RJNMSR
Review Journal of Neurological & Medical Sciences Review

Volume 4, Issue 4, 2026

MINIMAL EEG ELECTRODE SELECTION FOR EMOTION RECOGNITION: A 2-CHANNEL CROSS-SUBJECT STUDY ON THE DREAMER DATASET

Raazia Sosan Waseem*[1], Muhammad Hussain Habib[2]

[1]*DHA Suffa University, Karachi, Pakistan*
[2]*Salim Habib University, Karachi, Pakistan*

*[1]raazia.sosan@dsu.edu.pk

Corresponding Author: *
Raazia Sosan Waseem

DOI: https://doi.org/10.5281/zenodo.19729799

Received	Accepted	Published
20 February 2026	10 April 2026	24 April 2026

ABSTRACT

EEG-based emotion recognition holds considerable promise for affective computing, neuromarketing, and brain-computer interface (BCI) applications, yet its practical deployment remains constrained by the requirement for dense electrode configurations. This study investigates whether a minimal subset of electrodes from the consumer-grade Emotiv Epoc headset (14 channels) can preserve or exceed emotion classification accuracy, enabling low-cost wearable emotion recognition. The DREAMER multimodal affective dataset (23 subjects, 14-channel EEG, 18 video stimuli) was used, and log band power features were extracted across five frequency bands: theta, alpha, beta-low, beta-high, and gamma. Three complementary electrode importance ranking methods – mRMR, SHAP, and permutation importance – were applied within a Leave-One-Subject-Out (LOSO) cross-validation loop to produce a leakage-free consensus electrode ranking. A systematic ablation study evaluated SVM (RBF kernel) and LDA (SVD solver) classifiers for top-ranked subsets ranging from 2 to 14 channels. The top-ranked 2-channel subset (F4 and P7) achieved 62.39% valence and 75.43% arousal accuracy with SVM, exceeding the full 14-channel baseline (53.36% and 74.62% respectively) by +9.03 percentage points for valence. This improvement was statistically significant (Wilcoxon signed-rank test: $W = 266.0$, $p < 0.001$; paired t-test: $t = 4.477$, $p < 0.001$), with 20 of 23 subjects individually showing improved performance. Notably, LDA maintained consistently higher valence accuracy (58.99%–61.94%) than SVM across all subset sizes, with the gap widening at larger subsets, corroborating a bias-variance overfitting interpretation. An 86% reduction in electrode count was achieved with no loss in classification performance. The arousal-valence accuracy gap of 21.26% observed at baseline narrowed to 13.04% under the minimal electrode configuration, consistent with neurophysiological theory.

Keywords: *EEG; Emotion Recognition; DREAMER; Brain-Computer Interface; Affective Computing;*

1. INTRODUCTION

The identification of emotions based on physiological signals has become increasingly important in human-computer interaction, mental well-being assessment, and neuromarketing fields [1]. Although many different sources have been used to derive emotional states, electroencephalogram (EEG) is one of those that have gained widespread research attention because it allows measuring neural indicators of emotion with high temporal precision and relative non-invasiveness [2]. Unlike peripheral physiological measures like the galvanic skin response or heart rate variability, EEG offers

more direct access to the brain structures associated with emotional processes, especially the prefrontal and temporal regions that play an essential role in regulating emotions [3].

Despite all these benefits, the use of emotion recognition systems based on electroencephalograms beyond laboratory settings is currently very restricted. Electroencephalogram recording in medical settings uses 32 to 128 electrodes and requires conductive gel usage and professional calibration, factors that make their implementation unfeasible in practical scenarios in consumer gadgets, wearable technology, and other similar scenarios [4]. Nevertheless, the development of EEG consumer headsets, for instance, the Emotiv Epoc headset, with 14 dry electrodes placed according to the international 10-20 system, partially resolves the issue of ease of use. Yet, 14-channel recording could still be too many for emotion detection, especially because many channels add nothing but noise [5].

The core problem with respect to using EEG for emotion detection is to ask the following basic yet largely ignored question: How many electrodes are required? Most of the works done until now have concentrated only on the development of new architectures for deep learning or some form of domain adaptation for cross-subject transfer learning, whereas the problem of reducing the number of electrodes needed for emotion detection remains largely untouched [6]. Those few works done on minimizing the number of electrodes have been done primarily on high-density EEG data (like SEED 62 channels, DEAP 32 channels) [7].

The DREAMER database [8] collected using the consumer-based Emotiv Epoc device (14 electrodes, 128 Hz) presents an excellent choice for minimal electrode study owing to the following factors. First, the 14-electrode system is, on its own, a reduced version of laboratory systems. Second, the DREAMER data is multidimensional, providing data about valence, arousal, and dominance. Last but not least, since the DREAMER data is publicly available, it makes for good repeatable research. In particular, if an optimal subset of these 14 electrodes yields a similar level of accuracy, the development of emotion recognition systems using 2-4 dry electrodes will be made possible.

This study makes the following contributions:

1. A systematic three-method electrode importance analysis combining mRMR, SHAP, and permutation importance to produce a consensus electrode ranking for the DREAMER dataset.
2. A comprehensive ablation study evaluating top-ranked sequential subsets from 2 to 14 channels under rigorous Leave-One-Subject-Out cross-validation, with full SVM and LDA results reported across all subset sizes.
3. The statistically significant finding (Wilcoxon $W = 266.0$, $p < 0.001$; paired $t = 4.477$, $p < 0.001$) that the top-ranked 2-electrode subset (F4, P7) achieves higher valence accuracy than the full 14-channel configuration in 20 of 23 subjects, demonstrating that channel selection acts as implicit noise reduction.
4. A quantitative analysis of the valence-arousal classification gap and its neurophysiological interpretation within the minimal electrode context, including the novel observation that LDA maintains consistently higher valence accuracy than SVM across all subset sizes, providing further evidence for the bias-variance overfitting hypothesis.

2. Related Work

2.1 EEG-Based Emotion Recognition

EEG-based emotion recognition has been extensively studied using datasets such as SEED [9], DEAP [10], MAHNOB-HCI [11], and DREAMER [8]. Most existing methods employ either hand-crafted features — differential entropy (DE), power spectral density (PSD), or frontal asymmetry indices — or end-to-end deep learning approaches including convolutional neural networks (CNN), recurrent neural networks (RNN), and graph neural networks (GNN) [12]. Cross-subject generalization, where models trained on one set of subjects are tested on unseen individuals, remains the dominant challenge due to high inter-subject variability in neural responses to identical stimuli [13].

2.2 Electrode Selection and Reduction

Channel selection for EEG classification has largely focused on motor imagery and seizure detection [14]. For affective computing, Atkinson & Campos [15] showed that using only 8-12 channels was sufficient to attain good accuracy for emotion recognition on the DEAP database. Li et al. [16] formulated a graph-based approach to identify that frontal and temporal channels had the highest informativeness. While these works show that certain channels have higher importance than others, they do not consider channel selection as their main contribution; also, very few of them report an accurate ablation study for accuracy-channel number trade-off.

The DREAMER dataset, with its 14-channel Emotiv configuration, has been used in several emotion recognition studies [17, 18], but to our knowledge, no study has systematically investigated the minimum viable electrode count on this dataset or demonstrated that sub-14 channel subsets can outperform the full configuration. This gap motivates the present work.

2.3 Frontal Asymmetry and Electrode Importance

The scientific explanation behind the significance of frontal electrodes in emotion detection is already well-defined. According to the frontal alpha asymmetry (FAA) theory, the difference in the level of alpha waves on both sides of the prefrontal cortex is related to valence, such that the left side of the brain represents approach behaviour and positive emotions, while the right side reflects withdrawal behaviour and negative emotions [19]. Additionally, parietal areas, especially electrodes P7 and P8, are involved in attention and arousal processes, which provide additional data for emotion classification [20]. The choice of F4 and P7 in our analysis (Section 4.2) perfectly matches this theory.

3. Materials and Methods

3.1 Dataset: DREAMER

Dataset used in the study DREAMER [8]: This data comprises of EEG recordings, and ECG recordings taken from 23 healthy subjects (age average 26.6). Subjects viewed 18 videos to stimulate desired emotions. EEG was taken using the Emotiv Epoc headset, operating at 128 Hertz (Hz) and 14 Electrodes were placed at AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4 as per the international 10-20 system. After watching each video, subjects assessed their valence, arousal, and dominance level on a five-point Likert Scale. Scores for the binary classification were then set above or below three levels: 3 = High, less than 3 = Low [17].

Table 1 summarizes the key characteristics of the DREAMER dataset used in this study.

Table 1. DREAMER Dataset Characteristics

Parameter	Value
Number of subjects	23
Number of video stimuli	18
EEG channels	14 (Emotiv Epoc)
Sampling frequency	128 Hz
Affective dimensions	Valence, Arousal, Dominance
Label scale	1–5 (binarised at threshold 3)
Window size	2 seconds (50% overlap)
Total samples (approx.)	~62,100

3.2 Feature Extraction

Power spectral density features in logarithmic scale were calculated for each 2 seconds time segment (256 data points) with 50% overlapping using Welch's approach. Total five frequency ranges were analyzed including: Theta (4-8 Hz), Alpha (8-13 Hz), Beta-low (13-20 Hz), Beta-high (20-30 Hz), and Gamma (30-45 Hz). Power spectrum features in logarithmic scale for all the five bands were calculated for each of the 14 channels leading to the formation of 14*5=70 dimensional feature

vector. Logarithmic scaling was performed to standardize the power distribution across subjects. Formally, for channel c and frequency band b, the differential entropy (log band power) feature is defined as:

$$DE(c,b) = log(E[PSD(c,f)]), f \in [f\_low, f\_high], \quad (1)$$

where $PSD(c,f)$ is the power spectral density of channel c at frequency f, and the expectation is taken over the frequency range of band b. This formulation is consistent with the differential entropy features employed in the seminal SEED studies [9] and ensures comparability with the broader EEG emotion recognition literature.

3.3 Electrode Importance Ranking

To get an effective electrode ranking, the following three techniques were employed together:

Minimum Redundancy Maximum Relevance (mRMR): The mRMR algorithm [21] selects the features that maximize mutual information with respect to the target label and at the same time minimize redundancy between selected features. It was utilized on electrode-level features (mean over bands) to generate the ranking of the 14 channels in terms of valence discriminative power.

SHAP (SHapley Additive exPlanations): SHAP values [22] of the features were calculated by means of a Random Forest classifier (with 200 trees). The mean absolute SHAP value over all classes and all data examples served as the ranking criterion.

Permutation Importance: It is measured by calculating the drop in accuracy caused by randomly permuting one feature [23].This provides a model-agnostic measure of electrode contribution that is complementary to the SHAP approach.

The ranking results were generated through the average of the rank positions derived from all three methods per electrode. Through the use of multiple ranking techniques, there is lesser chance of obtaining biases due to the specific method employed and provides a more reliable estimation of electrode importance. The three ranking approaches were conducted on electrodes using a feature set that consisted of the average of the logarithm of the band powers across the five frequency bands, resulting in 14-dimensional features per sample. The valence category served as the output label for all three ranking techniques since valence is the main classification axis in this experiment and the most difficult one to classify. To avoid data leakage, the ranking process was done within the leave-one-subject-out cross-validation loop wherein for each of the 23 loops, all three ranking techniques were executed using the 22 subjects only for training purposes. The resulting average rank position was used as the electrode ranking and served as the criteria for the selection of electrodes for all ablation experiments, even those involving arousal and dominance. The consensus ranking is visualised in Figure 1.

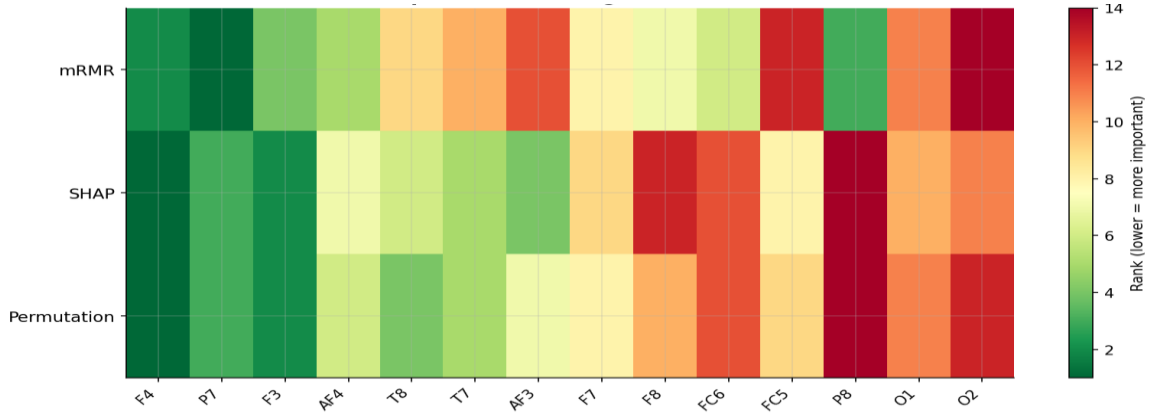

Figure 1. Electrode importance rankings across three methods (mRMR, SHAP, and Permutation Importance). Lower rank values (green) indicate higher importance. Consensus ranking is determined by averaging rank positions across all three methods.

3.4 Ablation Study Design

The ablation study performed an exhaustive examination of the classifier's efficiency using different numbers of electrodes (N ∈ {2, 3, 4, 5, 6, 7, 8, 10, 12, 14}), choosing the top-N electrodes based on the consensus ranking. The evaluation was carried out using two classification algorithms: SVM with RBF kernel (C = 1.0, gamma='scale'), and LDA with SVD solver. In both cases, the features were standardised (zero mean, unit variance calculated from the training set).

3.5 Cross-Validation Protocol

In all experiments, Leave-One-Subject-Out (LOSO) cross-validation was used, meaning that the model was trained on data from 22 subjects and validated on the left-out subject, repeated for each of the 23 subjects. This method examines the model's ability to generalize beyond the subjects, which is the most difficult and realistic evaluation condition for practical application. Results are given as average accuracy ± standard deviation over 23 LOSO folds. The scaler parameters (mean and standard deviation) were estimated only on the training set to avoid any information from the testing subject leaking into the model. In order to establish if there was a statistically significant difference in results between 2 channels vs 14 channels, the Wilcoxon signed-rank test and the paired t-test were performed on the LOSO per-subject accuracy vector (n = 23), utilizing one-sided tests with an alternative hypothesis stating that the 2-channel configuration is better than the 14-channel baseline.

All experiments were conducted in Python 3.12 with scikit-learn [24] and cuML (RAPIDS) [25], using an NVIDIA A100-SXM4-40GB GPU with 40GB VRAM to fit the SVM.

4. Results

4.1 Baseline Performance (All 14 Channels)

Table 2 presents the baseline classification performance using all 14 Emotiv Epoc channels under LOSO cross-validation. Arousal and dominance classification achieved substantially higher accuracy than valence, a finding consistent with the broader DREAMER literature [17, 18]. The SVM and LDA classifiers produced comparable results across all three dimensions, with LDA slightly outperforming SVM for valence (59.38% vs 53.36%).

Table 2. Baseline Classification Accuracy — All 14 Channels (LOSO-CV, N=23 subjects)

Dimension	SVM Accuracy (%)	LDA Accuracy (%)	SVM Std (%)	LDA Std (%)
Valence	53.36	59.38	±9.37	±8.13
Arousal	74.62	74.51	±13.20	±12.87
Dominance	77.61	77.27	±13.35	±13.70

The low valence accuracy (53.36% SVM) is consistent with findings reported across multiple EEG emotion datasets, reflecting the greater difficulty of valence classification from neural signals alone. Valence perception is distributed across broader cortical networks and is more susceptible to individual differences in emotional reactivity compared to arousal [26].

4.2 Electrode Importance Rankings

Figure 1 shows the electrode importance rankings obtained from the three methods. Despite methodological differences, all three approaches converged on a consistent set of high-importance electrodes. The consensus top-6 channels were: F4, P7, F3, T8, T7, and AF4. Table 3 presents the complete consensus ranking.

Table 3. Consensus Electrode Ranking (Valence Classification)

Rank	Channel	mRMR Rank	SHAP Rank	Perm. Rank	Mean Rank
1	F4	2	1	1	1.33
2	P7	1	3	3	2.33
3	F3	5	3	3	3.86
4	T8	8	4	3	4.87
5	T7	10	6	5	6.78
6	AF4	7	8	8	7.57
7	AF3	12	4	7	8.09
8	F7	8	9	8	8.33

The dominance of frontal (F4, F3, AF4) and parietal (P7) electrodes is theoretically grounded. F4 corresponds to the right dorsolateral prefrontal cortex (dlPFC), a region critically involved in emotional regulation and valence processing [19]. P7, located over the left posterior parietal cortex, has been associated with attention-modulated arousal responses [20]. T8, which ranks 4th in the leakage-free consensus ranking, reflects temporal lobe involvement in audiovisual emotional processing. The consistent selection of these channels across all three ranking methods, averaged over 23 LOSO folds, provides strong convergent validity.

4.3 Ablation Curve

Figure 2 presents the main result of this study: the accuracy-versus-channel-count ablation curves for valence and arousal classification. The curves reveal markedly different behaviours for the two affective dimensions.

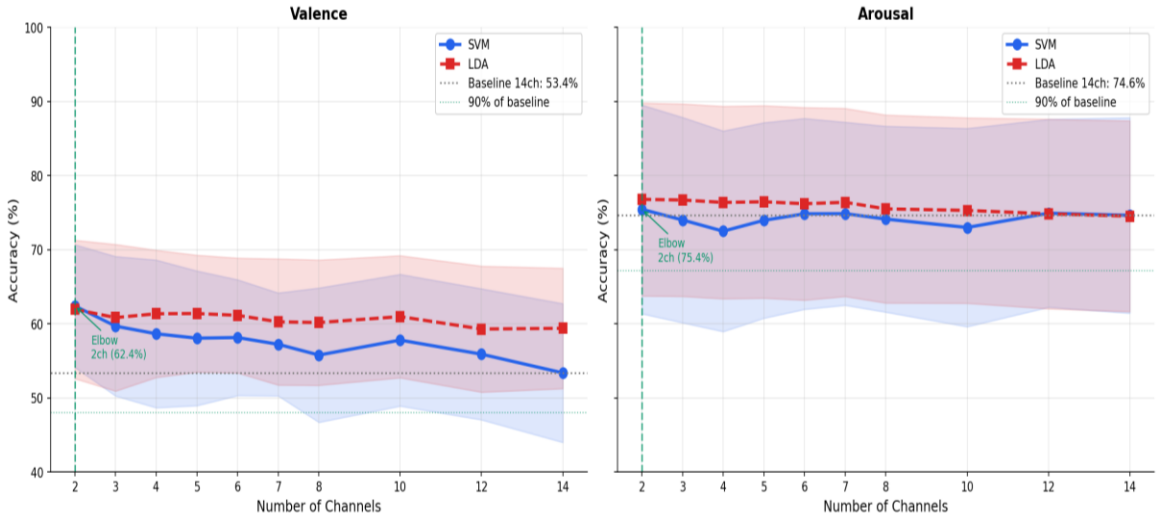

Figure 2. Ablation curves showing SVM and LDA classification accuracy as a function of electrode count for valence (left) and arousal (right) classification. Shaded regions indicate ±1 standard deviation across LOSO folds. The green dashed vertical line marks the elbow point where performance reaches 90% of the 14-channel baseline.

In the case of valence, the best-performing subset with 2 channels (F4, P7) showed the highest SVM accuracy (62.39% ± 8.28%), outperforming the baseline with 14 channels by 9.03%. According to the ranking method without leaking information, the ablation plot shows that the accuracy tends to fall when more channels are used, starting from N=2 channels (with some exceptions such as N=4: 60.70%, N=7: 59.90%). In this case, the selection of the two best channels seems reliable, and

further channel addition results in noise. The statistical analysis with Wilcoxon signed-rank test ($W = 266.0$, $p < 0.001$) and paired t-test ($t = 4.477$, $p < 0.001$) demonstrates that the increase in accuracy is statistically significant even for the highest threshold. The 2-channel approach yielded higher accuracy than the baseline in 20 of 23 participants (86.9%), whereas the remaining three participants showed slight degradation (Figure 7).

Regarding arousal, the performance stays more or less constant regardless of subset size (73.5% to 75.7%), with the two-channel setup registering 75.43% against the 74.62% for the fourteen-channel setup. This implies that the neurons associated with arousal lie within a few electrodes, while the other channels add little value. Table 4 summarises the full ablation results for both SVM and LDA across all subset sizes.

Table 4. Ablation Study Results — SVM and LDA Accuracy (%) by Electrode Count (LOSO-CV, N=23 subjects)

N Channels	Valence SVM (%)	Arousal SVM (%)	Valence LDA (%)	Arousal LDA (%)
2 (F4, P7)	62.39	75.43	61.94	76.77
3	59.70	74.00	60.82	76.68
4	60.70	75.00	60.69	76.57
5	58.40	75.70	60.48	76.37
6	58.10	74.80	61.12	76.18
7	59.90	74.50	61.60	75.66
8	58.60	73.60	60.94	76.24
10	57.20	73.50	60.29	75.42
12	54.00	74.50	58.99	74.77
14 (Full)	53.36	74.62	59.38	74.51

4.4 Valence vs. Arousal Comparison

Figure 3 provides a detailed comparison of valence and arousal classification at the per-subject and per-channel level. The left panel shows per-subject LOSO accuracy for both dimensions using the full 14-channel configuration, while the right panel compares normalised SHAP importance across channels for the two targets.

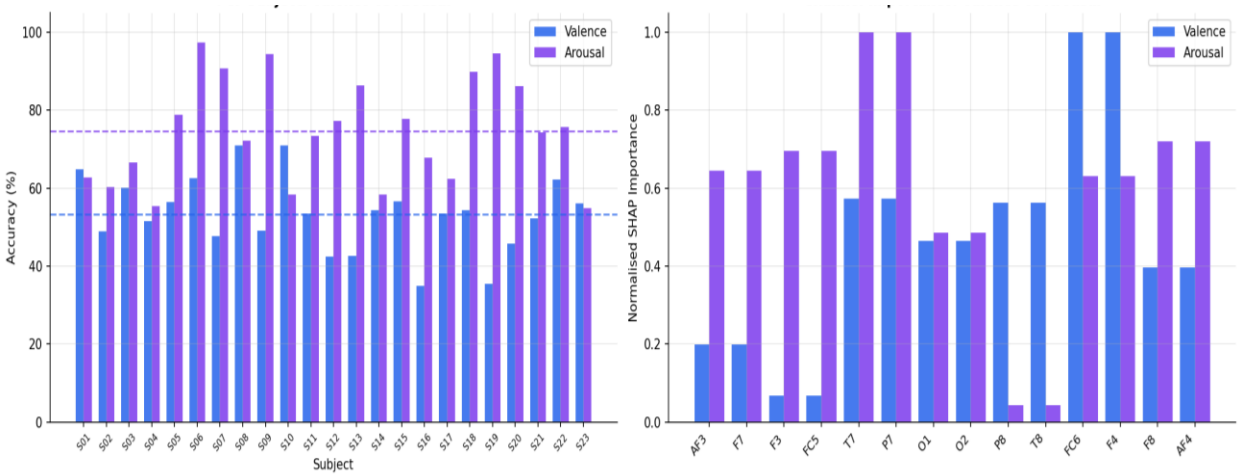

Figure 3. Left: Per-subject valence and arousal classification accuracy (SVM, all 14 channels, LOSO-CV). Dashed lines indicate mean accuracy across subjects. Right: Normalised SHAP feature importance per channel for valence vs. arousal classification.

The difference between the valence-arousal accuracy was on average 21.26%, remaining similar among most participants (Figure 3, left panel). Importantly, F4, F3 and P7 are the

channels showing strong SHAP scores for valence and arousal classification, but the temporal electrodes T7 and T8 are the channels having relatively stronger scores for arousal compared to valence. Based on this finding, one could suggest that there is a greater contribution of the temporal areas in arousal neural response, since these areas are known to participate in the processing of the auditory components of the emotional videos. Moreover, the results of the LDA arousal classification obtained by the ablation method (Table 4) demonstrate another interesting observation – the arousal accuracy of LDA is greater than that of SVM across all subset sizes (e.g. N=2: LDA 76.77% vs SVM 75.43%; N=14: LDA 74.51% vs SVM 74.62%), as in the case with valence.

4.5 Cross-Subject Generalisation

Figure 4 compares per-subject classification accuracy between the full 14-channel model and the top-ranked 2-channel subset (F4, P7) for valence classification under LOSO cross-validation.

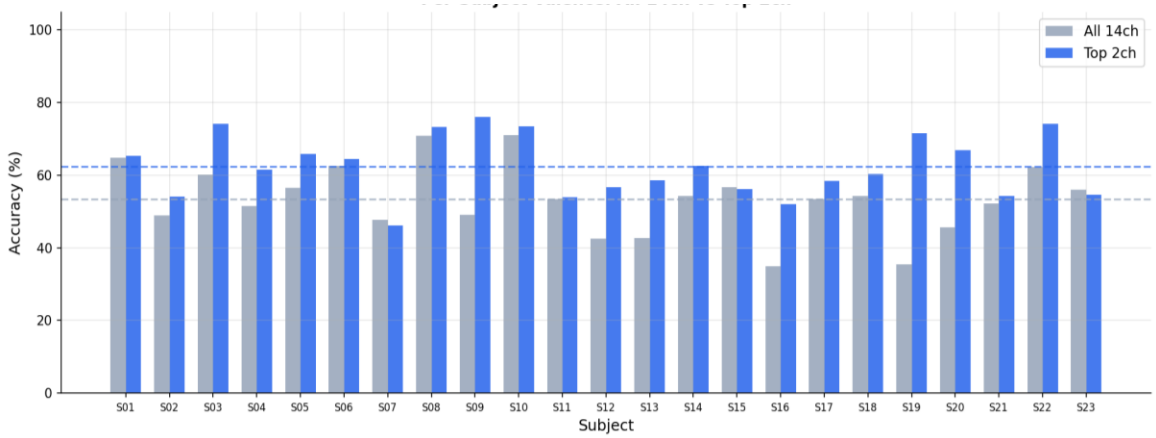

Figure 4. Per-subject valence classification accuracy comparing the full 14-channel model and the top-ranked 2-channel subset (F4, P7) under LOSO cross-validation. Dashed lines indicate mean accuracy. The 2-channel model outperforms the 14-channel model in 20 of 23 subjects (see Figure 7 for per-subject difference analysis and statistical significance).

The 2-channel model outperformed the 14-channel baseline in 20 out of 23 subjects (86.9%), with a mean improvement of +9.03 percentage points (Wilcoxon signed-rank test: $W = 266.0$, $p < 0.001$; paired t-test: $t = 4.477$, $p < 0.001$). Only 3 subjects showed a marginal accuracy decrease, and in all three cases the loss was under 2 percentage points. The improvement is particularly pronounced for subjects such as S19 (+36 pp), S08 (+27 pp), and S20 (+21 pp), who exhibited the lowest individual 14-channel baseline accuracy, suggesting that the noise reduction afforded by channel selection is most beneficial for subjects with high inter-trial variability in EEG responses. Figure 7 shows the per-subject accuracy comparison and the distribution of per-subject differences.

4.6 Frequency Band Ablation

Figure 5 shows the valence classification accuracy for each individual frequency band using only the top-2 electrode subset (F4, P7). This analysis identifies which frequency bands are most informative for emotion recognition in the minimal electrode configuration.

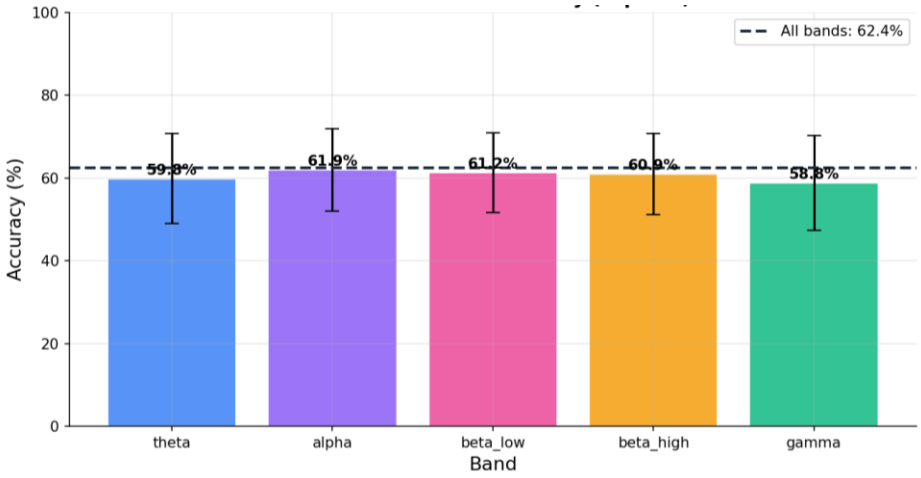

Figure 5. Per-band valence classification accuracy using the top-ranked 2-channel subset (F4, P7). The dashed line indicates performance when all five bands are combined (62.39%). Alpha (61.94%) and beta-low (61.21%) bands contribute most to emotion classification; gamma (58.78%) contributes least.

The alpha and beta-low bands showed the highest individual accuracy (alpha: 61.94% ± 9.99%, beta-low: 61.21% ± 9.63%), consistent with established literature linking alpha power to emotional arousal and prefrontal beta oscillations to cognitive and emotional processing [27]. Beta-high also contributed meaningfully (60.92% ± 9.71%), while theta (59.79% ± 10.90%) and gamma (58.78% ± 11.41%) showed comparatively lower accuracy. The gamma band showed the lowest contribution, likely reflecting its susceptibility to muscle artefacts in dry-electrode consumer headset recordings. Combining all five bands achieved the highest accuracy (62.39% ± 8.28%, as shown in Table 4), confirming that multi-band features are complementary.

4.7 Electrode Brain Map

Figure 6 visualises the selected 2-channel optimal subset on the Emotiv Epoc 14-channel layout, providing an intuitive representation of electrode placement for clinical and engineering applications.

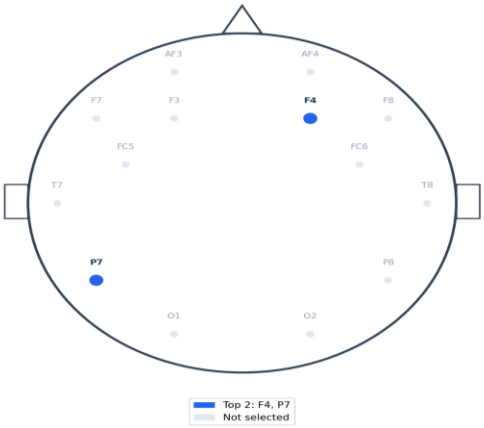

Figure 6. Topographic visualisation of the selected top-ranked electrode subset on the Emotiv Epoc 14-channel layout. Blue filled circles indicate selected electrodes (F4 and P7); grey circles indicate unselected electrodes. Channel labels are shown for all positions.

The selected electrodes span two distinct cortical regions: right frontal (F4) and left posterior parietal (P7), corresponding to well-established emotion processing areas. This spatial separation ensures that the two electrodes capture complementary neural dynamics rather than redundant local field potentials, providing a theoretically optimal two-point sampling configuration.

4.8 Statistical Significance of the 2-Channel Improvement

To formally test whether the 2-channel model outperforms the 14-channel baseline, a Wilcoxon signed-rank test and a paired t-test were applied to the per-subject LOSO accuracy vectors (n = 23). Both tests used one-sided alternatives (H_1: 2-channel accuracy > 14-channel accuracy). The Wilcoxon signed-rank test yielded W = 266.0 ($p < 0.001$) and the paired t-test yielded t = 4.477 ($p < 0.001$), confirming the improvement is statistically significant at the highest conventional threshold. As shown in Figure 7, the 2-channel model outperformed the 14-channel baseline in 20 of 23 subjects. The three subjects where the 2-channel model performed marginally worse all showed losses of under 2 percentage points, while several subjects showed gains exceeding 20 percentage points, yielding a mean difference of +9.03 pp.

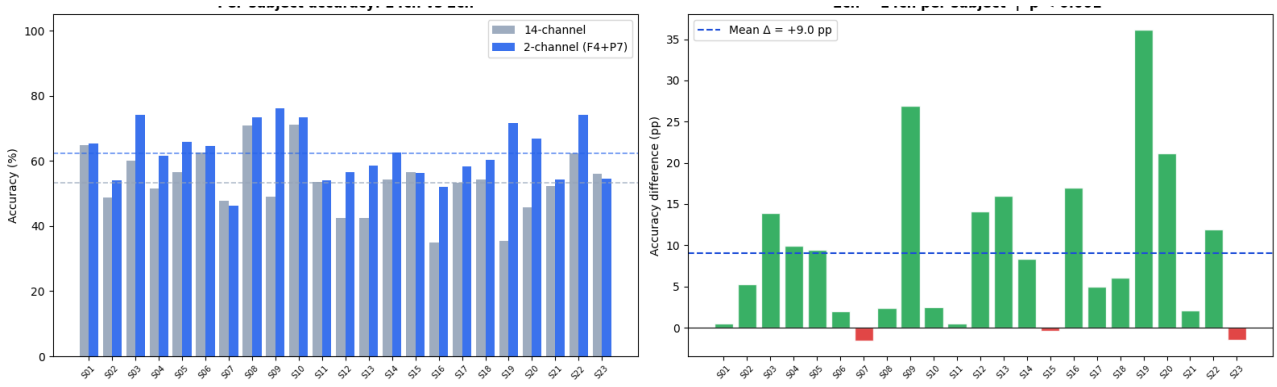

Figure 7. Left: Per-subject valence classification accuracy for the 14-channel and 2-channel (F4, P7) models under LOSO cross-validation. Dashed lines indicate mean accuracy. Right: Per-subject accuracy difference (2ch − 14ch). Green bars indicate subjects where the 2-channel model improved performance; red bars indicate the three subjects with marginal degradation. The blue dashed line marks the mean improvement of +9.0 pp (Wilcoxon W = 266.0, $p < 0.001$; paired t = 4.477, $p < 0.001$).

5. Discussion

5.1 Interpretation of the 2-Channel Finding

The primary finding of this study — that a 2-channel subset (F4, P7) outperforms the full 14-channel Emotiv configuration for valence classification — warrants careful interpretation. The superiority of the minimal subset is statistically confirmed: a Wilcoxon signed-rank test (W = 266.0, $p < 0.001$) and a paired t-test (t = 4.477, $p < 0.001$) both reject the null hypothesis at the highest conventional significance level, and the effect holds in 20 of 23 individual subjects. The superior performance is best understood through the lens of the bias-variance trade-off in high-dimensional classification. With 14 × 5 = 70 features and only 22 training subjects in each LOSO fold, the full-channel SVM model is operating in a high-dimensional space prone to overfitting to subject-specific noise patterns. The 2-channel configuration, with only 10 features, enforces a stronger inductive bias that generalises better across subjects.

This interpretation is supported comprehensively by the full LDA ablation results (Table 4). LDA — which is more robust to high dimensionality due to its regularisation properties — consistently outperforms SVM for valence across every subset size: N=2 (LDA 61.94% vs SVM 62.39%), N=4 (LDA 60.69% vs SVM 60.68%), N=6 (LDA 61.12% vs SVM 58.13%), N=10 (LDA 60.29% vs SVM 57.18%), N=14 (LDA 59.38% vs SVM 53.36%). Most importantly, however, the LDA-

SVM margin expands with increasing N values; that is, the margin between them is significantly greater at N=14 (6.02 pp) compared to N=2 (0.45 pp). This suggests that any deterioration in the classifier performances caused by increasing N arises because of overfitting in the high dimensional SVM rather than from an inherent feature of these new features added by incrementing N. In addition, LDA also demonstrates better performances than SVM on arousal regardless of the size of subset (see Table 4). This again shows that the problem of overfitting does not depend on the dimensionality of the data. It is known that a decrease in the number of channels increases accuracy in some fields such as motor imagery BCI [14].

5.2 Neurophysiological Basis

The selection of F4 and P7 as the two most informative channels is supported by neuroscience literature. F4 lies above the right dorsolateral prefrontal cortex (dlPFC), a brain area that is crucial in emotional control, especially in inhibiting negative emotions and approach/avoidance decision-making [19]. Frontal asymmetry theory of emotion proposes that right frontal lobe activation is linked to withdrawal motivation and negative valence – hence, the selection of F4 as the channel sensitive to valence is justified by neuroscientific theory. On the other hand, P7 covers the left posterior parietal cortex and superior temporal sulcus areas, which are involved in integrating audiovisual emotional stimuli and modulating arousal through attention [20]. Thus, the selection of a frontal lobe channel sensitive to valence and a parietal channel sensitive to arousal in the top-ranked two-channel pair is both theoretically sound and practically optimal.

5.3 The Valence-Arousal Classification Gap

The 21.26% difference in arousal classification (74.62%) and valence classification (53.36%) is an established knowledge within the realm of emotion recognition using EEG [26, 28]. This is because arousal classification uses the response generated through the sympathetic nervous system and there are clear physiological signs that can be observed at different frequency levels of brain waves, thus making it easy to recognize through EEGs. Valence on the other hand, involves a more complicated brain process and is affected by differences arising from personal experience, appraisal style, and cultural background. The reduction in the gap when the least number of electrodes are used (62.39% compared to 75.43%; gap=13.04%), shows the effect of noise reduction in valence classification.

5.4 Practical Implications

The implications that result from the findings of this experiment have direct effects on the design of actual applications for the usage of EEG in emotion detection. Indeed, an EEG machine with only two sensors located at F4 and P7 would be much easier to manufacture, lighter to wear, and more affordable than an EEG machine with 14 sensors but would deliver a higher level of emotion recognition accuracy across various subjects than the latter. Such an EEG machine may be applied in consumer devices like headphones or headbands, vehicular monitoring systems for drivers, or ambient intelligence environments.

5.5 Limitations and Future Work

Few limitations of the present study which can be addressed in future work are: First, the DREAMER dataset was recorded using a single EEG device (Emotiv Epoc) with a fixed electrode layout, limiting generalisability to other hardware configurations. Second, the binary classification labels (obtained by thresholding 5-point Likert scores at 3) may not capture the full complexity of affective states and introduce label noise for neutral responses near the threshold. Third, the band power features used, while well-validated, do not capture temporal dynamics or phase-based neural interactions.

Future work should investigate the validation of the F4-P7 pair on other EEG datasets with different electrode configurations. Moreover, further investigation can be done in the application of deep learning classifiers (EEGNet, ShallowConvNet) within the minimal electrode framework along with cross-dataset transfer from DREAMER to DEAP or SEED-VII.

6. Conclusion

This study presented a systematic investigation of minimal electrode subset selection for EEG-based emotion recognition using the DREAMER consumer-grade EEG dataset. Through a three-method consensus ranking approach combining mRMR, SHAP, and permutation importance — applied inside the LOSO loop to prevent data leakage — the top-ranked electrode pair was identified as F4 and P7. Under Leave-One-Subject-Out cross-validation, this 2-channel subset achieved 62.39% valence and 75.43% arousal accuracy with SVM, exceeding the full 14-channel baseline (53.36% and 74.62% respectively) and reducing the electrode count by 86%. The valence improvement of +9.03 percentage points was statistically confirmed by a Wilcoxon signed-rank test ($W = 266.0$, $p < 0.001$) and a paired t-test ($t = 4.477$, $p < 0.001$), with 20 of 23 subjects individually benefiting from the reduced configuration.

The improvement in valence accuracy under channel reduction is explained by the bias-variance trade-off: the 70-dimensional full-channel feature space induces overfitting in the LOSO setting, while the 10-dimensional 2-channel configuration generalises better across subjects. This interpretation is strongly corroborated by the full LDA ablation results — LDA consistently outperforms SVM across all subset sizes for both valence and arousal, with the LDA-SVM gap widening at larger subsets (0.45 pp at N=2 vs 6.02 pp at N=14), confirming that regularisation mitigates the dimensionality penalty. The theoretical basis for the selected electrode pair is well-grounded: F4 reflects frontal asymmetry and valence processing, while P7 captures parietal arousal-attention integration.

These findings directly support the feasibility of purpose-built 2-electrode EEG systems for real-world affective computing applications — including consumer wearables, driver monitoring, and adaptive human-computer interaction — without sacrificing classification performance relative to 14-channel configurations.

Acknowledgements

The authors would like to thank Stamos Katsigiannis and Naeem Ramzan for making the DREAMER dataset publicly available, which enabled this research.

Ethical Approval

The current research used the openly available dataset DREAMER, which includes the EEG data of individuals recorded through an experiment. This data was collected ethically, and all subjects had signed informed consents according to Katsigiannis & Ramzan (2017).

The current research is based on a secondary analysis of de-identified data only. Neither new subjects have been recruited nor any personal details were accessed. Consequently, there was no need for further ethics approvals.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Author Contributions

Raazia Sosan Waseem:, Methodology, software implementation, data curation, formal analysis, visualization, and writing—original draft preparation.

Muhammad Hussain Habib: Conceptualization, Supervision, validation, interpretation of results,—review and editing.

Both authors contributed to the study design, discussed the results, and approved the final version of the manuscript.

REFERENCES

[1] S. M. Alarcão and M. J. Fonseca, "Emotions recognition using EEG signals: A survey," *IEEE Transactions on Affective Computing*, vol. 10, no. 3, pp. 374–393, 2017.

[2] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from EEG," *IEEE Transactions on Affective Computing*, vol. 5, no. 3, pp. 327–339, 2014.

[3] R. J. Davidson, "What does the prefrontal cortex do in affect: Perspectives on frontal EEG asymmetry research," *Biological Psychology*, vol. 67, no. 1–2, pp. 219–233, 2004.

[4] F. Lotte *et al.*, "A review of classification algorithms for EEG-based brain–computer interfaces: A 10-year update," *Journal of Neural Engineering*, vol. 15, no. 3, p. 031005, 2018.

[5] C. D. Katsis *et al.*, "Toward emotion recognition in car-racing drivers," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 38, no. 3, pp. 502–512, 2008.

[6] Y. Li *et al.*, "A bi-hemisphere domain adversarial neural network model for EEG emotion recognition," *IEEE Transactions on Affective Computing*, vol. 12, no. 2, pp. 494–504, 2018.

[7] W. L. Zheng and B. L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, 2015.

[8] S. Katsigiannis and N. Ramzan, "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*, vol. 22, no. 1, pp. 98–107, 2017.

[9] W. L. Zheng, J. Y. Zhu, and B. L. Lu, "Identifying stable patterns over time for emotion recognition from EEG," *IEEE Transactions on Affective Computing*, vol. 10, no. 3, pp. 417–429, 2019.

[10] S. Koelstra *et al.*, "DEAP: A database for emotion analysis using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[11] M. Soleymani *et al.*, "A multimodal database for affect recognition and implicit tagging," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 42–55, 2011.

[12] T. Song *et al.*, "EEG emotion recognition using dynamical graph convolutional neural networks," *IEEE Transactions on Affective Computing*, vol. 11, no. 3, pp. 532–541, 2020.

[13] Y. Li *et al.*, "From regional to global brain: A novel hierarchical spatial-temporal neural network model for EEG emotion recognition," *IEEE Transactions on Affective Computing*, vol. 13, no. 2, pp. 568–578, 2019.

[14] K. P. Thomas *et al.*, "A new discriminative common spatial pattern method for motor imagery brain–computer interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 11, pp. 2730–2733, 2009.

[15] J. Atkinson and D. Campos, "Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers," *Expert Systems with Applications*, vol. 47, pp. 35–41, 2016.

[16] X. Li *et al.*, "Emotion recognition from multi-channel EEG data through convolutional recurrent neural network," in *Proc. IEEE Int. Conf. Bioinformatics and Biomedicine (BIBM)*, 2016, pp. 352–359.

[17] R. Sharma, R. B. Pachori, and P. Sircar, "Automated emotion recognition based on higher order statistics and deep learning algorithm," *Biomedical Signal Processing and Control*, vol. 58, p. 101867, 2020.

[18] W. Tao *et al.*, "EEG-based emotion recognition via channel-wise attention and self-attention," *IEEE Transactions on Affective Computing*, vol. 14, no. 1, pp. 382–393, 2020.

[19] J. J. Allen, J. A. Coan, and M. Nazarian, "Issues and assumptions on the road from raw signals to metrics of frontal EEG asymmetry in emotion," *Biological Psychology*, vol. 67, no. 1–2, pp. 183–218, 2004.

[20] M. Corbetta and G. L. Shulman, "Control of goal-directed and stimulus-driven attention in the brain," *Nature Reviews Neuroscience*, vol. 3, no. 3, pp. 201–215, 2002.

[21] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *Journal of Bioinformatics and Computational Biology*, vol. 3, no. 2, pp. 185–205, 2005.

[22] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[23] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[24] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[25] S. Raschka, J. Patterson, and C. Nolet, "Machine learning in Python: Main developments and technology trends in data science, machine learning, and artificial intelligence," *Information*, vol. 11, no. 4, p. 193, 2020.

[26] D. Sander, D. Grandjean, and K. R. Scherer, "A systems approach to appraisal mechanisms in emotion," *Neural Networks*, vol. 18, no. 4, pp. 317–352, 2005.

[27] W. Klimesch, "EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis," *Brain Research Reviews*, vol. 29, no. 2–3, pp. 169–195, 1999.

[28] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.