IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 17, April 2026
Impact Factor: 8.2

Neuro-Cognitive AI Fingerprinting for Pre-Radiographic Brain Tumor Detection Using Transformer-Based Phase-Amplitude Coupling Analysis

Shrikant Kudale, Rehan Surani, Seema Saud, Jayashree Prasad
Department of Computer Science & Engineering
MIT School of Computing, MIT-ADT University, Pune, India

Abstract: *Conventional brain tumor diagnosis waits for a mass to grow large enough to show up on an MRI. By that point, the damage is already done — treatment windows are narrower, and survival outcomes are far worse than they needed to be. This work takes a different angle. Rather than searching for a physical lump, we ask a more fundamental question: what does a growing tumor do to the brain's electrical behavior, and can an AI system learn to recognize that disruption early?*

Our method is built around Phase-Amplitude Coupling, or PAC — a measurable property of how high- and low-frequency brain waves coordinate with each other. Tumor-induced inflammation disrupts this coordination well before anything appears on structural imaging. We trained a Transformer-based deep learning model on synthetic EEG signals from 900 virtual subjects, paired with three behavioral biomarker features: eye-tracking variance, gait asymmetry, and speech syntax irregularity. On the held-out test set, the model reached 87.4% classification accuracy and an AUC-ROC of 0.93 for the highest-risk class. Attention analysis showed the model learned to focus on theta-gamma PAC features across frontal and temporal channels — precisely where neuroscience literature would predict disruption to appear. The results demonstrate that this detection framework is computationally viable, even if clinical validation lies ahead..

Keywords: Brain tumor detection, Phase-Amplitude Coupling, EEG biosignals, Transformer, early detection, neural oscillations, digital biomarkers, pre-radiographic screening

I. INTRODUCTION

Here is an uncomfortable truth about how brain tumors are currently diagnosed: by the time a radiologist sees something on an MRI, the tumor has typically been growing for months — sometimes longer. Glioblastoma, the most lethal primary brain tumor, carries a median survival of roughly 15 months after diagnosis, and that number is heavily shaped by when detection happens [1]. Not by the quality of treatment alone, and not by biology alone — but by timing. And timing, right now, is determined by structural imaging that can only see what is already large enough to register as a visible mass.

The frustrating part is that the brain does not stay quiet while this is happening. Long before a tumor becomes structurally identifiable, it disrupts the local neural environment — creating inflammation, altering metabolic activity, and interfering with the synchronized electrical rhythms that different brain regions use to communicate. Among the most well-studied of these rhythms is Phase-Amplitude Coupling, or PAC. Specifically, the phenomenon where the amplitude of gamma oscillations (30–80 Hz) gets modulated by the phase of slower theta waves (4–8 Hz) [2]. This coupling is not just a signal property — it has real functional significance. It underlies working memory, attentional control, and inter-regional coordination. And in glioma, it breaks down [3].

DOI: 10.48175/IJARSCT-34249

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 17, April 2026 Impact Factor: 8.2

That breakdown is the foundation of this project. If tumor-induced neuroinflammation produces measurable, class-specific PAC disruption before a scan can detect anything, then an AI trained to recognize that disruption could serve as a pre-radiographic warning signal. The question is whether the signature is distinctive enough to be learned reliably.

Beyond EEG, there is a growing body of work showing that neurological disease leaves footprints in behavioral patterns — gait timing, speech syntax, micro-saccade trajectories — that precede clinical presentation by months [4][5]. These behavioral signals are noisier than electrophysiology, but they provide a complementary perspective. Combining the two is the core multimodal bet this paper makes.

Specific contributions include: a synthetic data generation pipeline for 900 EEG subjects with controlled PAC disruption; a multi-modal feature extraction and fusion framework using the Modulation Index; a Transformer-based three-class risk classifier; an attention-weight explainability analysis; and a comparison against MLP and CNN baselines.

II. RELATED WORK

A. PAC in the Context of Brain Pathology

Canolty and Knight's foundational work established PAC as more than an artifact of signal mixing — it is a genuine mechanism of large-scale neural coordination, particularly theta-gamma coupling in hippocampal and prefrontal circuits during working memory tasks [2]. What happened next in the literature was gradual but important: researchers started noticing that PAC degrades in a range of neurological conditions.

For glioma specifically, there is now reasonable evidence that tumor infiltration reduces theta-gamma PAC in perilesional tissue, while alpha-beta coupling becomes abnormally elevated in surrounding regions — a kind of compensatory reorganization that also carries a detectable signature [3]. Jimenez-Marin and colleagues pushed this further by analyzing intraoperative ECoG recordings from glioma patients directly. PAC-derived features, they found, could separate tumor-adjacent tissue from healthy cortex with specificity above 80% [6]. That is a remarkable result — even if it came from invasive recordings under surgical conditions rather than anything resembling a screening scenario.

B. Machine Learning on EEG in Brain Tumor Research

Efforts to bring EEG into automated tumor detection remain relatively sparse. Mun and colleagues extracted power spectral density features from resting-state EEG and used an SVM to separate glioma patients from healthy controls at 74% accuracy [7]. Acharya et al. combined wavelet features with entropy measures and reached 88% using a random forest [8] — though both studies had modest cohort sizes and did not capture cross-frequency interactions.

On the deep learning side, EEGNet showed that compact convolutional architectures can generalize across multiple EEG paradigms [9]. More recently, BrainBERT adapted self-supervised Transformer pretraining to intracranial neural data, demonstrating that attention-based models can extract meaningful representations from high-dimensional neural time series [10]. That work is closer in spirit to what we are attempting here, though it targets different pathologies.

C. Behavioral Biomarkers

Zhan et al. demonstrated that fine-grained gait analysis using smartphone accelerometer data could predict Parkinson's disease severity with clinical-grade accuracy [11]. Fraser et al. showed that syntactic complexity measures from spontaneous speech could distinguish Alzheimer's patients with 81.9% accuracy [12]. Taphoorn and Klein reported measurable cognitive decline in glioma patients prior to clinical symptom onset [13].

D. What is Missing

No existing framework integrates EEG-based cross-frequency coupling analysis with multi-modal behavioral features under a sequence-aware architecture specifically targeting pre-radiographic tumor risk. Individual pieces exist. The integration does not. That is the gap this paper attempts to address, at least at the simulation level.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 17, April 2026
Impact Factor: 8.2

III. DATASET GENERATION

A. Synthetic EEG Signal Generation

No public dataset exists that provides early-stage tumor-associated EEG recordings alongside matched behavioral biomarker data. Building one from scratch would require a longitudinal clinical study spanning years — something well beyond the scope of what any single university project can undertake. So we did what the circumstances required: designed a synthetic generation pipeline that mirrors the statistical properties of real EEG, calibrated against published physiological values rather than invented from scratch.

Each virtual subject got a 19-channel EEG recording sampled at 256 Hz, running for 30 seconds. We modeled each channel as a mixture of five band-limited oscillatory components — delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (30–80 Hz) — with relative power drawn from Gaussian distributions parameterized against resting-state EEG norms from the literature [14].

The PAC component got handled separately, because that is where the three classes diverge. Normal subjects had gamma amplitude modulated by theta phase using a Modulation Index centered at MI = 0.18 (SD = 0.04) — right in line with healthy resting-state PAC values [2]. Early Disruption subjects had theta-gamma MI reduced by roughly 30% in five randomly selected channels, with alpha-beta coupling elevated by 15%. The High Risk class was more severe: theta-gamma MI dropped 55–70%, alpha-beta coupling climbed 35–40%, and background gamma power increased to simulate the neuroinflammatory hyperexcitability documented in the tumor literature. Each of the 900 subjects (300 per class) received a unique random seed — no two EEG recordings are identical, even within the same class.

B. Behavioral Biomarker Generation

Three scores accompany each synthetic subject, representing behavioral signals the kind a clinician or wearable device might realistically capture.

The Eye-Tracking Variance Score (ETVS) stands for saccade amplitude standard deviation during a simulated reading task. Healthy: centered at 1.2° (SD 0.18). Early Disruption: 1.65° (SD 0.22). High Risk: 2.3° (SD 0.31). The Gait Asymmetry Index (GAI) captures step-length asymmetry over a 10-meter walk. Healthy: 0.04 (SD 0.01); Early Disruption: 0.09 (SD 0.02); High Risk: 0.17 (SD 0.03). The Speech Syntax Complexity Score (SSCS) is an inverse measure of syntactic regularity — higher values mean greater fragmentation. Healthy: 0.22; Early Disruption: 0.34; High Risk: 0.47.

All three scores were min-max normalized using parameters computed from the training split alone. Applying those same parameters to validation and test data is standard practice, but worth stating explicitly: there was no leakage of test distribution statistics into the normalization.

IV. FEATURE EXTRACTION AND FUSION

A. EEG Preprocessing

Three steps before any feature gets computed. First, a 4th-order Butterworth bandpass filter per channel — one pass for theta (4–8 Hz), another for gamma (30–80 Hz). Second, each 30-second recording cut into 15 non-overlapping 2-second epochs. Third, any epoch with a peak amplitude exceeding ±150 µV flagged as artifactual and replaced via linear interpolation from its neighbors. That last step is deliberately conservative — we lose epochs, not subjects.

B. PAC via the Modulation Index

PAC was computed using the Modulation Index method introduced by Tort et al. [15], which remains the most widely validated approach for quantifying cross-frequency coupling. The instantaneous theta phase comes from the Hilbert transform. That phase gets divided into 18 bins of 20° each, covering the full cycle. Mean gamma amplitude within each bin gets computed, and the resulting distribution is compared to a uniform distribution via the Kullback-Leibler divergence. The MI is that divergence normalized by log(N bins).

Nineteen channels, fifteen epochs, two frequency pairs. That gives 285 MI values for theta-gamma coupling and another 285 for alpha-beta. Concatenated, each subject ends up with a 570-dimensional PAC feature vector.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 17, April 2026
Impact Factor: 8.2

C. Fusion

The 570-dimensional PAC vector concatenated with three behavioral scores gives a 573-dimensional fused representation per subject. That gets reshaped into a sequence of 15 temporal tokens, each of dimension 38 — representing one epoch's worth of channel-level PAC values (19 theta-gamma + 19 alpha-beta), with behavioral scores distributed across the first three tokens. The resulting tensor is what the Transformer encoder sees.

V. MODEL ARCHITECTURE

A. Transformer Encoder

The classification model is a Transformer encoder stack adapted from Vaswani et al. [16], adjusted for biosignal classification rather than language. The two tasks are more similar than they might initially appear: both involve learning meaningful patterns from token sequences with temporal structure, and the core self-attention mechanism transfers reasonably well.

An **Input Embedding Layer** first projects each 38-dimensional token into a 128-dimensional space. Learnable positional encodings of the same dimension are added to preserve temporal ordering across the 15-token sequence. A learnable [CLS] token is prepended, bringing the sequence length to 16.

Four **Transformer Encoder Blocks** follow in sequence. Each block runs multi-head self-attention with 8 heads (key, query, and value dimensions of 16 each), applies Pre-LN layer normalization before each sub-layer, passes through a position-wise feed-forward network with hidden dimension 256 and GELU activation, and applies dropout at rate 0.1 to both attention weights and feed-forward outputs.

The **Classification Head** takes the [CLS] token output and passes it through a two-layer MLP (128 → 64 → 3) with ReLU activation, ending in a three-class softmax. Total trainable parameters: approximately 412,000. Lightweight by modern standards, and compatible with standard CPU hardware.

B. Training

Models were trained for up to 80 epochs using Adam with learning rate 3×10^{-4} and weight decay 10^{-5}. A cosine annealing scheduler brought the learning rate down to 10^{-6} by the final epoch. Cross-entropy loss throughout. Dataset split: 70% training, 15% validation, 15% test. Batch size 32. Early stopping with patience 10 on validation loss.

C. Baselines

MLP Baseline: A three-layer fully connected network taking the flattened 573-dimensional input through hidden layers of 256 and 128 units (ReLU, batch normalization), terminating at a three-class output.

CNN Baseline: Three 1D convolutional layers with filter counts 32, 64, and 128 (kernel size 5, max pooling between layers), followed by global average pooling and a classification head identical to the MLP. Input treated as a (15, 38) sequence with the 38-dimensional axis serving as the channel dimension.

Both baselines used identical optimizer configurations to the Transformer.

VI. RESULTS AND DISCUSSION

A. Classification Performance

Table I summarizes performance on the held-out test set (n = 135 subjects, 45 per class).

TABLE I. Classification Performance Comparison

Model	Accuracy	F1-Score	AUC-ROC
MLP Baseline	71.9%	0.716	0.79
CNN Baseline	79.3%	0.790	0.86
Transformer (Ours)	87.4%	0.873	0.93

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 17, April 2026 Impact Factor: 8.2

The CNN-to-Transformer gap — roughly 8 percentage points in accuracy — is worth examining. Convolutional layers are good at catching local patterns within a short window, but the PAC-based disruption in our dataset is not localized to a single epoch. It accumulates and persists. Self-attention can learn to compare any token to any other across the full 15-epoch sequence, which turns out to matter a great deal here.

Per-class results for the Transformer are in Table II. Early Disruption is the hardest class to get right, which is clinically the most important failure mode to understand. The signals at that stage are subtle by design — the PAC reduction is modest and spread across only a few channels. An 0.82 F1 on this class, while not perfect, is encouraging for a proof-of-concept simulation.

TABLE II. Per-Class Performance (Transformer)

Class	Precision	Recall	F1-Score
Normal	0.91	0.94	0.92
Early Disruption	0.83	0.81	0.82
High Risk	0.90	0.89	0.89

B. Attention Weight Analysis

To understand what the model actually learned, attention weights were extracted from the final encoder layer, averaged across test subjects by predicted class. A consistent pattern emerged: for High Risk subjects, attention concentrated heavily on tokens 6 through 9 — which correspond to EEG epochs occurring between seconds 12 and 18 of the recording. Brief, transient disruptions did not drive predictions. Sustained disruption did. That finding aligns with what the neurological literature would predict.

In terms of which features drew attention, theta-gamma PAC values from frontal and temporal electrode positions (F3, F4, T5, T6) consistently received higher weights, consistent with documented glioma effects on oscillatory coupling in those regions. Among the behavioral features, the Gait Asymmetry Index was attended to approximately 2.3 times more than either ETVS or SSCS in both disruption classes. This does not mean gait is more sensitive than eye-tracking in general — it is an artifact of how the distributions were parameterized and how clearly the GAI separates classes in this particular synthetic dataset. In real data, that ranking may well shift.

C. Ablation: How Much Does Behavioral Fusion Help?

The behavioral-only result (61.3%) confirms what one would expect — these features are too nonspecific on their own. Gait asymmetry can reflect a sprained ankle. ETVS variation can come from fatigue. Without the electrophysiological context, the classifier is guessing as much as it is classifying. But with PAC features alongside them, the behavioral scores push accuracy up by 5.3 points, with the biggest gain on the Early Disruption class — exactly where the model needs the most help.

TABLE III. Ablation Study — Behavioral Biomarker Contribution

Configuration	Accuracy	Early Dis. F1	High Risk F1
PAC + Behavioral (Full)	87.4%	0.82	0.89
PAC Features Only	82.1%	0.74	0.87
Behavioral Only	61.3%	0.56	0.63

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 17, April 2026
Impact Factor: 8.2

D. Discussion

The results make a reasonable case for the core hypothesis: PAC shifts carry learnable neurological information, and a sequence model well-suited to temporal dependencies is the right architecture for extracting it. An AUC-ROC of 0.93 on the High Risk class is particularly meaningful in a screening context, where the cost of a missed positive vastly outweighs the cost of an unnecessary referral.

That said, some clear-eyed caveats are necessary. This is synthetic data, generated under assumptions about how PAC degrades with tumor proximity. Real EEG from actual patients carries interference from a dozen sources this simulation does not capture — sleep deprivation, anxiolytics, electrode impedance variability, comorbidities, and more. The 87.4% accuracy reported here is a simulation benchmark, not a clinical claim, and should not be interpreted as anything beyond that.

VII. LIMITATIONS AND FUTURE SCOPE

The most fundamental limitation is the absence of ground truth. To train a real pre-radiographic detector, one needs longitudinal data from patients monitored before their tumors become visible — which means enrolling people before they are sick, following them for years, and building a dataset where some fraction eventually receive an imaging-confirmed diagnosis. That kind of study takes a decade and requires neurosurgical partnerships that do not yet exist for this specific problem.

Specificity is the other major challenge. PAC disruption is not exclusive to brain tumors. It shows up in epilepsy. It shows up in anxiety disorders. It changes with sleep debt and alcohol. Any clinical deployment of this system would need to demonstrate that the model's learned representation is genuinely tumor-linked — not just "something is wrong neurologically" — before it could justify the kind of follow-up referrals it would trigger.

There is also the question of whether scalp EEG even retains enough signal. Most PAC effects documented in the glioma literature come from intracranial recordings. Scalp electrodes introduce spatial blurring and amplitude attenuation. Whether the relevant coupling information survives that journey to the surface is, honestly, an open and important empirical question.

For future directions: validation on the Temple University Hospital EEG Corpus [17] is the natural next step, even though it was not collected for tumor detection purposes. Architecturally, replacing pre-extracted PAC features with raw waveform inputs processed through convolutional front-ends could allow the model to discover coupling features the Modulation Index misses. Deployment on portable EEG headsets is a longer-term vision, but model compression through quantization and pruning could make the parameter count compatible with embedded hardware.

VIII. CONCLUSION

What this project set out to ask is simple even if the answer is not: can a machine learn to detect the neurological shadow a brain tumor casts before the tumor itself is visible on a scan? Based on the simulation results, the answer is a cautious yes — at least in principle.

The 87.4% accuracy and 0.93 AUC-ROC figures are not clinical benchmarks. They are demonstrations of viability. What they show is that the PAC-based signature is learnable, that a Transformer is well-suited to the temporal structure of this data, and that combining EEG features with behavioral biomarkers adds genuine discriminative value rather than just noise. Each of these is a meaningful finding on its own.

Brain tumors kill people partly because we find them too late. The idea that we might catch them functionally — through what the brain does rather than what the brain looks like — is not a new one. What is new is having the computational tools to potentially make it real. Getting from here to clinical deployment will take years, real cohort data, and regulatory work that dwarfs what a simulation study can claim. But the signal is there. Whether it is strong enough to build something clinically useful on remains to be found out, and this work suggests that question is worth pursuing seriously.

DOI: 10.48175/IJARSCT-34249

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 17, April 2026
Impact Factor: 8.2

REFERENCES

[1] R. Stupp et al., "Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma," New England Journal of Medicine, vol. 352, no. 10, pp. 987–996, 2005.
[2] R. T. Canolty and R. T. Knight, "The functional role of cross-frequency coupling," Trends in Cognitive Sciences, vol. 14, no. 11, pp. 506–515, 2010.
[3] L. Bao et al., "Disrupted gamma and theta oscillatory coupling in glioma patients," NeuroImage: Clinical, vol. 28, 102415, 2020.
[4] A. Vecht et al., "Seizures as a sign of brain tumor: mechanisms and biomarkers," Frontiers in Neurology, vol. 11, 582160, 2020.
[5] J. Torous et al., "New tools for new research in psychiatry: a customizable smartphone research platform," JMIR Mental Health, vol. 3, no. 2, e16, 2016.
[6] A. Jimenez-Marin et al., "Phase-amplitude coupling analysis of ECoG identifies tumor-related oscillatory disruptions," Journal of Neural Engineering, vol. 18, no. 4, 046019, 2021.
[7] S. Mun, E. Park, and K. Park, "Resting-state EEG features for classification of brain tumors," Biomedical Signal Processing and Control, vol. 62, 102089, 2020.
[8] U. R. Acharya et al., "Automated EEG-based screening using deep convolutional neural network," Computer Methods and Programs in Biomedicine, vol. 161, pp. 103–113, 2018.
[9] V. J. Lawhern et al., "EEGNet: a compact convolutional neural network for EEG-based BCIs," Journal of Neural Engineering, vol. 15, no. 5, 056013, 2018.
[10] Y. Wang, A. Liang, and D. Xie, "BrainBERT: self-supervised representation learning for intracranial recordings," arXiv:2302.14367, 2023.
[11] A. Zhan et al., "Using smartphones and machine learning to quantify Parkinson disease severity," JAMA Neurology, vol. 75, no. 7, pp. 789–795, 2018.
[12] K. C. Fraser, J. A. Meltzer, and F. Rudzicz, "Linguistic features identify Alzheimer's disease in narrative speech," Journal of Alzheimer's Disease, vol. 49, no. 2, pp. 407–422, 2016.
[13] M. J. Taphoorn and M. Klein, "Cognitive deficits in adult patients with brain tumors," Lancet Neurology, vol. 3, no. 3, pp. 159–168, 2004.
[14] E. Niedermeyer and F. H. Lopes da Silva, Electroencephalography: Basic Principles, Clinical Applications, 5th ed. Philadelphia: Lippincott Williams & Wilkins, 2005.
[15] A. B. Tort et al., "Measuring phase-amplitude coupling between neuronal oscillations of different frequencies," Journal of Neurophysiology, vol. 104, no. 2, pp. 1195–1210, 2010.
[16] A. Vaswani et al., "Attention is all you need," Advances in Neural Information Processing Systems, vol. 30, 2017.
[17] I. Obeid and J. Picone, "The Temple University Hospital EEG Data Corpus," Frontiers in Neuroscience, vol. 10, 196, 2016.