Efficient Decoding of Affective States from Video-elicited EEG Signals: An Empirical Investigation

KAYHAN LATIFZADEH* and NIMA GOZALPOUR*, University of Luxembourg, Luxembourg
V. JAVIER TRAVER, INIT, Universitat Jaume I, Spain
TUUKKA RUOTSALO, University of Copenhagen, Denmark and LUT University, Finland
ALEKSANDRA KAWALA-STERNIUK, Opole University of Technology, Poland
LUIS A. LEIVA, University of Luxembourg, Luxembourg

Affect decoding through brain-computer interfacing (BCI) holds great potential to capture users' feelings and emotional responses via non-invasive electroencephalogram (EEG) sensing. Yet, little research has been conducted to understand *efficient* decoding when users are exposed to *dynamic* audiovisual contents. In this regard, we study EEG-based affect decoding from videos in arousal and valence classification tasks, considering the impact of signal length, window size for feature extraction, and frequency bands. We train both classic Machine Learning models (SVMs and k-NNs) and modern Deep Learning models (FCNNs and GTNs). Our results show that: (1) affect can be effectively decoded using less than 1 minute of EEG signal; (2) temporal windows of 6 and 10 seconds provide the best classification performance for classic Machine Learning models but Deep Learning models benefit from much shorter windows of 2 seconds; and (3) any model trained on the Beta band alone achieves similar (sometimes better) performance than when trained on all frequency bands. Taken together, our results indicate that affect decoding can work in more realistic conditions than currently assumed, thus becoming a viable technology for creating better interfaces and user models.

CCS Concepts: • **Human-centered computing** → **User models**; *Ambient intelligence*; • **Computing methodologies** → *Machine learning approaches*.

Additional Key Words and Phrases: BCI; EEG; Neurophysiology; Affect; Emotions; Videos

ACM Reference Format:
Kayhan Latifzadeh, Nima Gozalpour, V. Javier Traver, Tuukka Ruotsalo, Aleksandra Kawala-Sterniuk, and Luis A. Leiva. 2024. Efficient Decoding of Affective States from Video-elicited EEG Signals: An Empirical Investigation. *ACM Trans. Multimedia Comput. Commun. Appl.* 1, 1, Article 1 (January 2024), 24 pages. https://doi.org/10.1145/3663669

1 INTRODUCTION

Human affect is guided by biologically-based action dispositions that play an important role in the determination of behavior [50]. For example, it is no secret that mood and emotions influence our daily lives, impacting human interactions [3, 13], decision-making [58], world perception [47], and physical well-being [36]. As researchers have increasingly focused on understanding affective reactions toward computing and digital information, the necessity to compute and decode affect has

*Equal contribution.

Authors' addresses: Kayhan Latifzadeh; Nima Gozalpour, University of Luxembourg, Luxembourg; V. Javier Traver, INIT, Universitat Jaume I, Spain; Tuukka Ruotsalo, University of Copenhagen, Denmark , LUT University, Finland; Aleksandra Kawala-Sterniuk, Opole University of Technology, Poland; Luis A. Leiva, University of Luxembourg, Luxembourg.

1551-6857/2024/1-ART1
https://doi.org/10.1145/3663669

also increased [47, 58]. Various measures, including subjective self-reports and neurophysiological measurements, have been applied for these purposes [3, 26, 37, 58].

However, our understanding of how to detect and decode affect in computing applications has only been explored recently, cf. [85]. Neural mechanisms of emotional functioning and Machine Learning models have been coupled to decode affective responses from the brain in computer-mediated environments [81]. In particular, brain-computer interfacing (BCI) and physiological sensing have been developed to detect affective states [80], which has been used to model the user's context [94] and adapt user interfaces [5], among other tasks.

Among the various BCI technologies available for affect decoding, electroencephalography (EEG) has gained a significant interest among researchers due to its simplicity, affordability, portability, and user-friendly nature [3, 47]. In addition, the number of publicly available EEG datasets has fostered many research advancements in this regard [19, 41, 44].

1.1 Problem statement

The main focus of previous studies has been on approaches for accurate affect decoding, including determining subject-specific, cross-subject, and temporally stable features [70, 118]. Another line of research has focused on feature engineering [103] and the development of Machine Learning models for affect decoding [4]. Consequently, previous research has focused on the *effectiveness* of affect decoding performance. Less effort, however, has been devoted to the *efficiency* of these approaches, i.e., to understand how dynamic and temporally varying stimuli could be exploited for affect decoding in a way that reflects more closely to real-world user behavior.

In order to make affect decoding efficient, we study how users react to video contents. This is a challenging but relevant research environment, since users watching videos typically skip parts or watch them partially, to only concentrate on parts that they find interesting. Furthermore, we rarely can afford to use lengthy recordings that have been assumed in idealistic laboratory settings, but rather only shorter sequences are available in realistic human-computer interaction (HCI) scenarios. Note that this is contrary to using maximal amount of data to optimize model performance, as done by all state-of-the-art decoding and classification approaches, which rely on features computed on the full-length EEG signal [93, 119].

1.2 Contributions

We study efficient EEG-based affect decoding from videos in arousal and valence classification tasks, considering the impact of: EEG signal length, window size for feature extraction, and EEG frequency bands. These three aspects are fundamental to ensure an effective decoding performance, as explained next.

In the same way as visual importance evolves dynamically over time [99], so do our affective experiences in both physical and digital environments [15]. A common and well-justified assumption is that user's affect do not change during exposure to short-term stimuli. However, this may not hold for longer and heterogeneous content, such as videos, where our affective reactions may be represented only in some specific temporal window and thus may emerge at a certain point in time. Therefore, by studying the impact of EEG signal length we can better understand the extent to which the exposure to dynamic contents may influence affect decoding.

We also systematically study the impact of sampling window size for feature extraction. In this regard, the community has not found a consensus since some work suggests that smaller window sizes [71, 112] are preferred over longer window sizes [49, 95]. Therefore, by studying the impact of window size we can better understand the amount of EEG samples that are required to achieve optimal decoding performance.

Finally, we also systematically study the impact of EEG frequency bands, which is particularly relevant because previous work has argued that higher-frequency bands generally carry the most discriminative information affect-wise [56, 109, 114], yet most of previous work has considered all frequency bands combined for affect decoding [117, 119]. If we could rely on just only one frequency band, for example, decoding would not only be simpler to achieve but also computationally more efficient.

In sum, our work addresses important knowledge gaps in the research literature by performing a systematic evaluation of key design decisions for affect decoding during dynamic contents perception. To this end, we pose the following research questions:

RQ1: *What is the optimal EEG signal length for affect decoding during dynamic content perception?*
RQ2: *What is the optimal temporal window for affect decoding during dynamic content perception*?
RQ3: *Which frequency bands (or their combinations) are associated with affective states during dynamic content perception?*

Our findings can be summarized as follows:

- Using short EEG signal samples (e.g. 30 seconds) achieves decoding performance comparable to using longer signals (e.g. 1 min).
- Temporal windows of 6 or 10 seconds for feature extraction are the ones that provide the best performance results for classic Machine Learning models but modern Deep Learning models benefit from much shorter windows of 2 seconds.
- Effective affect decoding can be achieved using a combined set of frequency bands, but a model trained with the Beta frequency band alone shows performance comparable to using all bands combined. Results ranged from 65% (valence classification in DEAP) to 88% (arousal classification in MAHNOB-HCI).

Taken together, our findings suggest that affect decoding can work in more realistic conditions than currently assumed, thus becoming a viable technology for creating better interfaces and user models. These findings may allow researchers and practitioners to build affective recognition models that are data efficient, faster to train, and less prone to signal noise.

2 RELATED WORK

The HCI community has increasingly studied automated methods to measure affective responses [18, 79], e.g., for making computing systems more natural toward humans' affective experiences [74]. The fundamental enabler for this direction of research is affect decoding, as this is the basis for several downstream tasks, user models, and adaptive systems [60]. For example, Menezes et al. [70] used EEG signals to model users' affective states based on Russell's Circumplex Model [84]. They used classic Machine Learning models, namely Support Vector Machine and Random Forest. Cimtay and Ekmekcioglu [19] aimed to improve subject-independent recognition accuracy by using pre-trained Convolutional Neural Networks (CNNs). Their approach used raw EEG signals with windowing, pre-adjustments, and feature normalization. They tested three data-sets: SEED [23], DEAP [44], and LUMED [19]. The proposed method achieved mean accuracy ranging from 78% to 81% across those datasets.

2.1 Dimensional theory of emotions

Affect can be understood through different frameworks, including "universal" discrete emotional states such as joy or fear [54], appraisal models that integrate the situational context of an event [27], and dimensional models [25], which position emotions on several continuous dimensions that can be analytically modeled. Due to their properties, dimensional models are particularly useful for understanding and decoding affect in human-computer interaction [63].

The dimensionality of human affect is most commonly represented on a two-dimensional valence/arousal plane [10], see Figure 1. Arousal (or intensity) is the level of autonomic activation that an event creates, and ranges from calm (or low) to excited (or high). Valence, on the other hand, is the level of pleasantness that an event generates and is defined along a continuum from negative (or low) to positive (or high). According to previous work, all human emotions can be located in this two-dimensional plane [50, 51]. For example, happiness has a high arousal and positive valence, fear has a high arousal and negative valence, and sadness has a low arousal and negative valence. Therefore, in this paper we will focus on valence/arousal classification as a proxy for affect decoding.

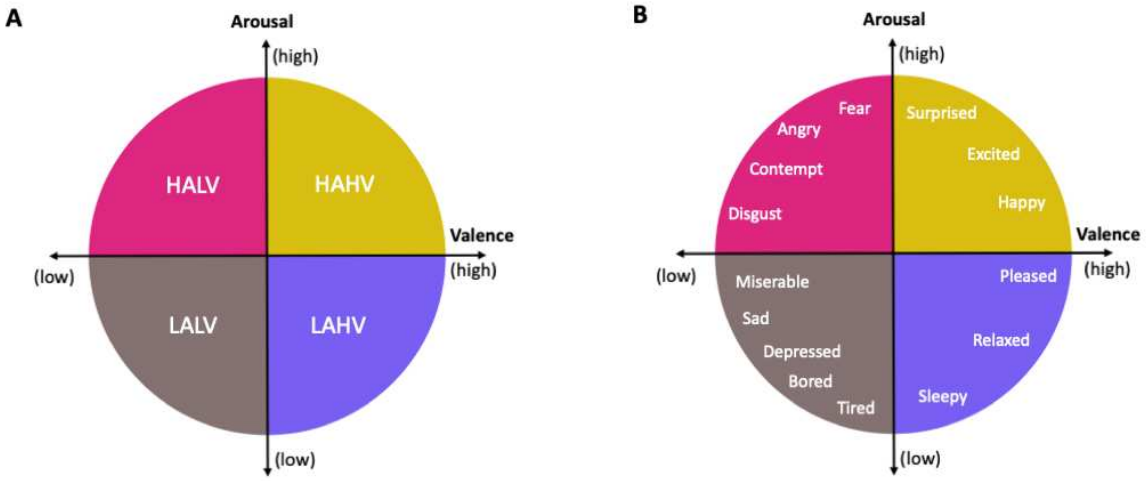

Fig. 1. Affect (right) and emotion (left) models according to the valence and arousal dimensions. **A**: Affect can be categorized as: high arousal and high valence (HAHV), high arousal and low valence (HALV), low arousal and high valence (LAHV), and low arousal and low valence (LALV). Valence is also considered as negative (low) or positive (high). **B**: Each of the four affect categories includes a subset of discrete emotional states.

2.2 Affect decoding from EEG

Affect can be reliably decoded from measurable neurophysiological activity, through BCI or other physiological sensing devices [58, 70, 90, 111]. Physiological signals are often preferred over behavioral signals, as they occur implicitly and are not affected by concealed feelings, as may be possible, for example, in the case of users intentionally avoiding or mocking facial expressions or other behavioral traits.

The large and increasing body of research on affect decoding from EEG signals, especially concerning emotion recognition, is accounted for in recent surveys [1, 3, 98, 101]. This increase is because electroencephalography offers several advantages as affect decoding technology, particularly its resistance to deceptive actions compared to visual or speech signals [19]. However, the non-stationary nature of EEG recordings poses a challenge for subject-independent decoding [19, 41]. Since a thorough review is not possible, nor is the goal of this work, a few relevant works are discussed next.

Regarding the *stimuli used to elicit affective responses*, static visual contents (images) are the most common ones [1], although recently researchers have started to consider dynamic, audiovisual contents such as videos [29, 68, 76, 116]. Interestingly, affective responses elicited from dynamic and interactive contents have been increasingly considered in broader contexts such as gaming [2,

22, 45, 48, 97, 120] and education [8, 108, 120]. Less common are affective responses elicited while performing cognitively demanding activities such as stock trading [96].

As for the *affective measures* considered, even though they are mapped to continuous values of arousal and valence, most research maps them into binary values (i.e., low vs. high) [49]. Predicting precise values of arousal and valence remains to be a challenging task at present [30]. Further, depending on the application or research purposes, affect decoding is often considered binary (e.g., like/dislike [29]). We can find previous work where affective responses were classified into three classes (positive, neutral, and negative) [77, 117], whereas previous research on emotion recognition has considered four (happy, sad, fear and relaxed) [68], or even eight [62] classes.

Regarding *decoding techniques*, many works rely on a variety of traditional Machine Learning algorithms. For example, Mehmood et al. [69] used a Random Forest classifier with Hjorth features (see Section 3.3). Deep Learning approaches have recently been explored as well [17, 96, 113, 117]. Importantly, the use of Deep Learning has its own challenges since training these models requires a large amount of data, and unfortunately collecting EEG signals is very time-consuming and costly [52]. To cope with these issues, previous work has suggested a myriad of data augmentation techniques, with moderate results [16, 38, 52, 65, 66, 87, 104, 115, 117]; for example, performance gains may only be marginal or not steady. Interesting recent work used self-supervision techniques [59, 115], and contrastive learning for cross-subject emotion recognition [89]. For example, Gilakjani and Al Osman [31] proposed a Contrastive Learning GAN-based Graph Neural Network that increased the number of trials for emotion recognition from EEG signals. They benchmarked the model on DEAP and MAHNOB-HCI datasets, achieving a binary classification accuracy ranging between 64.84% (valence, DEAP) and 71.69% (arousal, MAHNOB-HCI). They considered full-length EEG signals and did not explore the role of frequency bands or temporal windows. Li et al. [55] proposed a neural architecture search (NAS) framework for binary classification of valence and arousal based on Reinforcement Learning. They achieved around 98% binary classification accuracy for valence and arousal in DEAP and DREAMER [40] datasets. However they adopted a subject-dependent approach and considered the full-length EEG signals.

Most of previous work on EEG-based affect decoding have been conducted using publicly available datasets. These allow replication and reproducibility of research results, comparing e.g., feature engineering and classification approaches [6, 49, 77]. Researchers have reported classification accuracy ranging from 65% to 80% on the SEED and DEAP datasets, depending on the selected target variables and the feature selection procedures [62, 68]. All in all, previous work mark a trend toward different feature engineering approaches for affect decoding, mainly focused on models that can predict valence and arousal values. Yet, much of this work has aimed at maximizing classification performance rather than investigating model efficiency, which is crucial for their use in HCI problems and other downstream applications.

2.3 Extracting affect-related markers from EEG

EEG signals are highly complex and multi-dimensional [39], since not only they vary temporally according to a given stimulus, but they also include the response from different electrodes (the spatial channels). They involve non-linear brain dynamics [86] and their frequency analysis is also relevant since it has been found that different frequency bands carry discriminative information for specific tasks [39]. Furthermore, EEG datasets are typically available from a small set of participants from which brain signals were recorded during one or several experimental sessions. This makes it challenging to train competent Machine Learning models, as modern models such as those based on neural networks usually require large amounts of data.

In this context, it is natural to investigate which are the best and most efficient procedures to extract valuable information from EEG signals for affect decoding, and how the existing approaches

might be influenced by relevant factors. From a hardware perspective, there is a concern about how effective low-cost devices can be compared to clinical-grade ones [46, 78]. To cope with the inherent difficulty of capturing and processing brain signals, multimodal hybrid systems, which include additional signals, typically from eye tracking, have been proposed [46, 64, 105, 116]. However, they introduce non-trivial challenges such as multi-device and multi-computer synchronization [53].

Most of the existing body of work uses specific frequency bands that are widely agreed to work for specific tasks [39], which represents a limitation, since other potentially useful bands are ignored by default. This can be addressed by Machine Learning approaches so that the relevant bands are automatically attended [43, 117], although this combinatorial problem is not the focus of our work. Davis et al. [20] proposed crowdsourcing to combine affect-related annotations from multiple participants. Although interesting, they focused on using full-length signals from all participants, and no temporal analysis was performed. Overall, intra- and inter-subject variability is an important issue that requires specific mechanisms to increase transferability [85] so as to improve the applicability in EEG and related BCI technology.

Regarding the time domain, it has been found that EEG signals are stable among different sessions [118], which can be seen as a form of long-term analysis. However, the short-term analysis within a session for a single participant has not been sufficiently investigated. For Alzheimer's disease detection, for example, increasingly better performance has been found with longer signal lengths [100]. For motor imagery classification, longer temporal windows [12] and multiple frequency bands [91] have been found preferable. A clustering approach has been proposed for identifying time windows for event-related potentials (ERPs) of interest [67].

Overall, determining the optimal window size for sampling has been considered a controversial topic. We can find studies using widely different window sizes, ranging from 1 second to 30 seconds [56]. In most of these works, the time window refers to slices or chunks of the EEG signal that are used to extract features. Therefore, given a full-length signal, shorter time windows imply more chunks, more features to extract, and longer feature vectors to classify. Arguably, feature extraction is extremely consuming with very short temporal windows (2 seconds or less) [75], so longer windows should be preferred for practical applications.

Closer to our goals, previous work has found that the last half of the EEG signals recorded while watching videos turns out to be more discriminative affect-wise than other parts of the signal [49]. This is useful to inform affective computing systems. However, not only does this approach requires the lengthy EEG signals being recorded, but the brain responses at a given time are affected by previous exposure to a stimulus [33]. Therefore, to avoid uncontrolled carryover effects, our focus is on varying lengths starting from the onset of the stimulus. Ultimately, our findings may have important practical implications for the design of real-time BCI-based applications.

3 METHODOLOGY

We report experiments on two public datasets that recorded EEG data in response to emotional videos, and compare affect classification using different Machine Learning and Deep Learning models trained with different signal lengths, window sizes, and frequency bands.

3.1 Datasets

The first dataset we used is DEAP [44], which is one of the most well-known publicly available datasets in affective computing [6, 17, 29, 87, 106, 110]. The dataset consists of 1280 EEG recordings from 32 participants who watched 40 music videos of 60 s duration each. Participants reported their perceived affective responses in terms of arousal and valence in the $[1, 9]$ range after watching each video. The sample distribution over the valence/arousal plane is illustrated in Figure 2.

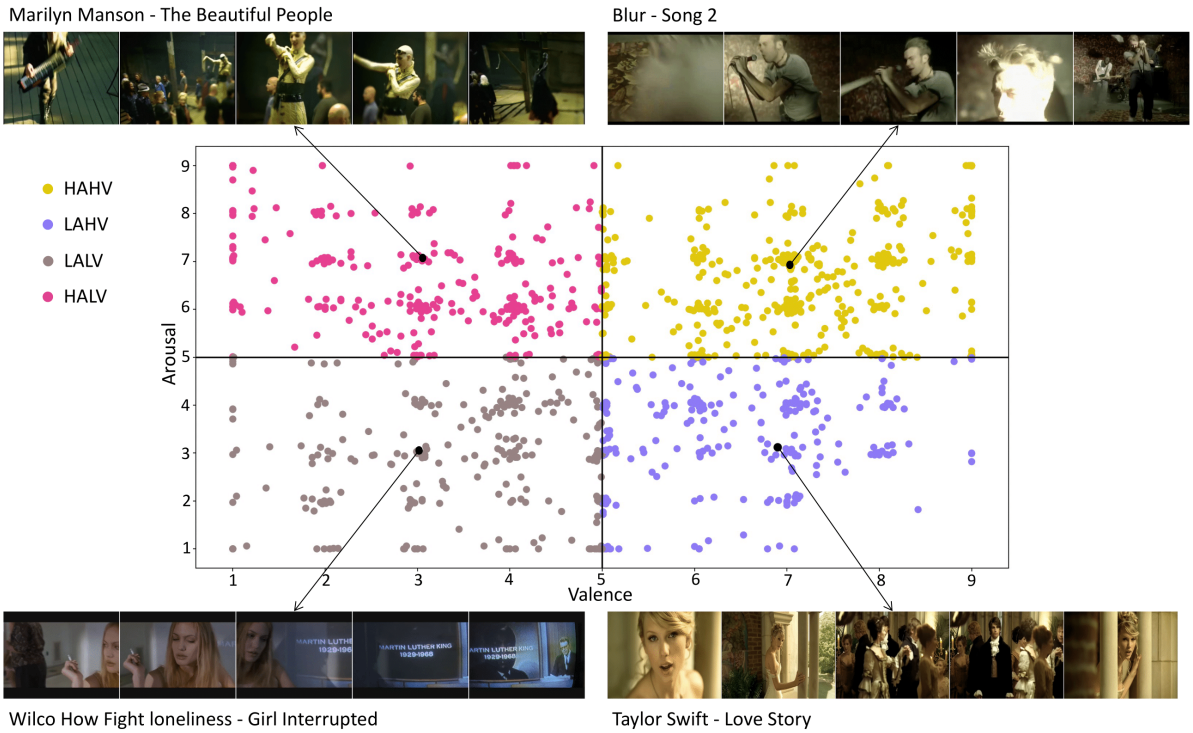

Fig. 2. Distribution of EEG samples from the DEAP dataset in the valence-arousal dimensions. high arousal and high valence (HAHV), high arousal and low valence (HALV), low arousal and high valence (LAHV), and low arousal and low valence (LALV). Each sample represents the reported arousal and valence scores by a participant who watched one of the music videos.

The EEG data comprises 32 channels that were downsampled to 128 Hz with electroocoulogram (EOG) artefacts removed. Also, a bandpass filter in the 4–45 Hz frequency range was applied to remove signal noise. For example, low frequency noise comes from sources such as movement of the head and electrode wires, and perspiration on the scalp. In contrast, high frequency noise comes from sources including electromagnetic interference, and muscle contractions (especially facial and neck muscles). Therefore, the dataset comprises reliable EEG data in the 4–45 Hz range.

The other dataset we used is MAHNOB-HCI [92], which is also very popular for affective computing tasks [32, 83, 107]. This dataset comprises EEG data from 27 participants who viewed 20 videos that lasted between 35 and 117 seconds. These participants provided valence and arousal ratings using a discrete scale ranging from 1 to 9. The sample distribution over the valence/arousal plane is illustrated in Figure 3. Various data sources were collected, including EEG, peripheral physiological signals, face and body video, eye gaze, and audio. The EEG signals were recorded at a sampling frequency of 256 Hz, utilizing 32 channels.

To make both datasets comparable for our experiments, we normalized the data in the following way. First all EEG signals were down-sampled to 128 Hz. Subsequently, we employed the Artifact Subspace Reconstruction (ASR) method [73] to remove transient and large-amplitude artifacts. Following that, a band-pass filter with a range of 4–45 Hz was applied. Additionally, we performed Common Average Referencing (CAR) [9] on the signals for each trial. To ensure uniform trial lengths comparable, we excluded all MAHNOB-HCI trials with a duration of less than 1 minute (since all DEAP trials were 1 min long) and signals over 1 minute were truncated, so only the initial 60 seconds were considered. As a result, we ended up with 449 trials out of the original 537 trials in

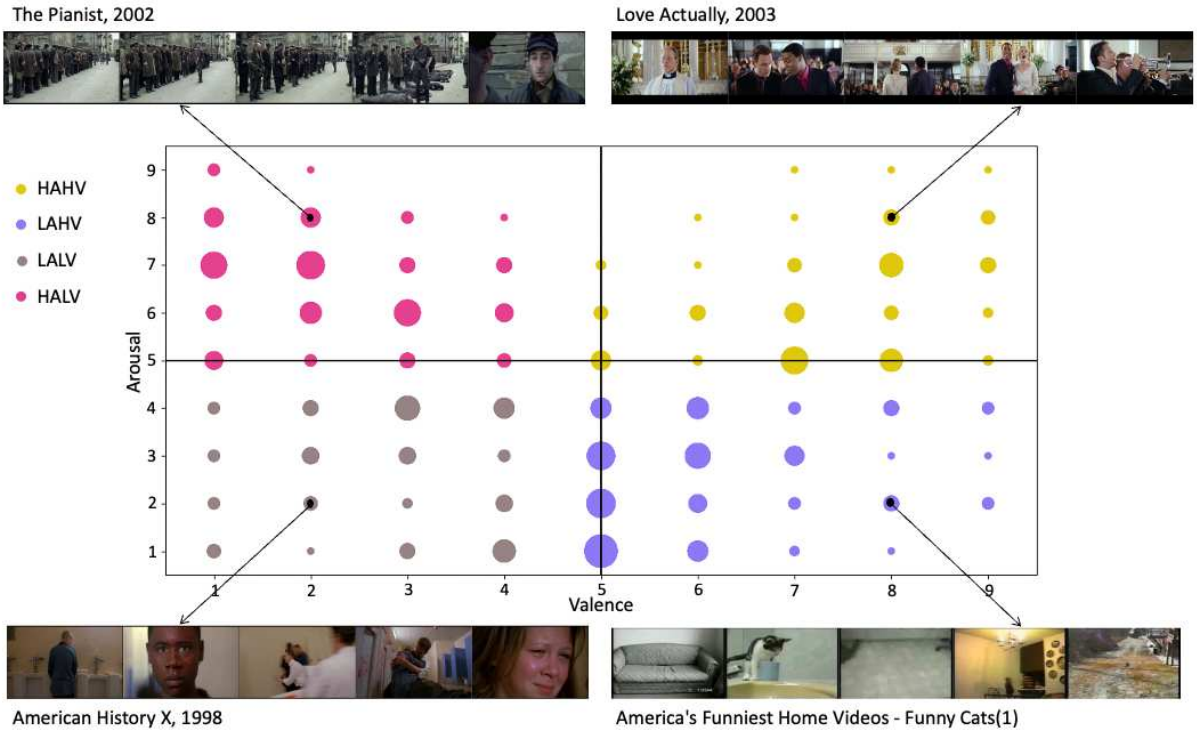

Fig. 3. Distribution of EEG samples from the MAHNOB-HCI dataset in the valence-arousal dimensions: high arousal and high valence (HAHV), high arousal and low valence (HALV), low arousal and high valence (LAHV), and low arousal and low valence (LALV). Since these ratings have discrete values from 1 to 9, radii represent sample size: a larger radius implies more instances. For example, at (valence,arousal) = (2,9) coordinate there is only a single instance; at the (6,4) coordinate there are 10 instances; and at the (5,1) coordinate there are 24 instances.

MAHNOB-HCI, all standardized to a signal length of 1 min. Both datasets had identical channel numbers and names, following the 10-20 standardized system for EEG [35, 41].

3.2 Experiment setup

The first design decision in our proposed investigation on efficient affect decoding is to determine the optimal temporal sampling window (w) to extract features. We study $w \in \{2, 4, 6, 10\}$ seconds, according to previous work [7, 42, 71, 112]. The second design decision is to determine the optimal EEG signal length (l), in seconds, to work with smaller signal chunks, which is defined as:

$$l = w \cdot n, \quad w, n \in \mathbb{N} \tag{1}$$

where n is a positive integer multiple of the window size, indicating the number of chunks into which the full EEG signal can be divided, i.e. $n \in \left\{1, 2, \ldots, \frac{60}{w}\right\}$. In other words, for a given window size there are n evenly spaced chunks of w seconds each. For example, for $w = 2$ s we can have up to $n = 30$ chunks of 2 s each, so l can be $2, 4, 6, \ldots, 60$ s. Similarly, for $w = 6$ s we can have up to $n = 10$ chunks of 6 s each, so l can be $6, 12, 18, \ldots, 60$ s. Equipped with this information, we experiment with increasing values of l until reaching the full signal value of 60 s. Figure 4 shows an example of splitting the original signal into different samples according to the selected parameters w and l.

The next step in our data processing pipeline is to extract the frequency bands from each EEG chunk: Theta (4–7 Hz), Alpha (8–12 Hz), Beta (13–30 Hz), and Gamma (31+ Hz) bands. For this, we apply the Fast Fourier transform (FFT) over the selected temporal window w and compute the Hjorth

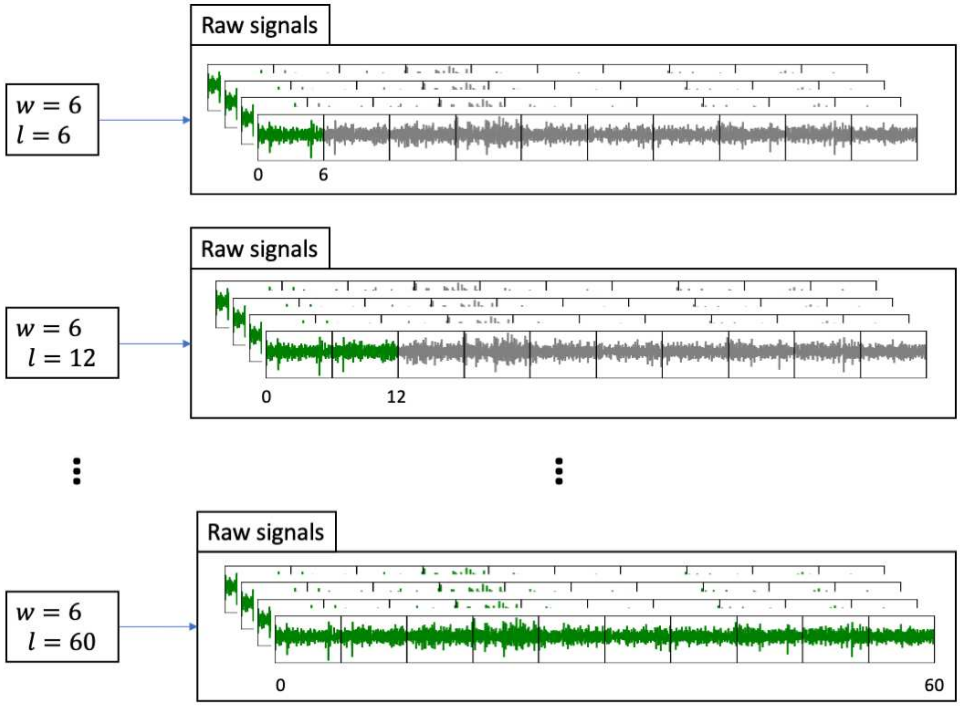

Fig. 4. Extracting samples (data instances) by making different signal lengths given a sampling window. This shows how we can study incrementally different signal lengths with a sampling window of constant size (here, 6 seconds). The green slices are the extracted EEG samples that we then use for affect decoding.

parameters (three features), spectral entropy, and signal energy. Figure 5 summarizes this process. As detailed below (Section 3.3), the signal energy is computed in the frequency domain while the other four features are computed in the original time domain after inverting the corresponding filtered Fourier transform.

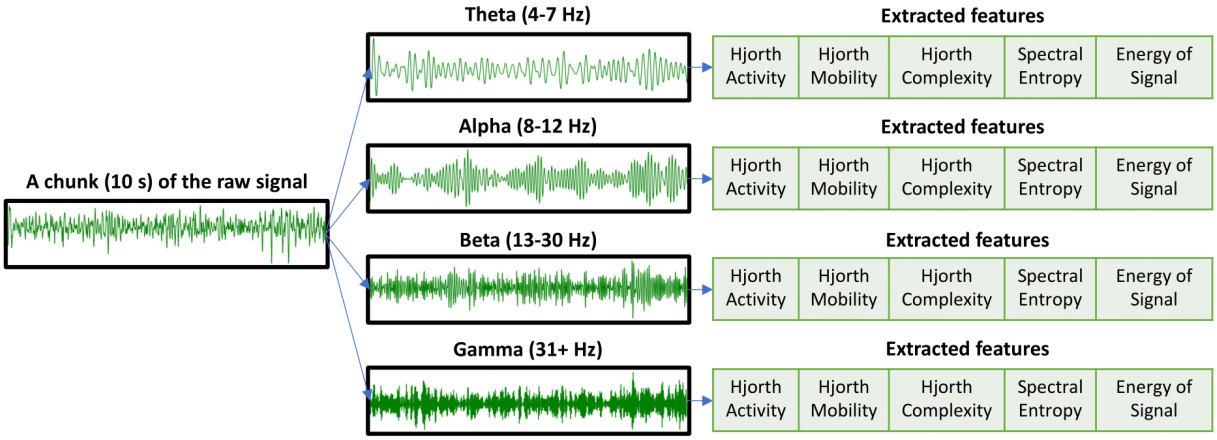

Fig. 5. Extraction of the frequency bands and features from EEG data. As an example, the raw signal shown here is a real chunk of $l = 10$ seconds of the original EEG signal from the DEAP dataset.

In sum, our design parameters are: sampling window (in seconds) w for feature extraction, signal length (duration, in seconds) l, and frequency band b. By systematically examining different sets and combinations of $\{w, l, b\}$, we sought to find the optimal parameters for affect decoding.

3.3 Feature extraction

We use three types of features, based on previous work, that have proved to work well with EEG signals: Hjorth Parameters [11, 30, 57, 69, 88, 102], Spectral Entropy [30, 82], and the Energy of the Signal [88, 102], as detailed below.

3.3.1 Hjorth Parameters. Hjorth parameters [34] describe time series in terms of their power spectrum: activity, mobility, and complexity. For a signal $x(t)$, the power spectrum is $S(m) = |X(m)|^2$, where $X(m)$ is the discrete Fourier transform of $x(t)$.

Hjorth activity is an indicator of the variance of the time series:

$$\text{Activity} = \text{Var}\left(x(t)\right) = \frac{1}{N-1}\sum_{i=1}^{N}|x_i - \mu_x|^2 \tag{2}$$

where $x(t)$ is the EEG signal, expressed as a discrete time series with N values, x_i, and μ_x is the mean of the signal values.

Hjorth mobility is proportional to the standard deviation of power spectrum:

$$\text{Mobility} = \sqrt{\frac{\text{Var}\left(\frac{d}{dt}x(t)\right)}{\text{Var}\left(x(t)\right)}} \tag{3}$$

Finally, Hjorth complexity is an indicator of change in the signal frequency:

$$\text{Complexity} = \frac{\text{Mobility}\left(\frac{d}{dt}x(t)\right)}{\text{Mobility}\left(x(t)\right)} \tag{4}$$

3.3.2 Spectral Entropy. The spectral entropy of a time series is an indicator of the spectral power distribution [21]. The probability distribution of an FFT-discretized signal $X(m)$ is $P(m) = \frac{S(m)}{\sum_i S(i)}$. The spectral entropy follows the definition of Shannon's entropy:

$$H = -\sum_{m=1}^{N} P(m)\log_2 P(m) \tag{5}$$

where N is the number of frequency points.

3.3.3 Energy of Signal. The last feature we extract from EEG signals is the Energy, calculated by the sum of the square of the signal magnitude, which measures the strength of a time series [28]:

$$E = \int_0^N |x(t)|^2\, dt \tag{6}$$

3.4 Classification models

According to literature reviews [30, 98], the most popular Machine Learning models for affect decoding from EEG signals are classifiers based on Support Vector Machine (SVM) and k-Nearest Neighbors (k-NN) [29, 30, 76]. Therefore we study these two classifiers for each combination of our design parameters $\{w, l, b\}$. Additionally, we also experimented with Fully-connected Neural Networks (FCNNs), given their increased relevance in the field, and Gated Transformer Networks (GTN), a kind of state-of-the-art Deep Learning models for multivariate time series classification. Note that FCNN models are given as input the same engineered features as SVM and kNN models,

however GTN models are given the raw values of each frequency band as input, from which suitable features are automatically derived.

3.4.1 Target labels. As in the vast majority of previous work, the considered classification tasks are binary: either predict low/high arousal or low/high valence. Arousal and valence scores are normalized in the $[0, 1]$ range. High arousal (or valence) classes are those with scores larger than or equal to 0.5, whereas low arousal (or valence) classes are those with scores lower than 0.5.

3.4.2 Data partitions. We randomly divide our data into training (90% of the data) and test (10%) sets. We use stratified sampling to ensure that both partitions are balanced in terms of our target classes. For the k-NN and SVM classifiers, the training set is divided into 10 folds (validation set) for fine-tuning model hyperparameters. For the FCNN and GTN classifiers, the training set is divided into training (90% of the samples) and validation (10%) sets.

3.4.3 SVM and k-NN Models hyperparameters. In SVM classifiers we consider the following hyperparameters: kernel $\in$ {Linear, RBF, Polynomial, Sigmoid}, regularization parameter $0.01 \leq C \leq 1000$, decay of non-linear kernels $0.01 \leq \gamma \leq 1000$, and degree of ploynominal kernel $2 \leq degree \leq 6$. In k-NN classifiers, the following hyperparameters are considered: number of neighbors $1 \leq k \leq 20$, neighboring voting weights $v \in \{\text{uniform}, \text{distance}\}$, and the metric p: Manhattan ($p = 1$), Euclidean ($p = 2$) and Minkowski L_p for $p > 2$. Hyperparameter fine-tuning is achieved with Bayesian optimization [14] as part of our training pipeline, since grid search becomes unfeasible given the combinatorial explosion of all possible values of the different model hyperparameters.

3.4.4 Architecture of fully-connected FCNNs. The model consists of two hidden dense layers with 800 neurons each and tanh as activation function. The model also incorporates a batch normalization layer after each dense layer for regularization, to prevent over-fitting and enhance generalization. The output layer of the model is a single neuron with sigmoid activation, meaning that it returns a probability and 0.5 is used to classify valence and arousal into low ($p \leq 0.5$) and high ($p > 0.5$) classes, respectively.

To train this model, we use the Adam optimizer with a learning rate of 10^{-3} and momentums $\beta_1 = \beta_2 = 0.99$. The learning rate is scheduled with a decay of 0.01. We use a batch size of 128 samples. The loss function to minimize is the binary cross-entropy. Training is capped at a maximum of 200 epochs, but we use early stopping with patience of 50 epochs (i.e., training will stop if the validation loss does not improve for 50 consecutive epochs and the optimal model weights will be retained).

3.4.5 Architecture of GTNs. This model consists of two Transformer branches [61]. The left branch finds spatial patterns between EEG channels, whereas the right branch find patterns in the temporal domain. We use embedding layers of 256 dimensions and 8 blocks of multi-head attention + fully-connected layers. The gating layer that merges both Transformer branches performs a weighted concatenation that helps the model learn channel-wise and step-wise correlations.

To train this model, we use the same optimizer proposed by Liu et al. [61]: Adagrad with scheduled learning rate 0.0001 and dropout of 0.2. We use binary cross-entropy loss, batch size of 8 samples, and 100 epochs with early stopping of 10 epochs.

4 RESULTS

We report in Figure 6 to Figure 9 Balanced Accuracy and F-measure (F_1 score, or the harmonic mean of Precision and Recall) considering the combinations of our design parameters (window size w, signal length l, and frequency band b). Balanced Accuracy and F-measure are the most popular metrics to inform about classification performance [24]. For the sake of brevity, we report results

of the best performing model in arousal and valence classification, using partial vs. full-length EEG signals. The model configurations are reported in Table 1 to Table 8. Classification results and configurations of all the other models are reported in the Supplementary Materials.

4.1 Summary of findings on the DEAP dataset

4.1.1 Arousal classification. When considering a single frequency band and partial EEG signals, the best classification result is 70%, achieved by SVM models using the Beta band, $w = 10$ seconds for feature extraction, and $l = 30$ seconds of EEG signal.

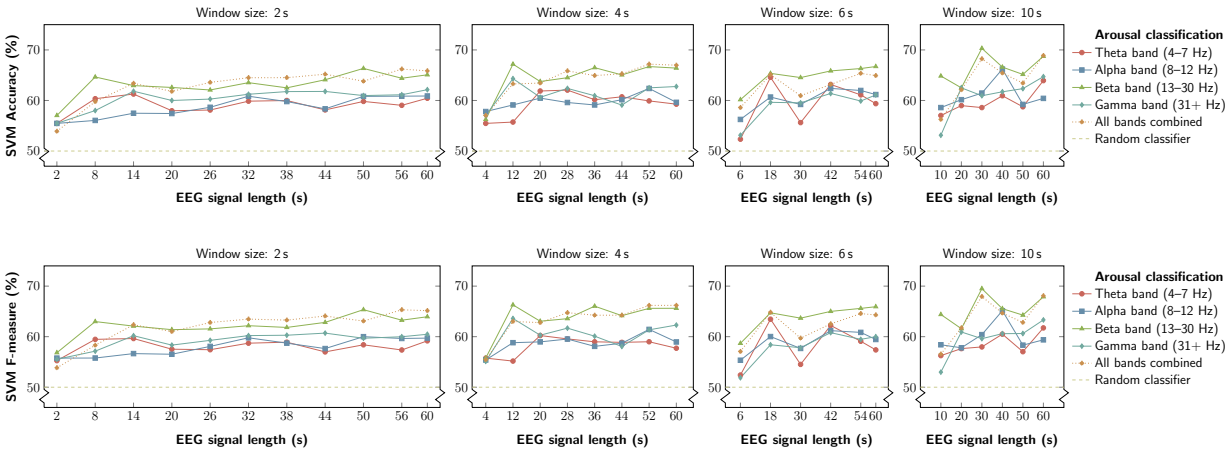

Fig. 6. Balanced accuracy (top) and F_1 score (bottom) of **arousal** classification with SVMs on DEAP as a function of the EEG signal length (horizontal axis, in seconds), using different sliding window sizes for feature extraction.

Regarding classification accuracy, we observed statistically significant differences between classifiers ($\chi^2(3, N = 5632) = 12.29, p < .01$). Post-hoc pairwise tests of proportions revealed that GTN performed worse than SVM ($p < .01$) and no differences were found between SVM, k-NN, and FCNN classifiers. Regarding frequency bands, no statistically significant differences were found between individual bands and the aggregated band set ($\chi^2(4, N = 2816) = 6.07, p > .05$). Regarding sampling window size w, no statistically significant differences were found between $\{2, 4, 6, 10\}$ seconds ($\chi^2(3, N = 5248) = 2.47, p > .05$). Finally, regarding signal length l, no statistically significant differences were found between $\{10, 20, \ldots, 60\}$ seconds ($\chi^2(5, N = 2688) = 6.59, p > .05$). All statistical tests are Bonferroni-Holm corrected, to guard against over-testing the data.

Table 1. Best performing SVM models for **arousal** classification in DEAP using **partial** EEG signals ($l < 60$).

			SVM hyperparams				
Freq. band	l	w	**kernel**	C	γ	**Bal. Acc.**	**F_1 score**
Theta	18	6	RBF	1000.000	0.049	0.646	0.635
Alpha	40	10	RBF	371.830	0.136	0.662	0.651
Beta	30	10	RBF	62.657	0.213	0.703	0.695
Gamma	12	4	RBF	1000.000	0.044	0.643	0.636
All bands	30	10	RBF	128.562	0.010	0.682	0.679

Table 2. Best performing SVM models for **arousal** classification in DEAP using **full-length** EEG signals ($l = 60$).

			SVM hyperparams				
Freq. band	l	w	**kernel**	C	γ	**Bal. Acc.**	**F_1 score**
Theta	60	10	RBF	109.299	0.235	0.639	0.617
Alpha	60	2	RBF	6.200	1.000	0.609	0.597
Beta	60	10	RBF	14.234	0.558	0.688	0.679
Gamma	60	10	RBF	1000.000	0.056	0.647	0.633
All bands	60	10	RBF	2.001	0.630	0.689	0.681

4.1.2 Valence classification. When considering a single frequency band and partial EEG signals, the best classification result is 65%, achieved by SVM models using the Beta band, $w = 6$ seconds for feature extraction, and $l = 48$ seconds of EEG signal.

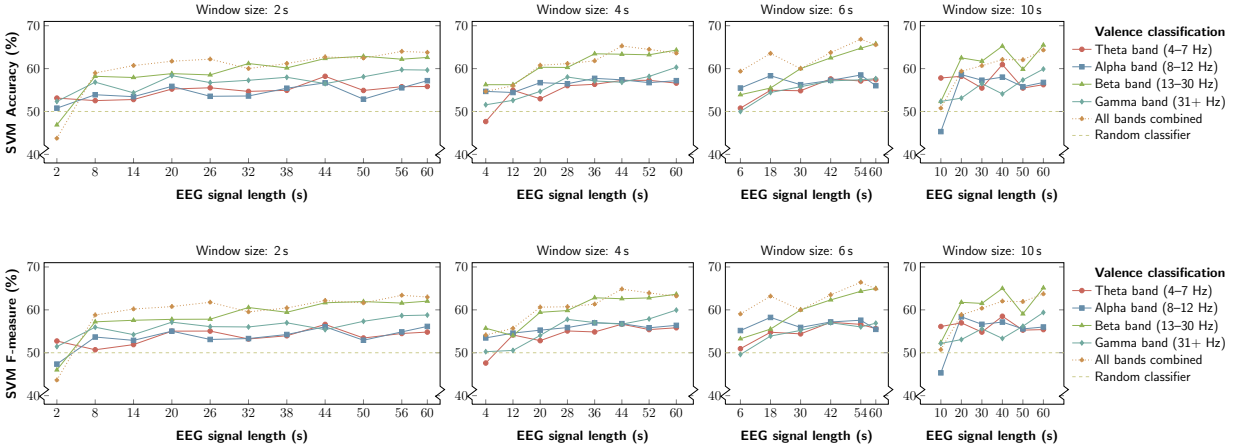

Fig. 7. Balanced accuracy (top) and F_1 score (bottom) of **valence** classification with SVMs on DEAP as a function of the EEG signal length (horizontal axis, in seconds), using different sliding window sizes for feature extraction.

Regarding classification accuracy, we observed statistically significant differences between classifiers ($\chi^2(3, N = 3968) = 11.41, p < .01$). Post-hoc pairwise tests of proportions revealed that GTN performed worse than SVM ($p < .001$) and no differences were found between SVM, k-NN, and FCNN classifiers. Regarding frequency bands, the Beta band demonstrated comparable performance to the combined band set ($\chi^2(1, N = 2432) = 0.30, p > .05$), while the remaining bands exhibited notably poorer performance ($\chi^2(4, N = 5632) = 24.20, p < .01$). Regarding sampling window size w, no statistically significant differences were found between $\{2, 4, 6, 10\}$ seconds ($\chi^2(3, N = 6656) = 2.91, p > .05$). Finally, regarding signal length l, no statistically significant differences were found between $\{6, 12, \ldots, 60\}$ seconds ($\chi^2(9, N = 7040) = 13.79, p > .05$). All statistical tests are Bonferroni-Holm corrected, to guard against over-testing the data.

4.2 Summary of findings on the MAHNOB-HCI dataset

4.2.1 Arousal classification. When considering a single frequency band and partial EEG signals, the best classification result is 88%, achieved by k-NN models using the Beta band, $w = 6$ seconds for feature extraction, and $l = 42$ seconds of EEG signal.

Table 3. Best performing SVM models for **valence** in DEAP using **partial** EEG signals ($l < 60$).

			SVM hyperparams				
Freq. band	l	w	**kernel**	C	γ	**Bal. Acc.**	**F_1 score**
Theta	16	4	RBF	5.264	1.000	0.600	0.591
Alpha	20	10	RBF	1000.000	0.072	0.586	0.584
Beta	48	6	RBF	5.749	1.000	0.656	0.651
Gamma	16	4	RBF	974.833	0.106	0.596	0.596
All bands	54	6	RBF	2.764	0.389	0.668	0.664

Table 4. Best performing SVM models for **valence** classification in DEAP using **full-length** EEG signals ($l = 60$).

			SVM hyperparams				
Freq. band	l	w	**kernel**	C	γ	**Bal. Acc.**	**F_1 score**
Theta	60	4	RBF	464.061	0.203	0.566	0.558
Alpha	60	4	RBF	271.481	0.215	0.572	0.564
Beta	60	10	RBF	14.206	0.477	0.655	0.651
Gamma	60	4	RBF	1000.000	0.475	0.603	0.600
All bands	60	6	RBF	2.887	0.448	0.655	0.649

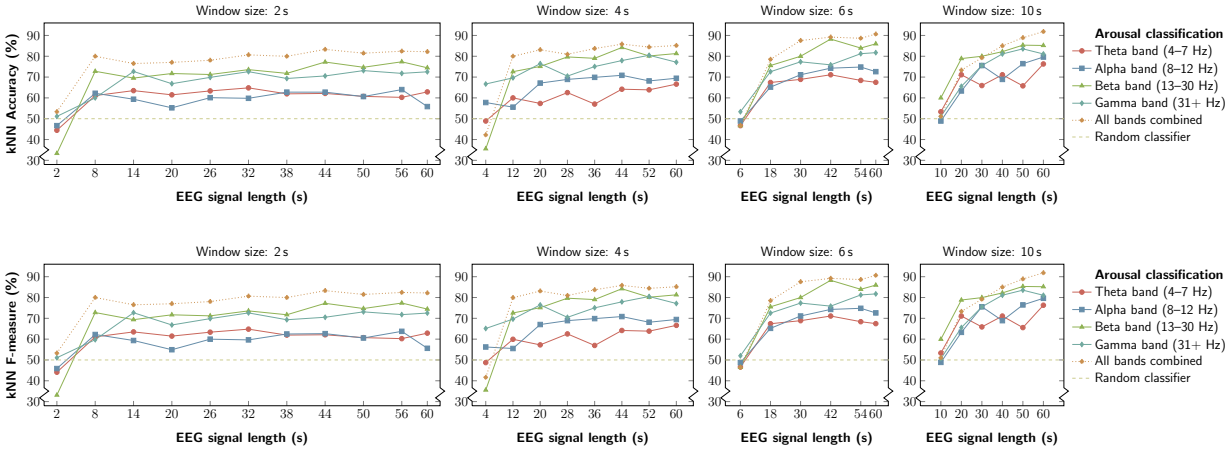

Fig. 8. Balanced accuracy (top) and F_1 score (bottom) of **arousal** classification with k-NNs on MAHNOB-HCI as a function of the EEG signal length (horizontal axis, in seconds), using different sliding window sizes for feature extraction.

Regarding classification accuracy, we observed statistically significant differences between classifiers ($\chi^2(3, N = 2661) = 14.38, p < .01$). Post-hoc pairwise tests of proportions revealed that GTN performed worse than k-NN ($p < .01$) and no differences were found between SVM, k-NN, and FCNN classifiers. Regarding frequency bands, the Beta band demonstrated comparable performance to the combined band set ($\chi^2(1, N = 585) = 2.08, p > .05$), while the remaining bands exhibited notably poorer performance ($\chi^2(4, N = 1485) = 32.49, p < .01$). Regarding sampling window size w, no statistically significant differences were found between $\{4, 6, 10\}$

seconds ($\chi^2(2, N = 1169) = 3.67, p > .05$) and all other comparisons yielded statistically significant results. Finally, regarding signal length l, durations of 40 seconds or longer achieved superior performance than shorter lengths ($\chi^2(5, N = 945) = 63.63, p < .01$). All statistical tests are Bonferroni-Holm corrected, to guard against over-testing the data.

Table 5. Best performing k-NN models for **arousal** classification in MAHNOB-HCI using **partial** EEG signals ($l < 60$).

			k-NN hyperparams				
Freq. band	l	w	k	p	v	**Bal. Acc.**	**F_1 score**
Theta	20	10	1.0	1.0	distance	0.711	0.711
Alpha	50	10	4.0	1.0	distance	0.764	0.764
Beta	42	6	1.0	1.0	distance	0.883	0.883
Gamma	48	6	2.0	1.0	distance	0.836	0.836
All bands	48	6	1.0	1.0	distance	0.908	0.908

Table 6. Best performing k-NN models for **arousal** classification in MAHNOB-HCI using **full-length** EEG signals ($l = 60$).

			k-NN hyperparams				
Freq. band	l	w	k	p	v	**Bal. Acc.**	**F_1 score**
Theta	60	10	1.0	1.0	distance	0.763	0.763
Alpha	60	10	6.0	1.0	distance	0.796	0.796
Beta	60	6	1.0	1.0	distance	0.860	0.860
Gamma	60	6	1.0	1.0	uniform	0.817	0.817
All bands	60	10	1.0	1.0	uniform	0.919	0.919

4.2.2 Valence classification. When considering a single frequency band and partial EEG signals, the best classification result is 85%, achieved by k-NN models using the Beta band, $w = 6$ seconds for feature extraction, and $l = 54$ seconds of EEG signal.

Regarding classification accuracy, we observed statistically significant differences between classifiers ($\chi^2(3, N = 2570) = 24.68, p < .01$). Post-hoc pairwise tests of proportions revealed that GTN performed worse than k-NN ($p < .001$) and no differences were found between SVM, k-NN, and FCNN classifiers. Regarding frequency bands, the Beta band demonstrated comparable performance to the combined band set ($\chi^2(1, N = 719) = 4.60, p > .05$), while the remaining bands exhibited notably poorer performance ($\chi^2(4, N = 1664) = 42.62, p < .01$). Regarding sampling window size w, no statistically significant differences were found between 6 and 10 seconds ($\chi^2(1, N = 719) = 0.62, p > .05$) and all other comparisons yielded statistically significant results. Finally, regarding signal length l, durations of 40 seconds or longer achieved superior performance than shorter lengths ($\chi^2(5, N = 945) = 95.32, p < .01$). All statistical tests are Bonferroni-Holm corrected, to guard against over-testing the data.

5 DISCUSSION

Overall, the analysis of biomedical data, in particular those generated by the human brain, is a challenging endeavor. In this paper, we have determined the most effective design parameters for

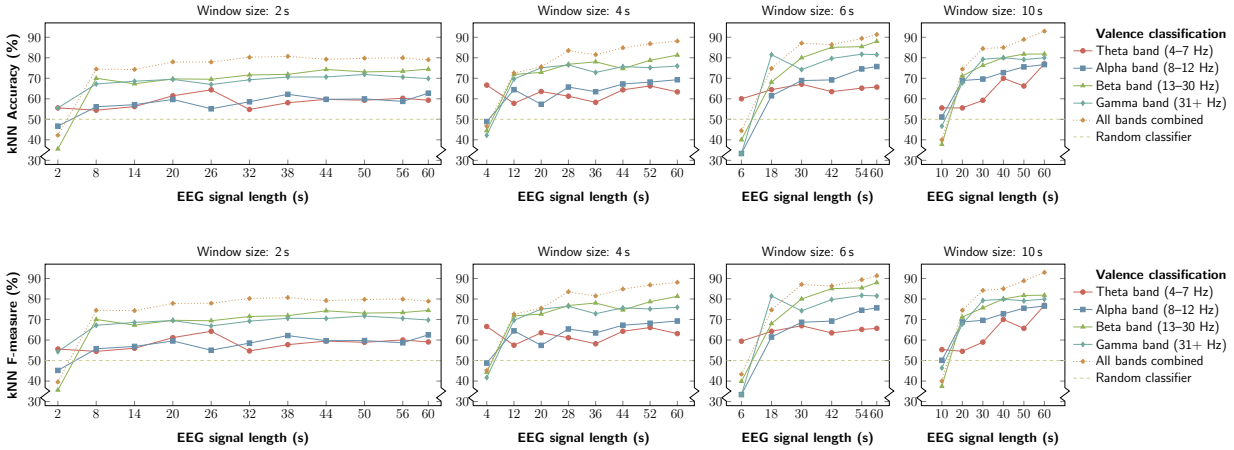

Fig. 9. Balanced accuracy (top) and F_1 score (bottom) of **valence** classification with k-NNs on MAHNOB-HCI as a function of the EEG signal length (horizontal axis, in seconds), using different sliding window sizes for feature extraction.

Table 7. Best performing k-NN models for **valence** classification in MAHNOB-HCI using **partial** EEG signals ($l < 60$).

			k-NN hyperparams				
Freq. band	l	w	k	p	v	**Bal. Acc.**	**F_1 score**
Theta	24	6	1.0	1.0	uniform	0.711	0.711
Alpha	50	10	1.0	1.0	distance	0.756	0.755
Beta	54	6	4.0	1.0	distance	0.854	0.854
Gamma	54	6	2.0	1.0	distance	0.817	0.818
All bands	54	6	2.0	1.0	distance	0.894	0.894

Table 8. Best performing k-NN models for **valence** classification in MAHNOB-HCI using **full-length** EEG signals ($l = 60$).

			k-NN hyperparams				
Freq. band	l	w	k	p	v	**Bal. Acc.**	**F_1 score**
Theta	60	10	1.0	1.0	distance	0.770	0.769
Alpha	60	10	1.0	1.0	distance	0.767	0.766
Beta	60	6	4.0	1.0	distance	0.880	0.880
Gamma	60	6	1.0	1.0	distance	0.815	0.815
All bands	60	10	1.0	1.0	distance	0.930	0.930

affect decoding from videos using EEG brain signals. Our analysis shows that sampling windows of 6 or 10 seconds outperform shorter windows, the latter being only beneficial for FCNN and GTN models. This observation has practical implications, as it can help researchers and practitioners to adopt larger window sizes in order to achieve better model performance with classic Machine Learning models. In terms of signal length, our analysis has challenged the notion for the necessity of a complete signal analysis. Concretely, we observed that a short signal length (in our case,

around 30 seconds) is sufficient to achieve peak performance. This is consistent with all the models considered. In terms of frequency band analysis, we observed that best results are achieved from either the Beta band alone or all bands combined. As discussed previously in Section 1, this finding has important implications for affect decoding, since decoding can not only be simpler to achieve but also be computationally more efficient.

In our experiments, Balanced Accuracy and F-measure achieved similar values, which means that recognition performance is not biased towards one of the two classes. In practical terms, the Precision and Recall metrics, from which the F-measure is computed, are meaningful metrics if the cost of false positives and false negatives are different in some particular application. Having an F-measure similar to the Balanced Accuracy suggests that these costs would not be a cause for concern in our affect decoding tasks.

Taken together, our findings suggest that affect decoding from EEG is possible within dynamic real-world scenarios without having to process the full-length EEG signal associated with the recorded stimulus and, in turn, without the user being exposed to the whole duration of such stimulus. Therefore, our findings suggest that practical affect decoding models can rely on more limited signal features and thus be more computationally efficient, faster to train, and less prone to signal variation. Based on our research findings, in the following we answer our research questions, regarding affect decoding during dynamic content perception.

RQ1: *What is the optimal EEG signal length for affect decoding during dynamic content perception?* Our results indicate that an EEG signal sequence lasting around 30 seconds, at most 42 seconds, is sufficient to attain a classification performance comparable to employing the full EEG signal. This not only can save processing time to make some automated decision, but also participants' effort (at a training stage) or user effort (during model deployment in potential applications). Figure 10 illustrates expected benefits through two envisioned usage scenarios.

RQ2: *What is the optimal temporal window for affect decoding during dynamic content perception?* Our results indicate that temporal windows of 6 and 10 seconds provide the best classification performance for classic Machine Learning models. On the contrary, modern Deep Learning models will benefit from using much shorter window sizes, around 2 seconds, since the resulting number of training instances will be much larger. Therefore, in practical settings, these can be the recommended window sizes for decoding affective states using different models.

RQ3: *Which frequency bands (or their combinations) are associated with affective states during dynamic content perception?* Our results indicate that a model trained on the Beta band alone achieves similar performance than a model trained on all frequency bands. Previous work also noted the effectiveness of using high-frequency bands for affect decoding in the context of dynamic contents [117], but most authors argue for using all bands combined (see e.g. Zhong et al. [119]). Our results also show that specific bands carry different amounts of information about affective states, and the Beta band alone shows comparable effectiveness w.r.t models using all bands. As a result, computational effort can be saved by considering processing of a single or smaller set of frequency bands, since using all bands may not always be required and processing power can be limited in certain application and usage scenarios.

In summary, our study offers valuable recommendations for efficient affect decoding from EEG signals when users are exposed to dynamic contents. By selecting the most adequate design parameters (sampling window size, signal length, and frequency band) researchers and practitioners can finally build computational models that are data efficient, faster to train, and less prone to signal noise.

5.1 Limitations and future work

The extent to which our findings can extrapolate to other scenarios has some limitations. First, DEAP and MAHNOB-HCI datasets are curated, so the selected clips are likely to elicit concrete feelings. This can partially explain why predicting using the initial duration of the EEG signals can already elicit the expected target affect (e.g. high valence) and, in turn, using a small part of the EEG signal from its onset turns out to be remarkably discriminative.

In this work we have treated signal segments as independent instances for classification, without any constraint regarding which participant and which video they correspond to. While this is standard practice [56, 71, 112], distinguishing those segments cross-participant or cross-video deserves to be explored, since this may significantly affect decoding performance [72].

We have experimented with several models that have been previously shown to perform well in similar classification scenarios, including modern representation learning techniques that learn optimized feature representations. However, we have observed that the learned representations do not lead to better classification performance, possibly because of a data scarcity issue, which is rather common in the research literature [1, 58].

Future work may consider crowdsourcing for obtaining further curated brain data. However, our knowledge of how to fuse participants' data for more reliable decoding of affective responses, is still limited. Specifically, the potential and intriguing relationship between the number of participants and the length of the signal required for robust decoding deserves further investigation.

Finally, additional signal modalities could also be explored, including for example peripheral sensors such as heart-rate variability or electrodermal activity. Fusing these with brain-responses might offer a more complete picture of the physiological activity associated with affective experiences.

5.2 Benefits and applications

Our results can benefit many research and application scenarios beyond standard offline decoding tasks. The allowance for shorter EEG signal lengths and optimized recognition windows means that researchers can conduct experiments faster and more efficient than before. As affect decoding performance tends to increase with longer exposure to video contents, the recorded and processed signals might also be decided and optimized according to the specific contents. For example, some contents might be more ambiguous in their affective signature and therefore they might require longer EEG signals. Alternatively, a sample of the signals from a crowd of users could be used for ensuring better consistency in the collected data.

More practically, our findings can help to summarize affective responses and therefore segment and annotate audiovisual content according to the experiences of users. In this regard, the models we have studied could be part of affect-aware adaptive systems that could operate under real-time constraints. A practical realization of annotation and adaptation applications could be the design of recommender systems, which could monitor users' affective responses during regular content consumption. Here, only the first chunks of the EEG recordings might need to be recorded or processed, thus saving valuable storage space and processing time, and allowing Machine Learning models to be built or updated more quickly. At recommendation time, the preferences of users might be derived from the estimation of how they consume contents during interaction, or even combined with other explicit requests; for example in response to a text query such as "melancholic music". Such practical systems could benefit from our findings, as affective responses could be recognized from users as a part of their everyday content consumption, without strong assumptions on signal recording, stimuli length, or extreme control of signal noise, with a potential for novel and better user experiences. Figure 10 illustrates two potential use cases in this regard.

Scenario 1: Tom is quietly browsing the web at home, while wearing an unobtrusive, plug-and-play, user-friendly EEG headset. He is interested in finding some music which fits his current mood. Instead of typing explicit text queries, he lets the music recommender decode his EEG signal. Instead of waiting the system to record one-minute long brain signal, the system can make online predictions in about 20 seconds, thus speeding up the process and consequently enhancing Tom's leisure experience.

Scenario 2: Helen is a busy PhD student who is recording brain signals from thirty participants in order to test a hypothesis of how an experimental condition influences human affect. She initially planned to record 2 minutes of brain signals per participant, but she learned that affect can be efficiently decoded with comparable accuracy with shorter segments, so she decides to shorten the recording time to 40 seconds, thus saving a significant amount of time that she will use to analyze the data for her thesis report. The participants are also happy since they will be wearing the EEG cap for shorter time and will finish the task earlier.

Fig. 10. A couple of envisioned use cases, highlighting the practical benefits derived from our findings.

6 CONCLUSION

Affect decoding is becoming a central research area in HCI, which has been conducted from various data sources, including behavioral, physiological, and brain activity. However, our understanding of how affect can be *efficiently* inferred using limited data captured in response to experiencing dynamic visual contents has remained elusive. Our results reveal that affective responses may be recognizable from limited time windows and limited EEG feature representations, as they occur in response to an exposure to dynamic stimuli. This may have impactful implications for the practical use of our findings in real-world HCI applications. For example, affect classifiers can be trained more efficiently by accounting only for a limited but significant part of brain signal associated to video-related data which contains the information that is most relevant for affect decoding. This opens up avenues for real-time systems that could benefit research practice, but also help to understand the users' cognitive and affective states when they consume information as a part of their everyday interactions with computers.

ACKNOWLEDGMENTS

Research supported by the Horizon 2020 FET program of the European Union (BANANA, grant CHIST-ERA-20-BCI-001), the European Innovation Council Pathfinder program (SYMBIOTIK, grant 101071147), the Academy of Finland (grants 313610, 322653, 328875), and the National Science Centre, Poland, under Grant Agreement no. 2021/03/Y/ST7/00008. This research is part of the project PCI2021-122036-2A, funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR.

REFERENCES

[1] A. Al-Nafjan, M. Hosny, Y. Al-Ohali, and A. Al-Wabil. 2017. Review and classification of emotion recognition based on EEG brain-computer interface system research: a systematic review. *Appl. Sci.* 7, 12 (2017), 1239.

[2] T. B. Alakus, M. Gonen, and I. Turkoglu. 2020. Database for an emotion recognition system based on EEG signals and various computer games – GAMEEMO. *Biomed. Signal Process. Control* 60 (2020), 101951.

[3] S. M. Alarcao and M. J. Fonseca. 2017. Emotions recognition using EEG signals: A survey. *IEEE Trans. Affect. Comput.* 10, 3 (2017), 374–393.

[4] A. Appriou, A. Cichocki, and F. Lotte. 2020. Modern machine-learning algorithms: for classifying cognitive and affective states from electroencephalography signals. *IIEEE Trans. Syst. Man Cybern. Syst.* 6, 3 (2020), 29–38.

[5] R. V. Aranha, C. G. Corrêa, and F. L. Nunes. 2019. Adapting software with affective computing: a systematic review. *IEEE Trans. Affect. Comput.* 12, 4 (2019), 883–899.

[6] J. Atkinson and D. Campos. 2016. Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers. *Expert Syst. Appl.* 47 (2016), 35–41.

[7] S. Bagherzadeh, K. Maghooli, A. Shalbaf, and A. Maghsoudi. 2022. Emotion recognition using effective connectivity and pre-trained convolutional neural networks in EEG signals. *Cogn. Neurodynamics* 16, 5 (2022), 1087–1106.

[8] L. Bai, J. Guo, T. Xu, and M. Yang. 2020. Emotional Monitoring of Learners Based on EEG Signal Recognition. *Procedia Comput. Sci.* 174 (2020), 364–368.

[9] O. Bertrand, F. Perrin, and J. Pernier. 1985. A theoretical justification of the average reference in topographic evoked potential studies. *Electroencephalogr. Clin. Neurophysiol.* 62, 6 (1985), 462–464.

[10] P. E. G. Bestelmeyer, S. A. Kotz, and P. Belin. 2017. Effects of emotional valence and arousal on the voice perception network. *Soc. Cogn. Affect. Neurosci.* 12, 8 (2017), 1351–1358.

[11] A. M. Bhatti, M. Majid, S. M. Anwar, and B. Khan. 2016. Human emotion recognition and analysis in response to audio music using brain signals. *Comput. Hum. Behav.* 65 (2016), 267–275.

[12] D. Blanco-Mora., A. Aldridge., C. Jorge., A. Vourvopoulos., P. Figueiredo., and S. Bermúdez i Badia. 2021. Finding the Optimal Time Window for Increased Classification Accuracy during Motor Imagery. In *Proc. BIODEVICES*. 144–151.

[13] S. Brave and C. Nass. 2007. Emotion in human-computer interaction. In *The human-computer interaction handbook*. 103–118.

[14] E. Brochu, V. M. Cora, and N. De Freitas. 2010. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv preprint arXiv:1012.2599* (2010).

[15] E. A. Butler. 2017. Emotions are temporal interpersonal systems. *Curr. Opin. Psychol.* 17 (2017), 129–134.

[16] S. Chang and H. Jun. 2019. Hybrid deep-learning model to recognise emotional responses of users towards architectural design alternatives. *J. Asian Archit. Build. Eng.* 18, 5 (2019), 381–391.

[17] J. Chen, P. Zhang, Z. Mao, Y. Huang, D. Jiang, and Y. Zhang. 2019. Accurate EEG-based emotion recognition on combined features using deep convolutional neural networks. *IEEE Access* 7 (2019), 44317–44328.

[18] L. S. Chen and T. S. Huang. 2000. Emotional expressions in audiovisual human computer interaction. In *Proc. ICME*. 423–426.

[19] Y. Cimtay and E. Ekmekcioglu. 2020. Investigating the use of pretrained convolutional neural network on cross-subject and cross-dataset EEG emotion recognition. *Sensors* 20, 7 (2020), 2034.

[20] K. M. Davis, L. Kangassalo, M. M. A. Spapé, and T. Ruotsalo. 2020. Brainsourcing: Crowdsourcing Recognition Tasks via Collaborative Brain-Computer Interfacing. In *Proc. CHI*, Regina Bernhaupt, Florian 'Floyd' Mueller, David Verweij, Josh Andres, Joanna McGrenere, Andy Cockburn, Ignacio Avellino, Alix Goguey, Pernille Bjøn, Shengdong Zhao, Briane Paul Samson, and Rafal Kocielnik (Eds.). 1–14.

[21] D. Devi, S. Sophia, and S. Boselin Prabhu. 2021. Chapter 4 - Deep learning-based cognitive state prediction analysis using brain wave signal. In *Cognitive Computing for Human-Robot Interaction*, Mamta Mittal, Rajiv Ratn Shah, and Sudipta Roy (Eds.). 69–84.

[22] G. Du, W. Zhou, C. Li, D. Li, and P. X. Liu. 2023. An Emotion Recognition Method for Game Evaluation Based on Electroencephalogram. *IEEE Trans. Affect. Comput.* 14, 1 (2023), 591–602.

[23] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu. 2013. Differential entropy feature for EEG-based emotion classification. In *Proc. NER*. 81–84.

[24] R. O. Duda, P. E. Hart, and D. G. Stork. 2001. *Pattern Classification* (second ed. ed.). John Wiley & Sons.

[25] E. Duffy. 1934. Emotion: an example of the need for reorientation in psychology. *Psychol. Rev.* 41, 2 (1934), 184.

[26] M. Egger, M. Ley, and S. Hanke. 2019. Emotion recognition from physiological signal analysis: A review. *Electron. Notes Theor. Comput.* 343 (2019), 35–55.

[27] P. C. Ellsworth and K. R. Scherer. 2003. *Appraisal processes in emotion.* Oxford University Press.

[28] O. Fasil and R. Rajesh. 2019. Time-domain exponential energy for epileptic EEG signal classification. *Neurosci. Lett.* 694 (2019), 1–8.

[29] F. Feradov, I. Mporas, and T. Ganchev. 2020. Evaluation of features in detection of dislike responses to audio–visual stimuli from EEG signals. *Computers* 9, 2 (2020), 33.

[30] F. Galvão, S. M. Alarcão, and M. J. Fonseca. 2021. Predicting exact valence and arousal values from EEG. *Sensors* 21, 10 (2021), 3414.

[31] S. S. Gilakjani and H. Al Osman. 2024. A Graph Neural Network for EEG-Based Emotion Recognition with Contrastive Learning and Generative Adversarial Neural Network Data Augmentation. *IEEE Access* 12 (2024), 113–130.

[32] W. M. B. Henia and Z. Lachiri. 2017. Emotion classification in arousal-valence dimension using discrete affective keywords tagging. In *Proc. ICEMIS*. 1–6.

[33] R. N. Henson. 2003. Neuroimaging studies of priming. *Prog. Neurobiol.* 70, 1 (2003), 53–81.

[34] B. Hjorth. 1970. EEG analysis based on time domain properties. *Electroencephalogr. Clin. Neurophysiol.* 29, 3 (1970), 306–310.
[35] G. E. Holder, G. G. Celesia, Y. Miyake, S. Tobimatsu, R. G. Weleber, et al. 2010. International Federation of Clinical Neurophysiology: recommendations for visual system testing. *Clin. Neurophysiol.* 121, 9 (2010), 1393–1409.
[36] W. Hu, G. Huang, L. Li, L. Zhang, Z. Zhang, and Z. Liang. 2020. Video-triggered EEG-emotion public databases and current methods: a survey. *Brain Sci. Adv.* 6, 3 (2020), 255–287.
[37] M. Imani and G. A. Montazer. 2019. A survey of emotion recognition methods with emphasis on E-Learning environments. *J. Netw. Comput. Appl.* 147 (2019), 102423.
[38] M. P. Kalashami, M. M. Pedram, and H. Sadr. 2022. EEG Feature Extraction and Data Augmentation in Emotion Recognition. *Comput. Intell. Neurosci.* 2022 (2022).
[39] N. Kamel and A. S. Malik. 2015. *EEG/ERP Analysis: Methods and Applications.* CRC Press, Taylor & Francis.
[40] S. Katsigiannis and N. Ramzan. 2017. DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices. *IEEE J. Biomed. Health Inform.* 22, 1 (2017), 98–107.
[41] A. Kawala-Sterniuk, N. Browarska, A. Al-Bakri, M. Pelc, J. Zygarlicki, M. Sidikova, R. Martinek, and E. J. Gorzelanczyk. 2021. Summary of over fifty years with brain-computer interfaces—a review. *Brain Sci.* 11, 1 (2021), 43.
[42] S.-H. Kim, H.-J. Yang, N. A. T. Nguyen, S. K. Prabhakar, and S.-W. Lee. 2021. WeDea: A new EEG-based framework for emotion recognition. *IEEE J. Biomed. Health Inform.* 26, 1 (2021), 264–275.
[43] D.-H. Ko, D.-H. Shin, and T.-E. Kam. 2021. Attention-based spatio-temporal-spectral feature learning for subject-specific EEG classification. In *Proc. BCI.*
[44] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras. 2011. DEAP: A database for emotion analysis; using physiological signals. *IEEE Trans. Affect. Comput.* 3, 1 (2011), 18–31.
[45] J. Kosiński, K. Szklanny, A. Wieczorkowska, and M. Wichrowski. 2018. An Analysis of Game-Related Emotions Using EMOTIV EPOC. In *Proc. FedCSIS.* 913–917.
[46] N. Kos'myna and F. Tarpin-Bernard. 2013. Evaluation and Comparison of a Multimodal Combination of BCI Paradigms and Eye Tracking With Affordable Consumer-Grade Hardware in a Gaming Context. *IEEE Trans. Comput. Intell. AI Games* 5, 2 (2013), 150–154.
[47] M. Kostyunina and M. Kulikov. 1996. Frequency characteristics of EEG spectra in the emotions. *Neurosci. Behav. Physiol.* 26, 4 (1996), 340–343.
[48] A. Kumar and A. Kumar. 2021. DEEPHER: Human Emotion Recognition Using an EEG-Based DEEP Learning Network Model. *Eng. Proc.* 10, 1 (2021).
[49] N. Kumar, K. Khaund, and S. M. Hazarika. 2016. Bispectral analysis of EEG for emotion recognition. *Procedia Comput. Sci.* 84 (2016), 31–35.
[50] P. J. Lang. 1995. The emotion probe. Studies of motivation and attention. *Am. Psychol.* 50 (1995), 372–385.
[51] R. J. Larsen and E. Diener. 1992. Promises and problems with the circumplex model of emotion. In *Review of personality and social psychology*, M. Clark (Ed.). Vol. 13. 25–59.
[52] E. Lashgari, D. Liang, and U. Maoz. 2020. Data augmentation for deep-learning-based electroencephalography. *J. Neurosci. Methods* 346 (2020), 108885.
[53] K. Latifzadeh and L. A. Leiva. 2022. Gustav: Cross-device Cross-computer Synchronization of Sensory Signals. In *Adj. Proc. UIST.*
[54] R. W. Levenson. 2003. Blood, sweat, and fears: The autonomic architecture of emotion. *Ann. N. Y. Acad. Sci.* 1000, 1 (2003), 348–366.
[55] C. Li, Z. Zhang, R. Song, J. Cheng, Y. Liu, and X. Chen. 2021. EEG-based emotion recognition via neural architecture search. *IEEE Trans. Affect. Comput.* 14, 2 (2021), 957–968.
[56] M. Li, H. Xu, X. Liu, and S. Lu. 2018. Emotion recognition from multichannel EEG signals using K-nearest neighbor classification. *Technol. Health Care* 26, S1 (2018), 509–519.
[57] X. Li, D. Song, P. Zhang, Y. Zhang, Y. Hou, and B. Hu. 2018. Exploring EEG features in cross-subject emotion recognition. *Front. Neurosci.* 12 (2018), 162.
[58] X. Li, Y. Zhang, P. Tiwari, D. Song, B. Hu, M. Yang, Z. Zhao, N. Kumar, and P. Marttinen. 2022. EEG based Emotion Recognition: A Tutorial and Review. *ACM Comput. Surv.* 55, 4 (2022).
[59] Y. Li, J. Chen, F. Li, B. Fu, H. Wu, Y. Ji, Y. Zhou, Y. Niu, G. Shi, and W. Zheng. 2022. GMSS: Graph-Based Multi-Task Self-Supervised Learning for EEG Emotion Recognition. *IEEE Trans. Affect. Comput.* 14, 3 (2022), 2512–2525.
[60] C. L. Lisetti and F. Nasoz. 2002. MAUI: a multimodal affective user interface. In *Proc. ACM MM.* 161–170.
[61] M. Liu, S. Ren, S. Ma, J. Jiao, Y. Chen, Z. Wang, and W. Song. 2021. Gated transformer networks for multivariate time series classification. *arXiv preprint arXiv:2103.14438* (2021).
[62] Y. Liu and O. Sourina. 2014. EEG-based subject-dependent emotion recognition algorithm using fractal dimension. In *Proc. SMC.* 3166–3171.

[63] I. Lopatovska and I. Arapakis. 2011. Theories, methods and current research on emotions in library and information science, information retrieval and human–computer interaction. *Inf. Process. Manag.* 47, 4 (2011), 575–592.

[64] J.-M. López-Gil, J. Virgili-Gomá, R. Gil, T. Guilera, I. Batalla, J. Soler-González, and R. García. 2016. Method for improving EEG based emotion recognition by combining it with synchronized biometric and eye tracking technologies in a non-invasive and low cost way. *Front. Comput. Neurosci.* 10 (2016), 85.

[65] Y. Luo and B.-L. Lu. 2018. EEG data augmentation for emotion recognition using a conditional Wasserstein GAN. In *Proc. EMBC*. 2535–2538.

[66] Y. Luo, L.-Z. Zhu, and B.-L. Lu. 2019. A GAN-based data augmentation method for multimodal emotion recognition. In *Proc. ISNN*. 141–150.

[67] R. Mahini, Y. Li, W. Ding, R. Fu, T. Ristaniemi, A. K. Nandi, G. Chen, and F. Cong. 2020. Determination of the time window of event-related potential using multiple-set consensus clustering. *Front. Neurosci.* 14 (2020), 521595.

[68] I. Mazumder. 2019. An analytical approach of EEG analysis for emotion recognition. In *Proc. DevIC*. 256–260.

[69] R. M. Mehmood, M. Bilal, S. Vimal, and S.-W. Lee. 2022. EEG-based affective state recognition from human brain signals by using Hjorth-activity. *Measurement* 202 (2022), 111738.

[70] M. L. R. Menezes, A. Samara, L. Galway, A. Sant'Anna, A. Verikas, F. Alonso-Fernandez, H. Wang, and R. Bond. 2017. Towards emotion recognition for virtual environments: an evaluation of EEG features on benchmark dataset. *Pers. Ubiquit. Comput.* 21 (2017), 1003–1013.

[71] Z. Mohammadi, J. Frounchi, and M. Amiri. 2017. Wavelet-based emotion recognition system using EEG signal. *Neural Comput. Appl.* 28, 8 (2017), 1985–1990.

[72] Y. Moreno-Alcayde, V. J. Traver, and L. Leiva. 2024. Sneaky Emotions: Impact of Data Partitions in Affective Computing Experiments with Brain-Computer Interfacing. *Biomed. Eng. Lett.* 14, 1 (2024), 103–113.

[73] T. Mullen, C. Kothe, Y.-M. Chi, A. Ojeda, T. Kerth, S. Makeig, G. Cauwenberghs, and T.-P. Jung. 2013. Real-time modeling and 3D visualization of source dynamics and connectivity using wearable EEG. In *Proc. EMBC*. 2184–2187.

[74] D. A. Norman. 2004. *Emotional design: Why we love (or hate) everyday things.* Civitas Books.

[75] D. Ouyang, Y. Yuan, G. Li, and Z. Guo. 2022. The Effect of Time Window Length on EEG-Based Emotion Recognition. *Sensors* 22, 13 (2022), 4939.

[76] M. S. Özerdem and H. Polat. 2017. Emotion recognition based on EEG features in movie clips with channel selection. *Brain Inform.* 4, 4 (2017), 241–252.

[77] E. S. Pane, A. D. Wibawa, and M. H. Pumomo. 2018. Channel selection of EEG emotion recognition using stepwise discriminant analysis. In *Proc. CENIM*. 14–19.

[78] V. Peterson, C. Galván, H. Hernández, and R. Spies. 2020. A feasibility study of a complete low-cost consumer-grade brain-computer interface system. *Heliyon* 6, 3 (2020), e03425.

[79] R. W. Picard. 2000. *Affective computing.* MIT press.

[80] R. W. Picard and J. Klein. 2002. Computers that recognise and respond to user emotion: theoretical and practical implications. *Interact. Comput.* 14, 2 (2002), 141–169.

[81] R. W. Picard, E. Vyzas, and J. Healey. 2001. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Trans. Pattern Anal. Mach. Intell.* 23, 10 (2001), 1175–1191.

[82] M. A. Rahman, M. F. Hossain, M. Hossain, and R. Ahmmed. 2020. Employing PCA and t-statistical approach for feature extraction and classification of emotion from multichannel EEG signal. *Egypt. Inform. J.* 21, 1 (2020), 23–35.

[83] S. Rayatdoost, D. Rudrauf, and M. Soleymani. 2020. Expression-guided EEG representation learning for emotion recognition. In *Proc. ICASSP*. 3222–3226.

[84] J. A. Russell. 1980. A circumplex model of affect. *J. Pers. Soc. Psychol.* 39, 6 (1980), 1161.

[85] S. Saha and M. Baumert. 2019. Intra- and Inter-subject Variability in EEG-based Sensorimotor Brain Computer Interface: A Review. *Frontiers Comput. Neurosci.* 13 (2019), 87.

[86] S. Saha, K. A. Mamun, K. Ahmed, R. Mostafa, G. R. Naik, S. Darvishi, A. H. Khandoker, and M. Baumert. 2021. Progress in brain computer interface: challenges and opportunities. *Front. Syst. Neurosci.* 15, 578875 (2021).

[87] E. S. Salama, R. A. El-Khoribi, M. E. Shoman, and M. A. W. Shalaby. 2018. EEG-based emotion recognition using 3D convolutional neural networks. *Int. J. Adv. Comput. Sci. Appl.* 9, 8 (2018).

[88] L. Shaw and A. Routray. 2016. Statistical features extraction for multivariate pattern analysis in meditation EEG using PCA. In *Proc. ISC*. 1–4.

[89] X. Shen, X. Liu, X. Hu, D. Zhang, and S. Song. 2023. Contrastive Learning of Subject-Invariant EEG Representations for Cross-Subject Emotion Recognition. *IEEE Trans. Affect. Comput.* 14, 3 (2023), 2496–2511.

[90] L. Shu, J. Xie, M. Yang, Z. Li, Z. Li, D. Liao, X. Xu, and X. Yang. 2018. A review of emotion recognition using physiological signals. *Sensors* 18, 7 (2018), 2074.

[91] N. Singh Malan and S. Sharma. 2021. Time window and frequency band optimization using regularized neighbourhood component analysis for Multi-View Motor Imagery EEG classification. *Biomed. Signal Process. Control* 67 (2021), 102550.

[92] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic. 2011. A multimodal database for affect recognition and implicit tagging. *IEEE Trans. Affect. Comput.* 3, 1 (2011), 42–55.
[93] T. Song, W. Zheng, P. Song, and Z. Cui. 2020. EEG Emotion Recognition Using Dynamical Graph Convolutional Neural Networks. *IEEE Trans. Affect. Comput.* 11, 3 (2020), 532–541.
[94] H. A. Spelt, J. H. Westerink, L. Frank, J. Ham, and W. A. IJsselsteijn. 2022. Physiology-based personalization of persuasive technology: a user modeling perspective. *User Model. User-Adapt. Interact.* 32, 1 (2022), 133–163.
[95] N. Thammasan, K.-i. Fukui, and M. Numao. 2016. Application of deep belief networks in EEG-based dynamic music-emotion recognition. In *Proc. IJCNN*. 881–888.
[96] S. Thejaswini, K. M. R. Kumar, and A. N. J. L. 2019. Analysis of EEG based emotion detection of DEAP and SEED-IV databases using SVM. In *Proc. ICETSE*.
[97] E. P. Torres, E. A. Torres, M. Hernández-Álvarez, and S. G. Yoo. 2021. Real-Time Emotion Recognition for EEG Signals Recollected from Online Poker Game Participants. In *Proc. Advances in Artificial Intelligence, Software and Systems Engineering*, Tareq Z. Ahram, Waldemar Karwowski, and Jay Kalra (Eds.). 236–241.
[98] E. P. Torres, E. A. Torres, M. Hernández-Álvarez, and S. G. Yoo. 2020. EEG-based BCI Emotion Recognition: A Survey. *Sensors* 20, 18 (2020), 5083.
[99] V. J. Traver, J. Zorío, and L. A. Leiva. 2021. Glimpse: A Gaze-Based Measure of Temporal Salience. *Sensors* 21, 9 (2021).
[100] K. D. Tzimourta, N. Giannakeas, A. T. Tzallas, L. G. Astrakas, T. Afrantou, P. Ioannidis, N. Grigoriadis, P. Angelidis, D. G. Tsalikakis, and M. G. Tsipouras. 2019. EEG Window Length Evaluation for the Detection of Alzheimer's Disease over Different Brain Regions. *Brain Sci.* 9, 4 (2019).
[101] K. P. Wagh and K. Vasanth. 2019. Electroencephalograph (EEG) based emotion recognition system: A review. *Innov. Electron. Commun. Eng.* 33 (2019), 37–59.
[102] K. P. Wagh and K. Vasanth. 2022. Performance evaluation of multi-channel electroencephalogram signal (EEG) based time frequency analysis for human emotion recognition. *Biomed. Signal Process. Control* 78 (2022), 103966.
[103] J. Wagner, J. Kim, and E. André. 2005. From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification. In *Proc. ICME*. 940–943.
[104] F. Wang, S.-h. Zhong, J. Peng, J. Jiang, and Y. Liu. 2018. Data augmentation for EEG-based emotion recognition with deep convolutional neural networks. In *Proc. MMM*. 82–93.
[105] P. Wang, Z. Song, H. Chen, T. Fang, Y. Zhang, X. Zhang, S. Wang, H. Li, Y. Lin, J. Jia, L. Zhang, and X. Kang. 2021. Application of Combined Brain Computer Interface and Eye Tracking. In *Proc. BCI*.
[106] Z.-M. Wang, S.-Y. Hu, and H. Song. 2019. Channel selection method for EEG emotion recognition using normalized mutual information. *IEEE Access* 7 (2019), 143303–143311.
[107] M. B. H. Wiem and Z. Lachiri. 2017. Emotion classification in arousal valence model using MAHNOB-HCI database. *Int. J. Adv. Comput. Sci. Appl.* 8, 3 (2017).
[108] T. Xu, Y. Zhou, Z. Wang, and Y. Peng. 2018. Learning Emotions EEG-based Recognition and Brain Activity: A Survey Study on BCI for Intelligent Tutoring System. *Procedia Comput. Sci.* 130 (2018), 376–382.
[109] X. Xu, F. Wei, Z. Zhu, J. Liu, and X. Wu. 2020. EEG feature selection using orthogonal regression: Application to emotion recognition. In *Proc. ICASSP*. 1239–1243.
[110] J. Yan, S. Chen, and S. Deng. 2019. A EEG-based emotion recognition model with rhythm and time characteristics. *Brain Inform.* 6, 1 (2019), 1–8.
[111] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang. 2007. A survey of affect recognition methods: audio, visual and spontaneous expressions. In *Proc. ICMI*. 126–133.
[112] J. Zhang, M. Chen, S. Zhao, S. Hu, Z. Shi, and Y. Cao. 2016. ReliefF-based EEG sensor selection methods for emotion recognition. *Sensors* 16, 10 (2016), 1558.
[113] Y. Zhang, J. Chen, J. H. Tan, Y. Chen, Y. Chen, D. Li, L. Yang, J. Su, X. Huang, and W. Che. 2020. An investigation of deep learning models for EEG-based emotion recognition. *Front. Neurosci.* 14 (2020), 622759.
[114] Y. Zhang, G. Yan, W. Chang, W. Huang, and Y. Yuan. 2023. EEG-based multi-frequency band functional connectivity analysis and the application of spatio-temporal features in emotion recognition. *Biomed. Signal Process. Control* 79 (2023), 104157.
[115] Z. Zhang, S.-h. Zhong, and Y. Liu. 2023. GANSER: A Self-supervised Data Augmentation Framework for EEG-based Emotion Recognition. *IEEE Trans. Affect. Comput.* 14, 3 (2023), 2048–2063.
[116] W.-L. Zheng, B.-N. Dong, and B.-L. Lu. 2014. Multimodal emotion recognition using EEG and eye tracking data. In *Proc. EMBC*. 5040–5043.
[117] W.-L. Zheng and B.-L. Lu. 2015. Investigating Critical Frequency Bands and Channels for EEG-based Emotion Recognition with Deep Neural Networks. *IEEE Trans. Auton. Ment. Dev.* 7, 3 (2015), 162–175.
[118] W.-L. Zheng, J.-Y. Zhu, and B.-L. Lu. 2019. Identifying Stable Patterns over Time for Emotion Recognition from EEG. *IEEE Trans. Affect. Comput.* 10, 3 (2019), 417–429.

[119] P. Zhong, D. Wang, and C. Miao. 2022. EEG-based Emotion Recognition Using Regularized Graph Neural Networks. *IEEE Trans. Affect. Comput.* 13, 3 (2022), 1290–1301.

[120] Y. Zhou, T. Xu, S. Li, and R. Shi. 2019. Beyond Engagement: An EEG-Based Methodology for Assessing User's Confusion in an Educational Game. *Univers. Access Inf. Soc.* 18, 3 (2019), 551–563.