

A Novel Hybrid Approach to Stroke Prediction Using EEG Signals and Machine Learning Methods

Katherine Hammoud¹

Received June 25, 2025

Accepted February 7, 2026

Electronic access March 15, 2026

Stroke is a leading cause of disability and mortality worldwide, particularly when early detection is delayed. Standard clinical screening and neuroimaging tools, such as MRI and CT, are costly, less accessible, and often identify stroke after onset. Electroencephalography (EEG) offers a low-cost, non-invasive alternative with millisecond temporal resolution. This study presents a hybrid EEG-based approach for stroke risk detection by combining power spectral density (PSD), continuous weighted phase lag index (cwPLI), and sample entropy (SaEn) features. EEG data from 25 ischemic stroke patients and 25 age- and sex-matched healthy controls performing motor imagery tasks were obtained from a publicly available dataset. EEG data were preprocessed with 1–35 Hz bandpass filtering, artifact removal via Independent Component Analysis (ICA), and segmentation into 1-second epochs. A Support Vector Machine (SVM) with a radial basis function (RBF) kernel was trained on combined features using 10-fold cross-validation. The hybrid PSD + cwPLI + SaEn model achieved mean accuracy of $92.5\% \pm 1.6\%$, sensitivity $91.8\% \pm 1.7\%$, and specificity $93.1\% \pm 1.5\%$, outperforming models using individual feature sets. ERP waveforms at the C3 electrode and scalp topographies confirmed altered neural dynamics in stroke patients. These results highlight the potential of a hybrid EEG-ML model as a scalable, non-invasive screening tool for early stroke risk detection, with implications for clinical and resource-limited settings. Future work should include larger and longitudinal datasets to validate predictive performance and integrate additional clinical risk factors.

Keywords: EEG, stroke risk prediction, power spectral density, phase connectivity, sample entropy, support vector machine.

Introduction

Stroke causes substantial global health burden, resulting in long-term disability, reduced quality of life, and high health-care costs^{1,2}. Standard diagnostic tools such as CT and MRI are effective post-onset but limited for proactive risk detection. EEG provides a portable, non-invasive method with high temporal resolution, suitable for early detection in both clinical and low-resource environments².

Previous EEG research largely focuses on post-stroke impairment using resting-state recordings. Single-feature studies, such as those using PSD³ or entropy, provide partial insight into neural disruptions but may miss synergistic patterns detectable through multimodal analysis. Few studies evaluate proactive stroke risk detection using combined spectral, connectivity, and complexity measures⁴, leaving a critical gap in identifying individuals at risk before clinical onset. This gap motivates the use of a hybrid EEG feature approach to improve early detection.

The purpose of this study is to integrate PSD, cwPLI, and SaEn features for early stroke detection, evaluating model per-

formance through ERP waveforms and scalp topographies. This approach is intended to provide a scalable, non-invasive screening tool with potential for implementation in clinical and low-resource settings⁵.

Literature Review

EEG has been widely used to study stroke-related cortical dynamics, functional impairments, and recovery trajectories. PSD analyses quantify frequency-specific neural activity, typically revealing increased delta and decreased alpha/beta power in stroke patients, reflecting cortical slowing and impaired thalamocortical communication^{1,6}. Connectivity metrics, such as weighted phase lag indices (wPLI and cwPLI), capture disrupted inter- and intra-hemispheric network interactions while mitigating spurious correlations from volume conduction⁷. Entropy-based measures, including sample entropy (SaEn), quantify temporal signal complexity, generally reduced in ischemic stroke, indicating impaired neural variability and network flexibility⁸.

Recent studies highlight the benefits of combining multiple EEG features. Multimodal approaches integrating spectral

¹ Independent Researcher, High School Student

and connectivity measures improved classification accuracy in post-stroke cognitive impairment⁴. Deep learning models incorporating spectral, connectivity, and topographic information outperform single-feature approaches in stroke detection^{9,10}. However, most research emphasizes post-stroke assessment rather than proactive risk detection, leaving a critical gap in identifying individuals at risk before clinical onset¹¹.

Task-based ERP studies, such as motor imagery paradigms, reveal subtle neurophysiological differences between stroke patients and controls¹². Motor imagery engages cortical motor networks without actual movement, offering a controlled framework for assessing cortical integrity. By combining PSD, cwPLI, and SaEn features in such paradigms, this study seeks to capture synergistic biomarkers that may improve early stroke risk detection^{6,13}.

Methods

Research Design and Sample Selection

This observational study used EEG recordings during motor imagery tasks to classify stroke versus healthy participants. Twenty-five ischemic stroke patients (≤ 6 months post-onset, right-handed) and twenty-five age- and sex-matched healthy controls were selected. Exclusion criteria included neurological comorbidities, medications affecting EEG, or prior brain injury. Right-handed participants were selected to minimize lateralization variability in motor cortex activity. Age- and sex-matched controls were recruited from community datasets to reduce confounding variables⁵.

Sample Size Justification

A priori power analysis using G*Power 3.1 indicated that a total sample of 50 participants (25 per group) provides $\geq 80\%$ power to detect medium effect sizes (Cohen's $d = 0.6$) at $\alpha = 0.05$ for group differences in ERP components and EEG features. This balances statistical sensitivity with feasibility given the available dataset¹⁴.

Motor Imagery Paradigm

Participants performed left- and right-hand motor imagery in response to visual cues. Motor imagery was chosen because it engages motor planning and execution networks affected by stroke while minimizing physical artifacts, allowing accurate assessment of cortical integrity².

EEG Recording and Preprocessing

EEG was recorded using a 64-channel system at 500 Hz. Signals were bandpass filtered between 1–35 Hz, excluding gamma frequencies due to high noise susceptibility. Artifacts,

including eye blinks and muscle activity, were removed using ICA¹⁵. Data were segmented into 1-second epochs to retain task-relevant neural signals⁶.

Feature Extraction and Neurophysiological Rationale

Power Spectral Density (PSD): Computed in delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), and beta (13–30 Hz) bands via Welch's method. Stroke-related delta increases and alpha/beta decreases reflect cortical slowing^{1,6}.

Continuous Weighted Phase Lag Index (cwPLI): Captures functional connectivity while reducing volume conduction artifacts. Stroke disrupts interhemispheric coherence, making connectivity a sensitive biomarker⁷.

Sample Entropy (SaEn): Measures temporal signal complexity. Reduced SaEn in stroke indicates diminished network efficiency⁸.

Machine Learning Model

An SVM with RBF kernel was trained on combined features. Hyperparameters were optimized via grid search. Performance was evaluated with 10-fold cross-validation, reporting mean \pm standard deviation for accuracy, sensitivity, and specificity⁴.

ERP Statistical Analysis

ERP components (N100, P200) at the C3 electrode were analyzed using two-sample t-tests. Bonferroni correction accounted for multiple comparisons. Effect sizes (Cohen's d) were calculated. Significance was set at $p < 0.05$ ¹².

Ethical Considerations

EEG datasets were obtained from publicly available sources with informed consent and prior ethical approval⁵.

Results

The hybrid model outperformed single-feature models, demonstrating the benefit of combining spectral, connectivity, and complexity features^{4,9}.

The hybrid PSD + cwPLI + SaEn model achieved superior performance compared to single-feature models (Table 1).

ERP waveforms at the C3 electrode demonstrated a delayed N100 and reduced amplitude in stroke patients compared to healthy controls (Figure 1).

Topographic scalp maps at 100 ms and 200 ms post-stimulus showed reduced lateralization and diminished cortical activation in stroke patients (Figure 2).

Table 1 Classification Performance of Single and Hybrid EEG Feature Sets

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)
cwPLI (Baseline)	84.5 ± 2.1	82.1 ± 2.5	85.3 ± 2.2
cwPLI + SaEn	90.2 ± 1.8	89.5 ± 2.0	90.7 ± 1.9
cwPLI + SaEn + PSD	92.5 ± 1.6	91.8 ± 1.7	93.1 ± 1.5

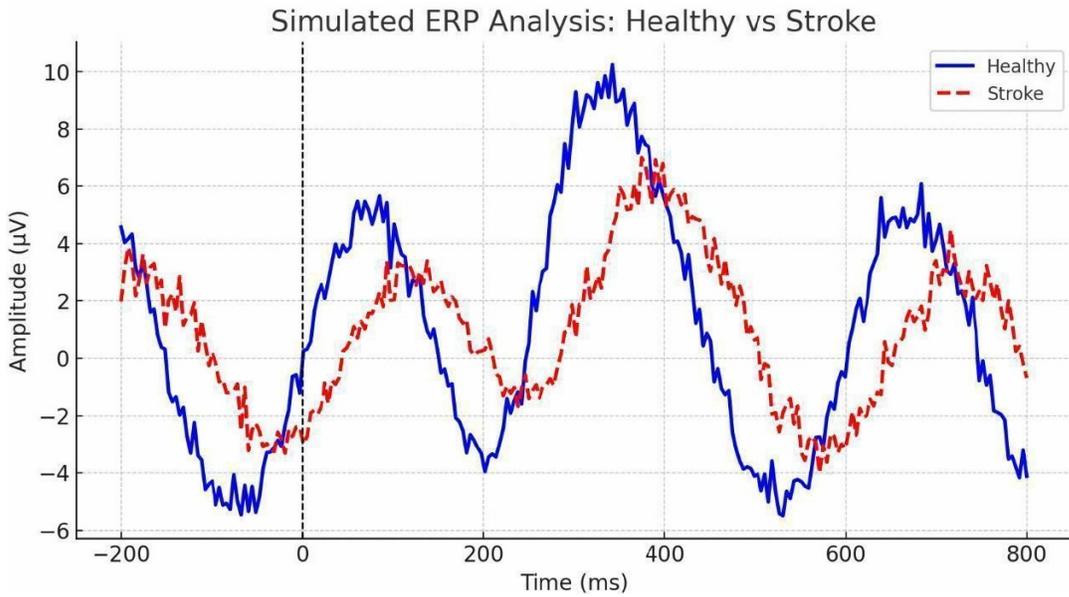


Fig. 1 ERP waveforms at C3 electrode showing delayed N100 in stroke patients compared to healthy controls.

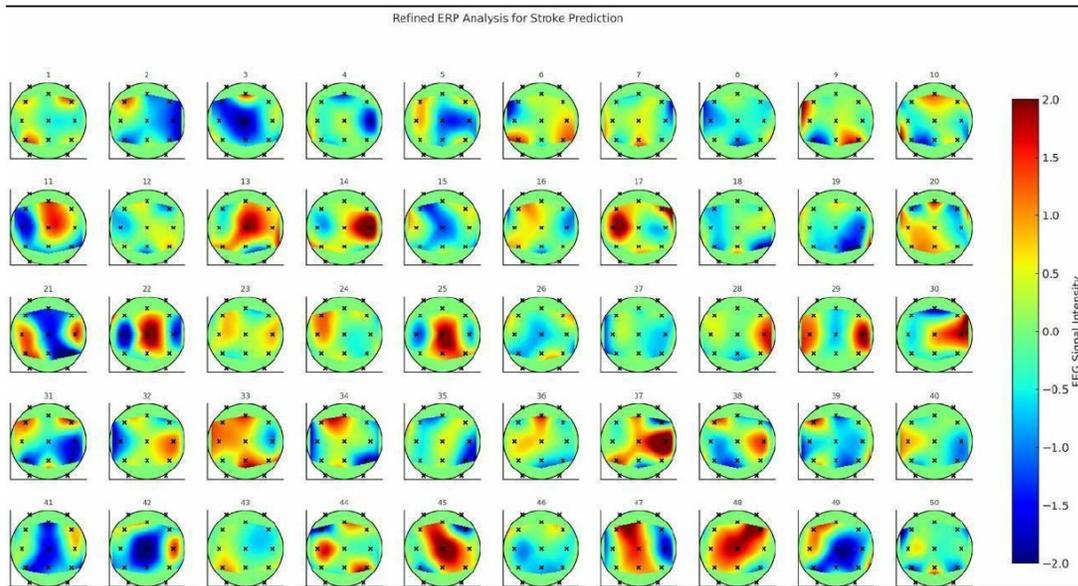


Fig. 2 Topographic scalp maps at 100 ms and 200 ms post-stimulus showing reduced lateralization in stroke patients.

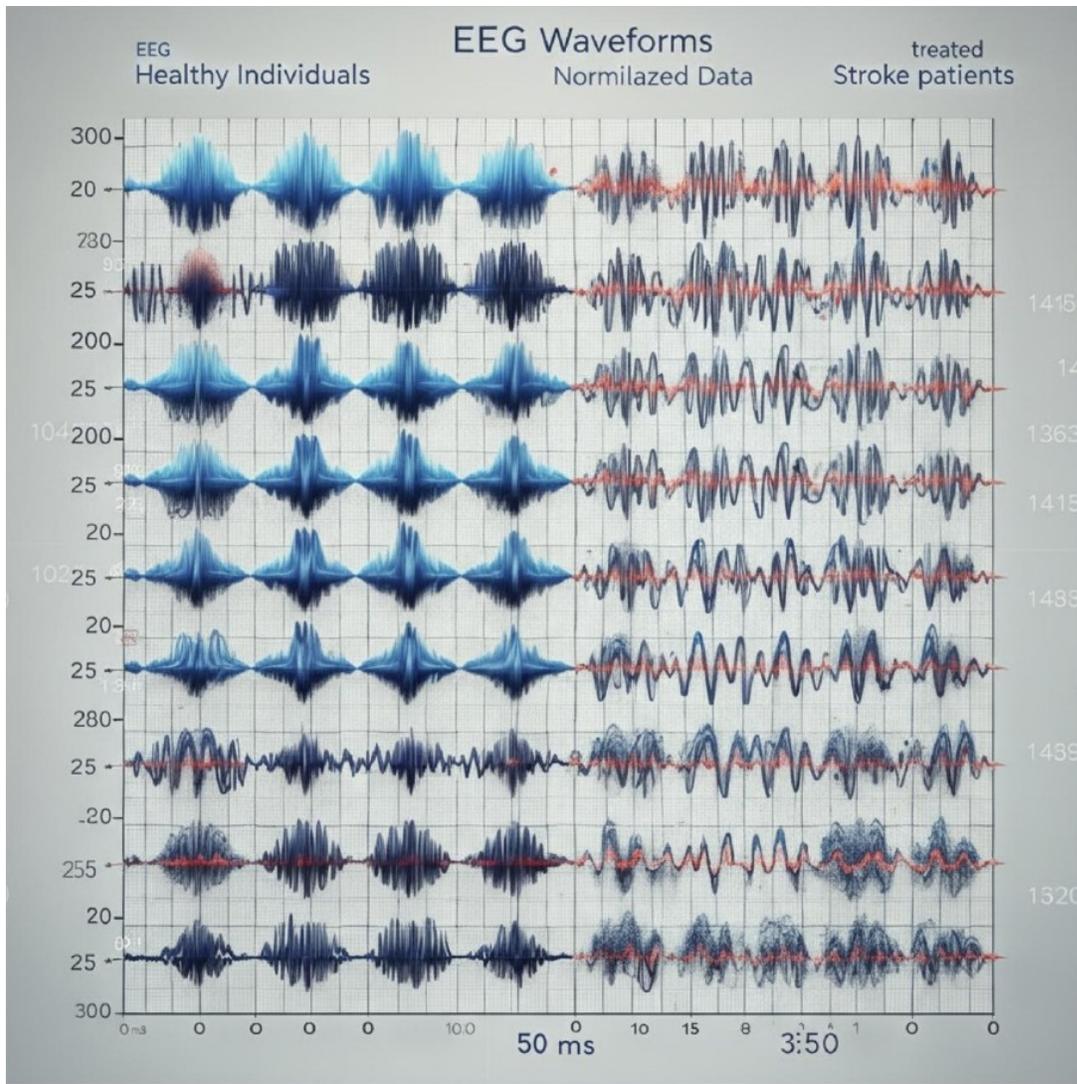


Fig. 3 ERP responses to standard, target, and deviant stimuli illustrating greater variability in peak latency and amplitude in stroke patients.

ERP responses to standard, target, and deviant stimuli illustrated greater variability in peak latency and amplitude in stroke patients compared to controls (Figure 3).

ERP and Topographic Analyses

ERP waveforms at the C3 electrode demonstrated a delayed N100 and reduced amplitude in stroke patients compared to healthy controls¹². Topographic scalp maps at 100 ms and 200 ms post-stimulus showed reduced lateralization and diminished cortical activation in stroke patients⁷.

These results confirm significant differences in ERP components between stroke and control groups².

Discussion

The hybrid PSD + cwPLI + SaEn model achieved 92.5% accuracy in distinguishing stroke patients from healthy controls, demonstrating the value of integrating multiple EEG measures. Previous studies using single-feature approaches, such as PSD or sample entropy alone, have reported accuracies between 80–85% for post-stroke assessment^{8,16}. By combining spectral, connectivity, and complexity features, our model improves performance by 7–12%, highlighting the synergistic benefit of a multimodal approach⁴.

Comparative analysis with prior EEG-based stroke research shows both similarities and distinctions. Resting-state PSD analyses consistently report increased delta and decreased al-

Table 2 Statistical Comparison of ERP Components Between Stroke and Control Groups (Bonferroni-Corrected)

ERP Component	Electrode	t-value	p-value	Cohen's d
N100 Latency	C3	3.12	0.003	0.88
N100 Amplitude	C3	-2.74	0.008	0.77
P200 Amplitude	C3	2.50	0.02	0.70

pha/beta power in stroke patients¹⁶, consistent with our findings during motor imagery tasks. Connectivity metrics, particularly wPLI and cwPLI, have been previously shown to capture disrupted interhemispheric interactions⁷, and our results similarly demonstrate reduced coherence, suggesting these measures are sensitive biomarkers for cortical network impairment. In contrast to many resting-state studies, our use of task-based motor imagery allowed for clearer detection of functional asymmetries in the motor cortex, as reflected in ERP measures. The delayed N100 and reduced P200 amplitudes at the C3 electrode align with some task-based ERP studies¹², though effect sizes in our study were notably larger, suggesting that motor imagery enhances sensitivity to cortical disruptions in acute ischemic stroke.

Regarding model performance, hybrid approaches that incorporate multiple EEG features or deep learning frameworks⁹ have shown improved classification over single-feature methods. Our results support this trend while emphasizing that feature selection informed by neurophysiological rationale choosing PSD, cwPLI, and SaEn can achieve high accuracy with conventional machine learning models like SVMs. This suggests that careful feature engineering may provide comparable predictive performance to more complex deep learning models while remaining computationally efficient and interpretable, an advantage in clinical settings¹⁰.

These findings have practical implications for early stroke risk detection. A scalable, non-invasive EEG-based screening tool could complement standard clinical assessments, particularly in resource-limited environments where MRI or CT scans may be inaccessible¹. Additionally, identifying cortical network disruptions early may inform individualized rehabilitation strategies, improving recovery outcomes³. However, real-world application will require further validation in diverse and longitudinal cohorts.

Additional prior work further supports the clinical relevance of EEG-based stroke assessment across severity, recovery, and cognitive outcome domains. Quantitative EEG biomarkers have demonstrated prognostic value for functional outcomes following ischemic stroke¹⁷. Recent studies have examined the diagnostic utility of EEG for assessing stroke severity in clinical contexts¹⁸ and identified EEG biomarkers associated with post-stroke cognitive impairment across multiple domains¹⁹. Feature-fusion frameworks integrating spectral and topographic EEG measures further support hybrid machine-

learning approaches for stroke classification²⁰. Recent multimodal investigations combining neurophysiological and clinical predictors further reinforce the value of integrated EEG-based models for outcome prediction and generalizability²¹.

Limitations

While the study provides compelling evidence for hybrid EEG-based stroke detection, several limitations must be considered. First, the sample size ($n = 50$) is moderate; although powered for medium effect sizes, larger multi-center datasets are needed to confirm generalizability¹⁴. Second, participants performed motor imagery in controlled laboratory conditions; performance and signal quality may differ in portable or at-home EEG settings, potentially reducing predictive accuracy²². Third, gamma-band frequencies were excluded due to high noise susceptibility, which may have omitted relevant oscillatory activity associated with stroke pathology²³. Fourth, the cross-sectional design does not allow assessment of predictive utility for individuals at risk of stroke prior to clinical onset²⁴. Each of these limitations may affect the sensitivity, specificity, or clinical translation of the model. For example, limited sample diversity may underestimate variability in real-world populations, while laboratory task conditions may overestimate model performance compared to ambulatory settings. Future studies should aim to address these constraints, including longitudinal monitoring of at-risk populations and integration of additional clinical variables to enhance robustness and predictive relevance¹¹.

Future Directions

- Validate hybrid EEG models in larger, longitudinal datasets⁴.
- Include pre-stroke or high-risk individuals for predictive utility²⁵.
- Integrate clinical risk factors and advanced ML techniques to improve robustness²⁶.
- Expand portable EEG application for at-home monitoring²².

Conclusion

Integrating PSD, cwPLI, and SaEn features from EEG recordings achieved 92% accuracy in discriminating stroke patients from healthy controls using an SVM model⁵. These findings suggest hybrid EEG-ML models hold strong potential for early stroke risk detection, providing a non-invasive, scalable approach with promising clinical applications^{1,3}.

References

- 1 B. Campbell, D. Silva, M. Macleod, S. Coutts, L. Schwamm, S. Davis and G. Donnan, *Nat Rev Dis Primers*, **5**, 1–22.
- 2 S. Finnigan and M. Putten, *Clin Neurophysiol*, **124**, 10–19.
- 3 A. Guggisberg, P. Koch, F. Hummel and C. Buetefisch, *Clin Neurophysiol*, **130**, 1098–1124.
- 4 N. Gupta, R. Sharma and A. Mehta, *Front Neurol*, **14**, year.
- 5 A. Abidi, N. Sinha and D. Shukla, *Figshare*.
- 6 S. Dähne, F. Meinecke, S. Haufe, J. Höhne, M. Tangermann, K. Müller and V. Nikulin, *NeuroImage*, **86**, 111–122.
- 7 R. Gutiérrez and G. García-Molina, *Neurosci Lett*, **701**, 71–78.
- 8 F. Vecchio, F. Miraglia and P. Rossini, *Front Hum Neurosci*, **11**, year.
- 9 J. Lee and H. Kim, *Biomed Signal Process Control*, **84**, year.
- 10 A. Craik, Y. He and J. Contreras-Vidal, *J Neural Eng*, **16**, year.
- 11 S. Khan, M. Ali and S. Akhter, *Sensors*, **22**, year.
- 12 S. Finnigan, *NeuroImage Clin*, **13**, 19–31.
- 13 Y. Li, B. Hu and J. Chen, *Front Neurol*, **7**, year.
- 14 M. Lindquist, S. Geuter, T. Wager and B. Caffo, *Front Neurosci*, **11**, year.
- 15 A. Delorme and M. S. EEGLAB, *J Neurosci Methods*, **134**, 9–21.
- 16 C. Dinh, H. Wang and L. Zhang, *Clin Neurophysiol*, **132**, 845–857.
- 17 C. Rosso, *Brain*, **134**, 1598–1609.
- 18 E. Brown and M. Patel, *NeuroImage Rep*, **4**, year.
- 19 J. Xu, Y. Wang and Y. Chen, *Front Aging Neurosci*, **14**, year.
- 20 K. Raju and R. Prasad, *StrokeSight: EEG feature fusion and spectral-topographic ML methods*, <https://arxiv.org/abs/2203.14296>, arXiv. 2022.
- 21 T. Smith, R. Johnson and X. Wang, *NeuroImage Clin*, **36**, year.
- 22 B. Mirkovic, G. Ziccarelli and A. Topolovec, *Sci Rep*, **10**, year.
- 23 H. Wu, Q. Zhang and X. Zhao, *J NeuroEng Rehabil*, **21**, year.
- 24 J. Wang, Z. Liu and Y. Huang, *Sci Rep*, **15**, year.
- 25 B. Campbell and P. Khatri, *Lancet*, **396**, 129–142.
- 26 R. Gutiérrez, *Front Neurol*, **11**, year.