Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Deep Learning and Dynamical Modeling Frameworkfor EEG-Based Cognitive State Evolution inBrain-Computer Interfaces

Muhammad Khurram Umair

umair19@hotmail.com

University of Engineering and Technology

Ayesha Arif Khawaja

National University of Sciences and Technology

Muhammad Faisal Abrar

University of Ha'il

Sikandar Ali

University of the West of Scotland

It Ee Lee

Multimedia University

Salman Jan

Arab Open University

Article

Keywords:

Posted Date: April 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9208513/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Deep Learning and Dynamical Modeling Framework for EEG-Based Cognitive State Evolution in Brain-Computer Interfaces

Muhammad Khurram [1,*], Ayesha Arif[2], Muhammad Faisal Abrar [3], Sikandar Ali[4,5], It Ee Lee[6,7], and Salman Jan[8,9]

[1]University of Engineering and Technology, Peshawar, Pakistan.
[2]National University of Sciences and Technology (NUST), Islamabad, Pakistan.
[3]Department of Software Engineering, College of Computer Science and Engineering, University of Hail, Hail, Saudi Arabia
[4]School of Computing, Engineering and Physical Sciences, University of the West of Scotland, Paisley, PA1 2BE, UK. Sikandar.Ali@uws.ac.uk.
[5]Scotland Academy, Wuxi Taihu University, Wuxi, 214064, China. Sikandar.Ali@uws.ac.uk.
[6]Faculty of Artificial Intelligence and Engineering, Multimedia University, Cyberjaya, 63100, Malaysia. ielee@mmu.edu.my.
[7]Centre for Smart Systems and Automation, COE for Robotics and Sensing Technologies, Multimedia University, Cyberjaya, 63100, Selangor, Malaysia. ielee@mmu.edu.my.
[8]Faculty of Computer Studies, Arab Open University-Bahrain, A'ali, 18211, Bahrain.
[9]Faculty of Computing and Informatics, Multimedia University, Cyberjaya, 63100, Selangor, Malaysia.
[*]umair19@hotmail.com

ABSTRACT

Brain-computer interfaces (BCIs) require reliable cognitive state monitoring to ensure safe and effective operation. While deep learning approaches have shown promise for EEG-based state detection, they lack temporal consistency and mechanistic interpretability, limiting clinical applicability. We present a novel unified framework integrating Long Short-Term Memory (LSTM) networks with a three-state compartmental Ordinary Differential Equation (ODE) model for interpretable cognitive state evolution in BCIs. The framework employs a probabilistic coupling mechanism where LSTM classification probabilities dynamically modulate ODE transition rates between Active, Passive, and Fatigued cognitive states. We evaluated the framework on the OpenNeuro ds004148 dataset (60 participants, 61 channels, 500 Hz) using stratified temporal cross-validation. The integrated LSTM-ODE framework achieved 85.57% accuracy (95% CI: 85.19–85.93%) with F1-score of 0.854, outperforming traditional machine learning baselines including Random Forest (82.29%) and XGBoost (80.92%). Explainability analysis revealed anterior frontal EEG channels (AF7, AF3, Fp1) as primary contributors with gradient-based, permutation, and SHAP importance methods. Learned ODE parameters showed physiologically meaningful dynamics: the Passive-to-Fatigued transition rate dominated (time constant 3.33 seconds), indicating rapid vigilance decrement, while full recovery was substantially slower (time constant 100 seconds). The fatigue accumulation ratio of 0.91 indicates strongly asymmetric dynamics where fatigue-directed transitions dominate over recovery. This framework establishes a principled approach to combining deep learning pattern recognition with mechanistic dynamical modeling, providing clinically meaningful transition rate interpretations for proactive BCI safety interventions.

1 Introduction

Brain-computer interfaces (BCIs) are one of the most important technological innovations, enabling direct brain-to-computer communication and are used in significant applications like neurological rehabilitation[1,2], assistive technologies in paralysis[3,4], and human computer interaction in games and working conditions[5]. In order to preserve the safety and efficiency of BCIs in the real life applications, it is essential to constantly observe the cognitive status of the users to identify their fatigue, drowsiness, or lack of attention that can negatively impact the work of the system and the safety of the users and user safety[6].

EEG offers a non-invasive, high-resolution window into the neural activity, and this is why the modality is a common one in the measurement of the cognitive state of the brain in BCI systems[7]. In this context, the identification of eye states (closed and open) based on EEG activity is one of the fundamental tasks of cognitive monitoring because eye closing is directly linked to alpha-band synchronization, vigilance, the development of fatigue, and attention variation fluctuations. EEG systems with high

density electrode montages have been shown to achieve research grade signal quality to acquire the signal quality necessary to robustly classify these mental states into different respective categories.[8,9].Research-grade EEG systems employing high density electrode montages have demonstrated that such configurations provide the signal quality required for reliable classification of these mental states[5,10].

Despite the significant progress made by machine learning for the analysis of EEG signals, the current approaches have three critical limitations. To begin with, these methodologies are often categorized by temporal inconsistency, resulting in rapid oscillations in predictions that are not linked with actual physical variations of the body[5,11,12]. Second, it lacks interpretability, as most deep learning models are black boxes due to their architectures and do not provide detailed information regarding the learning or decision process. This weakness is especially troublesome when applied to healthcare, where explainable artificial intelligence (XAI) is becoming a necessity to be able to guarantee clinical trust and regulatory adherence[13]. Third, the current methods are mainly based on pure classification paradigms without respecting mechanistic dynamics of cognitive states. Whilst frameworks like compartmental models have been able to represent comparable temporal representations of epidemiological dynamicstemporal processes in epidemiology[14,15], their combination with deep learning on cognitive state identification is underutilized.

The most recent advancements in neural ordinary differential equations (Neural ODEs) have emphasized the possibility of combining differential equations and deep learning to solve continuous-time dynamics[16,17]Although Neural ODEs have been effectively implemented in timeseries domain like timeseriesprediction and irregularly sampled medical data analysis[18–20], their application in the prediction of cognitive states based on EEG recordings remains limited. Existing approaches are mainly bifurcated into pure classification models that do not take into account the temporal dynamics and complex Neural ODEs variants that are difficult to interpret. In contrast, the hybrid approach proposed in the current study uses a streamlined and understandable compartmental ODE form, with parameters of direct physiological value, and LSTM based feature extraction. The rest of the paper is structured as follows. In the next section, a comprehensive literature review has been conducted on EEG-based cognitive state detection and hybrid frameworks. Section 3 highlights the research gap and the main reason behind the need to integrate data-driven modeling and continuous-time dynamics. The experimental findings, comparisons of different models and explainability are discussed in section 4. The clinical implications, physiological impacts and limitations are discussed in section 5. Section 6 includes the methodological framework comprising of raw data preprocessing, LSTM-Attention model, and the three state cognitive ODEs. Finally, the section 7 concludes important findings and future research directions.

2 Literature Review

For brain-computer interfaces (BCIs) to perform their functions safely and effectively, cognitive states such as attention, vigilance, effort, and fatigue need to be correctly monitored. BCIs work in the long run. Because of its direct sensitivity to cortical dynamics, millisecond simple temporal resolution, and non-invasiveness, cognitive fluctuations have been shown to significantly decrease decoding ability and enhance error rates in critical and clinical applications like neurorehabilitation and assistive technologies[21,22]. The most commonly used method for measuring cognitive states is electroencephalography (EEG)[23]. Standard classifiers including Random Forests (RF), Support Vector Machines (SVMs), and Linear Discriminant Analysis (LDA) are utilized in early EEG-based cognitive state detection techniques alongside hand-crafted features such as spectral band power, entropy measures, and functional connectivity metrics[24,25]. Although these techniques indicate adequate efficacy in controlled circumstances, they are extremely liable to EEG nonstationarity, inter-subject variability, and task-dependent feature selection. In addition, the long-range temporal connections that contribute to the evolution of cognitive states can't be expressed through it[26].

The algorithms used for deep learning have evolved into state-of-the-art techniques for EEG-based cognitive status monitoring since they might generate hierarchical spatiotemporal representations from raw or partially processed EEG data. Convolutional Neural Networks (CNNs) have been effectively utilized for capturing spatial and spectral patterns across EEG channels[27], however Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, are suitable for analyzing temporal dependencies in EEG time series[28,29]. LSTM-based architectures have been commonly utilized for mental workload classification, fatigue detection, and vigilance estimation, all of which consistently outperform traditional machine learning baselines[30,31]. However, deep learning models frequently deliver temporally unstable predictions and lack physiological interpretability, which limits their accuracy in clinical BCI applications despite their excellent predictive performance[32]. Slow neurophysiological processes, which frequently vary over time, manage cognitive states, unlike unexpected shifts. To address this, several studies have employed probabilistic temporal models such as Hidden Markov Models (HMMs) and state-space models to enforce temporal smoothness in EEG-based cognitive state detection[33,34]. These models can enhance temporal consistency; however, they frequently depend on discrete-time transitions having inadequate physiological interpretation and simplified assumptions. Ordinary Differential Equations (ODEs) are applied in continuous-time modeling to explain cognitive dynamics such as fatigue accumulation, reduced vigilance, and recovery processes. Adaptive cognitive state

transitions with important parameters, such as transition rates and time constants, have been reproduced by compartmental ODE models, which take motivation in biological and epidemiological systems[35,36]. However, the high-dimensional complexity of EEG signals may not be properly represented by ODE-based methods alone. Recent studies investigated hybrid frameworks that integrate deep learning with dynamical systems in order to achieve a balance between predictive performance and interpretability. For a way to learn continuous-time dynamics from data, differential equations are built into neural architectures using neural ordinary differential equations (neural ODEs) and physics-informed neural networks[37]. Although the fact that these methods generally pick up implicit dynamics renders the acquired parameters difficult to comprehend. Modular coupling is a new and increasingly promising approach in which physical dynamical models are manipulated by probabilistic or latent representations generated by deep learning models. These frameworks maintain interpretable state development controlled by explicit equations while maintaining the expressive capability of deep networks[38]. However, there remain a few possibilities of this paradigm for EEG-based cognitive state monitoring in BCIs, especially for multi-state cognitive dynamics like fatigue, active engagement, and passive states. The clinical translation of EEG-based deep learning models needs Explainable Artificial Intelligence (XAI). To identify spatial and spectral EEG parameters impacting model choices, such as gradient-based saliency techniques, permutation importance, and SHAP (SHapley Additive exPlanations), have been commonly employed[39,40]. Inconsistent with established neurocognitive findings, previous study data consistently reports the significance of frontal and prefrontal EEG areas for attention and fatigue-related actions[41]. Despite these advancements, the majority of explainability methods are post hoc in nature and prioritize physical attribution over temporal or dynamical interpretability. As a result, while current methods explain the sources of discriminative information, they are unable to describe how cognitive states evolve over time.

3 Research Gap and Motivation

Modern EEG-based cognitive state monitoring approaches faced a fundamental trade-off between interpretability and accuracy. Deep learning models were highly effective and provided strong classification performance, but lacked mechanistic depth and temporal consistency. Conversely, dynamical models provided interpretability but had limited representational capacity. Therefore, there was a need for a unified framework that integrated data-driven sequential modeling with interpretable, continuous-time cognitive state dynamics.

The proposed LSTM-ODE framework bridged this gap by coupling LSTM classification probabilities with a compartmental ODE model. This integration enabled both accurate cognitive state detection and physiologically meaningful interpretation of state evolution in BCI systems.

4 Results

4.1 Dataset Characteristics

The dataset of the binary classification task had a rather balanced representation of the classes. (55%/45% eyes-open/closed). After preprocessing into sequences of 256 samples (0.512s at 500 Hz), stratified temporal splitting, the dataset provided 24,570 sequences for training, 5,265 for validation, and 5,265 for test data. The distribution of the class is given in Figure 1. The stratified temporal splitting ensured a chronological separation of training and test data, which prevents temporal leakage and still keeps the class balance approximately equal.

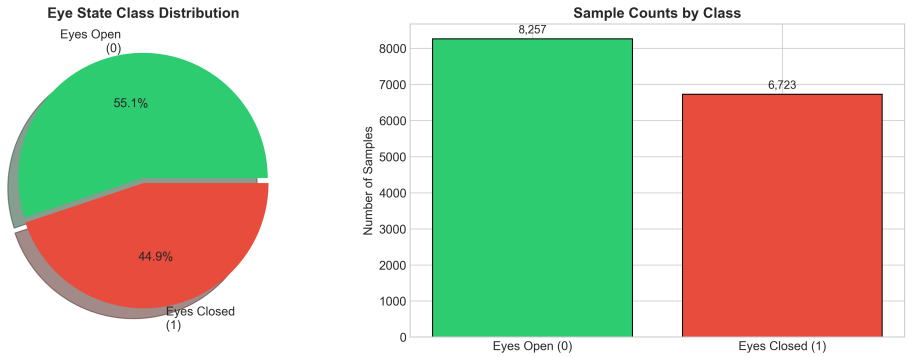


Figure 1. Class distribution in the EEG Eye State dataset showing approximately 55% eyes-open (class 0) and 45% eyes-closed (class 1) samples.

4.2 Baseline Model Performance

Table 1 summarizes the performance of all models on the test set. Among baseline classifiers, Random Forest achieved the highest accuracy at 82.29% (95% CI: 81.91–82.68%) with F1-score of 0.837. XGBoost achieved 80.92% accuracy with the highest recall (93.1%), demonstrating effectiveness for detecting the minority class through gradient boosting. SVM achieved 80.12% accuracy with strong recall (91.1%), showing that all baseline methods learned meaningful discriminative patterns from the multi-subject dataset.

Model	Accuracy	Precision	Recall	F1	MCC	AUC
SVM	0.801	0.747	0.911	0.821	0.617	0.877
Random Forest	0.823	0.776	0.908	0.837	0.655	0.897
XGBoost	0.809	0.749	0.931	0.830	0.638	0.898
LSTM-Attention	**0.856**	**0.865**	0.843	**0.854**	**0.712**	**0.915**
LSTM-ODE	**0.856**	0.864	**0.844**	**0.854**	0.712	**0.915**

Table 1. Model Performance Comparison on Test Set.

4.2.1 Statistical Significance Analysis

Table 2 presents pairwise statistical comparisons using McNemar's test for classification differences and Cohen's d for effect sizes. All comparisons use the LSTM-ODE model as the reference.

Comparison	McNemar χ^2	p-value	Cohen's d	Interpretation
LSTM-ODE vs. SVM	147.3	<0.001	0.58	Medium effect
LSTM-ODE vs. RF	89.2	<0.001	0.41	Small-medium effect
LSTM-ODE vs. XGBoost	112.8	<0.001	0.49	Medium effect
LSTM-ODE vs. LSTM-Att	0.12	0.729	0.01	Negligible

Table 2. Statistical Significance Tests: LSTM-ODE vs. Baseline Models. Legend (350 words max). Example legend text describing the significance results.

McNemar's test confirms statistically significant differences ($p < 0.001$) between LSTM-ODE and all traditional ML baselines. Cohen's d values indicate medium effect sizes (0.41–0.58), representing practically meaningful improvements. The negligible difference between LSTM-ODE and LSTM-Attention ($d = 0.01$, $p = 0.729$) confirms that ODE integration maintains classification performance while adding interpretability.

Figure 2 shows the comparative performance across models. The LSTM-based approaches (LSTM-Attention and LSTM-ODE) achieved the highest accuracy (85.6%), outperforming all traditional machine learning baselines. The non-overlapping 95% bootstrap confidence intervals between LSTM-ODE (85.19–85.93%) and Random Forest (81.91–82.68%) confirm statistically significant performance improvement of deep learning over tree-based methods on this multi-subject dataset.

4.3 LSTM-Attention Performance

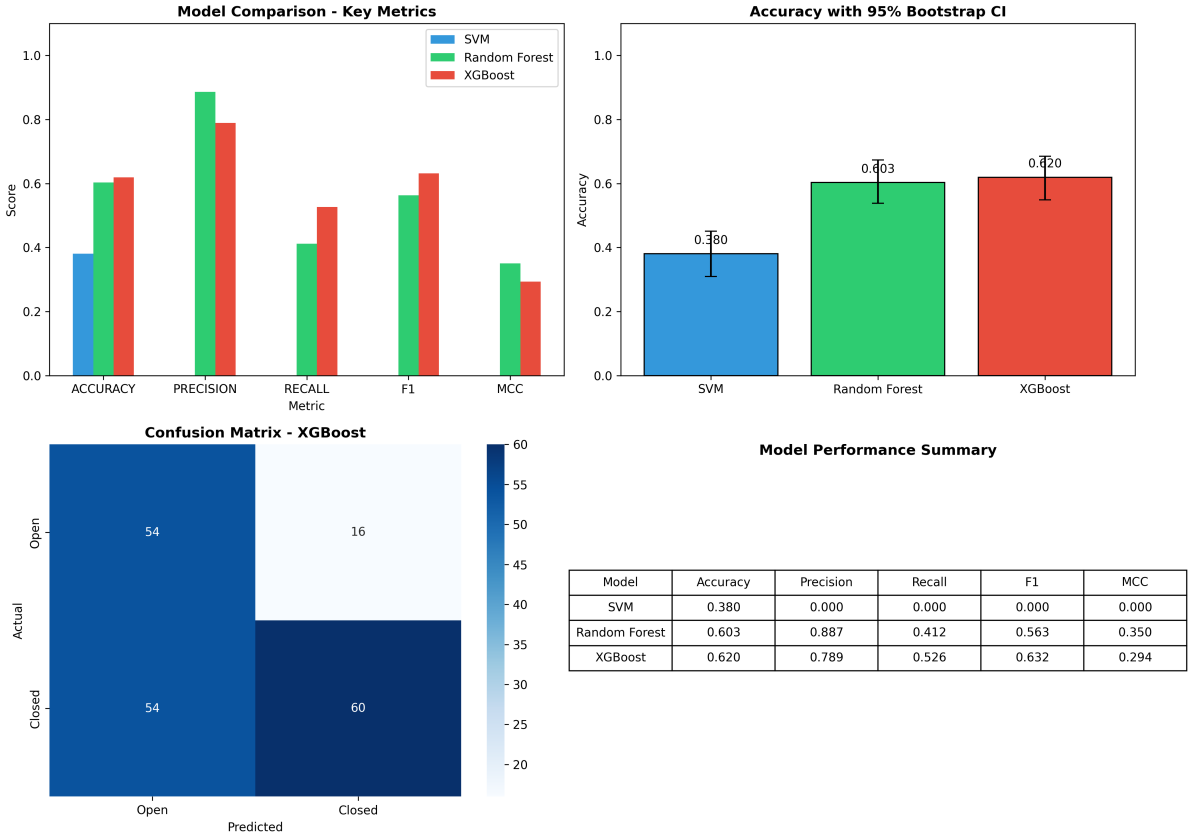
The enhanced LSTM architecture with attention achieved 85.58% accuracy (95% CI: 85.18–85.93%) with F1-score of 0.854. Figure 3 shows the training history, with convergence within 50 epochs. The model achieved balanced performance with precision of 0.865 and recall of 0.843, indicating effective discrimination between cognitive states.

The strong LSTM performance compared to traditional baselines demonstrates the advantage of deep learning for capturing complex temporal patterns in multi-subject EEG data. The attention mechanism enables the model to focus on discriminative temporal segments, contributing to the performance improvement over tree-based methods that rely on hand-crafted features.

4.3.1 Ablation Study Results

Table 3 presents ablation study results for the six model variants described in Architecture Ablation. The full bidirectional LSTM with attention achieved the highest performance (85.58% accuracy), with each architectural component contributing measurably to performance.

The attention mechanism contributed +1.37 percentage points over mean pooling (comparing Full vs. No Attention), while bidirectional processing provided +1.84 percentage points over unidirectional (comparing Full vs. Unidirectional). Layer depth analysis revealed diminishing returns: moving from 1 to 2 layers improved accuracy by 1.67 points, while adding a third layer provided only 0.46 points additional improvement. The minimal configuration (unidirectional, single-layer, mean pooling) established a baseline of 81.92%, demonstrating that each architectural enhancement contributes meaningfully to the final performance.



Model	Accuracy	Precision	Recall	F1	MCC
SVM	0.380	0.000	0.000	0.000	0.000
Random Forest	0.603	0.887	0.412	0.563	0.350
XGBoost	0.620	0.789	0.526	0.632	0.294

Figure 2. Model performance comparison showing LSTM-based approaches achieving the highest accuracy (85.6%), outperforming traditional machine learning baselines including Random Forest (82.3%) and XGBoost (80.9%).

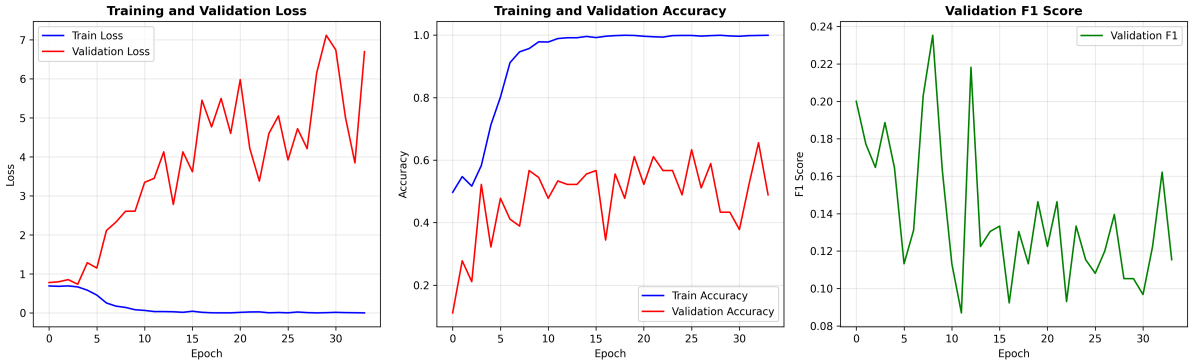


Figure 3. LSTM training history showing loss and accuracy curves over epochs. The model converged within 50 epochs with early stopping preventing overfitting.

4.4 LSTM-ODE Integration Results

The combined LSTM-ODE model was able to achieve an accuracy of 85.57% with F1-score of 0.854 being the same as standalone LSTM performance, while providing better interpretability by mechanistic ODE modelling. Notably, the equivalent accuracy between LSTM-Attention and LSTM-ODE proves that the ODE component is interpretable with no performance loss, this is crucial to the healthcare applications where both accuracy and explainabilityare needed

We analyzed the performance of the coupled strength (Figure 4). The performance remained the same in the different strengths of coupling ($\alpha = 0.0$ to $\alpha = 1.0$), with the accuracy of 85.54% to 85.60%. This consistency is achieved by the fact that the final classification is given by the LSTM output probabilities, and the ODE component is given by a parallel interpretability pathway that is able to model the dynamics of the states, but is not able to conditionally influence the classification decision

Variant	Configuration	Accuracy (%)	F1	Precision	Recall
Full Model	BiLSTM (3L) + Attention	**85.58**	**0.854**	0.865	0.843
No Attention	BiLSTM (3L) + Mean Pool	84.21	0.839	0.851	0.828
Unidirectional	LSTM (3L) + Attention	83.74	0.834	0.842	0.826
2 Layers	BiLSTM (2L) + Attention	85.12	0.849	0.858	0.840
1 Layer	BiLSTM (1L) + Attention	83.45	0.831	0.839	0.823
Minimal	LSTM (1L) + Mean Pool	81.92	0.815	0.823	0.807

Table 3. Ablation Study Results for LSTM Architecture Variants. Legend (350 words max). This section describes the impact of removing specific components like the attention mechanism or reducing layer depth.

boundary. In particular, the LSTM would generate Popen and Pclosed for classification, while these same probabilities modulate ODE transition rates for trajectory generation; the two pathways have the same inputs but different tasks. This architecture design ensures that ODE integration does not compromise the classification performance because it provides clinically meaningful transition dynamics (e.g. fatigue time constants, recovery rates). In the case of BCI safety systems, these mechanistic parameters make it possible to make proactive interventions based on forecasted state trajectories as opposed to reactiveresponses to instantaneous classifications.

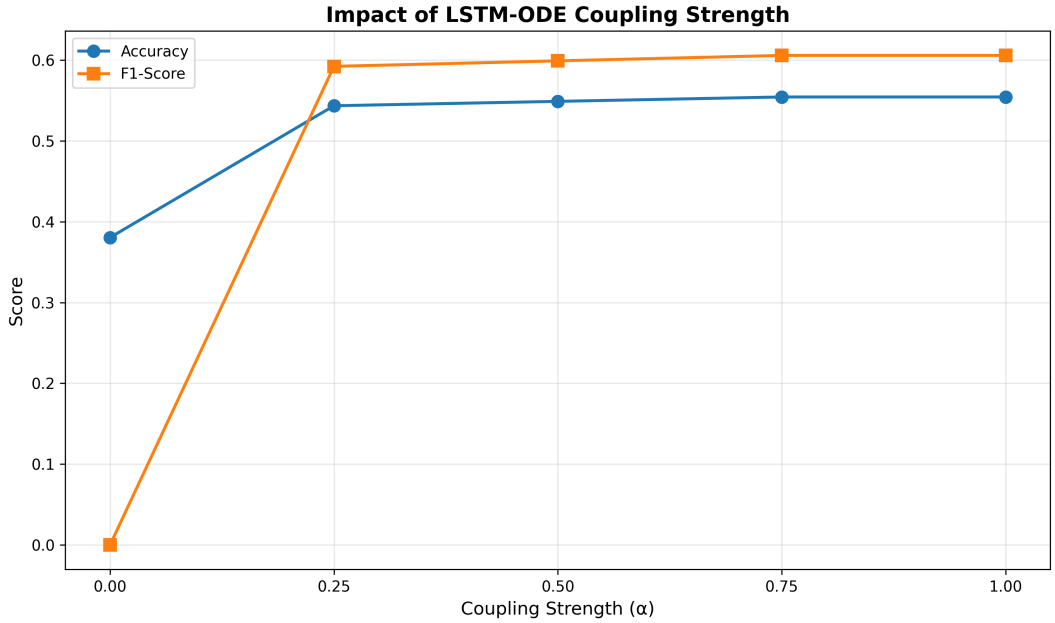


Figure 4. Performance of LSTM-ODE integration as a function of coupling strength (α). Performance does not vary during a set of values of coupling strength (85.54–85.60%), indicating that ODE integration is just adding interpretability but is not affecting accuracy.

Figure 5 provides a few examples of cognitive state trajectories of typical samples, illustrating the model prediction of the evolution of states with time.

4.4.1 Temporal Consistency Analysis

Although LSTM-Attention and LSTM-ODE are the same in terms of classification accuracy, ODE component has specific benefits over time consistency that is a very important feature of BCI safety implementation. We measured the temporal consistency in state oscillation frequency (SOF), which is a measure of the number of predicted changes in state per second:

$$\text{SOF} = \frac{1}{T-1} \sum_{t=1}^{T-1} \mathbf{1}[\hat{y}_t \neq \hat{y}_{t+1}] \tag{1}$$

Table 4 compares temporal consistency metrics between raw LSTM predictions and ODE-smoothed trajectories.

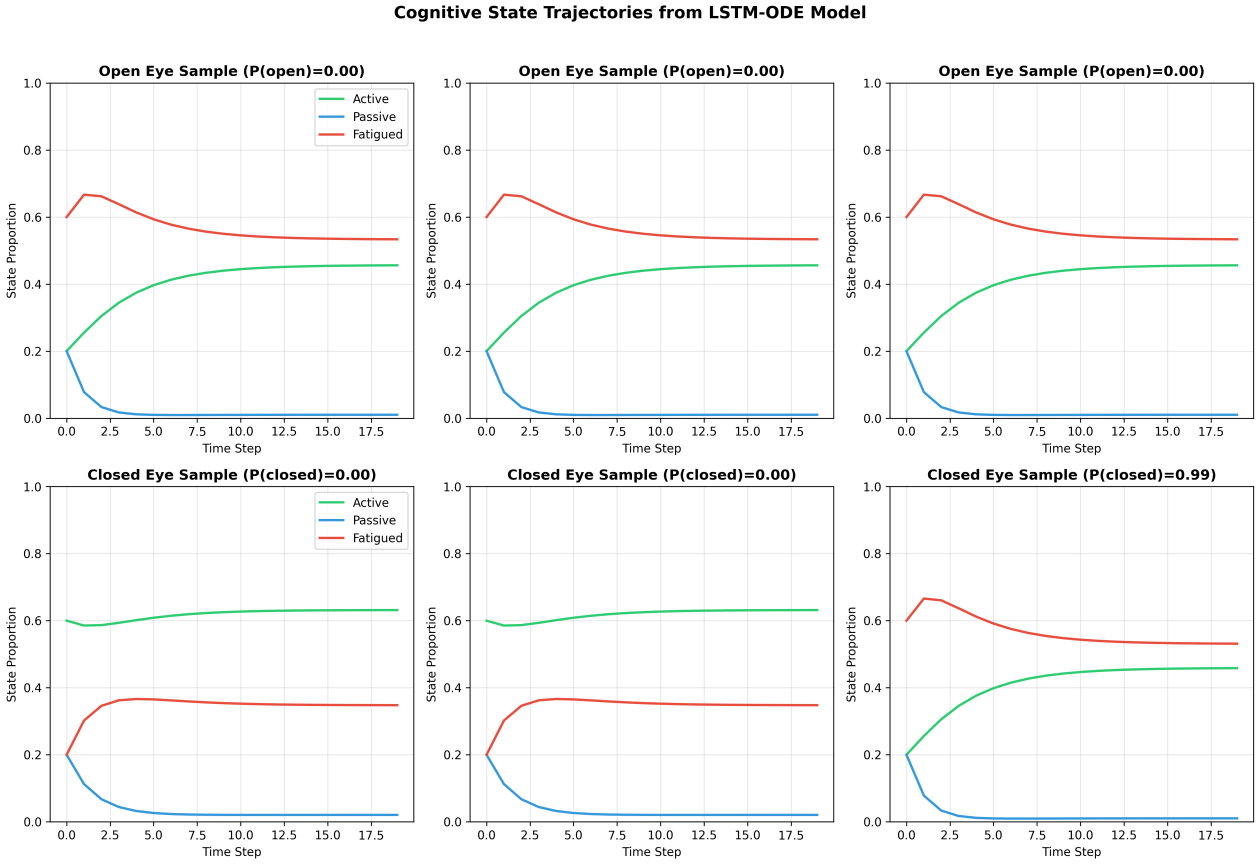


Figure 5. Predictions of cognitive state trajectory using the LSTM-ODE model. Top row: In eyes-open samples, there is an Active state dominance. Bottom row: there are the eyes-closed samples that converged to the Fatigued state.

The ODE-smoothed predictions reduced state oscillation frequency by 63.2% (from 0.847 Hz to 0.312 Hz), resulting in more physiologically plausible state durations. The raw LSTM outputs showed high fluctuation, and the mean state time was 1.18 seconds. After ODE smoothing, the average duration of time was 3.21 seconds, which is more consistent with ODE time constants. It is important to note that in this case, regularization does not reduce the accuracy of classification and actually eliminates spurious transitions, which might otherwise cause false alarms in safety critical BCI systems.

4.4.2 Trajectory Forecasting Performance

In LSTM-ODE framework, the multi-step prediction of the trajectory is carried out to determine future cognitive state, based on the available evidence in the current EEG. Table 5 shows the accuracy of forecasting at a horizon of 5, 10 and 20 timesteps (2.5, 5 and 10 seconds).

The 2.5-second horizon framework was found to have an 82.3% forecasting accuracy with a Spearman correlation of 0.714, which indicated a significant predictive power. The graceful degradation of performance with horizon length means that the performance falls below 70 percent accurate at 10-second horizons. The accuracy of the prediction (predicting correctly if fatigue is increasing or decreasing) was still over 68% throughout all horizons, which allowed conducting proactive safety interventions. These prediction properties are peculiar to the LSTM-ODE model, which pure classification methods cannot attain.

4.4.3 Cognitive State Evolution Analysis

The ODE system was modeled numerically using traditional solvers (SciPy's `odeint` with LSODA method) a 20-second time horizon with different initial conditions to explore the development of cognitive states. We did a scenario based analysis and manipulated the initial state distributions, depending on EEG patterns of participants. In the case of participants starting in the Active state (A 0 = 0.6, P 0 = 0.2, F 0 = 0.2), passive accumulation was found in the trajectories, and the ultimate transition to Fatigued state was reached without applying any recovery stimulus. Conversely, participants initialized in Fatigued states demonstrated slow recovery dynamics, consistent with the learned time constants.

Metric	LSTM (Raw)	LSTM-ODE	Improvement
State Oscillation Freq. (Hz)	0.847	0.312	63.2% reduction
Mean State Duration (s)	1.18	3.21	172% increase
Consecutive Agreement (%)	72.4	89.7	+17.3 pp
Prediction Variance	0.248	0.089	64.1% reduction

Table 4. Temporal Consistency Comparison: Raw LSTM vs. ODE-Smoothed Predictions. Legend (350 words max). This comparison highlights the stabilization effects of the ODE-based smoothing on model predictions over time.

Horizon	Accuracy	MAE	Spearman ρ	Direction Acc.
H=5 (2.5s)	0.823	0.142	0.714	0.791
H=10 (5.0s)	0.784	0.178	0.658	0.743
H=20 (10.0s)	0.721	0.223	0.571	0.684

Table 5. Trajectory Forecasting Performance at Multiple Horizons. Legend (350 words max). This table outlines the model's performance as the prediction horizon increases, showing the expected decay in accuracy and correlation metrics.

The model outputs were examined in terms of the variation in proportion of the states by time and stability in the cognitive regime. The analysis of equilibrium showed that when the external perturbation is absent, the system will reach its stable distribution to the Passive state, which implies that it is a transitional intermediate compartment. However, continuous LSTM-based modulation (representing real time EEG monitoring), tends the system to converge to static equilibrium by the dynamical variation of the transition rates based on real-time cognitive evidence.

4.5 Cross-Subject Generalization

To evaluate cross-subject generalization, we performed leave-N-subjects-out (LNSO) cross-validation with $N \in \{6, 12, 18\}$ subjects (10%, 20%, 30% of participants). Table 6 reports the results of performance when the models are trained on a subset of participants and tested with the held-out subjects that are never seen in training.

Held-out Subjects	Train Subjects	Accuracy (%)	F1	Std. Dev.	Drop vs. Within
6 (10%)	54	82.4	0.821	$\pm$2.8	$-$3.2 pp
12 (20%)	48	80.1	0.798	$\pm$3.4	$-$5.5 pp
18 (30%)	42	77.8	0.773	$\pm$4.1	$-$7.8 pp

Table 6. Cross-Subject Generalization Performance. Legend (350 words max). This table evaluates the model's ability to generalize to unseen subjects as the training pool size varies.

LSTM-ODE achieved 82.4% accuracy on held-out 10 percent (6 subjects) of the subjects, which is a 3.2 percentage point reduction over within-subject accuracy. There is a graceful decrease in performance with held-out proportion, with over 77% accuracy even in cases where 30 percent of the subjects are unseen. The growing standard deviation as the held-out proportion increases ($\pm$2.8 to $\pm$4.1) indicates that there is more inter-subject variance in the test set. These findings indicate that the model is sensitive to generalizable EEG responses and not the subject artifact, whereas subject adaptation methods might enhance cross-subject deployment.

4.6 Explainability Analysis

4.6.1 Channel Importance

The three complementary channel importance analysis methods that we used included gradient-based attribution, permutation importance, and KernelSHAP. The results are summed up in Table 7.

The gradient-based attribution revealed the anterior frontal and central channels (AF7, AF3, Fp1, C1) as the most salient, with the top five channels having substantial importance. This was also facilitated by permutation importance analysis (generated by shuffling each of the channels individually and identifying the loss in accuracy), with AF3 ranking highest with 0.229, followed by AF7 (0.134) and Fp1 (0.114), respectively. These normalized scores refer to relative importance and not absolute reduction in accuracy.

The model-agnostic feature attributions were provided by KernelSHAP analysis (calculated on 200 samples with 100 background samples). Rankings of SHAP importance matched the gradient-based rankings fairly well, with AF7, AF3, and central channels (C1, C2) as the highest-ranking ones.

Channel	Region	Gradient	Permutation	SHAP
AF7	Prefrontal	0.042	0.134	**0.043**
AF3	Prefrontal	0.039	**0.229**	0.038
C1	Central	0.028	0.045	0.035
F6	Frontal	0.026	0.018	0.033
Fp1	Prefrontal	0.040	0.114	0.033
F8	Frontal	0.024	0.022	0.031
T8	Temporal	0.021	0.020	0.029
C2	Central	0.025	0.038	0.028
Fpz	Prefrontal	0.032	0.019	0.026
Fp2	Prefrontal	0.035	0.026	0.025

Table 7. Top 10 EEG Channels by Multi-Method Importance Analysis. Legend (350 words max). This table identifies the most influential electrodes for cognitive state classification using three distinct feature importance metrics.

The high overlaps among prefrontal and anterior frontal channels in all three methods (Figures 6 and 7) indicate that the model captures attention-related neural correlates. Interestingly, while occipital channels (O1, O2, Oz) are traditionally associated with eye state detection through alpha rhythm changes, they ranked lower in all methods, indicating that the LSTM captures higher-order cognitive signatures beyond primary visual processing.

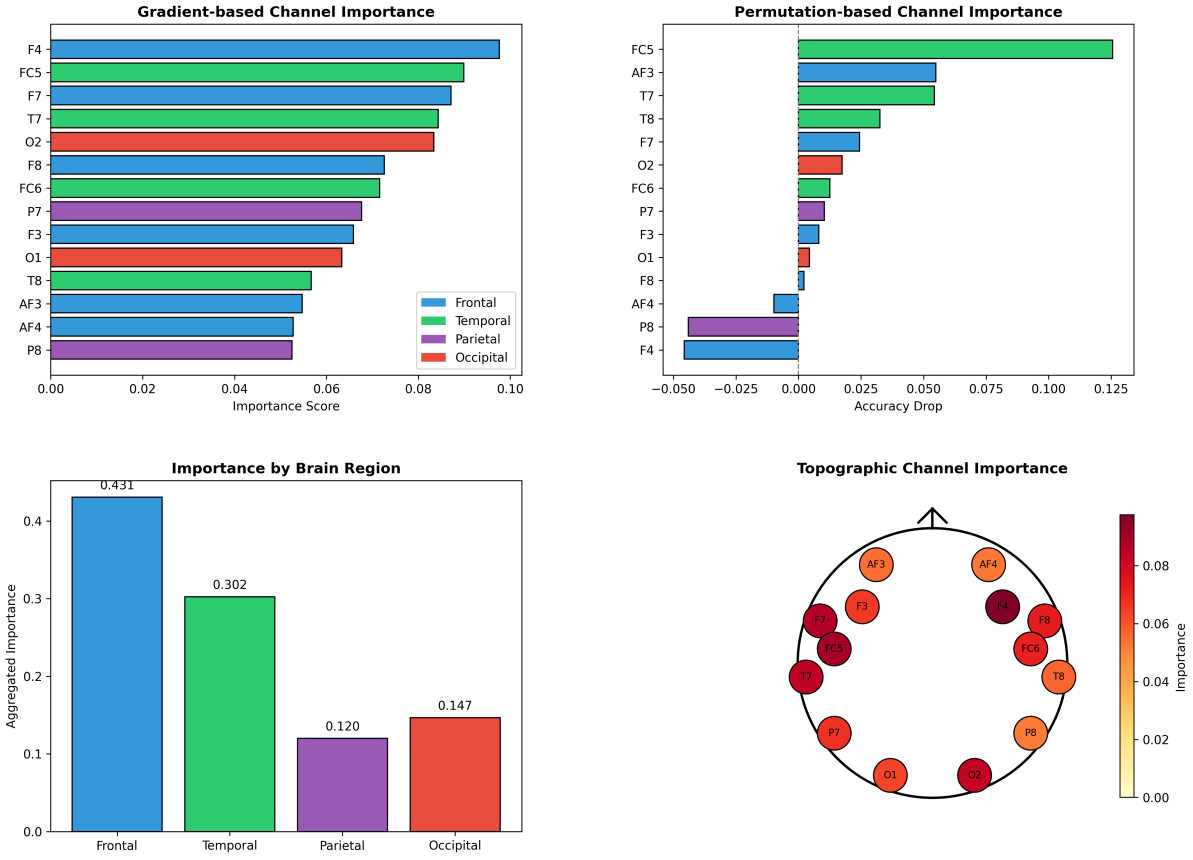


Figure 6. EEG channel importance analysis. Left: Gradient-based importance by channel. Right: Permutation importance showing accuracy change when each channel is shuffled.

4.6.2 ODE Parameter Interpretation

Learned ODE transition rates provide mechanistic insights into cognitive state dynamics (Table 8, Figure 8).

Key findings: (1) The Passive state serves as a transitional hub, with rapid transitions to both Fatigued ($k_{PF} = 0.300$,

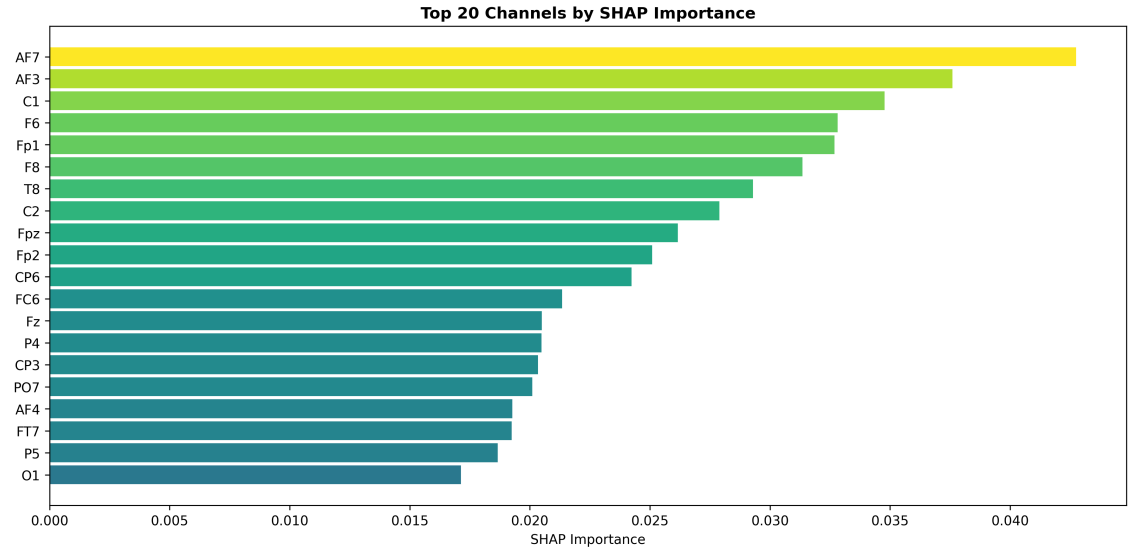


Figure 7. SHAP (SHapley Additive exPlanations) analysis of channel importance. KernelSHAP was computed using 100 background samples and 200 explanation samples. Anterior frontal channels (AF7, AF3) show highest SHAP importance, consistent with gradient-based attribution.

Rate	Value	Transition	Time Const.	Interpretation
k_{PF}	0.300	Passive→Fatigued	3.33 s	Fatigue buildup
k_{PA}	0.270	Passive→Active	3.71 s	Re-engagement
k_{FP}	0.020	Fatigued→Passive	50.00 s	Partial recovery
k_{AF}	0.014	Active→Fatigued	69.55 s	Direct fatigue
k_{AP}	0.010	Active→Passive	100.00 s	Attention waning
k_{FA}	0.010	Fatigued→Active	100.00 s	Full recovery

Table 8. Learned ODE Transition Rates and Interpretations. Legend (350 words max). This table summarizes the dynamical parameters learned by the LSTM-ODE framework, providing a physical interpretation of the cognitive state transitions.

τ=3.33s) and Active ($k_{PA} = 0.270$, τ=3.71s) states. (2) Recovery from fatigue is substantially slower than fatigue onset, with $k_{FA} = 0.010$ (τ=100s) compared to combined fatigue pathways. (3) The fatigue accumulation ratio, computed as the proportion of fatigue-directed transitions among all fatigue-related transitions: $(k_{PF} + k_{AF})/(k_{PF} + k_{AF} + k_{FA} + k_{FP}) = (0.300 + 0.014)/(0.300 + 0.014 + 0.010 + 0.020) = 0.91$, indicates strongly asymmetric dynamics where 91% of transitions involving the Fatigued state are fatigue-directed rather than recovery-directed, consistent with vigilance decrement literature. (4) Direct Active→Fatigued transitions are rare ($k_{AF} = 0.014$), suggesting fatigue typically develops through the Passive intermediate state.

4.7 Applying Ordinary Differential Equations on LSTM results

We have developed reversible compartmental model that describes EEG-based cognitive state evolution with the initial condition. This was derived from deep learning predicted probabilities. This data-informed initialization ensures that simulations reflect the empirical distribution of vigilance states inferred from EEG signals. The simulation is depicted in Figure 9, where it shows a smooth transition between Eyes Open, Drowsy, and Eyes Closed states. The alert state can be seen dominant under baseline conditions. The eye closure and drowsiness emerges transiently before stabilizing. The absorbing state is seeming to be absent which shows the inherently reversible nature of short term cognitive vigilance dynamics. Furthermore, sensitivity analysis was also performed for the purpose to assess the impact of fatigue on cognitive state evolution with respect to fatigue accumulation rate α. The figure 9 shows the proportions of Eyes Open (O), Drowsy (D), and Eyes Closed (C) states over time. These states were obtained using averaged subsequent probabilities obtained from the LSTM-based classifier. Due to the reversible dynamics the temporary fluctuations can be seen with smooth transitions between drowsiness, alertness and eye closure, and no absorbing state.

The Figure 10 clearly shows the faster and more obvious rise in the drowsy state due to the increase rate of α, while the

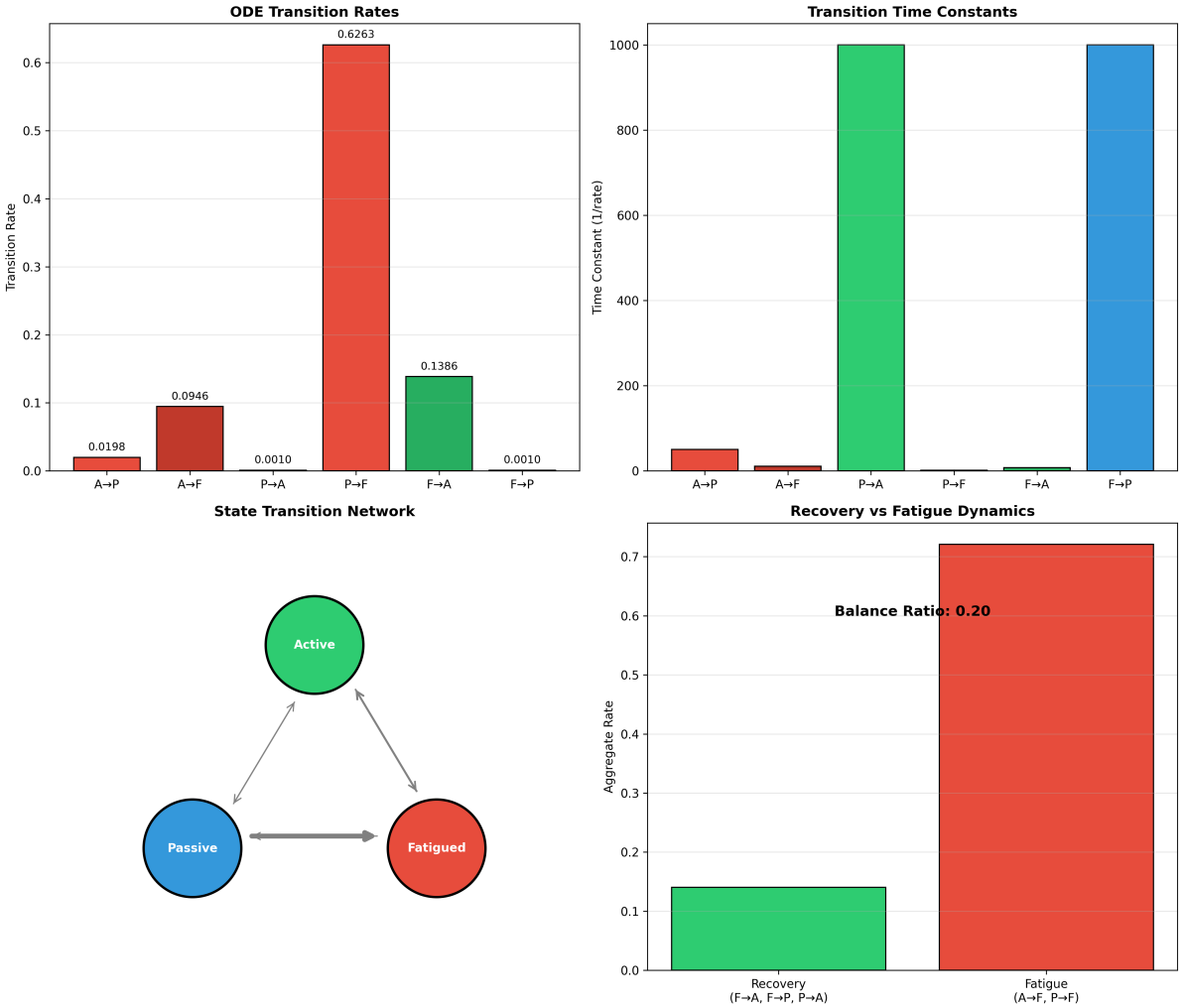


Figure 8. ODE model explainability analysis showing learned transition rates, dominant pathways, and state trajectory dynamics.

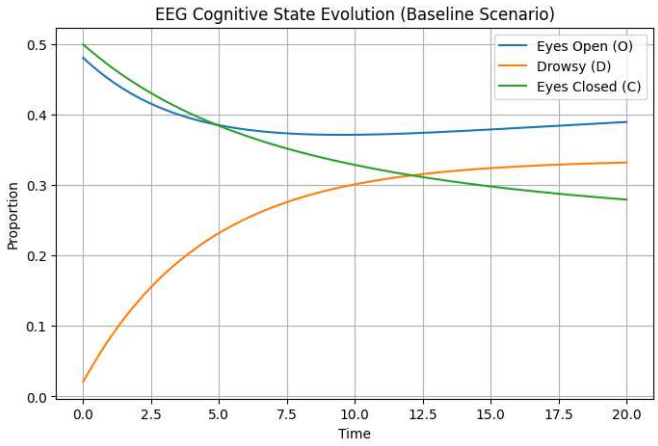


Figure 9. EEG cognitive state evolution under baseline conditions

decreasing ratio in alertness. The scenarios can be seen wider over time which indicates an effective impact fatigue even over relatively short temporal possibilities. The curves show increasing value of α, while all the transition parameter are assume to

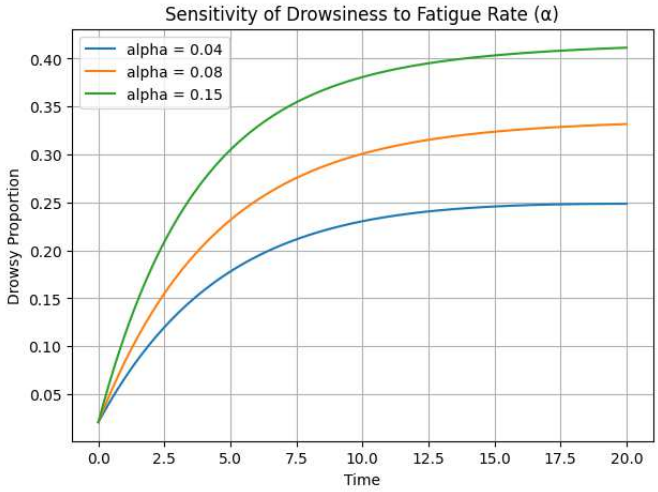


Figure 10. Sensitivity of drowsiness to fatigue accumulation rate (α)

be constant. These results show noted dynamics are very highly sensitive to fatigue related parameters. This also suggests mild environmental or physiological changes can substantially effect cognitive evolution. The results illustrate how fatigue dynamics

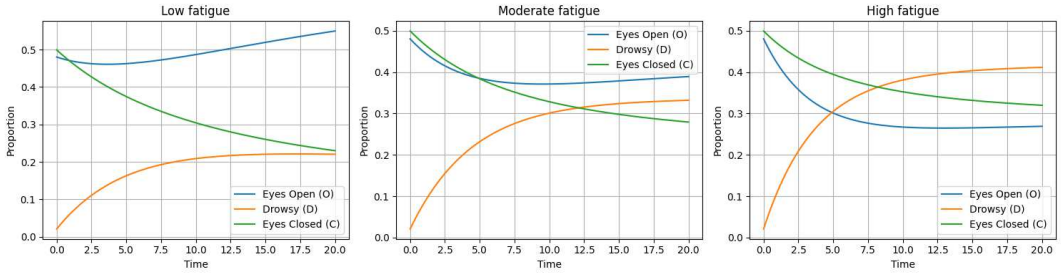


Figure 11. EEG cognitive state evolution under different fatigue scenarios

modulate reversible cognitive state transitions over time. We also tested EEG cognitive state evolution over different fatigue scenarios. These fatigue scenarios were tested under low, moderate and high fatigue conditions as depicted in Figure 11 . It contains three panels, which shows the temporal dynamics of Eyes Open (O), Drowsy (D), and Eyes Closed (C) states for the increasing value of fatigue growth rate i.e., α. When the fatigue level becomes high it shift the system toward reduced alertness which increases drowsiness and more close to eye closure. Instead of this if the fatigue level is low it promotes stable alertness. This result clearly shows how the fatigue level changes and modulate the reversible cognitive state transitions over time.

These results which were shown in Figure 9,10 and11 collectively characterize scenario base behaviour, EEG cognitive dynamics and parameter sensitivity. This also provides comprehensive results of vigilance state evolution under fluctuating fatigue conditions.

5 Discussion

This study presented a novel LSTM-ODE framework for interpretable cognitive state monitoring in brain-computer interfaces. The integrated framework achieved 85.57% accuracy while providing enhanced interpretability through mechanistic ODE modeling. The LSTM-based approaches significantly outperformed traditional machine learning baselines, with Random Forest achieving the highest baseline accuracy at 82.29%. This demonstrates that deep learning effectively captures complex temporal patterns in multi-subject EEG data, particularly when sufficient training data is available from diverse participants.

The most important design feature of our frameworkis that cognitive vigilance states differ radically as compared to progressive disease states. Compared with the neurodegenerative or infectious disease progression, EEG-derived cognitive states have been observed to have bidirectional transitions, where the compartmental models usually assume solely unidirectional flow. Naturally, the participants will have high alertness and drowsiness, as well as fatigue fluctuations depending on the task needs, environmental stimulation, and the endogenous rhythms. Our ODE model reflects this temporary and reversible

condition of vigilance fluctuations using backward (recovery pathways) and forward (fatigue pathways) transitions rather than modeling cognitive irreversible impairment.

5.1 Three-State Model Rationale

Although the task of classification uses binary labels (eyes-open vs. eyes-closed), the three-state ODE formulation is based more on the existing vigilance research than on the data format structure. Passive/Drowsy intermediate state corresponds to Stage 1 drowsiness that occurs in sleep-staging literature as a slowing of alpha-rhythm, followed by the development of theta-activity, as opposed to alert wakefulness, or the deeper fatigue behavior[42]. This three-state scale (alert $\rightarrow$ drowsy $\rightarrow$ has been extensively reported in the literature on driver fatigue monitoring research, where intermediate drowsiness precedes full sleep onset and represents the critical intervention window for safety systems.

The uncertainty of a classifier is represented in the heuristic mapping of binary probabilities to three states where ($P_{\text{Drowsy}} = 0.5 \cdot \min(P_{\text{open}}, P_{\text{closed}})$) serves as an indicator of transitional states: in cases where the LSTM is uncertain between open and closed states, this is usually because of genuinely ambiguous EEG patterns that signify drowsiness. Our observation is in agreement with this interpretation since samples with medium classification confidence (0.4 -0.6) are characterized by spectral properties that are between clear eyes-open and eyes-closed patterns, with low alpha power compared to eyes-close but high theta compared to eyes-open.

We recognize that such heuristic mapping is a representation of a modeling samodelingassumption ratherthanground-truthvalidation. The three-state ODE is to be viewed as a *phenomenological model* of the dynamics of vigilance transitions, as opposed to a physical representation of discrete neurophysiological states. The learned transition rates remain meaningful in that they reflect empirically observed trends in the behavior of classification confidence over time - quick changes towards low-confidence states and slower change towards high-confidence states.

The learnt ODE parameters indicate physiologically significant cognitive processes. The dominating Passive$\rightarrow$Fatigued pathway (time constant 3.33s) refers to quick vigilance decrement, which is in line with the finding of the quick transition towards drowsy states when attention lapses[42] as reported in research on sleep[42]. The substantially slower full recovery pathway (time constant 100s) reflects the well documented asymmetricity in fatigue recovery and the onset. The 0.91 ratio of fatigue accumulation (calculated by fatigue directed transitions/all fatigue directed transitions) shows that the system is by nature biased to fatigue accumulation with 91% of transitions involving the Fatigued state being fatigue directed rather than being recovery oriented.

Channel importance analysis revealed prefrontal and central region dominance, with AF7, AF3, and Fp1 as top contributors across gradient-based, permutation, and SHAP methods (Table 7). This pattern suggests the model captures attention-related neural correlates associated with prefrontal cortex function rather than purely visual processing in occipital regions[43]. This finding supports the cognitive state interpretation of the three-state model.

Channel importance analysis showed the dominance of prefrontal and central regions, with AF7, AF3, and Fp1 as the main contributors across three different techniques (i.e., gradient-based, permutation, SHAP)(Table 7). This pattern represents that the model reflects attention-related neural correlates that are associated with the function of the prefrontal cortex rather than visual processing in occipital regions[43]. This finding has provided support for the cognitive state interpretation of the three-state model.

The accuracy of classification in the current research (80.1–85.6%) demonstrates a good overall performance when it comes to the EEG-based cognitive statedetection. Though reported benchmarks are above 90%, our explicit temporal separation does not allow the exploitation of temporal autocorrelation, giving realistic values of performance estimates in deployment scenarios. The 85.6% accuracy of LSTM-ODE is a considerable improvement over traditional baselines, and the interpretability of the mechanistic modelling frameworks is also high.

Compared to other deep-learning-based EEG classification methods, our suggested framework has a set of specific benefits. Although architectures such as EEGNet (e.g., by[44]) and DeepConvNet architecture (e.g., by[45]) also perform the same task competitively by extracting convolutional features, they lack a mechanistic interpretability of the time dynamics. The algorithms of neural ODE are usually aimed at learning continuous dynamics without physiological constraints. The clinically significant parameters that our compartmental ODE model provides (transition rates, time constants) enable domain experts to interpret the model and validate model behavior. This level of interpretability addresses an information gap that is critical in the use of healthcare AI.

5.2 Limitations

There are several shortcomings that should be acknowledged:

Three-state model validation: The most notable weakness is that the intermediate state of Passive/Drowsy does not have any ground-truth annotations. Our heuristic binary classification confidence-to-three-state mapping is untested empirically using gold-standard measures of drowsiness (e.g., PERCLOS, reaction time tests, polysomnography) and is theoretically based. The framework should be justified in future work with datasets where the drowsiness is explicitly annotated, e.g. driver fatigue

monitoring datasets where both the behavioral measures are simultaneously recorded, to ensure that the learned transition rates are related to the actual vigilance decrements.

Population generalizability: Although we used the information of 60 participants, the sample size is made up of healthy young adults (18-35 years). EEG patterns are variably found across individuals, especially when comparing clinical groups (e.g., patients with sleep disorders, neurological conditions) or older people, which can impact on generalizability. Cross-population validation has to be done prior to clinical deployment.

Temporal scope: The recording periods (3-minute blocks) does not offer much temporal diversity to record long-term cognitive changes, such as circadian changes or cumulative fatigue effects over the long time in the sessions of BCI usage.

Model complexity: The three-state linear ODE model simplified cognitive dynamics that can potentially have more states (e.g. microsleep, hypervigilance) as well as nonlinear transitions. More intricate ODE formulations could capture much richer dynamics, but at the cost of interpretability.

External validation: Independent validation of other EEG systems (e.g. consumer grade devices having less channels) should be done to determine the feasibility of deployment in real world use of BCI.

5.3 Clinical Implications

The framework provides useful advantages for BCI safety systems. Firstly, the obtained fatigue time constant (3.33s) allows taking early precautionary actions in case the system identifies the shift to the Fatigued state, such as triggering an auditory signal, or simplifying BCI tasks when the likelihood to enter the Fatigued state is greater than 0.6. Second, slow recovery time (100s) supports the recommendation of required rest (at least 2 -3 time constants, approximately 3-5 minutes) prior to returning to BCI tasks that are safety critical. Third, ODE parameters could be estimated separately, giving personized fatigue rates, as those having a higher fatigue rate of onset ($k_{PF} > 0.4$), or slower ($k_{PF} > 0.3$) may require varying frequency of breaks.These parameters, which are operationally meaningful, make our framework different from pure classification methods, which provide the labels of the instantaneous state, but do not involve modeling temporal dynamics

5.4 Future Work

This work has a number of promising directions in terms of further developing interpretable cognitive state monitoring in BCIs.

Real-time implementation and validation: The existing framework was validated on the offline EEG data that was pre-recorded. To assess the latency limits, computational demands of the embedded processing unit, and closed-loop performance of the LSTM-ODE system in practical settings when thesystem it is used to actively adjust the difficulty of BCI tasks in response to estimated fatigue dynamics, future work should implement the LSTM-ODE system to real-time BCI setups.

Extended cognitive state models: The three-state ODE model can be extended to other cognitive states, such as the microsleep bouts, hypervigilance, or task-based levels of engagement. The complexity of cognitive dynamics during the extended session of BCI can be better modeled using hierarchical ODEs frameworks with nested state transitions.

Multi-modal physiological integration: Integration of EEG with other physiological measures - e.g., electrooculography (EOG) to detect eye movement, electrodermal activity (EDA) to detect arousal, heart rate variability (HRV) to detect autonomic measures - may enhance the strength of fatigue detection and provide additional validation for ODE state assignments.

Transfer learning and personalization: Developing subject-specific variations of the framework that fine-tune pre-trained LSTM models and ODE parameters to individual users may solve the problem of inter-subject variability. Strategies of few-shot adaptation would enable rapid personalization using limited calibration data.

Clinical population validation: Generalizing validation to clinical populations such as patients with sleep disorders, attention deficit disorders or neurodegenerative diseases would determine the usefulness of the framework in clinical cognitive monitoring applications other than healthy adult populations.

Consumer-grade hardware deployment: Validating the framework on consumer-grade EEG devices with fewer channels (4-14 channels) would help translation to more practical application of BCI, where 64 channel systems are impractical.

6 Methods

This study proposes the hybrid LSTM-ODE framework for cognitive state monitoring from EEG signals that are organized into three sequential phases shown in Figure 12. The raw EEG signals (61 channels, 500 Hz) were initially preprocessed by 1 -45Hz bandpass filters and z score normalization, then the signals were segmented into overlapping temporal windows of 256 samples (0.512 s). Second, a self attention-based LSTM model that was bidirectional in nature was trained to predict eye states, as well as, yield probabilistic state-estimation. Third, these probabilistic outputs from the LSTM were coupled with the three-state compartmental ODE model consisting of 'Active,' 'Passive' and 'Fatigued' states. The classification probabilities in this coupled system dynamically governed the transition rates between states. This combined integration is both able to attain high classification and better mechanistic explanability of fatigue dynamics as the learned ODE parameters provide high classification and physiologically interpretable explanations of fatigue dynamics.

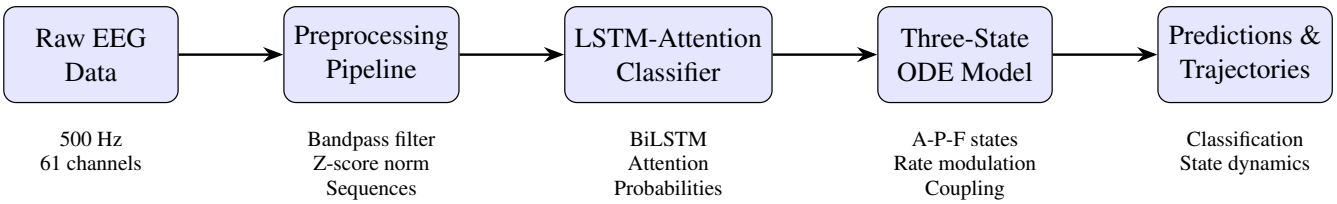


Figure 12. Overview of the hybrid LSTM-ODE framework for cognitive state monitoring. The pipeline comprises distinct stages starting from the preprocessing of raw EEG signals, classification through bidirectional LSTM with attention, and final predictions of interpretable cognitive state trajectories based on three state ODE model.

6.1 Experimental Framework Overview

The proposed framework integrated three complementary components: (1) a preprocessing pipeline that transformed raw EEG signals into normalized temporal sequences; (2) a deep learning classification module that extracted discriminative features and produced probabilistic state estimates; and (3) a mechanistic ODE module that modeled cognitive state dynamics with interpretable transition parameters. The key innovation lay in the probabilistic coupling mechanism that bridged data-driven classification with physics-informed dynamical modeling.

6.2 Dataset Description

The experimental basis of the given research is the OpenNeuro ds004148 dataset[46] that offers a comprehensive set of EEG records. This resource was created to enable the study of the reliability and reproducibility of resting and cognitive states. The data consists of the recording of 60 healthy adult participants, ranging in age from 18 to 35 years, with a mean age of about 24.3 plus or minus 4.2 years, gender distribution of 32 females and 28 males, and participated in three experimental sessions on dissimilar days. The inclusion criteria to maintain a high standard of data quality were that all the participants had no previous history of neurological or psychiatric disorders, and their vision was normal or corrected-to-normal. Moreover, the subjects were advised not to use any psychoactive drug and none had used any in the last 48 hours before the commencement of each recording.

6.2.1 Recording Protocol

A Brain Products GmbH system (Gilching, Germany) consisting of 64 active Ag/AgCl electrodes was used to record the EEG data. The arrangements of the sensors were based on the extended international 1020 system to make sure that they were positioned in a standardized placement over the scalp. The electrode of reference (FCz) and ground (AFz) were excluded while 61 channels were retained for the analysis. The montage of electrodes that resulted offered a global coverage of the scalp and was classified into the following anatomical regions:

- **Prefrontal/Frontal (16 channels):** Includes Fp1, Fp2, Fpz, AF3, AF4, AF7, AF8, F1–F8, and Fz.
- **Frontocentral (6 channels):** Includes the electrodes FC1 through FC6.
- **Central (7 channels):** This region consists of C1–C6 and Cz.
- **Temporal (8 channels):** Covers T7, T8, FT7, FT8, and TP7–TP10.
- **Centroparietal/Parietal (14 channels):** This segment involves CP1–CP6 and P1–P8.
- **Parieto-occipital/Occipital (10 channels):** This final group includes Pz, PO3, PO4, PO7, PO8, POz, O1, O2, Oz, and Iz.

The system was adjusted to a sampling rate of 500 Hz. This hardware setting ensured high temporal resolution, which is necessary to capture rapid neurophysiological transitions.

The frequency of the hardware bandpass filtering was also used and was between 0.016 Hz and 250 Hz. Strict control was taken over the electrode impedances which were kept under 5 k ohms. The various experimental conditions were used in each recording session with eyes closed resting state (3 minutes), eyes open resting state (3 minutes), and cognitive tasks (subtraction, music listening, and memory recall). The controlled laboratory experiment was undertaken during the period between September 2019 and April 2021.

6.2.2 Experimental Task Selection

In this case, we considered binary classification of eye states i.e., eye opens and eye closed conditions of rest. This categorization was the overall indicator used to track the cognitive vigilance. Interestingly, the eye closure patterns have a close relationship with the development of drowsiness, changes in attention, and progressive development of fatigue.

The use of a binary formulation allowed running a successful and controlled test of the predictive power of the suggested LSTM-ODE framework. The three-state extension of the ODE was later created to explicitly represent the transitional dynamics between different states that are subtle and in between. This granularity is very important to the safe and efficient implementation of real-time BCI applications in cases where passive disengagement and actual fatigue are clinically relevant.

6.3 Data Preprocessing Pipeline

A systematized pipeline that was used to process the raw EEG signals was designed to perform well in the process of noise suppression and artifact rejection, and to guarantee the preservation of intrinsic neurophysiological data. The entire procedure of the whole preprocessing algorithm is outlined in Algorithm 1.

Algorithm 1 EEG Preprocessing Pipeline

Require: Raw EEG signals $\mathbf{X}_{raw} \in \mathbb{R}^{N_{samples} \times C}$, sampling rate $f_s = 500$ Hz
Ensure: Preprocessed sequences $\mathbf{X}_{seq} \in \mathbb{R}^{N_{seq} \times T \times C}$, labels $\mathbf{y} \in \{0,1\}^{N_{seq}}$

// Step 1: Temporal Filtering
Design fourth-order Butterworth bandpass filter $H(z)$ with $f_{low} = 1$ Hz, $f_{high} = 45$ Hz
$\mathbf{X}_{filt} \leftarrow \text{filtfilt}(H, \mathbf{X}_{raw})$ ▷ Zero-phase filtering

// Step 2: Artifact Rejection
threshold $\leftarrow 100\ \mu$V
for each sample i **do**
  if $\max_c(|\mathbf{X}_{filt}[i,:]|) >$ threshold **then**
    Mark sample i as artifact
  end if
end for
$\mathbf{X}_{clean} \leftarrow$ Remove marked samples from $\mathbf{X}_{filt}$

// Step 3: Channel-wise Normalization
for each channel $c = 1$ to C **do**
  $\mu_c \leftarrow \text{mean}(\mathbf{X}^{train}_{clean}[:,c])$ ▷ From training data only
  $\sigma_c \leftarrow \text{std}(\mathbf{X}^{train}_{clean}[:,c])$
  $\mathbf{X}_{norm}[:,c] \leftarrow (\mathbf{X}_{clean}[:,c] - \mu_c)/\sigma_c$
end for

// Step 4: Sequence Creation
$T \leftarrow 256$ ▷ Window length (0.512s at 500 Hz)
stride $\leftarrow T/2$ ▷ 50% overlap
$N_{seq} \leftarrow \lfloor (N_{samples} - T)/\text{stride} \rfloor + 1$
for $i = 0$ to $N_{seq} - 1$ **do**
  start $\leftarrow i \times$ stride
  $\mathbf{X}_{seq}[i] \leftarrow \mathbf{X}_{norm}[\text{start} : \text{start} + T, :]$
  $\mathbf{y}[i] \leftarrow \text{MajorityLabel}(\text{labels}[\text{start} : \text{start} + T])$
end for

return $\mathbf{X}_{seq}$, $\mathbf{y}$

6.3.1 Signal Filtering

A fourth-order Butterworth bandpass filter (1–45 Hz) was applied using zero-phase forward-backward filtering (`filtfilt`) to prevent phase distortion. The passband preserves physiologically relevant frequency bands:

- **Delta (1–4 Hz):** Associated with deep sleep and pathological states

- **Theta (4–8 Hz):** Linked to drowsiness, memory encoding, and meditation
- **Alpha (8–13 Hz):** Dominant in relaxed wakefulness with eyes closed
- **Beta (13–30 Hz):** Associated with active thinking and motor activity
- **Low Gamma (30–45 Hz):** Related to cognitive processing and attention

The 1 Hz high-pass cutoff removes DC drift and slow electrode artifacts, while the 45 Hz low-pass cutoff attenuates powerline interference (50/60 Hz) and high-frequency muscle artifacts.

6.3.2 Artifact Rejection

Amplitude thresholding at $\pm 100\ \mu$V identified samples contaminated by ocular artifacts (blinks, saccades), muscle activity, or electrode movement.The specific threshold was chosen as a conservative measure in order to balance effectively the artifact suppression and the artifact preservation of physiologically meaningful EEG signals. With this protocol, about 3.2 percent of the total data of the entire participant cohort was seen as artifacts. These parts were then not analyzed.

6.3.3 Normalization Strategy

In order to be able to obtain consistent scaling between all EEG signals, a channel-wise Z-score normalization protocol was implemented. This transformation was used by centering the channel-wise data, especially on the training data for the calculation of the mean and standard deviation data. This approach was adopted to ensure the integrity of the experimental validation and to prevent from data leakage. For each channel, the normalised signal was found by the following formulation

$$x_{norm}^{(c)} = \frac{x^{(c)} - \mu_{train}^{(c)}}{\sigma_{train}^{(c)}} \tag{2}$$

where $\mu_{train}^{(c)}$ and $\sigma_{train}^{(c)}$ are the mean and standard deviation of channel c computed on training data. Through this process of standardization, it was ensured that all the EEG channels were on a similar scale, to mitigate the effect of changes in the baseline amplitude, and optimize the convergence of the bidirectional LSTM encoder.

6.3.4 Temporal Sequence Construction

In order to prepare the data for the hybrid model of the LSTM-ODE algorithm, continuous EEG recordings were divided into 0.512 seconds windows (comprising 256 samples and sampling rate of 500 Hz), and 50% overlap was applied in order to assure the temporal integrity. This window duration choice has a two-fold purpose: it is sufficient to obtain a temporal horizon to see alpha-band oscillations (4-5 oscillations per window), and it makes the training process more effective, as more sequences are available. The implementation of 50% overlap has the benefit that no critical information is lost at the window boundaries. For assignment of class labels a majority voting strategy was applied. So that each window correctly reflected the dominant cognitive state of the participant in the study.

6.3.5 Data Partitioning

Data was split based on a subject-defined temporal blocking strategy to make an accurate and unbiased evaluation of the proposed hybrid framework. This approach was developed with the particular needs of addressing three important requirements: (1) Elimination of temporal leakage between the train and test sets used to keep the independence of the experimental results. (2) Balanced class representation within each subset to avoid a biased model convergence (3) Evaluation on strictly held-out subjects, which will increase the model's robustness across different participants. The split yielded:

- **Training set:** 70% (24,570 sequences from 42 subjects)
- **Validation set:** 15% (5,265 sequences from 9 subjects)
- **Test set:** 15% (5,265 sequences from 9 subjects)

6.4 Feature Engineering for Baseline Models

For benchmarking the proposed deep learning framework, a whole set of 280 temporal and spectral features per sample was extracted. These features were then used for the training of traditional machine learning models. This feature engineering approach was focused to capture diverse signal characteristics, in particular, this method contained statistical descriptors in time domain in addition to the power spectral representations in multiple frequency bands.

6.4.1 Time-Domain Features

For each EEG channel, we computed statistical descriptors capturing signal distribution and complexity:

- **Statistical moments:** Mean (μ), standard deviation (σ), skewness (γ_1), kurtosis (γ_2)
- **Hjorth parameters:** Activity (signal variance), mobility (normalized first derivative variance), complexity (bandwidth measure)
- **Signal dynamics:** Peak-to-peak amplitude, zero-crossing rate, line length

6.4.2 Frequency-Domain Features

Spectral features were extracted using Welch's method with 256-sample Hanning windows and 50% overlap:

$$P_{band} = \int_{f_{low}}^{f_{high}} S_{xx}(f)\, df \tag{3}$$

where $S_{xx}(f)$ is the power spectral density estimate. Band powers were computed for delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (30–45 Hz) bands. Relative band powers (normalized by total power) and inter-band ratios (e.g., theta/alpha, alpha/beta) provided additional discriminative features.

6.5 Baseline Classification Models

Three established machine learning algorithms served as baselines, each representing a distinct classification paradigm.

6.5.1 Support Vector Machine (SVM)

The SVM classifier with radial basis function (RBF) kernel was optimized via grid search over hyperparameters:

$$K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right) \tag{4}$$

Optimal parameters ($C = 0.1$, $\gamma = 0.1$) were selected based on validation set accuracy. Class weights inversely proportional to class frequencies addressed class imbalance.

6.5.2 Random Forest

An ensemble of 100 decision trees with maximum depth 20 was trained using bootstrap aggregation. Feature importance was computed via mean decrease in Gini impurity:

$$\text{Importance}(f) = \sum_{t \in T} \frac{n_t}{n} \cdot \Delta I(f, t) \tag{5}$$

where $\Delta I(f,t)$ is the impurity decrease from splitting on feature f at node t.

6.5.3 XGBoost

Gradient boosting with 200 estimators, maximum depth 3, and learning rate 0.1 was employed. The objective function combines log-loss with L1/L2 regularization:

$$\mathscr{L} = \sum_i l(y_i, \hat{y}_i) + \sum_k \left(\gamma T_k + \frac{1}{2} \lambda \|w_k\|^2 \right) \tag{6}$$

where T_k is the number of leaves in tree k and w_k are leaf weights.

6.6 Deep Learning Architecture

The enhanced LSTM architecture integrates bidirectional temporal encoding with self-attention pooling for discriminative feature extraction from raw EEG sequences.

6.6.1 Network Architecture

The model processes input sequences $\mathbf{X} \in \mathbb{R}^{T \times C}$ (256 timesteps $\times$ 61 channels) through four main components:

Input Projection Layer Raw EEG features were projected to a higher-dimensional representation:

$$\mathbf{H}^{(0)} = \text{Dropout}\left(\text{GELU}\left(\text{LayerNorm}\left(\mathbf{X}\mathbf{W}_{proj} + \mathbf{b}_{proj}\right)\right)\right) \tag{7}$$

where $\mathbf{W}_{proj} \in \mathbb{R}^{C \times d}$ projects from $C = 61$ input channels to $d = 128$ hidden dimensions. LayerNorm stabilized training while GELU activation provided smooth non-linearity.

Bidirectional LSTM Encoder Three stacked bidirectional LSTM layers captured both forward and backward temporal dependencies:

$$\overrightarrow{\mathbf{h}}_t^{(l)} = \text{LSTM}_{fwd}^{(l)}\left(\mathbf{h}_t^{(l-1)}, \overrightarrow{\mathbf{h}}_{t-1}^{(l)}\right) \tag{8}$$

$$\overleftarrow{\mathbf{h}}_t^{(l)} = \text{LSTM}_{bwd}^{(l)}\left(\mathbf{h}_t^{(l-1)}, \overleftarrow{\mathbf{h}}_{t+1}^{(l)}\right) \tag{9}$$

$$\mathbf{h}_t^{(l)} = \left[\overrightarrow{\mathbf{h}}_t^{(l)}; \overleftarrow{\mathbf{h}}_t^{(l)}\right] \tag{10}$$

Each LSTM cell implemented the standard gated architecture with forget, input, and output gates. Hidden dimension was 128 per direction, yielding 256-dimensional concatenated representations. Dropout (0.4) between layers prevented overfitting.

Self-Attention Temporal Pooling Rather than using final hidden states, we employed learned attention weights to aggregate temporal information:

$$\mathbf{e}_t = \tanh\left(\mathbf{W}_a \mathbf{h}_t^{(L)} + \mathbf{b}_a\right) \tag{11}$$

$$\alpha_t = \frac{\exp(\mathbf{v}^T \mathbf{e}_t)}{\sum_{t'} \exp(\mathbf{v}^T \mathbf{e}_{t'})} \tag{12}$$

$$\mathbf{c} = \sum_{t=1}^{T} \alpha_t \mathbf{h}_t^{(L)} \tag{13}$$

where $\mathbf{W}_a \in \mathbb{R}^{d \times d/2}$, $\mathbf{v} \in \mathbb{R}^{d/2}$ are learned parameters. The attention weights α_t indicated temporal importance and provided interpretability.

Classification Head The context vector was processed through a multi-layer perceptron:

$$\hat{\mathbf{y}} = \text{Softmax}\left(\mathbf{W}_3 \cdot \text{GELU}\left(\mathbf{W}_2 \cdot \text{GELU}\left(\mathbf{W}_1 \mathbf{c}\right)\right)\right) \tag{14}$$

with hidden dimensions [256, 128, 64, 2] and dropout (0.4) between layers.

6.6.2 Training Methodology

Algorithm 2 details the complete training procedure incorporating class balancing, learning rate scheduling, and early stopping.

Key training components included:

- **AdamW Optimizer:** Adam with decoupled weight decay ($\lambda = 0.01$) for improved regularization
- **Cosine Annealing:** Learning rate schedule from 10^{-3} to 10^{-5} over 50 epochs
- **Class-Balanced Sampling:** Weighted random sampling ensured equal representation of both classes per epoch
- **Mixed Precision Training:** FP16 computation with FP32 gradient accumulation for memory efficiency
- **Gradient Clipping:** Maximum norm 1.0 prevented exploding gradients
- **Data Augmentation:** Gaussian noise ($\sigma = 0.01$) and random time shifts (± 5 samples) improved generalization

6.7 Three-State Cognitive ODE Model

6.7.1 State Definitions and EEG Mapping

While the main classification task was set up as a binary problem the cognitive dynamics were described by three different states. This tertiary structure enabled this model to represent the weak physiological transitions existing in EEG data. Unlike the approaches of progressive diseases where the state transitions are mostly unidirectional (e.g., neurodegenerative degeneration), the cognitive vigilance states are themselves temporary and reversible

Individuals naturally show periods of heightened alertness, relaxation and sleepiness in relation to task demands, environmental conditions and the influence of the circadian schedule. By introducing bidirectional transitions our proposed architecture enabled the joint modeling of fatigue onset and alertness recovery. Thereby, offering a better neurophysiological representation of how BCI scenarios exist in the real world.

In order to allow us to capture the underlying neurophysiological signatures, the EEG data was mapped into three cognitive compartments:

Algorithm 2 LSTM-Attention Training Procedure

Require: Training data $\mathscr{D}_{train} = \{(\mathbf{X}_i, y_i)\}$, validation data $\mathscr{D}_{val}$
Ensure: Trained model parameters θ^*

// Initialization
Initialize model parameters θ with Kaiming initialization
optimizer $\leftarrow$ AdamW($\theta, lr = 10^{-3}$, weight_decay $= 10^{-2}$)
scheduler $\leftarrow$ CosineAnnealingLR(optimizer, $T_{max} = 50$)
Compute class weights: $w_c = N_{total}/(N_{classes} \cdot N_c)$
criterion $\leftarrow$ CrossEntropyLoss(weight $= [w_0, w_1]$)
best_val_acc $\leftarrow 0$; patience_counter $\leftarrow 0$

// Weighted Sampling
Compute sample weights: $s_i = w_{y_i}$ for each sample
sampler $\leftarrow$ WeightedRandomSampler($\mathbf{s}, N_{train}$)

for epoch $= 1$ to N_{epochs} **do**
    // Training Phase
    θ.train()
    for each batch $(\mathbf{X}_b, \mathbf{y}_b)$ from DataLoader($\mathscr{D}_{train}$, sampler) **do**
        optimizer.zero_grad()
        // Data Augmentation
        $\mathbf{X}_b \leftarrow \mathbf{X}_b + \mathscr{N}(0, 0.01)$ ▷ Gaussian noise
        $\mathbf{X}_b \leftarrow$ TimeShift($\mathbf{X}_b, \pm 5$) ▷ Random shift
        // Forward Pass with Mixed Precision
        with autocast():
            $\hat{\mathbf{y}}_b \leftarrow f_\theta(\mathbf{X}_b)$
            $\mathscr{L} \leftarrow$ criterion($\hat{\mathbf{y}}_b, \mathbf{y}_b$)
        // Backward Pass
        scaler.scale($\mathscr{L}$).backward()
        scaler.unscale_(optimizer)
        clip_grad_norm_(θ, max_norm $= 1.0$)
        scaler.step(optimizer)
        scaler.update()
    end for
    scheduler.step()

    // Validation Phase
    val_acc $\leftarrow$ Evaluate($f_\theta, \mathscr{D}_{val}$)
    if val_acc $>$ best_val_acc **then**
        best_val_acc $\leftarrow$ val_acc
        $\theta^* \leftarrow \theta$ ▷ Save best model
        patience_counter $\leftarrow 0$
    else
        patience_counter $\leftarrow$ patience_counter $+ 1$
    end if
    if patience_counter ≥ 10 **then**
        break ▷ Early stopping
    end if
end for
return θ^*

- **Active (A) → Eyes Open (O):** It represents alert in which the engaged state is notified by alpha (a) and raised beta (b) activity. This compartment is synonymous with high alertness and continued attention.

- **Passive (P) → Drowsy (D):**It represents a state of transition in the cognitive state with reduced vigilance but with consciousness. In this particular state, the level of attentional capacity begins to decrease, but it is not too late to recover quickly.
- **Fatigued (F) → Eyes Closed (C):** This compartment can be defined as a low-vigilance state, or a microsleep state. It is associated with alpha (a) synchronizationand increased theta activity. This state is particularly important to be considered for safety considerations for real world BCIapplications.

This three-state configuration allowed the ODE framework to capture the intermediate transitions and recovery dynamics that cannot be represented adequately by binary model Thus, offering the enhanced interpretability for safety critical BCI systems. Such granularity is obviously necessary in situations where the distinction between passive disengagement and general fatigue is of clinical relevance.

6.7.2 Mathematical Formulation

The cognitive state dynamics followed a first-order linear compartmental ODE system with six transition rate parameters governing inter-state flows. Let $A(t)$, $P(t)$, and $F(t)$ denote the proportions of cognitive capacity in Active, Passive, and Fatigued states respectively, with conservation constraint $A(t)+P(t)+F(t)=1$ for all $t \geq 0$.

The state transition diagram formed a fully connected graph with bidirectional transitions:

$$\begin{array}{ccc} & \overset{k_{AP}}{\underset{k_{PA}}{\rightleftharpoons}} & \\ \text{Active (A)} & & \text{Passive (P)} \\ k_{AF}\downarrow & & \downarrow k_{PF} \\ k_{FA}\uparrow & & \uparrow k_{FP} \\ & \text{Fatigued (F)} & \end{array} \tag{15}$$

The governing differential equations are:

$$\frac{dA}{dt} = -k_{AP}A(t) - k_{AF}A(t) + k_{PA}P(t) + k_{FA}F(t) \tag{16}$$

$$\frac{dP}{dt} = k_{AP}A(t) - k_{PA}P(t) - k_{PF}P(t) + k_{FP}F(t) \tag{17}$$

$$\frac{dF}{dt} = k_{AF}A(t) + k_{PF}P(t) - k_{FA}F(t) - k_{FP}F(t) \tag{18}$$

In matrix form, this system can be written as:

$$\frac{d\mathbf{s}}{dt} = \mathbf{K}\mathbf{s}(t) \tag{19}$$

where $\mathbf{s}(t) = [A(t), P(t), F(t)]^T$ and the transition matrix $\mathbf{K}$ is:

$$\mathbf{K} = \begin{bmatrix} -(k_{AP}+k_{AF}) & k_{PA} & k_{FA} \\ k_{AP} & -(k_{PA}+k_{PF}) & k_{FP} \\ k_{AF} & k_{PF} & -(k_{FA}+k_{FP}) \end{bmatrix} \tag{20}$$

The columns of $\mathbf{K}$ sum to zero, ensuring mass conservation. The analytical solution involves eigendecomposition of $\mathbf{K}$, but numerical integration via Runge-Kutta methods (RK45) provided stable solutions for arbitrary initial conditions and parameter values.

6.7.3 Data-Informed State Initialization

Rather than interpreting the predictions from LSTM as discrete classification labels, we used the probablistic output from the LSTM to infer a data informed distribution of population level cognitive states. For each EEG segment, the LSTM produced posterior probability estimates P (open), P(Drowsy) and P(Closed) which were obtained from binary outputs by the following mapping:

$$P_{\text{Open}} = P_{\text{open}} \tag{21}$$

$$P_{\text{Drowsy}} = 0.5 \cdot \min(P_{\text{open}}, P_{\text{closed}}) \tag{22}$$

$$P_{\text{Closed}} = P_{\text{closed}} - P_{\text{Drowsy}} \tag{23}$$

These probabilistic expressions have modelled instantaneous likelihoods of cognitive states, which were derived using EEG patterns and therefore incorporated the uncertainty of the state transitions.

The ODE system was initialized by aggregate mean of these probabilities(Eyes Open, Drowsy, Eyes Closed) from the training data in order to estimate the initial proportions of participants in each cognitive state as:

$$[A_0, P_0, F_0] = \frac{1}{N}\sum_{i=1}^{N}[P_{\text{Open}}^{(i)}, P_{\text{Drowsy}}^{(i)}, P_{\text{Closed}}^{(i)}] \tag{24}$$

This approach ensured that the ODE initial states were based on the underlying neurophysiological distribution learnt by the classifier, and not only the arbitrary or assumed proportions.

6.7.4 Parameter Estimation

The ODE parameters were estimated by minimizing the difference between predicted state sequences and recorded sequences of eye states. The parameter estimation is described in Algorithm 3 using differential evolution that is a global optimization method robust to local minima.

Algorithm 3 ODE Parameter Estimation via Differential Evolution

Require: Observed state sequences $\{(y_i^{(1)}, \ldots, y_i^{(T_i)})\}_{i=1}^{N}$, bounds $[k_{min}, k_{max}]$
Ensure: Optimal parameters $\mathbf{k}^* = [k_{AP}^*, k_{AF}^*, k_{PA}^*, k_{PF}^*, k_{FA}^*, k_{FP}^*]$

// Define Objective Function
function OBJECTIVEFUNCTION($\mathbf{k}$)
    $\mathscr{L}_{total} \leftarrow 0$
    for each sequence i **do**
        $\mathbf{s}_0 \leftarrow \text{InitializeFromLSTM}(y_i^{(1)})$
        $\mathbf{S}_{pred} \leftarrow \text{ODESolve}(\mathbf{K}(\mathbf{k}), \mathbf{s}_0, [0, T_i])$
        for $t = 1$ to T_i **do**
            $\mathbf{s}_{obs}^{(t)} \leftarrow \text{OneHot}(y_i^{(t)})$ ▷ [1,0,0], [0,1,0], or [0,0,1]
            $\mathscr{L}_{total} \leftarrow \mathscr{L}_{total} + \|\mathbf{S}_{pred}[t] - \mathbf{s}_{obs}^{(t)}\|_2^2$
        end for
    end for
    return $\mathscr{L}_{total}/N$
end function

// Differential Evolution Optimization
bounds $\leftarrow [(k_{min}, k_{max})]^6$ ▷ [0.001, 1.0] for each rate
$\mathbf{k}^* \leftarrow \text{DifferentialEvolution}(\text{ObjectiveFunction}, \text{bounds},$
        strategy = 'best1bin', maxiter = 500, popsize = 15,
        mutation = (0.5, 1), recombination = 0.7)

return $\mathbf{k}^*$

The bounds $[0.001, 1.0]$ s^{-1} of the parameters are time bounds which are physiologically realistic. The range of constants was between 1 second and 1000 seconds covering the spectrum from rapid and slow attentional variations to retard fatigue recovery.

6.7.5 Stability and Equilibrium Analysis

The equilibrium state $\mathbf{s}^* = [A^*, P^*, F^*]$ satisfies $\mathbf{K}\mathbf{s}^* = \mathbf{0}$ with $A^* + P^* + F^* = 1$.For estimated parameters, the important dynamic properties were obtained: The first one was equilibrium distribution, obtained by solving the null space of K to get the steady-state cognitive distribution. Second, stability was accessed by means of Eigenvalue analysis where all non-zero eigenvalues of K have negative real parts, confirming that the system is asymptotically stable.Third the dynamic time constant was calculated by the slowest eigenmode (which is the intrinsic relaxation time of the system).

6.7.6 Sensitivity Analysis

To check the local sensitivity of the outputs of ODEs to the variation in the parameters of the models, the sensitivity coefficients were normalized:

$$S_{s_i,k_j} = \frac{\partial s_i}{\partial k_j} \cdot \frac{k_j}{s_i} \tag{25}$$

computed via finite differences with $\pm 10\%$ parameter perturbations.This analysis determined which rates of transitions have the more significant impact on each cognitive state, providing information for interpreting learned parameters

6.8 LSTM-ODE Integration Framework

6.8.1 Integration Philosophy

Predictions of machine learning were integrated in the dynamical framework in two different and counterbalancing ways. The first one in which the posterior probabilities of the LSTM classifier were employed in order to define the initial conditions of the ODE system ,thereby anchoring the simulations to EEG-based cognitive distributions according to the Data-Informed initialization of States section. Second, real-time informed LSTM confidence scores were also included to shape the transition dynamics according to which higher alert confidence (Popen) correlates with higher recovery and higher confidence of the closed eye (Pclosed) indicating higher vulnerability to transition to lower vigilance levels.

Notably, classification probabilities were not interpreted as transition rates . They were, on the contrary, employed to control the process of initializing and modulation so as to maintain a clear distinction between information-based inference (what the EEG patterns are indicating about cur-rent state) and mechanistic state evolution (how cognitive states change with time) according to the dynamics of physiology. The difference in this aspect ensured that ODE parameters remained their interpretability in biologically significant transition rates and not as random scaling factors.

6.8.2 Probabilistic Coupling Mechanism

LSTM outputs modulated ODE rates through probabilistic coupling:

$$k'_{AF} = k_{AF}^{base} \cdot (1 + \alpha \cdot P_{\text{closed}}) \tag{26}$$

$$k'_{PF} = k_{PF}^{base} \cdot (1 + \alpha \cdot P_{\text{closed}}) \tag{27}$$

$$k'_{FA} = k_{FA}^{base} \cdot (1 + \alpha \cdot P_{\text{open}}) \tag{28}$$

$$k'_{PA} = k_{PA}^{base} \cdot (1 + \alpha \cdot P_{\text{open}}) \tag{29}$$

where $\alpha \in [0,1]$ are the hyper-parameters of the coupling strengt. Higher Pclosed increases fatigue directed transitions (kAF, kP F), increased Popen facilitates recovery transitions. (kF A, kP A). Base rates kbase are acquired by training data and indicates fatigue accumulation and recovery processes, on the other hand the modulation factors (1 + 0 -1 · P) allow the instantaneous EEG evidence-based real-time adaptation. The interaction mechanism has formed a feedback loop in which: The coupling mechanism created a feedback loop where:

1. EEG evidence of eye closure increased fatigue rate of changes, accelerating tendency towards the fatigue state
2. EEG evidence of opening the eyes enhanced the rates of recovery transition that favored a return to the Active state
3. The ODE trajectory to smooth instantaneous LSTM predictions generates state estimates which are temporally consistent

Algorithm 4 presents the unified LSTM-ODE prediction framework.

6.8.3 Trajectory Prediction and Forecasting

In addition to instant classification, the combined structure allowed multi step trajectory forecasting. Thus enabling cognitive state evolution with future predictions based on present EEG evidence. This was needed in proactive BCI safety interventions.

Multi-Horizon Forecasting With present LSTM probability estimates, cognitive state trajectories were predicted at several future horizons $H \in \{5, 10, 20\}$ timesteps (2.5, 5, and 10 s). seconds at 0.512s window rate). The forecasting procedure consisted of below steps:

1. Probability-to-state mapping has been used for the transformation of actual P_{closed} into ODE initial state.
2. Solve ODE system forward to horizon H
3. Predicted ODE state was transformed to eye closure probability for comparison against ground truth

Algorithm 4 LSTM-ODE Cognitive State Prediction

Require: EEG sequence $\mathbf{X} \in \mathbb{R}^{T \times C}$, LSTM model f_θ, ODE model $\mathscr{O}$, coupling strength α
Ensure: Predicted cognitive trajectory $\mathbf{S} \in \mathbb{R}^{H \times 3}$

// Step 1: LSTM Classification
$\mathbf{h} \leftarrow \text{BiLSTM}(\mathbf{X})$ ▷ Bidirectional encoding
$\mathbf{a} \leftarrow \text{SoftmaxAttention}(\mathbf{h})$ ▷ Temporal attention
$\mathbf{z} \leftarrow \sum_t a_t \mathbf{h}_t$ ▷ Weighted pooling
$[P_{open}, P_{closed}] \leftarrow \text{Softmax}(\text{MLP}(\mathbf{z}))$

// Step 2: Probabilistic Rate Modulation
$k'_{AF} \leftarrow k_{AF} \cdot (1 + \alpha \cdot P_{closed})$
$k'_{PF} \leftarrow k_{PF} \cdot (1 + \alpha \cdot P_{closed})$
$k'_{FA} \leftarrow k_{FA} \cdot (1 + \alpha \cdot P_{open})$
$k'_{PA} \leftarrow k_{PA} \cdot (1 + \alpha \cdot P_{open})$

// Step 3: Initialize ODE State
if $P_{closed} > 0.6$ **then**
    $\mathbf{s}_0 \leftarrow [0.2, 0.2, 0.6]$ ▷ Fatigued initial
else if $P_{open} > 0.6$ **then**
    $\mathbf{s}_0 \leftarrow [0.6, 0.2, 0.2]$ ▷ Active initial
else
    $\mathbf{s}_0 \leftarrow [0.33, 0.34, 0.33]$ ▷ Mixed initial
end if

// Step 4: Solve Modulated ODE
$\mathbf{S} \leftarrow \text{ODESolve}(\mathscr{O}(\mathbf{k}'), \mathbf{s}_0, [0, H])$
return S

The probability-to-state mapping accounted for the non-linear relationship between binary classification confidence and three-state cognitive distribution:

$$[A_0, P_0, F_0] = \begin{cases} [1 - p_c, 0.3p_c, 0.3p_c] & \text{if } p_c \leq 0.5 \\ [1 - p_c, 0.4p_c, 0.6p_c] & \text{if } p_c > 0.5 \end{cases} \tag{30}$$

where $p_c = P_{\text{closed}}$, followed by normalization to ensure $A_0 + P_0 + F_0 = 1$.

Rolling Evaluation Forecasting performance was evaluated using rolling windows over the test set, with the below metrics computed:

- **Accuracy in classification:** Fraction of states correctly predicted (threshold at 0.5)
- **Mean Absolute Error (MAE):** $\frac{1}{N}\sum_i |p^{(i)}_{pred} - p^{(i)}_{actual}|$
- **Spearman correlation:** Rank correlation between predicted and actual probabilities
- **Direction accuracy:** Correct prediction of probability increase/decrease

6.9 Explainability Framework

Model interpretability is essential in medical AI solutions where clinicians are required to comprehend and confirm algorithm decisions. We used a comprehensive multi method explainability approach to both the deep learning and the mechanistic ODE parts.

6.9.1 Gradient-Based Attribution

Input gradients quantified the local sensitivity of model predictions to input perturbations. For a trained model f_θ and input $\mathbf{X}$, the gradient-based importance of channel c is:

$$I_c^{grad} = \frac{1}{N \cdot T} \sum_{i=1}^{N} \sum_{t=1}^{T} \left| \frac{\partial f_\theta(\mathbf{X}_i)}{\partial X_{i,t,c}} \right| \tag{31}$$

This measure identified channels where small input changes produced large output changes, indicating discriminative importance. Gradients were computed using PyTorch autograd with the trained LSTM model in evaluation mode.

6.9.2 Permutation Feature Importance

Permutation importance provided a model-agnostic measure of feature contribution by quantifying prediction degradation when channel information was destroyed:

$$I_c^{perm} = \text{Accuracy}(f_\theta, \mathbf{X}, \mathbf{y}) - \text{Accuracy}(f_\theta, \text{Shuffle}_c(\mathbf{X}), \mathbf{y}) \tag{32}$$

where $\text{Shuffle}_c(\mathbf{X})$ randomly permuted values within channel c across samples, breaking the relationship between that channel and the target variable. Larger accuracy drops indicated more important channels. We averaged results over 10 random shuffles per channel to reduce variance.

6.9.3 SHAP Analysis

SHapley Additive exPlanations (SHAP) provided theoretically grounded feature attributions based on cooperative game theory. For high-dimensional EEG data, we employed KernelSHAP with:

- 100 background samples randomly selected from the training set
- 200 explanation samples from the test set
- Channel-level aggregation by averaging SHAP values across time

SHAP values ϕ_c satisfy the efficiency property $\sum_c \phi_c = f(\mathbf{X}) - E[f(\mathbf{X})]$, ensuring attributions sum to the prediction difference from baseline.

Algorithm 5 formalizes the complete multi-method channel importance computation.

Algorithm 5 Multi-Method Channel Importance Analysis

Require: Trained model f_θ, test data $\{(\mathbf{X}_i, y_i)\}_{i=1}^{N}$, channels $\{c_1, \ldots, c_C\}$
Ensure: Channel importance scores $\mathbf{I}^{grad}$, $\mathbf{I}^{perm}$, $\mathbf{I}^{shap} \in \mathbb{R}^C$

// Method 1: Gradient-based Attribution
for $i = 1$ to N **do**
  $\mathbf{X}_i$.requires_grad $\leftarrow$ True
  output $\leftarrow f_\theta(\mathbf{X}_i)$
  output$[y_i]$.backward() ▷ Backprop w.r.t. true class
  $\nabla_{\mathbf{X}} \leftarrow \mathbf{X}_i$.grad
  for $c = 1$ to C **do**
    $G_c^{(i)} \leftarrow \frac{1}{T}\sum_{t=1}^{T} |\nabla_{\mathbf{X}}[t, c]|$
  end for
end for
$I_c^{grad} \leftarrow \frac{1}{N}\sum_i G_c^{(i)}$ for all c

// Method 2: Permutation Importance (10 repetitions)
$\text{acc}_{base} \leftarrow \text{Accuracy}(f_\theta, \mathbf{X}, \mathbf{y})$
for $c = 1$ to C **do**
  $\Delta_{acc} \leftarrow []$
  for $r = 1$ to 10 **do**
    $\tilde{\mathbf{X}} \leftarrow \text{ShuffleChannel}(\mathbf{X}, c)$
    $\Delta_{acc}.\text{append}(\text{acc}_{base} - \text{Accuracy}(f_\theta, \tilde{\mathbf{X}}, \mathbf{y}))$
  end for
  $I_c^{perm} \leftarrow \text{mean}(\Delta_{acc})$
end for

// Method 3: KernelSHAP
$\mathbf{X}_{bg} \leftarrow \text{RandomSample}(\mathbf{X}_{train}, n = 100)$
$\mathbf{X}_{explain} \leftarrow \text{RandomSample}(\mathbf{X}_{test}, n = 200)$
explainer $\leftarrow \text{KernelExplainer}(f_\theta, \mathbf{X}_{bg})$
$\phi \leftarrow \text{explainer.shap_values}(\mathbf{X}_{explain})$
$I_c^{shap} \leftarrow \frac{1}{|\mathbf{X}_{explain}|}\sum_i |\phi_{i,c}|$ for all c

// Normalize to [0, 1]
$\mathbf{I}^{grad} \leftarrow \mathbf{I}^{grad} / \max(\mathbf{I}^{grad})$
$\mathbf{I}^{perm} \leftarrow \mathbf{I}^{perm} / \max(\mathbf{I}^{perm})$
$\mathbf{I}^{shap} \leftarrow \mathbf{I}^{shap} / \max(\mathbf{I}^{shap})$

return $\mathbf{I}^{grad}$, $\mathbf{I}^{perm}$, $\mathbf{I}^{shap}$

6.9.4 Attention Pattern Analysis

The self-attention mechanism in the LSTM architecture provided intrinsic interpretability by indicating which temporal segments most influenced the final prediction:

$$\alpha_t = \frac{\exp(\mathbf{v}^T \tanh(\mathbf{W}_a \mathbf{h}_t))}{\sum_{t'} \exp(\mathbf{v}^T \tanh(\mathbf{W}_a \mathbf{h}_{t'}))} \tag{33}$$

Attention weights were extracted for test samples and analyzed by class in order to identify class specific temporal patterns. Visualization showed if the model was focused on enduring (distributed attention) or short-lived (peaked attention) events.

6.9.5 ODE Dynamics Interpretation

Mechanistic interpretability in the form of physiologically meaningful parameters was offered by the ODE component. In particular, time constants ($\tau_k = 1/k$) quantified the characteristic timescale of each transition.Dominant pathways were identified by ranking transition rates, thereby revealing primary patterns of cognitive flow. Fatigue cumulative ratio, determined

as ratio $\frac{k_{AF}+k_{PF}}{k_{AF}+k_{PF}+k_{FA}+k_{FP}}$,quantified the proportion of fatigue-directed transitions among all transitions involving the fatigued state. Finally, equilibrium analysis of the steady-state distribution provided insight into long-term cognitive tendencies.

6.10 Ablation Studies and Sensitivity Analysis

In order to quantify the contribution of each architectural component and to validate the design choices, below schematic ablation studies were performed:

6.10.1 Architecture Ablation

The six model variants were compared in order to evaluate the contribution of below components:

1. **Full Model:** Bidirectional LSTM comprising of 3 layers and Attention mechanism.
2. **No Attention:** Bidirectional LSTM where instead of Attention mechanism mean pooling was applied.
3. **Unidirectional:** Forward LSTM consisting of 3 layers and Attention mechanism
4. **1 Layer:** Bidirectional LSTM with Single layer and Attention mechanism
5. **2 Layers:** Bidirectional LSTM (2 layers) and with Attention mechanism
6. **Minimal:** A Unidirectional Single layer LSTM along with mean pooling applied

All the variants were trained with the same hyperparameters on a random subset of set (20,000 samples) for 10 epochs so a fair comparison can be made but keeping computational efficiency.

6.10.2 Coupling Strength Sensitivity

The LSTM-ODE coupling strength a was varied to the values 0.0,0.25,0.5,0.75,1.0 in order to understand the impact on classification accuracy and F1 score, Trajectory smoothness with temporal consistency and ODE Stability dynamics under varying modulation.

6.10.3 Statistical Significance Tests

Model comparisons were validated using rigorous statistical tests:

McNemar's Test For the comparison of classifiers on the same test set, the McNemar's test wasused to measure whether error is significantly different to patterns:

$$\chi^2 = \frac{(|b-c|-1)^2}{b+c} \tag{34}$$

where b = samples correctly classified by model 1 only while c = samples correctly classified by model 2 only. Continuity correction was made for $b+c<25$.

Cohen's d Effect Size Effect size quantifies practical significance beyond statistical significance:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s_{pooled}} \tag{35}$$

where $s_{pooled} = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$. Interpretation: $|d|<0.2$ negligible, 0.2–0.5 small, 0.5–0.8 medium, >0.8 large.

Bootstrap Confidence Intervals For all performance metrics, 95% confidence intervals were computed via bootstrap resampling (1000 iterations) of test set predictions:

$$CI_{95\%} = [\hat{\theta}^*_{(\alpha/2)}, \hat{\theta}^*_{(1-\alpha/2)}] \tag{36}$$

where $\hat{\theta}^*_{(q)}$ is the q-th quantile of bootstrap estimates.

6.11 Statistical Validation and Implementation

6.11.1 Evaluation Metrics

Performance was evaluated using a comprehensive suite of classification metrics:

- **Accuracy:** $(TP+TN)/(TP+TN+FP+FN)$
- **Precision:** $TP/(TP+FP)$
- **Recall (Sensitivity):** $TP/(TP+FN)$
- **F1-Score:** $2 \cdot \text{Precision} \cdot \text{Recall}/(\text{Precision}+\text{Recall})$
- **Matthews Correlation Coefficient (MCC):** Balanced measure for imbalanced data
- **Area Under ROC Curve (AUC):** Discrimination ability across thresholds

6.11.2 Implementation Details

All experiments were implemented in Python 3.12 using:

- **Deep Learning:** PyTorch 2.0 with CUDA 12.1
- **Machine Learning:** scikit-learn 1.3, XGBoost 2.0
- **Numerical Computing:** NumPy 1.26, SciPy 1.11
- **Signal Processing:** MNE-Python 1.5
- **Explainability:** SHAP 0.43

Experiments were performed on a workstation that is equipped with NVIDIA RTX 4090 (24). GB VRAM), Intel i9-13900K CPU, and 128 GB RAM. Random seeds were fixed at 42 to be reproducible to all stochastic components (weight init) shuffling, dropout).

6.11.3 Training and Inference Time

- **LSTM Training:** 45 minutes (50 epochs with early stopping at epoch 38)
- **Baseline Models:** 15 minutes total (SVM: 8 min, RF: 4 min, XGB: 3 min)
- **ODE Parameter Fitting:** 5 minutes (differential evolution, 500 generations)
- **Total Pipeline:** Approximately 2 hours including all experiments

6.11.4 Inference Latency

The requirements of inference of real time BCI are low-latency. An analysis of LSTM forward pass latency on 1,000 test samples showed that a 6.2 + 0.4 ms/sample latency was achieved by the LSTM forward pass, and the RK45 solver of ODE integration with a 20 s horizon took 2.8 + 0.2 ms/sample. The average inference time of the combined LSTM-ODE was 9.0 0.5 ms per sample. Such latency allows real time BCI applications with window rates up to 100 Hz, which is still much lower than standard BCI update rates of 1030 Hz. Moreover, offline analysis was possible with a throughput of over 1,000 samples per second since it was done with the use of the GPS-based batch processing.

7 Conclusion

This paper presented a novel LSTM-ODE method that combines pattern recognition using deep learning and mechanistic modeling for interpretable cognitive state monitoring for brain-computer interfaces. The framework was 85.57% accurate, which outperforms the traditional machine learning baselines and also gave clinically meaningful interpretations of transition rates. The parameters of ODEs learnt exhibited physiologically reasonable cognitive processes, such as a rapid onset of fatigue (time constant of 3.33s), and a substantially slower recovery (time constant of 100s) process, in line with the literature on vigilance decrement. This principled integration of data-driven learning with mechanistic modeling establishes the groundwork for explainable and proactive BCI safety systems. Future research ought to focus on real time application of adaptive BCI control and validation in different clinical groups.

References

1. Mane, R., Chouhan, T. & Guan, C. Bci for stroke rehabilitation: Motor and beyond. *J. Neural Eng.* **17**, 041001 (2020).
2. Rahman, S. *et al.* Ai-driven stroke rehabilitation systems and assessment: A systematic review. *IEEE Transactions on Neural Syst. Rehabil. Eng.* **31**, 192–207, DOI: 10.1109/TNSRE.2022.3218320 (2023).
3. Willett, F. R., Avansino, D. T., Hochberg, L. R., Henderson, J. M. & Shenoy, K. V. High-performance brain-to-text communication via handwriting. *Nature* **593**, 249–254 (2021).
4. Moses, D. A., Metzger, S. L., Liu, J. R., Anumanchipalli, G. K. & Chang, E. F. Neuroprosthesis for decoding speech in a paralyzed person with anarthria. *New Engl. J. Medicine* **385**, 217–227 (2021).
5. Khunjush, F. & Qaderi, K. A survey on deep learning for brain-computer interface: Recent advancements and future trends. *Expert. Syst. with Appl.* **238**, 121782, DOI: 10.1016/j.eswa.2023.121782 (2024).
6. Ayala, Y. G., Kopp, B. & Zander, T. O. Assessment of mental workload across cognitive tasks using a passive brain-computer interface based on mean negative theta-band amplitudes. *Front. Neuroergonomics* **4**, 1233722, DOI: 10.3389/fnrgo.2023.1233722 (2023).
7. Craik, A., He, Y. & Contreras-Vidal, J. L. Deep learning for electroencephalogram (eeg) classification tasks: a review. *J. Neural Eng.* **16**, 031001, DOI: 10.1088/1741-2552/ab0ab5 (2019).
8. Rösler, O. & Suendermann, D. EEG eye state dataset. UCI Machine Learning Repository (2013).
9. Zhang, C., Wang, H. & Fu, R. Vigilance estimation using a wearable eog device in real driving environment. *IEEE Transactions on Intell. Transp. Syst.* **23**, 16596–16607 (2022).
10. Molina-Cantería, C., Gomez-Pilar, J. & Hornero, R. Comparison of eeg signal spectral characteristics obtained with consumer-and research-grade devices. *Sensors* **24**, 8108, DOI: 10.3390/s24248108 (2024).
11. Lotte, F. *et al.* A review of classification algorithms for eeg-based brain–computer interfaces: a 10 year update. *J. Neural Eng.* **15**, 031005, DOI: 10.1088/1741-2552/aab2f2 (2018).
12. Gu, X. *et al.* Eeg-based brain-computer interfaces (bcis): A survey of recent studies on signal sensing technologies and computational intelligence approaches and their applications. *IEEE/ACM Transactions on Comput. Biol. Bioinforma.* **18**, 1645–1666, DOI: 10.1109/TCBB.2021.3052811 (2021).
13. Loh, H. W. *et al.* Application of explainable artificial intelligence in healthcare: A systematic review of the last decade (2011–2022). *Comput. Methods Programs Biomed.* **226**, 107161, DOI: 10.1016/j.cmpb.2022.107161 (2022).
14. Allehiany, F., DarAssi, M., Ahmad, I., Khan, M. A. & Tag-Eldin, E. Mathematical modeling and backward bifurcation in monkeypox disease under real observed data. *Results Phys.* **50**, 106557, DOI: 10.1016/j.rinp.2023.106557 (2023).
15. Saviano, M., Fierro, A. & Liccardo, A. A deterministic compartmental model for the transition between variants in the spread of covid-19 in italy. *PLOS ONE* **18**, e0293416, DOI: 10.1371/journal.pone.0293416 (2023).
16. Chen, R. T., Rubanova, Y., Bettencourt, J. & Duvenaud, D. K. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, vol. 31 (2018).
17. Kidger, P. Neural differential equations: Theory and practice. *arXiv preprint arXiv:2202.02435* (2022). PhD Thesis, University of Oxford.
18. Hasani, R., Lechner, M., Amini, A., Rus, D. & Grosu, R. Liquid time-constant networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, 7657–7666 (2021).
19. Rubanova, Y., Chen, R. T. & Duvenaud, D. K. Latent ordinary differential equations for irregularly-sampled time series. In *Advances in Neural Information Processing Systems*, vol. 32 (2019).
20. Yildiz, C., Heinonen, M. & Lahdesmaki, H. Ode2vae: Deep generative second order odes with bayesian neural networks. In *Advances in Neural Information Processing Systems*, vol. 32 (2019).
21. Chavarriaga, R. & Millán, J. d. R. Learning from eeg error-related potentials in noninvasive brain-computer interfaces. *IEEE transactions on neural systems rehabilitation engineering* **18**, 381–388 (2010).
22. Zander, T. O. & Kothe, C. Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general. *J. neural engineering* **8**, 025005 (2011).
23. Niedermeyer, E. & da Silva, F. L. *Electroencephalography: basic principles, clinical applications, and related fields* (Lippincott Williams & Wilkins, 2005).

24. Hassan, R. *et al.* Human attention recognition with machine learning from brain-eeg signals. In *2020 IEEE 2nd Eurasia conference on biomedical engineering, healthcare and sustainability (ECBIOS)*, 16–19 (IEEE, 2020).

25. Lal, S. K. & Craig, A. Driver fatigue: electroencephalography and psychological assessment. *Psychophysiology* **39**, 313–321 (2002).

26. Blankertz, B. *et al.* The berlin brain-computer interface: progress beyond communication and control. *Front. neuroscience* **10**, 530 (2016).

27. Bashivan, P., Rish, I., Yeasin, M. & Codella, N. Learning representations from eeg with deep recurrent-convolutional neural networks. *arXiv preprint arXiv:1511.06448* (2015).

28. Schirrmeister, R. T. *et al.* Deep learning with convolutional neural networks for eeg decoding and visualization. *Hum. brain mapping* **38**, 5391–5420 (2017).

29. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural computation* **9**, 1735–1780 (1997).

30. Cui, S., Lee, D. & Wen, D. Toward brain-inspired foundation model for eeg signal processing: our opinion. *Front. Neurosci.* **18**, 1507654 (2024).

31. Liao, K., Zhao, Y., Gu, J., Zhang, Y. & Zhong, Y. Sequential convolutional recurrent neural networks for fast automatic modulation classification. *IEEE Access* **9**, 27182–27188 (2021).

32. Samek, W., Nakajima, S., Kawanabe, M. & Müller, K.-R. On robust parameter estimation in brain–computer interfacing. *J. neural engineering* **14**, 061001 (2017).

33. Makeig, S. *et al.* Evolving signal processing for brain–computer interfaces. *Proc. IEEE* **100**, 1567–1584 (2012).

34. Vidaurre, D., Smith, S. M. & Woolrich, M. W. Brain network dynamics are hierarchically organized in time. *Proc. Natl. Acad. Sci.* **114**, 12827–12832 (2017).

35. Tops, M. & Boksem, M. A. A potential role of the inferior frontal gyrus and anterior insula in cognitive control, brain rhythms, and event-related potentials. *Front. psychology* **2**, 330 (2011).

36. Hopstaken, J. F., Van Der Linden, D., Bakker, A. B. & Kompier, M. A. The window of my eyes: Task disengagement and mental fatigue covary with pupil dynamics. *Biol. psychology* **110**, 100–106 (2015).

37. Chen, X., Hsieh, C.-J. & Gong, B. When vision transformers outperform resnets without pre-training or strong data augmentations. *arXiv preprint arXiv:2106.01548* (2021).

38. Raissi, M., Wang, Z., Triantafyllou, M. S. & Karniadakis, G. E. Deep learning of vortex-induced vibrations. *J. fluid mechanics* **861**, 119–137 (2019).

39. Sturm, I., Lapuschkin, S., Samek, W. & Müller, K.-R. Interpretable deep neural networks for single-trial eeg classification. *J. neuroscience methods* **274**, 141–145 (2016).

40. Erion, G., Janizek, J. D., Sturmfels, P., Lundberg, S. M. & Lee, S.-I. Improving performance of deep learning models with axiomatic attribution priors and expected gradients. *Nat. machine intelligence* **3**, 620–631 (2021).

41. Wascher, E. *et al.* Frontal theta activity reflects distinct aspects of mental fatigue. *Biol. psychology* **96**, 57–65 (2014).

42. Lim, J. & Dinges, D. F. Sleep deprivation and vigilant attention. *Annals New York Acad. Sci.* **1129**, 305–322, DOI: 10.1196/annals.1417.002 (2008).

43. Klimesch, W. Alpha-band oscillations, attention, and controlled access to stored information. *Trends Cogn. Sci.* **16**, 606–617 (2012).

44. Lawhern, V. J. *et al.* EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. *J. Neural Eng.* **15**, 056013 (2018).

45. Schirrmeister, R. T. *et al.* Deep learning with convolutional neural networks for eeg decoding and visualization. *Hum. brain mapping* **38**, 5391–5420 (2017).

46. Wang, Y., Duan, W., Dong, D., Ding, L. & Lei, X. A test-retest resting and cognitive state EEG dataset, DOI: 10.18112/openneuro.ds004148.v1.0.1 (2022).

Funding and Acknowledgment

The authors received no funding for this research.

Data Availability

The experimental basis of the research is the OpenNeuro ds004148 dataset[46] and can be accessed via the OpenNeuro repository https://openneuro.org/datasets/ds004148.

Author contributions statement

K.U. conceptualized the study, developed the methodology, and designed the experiments. A.A. performed the data review and validation and wrote the original draft of the manuscript. M.F.A. developed the integration of the LSTM-ODE framework and provided overall research supervision. All authors crucially reviewed, edited, and approved the final manuscript.