Evaluating Cepstral Feature Extraction for EEG Classification: A Comparison with Spectral Method

Luka Matič and Assoc. Prof. Peter Rogelj, PhD
Faculty of Mathematics, Natural Sciences and Information Technologies (FAMNIT)
University of Primorska, Koper, Slovenia
Email: 89252128@student.upr.si

***Abstract*—Electroencephalography (EEG) signal classification underpins a wide range of brain-computer interface (BCI) applications, and the choice of feature representation strongly influences classifier performance. This paper compares a spectral feature extraction method (Welch power spectral density over a five-channel sensorimotor montage) against a cepstral feature pipeline (triangular filterbank, logarithmic compression, discrete cosine transform, and multivariate Gaussian probability density modeling) on two binary EEG classification tasks drawn from the PhysioNet EEG Motor Movement/Imagery Dataset: eyes-open versus eyes-closed resting state, and movement versus rest (combining executed and imagined unilateral fist movement). Features were selected within cross-validated folds using Minimum Redundancy Maximum Relevance (MRMR) to avoid data leakage, and Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA) classifiers were evaluated using 5-fold cross-validation with ANOVA-based significance testing. Results show that cepstral features combined with LDA provide a statistically significant improvement of approximately 4% over spectral features for the eyes-state task, while classification of the movement task remains substantially harder (52–55% accuracy) due to high inter-subject variability. Common Spatial Patterns (CSP) was also explored as an alternative spatial filtering approach for the movement task but did not yield improved accuracy over plain Welch PSD, and was therefore not adopted in the final pipeline.**

***Index Terms*—EEG, cepstral analysis, spectral analysis, CSP, MRMR, motor imagery, feature extraction**

I. Introduction

Brain-computer interfaces (BCIs) rely on decoding mental states from electroencephalography (EEG) signals, and a central design choice in any EEG classification pipeline is the feature representation used to translate raw, noisy multichannel recordings into a compact, discriminative form. Spectral features characterize frequency content, typically via power spectral density (PSD) or spatial filters such as Common Spatial Patterns (CSP), and are well established due to their direct link to known EEG rhythms (alpha, beta, mu). Cepstral features, by contrast, originate from speech processing and apply a further transformation to the spectral envelope – a filterbank, logarithmic compression, and discrete cosine transform (DCT) – to capture smoothed, decorrelated spectral shape information analogous to Mel-Frequency Cepstral Coefficients (MFCCs).

While cepstral representations have shown promise in speech and audio classification, their relative merit for EEG classification is less consistently established across tasks of differing difficulty. This raises a natural question: does the additional transformation in cepstral analysis yield a measurable advantage over spectral features, and does this advantage hold across different EEG classification tasks?

This paper addresses this question empirically through two binary classification tasks performed on the PhysioNet EEG Motor Movement/Imagery Dataset (EEGMMIDB): (1) discriminating eyes-open from eyes-closed resting-state EEG, and (2) discriminating rest from movement, combining executed and imagined unilateral fist movement. For each task, spectral and cepstral feature pipelines are constructed, features are selected using Minimum Redundancy Maximum Relevance (MRMR) within cross-validation folds to prevent information leakage, and two classifiers – Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA) – are trained and evaluated using 5-fold cross-validation. Statistical significance of performance differences is assessed using analysis of variance (ANOVA).

The remainder of this paper is organized as follows. Section II describes the dataset and task definitions. Section III details the feature extraction, feature selection, and classification methodology. Section IV presents classification results for both tasks. Section V discusses the implications of these results, and Section VI concludes the paper.

II. Dataset and Task Definitions

A. Dataset

All experiments in this study use the PhysioNet EEG Motor Movement/Imagery Dataset (EEGMMIDB) [1]. The dataset consists of over 1500 one- and two-minute EEG recordings collected from 109 healthy subjects, each recorded using a 64-channel EEG montage sampled at 160 Hz according to the international 10-10 electrode placement system. For each subject, the dataset provides 14 experimental runs, each corresponding to a distinct combination of task type (baseline, real movement, or motor imagery) and limb (fists or feet). Recordings were preprocessed and epoched using EEGLAB's `pop_epoch` function. Rather than using the full 64-channel montage, analysis in this study was restricted to a five-channel sensorimotor subset (C3, Cz, C4, CP3, CP4), selected for their established relevance to motor cortex activity and to keep feature dimensionality manageable.

B. Task 1: Eyes-Open versus Eyes-Closed

The first classification task uses runs R01 (eyes open, baseline) and R02 (eyes closed, baseline) for each subject, with each continuous run segmented into non-overlapping 4-second epochs for feature extraction. This is a resting-state task with no motor component, and it is expected to be comparatively easier to classify, since eyes-closed recordings typically exhibit a strong and well-documented increase in occipital alpha-band (8–12 Hz) power relative to eyes-open recordings. This task therefore serves as a sanity-check baseline against which the more challenging movement task can be compared.

C. Task 2: Motor Imagery

The second classification task is a binary discrimination between resting-state segments (labeled T0) and movement segments in which the subject either physically performs or imagines performing a left- or right-fist movement (labeled T1 and T2 respectively, here pooled together as a single ”movement” class). This task draws on runs R03, R04, R07, R08, R11, and R12 of the EEGMMIDB protocol, which correspond to the unilateral fist condition and include both real-movement runs (R03, R07, R11) and motor-imagery runs (R04, R08, R12). Combining executed and imagined movement in this way follows directly from the pipeline structure used in this study and is noted here explicitly since it distinguishes this task from a pure motor-imagery formulation. Unlike Task 1, epochs for this task were extracted as fixed 2-second windows locked to the onset of each T0, T1, or T2 event marker, rather than by segmenting continuous recordings. Classification on this task is substantially harder than the eyes-state task, since the underlying event-related desynchronization/synchronization (ERD/ERS) patterns in the mu (8–12 Hz) and beta (13–30 Hz) bands are subtler, more spatially localized, and known to vary considerably across subjects.

III. Methods

This section describes the exact implementation of the two feature extraction pipelines, the feature selection procedure, the classifiers, and the validation methodology, as applied in this study.

A. Spectral Features

For each 4-second epoch (Task 1) or 2-second event-locked epoch (Task 2), the power spectral density (PSD) of each of the five selected channels (C3, Cz, C4, CP3, CP4) was estimated in MATLAB using `pwelch` with default window, overlap, and FFT length settings – this measures how much signal power is present at each frequency. The resulting PSD was then reduced to a feature vector by computing the mean power within a series of twelve 1 Hz-wide bins spaced every 2 Hz across the 8–30 Hz range (bin centers at 8, 10, 12, ..., 30 Hz), i.e., for each bin edge f_b, the feature was $\text{mean}\{P(f) : f_b \leq f < f_b + 1\}$. This was repeated independently for each of the five channels, yielding $12 \times 5 = 60$ spectral features per epoch.

1) Common Spatial Patterns (CSP): An Exploratory Alternative:

A separate script implementing Common Spatial Patterns (CSP) as an alternative spectral representation was also tested for the movement classification task. CSP-based features did not yield improved classification accuracy over the `pwelch`-based representation described above in preliminary experiments, and CSP was therefore not adopted in the final pipeline; all spectral results reported in Section IV use the Welch PSD features described above.

B. Cepstral Features

The cepstral feature pipeline used in this study follows the filterbank–log–DCT–MVGPDF method described by Bhattacharyya and Mukul [2]. For each epoch, the following steps were applied independently to each of the five channels:

1) **Framing**: Each epoch was divided into `NUM_FRAMES` $= 7$ overlapping frames with `OVERLAP` $= 0.5$ (50%), using a frame length computed as $\text{win\_len}/(7 \times 0.5 + 0.5)$ samples and a hop size of half the frame length. Each frame was windowed with a Hamming window before further processing – a standard tapering function that smooths the edges of each frame to reduce artifacts before taking its Fourier transform.
2) **Power spectrum**: For each windowed frame, an FFT of length $\text{NFFT} = 2^{\lceil \log_2(\text{WIN\_SEC} \times f_s) \rceil}$ was computed (with `WIN_SEC` $= 4$), and the squared magnitude of the first half of the spectrum (up to the Nyquist frequency) was retained as the frame's power spectrum, i.e., how much signal energy is present at each frequency.
3) **Triangular filterbank**: A bank of `NUM_FILT` $= 20$ overlapping triangular filters was constructed with center frequencies evenly spaced between `FREQ_LO` $= 8$ Hz and `FREQ_HI` $= 30$ Hz. Each filter's weight rises linearly from 0 to 1 between its left edge and center frequency, then falls linearly from 1 back to 0 between its center and right edge, forming a triangular shape; this smooths and pools energy from neighboring frequencies into 20 broader bands rather than treating each frequency independently. The power spectrum of each frame was projected onto this filterbank via matrix multiplication to obtain 20 filterbank energies per frame.
4) **Logarithmic compression**: The natural logarithm of each filterbank energy was taken (with a small additive constant, 10^{-10}, to avoid taking the log of zero). This compresses large energy differences into a smaller, more comparable range, similar to how loudness is perceived logarithmically rather than linearly.
5) **Discrete Cosine Transform (DCT)**: The 20 log filterbank energies of each frame were transformed via a custom DCT-II matrix into `NUM_CC` $= 12$ cepstral coefficients, using $\text{DCT}[n, m] = \cos(\pi n(m - 0.5)/20)$ for coefficient index $n = 1, \ldots, 12$ and filter index $m = 1, \ldots, 20$. This step reduces 20 correlated filterbank values down to 12 largely independent coefficients, concentrating the most informative shape of the spectrum into fewer numbers.

6) **Multivariate Gaussian probability density function (MVGPDF) modeling**: Across the 7 frames of an epoch, the mean vector μ (12-dimensional) and regularized covariance matrix $\Sigma = \mathrm{cov}(\mathrm{CC}) + 10^{-6}I$ of the frame-wise cepstral coefficient vectors were computed. The multivariate Gaussian density was then evaluated at each frame's coefficient vector relative to this fitted mean and covariance, producing one density value per frame that summarizes how typical or unusual that frame's coefficients are relative to the epoch as a whole.

The final cepstral feature vector for each channel consisted of the mean of the 7 per-frame MVGPDF density values together with the 12-dimensional mean cepstral coefficient vector μ, i.e., 13 features per channel, giving $13 \times 5 = 65$ cepstral features per epoch in total.

C. Feature Selection

Within each of the 5 cross-validation folds, `fscmrmr` (MATLAB's built-in Minimum Redundancy Maximum Relevance implementation) was run separately on the spectral training features and the cepstral training features, using only the training partition of that fold. `fscmrmr` ranks features by how strongly each one relates to the class label while penalizing features that are redundant with ones already selected, so the retained set is both informative and non-repetitive. The top 20 ranked features (or fewer, if fewer than 20 were available) were retained, and the same feature indices selected on the training data were then applied to the corresponding test partition. This fold-local selection was applied independently to each of the two feature sets, so the specific features retained could differ from fold to fold.

D. Classifiers

For each fold and each feature type, two classifiers were trained on the selected training features and evaluated on the corresponding test features:

- **SVM**: trained via `fitcsvm` with a linear kernel (`KernelFunction = 'linear'`), automatic kernel scale (`KernelScale = 'auto'`), box constraint $C = 1$, and standardized input features (`Standardize = true`).
- **LDA**: trained via `fitcdiscr` with `DiscrimType = 'linear'` (i.e., a shared-covariance linear discriminant classifier).

This produced four accuracy values per fold: SVM-Spectral, LDA-Spectral, SVM-Cepstral, and LDA-Cepstral.

E. Validation Procedure and Statistical Testing

All epochs across all subjects were pooled and partitioned into 5 stratified folds using MATLAB's `cvpartition(Y_all, 'KFold', 5)` with a fixed random seed (`rng(42)`) for reproducibility. For each fold, test-set classification accuracy was recorded for all four feature-classifier combinations, yielding a 5×4 accuracy matrix per task. To test whether cepstral features significantly outperformed spectral features, two separate one-way ANOVAs (`anova1`) were run per task: one comparing the 5 SVM-Spectral accuracies against the 5 SVM-Cepstral accuracies, and one comparing the 5 LDA-Spectral accuracies against the 5 LDA-Cepstral accuracies. Pairwise mean differences and their confidence intervals were obtained via `multcompare` on the resulting ANOVA statistics.

IV. Results

A. Eyes-Open versus Eyes-Closed

Table I summarizes 5-fold cross-validation accuracy for the eyes-open versus eyes-closed task across both feature types and both classifiers. Cepstral features outperformed spectral features for both classifiers: SVM accuracy increased from $67.19\% \pm 1.53\%$ (spectral) to $69.08\% \pm 0.92\%$ (cepstral), a mean improvement of approximately 1.9 percentage points ($p = 0.0450$), while LDA accuracy increased from $65.23\% \pm 2.69\%$ (spectral) to $69.11\% \pm 0.97\%$ (cepstral), a mean improvement of approximately 3.9 percentage points ($p = 0.0161$). Both differences are statistically significant at the $\alpha = 0.05$ level. Notably, spectral-feature accuracy was also more variable across folds (particularly for LDA, with a standard deviation of 2.69 percentage points) than cepstral-feature accuracy, which was consistently in the 68–70% range across all folds for both classifiers.

TABLE I
5-Fold CV Accuracy: Eyes-Open vs. Eyes-Closed

Feature Type	Classifier	Accuracy (%)	ANOVA p
Spectral	SVM	67.19 ± 1.53	0.0450
Cepstral	SVM	69.08 ± 0.92	
Spectral	LDA	65.23 ± 2.69	0.0161
Cepstral	LDA	69.11 ± 0.97	

B. Movement Task (T0 vs. T1+T2)

For the movement task, a total of 19,690 epochs were extracted, balanced evenly between the rest class (T0: 9,845 epochs) and the combined movement class (T1+T2: 9,845 epochs). Table II summarizes 5-fold cross-validation accuracy for this task. As expected, overall accuracy is substantially lower than for the eyes-state task, consistent with the greater difficulty of decoding movement-related activity relative to a resting baseline. Cepstral features again outperformed spectral features for both classifiers: SVM accuracy increased from $52.31\% \pm 0.50\%$ (spectral) to $55.22\% \pm 1.43\%$ (cepstral), a mean improvement of approximately 2.9 percentage points ($p = 0.0026$), while LDA accuracy increased from $52.80\% \pm 0.62\%$ (spectral) to $55.24\% \pm 1.47\%$ (cepstral), a mean improvement of approximately 2.4 percentage points ($p = 0.0093$). Both differences are statistically significant at the $\alpha = 0.05$ level, and in fact the SVM p-value here is smaller than for the eyes-state task, indicating a comparatively more consistent cepstral advantage across folds for this classifier despite the lower absolute accuracy.

TABLE II
5-FOLD CV ACCURACY: MOVEMENT VS. REST (T0 VS. T1+T2)

Feature Type	Classifier	Accuracy (%)	ANOVA p
Spectral	SVM	52.31 ± 0.50	0.0026
Cepstral	SVM	55.22 ± 1.43	
Spectral	LDA	52.80 ± 0.62	0.0093
Cepstral	LDA	55.24 ± 1.47	

C. Summary of Findings

Across both tasks and both classifiers, cepstral features consistently and significantly outperformed spectral features, with all four ANOVA comparisons reaching statistical significance ($p < 0.05$). The magnitude of the cepstral advantage was larger for LDA than for SVM on the eyes-state task (3.9 vs. 1.9 percentage points), whereas the two classifiers showed a more comparable advantage on the movement task (2.4 vs. 2.9 percentage points). Absolute accuracy was substantially higher for the eyes-state task (65–69%) than for the movement task (52–55%) across all feature-classifier combinations, reflecting the greater physiological separability of the eyes-open/eyes-closed conditions compared to the more subtle and subject-variable signatures of executed and imagined movement.

V. DISCUSSION

A. Why Cepstral Features Outperform Spectral Features

The consistent cepstral advantage across both tasks and classifiers suggests the filterbank–log–DCT–MVGPDF pipeline captures structure in the spectral envelope not fully accessible to raw per-bin power estimates. The triangular filterbank smooths the spectrum into overlapping bands, likely reducing sensitivity to noise in any single narrow 1 Hz bin, while the DCT decorrelates the log filterbank energies and concentrates variance into a few low-order coefficients, complementing MRMR selection. This may also explain the lower fold-to-fold variability observed for cepstral features – most notably for LDA on the eyes-state task, where the spectral standard deviation (2.69 points) was nearly three times the cepstral value (0.97 points), suggesting more consistent generalization rather than a few favorable folds.

B. Why Movement Classification Remains Difficult

Movement-task accuracy (52–55%) stayed far below the eyes-state task (65–69%) and only modestly above chance. Unlike the strong, spatially broad occipital alpha-band effect underlying eyes-open/eyes-closed, movement-related ERD/ERS patterns are weaker, more spatially localized, and vary considerably across subjects – a problem compounded here since all 109 subjects were pooled into a single cross-subject model rather than subject-specific classifiers. Combining executed movement (R03, R07, R11) with imagined movement (R04, R08, R12) into one class likely adds further within-class heterogeneity, since executed movement typically produces stronger sensorimotor activation than imagined movement.

C. Limitations

Three limitations should be noted. First, cross-validation was performed at the epoch level rather than with subject-level splits, so a subject's epochs may appear in both train and test partitions, potentially inflating accuracy relative to fully unseen subjects. Second, analysis used only a five-channel sensorimotor subset (C3, Cz, C4, CP3, CP4) rather than the full 64-channel montage. Third, an exploratory CSP implementation did not outperform Welch PSD for the movement task; further tuning may be needed before drawing firm conclusions about its merit here.

VI. CONCLUSION

This paper compared spectral (Welch PSD) and cepstral (filterbank–log–DCT–MVGPDF) features for two binary EEG classification tasks on PhysioNet EEGMMIDB: eyes-open versus eyes-closed, and movement versus rest (executed and imagined unilateral fist movement combined). With MRMR feature selection performed within cross-validation folds and both SVM and LDA classifiers, cepstral features significantly outperformed spectral features across both tasks and classifiers (1.9–3.9 percentage-point improvements, ANOVA $p < 0.05$ throughout). The eyes-state task was classified substantially more accurately (65–69%) than the movement task (52–55%), reflecting the stronger physiological basis of the eyes-state contrast relative to the subtler, subject-variable signatures of movement.

These results support cepstral features, originally developed for speech processing, as a viable and here superior alternative to spectral features for EEG classification. Future work could use subject-level cross-validation splits, the full 64-channel montage, further-tuned CSP as an additional baseline, and separate classification of executed versus imagined movement.

REFERENCES

[1] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: A general-purpose brain-computer interface (BCI) system," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1034–1043, 2004. [Online]. Available: https://www.physionet.org/content/eegmmidb/1.0.0/

[2] S. Bhattacharyya and A. Mukul, "Cepstrum based feature extraction method for EEG signal classification," *International Journal of Engineering and Technology (IJET)*, vol. 8, no. 1, 2016. [Online]. Available: https://www.enggjournals.com/ijet/docs/IJET16-08-01-008.pdf