# A comparison between different Multiclass Common Spatial Pattern approaches for identification of motor imagery tasks

Cecilia Lindig-León, Laurent Bougrain

# A comparison between different Multiclass Common Spatial Pattern approaches for identification of motor imagery tasks

cecilia.lindig-leon@inria.fr, laurent.bougain@loria.fr
Université de Lorraine, LORIA, UMR 7503, Vandoeuvre-lès-Nancy, F-54600

## Abstract

Common Spatial Patterns (CSP) is a feature extraction method suited for two-class problems. However, there are some alternatives to apply it for multiclass tasks by using a group of ensemble classifiers that divide the problem into different binary classification tasks, from which the final decision is inferred as the combination of their responses. Nevertheless, there is another approach to extend CSP for multiple classes to a one-step-method by approximating the joint diagonalization of their covariance matrices [1] (fig. 1). In this study, in order to identify whether CSP by Joint Approximate Diagonalization (JAD) represents an outperforming alternative to the standard multiclass CSP alternatives, four different methods were applied on dataset 2a used in BCI competition IV [2] (fig. 2).

**Join Approximate Diagonalization (JAD) algorithm**

**INPUT:** $Cov^k$ **(covariance matrices to be diagonalized)**
$W_{(1)} \leftarrow 0,\ CSPmatrix_{(1)} \leftarrow I,\ n \leftarrow 1$
$Cov^k_{(1)} \leftarrow CSPmatrix_{(1)} Cov^k CSPmatrix_{(1)}^T$
**repeat**
  *compute* $W_{(n)}$ *from* $Cov^k_{(n)}$
  **if** $\| W_{(n)} \|_F > \theta$ *then*
    $W_{(n)} \leftarrow \frac{\theta}{\| W_{(n)} \|_F} W_{(n)}$
  **end if**
  $CSPmatrix_{(n+1)} \leftarrow (I + W_{(n)})\, CSPmatrix_{(n)}$
  $Cov^k_{(n+1)} \leftarrow (I + W_{(n)}) Cov^k_{(n)} (I + W_{(n)})^T$
  $n \leftarrow n+1$
**until** *convergence*
**OUTPUT: CSPmatrix**

Fig. 1. Pseudo-code describing the JAD algorithm

## Material and methods

Dataset 2a, BCI competition IV:

- 9 subjects, 4 motor imageries: left arm, right arm, feet and tongue.
- 22 EEG channels, sampling rate of 250 Hz, [8-30 Hz] bandpass Butterworth filter (order 5), 576 trials per subject (2 sessions, 6 runs per session, 1 run=12 trials/class. Session 2 contains the unseen data.

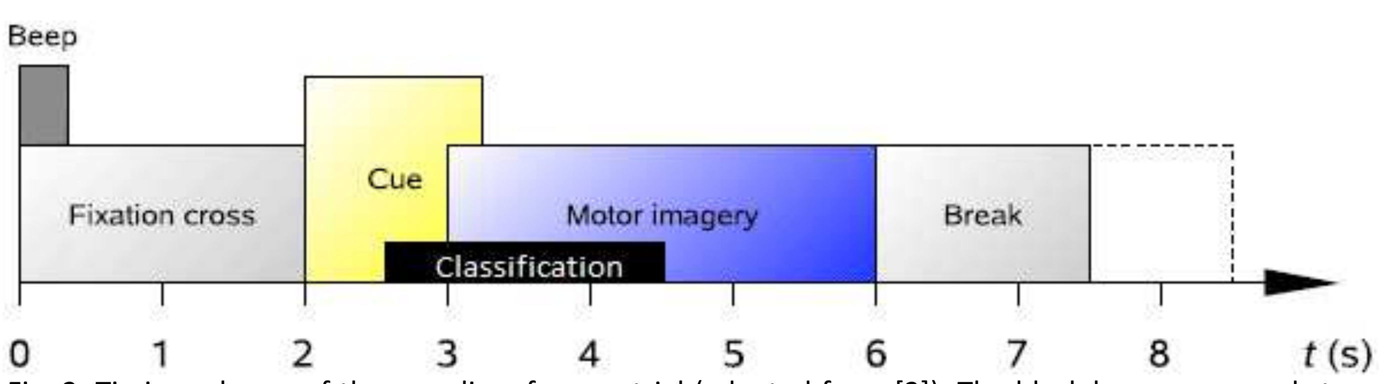


Fig. 2. Timing scheme of the paradigm for one trial (adapted from [2]). The black box corresponds to the used window to extract classification data.

| Cross Validation | | | | |
|---|---|---|---|---|
| | Method | | | |
| subject | oneVSone | oneVSall | BCSP | CSP by JAD |
| 1 | 0.793 ± 0.062 | 0.747 ± 0.062 | 0.772 ± 0.071 | 0.746 ± 0.057 |
| 2 | 0.379 ± 0.038 | 0.394 ± 0.056 | 0.425 ± 0.053 | 0.378 ± 0.042 |
| 3 | 0.880 ± 0.066 | 0.902 ± 0.071 | 0.805 ± 0.057 | 0.845 ± 0.065 |
| 4 | 0.439 ± 0.044 | 0.429 ± 0.066 | 0.494 ± 0.054 | 0.357 ± 0.037 |
| 5 | 0.198 ± 0.029 | 0.214 ± 0.036 | 0.388 ± 0.059 | 0.176 ± 0.027 |
| 6 | 0.424 ± 0.043 | 0.410 ± 0.052 | 0.440 ± 0.066 | 0.445 ± 0.041 |
| 7 | 0.812 ± 0.061 | 0.753 ± 0.079 | 0.773 ± 0.045 | 0.798 ± 0.060 |
| 8 | 0.939 ± 0.068 | 0.911 ± 0.066 | 0.793 ± 0.083 | 0.830 ± 0.064 |
| 9 | 0.871 ± 0.066 | 0.869 ± 0.071 | 0.790 ± 0.060 | 0.842 ± 0.063 |
| AVG | 0.637 ± 0.053 | 0.626 ± 0.062 | 0.631 ± 0.061 | 0.602 ± 0.051 |

Table 1. 10 by 10-fold cross validation performance in terms of the kappa value using one-vs-one; one-vs-all; BCSP and CSP by JAD methods to extract features from the training dataset. Classification was carried out by LDA.

The CSP Joint Approximate Diagonalization (JAD) method has been compared to the standard one-vs-one; one-vs-all and hierarchical (namely Binary-Common Spatial Patterns, BCSP, [3]) (Fig. 3) approaches. For feature extraction 8 filters were selected to perform CSP by JAD (see [1]) and this filter amount was kept the same for the other methods. Each classification task was carried out by using Linear Discriminant Analysis (LDA). Performance was evaluated in terms of the mean kappa value using 10x10-fold cross-validation (Table 1) and a session-to-session test (Table 2) over the referred dataset.

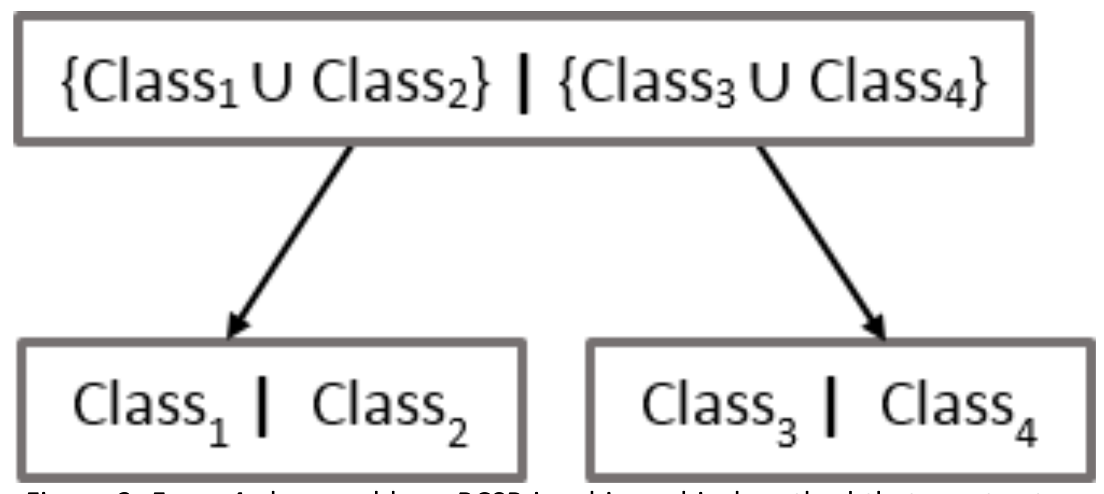


Figure 3. For a 4-class problem, BCSP is a hierarchical method that creates two new groups gathering together two classes each to design a first-step classification stage. Once that an unseen data has been identify as belonging to one of the new compound groups, another classification stage is used to determine the resulting label. In this study the first compound class corresponds to the union of left-hand and right-hand, and the second one, to the union of feet and tongue. For the second classification stage two classifiers to separate both compound classes were used (see [3]).

| Test | | | | |
|---|---|---|---|---|
| | Method | | | |
| subject | oneVSone | oneVSall | BCSP | CSP by JAD |
| 1 | 0.694 | 0.694 | 0.583 | 0.722 |
| 2 | 0.259 | 0.278 | 0.347 | 0.236 |
| 3 | 0.745 | 0.690 | 0.644 | 0.639 |
| 4 | 0.454 | 0.444 | 0.468 | 0.421 |
| 5 | 0.204 | 0.083 | 0.269 | 0.144 |
| 6 | 0.236 | 0.324 | 0.185 | 0.287 |
| 7 | 0.634 | 0.662 | 0.699 | 0.537 |
| 8 | 0.732 | 0.750 | 0.523 | 0.657 |
| 9 | 0.764 | 0.727 | 0.722 | 0.690 |
| AVG | 0.5247 | 0.517 | 0.493 | 0.4815 |

Table 2. Performance in terms of the kappa value using one-vs-one; one-vs-all; BCSP and CSP by JAD methods to extract features from the the testing dataset. Classification was carried out by LDA.

| Significant difference at 95 confidence level | | | | |
|---|---|---|---|---|
| | oneVsone | oneVSall | BCSP | CSP by JAD |
| oneVSone | | ✓ | ✗ | ✓ |
| oneVSall | ✓ | | ✓ | ✓ |
| BCSP | ✗ | ✓ | | ✓ |
| CSP by JAD | ✓ | ✓ | ✓ | |

Table 3. Results obtained for the paired t-tests with a significance level at 5% applied.

## Conclusions

These results show that multiclass CSP by JAD does not outperform the classical methods, in fact its performance is poorer. This is presumably related to the fact that the performed diagonalization is approximated and in some cases deficient. The best performance during the cross-validation test was achieved by the BCSP method. Nevertheless, after validation using both sessions' data its performance drops down to the third rank and it is preceded respectively by the one-vs-one and one-vs-all approaches; which shows that the BCSP model was overtrained.

Table 3 shows that most of the methods are significant different between one to another.

## References

[1] Grosse-Wentrup M. and Buss M., *Multi-class common spatial pattern and information theoretic feature extraction*. IEEE transactions of biomedical Engineering (2008).
[2] BCI competition IV, https://www.bbci.de:competition/iv/resutls/index.html.
[3] Y. Fang, M. Chen, X. Zheng and R. F. Harrison, Extending CSP to detect motor imagery in a four-class BCI. *Journal of information and computational science* (2012).