

Article

Not peer-reviewed version

EEG Channel Selection Based on Time-Frequency Hellinger Distance

[Seungjun Lee](#) *

Posted Date: 12 February 2025

doi: 10.20944/preprints202502.0938.v1

Keywords: EEG; Electroencephalography; motor imagery; Hellinger Distance



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

EEG Channel Selection Based on Time–Frequency Hellinger Distance

Seungjun Lee

Korea Science Academy of KAIST, 105-47 Baegyangwanmun-ro, Busanjin-gu, 47162, Busan, Republic of Korea;
22-086@ksa.hs.kr

Abstract: This paper presents a novel method for EEG channel selection based on the Hellinger Distance (HD) computed over time–frequency representations (TFRs). Here, we first convert raw EEG into the time–frequency domain using Short-Time Fourier Transform (STFT) and Continuous Wavelet Transform (CWT). Leveraging Hellinger Distance, we then identify and remove outlier epochs via a z-score threshold, and rank and select the most discriminative channels by measuring how well each channel's TFR distributions separate different classes. Our empirical evaluation uses the BCI Competition IVA dataset to compare the proposed HD-based approach (HD-CSP) against multiple variants of the Common Spatial Pattern algorithm, including standard CSP, L1 Norm CSP, SCSP, FBCSP, and E-CSP. Results indicate that HD-CSP consistently outperforms competing methods in all tested configurations, achieving notably high classification accuracy even when the number of channels is severely restricted. In particular, HD-CSP reaches around 70% accuracy with only three channels, while other approaches suffer significant performance drops. As the number of channels increases, HD-CSP maintains its superior accuracy, exceeding 80% in some configurations. Overall, the proposed method is superior on performance gains and ability to adapt to diverse channel configurations suggest broad applicability, especially in resource-constrained EEG settings where efficiency and accuracy are both priorities.

Keywords: EEG; Electroencephalography; motor imagery; Hellinger Distance

Highlights

- Robust Noise Mitigation via Hellinger Distance: Demonstrates a two-phase process—outlier epoch removal and channel selection—using the Hellinger Distance in the time–frequency domain, thereby effectively reducing artifacts and improving data quality.
- Superior Classification Accuracy in Low-Channel Regime: Achieves around 70% accuracy with only three EEG channels, outperforming other CSP variants and highlighting suitability for portable or resource-constrained BCI applications.
- Scalable Performance with Increased Channels: Maintains consistently high accuracy (exceeding 80% in some configurations) as channel count grows, indicating broad applicability across diverse EEG setups.

1. Introduction

Electroencephalography (EEG) has gained increasing attention in brain-computer interface (BCI) applications and clinical diagnostics due to its non-invasive nature and ability to capture real-time neural activity [1,2]. Despite these benefits, EEG signals are characteristically non-stationary and prone to various sources of noise, such as physiological artifacts and external electromagnetic interference. These challenges often lead to substantial variability in the recorded data, complicating the extraction of reliable features for classification or clinical interpretation.

Channel selection, a process of identifying the most informative subset of electrodes, has emerged as a key strategy for managing this complexity. By reducing the dimensionality of the data, channel

selection not only streamlines computational overhead but can also enhance classification accuracy by excluding channels with predominantly noisy or irrelevant signals. Traditional approaches to channel selection rely on criteria such as mutual information or variance-based metrics; however, they may not be robust in capturing subtle changes in EEG signals' spectral content or in handling highly non-stationary properties [3].

To address these gaps, this study proposes a novel pipeline centered on the Hellinger Distance computed over time-frequency representations (TFRs) of EEG. In abstract, we first transform raw EEG signals into the time-frequency domain utilizing Short-Time Fourier Transform (STFT) and Continuous Wavelet Transform (CWT). Then, we employ Hellinger Distance to perform two critical steps: automatic detection and removal of outlier epochs, and ranking and selection of the most informative channels based on their ability to discriminate between different classes or conditions. By leveraging Hellinger Distance for both noise mitigation and feature selection, our framework provides a unified approach to boosting classification performance.

2. Analysis

The overall goal is in two phase: first, to establish a procedure for detecting and eliminating noisy epochs that may adversely affect classification performance; and second, to identify a minimal, yet informative set of channels that maximizes discriminative power.

We begin by revisiting the core principles of time-frequency representations (Section 2.1), emphasizing the importance of capturing transient spectral fluctuations in EEG signals. Next, we introduce Hellinger Distance as a measure of distributional divergence (Section 2.2), illustrating its dual role in noise detection (Section 2.3) and channel selection (Section 2.4).

2.1. Time-Frequency Representation of EEG Signals

Let EEG signal be $x \in \mathbb{R}^{N \times C \times T}$, where N is the number of samples, C is the number of channels, and T is the number of timepoints, and $x(t) \in \mathbb{R}^{N \times C}$ be the signal where $time = t$.

As EEG signals are non-stationary, meaning its distribution varies over time, we analyze it by transforming it into a time-frequency domain to transient spectral features. By utilizing Short-Time Fourier Transform(STFT), Time-Frequency Representation(TFR) of a signal $X(t, f)$ can be calculated as

$$X(t, f) = \int_{-\infty}^{\infty} x(\tau) w(\tau - t) e^{-j2\pi f \tau} d\tau \quad (1)$$

where $w(\cdot)$ is the window function centered at time τ [2]. Analogous to STFT, Continuous Wavelet Transform(CWT) can be utilized by calculating

$$W(s, t) = \int_{-\infty}^{\infty} x(u) \psi_{s,t}(u) du, \quad (2)$$

where the wavelet function $\psi_{s,t}(x) \equiv \frac{1}{\sqrt{s}} \psi^*\left(\frac{x-t}{s}\right)$. In this study, Morlet wavelet function was used, which is defined as

$$\psi(t) = \pi^{-\frac{1}{4}} e^{-\frac{1}{2}t^2} e^{-j\omega_0 t}, \quad (3)$$

where ω_0 is the central frequency and σ controls the time spread. [4].

Continuous Wavelet Transform(CWT) can be utilized using $\psi(t)$ as

$$W(a, b) \equiv \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi^*\left(\frac{t-b}{a}\right) dt, \quad (4)$$

where a is the inverse frequency and b is the time shift. [5]

2.2. Utilization of Hellinger Distance for Z-Score Calculation

Hellinger Distance is a measure of difference between two probability distributions. For two discrete distributions: $P = \{p_i\}$ and $Q = \{q_i\}$, Hellinger distance $H(P, Q)$ is defined as

$$H(P, Q) = \frac{1}{\sqrt{2}} \sqrt{\sum_i (\sqrt{p_i} - \sqrt{q_i})^2} \quad (5)$$

In this study, Hellinger distance is utilized in two different ways: noise removal and channel selection.

2.3. Noise Removal

For calculating the z-score for each epoch, we calculate H_i . where $H_i = H(W_i, \bar{W}_i)$ for each epoch i , where $\bar{W}_i$ represents overall average of W_i . This H_i serves as a measure of how discriminative epoch i 's features are between conditions. By subtracting μ_H , which is the average of H_i , and dividing with σ_H , which is the standard deviation of H_i , we get each epoch i 's Z-score:

$$Z_i = \frac{H_i - \mu_H}{\sigma_H} \quad (6)$$

Epochs having large Z-score indicate how extreme the signal of each trial is compared to other epochs. Therefore, we assume epochs having high Z-score as outliers(in this case noise), and remove it.

Therefore, we can define X^* , which indicates non-noisy trials, as

$$X^* = \{X_i | Z_i < \alpha\}, \quad (7)$$

where X_i indicates the raw time-domain EEG data corresponding to the i th epoch.

2.4. Channel Selection

After removing noisy epochs, the most informative EEG channels classification is identified. The goal of this process is to retain channels that maximize class separability while discarding those that contribute redundant or noisy information. This is achieved by computing the Hellinger Distance between class-specific time-frequency representations (TFRs) of each EEG channel.

The TFR features extracted from each channel serve as probability distributions of spectral power. Let the normalized power spectrum of a channel c for class 1 and class 2 be represented as $P_c(f)$ and $Q_c(f)$, respectively. The Hellinger Distance for a given channel c is computed as:

$$H_c = \frac{1}{\sqrt{2}} \sqrt{\sum_f \left(\sqrt{P_c(f)} - \sqrt{Q_c(f)} \right)^2}. \quad (8)$$

The computed Hellinger Distance H_c for each channel provides a measure of how well the channel distinguishes between the two classes. A higher H_c value indicates greater separability. To rank channels, we define the effect size E_c as:

$$E_c = \max_{k \in \text{class}} H_c(P_c^k, \bar{P}_c), \quad (9)$$

where $\bar{P}_c$ is the overall mean spectral distribution across all epochs. The channels are then sorted in descending order based on E_c .

3. Results

By [3], CSP based motor imagery classification algorithms can be enumerated as CSP, L1 Norm of CSP, SCSP, FBCSP, and E-CSP. The accuracy of each algorithm, including HD-CSP(Hellinger Distance CSP), is computed using the BCI competition IVa dataset, which is a publicly available dataset consisting of imagery of four classes: left hand movement, right hand movement, leg movement, and

tongue movement [1], to evaluate HD-CSP’s performance. Each subject’s dataset is divided in 5-fold cross manner for evaluation and training. CSP filter is applied to transform the signal, and train an Support Vector Machine (SVM) classifier.

In our comparative evaluation of multiple variants of CSP (standard CSP, L_1 Norm CSP, SCSP, FBCSP, E-CSP, and HD-CSP) under different number of EEG channels (Table 1), HD-CSP consistently outperformed competing approaches in all tested conditions, except one case in 11 channel configuration. The results show that HD-CSP not only yields superior classification accuracy overall but also maintains a stable advantage across different channel counts (Figure 1).

Table 1. Classification Accuracy among Various CSP Based EEG Channel Selection Methods for Different Numbers of Channels.

Number of Channels		1	2	3	4	5	6
CSPRank		0.243±0.002	0.427±0.026	0.476±0.079	0.493±0.068	0.51±0.056	0.517±0.068
L1 Norm applied CSPRank		0.243±0.002	0.344±0.055	0.433±0.038	0.458±0.044	0.451±0.032	0.453±0.014
SCSPRank		0.243±0.002	0.303±0.029	0.361±0.071	0.384±0.065	0.365±0.055	0.422±0.019
E-CSP		0.385±0.054	0.406±0.036	0.451±0.044	0.458±0.043	0.49±0.059	0.601±0.053
HD-CSP		0.396±0.019	0.556±0.049	0.716±0.068	0.711±0.053	0.708±0.061	0.709±0.084
7	8	9	10	11	12	13	14
0.451±0.018	0.417±0.015	0.479±0.048	0.49±0.07	0.518±0.053	0.514±0.076	0.538±0.051	0.524±0.06
0.434±0.031	0.427±0.02	0.423±0.022	0.439±0.026	0.416±0.03	0.447±0.059	0.446±0.032	0.48±0.088
0.43±0.028	0.462±0.029	0.464±0.05	0.488±0.052	0.477±0.058	0.45±0.082	0.45±0.039	0.468±0.069
0.621±0.071	0.673±0.055	0.701±0.057	0.691±0.063	0.753±0.051	0.722±0.066	0.733±0.071	0.733±0.033
0.739±0.079	0.736±0.043	0.731±0.056	0.758±0.048	0.743±0.044	0.761±0.079	0.767±0.043	0.768±0.058
15	16	17	18	19	20	21	22
0.458±0.082	0.462±0.065	0.469±0.042	0.583±0.048	0.635±0.051	0.736±0.063	0.778±0.051	0.809±0.028
0.44±0.063	0.438±0.064	0.47±0.026	0.526±0.062	0.579±0.07	0.607±0.069	0.738±0.032	0.755±0.032
0.516±0.059	0.571±0.042	0.551±0.07	0.522±0.08	0.648±0.051	0.712±0.018	0.726±0.018	0.75±0.048
0.736±0.061	0.754±0.059	0.747±0.054	0.75±0.047	0.75±0.041	0.754±0.06	0.761±0.057	0.757±0.056
0.778±0.04	0.778±0.061	0.791±0.049	0.792±0.033	0.814±0.05	0.802±0.049	0.812±0.031	0.817±0.043

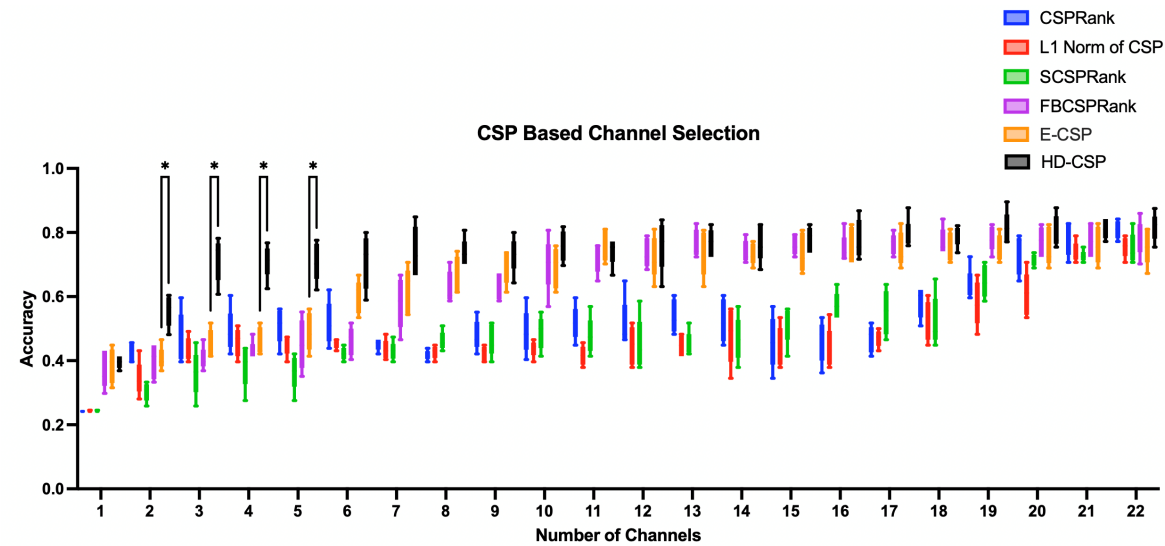


Figure 1. Development of CSP Based Channel Selection Methods Across Different Number of Channel Configuration.

Notably, in the most restrictive scenario of only three channels, HD-CSP still manages to reach around 70% classification accuracy. This outcome is particularly compelling because many methods suffer a significant drop in performance when the dimensionality of the EEG data is limited. Except for this model, the rest of the models show an accuracy of 36% - 46%. The robust performance of HD-CSP in this low-channel regime illustrates its suitability for practical applications where hardware or time constraints require the use of fewer electrodes.

Moreover, even as the number of channels increases, HD-CSP remains consistently superior. Its classification accuracy continues to surpass that of standard CSP, L_1 Norm CSP, SCSP, FBCSP, and E-CSP for all tested configurations, reaching accuracy of 80% in 19 channel configuration, which none of other models have reached. These findings highlight HD-CSP's efficacy in balancing feature extraction quality and noise handling, suggesting it can serve as a strong default method for both minimal-channel and full-cap montage scenarios.

References

1. Brunner, C.; Leeb, R.; Müller-Putz, G.; Schlögl, A.; Pfurtscheller, G. BCI Competition 2008–Graz data set A. *Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology* **2008**, *16*, 1–6.
2. Goyal, D.; Pabla, B.S. Condition based maintenance of machine tools—A review. *CIRP Journal of Manufacturing Science and Technology* **2015**, *10*, 24–35. <https://doi.org/https://doi.org/10.1016/j.cirpj.2015.05.004>.
3. Lee, S. A Mathematical Review on EEG Channel Selection Techniques for Motor Imagery Classification, 2025. <https://doi.org/10.21203/rs.3.rs-5822489/v1>.
4. Wallisch, P.; Lusignan, M.; Benayoun, M.; Baker, T.I.; Dickey, A.S.; Hatsopoulos, N.G. Chapter 9 - Wavelets. In *Matlab for Neuroscientists*; Wallisch, P.; Lusignan, M.; Benayoun, M.; Baker, T.I.; Dickey, A.S.; Hatsopoulos, N.G., Eds.; Academic Press: London, 2009; pp. 133–140. <https://doi.org/https://doi.org/10.1016/B978-0-12-374551-4.00009-9>.
5. Dişli, F.; Gedikpınar, M.; Fırat, H.; Şengür, A.; Güldemir, H.; Koundal, D. Epilepsy Diagnosis from EEG Signals Using Continuous Wavelet Transform-Based Depthwise Convolutional Neural Network Model. *Diagnostics (Basel, Switzerland)* **2025**, *15*. <https://doi.org/10.3390/diagnostics15010084>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.