Decoding the Neural Signatures of Valence and Arousal From Portable EEG Headset

Nikhil Garg[1,2,3,*], Rohit Garg[1,3], Parrivesh NS[1,3], Apoorv Anand[1,3], V.A.S. Abhinav[1,3], Veeky Baths[1,3,*]

1 Cognitive Neuroscience Lab, BITS Pilani - K.K.Birla Goa Campus, Goa, India
2 Institut Interdisciplinaire d'Innovation Technologique (3IT), Laboratoire Nanotechnologies Nanosystèmes (LN2) – CNRS UMI-3463 – 3IT, Université de Sherbrooke, Sherbrooke, Canada
3 Ironwork Insights Pvt. Ltd., India

* Corresponding authors
* veeky@goa.bits-pilani.ac.in
* f20160597@goa.bits-pilani.ac.in

Abstract

This paper focuses on classifying emotions on the valence-arousal plane using various feature extraction, feature selection and machine learning techniques. Emotion classification using EEG data and machine learning techniques has been on the rise in the recent past. We evaluate different feature extraction techniques, feature selection techniques and propose the optimal set of features and electrodes for emotion recognition. The images from the OASIS image dataset were used for eliciting the valence and arousal emotions, and the EEG data was recorded using the Emotiv Epoc X mobile EEG headset. The analysis is additionally carried out on publicly available datasets: DEAP and DREAMER. We propose a novel feature ranking technique and incremental learning approach to analyze the dependence of performance on the number of participants. Leave-one-subject-out cross-validation was carried out to identify subject bias in emotion elicitation patterns. The importance of different electrode locations was calculated, which could be used for designing a headset for emotion recognition. Our study achieved root mean square errors (RMSE) of less than 0.75 on DREAMER, 1.76 on DEAP, and 2.39 on our dataset.

Introduction

The role of human emotion in cognition is very vital and has been studied for a long time, with different experimental and behavioural paradigms. For a very long time, psychology and marketing researchers have tried to understand human perception through surveys. In recent times, with the increasing need to learn about human perception, without external and internal human biases such as varying facial expressions and conception of various emotions across people [1], the use of brain electrophysiological and imaging data have been on the rise. Moreover, with BCI algorithms being developed for EEG data, the scope of using electrophysiological data for emotion prediction has widened. This would require standardizing emotion metrics to train any pattern recognition algorithm. One such metric in play is the valence-arousal 2D space proposed by Russel [2]. The circumplex model, obtained by

asking the participants to rate their response to 28 terms of the English language corresponding to various states of the human mind, on a 2D surface with eight emotional states already pre-indexed with the orthogonal axes corresponding to valence and arousal, laid the foundation for quantification of emotion. Following this, the effort to associate emotion quantitatively with words, pictures (visual stimuli), sounds (auditory stimuli), and videos (audio-visual stimuli) [3–11] and now we majorly deal with valence, arousal, likeness , and dominance.

With the advancements in BCI technology, the research work in machine learning-based neuropsychology started growing significantly. Two new fields, consumer neuroscience, and neuromarketing were born as an effect of the commercialization of these technologies. Some of the most commonly used techniques are EEG, fMRI, skin conductance, and eye-tracking. The temporal resolution of EEG can go to around 20KHz compared to that of fMRI, which goes to about 0.5Hz but can provide better spatial resolutions. Since emotions are triggered almost instantly, we prefer EEG studies due to better temporal resolutions for emotion studies. Electroencephalography or EEG is a technique used to obtain the electrical activity of the brain by placing electrodes at various positions on the scalp of the subject, generally studied under various experimental and behavioural tasks like performing basic mental ability tasks, viewing images, videos, smelling or even sleep [10, 12–15].

Each EEG frequency band is dominant during different roles, corresponding to various emotional and cognitive states. Besides using energy spectral values, researchers use many other features such as frontal asymmetry, differential entropy, and indexes for attention, approach motivation, and memory. Approach emotions such as happiness are associated with left hemisphere brain activity, whereas withdrawal such as disgust emotions are associated with right hemisphere brain activity [16, 17]. The left to right alpha activity is therefore used for approach motivation. Differential entropy combined with asymmetry gives out features such as differential and rational asymmetry for EEG segments are some recent developments as forward fed features for neural networks [18, 19]. The use of machine learning and neural networks in EEG emotion classification has been on the rise due to the algorithms' ability to detect deep features and classify EEG based on emotions. We have summarized the latest studies using EEG to recognize the emotional state in Table 1

Machine learning-based emotion recognition hence performs weighted Spatio-temporal averaging of EEG signals. While several feature extraction methods were reported in the past, it is crucial to understand which methods are suited for emotion recognition and optimize the set of features for performance. Moreover, the electrodes' relative importance can help explain the significance of different regions for emotion elicitation. This could, in turn, help in optimizing the electrode locations while conducting EEG-based studies. In this study, we try to answer these questions through different feature ranking methods for decoding valence and arousal with the help of two open-source datasets: DEAP and DREAMER and our self-curated data from a portable EEG headset called as OASIS EEG dataset in the rest of the paper.

Materials and Methods

1 Datasets

1.1 DEAP

DEAP dataset [13] has 32 subjects; each subject was shown 40 music videos one min long. Participants rated each video in terms of arousal levels, valence, like/dislike, dominance, and familiarity. Data was recorded using 40 EEG electrodes placed

Table 1. Table summarising various machine learning algorithms and features used to classify emotions on various datasets, reported with accuracy

Dataset	Algo	Attributes	Accuracy	Ref
DEAP	kNN	i) Gray-Level Co-occurrence Matrix ii) Spectral Power Density	79.58 (Average)	[20]
DEAP	MLP	i) Relative Power Energy ii) Logarithmic Relative Power Energy iii) Absolute Logarithmic Relative Power Energy	63.47,69.42,63.57 (VAD)	[21]
	SVM		56.34,52.79,57.71 (VAD)	
	kNN		67.51,68.55,65.10 (VAD)	
DEAP	SVM	i) Hjorth parameters ii) Entropy iii) Power of Frequency bands iv) RASM & DASM v) Energy of frequency bands using wavelets vi) Only Fp1, Fp2, F3 & C4 electrodes	65.72 (10 fold CV) 65.92 (LOOCV) (9 emotions)	[22]
DEAP	GNB	i) Spectral power ii) Spectral power differential asymmetry	61.6, 64.7, 61.8 (AVL)	[13]
Movie Clips	kNN	Mean, standard deviation, mean of absolute first difference, mean of absolute second difference of raw signal and mean of absolute second difference of normalised signal	59.84	[23]
	MLP		63.07	
	SVM		66.51	
Video Clips	kNN	Absolute logarithmic Recoursing Energy Efficiency of alpha, beta, gamma bands decomposed using db4 wavelet function	83.26	[24]
	LDA		75.21	
Movie Clips	SVM	i) Power spectrum and wavelet decomposition of frequency bands ii) Entropy, Hurts exponent, Katz fractal dimension iii) Feature smoothening using LDS iv) Feature reduction using PCA, LDA and CFS	Best accuracy of 87.53	[23]
DEAP	2k-NN	i) Spectral power of frequency bands ii) Spectral power difference of symmetric electrodes iii) Histogram parameters of segment level probability vectors iv) Dirichlet distribution parameters	76.9,68.4,73.9,75.3 (VADL)	[25]

according to standard 10-20 montage system. The sampling frequency was 128Hz. In this analysis, we consider only 14 channels (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4) for the sake of uniformity with other two datasets.

1.2 DREAMER

DREAMER [12] dataset has 23 subjects; each subject was shown 18 videos at a sampling frequency 128Hz. Audio and visual stimuli in the form of film clips were employed to elicit emotional reactions to the participants of this study and record EEG and ECG data. After viewing each film clip, participants were asked to evaluate their emotion by reporting the felt arousal (ranging from uninterested/bored to excited/alert), valence (ranging from unpleasant/stressed to happy/elated), and dominance. Data was recorded using 14 EEG electrodes.

1.3 OASIS EEG dataset

1.3.1 Stimuli selection

The OASIS image dataset [11] consists of a total 900 images from various categories such as natural locations, people, events, inanimate objects with various valence and arousal elicitation values. Out of 900 images, 40 images were selected to cover the whole spectrum of valence and arousal ratings as shown in Fig 1.

1.3.2 Participants and device

The experiment was conducted in a closed room with the only source of light being the digital 21" Samsung 1080p monitor. Data was collected from fifteen participants of mean age 22 with ten males and five females using EMOTIV Epoc EEG headset consisting of 14 electrodes according to the 10-20 montage system at a sampling rate of 128Hz and only the EEG data corresponding to the image viewing time were segmented using markers and used for analysis.

The study was approved by the Institutional Ethics Committee of BITS, Pilani (IHEC-40/16-1). All EEG experiments/methods were performed in accordance with the relevant guidelines and regulations as per the Institutional Ethics Committee of BITS, Pilani. All participants were explained the experiment protocol and written consent for recording the EEG data for research purpose was obtained from each of the subject.

1.3.3 Protocol

The subjects were explained the meaning of valence and arousal before the start of the experiment and were seated at a distance of 80-100 cms from the monitor.

The images were shown for 5 seconds through Psychopy [26], and the participants were asked to rate valence and arousal on a scale of 1 to 10 before proceeding to the next image, as shown in Fig 2. The ratings given in the OASIS image dataset were plotted against the ratings reported by the participants in order to draw correlation between them, as shown in Fig 3.

Fig 1. Valence and arousal ratings of OASIS dataset. Valence and arousal ratings of entire OASIS image dataset (blue) and of the images selected for our experiment (red).

Fig 2. EEG Data collection protocol Experiment protocol for the collection of EEG data. 40 images from OASIS dataset were shown to elicit emotion in valence and arousal plane. After presenting each image, ratings were collected from participant.

Fig 3. Comparison of actual and self-reported valence and arousal ratings. Valence and arousal ratings reported by the participants during the EEG data collection and ratings from the OASIS image dataset.

2 Preprocessing

2.1 Referencing

The average amplitude of all electrodes for a particular time point was calculated and subtracted from the data of all electrodes. This was done for all time points across all trials.

2.2 Filtering

A Butterworth bandpass filter of 4th order was applied to filter out frequencies between 0.1Hz and 40Hz

2.3 Motion Artifact

Motion artifacts can be removed by using Pearson Coefficients [27]. The gyroscopic data(accelerometer readings) and EEG data were taken corresponding to each trial. Each of these trials of EEG data was separated into its independent sources using Independent Component Analysis (ICA) algorithm. For the independent sources obtained corresponding to a single trial, Pearson coefficients were calculated between each source signal and each axis of accelerometer data for the corresponding trial. The mean and standard deviations of Pearson coefficients were then calculated for each axis obtained from overall sources. The sources that had Pearson coefficient 2 standard deviations above mean for any one axis were high pass filtered for 3Hz using a Butterworth filter as motion artifacts exist at these frequencies. The corrected sources were then projected back into the original dimensions of the EEG data using the mixing matrix given by ICA.

2.4 Rejection and repair of bad trials

Auto Reject is an algorithm developed by Mainak et al. [28] for the rejection of bad trials in Magneto-/Electro- encephalography (M/EEG data), using a cross-validation framework to find the optimum peak to peak threshold to reject data.

- We first consider a set of candidate thresholds ϕ.
- Given a matrix of dimensions (epochs x channels x time points) by X $\in$ R N$\times$P, where N is the number of trials/epochs P is the number of features. P = Q*T, Q being the number of sensors, and T the number of time points per sensor.
- The matrix is split into K folds. Each of the K parts will be considered the training set once, and the rest of the K-1 parts become the test set.
- For each candidate threshold, i.e. for each

$$T_l \in \phi$$

we apply this candidate peak to peak threshold(ptp) to reject trials in training set known as bad trials, and the rest of the trials become the good trials in the training set.

$$ptp(X_i) = max(X_i) - min(X_i)$$

where X_i indicates a particular trial.

- A is the peak to peak threshold of each trial, G_l is the set of trials whose ptp is less than the candidate threshold being considered

$$A = \{ptp(X_i)|i \in train_k\}$$

$$G_l = \{i \in train_k|ptp(X_i) < T_l\}$$

- Then the mean amplitude of the good trials (for each sensor and their corresponding set of time points) is calculated

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

- While the median amplitude of all trials is calculated for the test set $\tilde{X}_{val_k}$
- Now the Frobenius norm is calculated for all K folds giving K errors $e_k \in E$; mean of all these errors is mapped to the corresponding candidate threshold.

$$e_{kl} = ||\overline{X}_{G_l} - \tilde{X}_{val_k}||_{Fro}$$

- The following analysis was done taking all channels into consideration at once, thus it is known as auto-reject global
- Similar process can be considered where analysis can be done for each channel independently i.e data matrix becomes(epochs x 1 x time points) known as the local auto-reject, where we get optimum thresholds for each sensor independently.
- The most optimum threshold is the one that gives the least error

$$T_* = T_{l_*} \;\; with \;\; l_* = argmin\ l \;\; \frac{1}{K}\sum_{i=1}^{K} e_{kl}$$

As bad trials were already rejected in the DEAP and DREAMER dataset, we do not perform automatic trial rejection in them.

3 Feature extraction

In this work, the following set of 36 features were extracted from the EEG signal data with the help of EEGExtract library [29] for all three datasets:

- Shannon Entropy (S.E.)
- Subband Information Quantity for Alpha [8 Hz - 12 Hz], Beta [12 Hz - 30 Hz], Delta [0.5 Hz - 4 Hz], Gamma [30 Hz - 45 Hz] and Theta[4 Hz - 8 Hz] band (S.E.A., S.E.B., S.E.D., S.E.G., S.E.T.)
- Hjorth Mobility (H.M.)

- Hjorth Complexity (H.C.)
- False Nearest Neighbour (F.N.N)
- Differential Asymmetry (D.A., D.B., D.D., D.G., D.T.)
- Rational Asymmetry (R.A., R.B., R.D., R.G., R.T.)
- Median Frequency (M.F.)
- Band Power (B.P.A., B.P.B., B.P.D., B.P.G., B.P.T.)
- Standard Deviation (S.D.)
- Diffuse Slowing (D.S.)
- Spikes (S.K.)
- Sharp spike (S.S.N.)
- Delta Burst after Spike (D.B.A.S.)
- Number of Bursts (N.B.)
- Burst length mean and standard deviation (B.L.M., B.L.S.)
- Number of Suppressions (N.S.)
- Suppression length mean and standard deviation (S.L.M., S.L.S.)

These features were extracted with a 1s sliding window and no overlap. The extracted features can be categorized into two different groups based on the ability to measure the complexity and continuity of the EEG signal. The reader is encouraged to refer to the work done by Ghassemi et al. [30] for an in-depth discussion of these features.

3.1 Complexity Features

Complexity features represent the degree of randomness and irregularity associated with the EEG signal. Different features in the form of entropy and complexity measures were extracted to gauge the information content of non-linear and non-stationary EEG signal data.

3.1.1 Shannon Entropy

Shannon entropy [31] is a measure of uncertainty (or variability) associated with a random variable. Let X be a set of finite discrete random variables $X = \{x_1, x_2, ..., x_m\}, x_i \in R^d$, Shannon entropy, $H(X)$, is defined as

$$H(X) = -c \sum_{i=0}^{m} p\left(x_i\right) \ln p\left(x_i\right) \tag{1}$$

where c is a positive constant and p(x_i) is the probability of $(x_i) \in$ X such that:

$$\sum_{i=0}^{m} p\left(x_i\right) = 1 \tag{2}$$

Higher values of entropy are indicative of high complexity and less predictability in the system. [32]

3.1.2 Subband Information Quantity

Subband Information Quantity (SIQ) refers to the entropy of the decomposed EEG wavelet signal for each of the five frequency bands. [33]. In our analysis, the EEG signal was decomposed using a butter-worth filter of order 7 followed by an FIR/IIR filter. Shannon entropy ($H(X)$) of this resultant wave signal is the desired SIQ of a particular frequency band. Due to its tracking capability for dynamic amplitude change and frequency component change, this feature has been used to measure the information contained in the brain [34].

3.1.3 Hjorth Parameters

Hjorth Parameters indicate time-domain statistical properties introduced by Bo Hjorth in 1970 [35]. Variance-based calculation of Hjorth parameters incurs a low computational cost which makes them appropriate for performing EEG signal analysis. We make use of complexity and mobility [36] parameters in our analysis. Horjth mobility signifies the mean frequency or the proportion of standard deviation of the power spectrum. It is defined as :

$$\text{Hjorth}_{\text{Mobility}} = \sqrt{\frac{\text{var}\left(\frac{dx(t)}{dt}\right)}{\text{var}(x(t))}} \tag{3}$$

where var(.) denotes the variance operator and $x(t)$ denotes the EEG time-series signal.

Hjorth complexity signifies the change in frequency. This parameter has been used to get a measure of similarity of the signal to a sine wave. It is defined as:-

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

3.1.4 False Nearest Neighbour

False Nearest Neighbour is a measure of signal continuity and smoothness. It is used to quantify the deterministic content in the EEG time series data without assuming chaos [37, 38].

3.1.5 Asymmetry features

We incorporate Differential Entropy (DE) [39] in our analysis to construct two features for each of the five frequency bands, namely, Differential Asymmetry (DASM)and Rational Asymmetry (RASM). Mathematically, DE ($h(X)$) is defined as :

$$\begin{aligned} h(X) = -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{(x-\mu)^2}{2\sigma^2} \log \frac{1}{\sqrt{2\pi\sigma^2}} \\ \exp \frac{(x-\mu)^2}{2\sigma^2} dx = \frac{1}{2} \log 2\pi e \sigma^2 \end{aligned} \tag{5}$$

where X follows the Gauss distribution $N(\mu, \sigma^2)$, x is a variable and π and exp are constant.

Differential Asymmetry(or DASM) [18] for each frequency band were calculated as the difference of differential entropy of each of 7 pairs of hemispheric asymmetry electrodes.

$$DASM = h\left(X_i^{\text{left}}\right) - h\left(X_i^{\text{right}}\right) \tag{6}$$

Rational Asymmetry(or RASM) [18] for each frequency band were calculated as the ratio of differential entropy between each of 7 pairs of hemispheric asymmetry electrodes.

$$RASM = h\left(X_i^{\text{left}}\right)/h\left(X_i^{\text{right}}\right) \tag{7}$$

3.2 Continuity Features

Continuity features signify the clinically relevant signal characteristics of EEG signals [30, 40]. These features have been acclaimed to serve as qualitative descriptors of states of the human brain and hence, are important towards the process of emotion recognition.

3.2.1 Median Frequency

Median Frequency refers to the 50% quantile or median of the power spectrum distribution. Median Frequency has been studied extensively in the past due to its observed correlation with awareness [41] and its ability to predict imminent arousal [42]. It is a frequency domain or spectral domain feature.

3.2.2 Band Power

Band power refers to the average power of the signal in a specific frequency band. The band powers of delta, theta, alpha, beta , and gamma were used as spectral features. To calculate band power, initially, a butter-worth filter of order 7 was applied on the EEG signal. IIR/FIR filter was applied further on the EEG signal in order to separate out signal data corresponding to a specific frequency band. Average of the power spectral density was calculated using a periodogram of the resulting signal. Signal Processing sub module (scipy.signal) of SciPy library [43] in python was used to compute the band power feature.

3.2.3 Standard Deviation

Standard Deviation has proved to be an important time-domain feature in the past experiments [44, 45]. Mathematically, it is defined as the square root of variance of EEG signal segment.

3.2.4 Diffuse Slowing

Previous studies [46] have shown that diffuse slowing is correlated with impairment in awareness, concentration , and memory and hence, it is an importance feature for estimation of valence/arousal levels from EEG signal data.

3.2.5 Spikes

Spikes [40] refer to the peaks in the EEG signal up to a threshold, fixed at mean + 3 standard deviation. The number of spikes was computed by finding local minima or peaks in EEG signal over 7 samples using scipy.signal.find_peaks method from SciPy library [43].

3.2.6 Delta Burst after spike

The change in delta activity after and before a spike computed epoch wise by adding mean of 7 elements of delta band before and after the spike, used as a continuity feature.

3.2.7 Sharp spike

Sharp spikes refer to spikes which last less than 70ms and is a clinically important feature in study of electroencephalography [40].

3.2.8 Number of Bursts

The number of amplitude bursts(or simply number of bursts) constitutes a significant feature [40].

3.2.9 Burst length mean and standard deviation

Statistical properties of the bursts, mean μ and standard deviation σ of the burst lengths, have been used as continuity features.

3.2.10 Number of Suppressions

Burst Suppression refers to a pattern where high voltage activity is followed by an inactive period and is generally a characteristic feature of deep anaesthesia [47]. We use the number of contiguous segments with amplitude suppressions as a continuity feature with a threshold fixed at 10μ [29].

3.2.11 Suppression length mean and standard deviation

Statistical properties like mean μ and standard deviation σ of the suppression lengths, used as a continuity feature.

4 Feature selection

Selecting the correct predictor variables or feature vectors can improve the learning process in any machine learning pipeline. In this work, initially, sklearn's [48] VarianceThreshold feature selection method was used to remove zero-variance or constant features from the set of 36 extracted EEG features. Next, a subset of 25 features common to all 3 datasets (DREAMER, DEAP, and OASIS) was selected after applying the VarianceThreshold method for further analysis. This was done to validate our approach and compare the effect of feature selection techniques on a common set of features. The set of 11 features (S.E., F.N.N., D.S., S.K., D.B.A.S., N.B., B.L.M., B.L.S., N.S., S.L.M., S.L.S.) were excluded from further analysis. Our study compared three feature selection techniques that differ in the way they automatically rank the features.

4.1 Method A: Regressor approach

In this approach, random forest regressor was used to compute RMSE and rank electrodes and features. Random forest regressor is used as the regressor of choice mainly because it avoids overfitting in small datasets , and does not need feature scaling.

The random forest regressor model from the cuML library [49] was trained separately for 80% of each of the EEG datasets. For ranking electrodes, all the 25 feature vectors belonging to a single electrode were considered for training the model and accuracy was computed for each electrode. Hence all 14 electrodes were ranked based on RMSE obtained on the test set of each dataset. The same procedure was repeated for ranking features by using one feature and signals from all the electrodes.

4.2 Method B: F-Value-Based Feature Selection (SelectKBest approach)

SelectkBest [48] is a filter-based, univariate feature selection method intended to select and retain first k-best features based on the scores produced by univariate statistical tests. In our work, f_regression was used as the scoring function since valence and arousal are continuous numeric target variables. It uses Pearson correlation coefficient as defined in Eq 8 to compute the correlation between each feature vector in the input matrix, X and target variable, y as follows:

$$\rho_i = \frac{(X[:,i] - \text{mean}(X[:,i])) * (y - \text{mean}(y))}{\text{std}(X[:,i]) * \text{std}(y)} \tag{8}$$

The corresponding F-value is then calculated as:

$$F_i = \frac{\rho_i^2}{1 - \rho_i^2} * (n - 2) \tag{9}$$

where n is the number of the samples.

SelectkBest method then ranks the feature vectors based on F-scores returned by f_regression method. Higher scores correspond to better features.

4.3 Method C: Random Forest Importances approach

Random forest is an ensemble method that combines multiple individual decision trees. Every decision tree comprises internal nodes (non-leaf nodes) and leaf nodes. In the internal nodes of these decision trees, the selected feature is used individually to split the data. In this work, features were ranked using feature importances based on impurity reduction averaged over all decision trees constructed by random forest regressor [48]. Higher impurity scores correspond to more important features [50, 51].

5 Results

5.1 Electrodes ranking and selection

The electrodes were ranked for the three datasets, using the three feature selection techniques, as discussed in Section 4 and the ranks are tabulated for valence (Table 2) and arousal (Table 3) labels. To produce a ranking for Top N electrodes taken together, feature data for top i electrodes were initially considered. The resultant matrix was split in the ratio 80:20 for training and evaluation of the random forest regressor model. The procedure was repeated until all the 14 electrodes were taken into account. The RMSE values for the same are shown in Figs 4 (a), 4 (b), and 4 (c). It should be noted that, unlike feature analysis, data corresponding to 5 features each of DASM and RASM was excluded from the Top N electrode-wise RMSE study since these features are constructed using pairs of opposite electrodes.

5.2 Features ranking and selection

Each extracted feature was used to generate its corresponding feature matrix of shape (nbChannels, nbSegments). These feature matrices were then ranked using three feature selection techniques. Initially, a feature matrix for the best feature was generated. The ranks were tabulated for valence(Table 4) and arousal(Table 5) labels respectively. This data was split into 80:20 train-test data, the training data was used to perform regression with Random Forest Regressor and predicted values on test data were

Table 2. Electrode Ranking for valence label based on feature selection method. Method A: Regressor approach, Method B: SelectKBest approach, Method C: RandomForest Importances approach

Electrode	DREAMER			DEAP			OASIS		
	A	B	C	A	B	C	A	B	C
AF3	6	13	10	4	8	6	11	6	6
F7	10	12	8	2	1	4	6	1	7
F3	11	10	7	11	11	12	9	5	10
FC5	2	9	4	8	14	11	4	7	2
T7	4	7	1	13	13	9	14	2	1
P7	1	2	2	14	3	14	13	12	8
O1	8	8	11	3	7	13	8	11	3
O2	12	3	13	12	9	8	5	13	5
P8	9	6	9	6	5	10	12	14	4
T8	14	1	5	10	12	7	7	3	12
FC6	5	4	12	7	4	3	1	10	14
F4	3	14	3	1	6	1	2	9	9
F8	13	5	14	9	2	5	3	4	13
AF4	7	11	6	5	10	2	10	8	11

Table 3. Electrode Ranking for arousal label based on feature selection method. Method A: Regressor approach, Method B: SelectKBest approach, Method C: RandomForest Importances approach

Electrode	DREAMER			DEAP			OASIS		
	A	B	C	A	B	C	A	B	C
AF3	5	10	10	6	10	5	11	7	9
F7	4	14	3	2	1	1	10	1	12
F3	2	11	2	9	7	10	8	5	2
FC5	3	5	5	13	14	3	5	10	6
T7	9	1	7	1	9	2	14	3	3
P7	1	4	9	4	5	13	13	12	7
O1	7	12	13	3	13	14	6	11	4
O2	12	8	8	5	6	8	3	14	10
P8	10	7	6	8	4	7	12	13	5
T8	13	2	4	12	11	6	2	2	1
FC6	8	6	12	10	3	9	1	9	11
F4	6	13	11	14	8	11	4	6	8
F8	14	3	14	11	2	12	7	4	13
AF4	11	9	1	7	12	4	9	8	14

Table 4. Feature Ranking for valence label based on feature selection method. Method A: Regressor approach, Method B: SelectKBest approach, Method C: RandomForest Importances approach

Feature	DREAMER			DEAP			OASIS		
	A	B	C	A	B	C	A	B	C
B.P.A.	8	5	9	12	17	21	8	15	13
B.P.B.	6	8	7	6	7	9	4	13	10
B.P.D.	23	22	19	18	22	19	6	23	3
B.P.G.	7	18	6	1	13	2	12	20	12
B.P.T.	4	20	5	15	18	24	10	17	11
D.A.	17	11	18	21	10	16	15	9	20
D.B.	10	10	14	9	4	6	23	11	21
D.D.	25	25	22	24	16	10	14	21	14
D.G.	12	16	11	5	6	4	19	18	18
D.T.	15	14	13	17	9	18	21	16	15
H.C.	2	4	2	11	20	17	3	4	4
H.M.	1	3	1	10	19	7	1	2	2
M.F.	19	12	21	23	25	11	11	5	7
R.A.	16	13	17	22	21	14	18	12	23
R.B.	11	9	16	8	2	5	25	10	22
R.D.	24	24	23	25	24	13	20	22	16
R.G.	13	17	12	4	1	3	17	19	19
R.T.	14	15	15	19	11	15	24	14	17
S.E.A.	20	7	10	13	12	22	9	24	6
S.E.B.	9	2	8	3	5	8	22	3	5
S.E.D.	18	19	24	16	15	20	16	7	24
S.E.G.	5	1	4	2	8	1	7	1	9
S.E.T.	22	21	20	14	14	23	13	8	8
S.S.N.	21	23	25	20	23	25	5	25	25
S.D.	3	6	3	7	3	12	2	6	1

compared with actual test labels, and RMSE was computed. In the second run, feature matrices of best and second-best features were combined, data was split into train and test data, model was trained, and predictions made by model on test data were used to compute RMSE. This procedure was followed until all the features are taken into account. The RMSE values for the feature analysis procedure, as described above, are shown in Fig 4.

Fig 4. Model evaluation for feature and electrode selection. The random forest regressor was trained on the training set (80%) corresponding to top N electrodes (ranked using Regressor approach (Method A), F-Value-Based Feature Selection (Method B), and Random Forest Importances approach (Method C)) and RMSE was computed on the test set (20%) for valence and arousal label on DREAMER (a), DEAP (b), OASIS EEG dataset (c). A similar analysis was performed for top N features for DREAMER (d), DEAP (e), and OASIS EEG (f) dataset.

The identification of an optimum set of electrodes and features is a critical step. By optimum set, we imply the minimum number of electrodes and features that produce minimum RMSE during model evaluation, as shown in Fig 4. We can observe a general decline in RMSE value when the number of electrodes under consideration is increased. DREAMER dataset shows a much greater and smoother convergence than the other

Table 5. Feature Ranking for arousal label based on feature selection method. Method A: Regressor approach, Method B: SelectKBest approach, Method C: RandomForest Importances approach

Feature	DREAMER			DEAP			OASIS		
	A	B	C	A	B	C	A	B	C
B.P.A.	7	7	8	15	18	22	10	25	16
B.P.B.	5	9	7	3	8	8	5	11	5
B.P.D.	23	22	20	21	23	17	21	12	8
B.P.G.	4	21	5	2	3	4	11	5	9
B.P.T.	6	20	6	11	19	24	6	24	11
D.A.	17	4	18	20	12	18	17	16	20
D.B.	12	12	12	8	4	7	8	21	13
D.D.	25	24	23	24	16	12	25	20	19
D.G.	13	16	13	6	5	3	18	14	21
D.T.	14	13	17	16	10	16	22	23	15
H.C.	2	2	3	14	20	14	2	4	1
H.M.	1	6	1	10	17	10	1	1	3
M.F.	18	14	22	25	24	11	15	7	12
R.A.	15	5	19	19	21	19	19	15	17
R.B.	11	11	15	9	1	6	9	19	14
R.D.	24	23	24	23	25	15	24	18	23
R.G.	10	17	14	7	2	2	16	13	22
R.T.	16	15	16	17	11	13	23	22	18
S.E.A.	20	10	11	13	13	20	14	10	6
S.E.B.	8	3	9	4	6	5	7	3	7
S.E.D.	19	18	10	18	15	21	20	8	24
S.E.G.	9	1	2	1	9	1	4	2	2
S.E.T.	21	19	21	12	14	23	13	9	10
S.S.N.	22	25	25	22	22	25	12	17	25
S.D.	3	8	4	5	7	9	3	6	4

Table 6. RMSE values for valence and arousal label on the test set(20%) of DEAP, DREAMER and OASIS dataset for optimum set of features obtained from different feature selection techniques

Dataset	Method	Valence		Arousal	
		N	RMSE	N	RMSE
DREAMER	A	6	**0.883**	9	0.746
	B	11	0.905	9	0.749
	C	7	0.884	6	**0.742**
DEAP	A	6	1.895	8	1.765
	B	8	1.902	9	1.769
	C	8	1.898	10	1.769
OASIS	A	3	2.677	1	2.293
	B	2	2.728	1	2.300
	C	2	2.698	4	2.314

Table 7. RMSE values for valence and arousal label on the test set (20%) of DEAP, DREAMER and OASIS dataset for optimum set of electrodes obtained from different feature selection techniques

Dataset	Method	Valence		Arousal	
		N	RMSE	N	RMSE
DREAMER	A	14	**0.911**	13	**0.757**
	B	14	0.914	14	0.759
	C	14	0.913	13	**0.757**
DEAP	A	14	1.936	14	1.805
	B	14	1.938	14	1.806
	C	14	1.936	14	1.805
OASIS	A	8	2.757	5	2.386
	B	7	2.765	14	2.417
	C	14	2.771	11	2.395

two datasets because more training data was available for training the model. In general, the minimum RMSE is observed when all 14 electrodes are selected. OASIS dataset can be excluded from this inference since it contained only 15 participants at the time of the experiment.

Figs 4 (d), 4 (e), and 4 (f) reveal a general pattern about optimal set of features. On increasing the number of features in consideration, initially, there is a steady drop in the RMSE values followed by a gradual increase after a certain critical point in the graph. Hence, a minima can be observed in the graph. As discussed above, the OASIS dataset can be ruled out of this generalization. The minimum RMSE values and the corresponding number of best features and electrodes selected are summarized in Table 6 and Table 7 respectively.

Incremental Learning

Best Features as given by the feature analysis described before were used to generate a feature matrix for valence and arousal correspondingly for each dataset. The feature matrix generated was then used to train a support vector regressor as part of the incremental learning algorithm.

Incremental learning was performed based on the collection of subject data. Initially, the first subject data was taken, their trial order shuffled and then split using 80:20 train test size, the model was trained using train split, and predictions were made for

test data, next 2nd subject data was taken together with the 1st subject, trial order shuffled, again a train-test split taken and the random forest regressor model was trained using the train split, and predictions were made for the test split. This procedure was repeated until data of all the subjects were used for RMSE computation. RMSE values for calculated for each step of training, i.e. training data consisted of subject1 data, then the combination of subject 1, 2 data, then the combination of subject 1, 2, 3 data, etc. The plots generated for RMSE values for the individual steps of training show a general decreasing trend as evident from Fig 5.

Fig 5. Incremental learning performance. Valence and arousal RMSE readings obtained with incremental learning for DREAMER (a), DEAP (b) and OASIS (c) EEG dataset using random forest regressor (rfr).

Leave-one-subject-out cross-validation

Subject generalization is a crucial problem to solve in identifying patterns in EEG signals. To prevent over-fitting and avoiding subject-dependent patterns. We train the model with data of all the subjects except a single subject and evaluate the model on this remaining subject. Hence, the model is evaluated for each subject to identify subject bias and prevent any over-fitting. Also, when building a machine learning model, it is a standard practice to validate the results by leaving aside a portion of data as the test set. In this work, we have used the leave-one-subject-out cross-validation technique due to its robustness for validating results for datasets built at the participant level. Leave-one-subject-out cross-validation is a type of k-fold cross-validation technique, where the number of folds, k, is equal to the number of participants in a dataset. The cross-validated RMSE values for the three datasets for all the participants are plotted in Figs 6 and 7.

Fig 6. Subject wise performance analysis for valence label.
Leave-one-subject-out cross-validation performance analysis for valence label for (a) DREAMER (b) DEAP (c) OASIS datasets. In this cross-validation technique, one subject was chosen as the test subject, and the models were trained on the data of the remaining subjects.

Fig 7. Subject wise performance analysis for arousal label.
Leave-one-subject-out cross validation performance analysis for arousal label for (a) DREAMER (b) DEAP (c) OASIS datasets. In this cross-validation technique, one subject was chosen as the test subject and the models were trained on the data of remaining subjects.

Mean and standard deviation of RMSE values for valence and arousal label after cross validation have been summarized in Table 8. The best RMSE values lie within the standard deviation range respect to the leave-one-subject-out cross validation results and hence, inferences drawn from them can be validated.

Discussion and future scope

The relation between the performance (RMSE) and the number of participants is critical for any study concerning emotion recognition from the EEG dataset. As in Fig 5, we observe an improvement in performance with an increasing number of

Table 8. Mean and Standard Deviation (Std. Dev.) of RMSE values for Valence and Arousal Label Data after Leave-one-subject-out-cross-validation

Dataset	Label	Mean	Std. Dev.
DREAMER	Valence	1.336	0.127
	Arousal	1.108	0.176
DEAP	Valence	2.102	0.415
	Arousal	2.012	0.541
OASIS	Valence	2.880	0.510
	Arousal	2.630	0.777

participants. This explains the fact that the machine learning algorithm needs data of more participants for generalization. Interestingly, the performance degrades for the OASIS dataset (Fig 5 (c)) while increasing the number of participants. This could be explained as the model overfits when trained with data from a few subjects. This could be verified from the fact that the degradation in performance is only up to a certain number of subjects as in Fig 5 (a). Hence, with data from more participants in the OASIS EEG dataset, we should eventually observe an increase in performance.

As in tables 2 and 3, 9 rankings were obtained as a result of 3 ranking methods and 3 datasets. For the valence labels, out of the top 25 % electrodes, 55.55 % were in the frontal regions (F3, F4, F7, F8, AF3, AF4, FC5, FC6), 18.51% in the temporal regions (T8, T7) , 14.81 % the parietal regions (P7, P8), and 11.11 % in the occipital regions (O1, O2). For the top 50% electrodes, 61.90 % were in the frontal regions, 15.87 % in the temporal regions, 12.69 % in the parietal regions, and 9.52 % in the occipital regions. For top 75 % electrodes, 61.11 % were in the frontal regions , 13.33 % in the temporal regions, 13.33 % in the parietal regions , and 12.22 % in the occipital regions.

For the arousal labels, out of the top 25 % electrodes, 55.55 % were in the frontal regions, 33.33 % in the temporal regions, 3.7 % in the parietal regions, and 7.40 % in the occipital regions. For top 50% electrodes, 53.96 % were in the the frontal regions, 19.04 % in the temporal regions, 15.87 % in the parietal regions , and 11.11 % in the occipital regions. For top 75 % electrodes, 57.77 % were in the frontal regions , 15.55 % in the temporal regions, 14.44 % in the parietal regions , and 12.22 % in the occipital regions. Therefore, the frontal region was the most significant brain region for recognizing valence and arousal, followed by temporal, parietal, and occipital.

The optimum set of features for the DREAMER dataset were observed to be (H.M., H.C., S.D., B.P.T., S.E.G., B.P.B.) using method A and (H.M., S.E.G., H.C., S.D., B.P.G., B.P.T.) using method C for valence and arousal respectively. The minimum RMSE values obtained using these optimal features on the DREAMER dataset were 0.883 and 0.742 for valence and arousal dimensions, respectively, as evident from Table 6. Therefore these features were critical for recognizing emotion states and would be used in future studies to evaluate classifiers like Artificial Neural Networks, ensembles etc.

As in tables 4 and 5, band power and subband information quantity features for gamma and beta frequency bands performed better in the estimation of both valence and arousal than other frequency bands. Hence the gamma and beta frequency bands are the most critical for emotion recognition.

It can be inferred from tables 4 that H.M. was mostly ranked among the top 3 features for predicting valence labels and arousal labels. Similarly, H.C. was ranked among the top 4 features. This inference is consistent with the previous studies that claim the importance of time-domain Hjorth parameters in accurate EEG classfication tasks [52, 53].

In past, statistical properties like standard deviation derived from the reconstruction of EEG signals have been claimed to be significant descriptor of the signal and provide

a supporting evidence to the results obtained in this study [54, 55]. It was observed that S.D. was ranked among the top 5 ranks in general. It was even ranked as the top feature for valence label for the OASIS dataset.

Table 6 indicates that the minimum RMSE values obtained on the test set (20%) using optimum set of features were 0.883 and 0.742 on DREAMER dataset, 1.895 and 1.765 on DEAP dataset and 2.677 and 2.293 on OASIS dataset for valence and arousal respectively. For leave-one-subject-out cross-validation, we achieved the best RMSE of 1.030, 0.750 on DREAMER, 1.170, 1.163 on DEAP and 1.867, 1.185 on OASIS dataset for valence and arousal respectively as shown in Figs 6 and 7.

We performed the evaluation of EEG activity induced by videos (DEAP), and static images (DREAMER & OASIS), but not on audio stimulus. The OASIS dataset collection was limited to 15 participants due to the Covid 19 pandemic. In future studies, we wish to collect the data for at least 40 participants to draw stronger inferences. Future work would also include analysis of end to end neural networks and transfer learning for the purpose of emotion recognition.

Author contributions

N.G. conceptualized the research. P.N. performed data collection. N.G., R.G., A.A. and V.A. performed the experiments and analyzed the data. V.B. supervised the study. All authors approved and contributed to manuscript.

Acknowledgments

This work was funded by the Department of Science and technology, Government of India, vide Reference No:SR/CSI/50/2014(G)through the Cognitive Science Research Initiative (CSRI)

References

1. Ekman P. Universals and Cultural Differences in Facial Expressions of Emotion BT - Nebraska Symposium on Motivation; 1972. Available from: `papers3://publication/uuid/FDC5E29A-0E28-4DDF-B1A4-F53FEE0B4F70`.

2. Russell JA. A circumplex model of affect. Journal of Personality and Social Psychology. 1980;39(6):1161–1178. doi:10.1037/h0077714.

3. Moors A, De Houwer J, Hermans D, Wanmaker S, van Schie K, Van Harmelen AL, et al. Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. Behavior Research Methods. 2013;45(1):169–177. doi:10.3758/s13428-012-0243-8.

4. Mohammad SM. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers). 2018;1:174–184. doi:10.18653/v1/p18-1017.

5. Leite J, Carvalho S, Galdo-Alvarez S, Alves J, Sampaio A, Gonçalves ÓF. Affective picture modulation: Valence, arousal, attention allocation and motivational significance. International Journal of Psychophysiology. 2012;83(3):375–381. doi:10.1016/j.ijpsycho.2011.12.005.

6. Lane RD, Chua PML, Dolan RJ. Common effects of emotional valence, arousal and attention on neural activation during visual processing of pictures. Neuropsychologia. 1999;37(9):989–997. doi:10.1016/S0028-3932(99)00017-2.

7. Gerber AJ, Posner J, Gorman D, Colibazzi T, Yu S, Wang Z, et al. An affective circumplex model of neural systems subserving valence, arousal, and cognitive overlay during the appraisal of emotional faces. Neuropsychologia. 2008;46(8):2129–2139. doi:10.1016/j.neuropsychologia.2008.02.032.

8. Warriner AB, Kuperman V, Brysbaert M. Norms of valence, arousal, and dominance for 13,915 English lemmas. Behavior Research Methods. 2013;45(4):1191–1207. doi:10.3758/s13428-012-0314-x.

9. Eerola T, Vuoskoski JK. A comparison of the discrete and dimensional models of emotion in music. Psychology of Music. 2011;39(1):18–49. doi:10.1177/0305735610362821.

10. Lang P. International Affective Picture System (IAPS) : Technical Manual and Affective Ratings; 1995.

11. Kurdi B, Lozano S, Banaji MR. Introducing the Open Affective Standardized Image Set (OASIS). Behavior Research Methods. 2016;49(2):457–470. doi:10.3758/s13428-016-0715-3.

12. Katsigiannis S, Ramzan N. DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals from Wireless Low-cost Off-the-Shelf Devices. IEEE Journal of Biomedical and Health Informatics. 2018;22(1):98–107. doi:10.1109/JBHI.2017.2688239.

13. Koelstra S, Mühl C, Soleymani M, Lee JS, Yazdani A, Ebrahimi T, et al. DEAP: A database for emotion analysis; Using physiological signals. IEEE Transactions on Affective Computing. 2012;3(1):18–31. doi:10.1109/T-AFFC.2011.15.

14. Ko LW, Su CH, Yang MH, Liu SY, Su TP. A pilot study on essential oil aroma stimulation for enhancing slow-wave EEG in sleeping brain. Scientific Reports. 2021;11(1):1–11. doi:10.1038/s41598-020-80171-x.

15. Moss M, Cook J, Wesnes K, Duckett P. Aromas of rosemary and lavender essential oils differentially affect cognition and mood in healthy adults. International Journal of Neuroscience. 2003;113(1):15–38. doi:10.1080/00207450390161903.

16. Coan JA, Allen JJB, Harmon-Jones E. Voluntary facial expression and hemispheric asymmetry over the frontal cortex. Psychophysiology. 2001;38(6):912–925. doi:10.1111/1469-8986.3860912.

17. Davidson RJ, Ekman P, Saron CD, Senulis JA, Friesen WV. Approach-Withdrawal and Cerebral Asymmetry: Emotional Expression and Brain Physiology I. Journal of Personality and Social Psychology. 1990;58(2):330–341. doi:10.1037/0022-3514.58.2.330.

18. Duan RN, Zhu JY, Lu BL. Differential entropy feature for EEG-based emotion classification. In: 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER). IEEE; 2013. p. 81–84.

19. Torres P EP, Torres EA, Hernández-Álvarez M, Yoo SG. EEG-based BCI emotion recognition: A survey. Sensors (Switzerland). 2020;20(18):1–36. doi:10.3390/s20185083.

20. Jadhav N, Manthalkar R, Joshi Y. Electroencephalography-based emotion recognition using gray-level co-occurrence matrix features. Advances in Intelligent Systems and Computing. 2017;459 AISC:335–343. doi:10.1007/978-981-10-2104-6$_3$0.

21. Verma GK, Tiwary US. Affect representation and recognition in 3D continuous valence–arousal–dominance space. Multimedia Tools and Applications. 2017;76(2). doi:10.1007/s11042-015-3119-y.

22. Khateeb M, Anwar S, Alnowami M. Multi-Domain Feature Fusion for Emotion Classification Using DEAP Dataset. IEEE Access. 2021;9:12134–12142. doi:10.1109/ACCESS.2021.3051281.

23. Wang X, Nie D, Lu BL. EEG-Based Emotion Recognition Using Frequency Domain Features and Support Vector Machines. In: ICONIP; 2011.

24. M M, Ramachandran N, Sazali Y. Classification of human emotion from EEG using discrete wavelet transform. J Biomedical Science and Engineering. 2010;334054:390–396. doi:10.4236/jbise.2010.34054.

25. Wang XW, Nie D, Lu BL. Emotional state classification from EEG data using machine learning approach. Neurocomputing. 2014;129:94–106. doi:10.1016/j.neucom.2013.06.046.

26. Peirce J, Gray J, Simpson S, MacAskill M, Höchenberger R, Sogo H, et al. PsychoPy2: Experiments in behavior made easy. Behavior Research Methods. 2019;51:195 – 203.

27. Onikura K, Katayama Y, Iramina K. Evaluation of a Method of Removing Head Movement Artifact from EEG by Independent Component Analysis and Filtering. Advanced Biomedical Engineering. 2015;4(0):67–72. doi:10.14326/abe.4.67.

28. Jas M, Engemann DA, Bekhti Y, Raimondo F, Gramfort A. Autoreject: Automated artifact rejection for MEG and EEG data. NeuroImage. 2017;159:417–429. doi:10.1016/j.neuroimage.2017.06.030.

29. Saba-Sadiya S, Chantland E, Alhanai T, Liu T, Ghassemi MM. Unsupervised EEG Artifact Detection and Correction. Frontiers in Digital Health. 2020;2:57.

30. Ghassemi MM. Life after death: techniques for the prognostication of coma outcomes after cardiac arrest. Massachusetts Institute of Technology; 2018.

31. Shannon CE. A mathematical theory of communication. The Bell System Technical Journal. 1948;27(3):379–423. doi:10.1002/j.1538-7305.1948.tb01338.x.

32. Phung DQ, Tran D, Ma W, Nguyen P, Pham T. Using Shannon Entropy as EEG Signal Feature for Fast Person Identification. In: ESANN. vol. 4; 2014. p. 413–418.

33. Jia X, Koenig MA, Nickl R, Zhen G, Thakor NV, Geocadin RG. Early electrophysiologic markers predict functional outcome associated with temperature manipulation after cardiac arrest in rats. Critical care medicine. 2008;36(6):1909.

34. Shin HC, Tong S, Yamashita S, Jia X, Geocadin G, Thakor V. Quantitative EEG and effect of hypothermia on brain recovery after cardiac arrest. IEEE Transactions on Biomedical Engineering. 2006;53(6):1016–1023. doi:10.1109/TBME.2006.873394.

35. Hjorth B. EEG analysis based on time domain properties. Electroencephalography and clinical neurophysiology. 1970;29(3):306–310.

36. Das K, Pachori R. Schizophrenia Detection Technique Using Multivariate Iterative Filtering and Multichannel EEG Signals. Biomedical Signal Processing and Control. 2021;67. doi:10.1016/j.bspc.2021.102525.

37. Kennel MB, Brown R, Abarbanel HD. Determining embedding dimension for phase-space reconstruction using a geometrical construction. Physical review A. 1992;45(6):3403.

38. Hegger R, Kantz H. Improved false nearest neighbor method to detect determinism in time series data. Physical Review E. 1999;60(4):4970.

39. Zheng WL, Dong BN, Lu BL. Multimodal emotion recognition using EEG and eye tracking data. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE; 2014. p. 5040–5043.

40. Hirsch L, LaRoche S, Gaspard N, Gerard E, Svoronos A, Herman S, et al. American clinical neurophysiology society's standardized critical care EEG terminology: 2012 version. Journal of clinical neurophysiology. 2013;30(1):1–27.

41. Schwilden H. Use of the median EEG frequency and pharmacokinetics in determining depth of anaesthesia. Baillière's clinical anaesthesiology. 1989;3(3):603–621.

42. Drummond J, Brann C, Perkins D, Wolfe D. A comparison of median frequency, spectral edge frequency, a frequency band power ratio, total power, and dominance shift in the determination of depth of anesthesia. Acta Anaesthesiologica Scandinavica. 1991;35(8):693–699.

43. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods. 2020;17:261–272. doi:10.1038/s41592-019-0686-2.

44. Amin HU, Mumtaz W, Subhani AR, Saad MNM, Malik AS. Classification of EEG signals based on pattern recognition approach. Frontiers in computational neuroscience. 2017;11:103.

45. Panat A, Patil A, Deshmukh G. Feature Extraction of EEG Signals in Different Emotional States. In: IRAJ conference; 2014.

46. Boutros NN. Diffuse electroencephalogram slowing in psychiatric patients: a preliminary report. Journal of Psychiatry and Neuroscience. 1996;21(4):259.

47. Ching S, Purdon PL, Vijayan S, Kopell NJ, Brown EN. A neurophysiological–metabolic model for burst suppression. Proceedings of the National Academy of Sciences. 2012;109(8):3095–3100. doi:10.1073/pnas.1121461109.

48. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 2011;12:2825–2830.

49. Raschka S, Patterson J, Nolet C. Machine Learning in Python: Main developments and technology trends in data science, machine learning, and artificial intelligence. arXiv preprint arXiv:200204803. 2020;.

50. Breiman L. Random forests. Machine learning. 2001;45(1):5–32.

51. Gromping U. Variable importance assessment in regression: linear regression versus random forest. The American Statistician. 2009;63(4):308–319.

52. Türk Ö, Şeker M, Akpolat V, Özerdem MS. Classification of mental task EEG records using Hjorth parameters. In: 2017 25th Signal Processing and Communications Applications Conference (SIU). IEEE; 2017. p. 1–4.

53. Cecchin T, Ranta R, Koessler L, Caspary O, Vespignani H, Maillard L. Seizure lateralization in scalp EEG using Hjorth parameters. Clinical neurophysiology. 2010;121(3):290–300.

54. Malini A, Vimala V. An epileptic seizure classifier using EEG signal. In: 2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16). IEEE; 2016. p. 1–4.

55. Panda R, Khobragade PS, Jambhule PD, Jengthe SN, Pal PR, Gandhi TK. Classification of EEG signal using wavelet transform and support vector machine for epileptic seizure diction. In: 2010 International Conference on Systems in Medicine and Biology; 2010. p. 405–408.

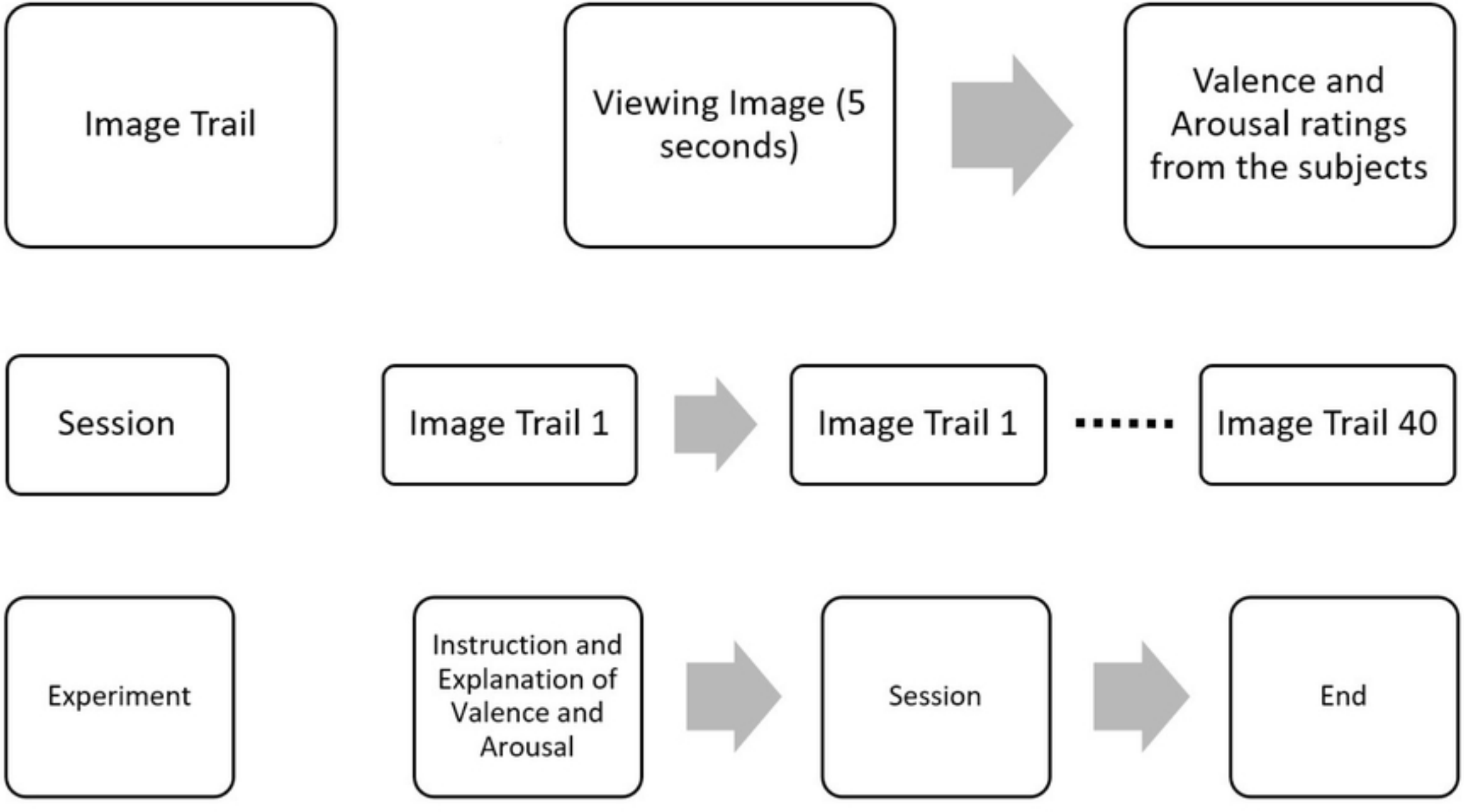
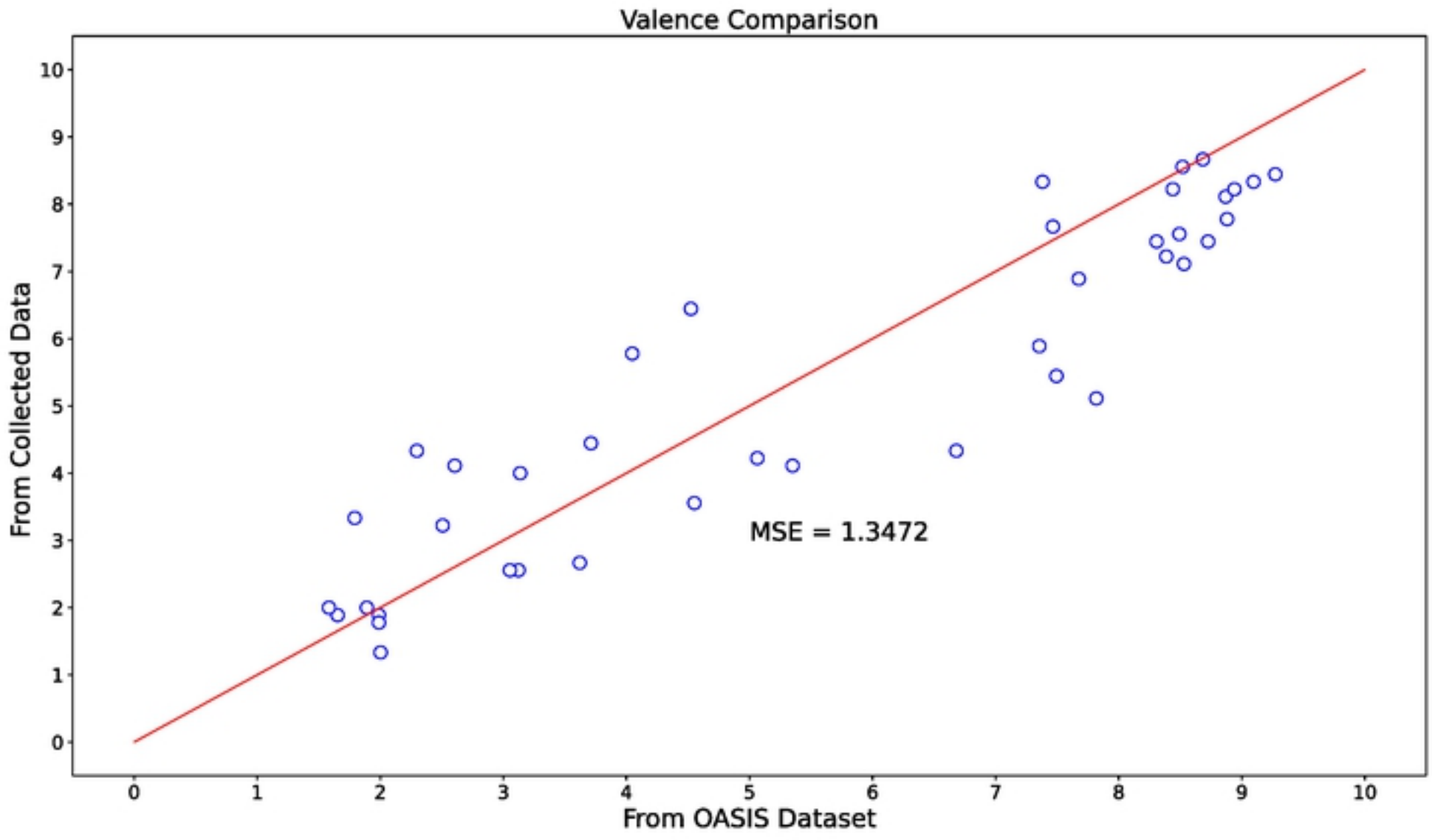
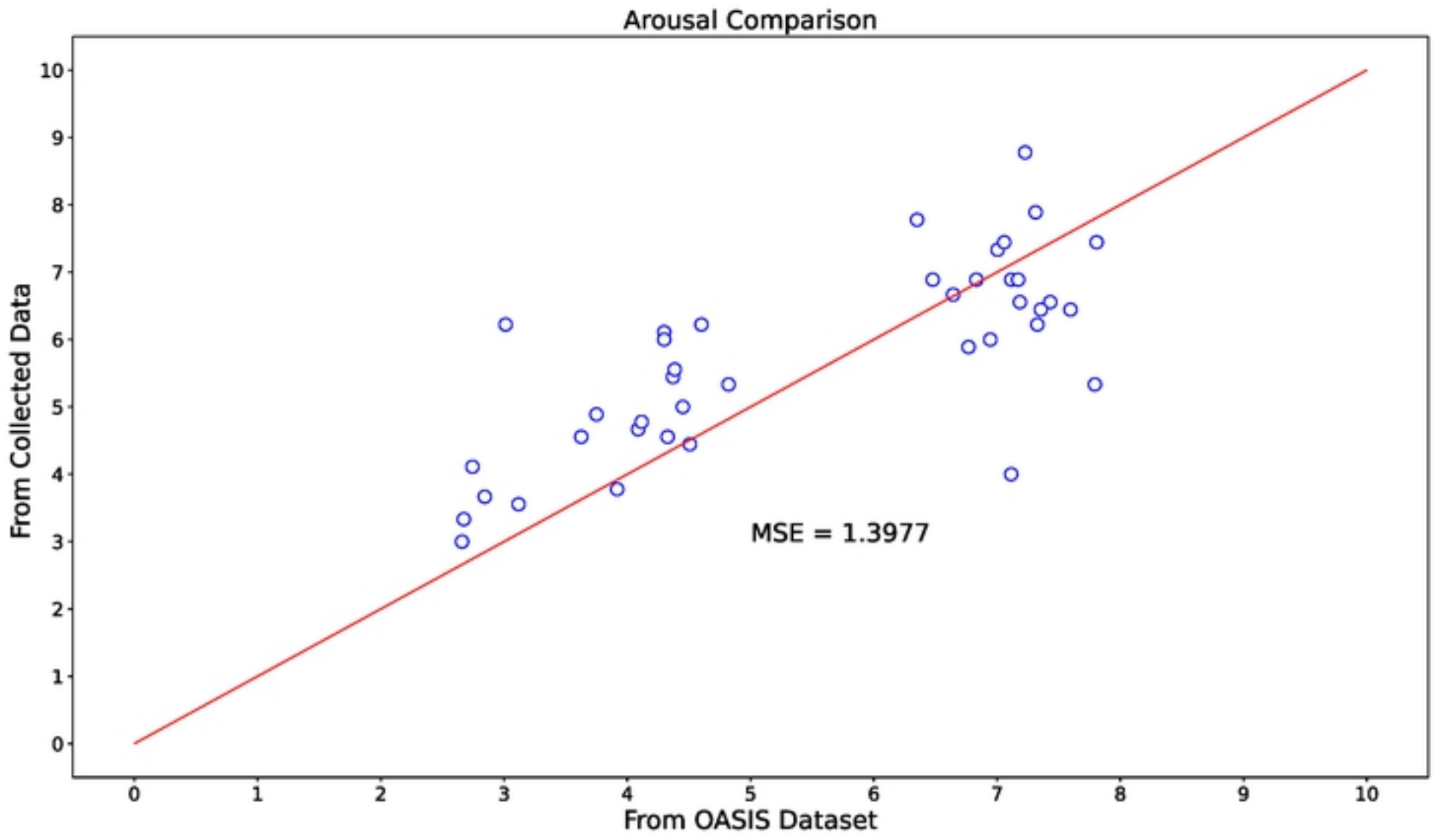


Image Trail
Viewing Image (5 seconds)
Valence and Arousal ratings from the subjects
Session
Image Trail 1
Image Trail 1
Image Trail 40
Experiment
Instruction and Explanation of Valence and Arousal
Session
End

Figure 2


Valence Comparison
From Collected Data
From OASIS Dataset
MSE = 1.3472

(a)


Arousal Comparison
From Collected Data
From OASIS Dataset
MSE = 1.3977

(b)

Figure 3

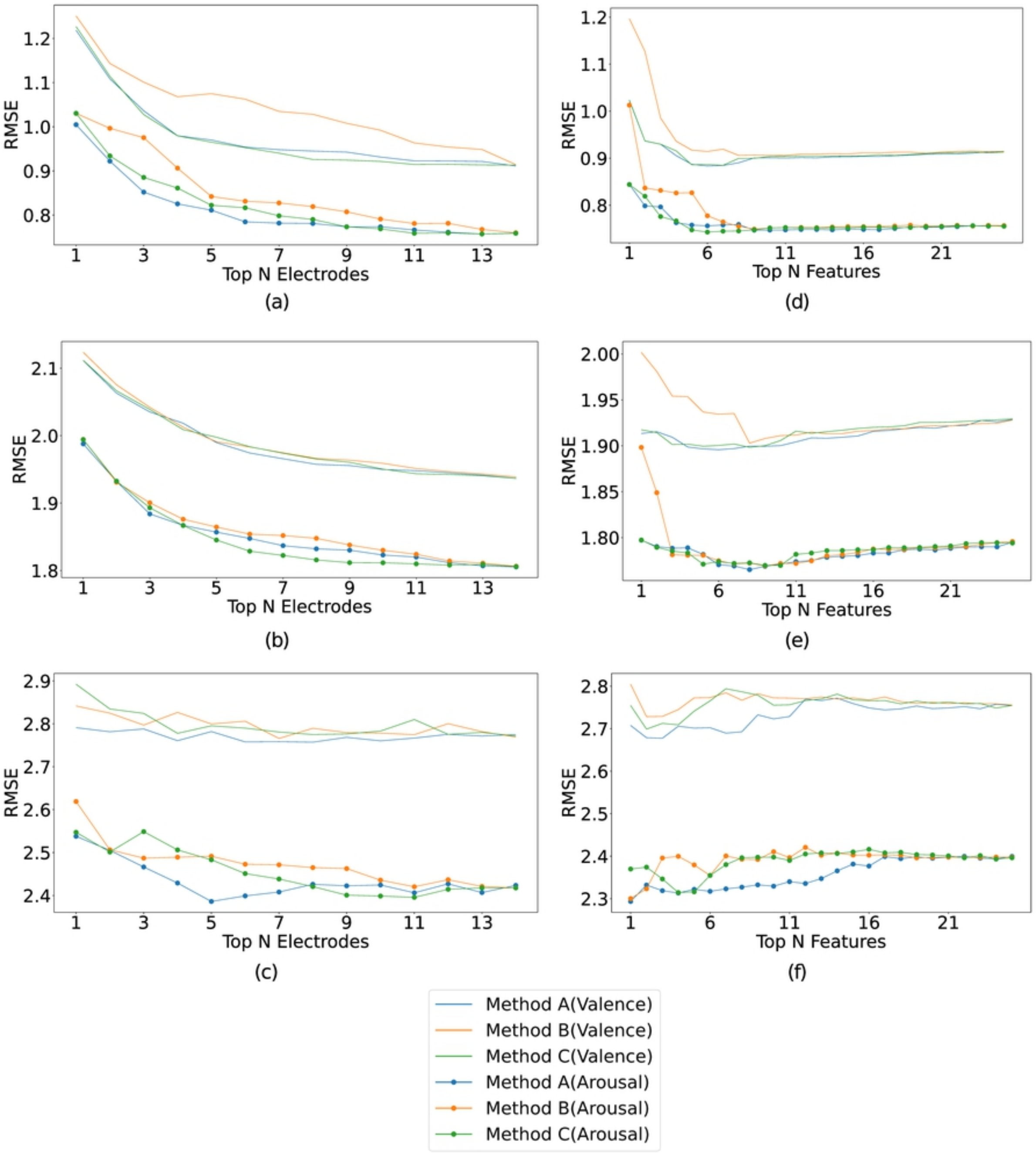
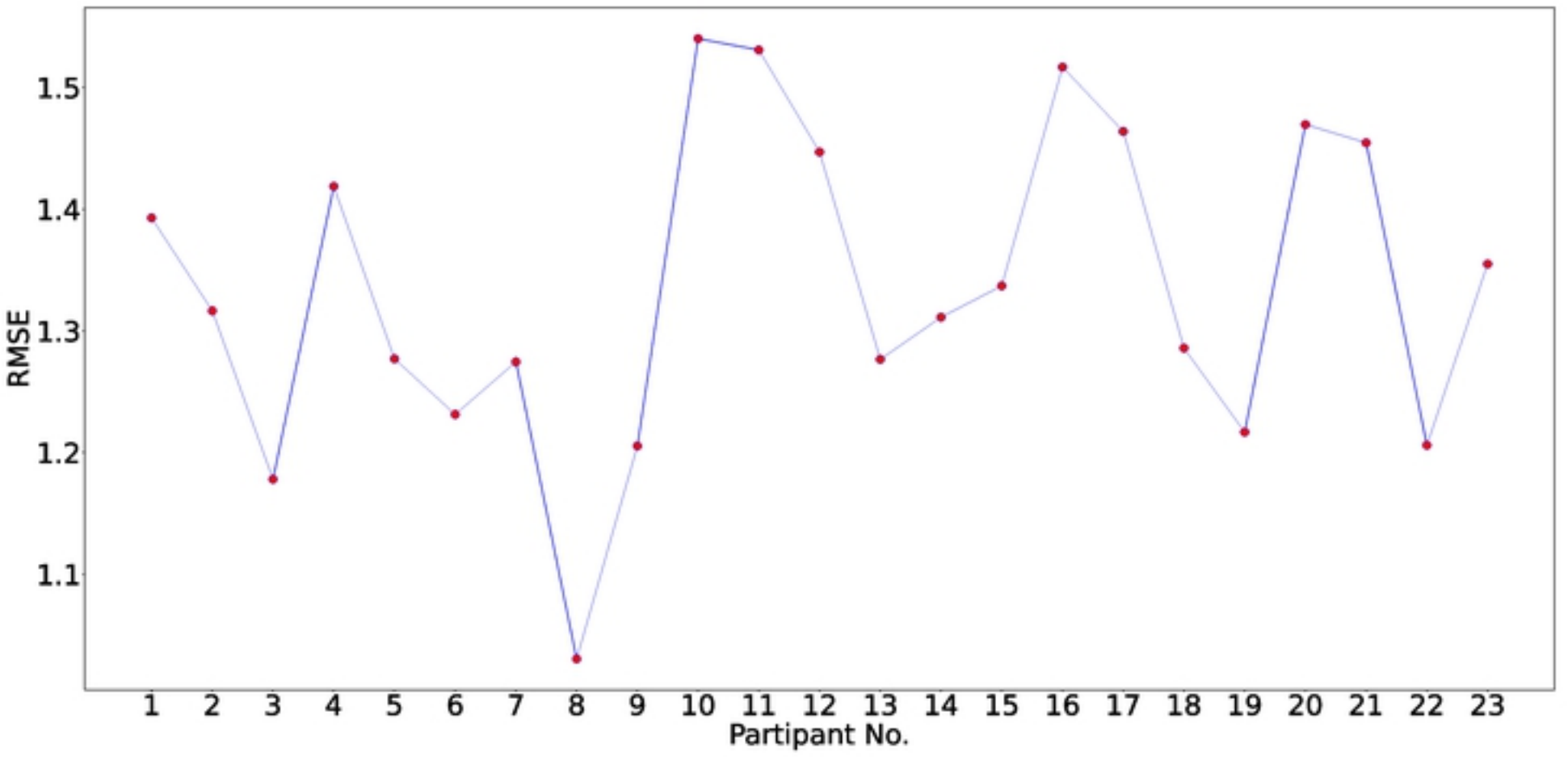
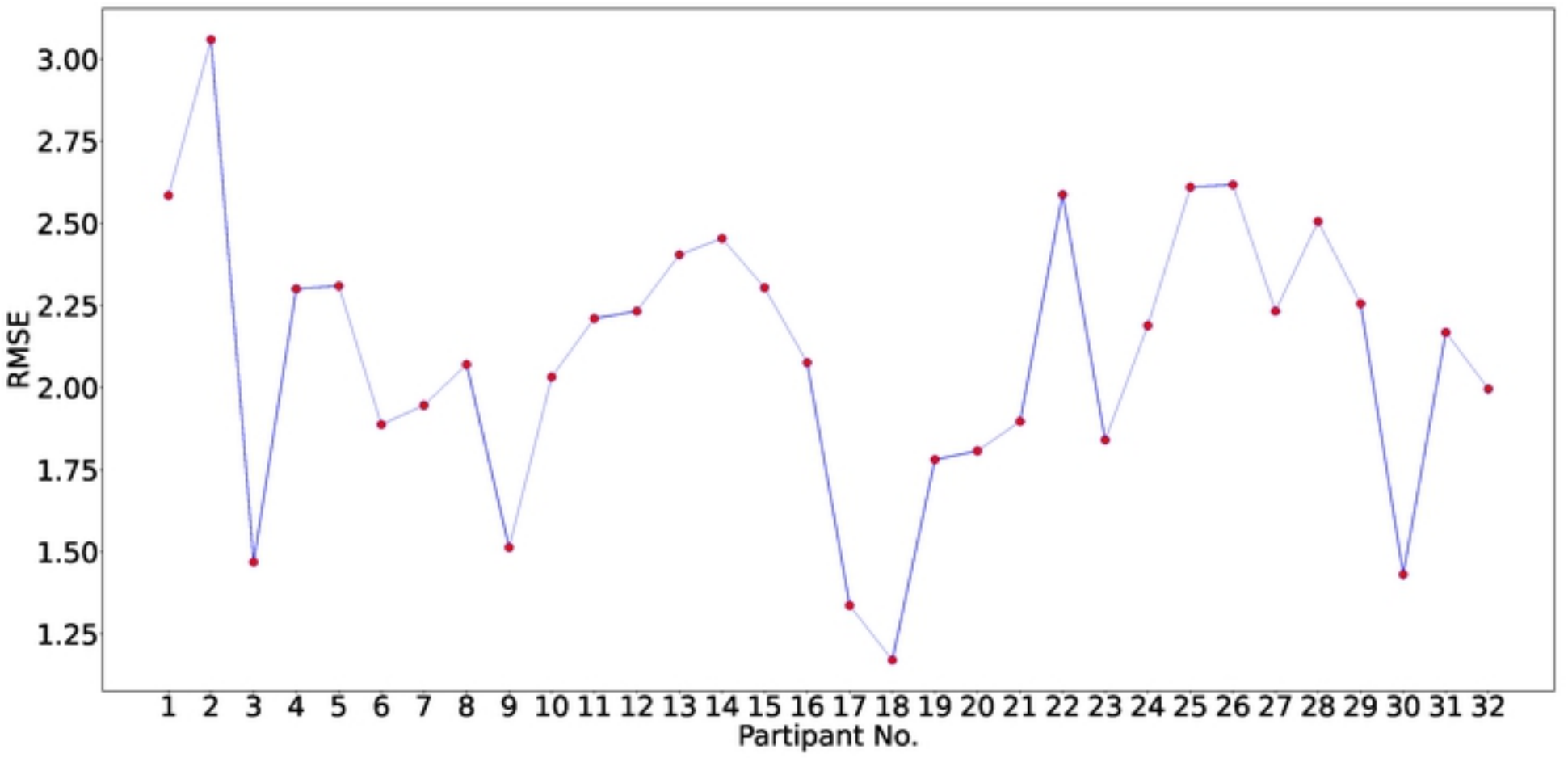

RMSE
Top N Electrodes
(a)
Top N Features
(d)
(b)
(e)
(c)
(f)
Method A(Valence)
Method B(Valence)
Method C(Valence)
Method A(Arousal)
Method B(Arousal)
Method C(Arousal)

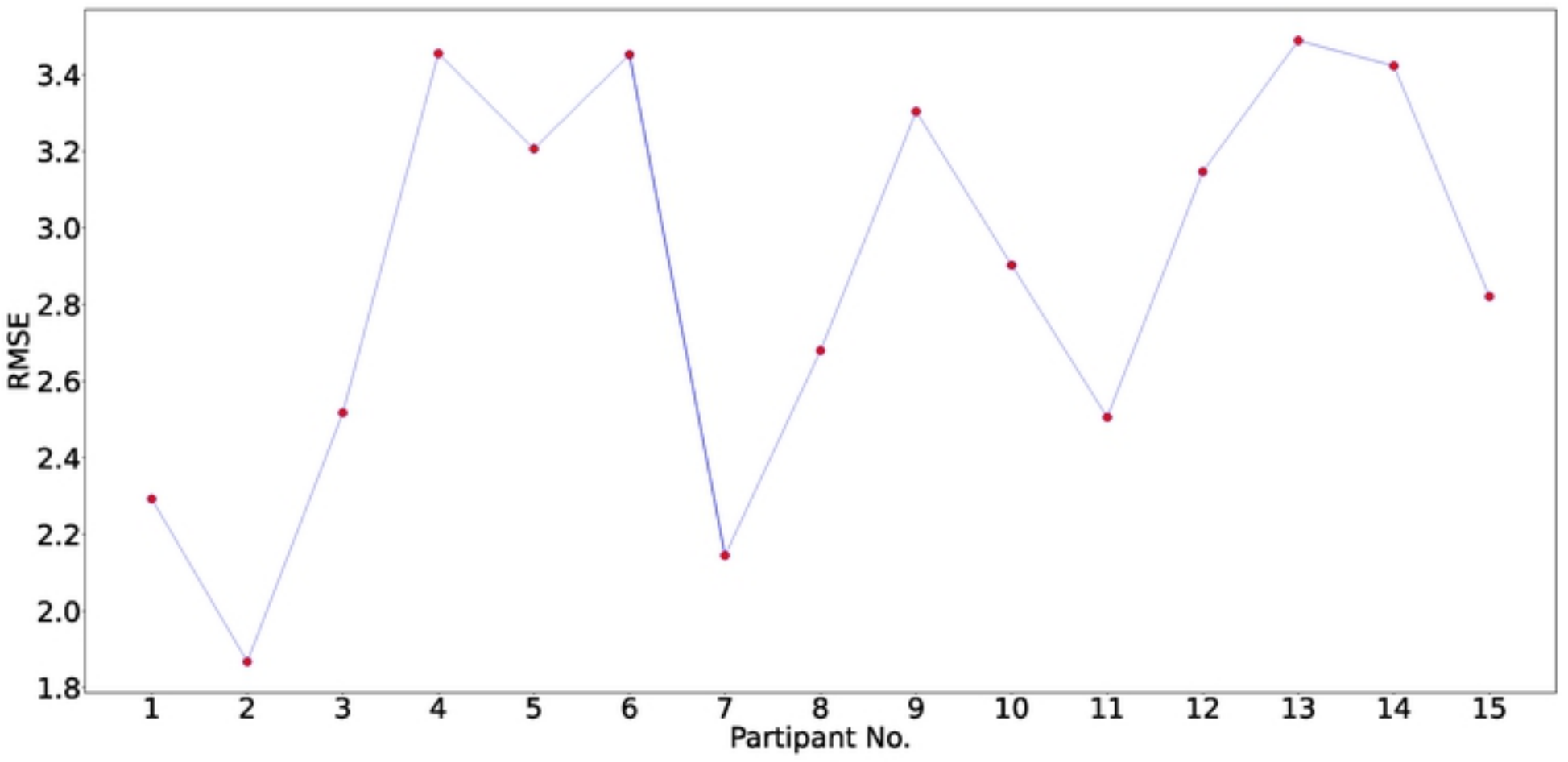
Figure 4


RMSE
1.5
1.4
1.3
1.2
1.1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Partipant No.

(a)

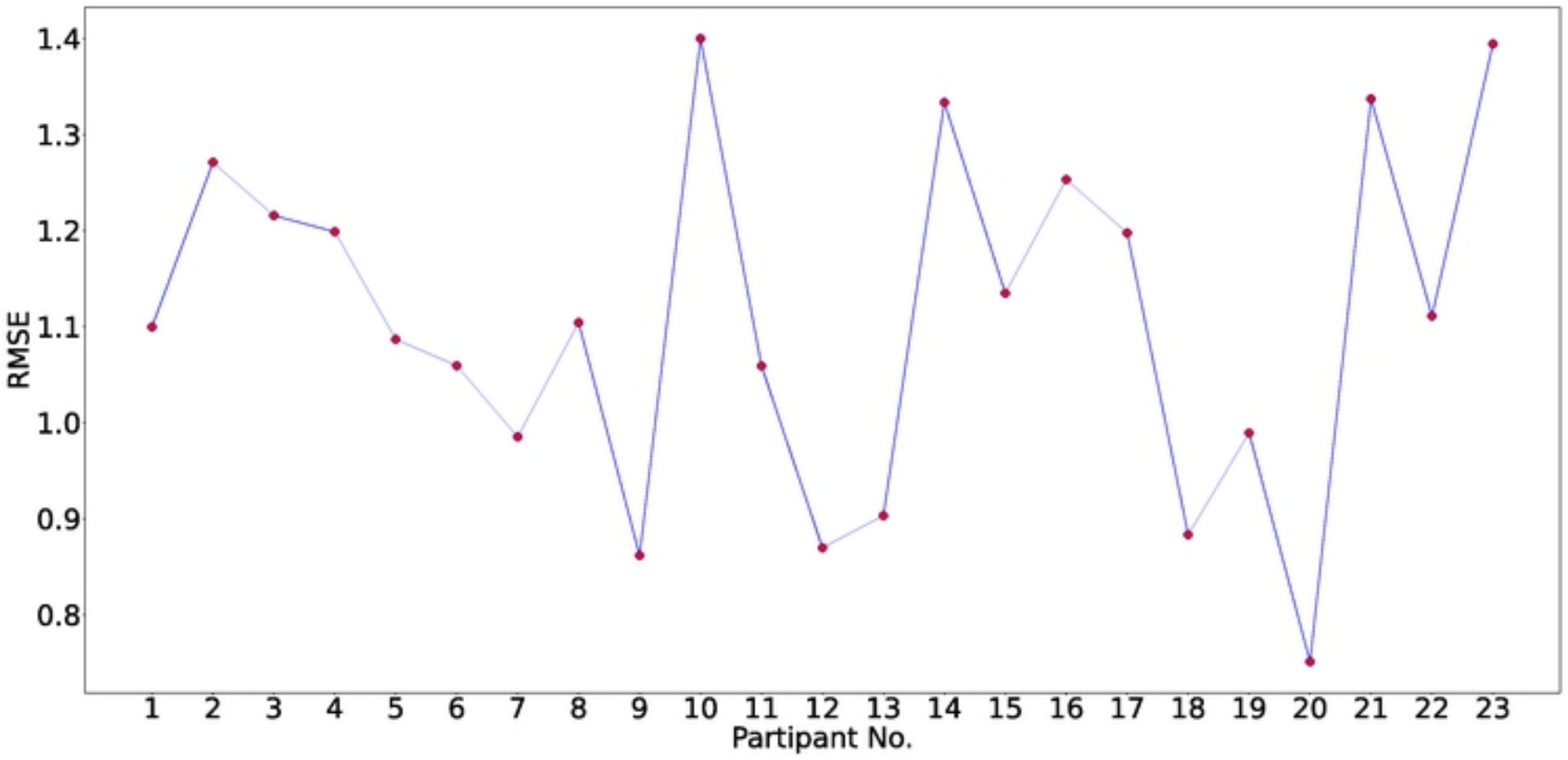

RMSE
3.00
2.75
2.50
2.25
2.00
1.75
1.50
1.25
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
Partipant No.

(b)

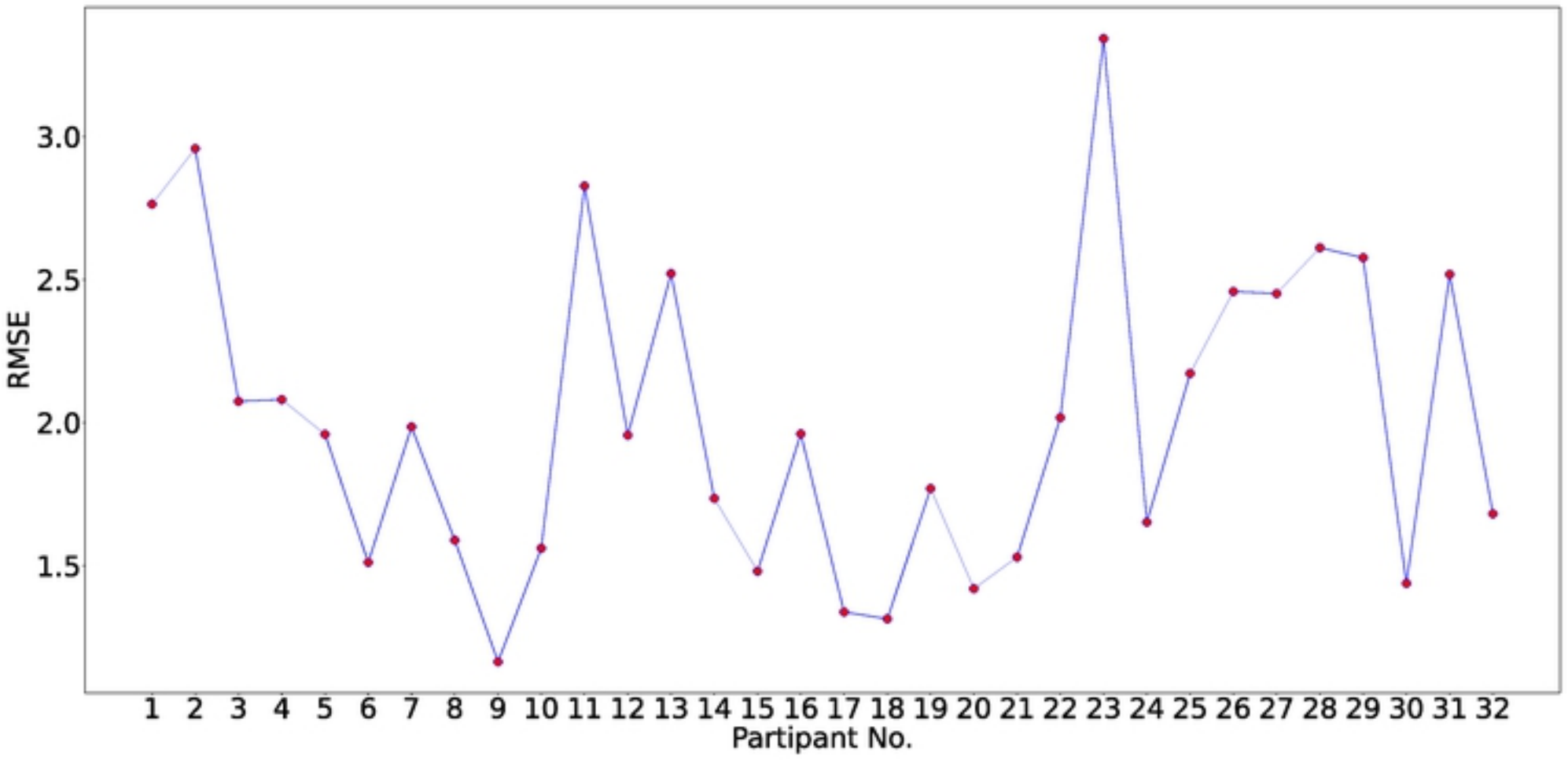

RMSE
3.4
3.2
3.0
2.8
2.6
2.4
2.2
2.0
1.8
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Partipant No.

(c)

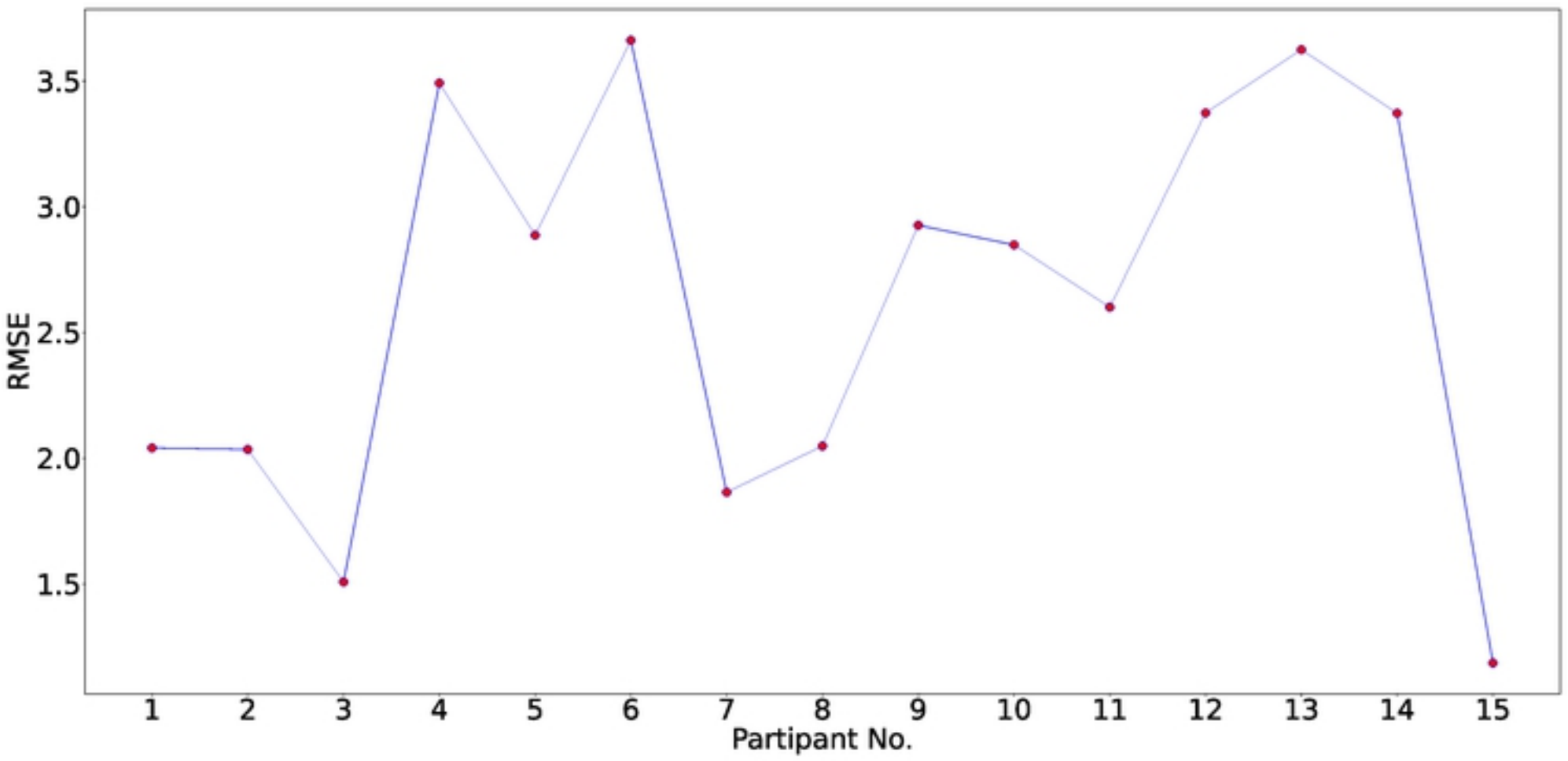
Figure 6


RMSE
1.4
1.3
1.2
1.1
1.0
0.9
0.8
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Partipant No.

(a)


RMSE
3.0
2.5
2.0
1.5
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
Partipant No.

(b)


RMSE
3.5
3.0
2.5
2.0
1.5
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Partipant No.

(c)

Figure 7

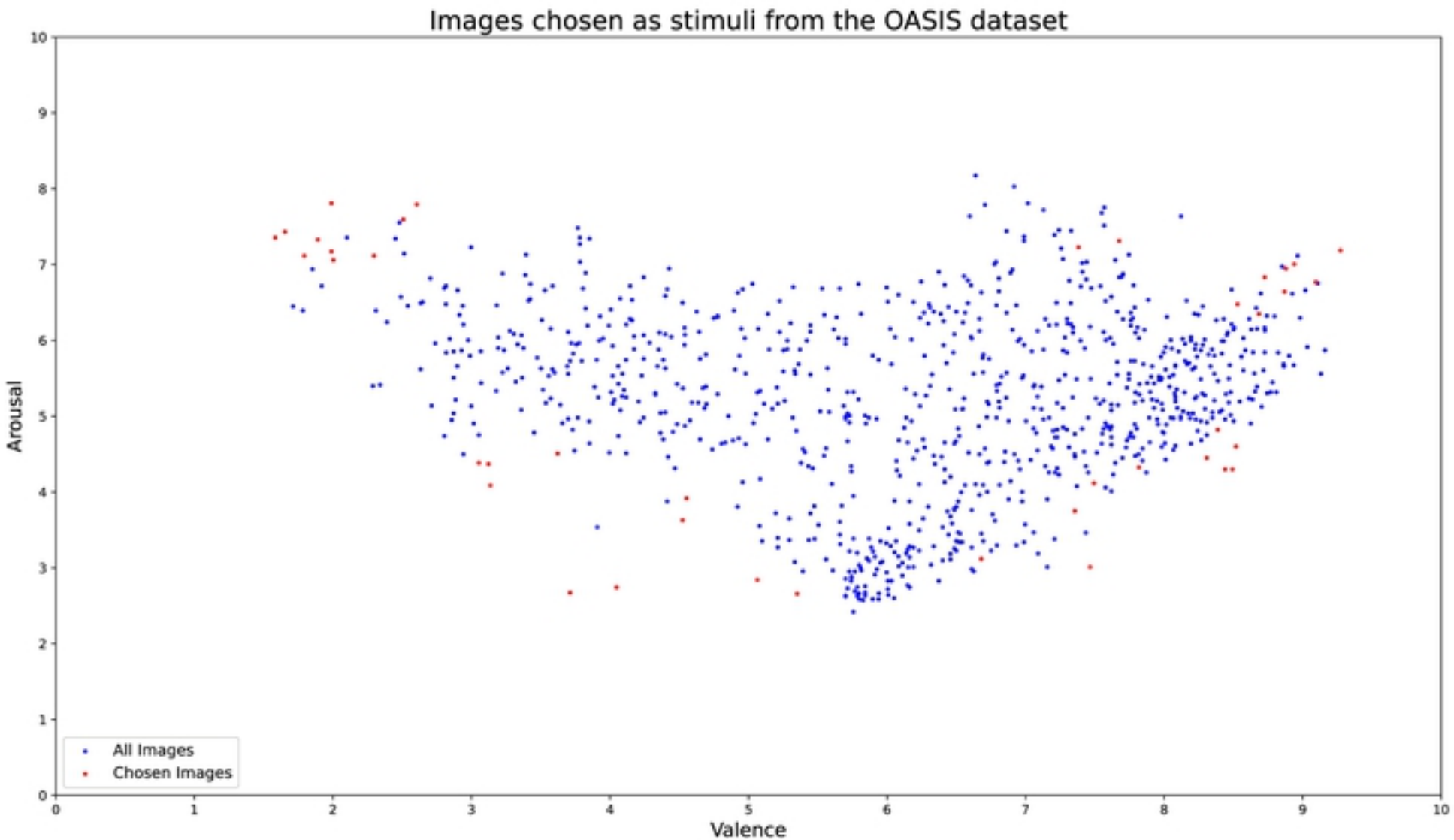
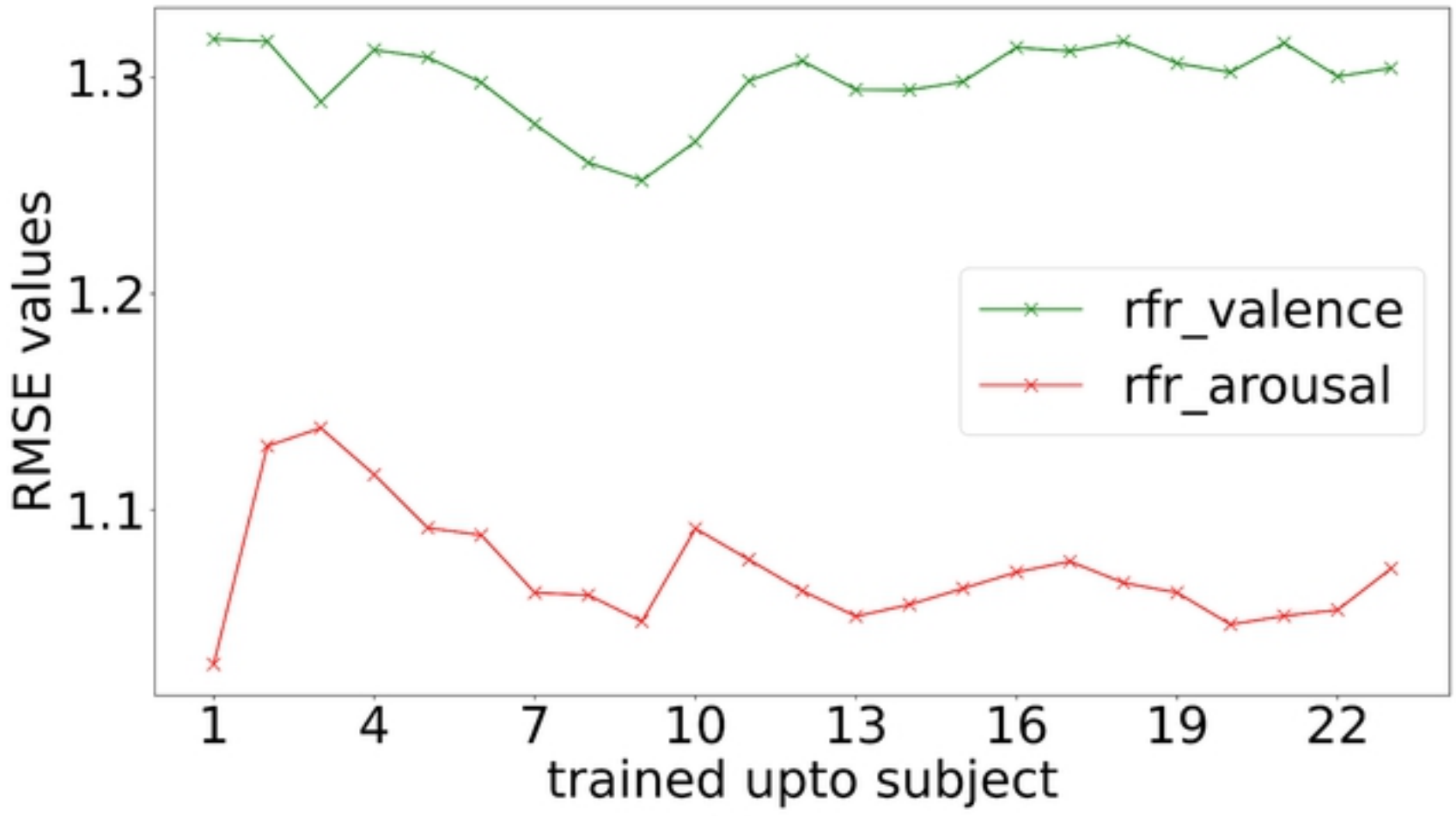
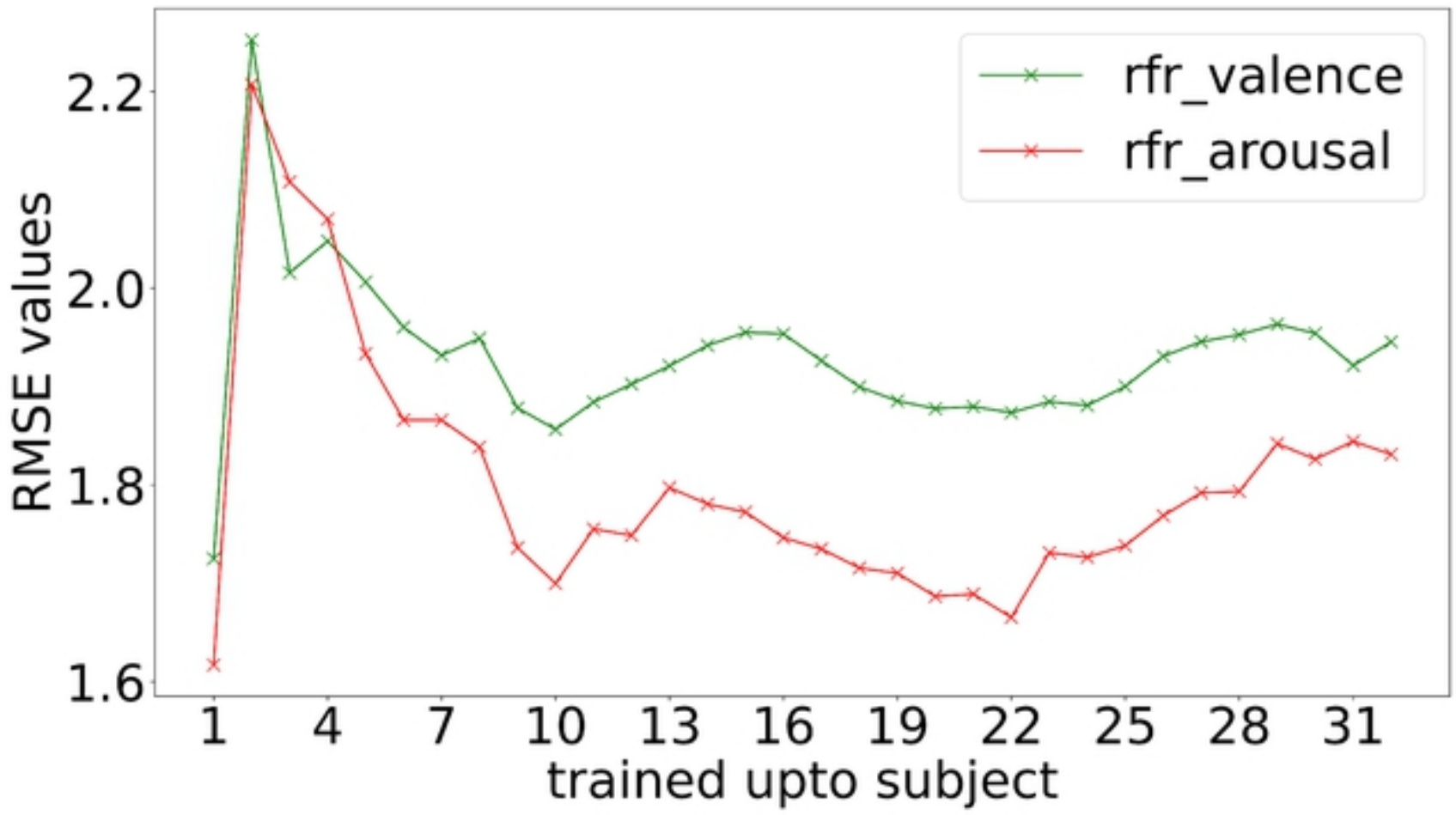

Images chosen as stimuli from the OASIS dataset
Arousal
Valence
All Images
Chosen Images
0
1
2
3
4
5
6
7
8
9
10

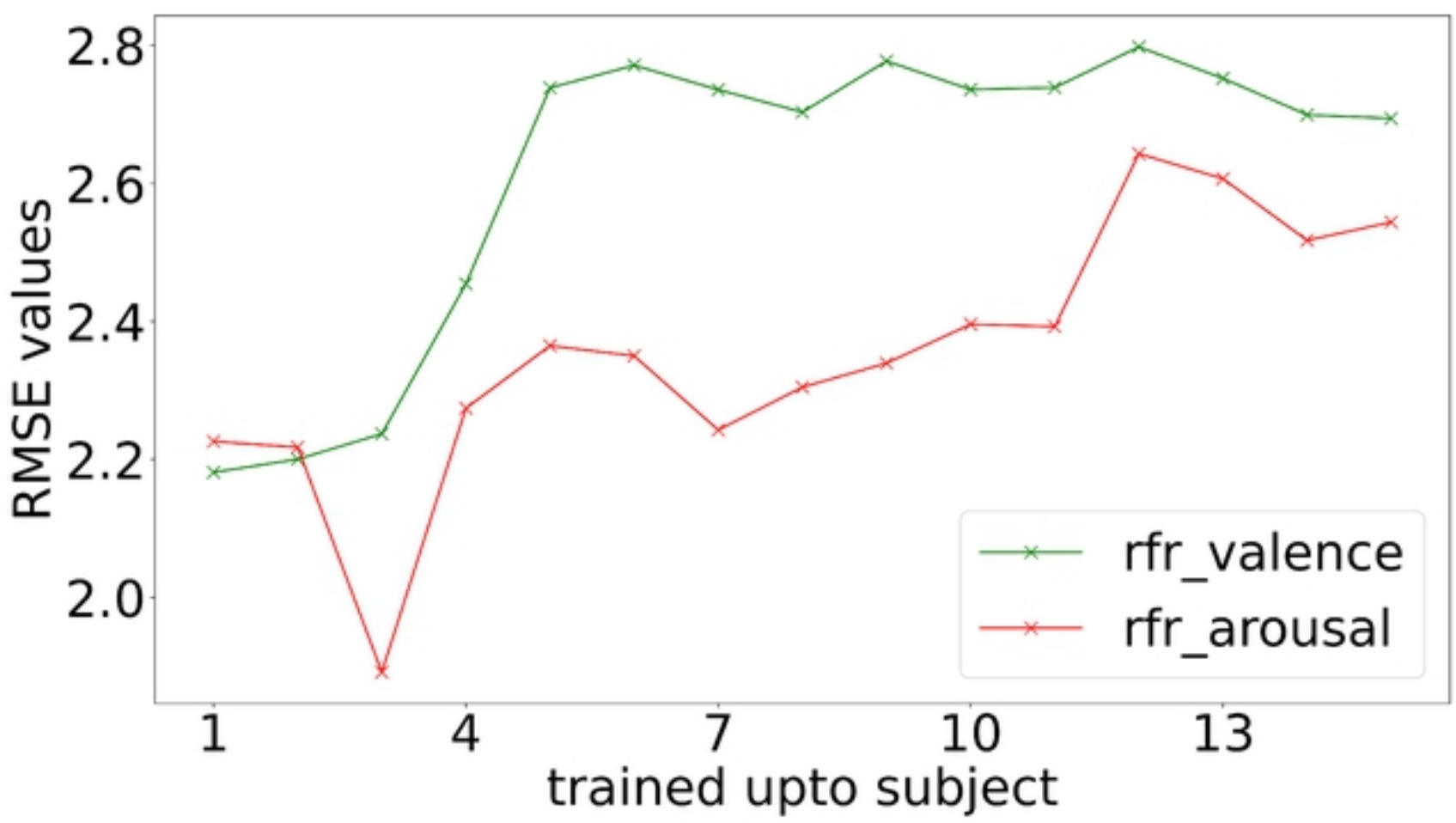
Figure 1


1.3
1.2
1.1
RMSE values
rfr_valence
rfr_arousal
1
4
7
10
13
16
19
22
trained upto subject

(a)


2.2
2.0
1.8
1.6
RMSE values
rfr_valence
rfr_arousal
1
4
7
10
13
16
19
22
25
28
31
trained upto subject

(b)


2.8
2.6
2.4
2.2
2.0
RMSE values
rfr_valence
rfr_arousal
1
4
7
10
13
trained upto subject

(c)

Figure 5