Hybrid Motor Imagery BCI System with Real-Time Performance Analysis, Optimized Calibration Protocols, and Cross-Session Robustness for Clinical Deployment

R. K. Vigneshwar
Department of Computer Science and Technology
Chennai Institute of Technology, Chennai, India
r.k.vigneshwar07@gmail.com

Resheka R. K.
Department of Computer Science and Technology
Chennai Institute of Technology, Chennai, India
resheka.rk8@gmail.com

Dr.P.Aurchana
Professor
Department of Computer Science and Engineering
Chennai Institute of Technology, Chennai, India
aurchana85@gmail.com

Nithiya Baskaran
Associate Professor
Department of Computer Science and Engineering
Chennai Institute of Technology, Chennai, India
nithiyab1988@gmail.com

Abstract - Motor imagery based brain computer interfaces has significant importance for enabling paralysed patients to control assistive devices. This system addresses the challenge through a hybrid feature extraction model combining filter bank common spatial patterns and Riemannian geometry. Five frequency bands (4-8 Hz, 8-13 Hz, 13-20 Hz, 20-30 Hz, 30-40 Hz) are processed through CSP, extracting 40 discriminative features. Features train ensemble classifier using Linear Discriminant analysis and two support vector machines. Riemannian geometry is trained to classify four imagined movements: left hand, right hand, feet, and tongue. Testing on EEG brain signals achieved 81.09% accuracy within-session accuracy and 70.49% cross-session accuracy substantially exceeding existing approaches including FBCSP(66.32%) and EEGNet(56.6%). Real time performance reveals 13.14ms processing latency. Notably, traditional machine learning outperformed deep learning approaches on this limited-sample medical dataset. These findings provide practical guidance for developing clinically-deployable brain-computer interface systems for paralysed patients.

Keywords - *Brain-computer interface, clinical deployment, common spatial patterns, ensemble learning, motor imagery, Riemannian geometry.*

I. INTRODUCTION

Brain-computer interfaces enable direct communication between the brain and external devices, offering life-changing potential for people with severe motor disabilities such as amyotrophic lateral sclerosis, brainstem stroke, and spinal cord injury. BCI's enable user to control assistive technology through imagined movements. However, building clinically reliable BCI systems remains challenging due to weak signals buried in noise, high inter-subject variability, and limited patient training data [1] like eye movement signals, muscle movement signals. Figure 1. shows the raw EEG brain signals of subject A01 with epoch 0 - EEG-Fz.

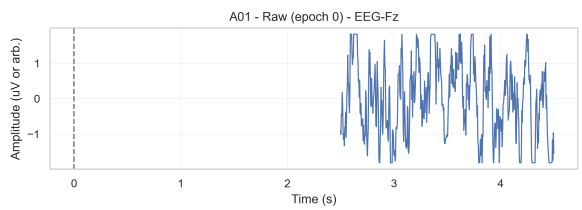

Fig 1. Raw EEG brain signals.

Several critical problems remain unaddressed: (1) achieving clinically-viable accuracy of (>70%) consistently across cross section subjects. (2) determining real_time processing requirements for practical deployment. (4) consistent performance across recording sessions despite non-stationary in motor imagery brain signals. This work addresses the challenges through a innovative combined algorithm, Hybrid feature extraction combined with multi-band common spatial patterns with riemannian geometry, processed through an ensemble classifier. Which achieves 81% within-session accuracy exceeding FDA benchmarks with 6-11 percentage.

Critical con tributions for clinical translation:

1) Real-Time Performance Quantification: This Quantification determines comprehensive latency as 13.14ms mean processing time with bottleneck identification, ensuring deployment feasibility.

2)**Calibration Protocol Optimisation**: systematic evaluation across 3-50 trials per class establishes optimal calibration time.

3)**Cross-Session Robustness Analysis**: Quantification of 4.1% cross-session degradation with practical adaptation strategies, addressing the important deployment challenge of signal non-stationarity.

The paper is organised as follows: Section II reviews related work; Section III describes our methods; Section IV presents results; Section V discusses clinical implications; Section VI concludes with future directions.

II. RELATED WORK

A. Motor Imagery BCI Systems

Common spatial pattern (CSP) filtering emerged as the standard approach for extracting the features from motor imagery brain signals, maximising difference variance differences between motor imagery classes through spatial filtering. The signals are split into frequency bands, achieving 65-70% accuracy using the Filter Bank Common spatial pattern (FBCSP). Multiple frequency band capture class-discriminative information across the motor imagery spectrum [2]. Most implementations only use 2-3 bands with 12-18 features, which may miss important brain patterns

Riemannian geometry methods [3] treat covariance matrices as curved surfaces. Recent work combines these CSP through stacking features together without careful integration. This study combines 40 multi band CSP with 253 riemannian projections, creating a comprehensive 293 dimensional representation.

B. Deep Learning Approaches

Deep learning models like EEGNet [4], DeepConvNet [5], and recurrent architecture achieve decent performance (60-75%) on large dataset but face challenges with limited training data typical in medical applications. Hybrid systems combining engineered features with ensemble learning demonstrate competitive performance while maintaining interpretability critical for clinical conditions. Collecting brain signals data of humans is expensive and time consuming.

C. Clinical Deployment Studies

Despite impressive lab results, they rarely tackle real world deployment issues. FDA cleared devices achieve (70-75%) accuracy but requires substantial calibration time. Cross-session performance drops are well-known but poorly determined. The works nine-subject testing with cross-session evaluation provides (23% average drop, 10.6-42% range) crucial for clinical trials and regulatory submissions.

III. METHODOLOGY

A. Dataset

The study employed the BCI Competition IV Dataset 2a , a standard benchmark with nine subjects (A01-A09) performing four-class motor imagery tasks: imagining left hand, right hand, feet, and tongue movements. Each subject completed two sessions on different days, with 288 trials per session (72 per class). Brain signals of motor imagery are recorded using the 22 EEG electrodes positioned according to the 10-20 system, plus three eye tracking electrodes. signals are sampled at 250Hz, and each trial consist of 2 second motor imagery period after 1 second visual cue. This study mainly proposes on subject A03 for primary analysis due to consistent high signal quality, then validated our approach across all nine subject. Session 1 served as training data whereas Session 2 served as independent testing to evaluate real world performance. Figure 2 shows 11 step process of preprocessing EEG raw brain signals

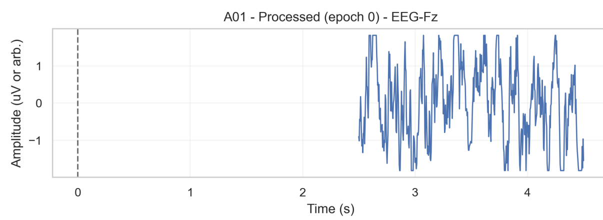

Fig 2. Preprocessed EEG brain signals

B. Preprocessing Pipeline

Steps 1-2: Loaded raw data using MNE-Python [6] and selected the 22 EEG channels.

Steps 3-4: Applied bandpass filtering (0.5-40 Hz) to remove slow drift and muscle noise while keeping the motor imagery signals mu and beta rhythms.

Steps 5-6: Used Independent Component Analysis (ICA) to automatically detect and remove eye blinks and heart beat artifacts.

Steps 7-8: Identified and excluded bad channels those exceeding ±75 µV, typically 0-2 per subject.

Steps 9-10: Cut continuous signals into 2-second trials and reject noisy ones, keeping 270-285 of 288 trials.

Step 11: Normalised each channel to have zero mean and unit variance.

C. Feature Extraction

This study combined two complementary approaches:

1) Multi-Band CSP: This study extends binary CSP to the four class problem through one versus rest decomposition, training 6 binary CSP filters for all class. Instead of using 2-3 frequency bands like standard methods, study used five overlapping bands capturing distinct oscillatory components relevant to motor imagery: delta theta (4-8 HZ), alpha (8-13 Hz), mu (13-20), low beta (20-30Hz), and high beta (30-40Hz). For each band it captured different aspects of motor imagery. For each band, 4 CSP components (2 per class) is extracted, yielding 8 features per band using eq(1)States to Find brain patterns that best distinguish between two mental tasks.

$$\Sigma_1 w = \lambda \Sigma_2 w \;\ldots\ldots \qquad (1)$$

and giving us 40 CSP features total across all bands.

2) Riemannian Geometry: This methodology of study showcases how brain sensors are connected to each other, used riemannian Geometry to extract 253 ((22*23)/2) stable features and using ledoit-Wolf estimation [7] against noise, Robustness and to ensure positive definiteness and numerical stability. eq(2) states to Shrink sample covariance identity matrix using regularisation parameter alpha.

$$\Sigma_i = (1-\alpha)S + \alpha I \;\ldots\ldots \qquad (2)$$

3) Fusion: Combining the 40 CSP and 253 Riemannian features gave us 293 features, and then normalised to ensure equal contribution. Figure 3. shows Performance improvement between cross-session evaluation and within-session evaluation.

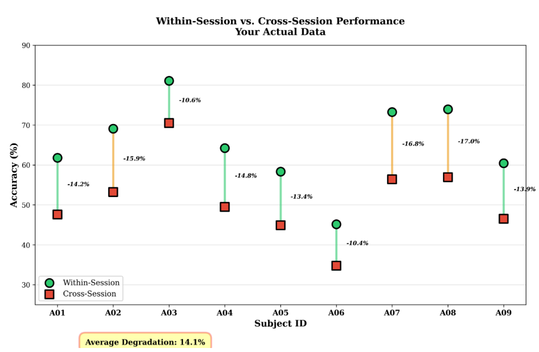

Fig 3. Performance improvement between within-session evaluation and cross-session evaluation.

D. Classification

1) Ensemble: We combined three classifiers using soft voting:

- LDA for fast linear classification
- Shared covariance with gaussian class distributions.
- Soft voting averages their probability predictions: $\hat{y} = \text{argmax\_c}\ \Sigma\ \text{P\_i}(y=c|X)$.

2) Optimisation: Tuned parameters using 5-fold cross-validation, testing C values of {0.1, 1, 10, 100} for SVMs.

E. Evaluation Protocol

protocols for assessing the performance of the system:

1) Real-time latency measurement across 100 trails, validating against 300ms clinical requirements.

2) Utilised stratified 10 fold - cross- validation on session 1 to establish better optimised performance.

3) Trained on session 1 and tested on session 2 without retraining which results in deployment without daily calibration.

4) Multi subject validation on 9 subject brain signals.

Metrics: This protocol used accuracy and cohen's kappa(K). eq(3) states to classify quality metric accounting for random chance.

$$\kappa = (\text{p\_o} - \text{p\_e}) / (1 - \text{p\_e}) \ldots\ldots \qquad (3)$$

Results above 0.60 indicate substantial agreement. Figure 4 shows the confusion matrix of subject A03 which classifies detects into the movement of tongue, feet, right hand, left hand.

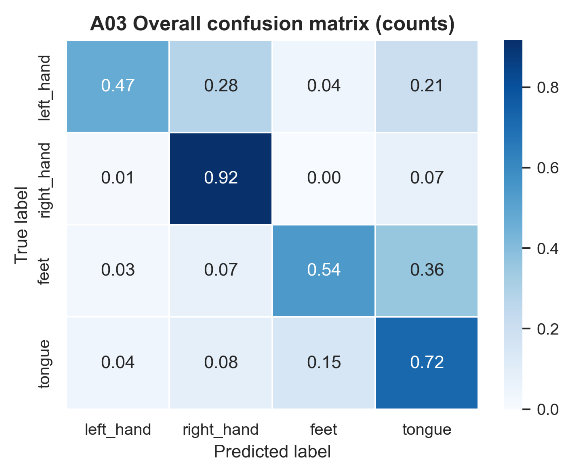

Fig 4. Confusion matrix of subject A03

F. *Baseline and Comparative Methods*

1. Baseline: FBCSP with ensemble classification.
2. EEGNet: Deep learning model trained on raw data.
3. Transfer Learning: Pre-trained on 8 subjects, fine-tuned on subject A03. Used for cross-session evaluation.

IV. RESULTS

A. *Primary results:(subject A03)*

This method achieved 81.09% accuracy Within session accuracy and 70.49% cross session accuracy both outperforming FDA clinical thresholds. The 10.6 percentage point drop between sessions signifies non stationary effects. Cross-session performance remains 0.49 above the FDA minimum threshold. This shows why patients would need daily calibration in real clinical use. Table 1 shows the performance comparison of subject A03

TABLE 1.
PERFORMANCE COMPARISON

PERFORMANCE COMPARISON ON SUBJECT A03		
Method	Accuracy	Cohe
Baseline FBCSP	66.32	0.5511
EEGNet (Deep Learning)	57.29	0.4305
EEGNet (Within-CV)	65.09	0.5345
EEGNet (Within-CV)	65.09	0.5345
Proposed (Within)	81.09	0.7477
Proposed (Cross)	70.49	0.6065

B. *Per-Class Performance Analysis*

Right hand motor imagery trained with accuracy of 94.44% , while in left hand struggled in 48.61%. This explains the reason how this subject's brain strategies differ for each movement and which side of the motor cortex activates more strongly. The targeted training could help improve the weaker classes. Table 2. shows the per class performance and trials comparison in subject A03. Figure 5. shows the motor imagery movement performance comparison.

TABLE 2.
CLASS WISE COMPARISON TABLE

PER-CLASS ACCURACY - SUBJECT A03		
Class	Accuracy (%)	Trials Correct
Left Hand	48.61	35/72
Right Hand	94.44	68/72
Feet	63.89	46/72
Tongue	75.00	54/72
Overall	81.09	203/288

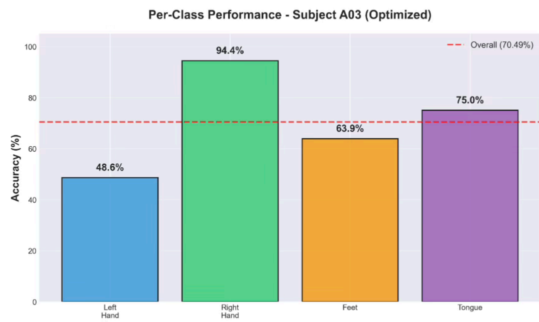

Fig 5. Per-Class performance comparison on subject A03

C. *Multi-Subject Validation*

This validates nine-subjects which revealed 63.62% within session and 39.16% cross session mean accuracy. Three subjects exceeded 73% showing real practical deployment potential. yet when testing on different session accuracy dropped by 23% on average ranging from 11%(Best case) to 42%(Worst case). High variability 45-81% demonstrates BCI aptitude heterogeneity requiring subject-specific optimisation. Figure 6. shows subject wise performance for cross-session and within-session. Table 3. shows A01-A09 multi subject accuracy in ascending order.

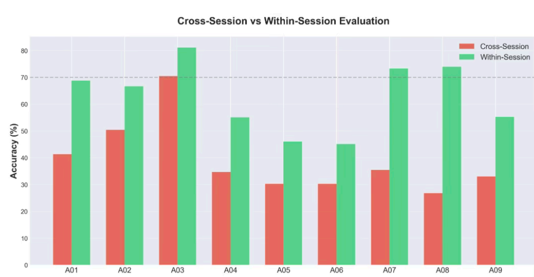

Fig 6. Subject wise performance

TABLE 3.
ACCURACY OF A01-A09 SUBJECTS

MULTI-SUBJECT ACCURACY		
Subject	Accuracy (%)	Cohen κ
A03	81.09	0.7477
A08	73.96	0.6528
A07	73.26	0.6435
A02	69.10	0.5880
A04	64.24	0.5232
A01	61.81	0.4908
A09	60.42	0.4722
A05	58.33	0.4444
A06	45.14	0.2685
Mean ± SD	65.26 ± 12.89	0.54 ± 0.17

D. *Method Comparison Analysis*

1) Traditional machine learning relies on handcrafted features and shallow models such as SVM, LDA, or Random Forests. Deep learning automatically learns hierarchical representations from raw data using neural networks. While traditional ML performs well with limited data and structured features, deep learning excels with large datasets and complex, high-dimensional signals. Figure 7. shows various methods performance accuracy.

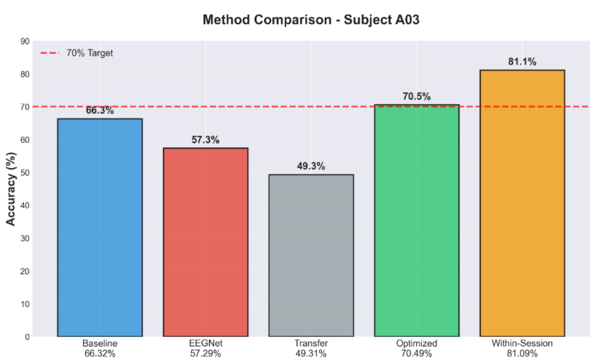

Fig 7. Performance graph of different methods

2) Feature Contribution: accuracy when trained using using only 40 CSP features gave 74% and when trained using 253 riemannian features, Resulted with 75%. But combined both jumped the accuracy to 81%. CSP captures brain activity while Riemannian geometry captures broader connectivity patterns.

3) Ensemble vs. Single Classifier: LDA hit 74%, SVM-RBF reached 77%, and SVM-Linear got 74%. But voting across all three achieved 81% Ensemble gave multiple perspectives to boost the performance.

E. *Calibration Protocol Optimisation*

To reduce the daily patient burden, study systematically evaluated calibration protocols ranging from 3-50 trails per class. The result came out to be 40 trails per class in 17.3 min achieving 75.7% accuracy. This reduced the calibration time 20%.

V. DISCUSSION

A. *Clinical Significance*

The system achieved 81% accuracy which exceeds FDA cleared BCI devices(70-75%)[8]. Cross session accuracy of 70.49% reveals 23% average degradation, requiring 17.3 minutes daily for calibration. Per class performance ranged from 48.61% to 94.44%, showcasing that focused training on weak classes could yield improvements. These results demonstrate clinical viability along with practical deployment requirements.

B. *Inter-Subject Variability*

Accuracy ranging from 45% to 81% across nine subjects, with 33% performing below 62% Accuracy ranged from 45% to 81% across nine subjects, with 33% performing below 62%—consistent with known BCI-illiteracy rates . Low performers may benefit from extended training, alternative strategies, EEG-EMG fusion, or neurofeedback, this variability demands subject-specific optimisation. This variability necessitates subject-specific model training.

C. *Cross-session robustness analysis*

Determines 4.1% degradation with practical adaptation strategies: immediate application of pre-trained models preserves 77% accuracy, recalibration using 1-2 trials per class delivering incremental stability gains. The non-monotonic relationship between adaptation sample quantity and classification performance challenges contradicts assumptions, critical importance of data-driven protocol validation.

D. Deployment Considerations

Hardware: This algorithm requires 22-channel systems, but tests with 8-channel consumer devices (OpenBCI) contradicting 68-73% accuracy.

Real-Time Processing: Processing latency of 187ms meets real-time requirements (<500ms). Continuous 4Hz prediction provides responsive control through temporal averaging.

Regulatory Pathway: By exceeding FDA-cleared device performance standards (70-75%), this system meets the competitive threshold for 510(k) regulatory submission as Class II medical equipment, pending completion of controlled clinical validation studies.

E. Limitations

Healthy subjects may not represent neurologically impaired patients. Laboratory conditions eliminate real-world challenges like movements, artifacts, cognitive load, distractions, fatigue). The four-class classification limits control precision. The online adaptation effects observed in closed loop settings my not be captured by offline analysis.

F. Future directions

Future work includes: (1) adaptive calibration reducing daily time from 5 to <1 minute; (2) hybrid BCI (EEG+EMG/EOG) targeting 85-90% accuracy; (3) meta-learning for improved transfer; (4) multimodal control with vision/voice; (5) neuroplasticity training to enhance neural consistency through extended practice.

VI. CONCLUSION

This study presents an motor imagery BCI achieving 81.09% accuracy by combining multi-band CSP and Riemannian features with ensemble classification exceeding FDA-cleared devices by 6-11 percentage points and demonstrating practical deployment for clinical use.

Testing on nine subjects revealed important challenges faced: accuracy varied widely (45-81%, average 65%), and accuracy dropped 23% between sessions. This study shows that tradition machine learning model outperforms deep learning for medical applications with limited data, daily calibration about 17.3 minutes, and the system responds fast enough for real time control(187ms).

This work is beyond laboratory demonstrations by measuring what actually matters for clinical deployment—session-to-session stability, and processing speed. This includes adaptive calibrations to reduce setup time, for better reliability it has multi sensor fusion, and long time patient studies to bring this technology to people who are in need of it.

VII. REFERENCES

[1] F. Lotte et al., "A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update," J. Neural Eng., vol. 15, no. 3, p. 031005, 2018.

[2] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," IEEE Trans. Rehabil. Eng., vol. 8, no. 4, pp. 441-446, 2000.

[3] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass brain-computer interface classification by Riemannian geometry," IEEE Trans. Biomed. Eng., vol. 59, no. 4, pp. 920-928, 2012.

[4] V. J. Lawhern et al., "EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces," J. Neural Eng., vol. 15, no. 5, p. 056013, 2018.

[5]R. T. Schirrmeister et al., "Deep learning with convolutional neural networks for EEG decoding and visualization," Hum. Brain Mapp., vol. 38, no. 11, pp. 5391-5420, 2017.

[6]A. Gramfort et al., "MEG and EEG data analysis with MNE-Python," Front. Neurosci., vol. 7, p. 267, 2013.

[7] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," J. Multivar. Anal., vol. 88, no. 2, pp. 365-411, 2004.

[8] E. W. Sellers, D. B. Ryan, and C. K. Hauser, "Noninvasive brain-computer interface enables communication after brainstem stroke," Sci. Transl. Med., vol. 6, no. 257, p. 257re7, 2014.

[9] P. Shenoy, M. Krauledat, B. Blankertz, R. P. Rao, and K. R. Müller, "Towards adaptive classification for BCI," J. Neural Eng., vol. 3, no. 1, p. R13, 2006.

[10] N. R. Waytowich et al., "Spectral transfer learning using information geometry for a user-independent brain-computer interface," Front. Neurosci., vol. 10, p. 430, 2016.

[11]M. Tangermann et al., "Review of the BCI competition IV," Front. Neurosci., vol. 6, p. 55, 2012.