ORIGINAL ARTICLE

Long-Term EEG-Based Modeling and Classification of Migraine Phases Using Hidden Markov Models

Safoura Ashoorisefat [1], Mohammad Pooyan [2*], Alia Saberi [1]

[1] *Department of Biomedical Engineering, Qa., c., Islamic Azad University, Qazvin, Iran*

[2] *Biomedical Engineering Department, Engineering Faculty, Shahed University, Tehran, Iran*

[3] *Neurology Department, Guilan University of Medical Sciences, Rasht, Iran*

*Corresponding Author: Mohammad Pooyan

Received: 05 August 2025 / Accepted: 06 December 2025

Email: pooyan@shahed.ac.ir

Abstract

Purpose: Migraine is a complex neurological disorder characterized by dynamic alterations in brain activity during multiple phases: interictal (baseline), preictal, ictal, and postictal. This study aims to model and differentiate these migraine phases using Electroencephalogram (EEG) and a Hidden Markov Model (HMM).

Materials and Methods: EEG signals were collected from each subject over several months through frequent, short sessions often multiple times per day. The recordings were temporally aligned with self-reported symptom diaries, allowing for precise labeling of migraine phases. A comprehensive set of features was extracted from the EEG signals, including spectral, temporal, and nonlinear measures such as Dynamic Mode Decomposition (DMD) and Katz Fractal Dimension (KFD) across various frequency bands. Despite the limited number of participants, the dense long-term recordings captured multiple migraine episodes, enabling reliable phase modeling.

Results: The HMM identified migraine-specific neural patterns, achieving an average classification accuracy of approximately 87% for all 15 patients, with individual patient performance ranging from 70% to 95%, depending on signal length and normalization. Only three patients are shown in detail in the results section as illustrative examples.

Conclusion: The HMM identified distinguishable neural patterns corresponding to migraine states, suggesting the feasibility of temporal EEG modeling for clinical applications in personalized migraine management.

Keywords: Electroencephalogram Signal; Migraine; Hidden Markov Model; Dynamic Mode Decomposition; Wavelet Transform; Phase Classification; Prediction.

DOI:

1. Introduction

Migraine is a complex and disabling neurological disorder that affects a significant portion of the global population, with a higher prevalence among women [1]. Characterized by recurring headaches often accompanied by nausea, photophobia, phonophobia, and, in some cases, transient neurological disturbances (aura), migraine is typically classified into episodic (fewer than 15 headache days per month) and chronic forms (15 or more headache days per month) [2, 3]. The chronic nature and variable symptomatology make early diagnosis and personalized intervention both challenging and clinically significant.

Diagnosis primarily relies on patient-reported symptoms and clinical interviews, which introduces subjectivity and impedes timely intervention [4, 5]. Notably, many patients tend to ignore or underreport early symptoms (prodrome or aura), reducing the window for effective treatment. This emphasizes the growing need for objective, continuous, and patient-specific diagnostic approaches that can detect distinct migraine phases and forecast attacks before peak onset.

Electroencephalography (EEG) has emerged as a promising non-invasive tool to capture the brain's electrophysiological activity with high temporal resolution. Prior studies have shown that EEG can reflect subtle changes during different migraine phases, including the interictal, preictal, ictal, and postictal states [6-9]. Moreover, resting-state EEG abnormalities in migraine patients have been documented, highlighting the potential of EEG to reveal underlying pathophysiological changes [10].

To overcome these limitations, recent research has explored advanced feature extraction techniques such as Empirical Mode Decomposition (EMD) [11], Variational Mode Decomposition (VMD) [12], and fractal dimension analysis [8, 13], as well as modern classification algorithms. For instance, Akben *et al.* (2016) [14] used machine learning to classify migraine from EEG, but did not model the temporal sequence of phases. Likewise, Angelini *et al.* [15] focused on phase synchronization in visual evoked potentials during migraine, yet failed to establish predictive frameworks.

Hidden Markov Models (HMMs), known for their capability to model sequential data with hidden states, have been widely applied in speech recognition, epilepsy detection, and emotion classification [16-18]. In migraine research, their use remains limited. Jindal *et al.* [7] applied non-linear EEG features and simple classifiers for migraine detection, but did not integrate phase-specific modeling. Dash *et al.* [17, 19-21] implemented an HMM framework for seizure prediction, showing the potential of this method in capturing dynamic neurological states—though their work was not migraine-specific.

In this study, a novel framework based on Hidden Markov Models (HMMs) is developed using long-term EEG data recorded from clinically diagnosed migraine patients, aiming to accurately identify different migraine phases in real time. Unlike previous approaches that mostly focused on cross-sectional analysis or generic classification, the proposed framework leverages a diverse set of statistical, frequency-based, nonlinear, and dynamical features—including entropy, fractal dimension, and Dynamic Mode Decomposition (DMD) [22, 23]—to achieve a deeper and more precise analysis of brain signals during migraine phases. By combining these features with the temporal modeling capacity of HMMs, the system is capable of learning patient-specific patterns and predicting phase transitions. This intelligent approach may facilitate early detection and targeted clinical interventions, reducing reliance on subjective assessments, enhancing diagnostic accuracy, and advancing personalized medicine in migraine management.

2. Materials and Methods

2.1. Participants and Study Method

The study population consisted of patients diagnosed with migraine headaches, recruited between 2021 and 2023. A total of 15 patients participated in the study, including 9 females and 6 males, aged between 25 and 40 years. The research methodology was entirely patient-centered and long-term in design. A two-part questionnaire was administered. The first part collected demographic data, including age, gender, education level, occupation, and duration of illness. The second part

was based on the International Headache Society (IHS) criteria, covering clinical aspects such as average frequency of migraine attacks per month and year, headache laterality, and associated symptoms [24].

Participants presenting with headache complaints were referred to a neurology clinic, where the diagnosis of migraine was confirmed by a neurologist. Informed consent was obtained from all participants. Ethical approval for this study was granted by the Ethics Committee of Iran University of Medical Sciences.

Importantly, EEG data collection was performed repeatedly over an extended monitoring period. Each patient was followed individually for several consecutive months. EEG signals were recorded not only at predetermined intervals but also dynamically based on the patient's condition and symptom fluctuations. Participants were instructed to maintain detailed symptom diaries, and EEG sessions were scheduled accordingly—sometimes on a daily basis and, when necessary, even multiple times per day. This approach allowed for the collection of dense, phase-specific data reflective of real-life migraine cycles, enhancing the temporal accuracy and clinical relevance of the recordings. The preictal phase was defined according to IHS (2018) [24] as the period preceding headache onset, typically from several hours up to 72 hours. In our study, preictal windows were individualized based on patient-reported prodromal symptoms and aligned EEG recordings, rather than using a fixed time window. Patient demographics, monitoring periods, EEG session frequency during migraine and non-migraine days, average session duration, and monthly headache days are summarized in Table 1. Inclusion criteria were: clinically diagnosed migraine without aura, age 25–40 years, and ability to maintain a daily symptom diary. Exclusion criteria included other neurological disorders, use of neuroactive medications, or incomplete EEG/diary data. Table 1 provides an overview of data density and temporal coverage for accurate interpretation and reproducibility.

2.2. Tools and Equipment

EEG recordings were performed in accordance with the international 10–20 electrode placement system. Recordings were conducted in a quiet, dark room to minimize external stimuli and enhance signal reliability. Each session lasted approximately 15 minutes, during which participants were instructed to lie still with their eyes closed to reduce artifacts.

The sampling rate was set at 512 Hz. Although each migraine phase (interictal, preictal, ictal, postictal) may span several hours to days, recordings were performed repeatedly over multiple months, capturing EEG signals from each subject daily or even several times per day, depending on symptom presentation. This long-term and high-frequency recording strategy

Table 1. Patient demographics and EEG recording summary

patient	sex	Age	Monitoring period (months)	Migraine frequency (day/ month)	Sessions per day (pre/ictal/post)	Sessions per day (interictal)	TotalEEG sessions	Ave. session duration (min)
P1	F	32	3	8	3-5	1	150	15
P2	M	40	3	9	3-5	1	165	15
P3	F	28	3	7	3-5	1	140	15
P4	F	35	3	8	3-4	1	148	15
P5	F	30	2	6	3-4	1	98	15
P6	M	38	2	7	3-5	1	112	15
P7	F	26	3	10	3-5	1	180	15
P8	M	40	3	8	3-4	1	150	15
P9	F	34	2	7	3-5	1	108	15
P10	M	36	3	9	3-4	1	162	15
P11	F	29	2	6	3-5	1	104	15
P12	M	40	3	8	3-5	1	156	15
P13	F	27	3	7	3-4	1	144	15
P14	M	40	2	6	3-5	1	100	15
P15	F	31	3	9	3-5	1	168	15

allowed for detailed temporal tracking of migraine dynamics.

To map EEG data with clinical migraine phases, participants maintained structured symptom diaries throughout the study. They documented physical and neurological sensations experienced before the onset of migraine, during headache episodes, and after symptom resolution. These diaries enabled precise phase annotation for each EEG segment and contributed to constructing personalized migraine timelines. Only patients with migraine without aura were included in this study.

2.3. Data Labelling and Temporal Alignment

Accurate identification of migraine phases is essential for modeling their temporal dynamics using EEG. To ensure reliable labeling, each participant maintained a structured symptom diary over several months. The diary included timestamped reports of migraine onset and resolution, pain intensity, and a wide range of emotional and neurological symptoms associated with each phase—such as fatigue, irritability, sensitivity to light or sound, and cognitive changes. These continuous self-reports enabled fine-grained temporal alignment between EEG recordings and the four clinically defined migraine phases: interictal (baseline), preictal, ictal, and postictal.

EEG recording sessions were scheduled based on the participant's diary entries or their perceived state at the time. For instance, when a subject reported early warning signs suggestive of the preictal phase, an EEG was recorded shortly after to capture the relevant neural activity. Similarly, ictal and postictal recordings were conducted during or immediately after headache episodes, while interictal data were collected on symptom-free days.

This personalized and long-term data acquisition strategy enabled the model to learn phase-specific neural signatures with high temporal resolution. The tight correspondence between symptom timelines and EEG recordings supported robust phase labeling, allowing the HMM to be trained on data that authentically represented the transitions and dynamics of the migraine cycle—even within a limited patient population.

2.4. EEG Data Preprocessing

Electroencephalogram (EEG) signals are often contaminated by various types of noise and artifacts, particularly the 50 Hz power line interference. To mitigate this, a notch filter was applied to suppress line noise artifacts [8, 25, 26]. Additionally, EEG signals frequently contain non-neural artifacts such as those originating from muscle activity, eye blinks, and other physiological or environmental sources. Therefore, advanced artifact removal techniques are crucial for reliable analysis.

In this study, two signal decomposition techniques were employed during the preprocessing phase: Independent Component Analysis (ICA) and wavelet-enhanced ICA (wICA). These methods are designed to separate neural signals from non-neural artifacts while preserving underlying brain activity.

ICA assumes that the observed EEG signal is a linear mixture of independent source signals, modeled as follows (Equation 1):

$$\boldsymbol{X}(t) = \boldsymbol{A}.\boldsymbol{S}(t) + \boldsymbol{n}(t) \quad (1)$$

where $\boldsymbol{S}(t) = [s_1(t)\, s_2(t) \ldots s_J(t)]^T$ represents the J latent source signals, A is the unknown JI×J mixing matrix, and $\boldsymbol{n}(t) = [n_1(t) n_2(t) \ldots n_I(t)]^T$ denotes additive noise. The objective of ICA is to estimate a demixing matrix W that retrieves the sources (Equation 2):

$$\widehat{\boldsymbol{S}}(t) = W.\boldsymbol{X}(t) \quad (2)$$

The ICA-based artifact correction process includes the following steps [25]:

1. Filtering: The raw EEG data X(t) is first filtered using a notch filter to remove power line interference.
2. Decomposition: ICA (e.g., FastICA) is applied to decompose the filtered data into statistically independent components $\hat{s}_{ICA}(t)$.
3. Component Selection: Artifact components are identified via visual inspection or automated methods. Only non-artifactual components $\hat{s}_{ICA-EEG}(t)$ are retained.

4. Signal Reconstruction: The clean EEG signal is reconstructed by applying the inverse ICA transformation (Equation 3):

$$\hat{\boldsymbol{X}}_{clean}(t) = \boldsymbol{W}^{-1}\hat{\boldsymbol{S}}_{ICA-EEG}(t) \quad (3)$$

Although ICA is effective, it may result in the unintended removal of useful brain signals. To address this limitation, wavelet-enhanced ICA (wICA) was employed as an additional step. wICA enables partial denoising of identified artifact components instead of complete removal, thus minimizing the risk of discarding valuable neural information. In this hybrid approach, wavelet denoising is performed on artifact-laden independent components before signal reconstruction, helping to preserve subtle neural dynamics [7, 8, 25].

In this study, the Daubechies 4 (db4) wavelet family was used with a decomposition level of 5, and thresholding was performed using the VisuShrink algorithm to suppress artifacts while preserving neural signals [27, 28].

2.5. Feature Extraction

After the signal preprocessing stage, the analysis stage is performed according to the following flowcharts (Figure 1 and Figure 2).

Figure 1 shows the steps involved in training the model, including EEG signal segmentation, feature extraction, feature selection using a Genetic Algorithm, and HMM-based parameter training.

Figure 2 depicts the testing workflow, where preprocessed EEG segments are input to the trained HMM models, and phase classification is performed based on the computed log-likelihood for each migraine phase.

These figures provide a visual summary of the complete training and testing processes used in this study.

After denoising, wavelet decomposition was applied to segment the EEG into five standard frequency bands: delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (>30 Hz). Each signal was divided into overlapping 2-second and 5-second segments, and a comprehensive set of 26 features was extracted from each segment across all frequency bands.

The extracted features were categorized as follows:

- Statistical features: standard deviation, skewness, and kurtosis.
- Frequency-domain features: 14 features, including 7 major peaks and 7 frequency

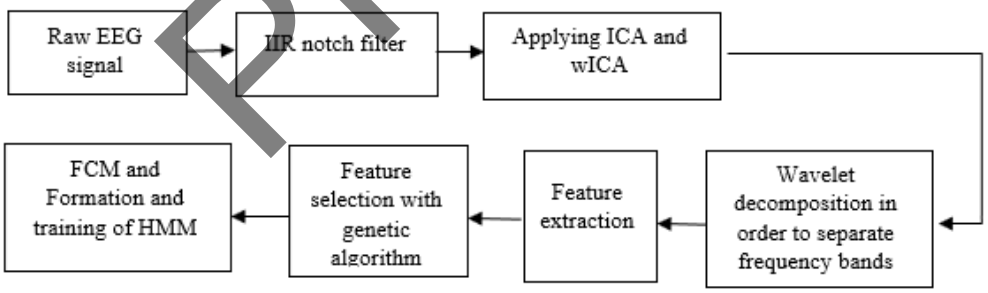

Figure 1. Steps of training the proposed model

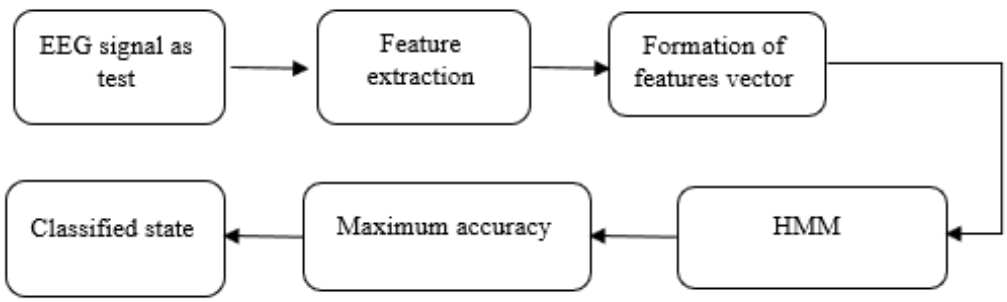

Figure 2. Steps of testing the proposed model

locations from the power spectral density (PSD).

- Wavelet-based features: 5 features derived from multilevel decomposition.
- Nonlinear features: Shannon entropy and Katz Fractal Dimension (KFD) [23].
- Spatiotemporal features: Dynamic Mode Decomposition (DMD) power [22].

These steps summarize the preprocessing and feature extraction stages before applying advanced spatiotemporal analysis with DMD.

2.6. Dynamic Mode Decomposition (DMD)

DMD is a data-driven technique that extracts spatiotemporal patterns from high-dimensional data. Originally developed in fluid mechanics [19], DMD identifies dominant dynamic modes that represent both spatial structures and their temporal evolution.

The DMD process involves the following steps:

1. Construct time-shifted data matrices (Equation 4):

$$X=[x_1, x_2,\ldots,x_{m-1}],\quad X'=[x_2, x_3,\ldots,x_m] \tag{4}$$

2. Perform Singular Value Decomposition (SVD) (Equation 5):

$$X = U\Sigma V^* \tag{5}$$

3. Estimate the low-dimensional approximation of the linear map (Equation 6):

$$\tilde{A} = U^* \acute{X} V \sum^{-1} \tag{6}$$

4. Perform eigen-decomposition (Equation 7):

$$\tilde{A}W = W\Lambda \tag{7}$$

W: eigenvectors

Λ: diagonal matrix of eigenvalues

5. Compute DMD modes (Equation 8):

$$\Phi \approx \acute{X} V \sum^{-1} W \tag{8}$$

Each DMD mode ϕ_i is associated with a corresponding eigenvalue λ_i, characterizing the growth/decay and oscillation frequency of the underlying dynamics. The DMD power was calculated as the magnitude of each mode, representing its contribution to the overall signal structure [22].

2.7. Feature Normalization

To ensure uniformity across features and facilitate model convergence, min-max normalization was applied to all extracted features using the following Equation 9:

$$S_{Nf} = \frac{(S_f - \bar{S}_f)}{(S_{fmax} - S_{fmin})} \tag{9}$$

where:

- S_f : original feature value,
- $\bar{S}_f$: mean of the feature,
- S_{fmax} : maximum value,
- S_{fmin} : minimum value.

This normalization maps the features to a standard range, enhancing their suitability for classification tasks in subsequent modeling steps.

For optimal feature selection, a Genetic Algorithm (GA) was employed to identify the most discriminative features among the extracted set. GA iteratively evaluates subsets of features based on classification performance, selecting those that maximize predictive accuracy for migraine phase detection.

2.8. FCM Clustering

For each migraine phase, 15 minutes of EEG data were recorded and divided into either 450 two-second segments or 180 five-second segments. From each segment, 60 features (12 features for each frequency band) were extracted. Two scenarios were considered: in the first scenario, each frequency band was employed separately, and only features from the selected frequency band were used. In the second

scenario, all features from the five frequency bands were utilized together. Each segment consists of a feature vector that belongs to a cluster. Clustering involves grouping similar points (feature vectors) based on similarity, which is usually measured by distance.

In the FCM (fuzzy C-means) clustering method [23], the cluster membership of each feature vector is assigned based on the degree of proximity to cluster centers. The FCM algorithm was used to cluster the data, and the feature vectors were divided into five clusters for use in the training phase. The number of clusters is independent of the number of classes. In this study, four classes corresponding to the four disease phases were considered, and in each class, EEG segments belonged to different clusters. Although the feature vector space is multidimensional, each vector belongs to one of five clusters. The number of clusters was determined empirically based on the data.

As shown in Figure 3, a two-dimensional scatter plot illustrates the results of Fuzzy C-Means clustering, showing five distinct clusters using different colors. Data points are marked with different shapes to denote class membership.

For model training and evaluation, the feature vectors from FCM clustering were split at the subject level to prevent information leakage. For each participant, 70% of the segments were used for training and 30% for testing. The split was performed chronologically within each migraine phase to preserve temporal structure. All reported accuracies were obtained using 10-fold cross-validation at the segment level. For each fold, feature selection (GA), clustering (FCM), and HMM training were repeated to avoid bias, and the final accuracy represents the average performance across folds.

2.9. Hidden Markov Model

A Hidden Markov Model (HMM) is a statistical model that represents a Markov process with hidden (unobservable) states. It is typically defined by the parameter set $\lambda = (\pi, A, B)$, where π is the vector of initial state probabilities, A is the state transition probability matrix, and B represents the observation probability distribution.

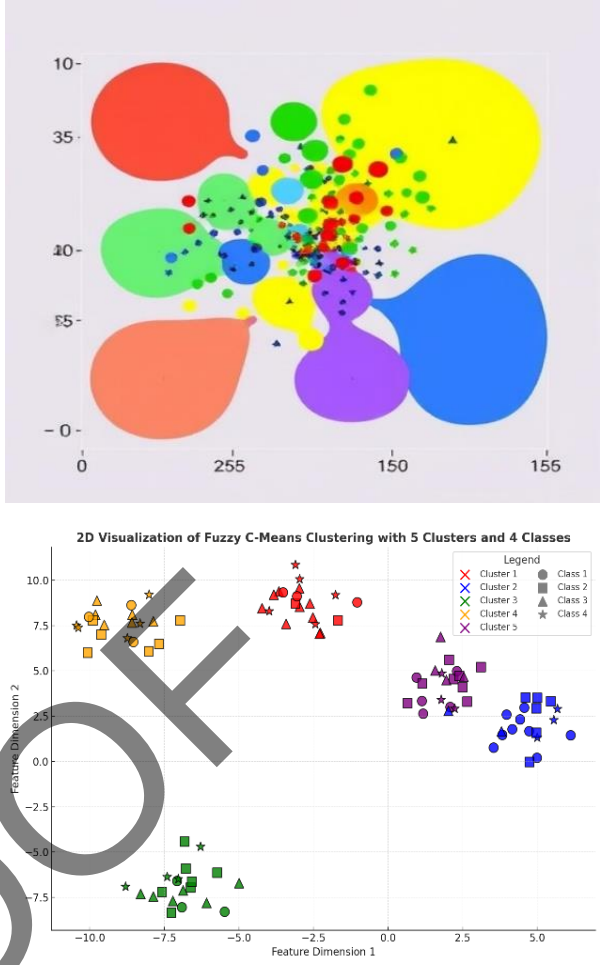

Figure 3. A two-dimensional scatter plot illustrating the results of Fuzzy C-Means clustering. Five distinct clusters are shown using different colors (red, blue, green, yellow, and purple). Data points are also marked with different shapes to denote class membership: Class 1 – circles, Class 2 – squares, Class 3 – triangles, Class 4 – stars. Each class contains data points across multiple clusters, with a higher concentration in one or two dominant clusters

In Figure 4, a typical two-layer HMM with four hidden states (S1 to S4) is illustrated. Each hidden state can emit four distinct observations (O1 to O4). The figure shows both the transition probabilities (A) between hidden states and the emission probabilities (B) associated with observable outputs.

Figure 5 demonstrates how an observable sequence is generated from an underlying hidden state sequence in an HMM. Although only the observations are visible to the observer, the hidden states drive the process. The HMM framework enables inference of the most probable hidden state path given the observed data.

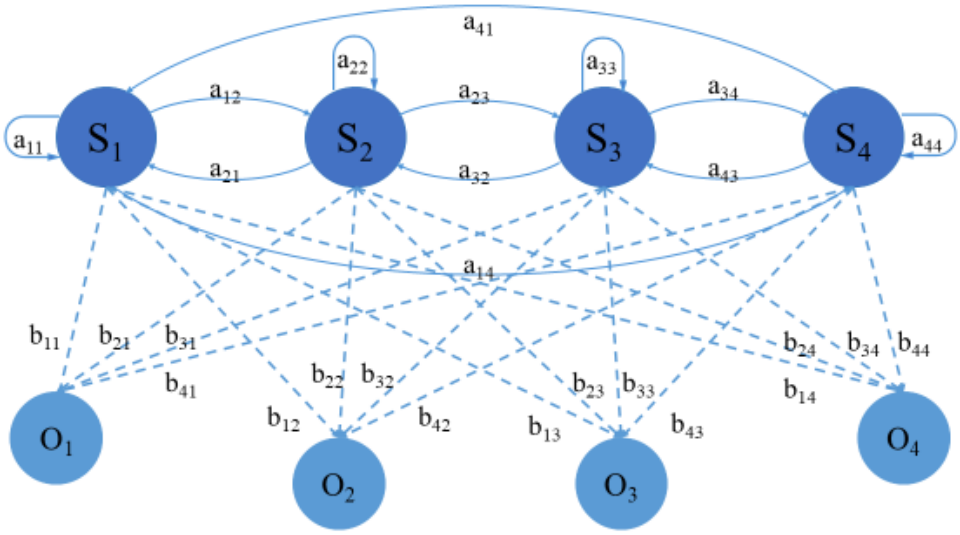

Figure 4. A typical 2-layer HMM with 4 hidden states and 4 observations

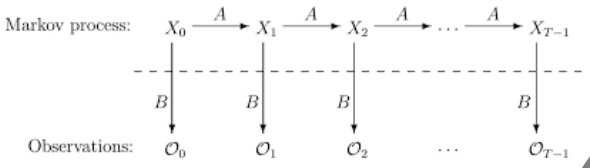

Figure 5. Typical Markov chain representation

2.9.1. HMM-Based Modeling of Migraine Phases Using FCM-Clustering of EEG Features

In this study, a Hidden Markov Model (HMM) was employed to model migraine phases based on EEG feature clustering. The HMM consists of hidden states and observable outputs, where each observation probabilistically depends on its generating hidden state. Feature vectors extracted from EEG signals were clustered using the Fuzzy C-Means (FCM) algorithm into five distinct clusters, each characterized by 12 features.

The model comprises two layers:

- The hidden layer includes four states representing clinically recognized migraine phases: no migraine, prelude, pain, and postlude. State transitions are governed by a transition probability matrix.
- The observation layer contains five symbols corresponding to the FCM clusters, reflecting typical signal patterns during different migraine phases.

In this study, an ergodic HMM structure was employed, allowing transitions between any pair of migraine phases. This choice was made to accommodate the complex and non-linear progression of migraine, ensuring the model could capture atypical phase sequences that may occur in real patients.

Each hidden state has a specific emission probability distribution over the five observation symbols. Model parameters—including transition and emission probabilities—are estimated iteratively via the Expectation-Maximization (EM) algorithm, alternating optimization between layers [16, 23, 27, 29, 30].

The model assumes the Markov property, where the next state depends only on the current state, simplifying the representation of disease dynamics. This assumption aligns with standard practices in biomedical signal modeling [30].

The structure of the HMM for migraine disease is shown in Figure 6.

2.10. Hidden Markov Model Training

The essential matrices for Hidden Markov Model (HMM) training include the initial state probability vector (π), the state transition probability matrix (A), and the emission probability matrix (B) [31]. To define the emission matrix, it is necessary to specify the number of hidden states and observation symbols. In this study, four hidden states were considered,

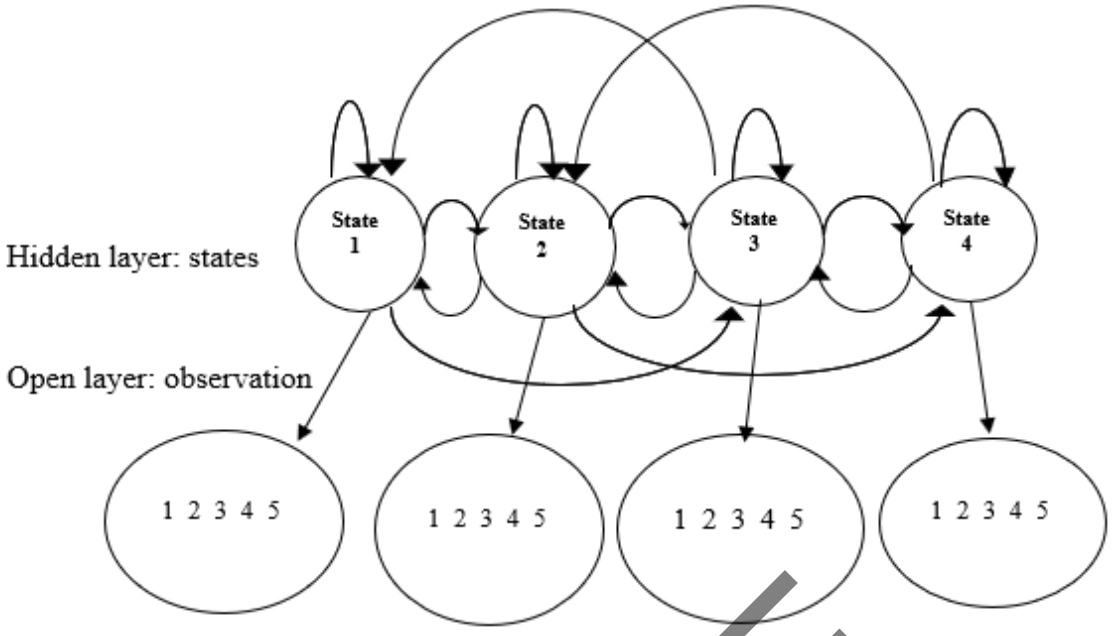

Figure 6. Structure of hidden Markov model for migraine disease

corresponding to the migraine phases, while the observations were based on feature vectors clustered into five groups using the Fuzzy C-Means (FCM) algorithm.

The EEG signals were segmented into frames — 450 frames using two-second framing and 180 frames using five-second framing. After feature extraction and optimization via a genetic algorithm, 12 optimal features were selected. To construct the emission matrix, the mean and variance vectors of each cluster were calculated from the FCM clustering results. Since the number of observations within each cluster may vary, emission probabilities were estimated by calculating the distance between each observation vector and the corresponding cluster centers, assigning observation symbols accordingly. Subsequently, the frequency of observations within each cluster per state was counted and normalized to obtain emission probabilities.

For the initial state transition probability matrix, equal probabilities (0.25) were assigned to each migraine state, reflecting a uniform initial assumption. The Baum-Welch algorithm, an Expectation-Maximization (EM) based approach, was employed to iteratively estimate and refine the transition and emission probability matrices [31]. This algorithm utilizes forward and backward procedures to maximize the likelihood of observed data given the model parameters.

Once the initial probability, transition, and emission matrices were defined, the algorithm then iteratively updated these parameters by computing intermediate variables, including forward and backward probabilities, to optimize model fitting to the observed data [31]. In this study, HMMs were trained sequentially across patients. Initially, the model was trained using data from Patient 1, and then the learned parameters were gradually updated as data from Patients 2 to 15 were incorporated. For each migraine phase, feature vectors from all patients were combined to form the final emission matrices (λ), resulting in four generalized HMMs corresponding to the four phases. This approach allowed the model to capture both patient-specific patterns and cohort-level variability. During testing, 30% of the data were held out for validation to evaluate the model's generalization across patients.

As shown in Figure 7, the flowchart depicts the roles and relationships of the key intermediate variables ζ (zeta), γ (gamma), α (alpha), and β (beta) in the Baum-Welch algorithm for model parameter estimation.

2.10.1. Forward Algorithm

As shown in Figure 6, several variables must be calculated for migraine modeling. The training algorithm for the Hidden Markov Model (HMM) uses these variables to estimate model probabilities. To

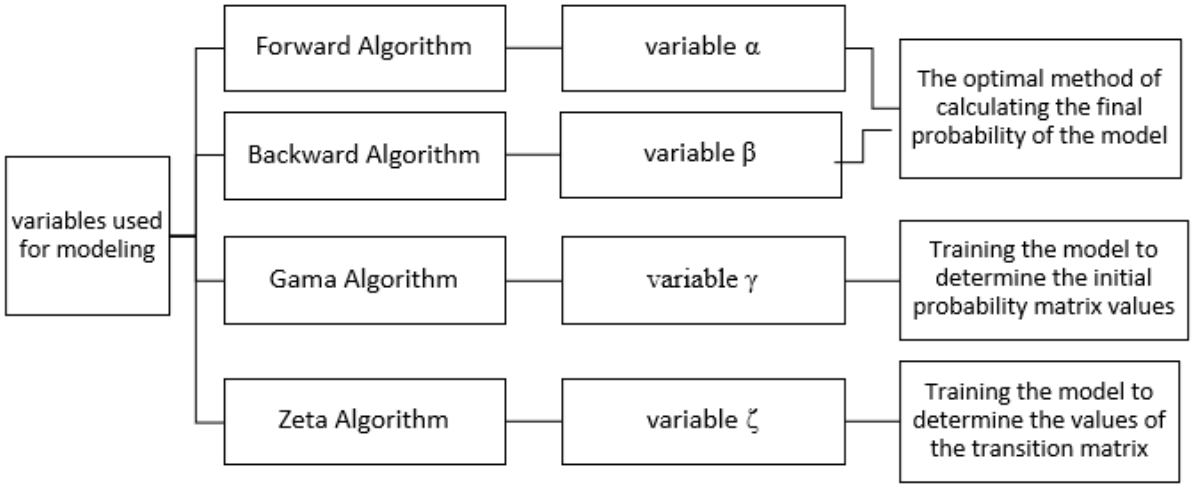

Figure 7. Flowchart depicting the roles and relationships of the key intermediate variables ζ (zeta), γ (gamma), α (alpha), and β (beta) in the Baum-Welch algorithm for model parameter estimation

compute the final probability of the model, an optimal method must be selected. Ignoring repeated states in the selection of the state sequence, the forward algorithm applies a recursive approach to calculate the total probability of all possible hidden state sequences leading to a given observation sequence.

The forward algorithm progresses sequentially through time, computing the probability of partial observations up to time t for each state j, based on the transition and emission probabilities. The calculation follows the inductive relation below, ultimately yielding the forward variable $\alpha_j(t)$ [29, 32] (Equation 10):

$$\begin{aligned} &\alpha_j(t) \leftarrow 0 \quad t = 0 \quad j \neq initial\ state \\ &\alpha_j(t) \leftarrow 1 \quad t = 0 \quad j = initial\ state \\ &\alpha_j(t+1) \leftarrow \sum_j [\alpha_i(t) \times aatr_i] \times bbtr_{jo(t)} \end{aligned} \tag{10}$$

$$P(O|\lambda) \leftarrow \sum_{i=1}^{N} \alpha_T(i)$$

2.10.2. Backward Algorithm

The backward algorithm complements the forward algorithm by calculating the probability of partial observation sequences from the end time T backward to time t for each state i. It sums the probabilities of all possible future state sequences given the current state and observations [16].

This recursive process computes the backward variable $\beta_i(t)$ as follows (Equation 11):

$$\begin{aligned} &\beta_i(t) \leftarrow 0 \quad t \neq T \\ &\beta_i(t) \leftarrow 1 \quad t = T \\ &\beta_i(t) \leftarrow \sum_j [\beta_i(t+1) \times aatr_i] \times bbtr_{jo(t+1)} \end{aligned} \tag{11}$$

The total likelihood of the observation sequence given the model parameters λ=(π,A,B) is then calculated by (Equation 12):

$$P(O|\lambda) \leftarrow \sum_{j=1}^{N} \pi_j \beta_1(i) b_j(O_1) \tag{12}$$

where π_j is the initial probability of state j.

Together, the forward and backward algorithms allow for efficient evaluation of the likelihood of observed data sequences, which is essential for parameter estimation and model training in HMM frameworks.

2.10.3. Parameter Estimation Using Gamma and Zeta Variables

To estimate the parameters of the Hidden Markov Model (HMM), two key intermediate variables, gamma (γ) and zeta (ζ), were calculated using the Baum-Welch training procedure [16, 30]. These variables facilitate the re-estimation of the model's transition and emission probabilities based on observed EEG sequences.

The gamma variable $(\gamma_t(i))$ represents the probability of being in state i at time t given the observation sequence and the current model parameters. Its inputs include the transition matrix A, the emission matrix B, the initial state distribution π, and the observation sequence, combined through the forward and backward variables.

The zeta variable ($\zeta_t(i.j)$ indicates the probability of being in state i at time t and transitioning to state j at time t+1, again conditioned on the observations and model parameters. This variable is critical for updating the transition matrix A.

The re-estimation formulas are as follows (Equation 13):

$$\gamma_t(i) \quad \frac{\alpha_t(i) \times \beta_t(i)}{\sum_{m=1}^{N} \alpha_t(m) \times \beta_t(m)}$$
$$1 =< i =< N$$
$$1 =< t =< T$$
$$\zeta_t(i.j) \quad \frac{\alpha_t(i) \times a_{ij} \times b_j(O_{t+1}) \times \beta_{t+1}(j)}{\sum_{m=1}^{N} \alpha_t(m) \times \beta_t(m)} \tag{13}$$
$$1 =< i =< N$$
$$1 =< t =< T$$

These probability estimates were then used to update the model parameters iteratively until convergence. The gamma values contributed to the re-estimation of the initial state distribution and the emission matrix, while the zeta values were used to update the transition matrix.

2.10.4. Re-Estimation of the Initial Probability Matrix (π) Using Gamma Values

The initial state probability matrix π was re-estimated using the Baum-Welch algorithm and a set of training sequences [29]. The gamma variable $\gamma_t(i)$ represents the probability of the HMM being in state i at time t, given the observation sequence and the current model parameters. For re-estimating π, gamma values were specifically computed at time t=1for each training sequence.

For each sequence l∈{1,2,...,L}, the value $\gamma_1^L(i)$—the probability that the model starts in state i for the l-th sequence—was calculated. The updated initial state probability for state i, denoted $\bar{\pi}_i$, was then obtained by averaging across all L training sequences, as follows (Equation 14):

$$\bar{\pi}_i = \frac{1}{L}\sum_{l=1}^{L} \gamma_1^L(i) \tag{14}$$

To compute this, each sequence from the training dataset was processed by extracting a 12-dimensional optimal feature vector to form the observation sequence. Based on this sequence, the emission probability matrix B was constructed via clustering methods, while the initial transition matrix A_{init} and the initial state distribution π_{init} were predefined.

The forward-backward algorithm was then applied to compute gamma values at t=1. These values were accumulated and averaged over all training sequences using Equation 14, yielding the updated estimates of π for all states i∈{1,2,...,N}, where N=4 in this study [29].

The flowchart of this training process is illustrated in Figure 7, which summarizes the main steps involved in updating the initial state probability matrix using the Baum-Welch algorithm.

2.10.5. Re-Estimation of Transition and Emission Probability Matrices:

The transition probability matrix A was re-estimated using the zeta variable (ζ), which quantifies the probability of transitioning from state i at time t to state j at time t+1, conditioned on the observation sequence and current model parameters. This value represents part of the joint probability over the entire sequence, contributing to the estimation of transitions across all states [16, 30].

For each training sequence of length T, $\zeta_t^L(i.j)$ and $\gamma_t^L(i)$ were computed across time. These were summed across all training sequences L, and the updated transition probability $\bar{a}_{ij}$ was calculated using Equation 15:

$$\bar{a}_{ij} = \frac{\sum_{l=1}^{L}\sum_{t=1}^{T-1} \zeta_t^L(i.j)}{\sum_{l=1}^{L}\sum_{t=1}^{T-1} \gamma_t^L(i)} \tag{15}$$

To re-estimate the emission probability matrix B, Gaussian distribution functions were utilized. Specifically, the multivariate normal distribution was used to model the likelihood of observing a feature vector O_t in state j. This required computation of the

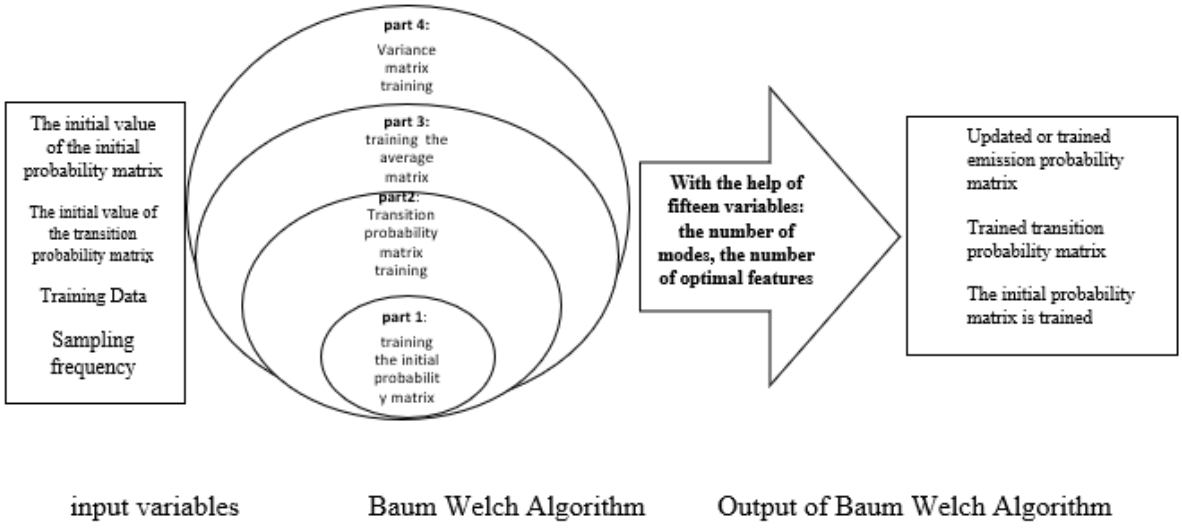

Figure 8. Overview of the training procedure for the Markov Migraine Model using the Baum-Welch algorithm. The diagram illustrates the iterative re-estimation of model parameters (π, A, B) based on gamma (γ) and zeta (ζ) variables computed from observation sequences

mean vector μ_j and covariance matrix $\sum_j$ for each state j [33].

The emission probability $b_j(O_t)$ is given by (Equation 16):

$$b_j(O_t) = \aleph(O_t.\mu_j.\sum_j) = \frac{1}{2\pi^{\frac{D}{2}}|\sum_j|^{\frac{1}{2}}} e^{-\frac{1}{2}(O_t-\mu_j)^T\sum_j^{-1}(O_t-\mu_j)} \quad (16)$$

The parameters were updated using the following Equations 17, 18:

$$\bar{\mu}_j = \frac{\sum_{l=1}^{L}\sum_{t=1}^{T}\gamma_t^l(j)\cdot O_t}{\sum_{l=1}^{L}\sum_{t=1}^{T}\gamma_t^l(i)} \quad (17)$$

$$\bar{\sum}_j = \frac{\sum_{l=1}^{L}\sum_{t=1}^{T}\gamma_t^l(j)\cdot(O_t-\bar{\mu}_j)(O_t-\bar{\mu}_j)^T}{\sum_{l=1}^{L}\sum_{t=1}^{T}\gamma_t^l(i)} \quad (18)$$

During training, each observation sequence was used to extract a 12-dimensional feature vector. The emission probabilities were estimated using these features via clustering and Gaussian modeling. The initial matrices A_{init} and π_{init} were set, and the forward-backward algorithm was executed to compute γ and ζ. This process was repeated for all L training sequences, and parameter updates were performed iteratively.

As the training proceeded, the final log-likelihood of the model was computed at each iteration. Once the model parameters (initial, transition, and emission) stabilized—meaning the log-likelihood ceased to improve beyond a threshold—the training process was considered converged [16, 30, 31].

The complete flow of this training procedure is depicted in Figure 8, summarizing the core stages of model training using the Baum-Welch algorithm.

3. Results

3.1. Model Training and Evaluation

Hidden Markov Models (HMMs) were trained individually for each patient using the Baum-Welch algorithm. The optimized parameters for each model were denoted by λ and stored for evaluation. During testing, the trained models were applied to EEG sequences to infer the most probable migraine phase. To mitigate numerical underflow in the product of probabilities during inference, logarithmic likelihoods were used for decision-making instead of raw probability values.

EEG signals were recorded a long-term period from fifteen patients suffering from high-frequency episodic migraine (attack intervals <15 days). For each patient, four migraine phases—interictal, preictal, ictal, and postictal—were labeled, enabling subject-specific HMM training.

Each session was divided into short-duration windows of 2 seconds and 5 seconds to evaluate the effect of signal length on classification performance.

Performance across four experimental settings is presented in Tables 2–5, where Correct Classification Rates (CCR) are reported for each migraine phase and patient. The configurations tested were:

- Unbalanced & Unnormalized data (Table 2)
- Balanced & Unnormalized data (Table 3)
- Unbalanced & Normalized data (Table 4)
- Balanced & Normalized data (Table 5)

To provide a clear illustration, detailed performance results of three representative patients are presented in Figures 9 and 10 and Tables 2–5, while overall accuracy and trends were computed across all 15 patients. Comparing Tables 2 and 4 (unbalanced data), normalization of feature vectors consistently improved classification accuracy across patients. For example:

- For Patient 1 (2s signals), total CCR increased from 82.0% (unnormalized) to 82.5% (normalized).
- For Patient 2 (2s signals), the CCR slightly decreased from 82.63% to 78.33%, showing that normalization may not benefit all subjects equally in short windows.

On average, normalized features led to improved classification performance by approximately 3%, particularly in longer (5s) recordings.

3.2. Impact of Data Balancing

Balancing the training dataset improved robustness of classification across classes and reduced bias toward more frequent phases. In the balanced and normalized condition (Table 5), the classification performance reached its highest values:

- Patient 2 (5s): CCR = 88.33%
- Patient 3 (2s and 5s): CCR = 95%
- Patient 1 (2s): CCR improved from 82.0% (unbalanced) to 87.5%

It is important to note that in some comparisons (e.g., Patient 2: CCR of 82.63% for 2s vs. 93.68% for 5s), the observed improvement may partly result from increased signal duration rather than balancing alone. This underscores the combined influence of window length and data balancing on performance.

3.3. Overal Accuracy Trends

Best performance was achieved under the balanced and normalized configuration, with an average CCR of ~90% across all patients.

- Lower accuracy was observed in the unbalanced-unnormalized setting, especially for shorter 2s windows.
- Longer signal durations (5s) generally yielded better CCR due to improved temporal context for the HMM.

3.4. Robustness and Limitations

Despite inherent challenges such as inter-session variability, participant fatigue, and environmental noise during EEG acquisition, the proposed Hidden Markov Model (HMM) framework demonstrated strong correspondence with clinically established migraine phase transitions, underscoring its robustness under real-world conditions.

A primary limitation of this study is the relatively small sample size, involving only fifteen patients. However, unlike cross-sectional designs with large cohorts and limited data per subject, this study adopted a long-term approach. Each participant was continuously monitored over several months, capturing multiple complete migraine cycles. EEG recordings were obtained across four distinct phases: interictal (headache-free baseline), preictal (prior to onset), ictal (during migraine), and postictal (following the attack). To ensure accurate phase annotation, participants maintained detailed daily headache diaries, enabling precise temporal alignment between subjective reports and EEG recordings.

This individualized, high-density dataset allowed for fine-grained modeling of migraine dynamics at the single-subject level. While the limited number of participants may restrict generalizability, the depth and temporal resolution of the data provided sufficient statistical power to reliably characterize phase-specific neural patterns. Nevertheless, future work should include cross-subject generalization and

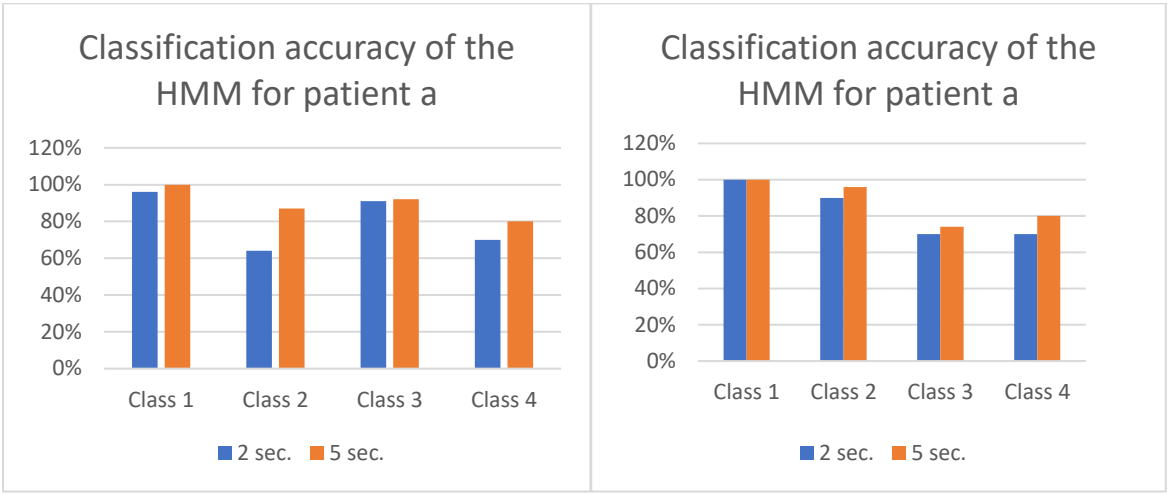

Figure 9. Classification accuracy for unbalanced and unnormalized data (left) and balanced and normalized data (right) for patient a

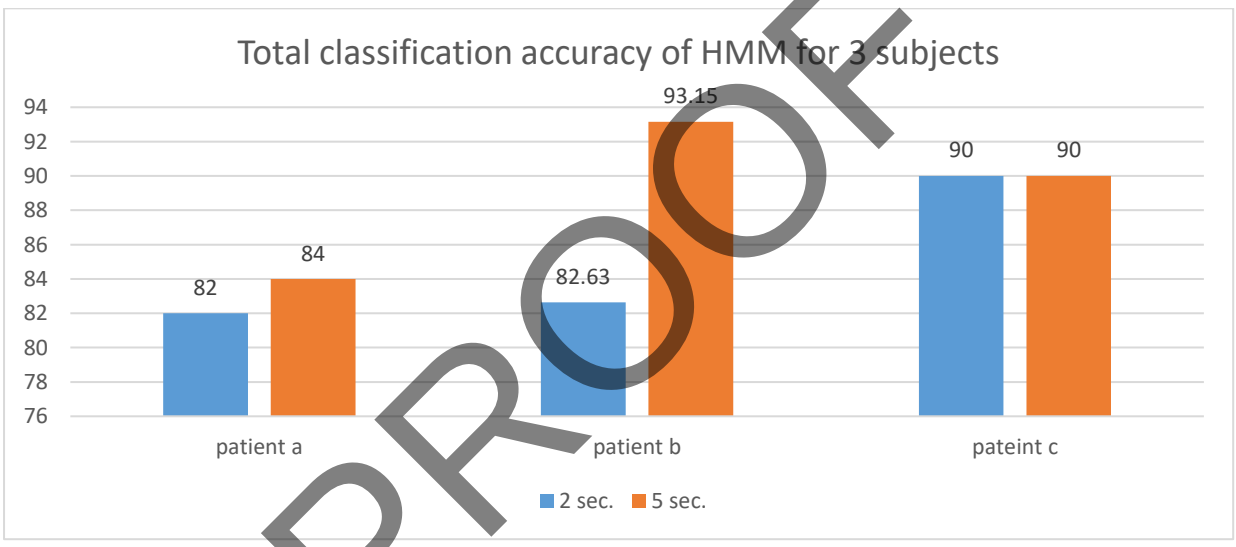

Figure 10. Total classification accuracy of HMM for three representative subjects

Table 2. Correct Classification Rates (CCR) for unbalanced and unnormalized data

Subject ID	Duration of signal recording	Total CCR	Class 1 CCR	Class 2 CCR	Class3 CCR	Class4 CCR
Patient 1	2s	%82.00	%96.66	%64.28	%91.25	%70.00
Patient 1	5s	%84.00	%100	%87.14	%92.00	%80.00
Patient 2	2s	%82.63	%96.66	%70.00	%90.00	%96.66
Patient 2	5s	%93.15	%96.66	%92.50	%90.00	%96.66
Patient 3	2s	%90.00	%100	%80	%100	%96.66
Patient 3	5s	%90.00	%100	%80	%100	%96.66

statistical validation on larger cohorts to further reinforce the current findings.

4. Discussion

This study aimed to model migraine phase transitions using EEG data and a patient-specific

Table 3. Correct Classification Rates (CCR) for balanced and unnormalized data

Subject ID	Duration of signal recording	Total CCR	Class1 CCR	Class2 CCR	Class3 CCR	Class4 CCR
Patient 1	2s	%78.00	%100	%87.14	%98.75	%75.00
Patient 1	5s	%80.00	%100	%100	%80.00	%79.16
Patient 2	2s	%82.63	%100	%73.50	%70.00	%100
Patient 2	5s	%93.68	%100	%92.50	%90.00	%100
Patient 3	2s	%97.00	%100	%100	%100	%73.33
Patient 3	5s	%97.00	%100	%100	%100	%73.33

Table 4. Correct Classification Rates (CCR) for unbalanced and normalized data

Subject ID	Duration of signal recording	Total CCR	Class 1 CCR	Class 2 CCR	Class3 CCR	Class4 CCR
Patient 1	2s	%82.50	%70.00	%70.00	%90.00	%90.00
Patient 1	5s	%85.00	%80.00	%70.00	%90.00	%91.00
Patient 2	2s	%78.33	%76.66	%80.00	%76.66	%96.66
Patient 2	5s	%83.33	%86.66	%81.00	%77.33	%93.33
Patient 3	2s	%85.00	%90.00	%90.00	%80.00	%90.00
Patient 3	5s	%90.00	%100	%92.00	%80.00	%100

Table 5. Correct Classification Rates (CCR) for balanced and normalized data

Subject ID	Duration of signal recording	Total Correct classification Rate	Class1 Correct classification Rate	Class2 Correct classification Rate	Class3 Correct classification Rate	Class4 Correct classification Rate
Patient 1	2s	%87.50	%100	%90.00	%70.00	%70.00
Patient 1	5s	%89.50	%100	%96.00	%74.00	%80.00
Patient 2	2s	%86.66	%76.66	%76.66	%70.00	%93.33
Patient 2	5s	%88.33	%76.66	%80.00	%76.66	%94.00
Patient 3	2s	%95.00	%70.00	%90.00	%90.00	%100
Patient 3	5s	%95.00	%70.00	%100	%100	%100

Hidden Markov Model (HMM) framework. using EEG data and a patient-specific Hidden Markov Model (HMM) framework, similar to previous EEG-based migraine modeling approaches [7,14, 34]. To illustrate the model's behavior, results from three randomly selected patients out of a cohort of fifteen were analyzed and presented. These representative cases demonstrated that the proposed model can reliably differentiate migraine phases—interictal, preictal, ictal, and postictal—based on short EEG segments.

Performance was systematically evaluated under four data preprocessing configurations: unbalanced/unnormalized, balanced/unnormalized, unbalanced/normalized, and balanced/normalized, offering a thorough assessment of the model's robustness. Similar preprocessing and normalization strategies have been reported to improve EEG classification accuracy in migraine and epilepsy studies [8, 11, 17].

Among the selected patients, the highest classification accuracy was achieved with balanced and normalized data, yielding total correct classification rates (CCR) up to 95% (Table 4). Notably, Patient 3 exhibited near-perfect phase classification for both 2- and 5-second EEG segments under this preprocessing approach, highlighting the significant impact of data balancing and normalization in enhancing model generalizability across temporal windows and migraine states. Comparable improvements after normalization were also reported by Kazemi and Katibeh (2018) in pediatric migraine EEG classification [8].

Conversely, unbalanced and unnormalized data (Table 2) led to decreased overall CCRs (82–90%), underscoring the model's sensitivity to data distribution and preprocessing quality. Yet, even

under these suboptimal conditions, the model maintained robust performance, especially in classifying ictal (Class 1) and postictal (Class 2) phases, suggesting the presence of intrinsic, discriminative EEG features characterizing migraine dynamics. This robustness is consistent with prior HMM-based EEG studies showing resilience to noise and variability [17,35].

A critical advantage of this research lies in the use of short-duration EEG segments (2–5 seconds), which supports the feasibility of real-time, portable EEG-based migraine monitoring systems. Such technology could enable early detection of the preictal phase, facilitating timely interventions to mitigate or prevent full-blown attacks. The feasibility of using short EEG windows for migraine or seizure prediction has also been demonstrated in related studies [34, 36].

Furthermore, the integration of detailed daily symptom diaries maintained by patients during recording provided essential clinical context for EEG interpretation. This allowed for precise phase labeling and robust alignment between subjective symptomatology and electrophysiological markers. Consistent symptom patterns identified in the preictal and postictal phases—including fatigue, mood fluctuations, and cognitive disturbances—may serve as valuable predictors for imminent migraine episodes. Future research should explore the fusion of these symptom profiles with EEG features to enhance phase prediction accuracy and enable modeling of ictal pain intensity levels. Similar symptom–EEG correlations were noted by Ashina *et al.* (2021) and Sutherland *et al.* (2019), emphasizing multimodal monitoring in migraine research [1,6].

All participants in this study experienced migraine without aura, thus the dataset lacked aura-related features. Future extensions could incorporate patients with migraine with aura, adding the aura phase as an explicit intermediate state to the current four-state model (interictal, pre-ictal, ictal, postictal). This five-state framework would facilitate earlier forecasting of ictal onset in a broader patient population and improve the personalized applicability of EEG-based migraine monitoring. Previous studies have also shown distinct EEG features in migraine with aura, suggesting the need to model this additional phase [3,37].

Moreover, beyond the ictal phase—during which patients endure moderate to severe pain lasting hours to days—the preictal and postictal phases impose substantial non-pain-related burdens, often spanning much of the month. Symptoms such as fatigue, mood disturbances, cognitive impairment, and functional decline affect patients' quality of life and impact families, social relationships, and work productivity. Expanding patient-specific modeling to reliably detect and predict these debilitating phases holds great promise for enabling early support, personalized interventions, and broader societal benefits. These non-pain-related burdens have been widely reported in migraine populations and contribute significantly to disease-related disability [2,4,38].

In summary, the proposed HMM-based approach exhibits strong potential for accurate and reliable migraine phase detection. With further refinement and validation, this methodology could underpin individualized, clinically practical systems for continuous migraine monitoring and proactive management. Future work will build on existing EEG-based migraine modeling frameworks [34] to develop real-time predictive systems.

5. Conclusion

This study demonstrates the feasibility and effectiveness of utilizing patient-specific Hidden Markov Models (HMMs) for the automatic classification of migraine phases using EEG data. The proposed framework was able to capture temporal patterns during interictal, preictal, ictal, and postictal states, offering a promising approach for modeling the cyclic nature of migraine at an individualized level.

A key finding is the critical importance of data preprocessing—particularly class balancing and signal normalization—in significantly improving model accuracy and stability. Under balanced and normalized conditions, the system achieved superior classification performance across multiple patients and time segments, underscoring the robustness of the approach in realistic clinical scenarios.

The integration of detailed symptom diaries further enriched the modeling process, enabling precise alignment between self-reported experiences and EEG-derived biomarkers. This multimodal design

strengthens the clinical interpretability of the results and lays the groundwork for more comprehensive, hybrid monitoring systems in future research.

Given the chronic and heterogeneous nature of migraine, this methodology provides a strong foundation for the development of non-invasive, real-time migraine monitoring tools. Such systems may enable early detection of high-risk states (e.g., preictal phase), facilitate timely interventions, and ultimately support more personalized management strategies for migraine sufferers.

Future research should focus on validating the proposed model in larger and more diverse cohorts, exploring cross-subject generalization, incorporating additional physiological and behavioral modalities (such as heart rate, sleep patterns, or mood fluctuations), and investigating its integration into wearable platforms for continuous, ambulatory monitoring. Additionally, extending the model to account for the aura phase in migraine with aura subtypes may further enhance its clinical relevance and applicability.

References

1- Heidi G Sutherland, Cassie L Albury, and Lyn R Griffiths, "Advances in genetics of migraine." *The journal of headache and pain,* Vol. 20 (No. 1), p. 72, (2019).

2- Todd J Schwedt, "Chronic migraine." *Bmj,* Vol. 348(2014).

3- Nirosen Vijiaratnam *et al.*, "Migraine: Does aura require investigation?" *Clinical Neurology and Neurosurgery,* Vol. 148pp. 110-14, (2016).

4- Stephen D Silberstein, "Considerations for management of migraine symptoms in the primary care setting." *Postgraduate medicine,* Vol. 128 (No. 5), pp. 523-37, (2016).

5- Abigail Ortiz *et al.*, "Cross-prevalence of migraine and bipolar disorder." *Bipolar Disorders,* Vol. 12 (No. 4), pp. 397-403, (2010).

6- Sait Ashina, Enrico Bentivegna, Paolo Martelletti, and Katharina Eikermann-Haerter, "Structural and functional brain changes in migraine." *Pain and therapy,* Vol. 10 (No. 1), pp. 211-23, (2021).

7- K Jindal *et al.*, "Migraine disease diagnosis from EEG signals using Non-linear Feature Extraction Technique." in *2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, (2018): *IEEE*, pp. 1-4.

8- S Kazemi and P Katibeh, "Comparison of parametric and non-parametric EEG feature extraction methods in detection of pediatric migraine without aura." *Journal of biomedical physics & engineering,* Vol. 8 (No. 3), p. 305, (2018).

9- Louise O'Hare, Federica Menchinelli, and Simon J Durrant, "Resting-state alpha-band oscillations in migraine." *Perception,* Vol. 47 (No. 4), pp. 379-96, (2018).

10- Feng Li, Jing Xiang, Ting Wu, Donglin Zhu, and Jingping Shi, "Abnormal resting-state brain activity in headache-free migraine patients: A magnetoencephalography study." *Clinical Neurophysiology,* Vol. 127 (No. 8), pp. 2855-61, (2016).

11- Rajeev Sharma and Ram Bilas Pachori, "Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode functions." *Expert Systems with Applications,* Vol. 42 (No. 3), pp. 1106-17, (2015).

12- Konstantin Dragomiretskiy and Dominique Zosso, "Variational mode decomposition." *IEEE transactions on signal processing,* Vol. 62 (No. 3), pp. 531-44, (2013).

13- S Sethi and R Upadhyay, "Classification of mental tasks using S-transform based fractal features." in *2017 International Conference on Computer, Communications and Electronics (Comptelix)*, (2017): *IEEE*, pp. 38-43.

14- S Batuhan Akben, Deniz Tuncel, and Ahmet Alkan, "Classification of multi-channel EEG signals for migraine detection." *Biomed Res,* Vol. 27 (No. 3), pp. 743-48, (2016).

15- Leonardo Angelini *et al.*, "Steady-state visual evoked potentials and phase synchronization in migraine patients." *Physical review letters,* Vol. 93 (No. 3), p. 038103, (2004).

16- Lawrence R Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition." *Proceedings of the IEEE,* Vol. 77 (No. 2), pp. 257-86, (1989).

17- Deba Prasad Dash, Maheshkumar H Kolekar, and Kamlesh Jha, "Multi-channel EEG based automatic epileptic seizure detection using iterative filtering decomposition and Hidden Markov Model." *Computers in biology and medicine,* Vol. 116p. 103571, (2020).

18- Kevin P Murphy, Machine learning: a probabilistic perspective. *MIT press*, (2012).

19- Steven Tobochnik, Robyn Fahlstrom, Catherine Shain, Melodie R Winawer, and EPGP Investigators, "Familial aggregation of focal seizure semiology in the Epilepsy Phenome/Genome Project." *Neurology,* Vol. 89 (No. 1), pp. 22-28, (2017).

20- Melodie R Winawer, Robert Connors, and EPGP Investigators, "Evidence for a shared genetic susceptibility to migraine and epilepsy." *Epilepsia,* Vol. 54 (No. 2), pp. 288-95, (2013).

21- Fernando Tenório Gameleira, Luiz Ataíde Jr, and Maria Cristina Falcão Raposo, "Relations between epileptic seizures and headaches." *Seizure,* Vol. 22 (No. 8), pp. 622-26, (2013).

22- Peter J Schmid, "Dynamic mode decomposition of numerical and experimental data." *Journal of fluid mechanics,* Vol. 656pp. 5-28, (2010).

23- James C Bezdek, Robert Ehrlich, and William Full, "FCM: The fuzzy c-means clustering algorithm." *Computers & geosciences,* Vol. 10 (No. 2-3), pp. 191-203, (1984).

24- Headache Classification Committee of the International Headache Society, "The international classification of headache disorders." *Cephalalgia,* Vol. 38 (No. 1), p. 1, (2018).

25- Tatjana Zikov, Stephane Bibian, Guy A Dumont, Mihai Huzmezan, and Craig R Ries, "A wavelet based de-noising technique for ocular artifact correction of the electroencephalogram." in *Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society][Engineering in Medicine and Biology*, (2002), Vol. 1: *IEEE*, pp. 98-105.

26- Anna Zduńska, Joanna Cegielska, and Jan Kochanowski, "Variability of the blink reflex in patients with migraine." *Neurologia i Neurochirurgia Polska,* Vol. 47 (No. 4), pp. 352-56, (2013).

27- Nazareth P Castellanos and Valeri A Makarov, "Recovering EEG brain signals: Artifact suppression with wavelet enhanced independent component analysis." *Journal of neuroscience methods,* Vol. 158 (No. 2), pp. 300-12, (2006).

28- Ian Daly, Reinhold Scherer, Martin Billinger, and Gernot Müller-Putz, "FORCe: Fully online and automated artifact removal for brain-computer interfacing." *IEEE transactions on neural systems and rehabilitation engineering,* Vol. 23 (No. 5), pp. 725-36, (2014).

29- Leonard E Baum, Ted Petrie, George Soules, and Norman Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains." *The annals of mathematical statistics,* Vol. 41 (No. 1), pp. 164-71, (1970).

30- Jeff A Bilmes, "A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models." *International computer science institute,* Vol. 4 (No. 510), p. 126, (1998).

31- Alexandre Bureau, James P Hughes, and Stephen C Shiboski, "An S-Plus implementation of hidden Markov models in continuous time." *Journal of Computational and Graphical Statistics,* Vol. 9 (No. 4), pp. 621-32, (2000).

32- Christopher M Bishop and Nasser M Nasrabadi, Pattern recognition and machine learning. (No. 4). *Springer*, (2006).

33- Asma Bashir, Richard B Lipton, Sait Ashina, and Messoud Ashina, "Migraine and structural changes in the brain: a systematic review and meta-analysis." *Neurology,* Vol. 81 (No. 14), pp. 1260-68, (2013).

34- Bihua Bie *et al.*, "Electroencephalographic signatures of migraine in small prospective and large retrospective cohorts." *Scientific Reports,* Vol. 14 (No. 1), p. 28673, (2024).

35- K Jindal and R Upadhyay, "Epileptic seizure detection from EEG signal using Flexible Analytical Wavelet Transform." in *2017 International Conference on Computer, Communications and Electronics (Comptelix)*, (2017): *IEEE*, pp. 67-72.

36- Khakon Das, Debashis Daschakladar, Partha Pratim Roy, Atri Chatterjee, and Shankar Prasad Saha, "Epileptic seizure prediction by the detection of seizure waveform from the pre-ictal phase of EEG signal." *Biomedical Signal Processing and Control,* Vol. 57p. 101720, (2020).

37- Ming-Lin Li *et al.*, "A state-of-the-art review of functional magnetic resonance imaging technique integrated with advanced statistical modeling and machine learning for primary headache diagnosis." *Frontiers in Human Neuroscience,* Vol. 17p. 1256415, (2023).

38- Michael J Marmura, "Triggers, protectors, and predictors in episodic migraine." *Current pain and headache reports,* Vol. 22 (No. 12), p. 81, (2018).