AUC ROC |
|---|---|---|---|
| **high/low arousal and high/low valence classification of all images** | | | |
| LReg | 4 | 0.372 ± 0.014 | 0.607 ± 0.011 |
| SLDA | 4 | 0.392 ± 0.015 | 0.610 ± 0.012 |
| **high/low valence classification of all images** | | | |
| LReg | 2 | 0.582 ± 0.016 | 0.598 ± 0.022 |
| SLDA | 2 | 0.582 ± 0.014 | 0.596 ± 0.021 |
| **high/low arousal classification of all images** | | | |
| LReg | 2 | 0.553 ± 0.015 | 0.576 ± 0.017 |
| SLDA | 2 | 0.555 ± 0.016 | 0.573 ± 0.017 |
| **high/low arousal classification of low-valence images** | | | |
| LReg | 2 | 0.627 ± 0.023 | 0.628 ± 0.029 |
| SLDA | 2 | 0.625 ± 0.022 | 0.615 ± 0.027 |
| **high/low arousal classification of high-valence images** | | | |
| LReg | 2 | 0.569 ± 0.025 | 0.579 ± 0.033 |
| SLDA | 2 | 0.575 ± 0.026 | 0.584 ± 0.034 |
| **high/low valence classification of low-arousal images** | | | |
| LReg | 2 | 0.563 ± 0.017 | 0.579 ± 0.022 |
| SLDA | 2 | 0.574 ± 0.020 | 0.569 ± 0.022 |
| **high/low valence classification of high-arousal images** | | | |
| LReg | 2 | 0.646 ± 0.019 | 0.692 ± 0.021 |
| SLDA | 2 | 0.654 ± 0.020 | 0.686 ± 0.022 |

We consider two distinct evaluation aspects: retrieval performance and ranking similarity. Retrieval performance, informed by Precision at different rank positions $K \in \{1, 5, 10, 20\}$, indicates whether the retrieved images evoke similar emotions as to a given query image. Ranking similarity, as informed by Rank-biased Overlap (RBO) at different rank positions $K$, indicates whether the provided rankings agree with the ground-truth IAPS rankings.

The ground-truth rankings are computed in a 2-dimensional vector space (arousal and valence scores) using arousal and valence scores from the original IAPS database [33] and Euclidean distance as dissimilarity metric. The fNIRS rankings are computed based on the softmax vectors provided by the classifier we trained in our previous experiments. It estimates arousal and valence scores for each image assessed by each user by deviating from the centroid that represents the predicted class. Then, the final ranking will comprise the closest images to these predicted arousal and valence values according to the Euclidean distance to the query image $x$. Finally, we also report ranking results with a random classifier, to provide an empirical lower bound for our experiments.

## 4.2 Ranking results

Table 2 summarizes the retrieval performance results. We can see that fNIRS outperforms the random baselines by a large margin, evidencing thus that human affective processing can be successfully incorporated in retrieval models. Differences between any of the fNIRS models and the random model are statistically significant ($p < .05$) in all scenarios. Differences between the two classifiers for fNIRS (LReg and SLDA) are not statistically significant. All significance comparisons are done with the Chi-square test of proportions, Bonferroni-Holm corrected for multiple comparisons.

**Table 2: Precision@K retrieval results: Mean ± Std. Err.**

| Model | Prec@1 | Prec@5 | Prec@10 | Prec@20 |
|---|---|---|---|---|
| **high/low arousal and high/low valence retrieval of all images** | | | | |
| fNIRS LReg | 0.450 ± 0.08 | 0.455 ± 0.08 | 0.455 ± 0.08 | 0.459 ± 0.08 |
| fNIRS SLDA | 0.525 ± 0.06 | 0.522 ± 0.06 | 0.520 ± 0.06 | 0.518 ± 0.06 |
| Random | 0.214 ± 0.05 | 0.229 ± 0.02 | 0.250 ± 0.01 | 0.247 ± 0.01 |
| **high/low valence retrieval of all images** | | | | |
| fNIRS LReg | 0.450 ± 0.08 | 0.455 ± 0.08 | 0.455 ± 0.08 | 0.459 ± 0.08 |
| fNIRS SLDA | 0.525 ± 0.06 | 0.522 ± 0.06 | 0.520 ± 0.06 | 0.518 ± 0.06 |
| Random | 0.214 ± 0.05 | 0.229 ± 0.02 | 0.250 ± 0.01 | 0.247 ± 0.01 |
| **high/low arousal retrieval of all images** | | | | |
| fNIRS LReg | 0.569 ± 0.02 | 0.568 ± 0.02 | 0.562 ± 0.02 | 0.561 ± 0.02 |
| fNIRS SLDA | 0.593 ± 0.02 | 0.592 ± 0.02 | 0.592 ± 0.02 | 0.592 ± 0.02 |
| Random | 0.458 ± 0.05 | 0.482 ± 0.02 | 0.488 ± 0.02 | 0.488 ± 0.01 |
| **high/low arousal retrieval of low-valence images** | | | | |
| fNIRS LReg | 0.588 ± 0.02 | 0.593 ± 0.02 | 0.594 ± 0.02 | 0.599 ± 0.02 |
| fNIRS SLDA | 0.649 ± 0.02 | 0.647 ± 0.02 | 0.648 ± 0.02 | 0.650 ± 0.02 |
| Random | 0.500 ± 0.07 | 0.470 ± 0.03 | 0.488 ± 0.02 | 0.509 ± 0.01 |
| **high/low arousal retrieval of high-valence images** | | | | |
| fNIRS LReg | 0.631 ± 0.02 | 0.628 ± 0.02 | 0.629 ± 0.02 | 0.628 ± 0.02 |
| fNIRS SLDA | 0.661 ± 0.02 | 0.659 ± 0.02 | 0.656 ± 0.02 | 0.655 ± 0.02 |
| Random | 0.417 ± 0.06 | 0.453 ± 0.03 | 0.455 ± 0.02 | 0.464 ± 0.01 |
| **high/low valence retrieval of low-arousal images** | | | | |
| fNIRS LReg | 0.624 ± 0.02 | 0.623 ± 0.02 | 0.621 ± 0.02 | 0.616 ± 0.02 |
| fNIRS SLDA | 0.617 ± 0.02 | 0.619 ± 0.02 | 0.620 ± 0.02 | 0.621 ± 0.02 |
| Random | 0.467 ± 0.06 | 0.490 ± 0.03 | 0.487 ± 0.02 | 0.473 ± 0.01 |
| **high/low valence retrieval of high-arousal images** | | | | |
| fNIRS LReg | 0.673 ± 0.02 | 0.671 ± 0.02 | 0.670 ± 0.02 | 0.671 ± 0.02 |
| fNIRS SLDA | 0.685 ± 0.02 | 0.684 ± 0.02 | 0.684 ± 0.02 | 0.684 ± 0.02 |
| Random | 0.433 ± 0.06 | 0.477 ± 0.03 | 0.480 ± 0.02 | 0.481 ± 0.01 |

**Table 3: Rank-biased overlap results: Mean ± Std. Err.**

| Model | RBO@1 | RBO@5 | RBO@10 | RBO@20 |
|---|---|---|---|---|
| **Predict high/low arousal and high/low valence, all images** | | | | |
| fNIRS LReg | 0.00 ± 0.00 | 0.07 ± 0.03 | 0.14 ± 0.04 | 0.22 ± 0.05 |
| fNIRS SLDA | 0.04 ± 0.02 | 0.08 ± 0.02 | 0.13 ± 0.03 | 0.23 ± 0.03 |
| Random | 0.00 ± 0.00 | 0.03 ± 0.01 | 0.04 ± 0.01 | 0.09 ± 0.01 |
| **Predict high/low valence for all images** | | | | |
| fNIRS LReg | 0.00 ± 0.00 | 0.02 ± 0.00 | 0.05 ± 0.00 | 0.11 ± 0.01 |
| fNIRS SLDA | 0.00 ± 0.00 | 0.02 ± 0.00 | 0.05 ± 0.00 | 0.11 ± 0.01 |
| Random | 0.00 ± 0.00 | 0.02 ± 0.01 | 0.05 ± 0.01 | 0.09 ± 0.01 |
| **Predict high/low arousal for all images** | | | | |
| fNIRS LReg | 0.00 ± 0.00 | 0.02 ± 0.00 | 0.04 ± 0.00 | 0.09 ± 0.01 |
| fNIRS SLDA | 0.01 ± 0.00 | 0.02 ± 0.00 | 0.04 ± 0.00 | 0.08 ± 0.01 |
| Random | 0.00 ± 0.00 | 0.02 ± 0.01 | 0.04 ± 0.01 | 0.09 ± 0.01 |
| **Predict high/low arousal for low-valence images** | | | | |
| fNIRS LReg | 0.03 ± 0.01 | 0.06 ± 0.01 | 0.12 ± 0.01 | 0.24 ± 0.01 |
| fNIRS SLDA | 0.02 ± 0.01 | 0.07 ± 0.01 | 0.13 ± 0.01 | 0.27 ± 0.01 |
| Random | 0.02 ± 0.02 | 0.05 ± 0.01 | 0.09 ± 0.01 | 0.18 ± 0.01 |
| **Predict high/low arousal for high-valence images** | | | | |
| fNIRS LReg | 0.02 ± 0.01 | 0.07 ± 0.01 | 0.13 ± 0.01 | 0.27 ± 0.01 |
| fNIRS SLDA | 0.02 ± 0.01 | 0.06 ± 0.01 | 0.12 ± 0.01 | 0.26 ± 0.01 |
| Random | 0.02 ± 0.02 | 0.05 ± 0.01 | 0.09 ± 0.01 | 0.17 ± 0.01 |
| **Predict high/low valence for low-arousal images** | | | | |
| fNIRS LReg | 0.02 ± 0.01 | 0.07 ± 0.01 | 0.13 ± 0.01 | 0.24 ± 0.01 |
| fNIRS SLDA | 0.04 ± 0.01 | 0.07 ± 0.01 | 0.12 ± 0.01 | 0.23 ± 0.01 |
| Random | 0.00 ± 0.00 | 0.02 ± 0.01 | 0.06 ± 0.01 | 0.15 ± 0.01 |
| **Predict high/low valence for high-arousal images** | | | | |
| fNIRS LReg | 0.03 ± 0.01 | 0.09 ± 0.01 | 0.15 ± 0.01 | 0.28 ± 0.01 |
| fNIRS SLDA | 0.04 ± 0.01 | 0.10 ± 0.01 | 0.16 ± 0.01 | 0.28 ± 0.01 |
| Random | 0.03 ± 0.02 | 0.05 ± 0.01 | 0.08 ± 0.01 | 0.17 ± 0.01 |

Similarly, Table 3 summarizes the ranking similarity results. We can see that fNIRS rankings overlap more with ground-truth IAPS rankings as compared with the rankings provided by a random model. Both fNIRS models increased their RBO scores with increasing values of rank position $K$, as expected. Meanwhile, the random model increased way more slowly with increasing values of $K$. As per the Chi-square test of proportions, differences between any of the fNIRS models and the random model are statistically significant in all scenarios ($p < .05$) except RBO@1, where only the first scenario (high/low arousal and high/low valence, all images) was statistically significant, and RBO@5 for the second and third scenarios (high/low valence and arousal, all images). Differences between both fNIRS models are not statistically significant.All significance comparisons are Bonferroni-Holm corrected.

## 5 DISCUSSION AND CONCLUSIONS

We have reported the first study where affective states were incorporated to the notion of relevance and directly decoded from human brain activity. We also incorporated affective information to ranking models for image search. The classification results from single-trial decoding experiments show that affective states can be decoded with reasonably high accuracy using fNIRS. The results suggest that, while affective classification is overall a challenging task, classification of different data splits, such as only high-arousal images (that are generally relevant and attention grasping) and low-arousal images (that is known to evoke diminutive emotional responses), show higher classification performance. This is in-line with previous findings [12] but our results show higher accuracy than previously reported studies [44]. This finding is particularly encouraging for IR research, as search engines are already good at detecting topical matches, but less good at predicting which results are likely to be associated with positive or negative experiences.

Our results rely on visual stimuli from a standard dataset that is accepted and widely used by the scientific community for studying emotional reactions of human subjects [33]. Despite this, there is a possibility of bias related to the stimuli that cannot be fully ignored, such that ground-truth values were collected from a different participant population and that images were produced 15 years ago.

While our experiments are limited to a single data collection, our results demonstrate that additional signals of human cognitive and affective processing can be decoded via brain-computer interfacing, and they allow incorporating a new affective dimension of relevance in IR models. All in all, our work shows evidence that emotions and affective dimensions of how information is experienced by users may play a crucial role in future IR research and practice.

## ACKNOWLEDGMENTS

This work is supported by the Academy of Finland (grants 352915, 350323, 336085, 322653), the Horizon 2020 FET program of the European Union (BANANA, grant CHIST-ERA-20-BCI-001), and the EIC Pathfinder program (SYMBIOTIK, grant 101071147).

## REFERENCES


[1] Eugene Agichtein, Eric Brill, and Susan Dumais. 2006. Improving Web Search Ranking by Incorporating User Behavior Information. In *Proc. SIGIR.*

[2] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. 2019. Optuna: A next-generation hyperparameter optimization framework. In *Proc. KDD.*

[3] Ioannis Arapakis, Joemon M Jose, and Philip D Gray. 2008. Affective feedback: an investigation into the role of emotions in the information seeking process. In *Proc. SIGIR.*

[4] Ioannis Arapakis and Luis A. Leiva. 2020. Learning Efficient Representations of Mouse Movements to Predict User Attention. In *Proc. SIGIR.*

[5] Pedro Avero and Manuel G Calvo. 2006. Affective priming with pictures of emotional scenes: The role of perceptual similarity and category relatedness. *Span. J. Psychol.* 9, 1 (2006).

[6] Hasan Ayaz, Meltem Izzetoglu, Kurtulus Izzetoglu, and Banu Onaral. 2019. The use of functional near-infrared spectroscopy in neuroergonomics. In *Neuroergonomics*. Elsevier.

[7] Michela Balconi, Elisabetta Grippa, and Maria Elide Vanutelli. 2015. Resting lateralized activity predicts the cortical response and appraisal of emotions: an fNIRS study. *Soc. Cogn. Affect. Neurosci.* 10, 12 (2015).

[8] Michela Balconi, Elisabetta Grippa, and Maria Elide Vanutelli. 2015. What hemodynamic (fNIRS), electrophysiological (EEG) and autonomic integrated measures can tell us about emotional processing. *Brain Cogn.* 95 (2015).

[9] Oswald Barral, Manuel JA Eugster, Tuukka Ruotsalo, Michiel M Spapé, Ilkka Kosunen, Niklas Ravaja, Samuel Kaski, and Giulio Jacucci. 2015. Exploring peripheral physiology as a predictor of perceived relevance in information retrieval. In *Proc. IUI.*

[10] Lukas Brückner, Ioannis Arapakis, and Luis A. Leiva. 2021. When Choice Happens: A Systematic Examination of Mouse Movement Length for Decision Making in Web Search. In *Proc. SIGIR.*

[11] Scott C Bunce, Meltem Izzetoglu, Kurtulus Izzetoglu, Banu Onaral, and Kambiz Pourrezaei. 2006. Functional near-infrared spectroscopy. *IEEE Eng. Med. Biol. Mag.* 25, 4 (2006).

[12] Luis Carretié. 2014. Exogenous (automatic) attention to emotional stimuli: a review. *Cogn. Affect. Behav. Neurosci.* 14, 4 (2014).

[13] Ming Chen, Lu Zhang, and Jan P Allebach. 2015. Learning deep features for image emotion classification. In *Proc. ICIP.*

[14] Xu Cui, Signe Bray, and Allan L Reiss. 2010. Functional near infrared spectroscopy (NIRS) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics. *Neuroimage* 49, 4 (2010), 3039–3046.

[15] Keith M Davis, Carlos de la Torre-Ortiz, and Tuukka Ruotsalo. 2022. Brain-Supervised Image Editing. In *Proc. CVPR.*

[16] Keith M. Davis, Michiel Spapé, and Tuukka Ruotsalo. 2022. Contradicted by the Brain: Predicting Individual and Group Preferences via Brain-Computer Interfacing. *IEEE Trans. Affect. Comput.* (2022).

[17] Keith M. Davis III, Michiel Spapé, and Tuukka Ruotsalo. 2021. Collaborative Filtering with Preferences Inferred from Brain Signals. In *Proc. WWW.*

[18] Carlos de la Torre-Ortiz, Michiel M Spapé, Lauri Kangassalo, and Tuukka Ruotsalo. 2020. Brain relevance feedback for interactive image generation. In *Proc. UIST.*

[19] David T Delpy, Mark Cope, Pieter van der Zee, Simon Arridge, Susan Wray, and JS Wyatt. 1988. Estimation of optical pathlength through tissue from direct time of flight measurement. *Phys. Med. Biol.* 33, 12 (1988).

[20] Manuel JA Eugster, Tuukka Ruotsalo, Michiel M Spapé, Oswald Barral, Niklas Ravaja, Giulio Jacucci, and Samuel Kaski. 2016. Natural brain-information interfaces: Recommending information by relevance inferred from human brain signals. *Sci. Rep.* 6, 1 (2016).

[21] Manuel J.A. Eugster, Tuukka Ruotsalo, Michiel M. Spapé, Ilkka Kosunen, Oswald Barral, Niklas Ravaja, Giulio Jacucci, and Samuel Kaski. 2014. Predicting Term-Relevance from Brain Signals. In *Proc. SIGIR.*

[22] Frank A Fishburn, Ruth S Ludlum, Chandan J Vaidya, and Andrei V Medvedev. 2019. Temporal derivative distribution repair (TDDR): a motion correction method for fNIRS. *Neuroimage* 184 (2019).

[23] Don C Fowles. 1980. The three arousal model: Implications of Gray's two-factor learning theory for heart rate, electrodermal activity, and psychopathy. *Int. J. Psychophysiol.* 17, 2 (1980), 87–104.

[24] E Grippa, Maria Elide Vanutelli, I Venturella, E Molteni, and Michela Balconi. 2014. Hemodynamic responses (fNIRS) and EEG modulation of prefrontal cortex during emotion processing. In *Proc. SIPF.*

[25] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *Proc. CVPR.*

[26] Fabian Herold, Patrick Wiegel, Felix Scholkmann, and Notger G Müller. 2018. Applications of functional near-infrared spectroscopy (fNIRS) neuroimaging in exercise–cognition science: a systematic, methodology-focused review. *J. Clin. Med.* 7, 12 (2018).

[27] Xin Hu, Chu Zhuang, Fei Wang, Yong-Jin Liu, Chang-Hwan Im, and Dan Zhang. 2019. fNIRS evidence for recognizably different positive emotions. *Front. Hum. Neurosci.* 13 (2019).

[28] Giulio Jacucci, Oswald Barral, Pedram Daee, Markus Wenzel, Baris Serim, Tuukka Ruotsalo, Patrik Pluchino, Jonathan Freeman, Luciano Gamberini, Samuel Kaski, et al. 2019. Integrating neurophysiologic relevance feedback in intent modeling for information retrieval. *J. Assoc. Inf. Sci. Technol.* 70, 9 (2019).

[29] Thorsten Joachims. 2002. Optimizing search engines using clickthrough data. In *Proc. KDD.*

[30] Lauri Kangassalo, Michiel Spapé, Giulio Jacucci, and Tuukka Ruotsalo. 2019. Why do users issue good queries? Neural correlates of term specificity. In *Proc. SIGIR.*

[31] Lauri Kangassalo, Michiel Spapé, and Tuukka Ruotsalo. 2020. Neuroadaptive modelling for generating images matching perceptual categories. *Sci. Rep.* 10, 1 (2020).

[32] Aditya Khosla, Atish Das Sarma, and Raffay Hamid. 2014. What makes an image popular?. In *Proc. WWW.*

[33] Peter J. Lang, Margaret M. Bradley, and Bruce N. Cuthbert. 2008. International Affective Picture System (IAPS): Affective ratings of pictures and instruction manual.

[34] Jana Machajdik and Allan Hanbury. 2010. Affective image classification using features inspired by psychology and art theory. In *Proc. MM.*

[35] Daniel McDuff, Paul Thomas, Nick Craswell, Kael Rowan, and Mary Czerwinski. 2021. Do Affective Cues Validate Behavioural Metrics for Search?. In *Proc. SIGIR.*

[36] Dominika Michalkova, Mario Parra-Rodriguez, and Yashar Moshfeghi. 2022. Information Need Awareness: an EEG study. In *Proc. SIGIR.*

[37] Yashar Moshfeghi and Joemon M Jose. 2013. An effective implicit relevance feedback technique using affective, physiological and behavioural features. In *Proc. SIGIR.*

[38] Sakrapee Paisalnan, Yashar Moshfeghi, and Frank Pollick. 2021. Neural Correlates of Realisation of Satisfaction in a Successful Search Process. *Proc. Assoc. Inf. Sci.* 58, 1 (2021).

[39] Zuzana Pinkosova, William J. McGeown, and Yashar Moshfeghi. 2020. The Cortical Activity of Graded Relevance. In *Proc. SIGIR.*

[40] Luca Pollonini, Cristen Olds, Homer Abaya, Heather Bortfeld, Michael S Beauchamp, and John S Oghalai. 2014. Auditory cortex activation to natural speech and simulated cochlear implant speech measured with functional near-infrared spectroscopy. *Hear. Res.* 309 (2014).

[41] Tianrong Rao, Xiaoxu Li, and Min Xu. 2020. Learning multi-level deep representations for image emotion classification. *Neural Process. Lett.* 51, 3 (2020).

[42] James A Russell. 1980. A circumplex model of affect. *J. Pers. Soc. Psychol.* 39, 6 (1980).

[43] Stefanie Schmidt and Wolfgang G Stock. 2009. Collective indexing of emotions in images. A study in emotional information retrieval. *J. Assoc. Inf. Sci. Technol.* 60, 5 (2009).

[44] Lucas R Trambaiolli, Juliana Tossato, André M Cravo, Claudinei E Biazoli Jr, and João R Sato. 2021. Subject-independent decoding of affective states using functional near-infrared spectroscopy. *Plos one* 16, 1 (2021), e0244840.

[45] Anirudh Unni, Klas Ihme, Henrik Surm, Lars Weber, Andreas Lüdtke, Daniela Nicklas, Meike Jipp, and Jochem W Rieger. 2015. Brain activity measured with fNIRS for the prediction of cognitive workload. In *Proc. CogInfoCom*. IEEE.

[46] Carmen Vidaurre and Benjamin Blankertz. 2010. Towards a Cure for BCI Illiteracy. *Brain Topogr.* 23, 2 (2010).

[47] Ji Wan, Dayong Wang, Steven Chu Hong Hoi, Pengcheng Wu, Jianke Zhu, Yongdong Zhang, and Jintao Li. 2014. Deep learning for content-based image retrieval: A comprehensive study. In *Proc. MM.*

[48] Jiang Wang, Yang Song, Thomas Leung, Chuck Rosenberg, Jingbin Wang, James Philbin, Bo Chen, and Ying Wu. 2014. Learning fine-grained image similarity with deep ranking. In *Proc. CVPR.*

[49] Victoria Yanulevskaya, Jan C van Gemert, Katharina Roth, Ann-Katrin Herbold, Nicu Sebe, and Jan-Mark Geusebroek. 2008. Emotional valence categorization using holistic image features. In *Proc. ICIP.*

[50] Sicheng Zhao, Guiguang Ding, Qingming Huang, Tat-Seng Chua, Björn W Schuller, and Kurt Keutzer. 2018. Affective Image Content Analysis: A Comprehensive Survey. In *Proc. IJCAI.*

[51] Sicheng Zhao, Yue Gao, Xiaolei Jiang, Hongxun Yao, Tat-Seng Chua, and Xiaoshuai Sun. 2014. Exploring principles-of-art features for image emotion recognition. In *Proc. MM.*

[52] Sicheng Zhao, Xingxu Yao, Jufeng Yang, Guoli Jia, Guiguang Ding, Tat-Seng Chua, Bjoern W Schuller, and Kurt Keutzer. 2021. Affective image content analysis: Two decades review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intellig.* (2021).