A Multimodal Cluster-Aware Learning Framework for Pilot Cognitive-State Recognition Using EEG and Peripheral Physiological Signals

Quynh Anh Nguyen
Faculty of Information Technology, Electric Power University, Hanoi, Vietnam
anhnq@epu.edu.vn (corresponding author)

Nam Anh Dao
Faculty of Information Technology, Electric Power University, Hanoi, Vietnam
anhdn@epu.edu.vn

Received: 17 March 2026 | Revised: 19 April 2026 and 1 May 2026 | Accepted: 3 May 2026

 DOI: https://doi.org/10.48084/etasr.18805

ABSTRACT

Loss of state awareness remains a major contributor to aviation incidents, motivating the need for reliable and interpretable monitoring of pilot cognitive states. This study presents a multimodal cluster-aware learning framework for pilot cognitive-state recognition using Electroencephalography (EEG), Electrocardiography (ECG), Galvanic Skin Response (GSR), and respiration signals. The framework combines modality-aware preprocessing, hybrid dimensionality reduction through principal component analysis and an autoencoder, latent-structure discovery via k-means clustering, and cluster-specific classification using Light Gradient Boosting Machine (LGBM). Experiments were conducted on the public Kaggle benchmark Reducing Commercial Aviation Fatalities, which contains synchronized psychophysiological recordings from 18 pilots under four induced mental states. Under a stratified 5-fold benchmark protocol, the proposed method achieved an overall accuracy of 0.956, outperforming XGBoost, multilayer perceptron, and several representative baselines. The results indicate that multimodal fusion and cluster-aware classification improve the representation of heterogeneous psychophysiological patterns while maintaining computational tractability. Although additional subject-independent validation remains necessary to establish stronger generalization claims, the proposed framework provides a structured and scalable foundation for cognitive-state monitoring in aviation environments.

Keywords-cognitive state detection; multimodal physiological signals; EEG-based monitoring; aviation safety; machine learning; dimensionality reduction; clustering methods

I. INTRODUCTION

Electroencephalography (EEG) is one of the most widely used noninvasive modalities for monitoring human cognitive states because it captures neural activity with high temporal sensitivity [1]. As wearable and semi-portable biosignal acquisition systems continue to mature, physiological monitoring is moving beyond controlled laboratory settings to operational environments such as transportation, defense, industrial control, and aviation. In these contexts, timely detection of changes in attention, workload, or surprise is increasingly important for the design of adaptive human-machine systems.

Aviation is a particularly important application domain because flight crews operate under time pressure, high information density, and safety-critical constraints. Human factors remain a dominant contributor to accidents and incidents, with loss of situational awareness, attentional narrowing, startle responses, and cognitive overload repeatedly identified as risk mechanisms [2, 3]. For this reason, continuous assessment of pilot mental state has become a major topic in aviation neuroergonomics [4, 5]. Reliable cognitive-state monitoring could support early warning systems, adaptive cockpit interfaces, and more resilient human-machine teaming under abnormal operating conditions.

The public Kaggle competition "Reducing Commercial Aviation Fatalities" provides an important benchmark for this problem by offering synchronized psychophysiological recordings from 18 pilots exposed to experimentally induced operational states [4]. The dataset includes four classes: normal/no event (A), channelized attention (CA), diverted attention (DA), and startle/surprise (SS). While EEG is the

primary neural source for inferring cognitive state, peripheral signals such as Electrocardiography (ECG), Galvanic Skin Response (GSR), and respiration reflect autonomic and emotional processes that are also affected by mental workload and unexpected events [5].

Prior studies have demonstrated the promise of both EEG-only and multimodal approaches, but important limitations remain. In [6, 7], multimodal EEG models were based on Riemannian geometry features and later convolutional architectures with SHAP-based interpretation, obtaining strong results but still relying on carefully engineered spectral representations and nontrivial preprocessing. In [8], EEG and GSR were combined to support attention management, but shallow models were limited in their ability to represent complex nonlinear patterns. In [9], deep learning was applied for pilot fatigue detection using EEG rhythms, whereas in [10], a hybrid CNN-LSTM framework was proposed to detect multiple abnormal mental states in flight environments. These studies collectively confirm the value of physiological sensing, but they also highlight persistent challenges related to subject variability, limited data balance, operational noise, and model interpretability [6-10].

Other studies have explored multimodal machine learning for pilot-state recognition. In [11], EEG, ECG, and eye-movement signals were fused using Support Vector Machines (SVMs), demonstrating the value of multimodal integration but also the constraints of fixed-margin decision boundaries when distributions become highly nonlinear. In [12], a CNN-based framework was proposed with improved clustering and semi-supervised learning, but it remained primarily EEG-centric and computationally more intensive. Similar observations arise in related physiological classification tasks, such as sleep-state analysis [13]. More broadly, compact representation learning and dimensionality reduction remain important for high-dimensional biosignal analysis [14]. Related studies have also investigated cognitive-state classification from neural connectivity patterns, highlighting the value of structured brain-signal representations for discriminative learning [15].

The existing literature points to four practical gaps. First, many studies still rely heavily on handcrafted features or EEG-dominant representations. Second, some pipelines involve preprocessing and model configurations that are difficult to translate into more practical settings. Third, the latent structure of multimodal physiological data is rarely explicitly exploited before classification. Fourth, many models remain sensitive to subject variability and overlapping state distributions. These gaps motivate a framework that is computationally efficient, multimodal, and able to adapt to heterogeneous physiological patterns without relying on a single global decision boundary.

To address these gaps, this paper proposes a multimodal cluster-aware learning framework that combines signal filtering, hybrid dimensionality reduction, latent-structure discovery, and cluster-dependent classification. Specifically, Butterworth filtering is used to suppress noise in the ECG and respiration channels; Principal Component Analysis (PCA) and an autoencoder jointly provide linear and nonlinear compact representations; k-means clustering uncovers latent physiological regimes in the learned feature space; and a Light Gradient Boosting Machine (LGBM) classifier is trained to predict mental state under different cluster conditions. The main contribution is not the introduction of a wholly new classifier, but rather the demonstration that multimodal fusion combined with structured representation learning and cluster-aware decision making yields a robust and scalable solution for pilot cognitive-state detection.

II. METHODOLOGY

Figures 1 and 2 illustrate the overall methodology. The proposed framework is organized into five sequential stages, namely multimodal signal acquisition, signal preprocessing, feature extraction and representation learning, multimodal feature fusion with clustering-based partitioning, and cluster-specific hierarchical classification. This structured design was developed to address the main challenges of aviation psychophysiological data analysis, including signal noise, modality heterogeneity, feature redundancy, nonlinear variability, and the existence of complex class boundaries across cognitive states. Instead of adopting a single end-to-end model, the proposed approach decomposes the learning process into a set of interpretable modules, where each stage performs a specific function to improve the stability and discriminative capability of the overall system.

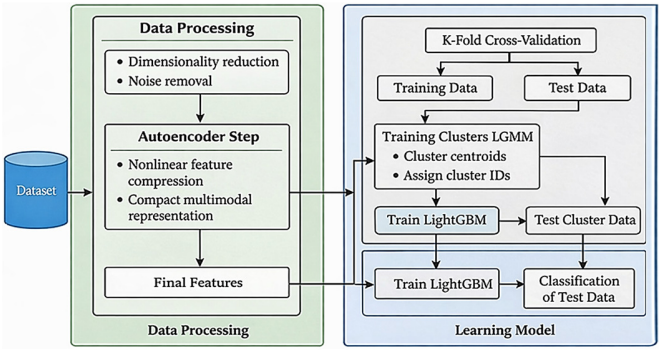

Fig. 1. Workflow of the proposed multimodal cluster-aware classification.

The framework starts with synchronized EEG, ECG, GSR, and respiration signals collected from the pilot-monitoring environment. Among these modalities, EEG provides the principal neural representation of cognitive-state dynamics, as it reflects cortical activity associated with attention, vigilance, mental workload, and transient state changes. In contrast, ECG, GSR, and respiration provide complementary peripheral physiological information related to autonomic regulation and psychophysiological response. Since these signals differ substantially in sampling characteristics, spectral content, and sensitivity to artifacts, the preprocessing stage is designed in a modality-aware manner. This stage includes memory-efficient signal handling, noise reduction, artifact attenuation, temporal alignment, and normalization. For the non-EEG channels, band-limited filtering is applied to suppress high-frequency disturbances while preserving the physiologically meaningful components of the waveform [16]. For multimodal learning, such preprocessing is necessary not only to improve signal quality but also to reduce irrelevant variability before feature construction and fusion.

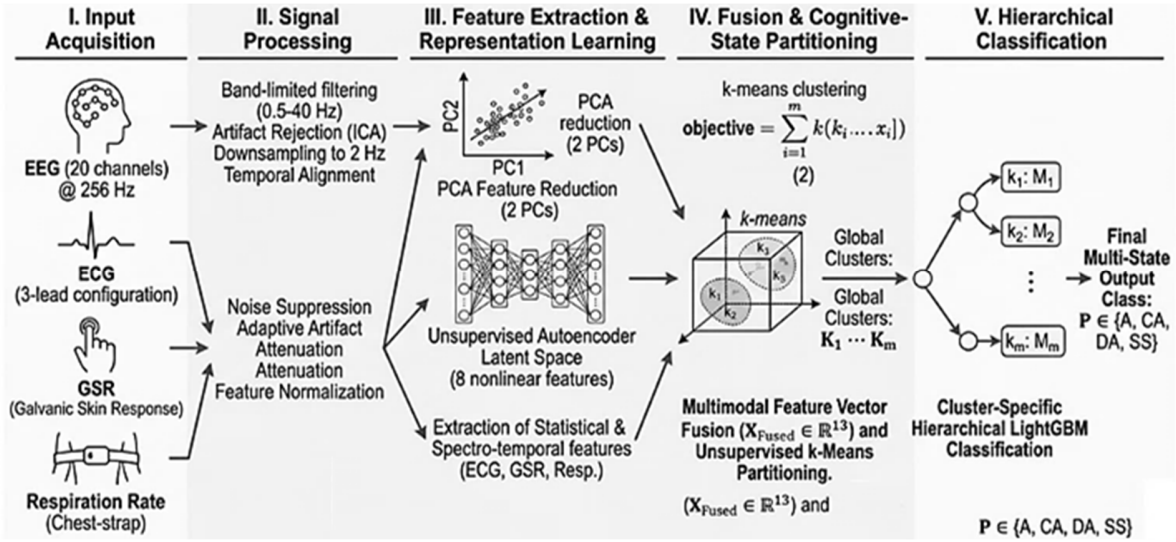

Fig. 2. Proposed multimodal cluster-aware cognitive-state recognition framework.

After preprocessing, the method performs feature extraction and representation learning to obtain a compact and informative multimodal representation. Since EEG contains the largest number of channels and therefore dominates the raw feature space, direct use of the original EEG signals may introduce substantial redundancy and reduce the balance of multimodal fusion. To address this issue, a hybrid representation strategy is employed. First, PCA is used to extract a low-dimensional linear representation that preserves the dominant variance structure of the EEG channels. In parallel, an unsupervised autoencoder is trained to learn nonlinear latent features from the same EEG input. The use of PCA and the autoencoder in parallel allows the framework to retain both global linear structure and more complex nonlinear dependencies within the EEG data. These learned EEG features are then combined with statistical and spectro-temporal descriptors derived from ECG, GSR, and respiration, resulting in a fused multimodal feature vector. In this way, the final representation preserves complementary information from both central and peripheral physiological processes while controlling the dimensionality of the input to the downstream classifier.

Once the fused feature space has been constructed, k-means clustering is applied to partition the training data into local groups with similar multimodal characteristics. The purpose of this step is not to replace the original class labels, but to identify latent substructures in the physiological feature space that may correspond to different local response regimes. In practice, multimodal physiological data often exhibit substantial heterogeneity due to inter-subject variability, sensor noise, and within-state fluctuations. Under such conditions, a single global decision boundary may be insufficient to achieve reliable separation among cognitive states. By introducing clustering prior to classification, the proposed framework attempts to reduce inter-sample heterogeneity and transform the original problem into a set of more locally consistent classification tasks. As shown in Figure 1, cluster centroids are learned from the training data during cross-validation, and each test sample is subsequently assigned to its nearest cluster in the fused feature space.

The final classification stage is based on a cluster-specific hierarchical LightGBM strategy. After cluster assignment, separate LightGBM models are trained using the corresponding cluster-level training subsets, allowing the classifier to learn more localized decision boundaries. During inference, each test sample is first mapped to the nearest centroid and then classified by the associated cluster-specific model. This strategy is motivated by the assumption that physiological patterns belonging to the same local cluster are more homogeneous and therefore easier to discriminate than samples distributed across the entire global space. LGBM was selected as the main classifier because it offers strong performance on structured feature representations, good robustness to moderate nonlinear relationships, and lower computational complexity than deeper sequential architectures when the available dataset is limited. In addition, benchmark models based on XGBoost and Multilayer Perceptron (MLP) were trained on the same feature representation to ensure a fair comparative evaluation. This design allows the contribution of the proposed framework to be interpreted more clearly as the combined effect of multimodal fusion, hybrid representation learning, latent-space partitioning, and cluster-aware classification, rather than because of classifier choice alone.

Overall, the proposed method integrates multimodal physiological sensing, compact feature learning, unsupervised latent-structure discovery, and localized supervised classification into a unified pipeline for pilot cognitive-state recognition. As shown in Figures 1 and 2, the framework is intended to improve both representational efficiency and class discrimination, while maintaining a transparent and modular architecture suitable for experimental analysis and future extension.

A. Data Analysis and Preprocessing

The multimodal dataset contains EEG, ECG, GSR, and respiration recordings collected from multiple pilots. These

signals are first converted into an optimized numerical format to reduce memory usage and computational overhead. Noise is then removed from the ECG and respiration signals using a Butterworth filter, chosen because of its maximally flat passband and its ability to suppress unwanted high-frequency components without distorting waveform morphology. The filter's frequency response is expressed as:

$$H(\omega) = \frac{1}{\sqrt{1+\left(\frac{\omega}{\omega_c}\right)^{2n}}} \quad (1)$$

Filtered signals are produced by convolving the filter with the raw physiological recordings [16]:

$$\widehat{X^{ECG}} = H(\omega) * X^{ECG} \quad (2)$$

$$\widehat{X^{R}} = H(\omega) * X^{R} \quad (3)$$

This preprocessing step ensures that the downstream dimensionality-reduction and learning procedures operate on a cleaner and more reliable representation of the data.

B. Feature Extraction and Dimensionality Reduction

EEG contains many channels and produces high-dimensional input, which can negatively impact model performance. To reduce redundancy while preserving essential information, the proposed method combines PCA with an autoencoder. PCA is first applied by forming the covariance matrix of the EEG feature space:

$$C = \frac{1}{N}(E - \bar{E})(E - \bar{E})^T \quad (4)$$

The principal components are obtained by solving the eigenvalue problem [14]:

$$Cv_k = \lambda_k v_k \quad (5)$$

The reduced EEG representation is projected into a lower-dimensional basis through:

$$Z_{PCA} = V_d^T E \quad (6)$$

In parallel, a nonlinear latent representation is extracted using an autoencoder [17]. The encoder maps the input to a compressed hidden space:

$$h = \sigma(W_e x + b_e) \quad (7)$$

and the decoder reconstructs the signal:

$$\hat{x} = \sigma(W_d h + b_d) \quad (8)$$

The autoencoder is trained by minimizing reconstruction error:

$$\mathcal{L}_{\mathcal{AE}} = \frac{1}{N}\sum_{i=1}^{N} | x_i - \hat{x_i}|^2 \quad (9)$$

The final feature representation used in the training stage is obtained by concatenating PCA components with the autoencoder's latent variables:

$$A = [Z_{PCA}, h] \quad (10)$$

This hybrid dimensionality-reduction strategy preserves both linear and nonlinear structure in the physiological data.

1) Autoencoder Configuration

The autoencoder was implemented as a fully connected feedforward network with a symmetric encoder-decoder structure. The encoder maps the 20-dimensional EEG input to a compact latent representation of dimension 8 through two dense layers. The decoder mirrors this structure to reconstruct the original EEG input from the latent code (see Appendix: Table IX).

In this implementation, the encoder architecture was defined as 20-16-8, and the decoder was defined as 8-16-20. Hidden layers used ReLU activation, whereas the output layer employed linear activation to preserve the continuous nature of the EEG signal. The model was trained by minimizing the mean squared reconstruction error between the input and the reconstructed signal.

The autoencoder was optimized using the Adam optimizer with a learning rate of 0.001, a batch size of 256, and a maximum training budget of 100 epochs. To reduce overfitting and improve stability, early stopping based on validation loss was applied with a patience of 10 epochs. The latent vector produced by the encoder was then concatenated with the PCA features and the non-EEG modalities to form the final fused multimodal feature representation.

This configuration was selected to provide a compact nonlinear embedding while maintaining low computational overhead, which is consistent with the practical objective of the proposed framework.

C. Learning Model

To ensure robust evaluation, k-fold cross-validation is used to partition the dataset into training and test segments. The training features are then clustered using k-Means to reveal latent structure among pilot mental states [18]. The clustering objective minimizes within-cluster variance:

$$\arg\min \sum_{j=1}^{K} \sum_{A_i \in C_j} | A_i - \mu_j|^2 \quad (11)$$

Cluster assignments are computed by selecting the nearest centroid:

$$c_i = \arg\min_j | A_i - \mu_j|^2 \quad (12)$$

Centroids are updated by averaging the points assigned to each cluster:

$$\mu_j = \frac{1}{|C_j|}\sum_{A_i \in C_j} A_i \quad (13)$$

Test samples are assigned to clusters using the same distance-based rule:

$$c_i^{test} = \arg\min_j | A_i^{test} - \mu_j|^2 \quad (14)$$

After clustering, classification is performed using LGBM [19]. The model minimizes an objective composed of prediction loss and regularization:

$$\mathcal{L}_{\mathcal{LGBM}} = \sum l\big(y_i, (A_i)\big) + \sum \Omega(f_t) \quad (15)$$

During boosting, the gradient of the loss is computed for each sample:

$$g_i = \frac{\partial l(y_i, F(A_i))}{\partial F(A_i)} \quad (16)$$

New trees are fitted to these gradients, and LGBM grows trees leaf-wise, selecting the split that maximizes the loss reduction:

$$\Delta\mathcal{L} = \mathcal{L}_{left} + \mathcal{L}_{right} - \mathcal{L}_{parent} \quad (17)$$

The final mental-state prediction for each test example depends on the classifier associated with the cluster to which it was assigned:

$$\hat{y} = F_{c^{test}}(A^{test}) \quad (18)$$

This cluster-aware classification strategy improves robustness by allowing each LGBM model to specialize in a subset of physiological patterns that share a similar structure.

III. RESULTS AND DISCUSSION

This section presents a comprehensive evaluation of the proposed clustering-assisted multimodal cognitive-state detection framework. The analysis covers the dataset characteristics, preprocessing outcomes, model training protocol, and classification performance across multiple baseline algorithms. The section concludes with a comparative assessment against state-of-the-art methods. All experiments were conducted using the data described below, and all results were reproduced under a consistent cross-validation scheme to ensure fairness.

A. Signal Preprocessing

The raw dataset consists of high-resolution synchronized recordings sampled at 256 Hz [4]. To reduce memory overhead and speed up subsequent model training, the signals are converted to 32-bit floating-point arrays. This conversion reduces storage requirements from 1039.79 MB to 241.38 MB, corresponding to an approximately 76.8% reduction. Although this step does not alter information content materially, it improves computational tractability and allows larger training batches to be used. ECG and respiration are especially sensitive to motion noise and other high-frequency disturbances. A Butterworth filter is employed for these channels because of its maximally flat passband and smooth attenuation characteristics. The filtering stage is intended to preserve waveform morphology while reducing signal components that are unlikely to contribute meaningfully to state classification. GSR is kept in the multimodal stack as a low-dimensional autonomic indicator, and EEG channels are carried forward to the dimensionality-reduction stage, where redundancy is addressed more directly.

B. Hybrid Dimensionality Reduction

PCA is first used to extract a compact linear representation of the 20 EEG channels. In this study, two principal components are retained, reducing the raw EEG dimensionality from 20 to 2 while preserving the dominant linear variance structure. However, purely linear compression may not be sufficient for EEG because cognitive-state signatures are often nonlinear and subject-dependent. To complement PCA, an autoencoder is trained to learn an 8-dimensional nonlinear latent code from the same EEG input, as summarized in Table I and illustrated in Figure 3 [16]. The resulting latent space captures information that may not be preserved by variance-based projection alone.

TABLE I. DATASET SUMMARY AND MODALITIES

Property	Description
Number of pilots	18
Number of crews	9
Sampling frequency	256 Hz
Total EEG channels	20 (FP1, FP2, F3, F4, F7, F8, Fz, C3, C4, Cz, P3, P4, Pz, POz, O1, O2, T3, T4, T5, T6)
ECG channels	1
GSR channels	1
Respiration channels	1
Total modalities	4
Mental states	A, CA, DA, SS
Total raw samples	3.12 million (approx.)
Duration of recording	~3.4 hours per modality
Class distribution (approx.)	A: 35%, CA: 25%, DA: 28%, SS: 12%

Note: All statistics in are derived from the public Kaggle benchmark dataset "Reducing Commercial Aviation Fatalities" [4]

The final feature vector is formed by concatenating the PCA representation, the autoencoder latent embedding, and the non-EEG channels (ECG, GSR, and respiration). This yields a compact 13-dimensional multimodal representation. This strategy combines the interpretability and efficiency of PCA with the flexible nonlinear encoding capability of the autoencoder, thereby creating a feature space that is both tractable and expressive for downstream clustering and classification.

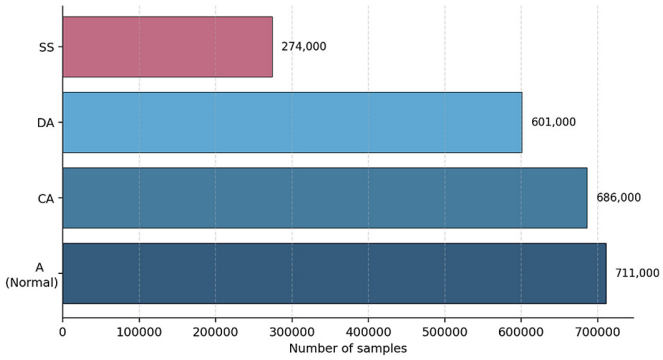

Fig. 3. Class distribution across the four cognitive-state labels in the dataset [4].

C. Cluster-Aware Classification

Once the 13-dimensional multimodal feature vector has been constructed, the training data is further organized through k-means clustering to partition the representation space into K local groups. In the simplest case, $K = 0$, the framework falls back to direct classification without any clustering stage, which serves as the reference baseline. When clustering is enabled, each test sample is first mapped to the nearest centroid learned from the training data, and the corresponding local LGBM classifier is then used to infer the pilot's cognitive state. In this way, the framework does not force all samples to be interpreted through a single global decision surface. Instead, it allows the classification process to adapt to local regions of the

multimodal feature space where physiological patterns may be more internally consistent. The motivation behind this design is grounded in a practical observation: pilot-state data are rarely distributed in a clean, globally uniform manner. Even when two samples share the same cognitive-state label, their physiological manifestations may still differ in subtle but meaningful ways because of individual variability, task context, fatigue level, or autonomic response intensity. In other words, the same nominal state may appear differently across pilots or even across operating conditions for the same pilot. Treating all such observations as if they were drawn from a single homogeneous distribution can make the classification problem unnecessarily rigid and may blur the boundaries that matter most.

The cluster-aware strategy was introduced as a middle ground between two extremes. On one side, a fully global model is attractive because of its simplicity, but it can easily become too coarse to reflect the local structure embedded in multimodal physiological data. On the other side, a strictly subject-specific model may fit individual patterns more closely, yet such a design is difficult to scale and less suitable for broader deployment. By partitioning the latent space into local physiological regimes before classification, the proposed framework attempts to preserve the practicality of a shared model while still allowing a degree of specialization. This balance is especially valuable in aviation applications, where physiological heterogeneity is not noise in the trivial sense, but part of the reality that any reliable monitoring system must learn to accommodate. Table II and Figure 4 show that the framework reduces the feature dimension from 23 raw channels to a compact 13-dimensional multimodal descriptor. This reduction is important for controlling complexity while still preserving complementary information from central and peripheral physiology.

TABLE II. FEATURE DIMENSIONALITY BEFORE AND AFTER

Processing stage	EEG dim.	ECG	GSR	Resp.	Total features
Raw input	20	1	1	1	23
After PCA	2	1	1	1	5
Autoencoder latent dimensions	8	–	–	–	8
Final concatenated vector	–	–	–	–	13

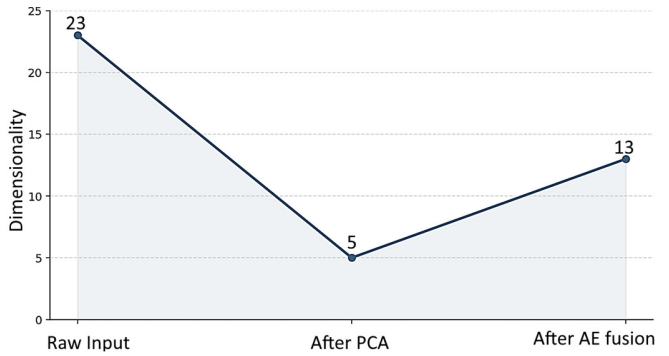

Fig. 4. Dimensionality transition from raw multimodal input to the compact multimodal representation.

D. Experimental Protocol

The proposed multimodal cluster-aware framework was evaluated on the public Kaggle benchmark "Reducing Commercial Aviation Fatalities" [4], which contains synchronized multimodal physiological recordings from 18 pilots under four induced cognitive-state conditions. The evaluation was designed not only to measure overall classification accuracy, but also to understand how the proposed framework behaves under class imbalance, how stable it remains across folds, and whether the selected clustering structure is supported by an independent quality criterion. In this sense, the experimental design was intended to move beyond a single summary metric and provide a more faithful picture of the strengths and limitations of the framework.

The experiments were conducted using a stratified 5-fold cross-validation protocol so that each fold preserved the class proportions of the four target cognitive states: normal/no event (A), Channelized Attention (CA), Diverted Attention (DA), and Startle/Surprise (SS). In each iteration, four folds were used for training, and the remaining fold was used for testing. This protocol was adopted to provide a benchmark-level comparison under consistent class distribution. However, it should be acknowledged that this evaluation does not fully guarantee subject independence, since the dataset contains recordings from only 18 pilots, and samples from the same subject may still appear across different folds. For this reason, the reported results should be interpreted primarily as evidence of benchmark effectiveness rather than as definitive proof of cross-subject generalization.

To ensure fairness, all candidate classifiers were trained and evaluated on the same fused multimodal representation**.** This representation consisted of PCA-compressed EEG features, the nonlinear EEG latent embedding learned by the autoencoder, and the peripheral physiological descriptors derived from ECG, GSR, and respiration. On this common feature space, three classifiers were compared: LGBM**,** XGBoost**,** and MLP. In addition, several representative literature-inspired baselines were reconstructed and evaluated on the same benchmark to contextualize the obtained performance relative to existing approaches. These baseline models should be regarded as representative reimplementations inspired by prior work rather than exact replications, since the original studies differ in preprocessing assumptions, input modalities, and model configuration.

The first part of the evaluation focused on the effect of latent-space partitioning**.** The number of clusters K was varied systematically, with $K = 0$ corresponding to the no-clustering baseline. In the clustered setting, k-means was fitted only on the training representation of each fold, and the test samples were subsequently assigned to the nearest learned centroids before classification. This design allowed the framework to examine whether the multimodal latent space benefits from local specialization prior to supervised learning. The selected value $K = 8$ was retained as the most effective empirical configuration because it yielded the highest downstream LGBM classification accuracy under the present protocol.

However, classification accuracy alone is not sufficient to justify the selection of the clustering configuration. Therefore, an additional independent clustering-quality analysis was performed using the silhouette coefficient**.** The results in Table III show that the fused multimodal representation exhibits meaningful local structure across all tested values of K, with silhouette scores ranging from 0.412 to 0.452. Although the highest silhouette score was observed at $K = 7$, the best downstream LGBM accuracy was achieved at $K = 8$. This difference is not unexpected, because cluster compactness and supervised discriminative performance do not necessarily attain their optima at the same operating point. Thus, $K = 8$ was retained as the most effective classification-oriented setting, while the silhouette analysis provided additional evidence that the selected cluster structure reflects a nontrivial representation of the multimodal feature space.

TABLE III. SILHOUETTE-BASED CLUSTERING QUALITY AND DOWNSTREAM LIGHTGBM ACCURACY ACROSS CANDIDATE VALUES OF K

Number of clusters (K)	Silhouette score	LGBM accuracy
3	0.412	0.945
4	0.436	0.950
5	0.448	0.952
6	0.441	0.949
7	0.452	0.954
8	**0.447**	**0.956**

To obtain a more complete view of model behavior, the evaluation was extended beyond overall accuracy. The revised protocol also considered class-wise precision, recall, and F1-score**,** together with macro-F1 and balanced accuracy**.** This extension is especially important in this benchmark because the SS class is underrepresented, but is operationally more critical than its frequency in the data might suggest.

Table IV presents the class-wise results at the best clustering setting ($K = 8$). The proposed framework achieves strong and relatively balanced discrimination for the major classes A, CA, and DA, while the SS class remains comparatively more difficult, which is consistent with its lower representation and greater physiological overlap with adjacent states. Nevertheless, the obtained macro-F1 of 0.942 and balanced accuracy of 0.939 indicate that the reported improvement is not driven solely by majority-class dominance.

TABLE IV. CLASS-WISE PRECISION, RECALL, AND F1-SCORE OF THE PROPOSED MODEL AT THE BEST CLUSTERING SETTING

Class	Precision	Recall	F1-score	Support (%)
A	0.968	0.971	0.969	35
CA	0.951	0.944	0.947	25
DA	0.948	0.950	0.949	28
SS	0.915	0.892	0.903	12
Macro average	**0.946**	**0.939**	**0.942**	—
Weighted average	**0.955**	**0.956**	**0.955**	—

To further improve the reliability of reporting, Table V summarizes fold-wise results using mean ± standard deviation. At the best clustering setting ($K = 8$), the proposed framework achieved an accuracy of 0.956 ± 0.007, precision of 0.955 ± 0.008, F1-score of 0.953 ± 0.009, macro-F1 of 0.942 ± 0.011, and balanced accuracy of 0.939 ± 0.012. The relatively small standard deviations suggest that the framework remains stable under the adopted benchmark protocol and that the reported gains are not the result of a single particularly favorable partition.

TABLE V. AGGREGATE PERFORMANCE OF THE PROPOSED FRAMEWORK ACROSS FOLDS AT THE BEST CLUSTERING SETTING

Metric	Mean ± Stdev
Accuracy	0.956 ± 0.007
Precision	0.955 ± 0.008
F1-score	0.953 ± 0.009
Macro-F1	0.942 ± 0.011
Balanced Accuracy	0.939 ± 0.012

1) Effect of Clustering

Figure 5 shows a clear benefit from cluster-aware learning. The baseline model without clustering ($K = 0$) yields an accuracy of 0.923. Introducing latent clustering improves the result consistently, with the best performance of 0.956 is obtained at $K = 8$. The trend suggests that the multimodal feature space contains meaningful internal structures that can be exploited before classification. The slight non-monotonic variation between $K = 5$ and 6 also indicates that clustering is useful only when the partition granularity is appropriate; overly coarse or suboptimal partitions may weaken the benefit.

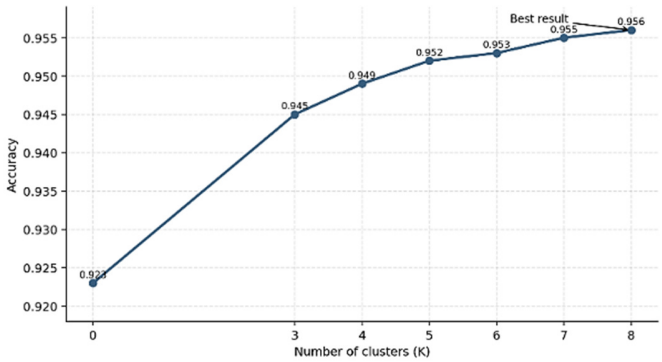

Fig. 5. Variation in LGBM accuracy as a function of the number of latent clusters K.

2) Comparison Across Classifiers

As reported in Table VI and Figure 6, LGBM provides the strongest and most stable performance across clustering settings. XGBoost performs competitively in some cases and even slightly exceeds LightGBM when $K = 0$, but it does not benefit from clustering as consistently. The MLP model remains markedly weaker throughout, suggesting that a relatively simple feedforward architecture is less suited to the compact multimodal representation used here, especially under heterogeneous class structure.

Table VII and Figure 7 show that the accuracy advantage of LGBM is accompanied by strong precision and F1-score at the best clustering level. This is important because it suggests that the improvement is not driven solely by majority-class bias.

Nevertheless, since the dataset is not perfectly balanced, future studies should also report per-class recall, macro-F1, and confusion matrices to better assess performance on the minority SS state.

TABLE VI. CLASSIFICATION ACCURACY OF LIGHTGBM, XGBOOST, AND MLP ACROSS K

K	LightGBM	XGBoost	MLP
0	0.923	0.935	0.814
3	0.945	0.901	0.830
4	0.950	0.909	0.847
5	0.952	0.931	0.823
6	0.949	0.927	0.835
7	0.954	0.942	0.863
8	0.956	0.936	0.870

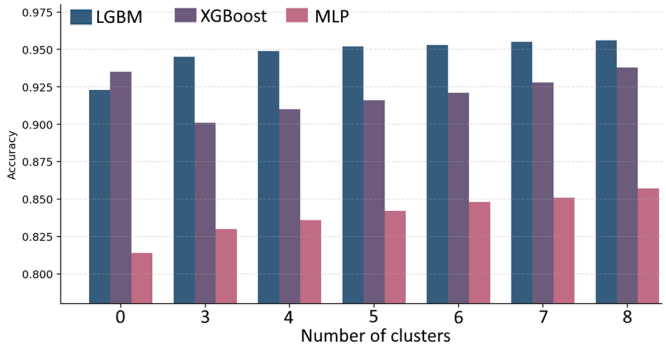

Fig. 6. Comparative classifier accuracy across clustering configurations.

TABLE VII. PRECISION AND F1-SCORE FOR THE THREE CLASSIFIERS AT $K = 8$

Metric	LightGBM	XGBoost	MLP
Precision	0.956	0.933	0.821
F1-score	0.953	0.936	0.842

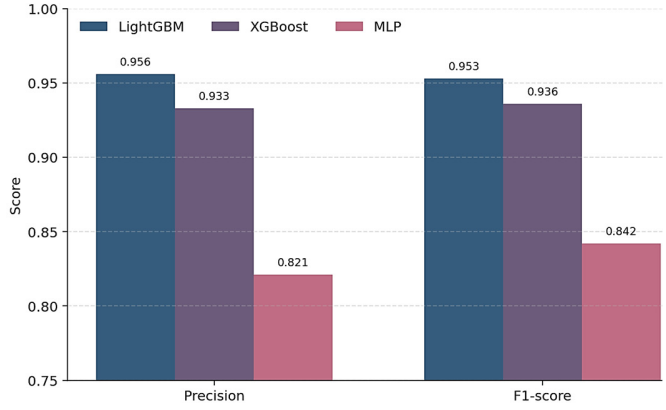

Fig. 7. Precision and F1-score comparison at the best clustering setting ($K = 8$).

Figures 8 and 9 illustrate the class-wise prediction behavior of the proposed four-state classifier in terms of absolute counts and normalized percentages. The dominant diagonal structure confirms that most samples are classified correctly, while the remaining errors are concentrated mainly between cognitively adjacent states. This is consistent with the class-wise analysis in Table IV, where the minority startle/surprise state exhibits lower recall than the majority categories. The confusion matrices provide a more informative view of the model behavior than overall accuracy alone and further support the interpretation that the framework achieves strong but not perfectly uniform discrimination across all cognitive states.

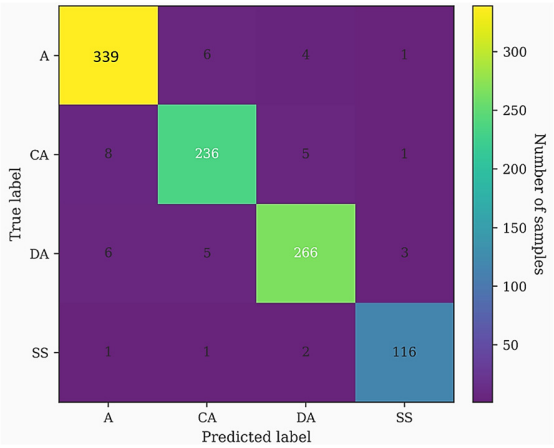

Fig. 8. Confusion matrix of the proposed four-state cognitive-state classifier.

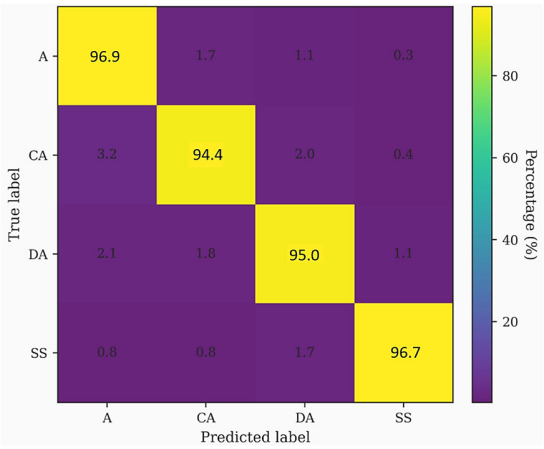

Fig. 9. Normalized confusion matrix of the proposed four-state cognitive-state classifier (%).

3) Comparison with Representative Literature-Inspired Baselines

It should be noted that the original studies considered in the comparative analysis were not all conducted on the same benchmark dataset. Some were developed on aviation-related EEG or multimodal physiological datasets with different subject pools, modality configurations, and preprocessing pipelines [8-12]. The models reported here should be interpreted as literature-inspired reconstructions evaluated on the current Kaggle benchmark rather than as direct reproductions of the original published results. This clarification is important for ensuring a fair and transparent comparison under a unified data setting.

To contextualize the obtained results, four representative models were reconstructed and evaluated under the same data

conditions. Table VIII shows that the psychophysiological model adapted from [8] reaches an accuracy of 0.842, indicating the limitations of low-capacity classifiers for highly variable physiological data. The CNN-LSTM model adapted from [10] improves accuracy to 0.906, but its performance remains below the proposed framework, likely because deep temporal models typically benefit from larger and more balanced datasets than the present benchmark provides. The multimodal SVM model adapted from [11] achieves 0.933 accuracy, demonstrating the value of multimodal integration but also suggesting that fixed-margin classifiers are less flexible than cluster-aware boosting when the feature space is heterogeneous. The CNN-based model adapted from [12] attains 0.914 accuracy, which confirms the usefulness of spatial EEG learning but also highlights the added value of including peripheral physiological channels.

TABLE VIII. PERFORMANCE COMPARISON BETWEEN THE PROPOSED FRAMEWORK AND RECONSTRUCTED LITERATURE-INSPIRED BASELINE MODELS

	Model	Study	Original study dataset	Modality	Accuracy
1	Psychophysiological sensing + shallow classifier	[8]	Aviation psychophysiological experimental dataset	EEG + GSR	0.842
2	Hybrid CNN-LSTM neural network	[10]	Flight-environment EEG abnormal mental-state dataset	EEG	0.906
3	Multimodal SVM fusion	[11]	Pilot behavior multimodal dataset with EEG, ECG, and eye movements	EEG + ECG + Eye	0.933
4	CNN spatial EEG classifier	[12]	Pilot mental-state EEG dataset	EEG	0.914
5	Proposed clustering + LightGBM model	This work	Kaggle "Reducing Commercial Aviation Fatalities" [4]	EEG + ECG + GSR + Resp.	0.956

E. Discussion

The reported results reinforce the importance of multimodal physiological fusion for pilot cognitive-state recognition. While EEG remains the dominant source of neural information for representing attentional and cognitive dynamics, the addition of ECG, GSR, and respiration provides complementary descriptors of autonomic activation and physiological regulation. The consistent gain over EEG-dominant baselines indicates that pilot cognitive state is more appropriately modeled as an integrated neurophysiological construct rather than as a purely cortical phenomenon. This interpretation is well aligned with the neuroergonomics literature, where workload modulation, attentional narrowing, diverted attention, and startle-related responses are typically manifested through coupled central and peripheral processes. In that sense, the results of this study support the view that physiologically complementary sensing is not only beneficial but structurally necessary for more accurate state representation [5, 8].

Second, the proposed hybrid compression strategy appears to be well-suited to the statistical characteristics of high-frequency biosignal data. PCA offers an efficient summary of dominant linear variance, whereas the autoencoder captures nonlinear structures that would likely be suppressed under purely linear projection. Their joint use yields a compact and expressive latent space that supports effective downstream learning without resorting to full end-to-end deep temporal architectures. This is an important practical consideration. In aviation psychophysiology, the data are often high-dimensional and heterogeneous, whereas the number of subjects remains modest. Under such constraints, overly deep models may be difficult to justify in terms of stability, interpretability, and computational burden. The proposed framework adopts a more disciplined engineering compromise by preserving informative structures while maintaining tractable model complexity.

Third, the results show that latent-space partitioning is central to the framework, rather than a peripheral preprocessing choice. The improvement from 0.923 without clustering to 0.956 at $K = 8$ indicates that the fused multimodal representation contains exploitable local structure. This observation has methodological significance. It suggests that pilot-state data are not optimally described by a single global decision boundary but instead occupy partially separable local regions shaped by heterogeneous physiological regimes, subject-specific response patterns, and task-driven variability. By introducing a modest unsupervised partitioning stage before classification, the proposed method enables localized specialization and thereby improves robustness. Accordingly, the main contribution of this study is not simply an accuracy increment, but the demonstration that cluster-aware multimodal learning is an effective strategy for structured cognitive-state inference in aviation biosignals.

It is also worth noting that the performance gain should be attributed to the framework, rather than to the classifier alone. Comparative experiments indicate that the benefit arises from the combined effect of multimodal fusion, hybrid representation learning, and cluster-based specialization, with LGBM providing the most effective final decision model within that design. This distinction clarifies the source of novelty. The contribution lies not in the isolated use of a particular learner, but in a pipeline that is better matched to the geometry and heterogeneity of the underlying physiological feature space.

However, these findings should be interpreted with suitable caution. The evaluation is conducted on a public benchmark involving 18 pilots under a stratified 5-fold protocol. Although such a setting is appropriate for comparative benchmarking, it does not fully resolve the broader issues of cross-subject generalization, class imbalance, and deployment realism. Aggregate performance under fold-based evaluation may still reflect latent subject-specific regularities that are less informative under subject-independent conditions. For this reason, this study should be read primarily as evidence of strong benchmark-level efficacy, not as definitive proof of operational readiness in cockpit environments.

From a systems perspective, the proposed framework exhibits several attractive properties. It is lighter than many deep temporal alternatives, explicitly encodes multimodal complementarity, and retains a modular architecture that can be extended to other safety-critical monitoring problems involving heterogeneous biosignals. These features make the method

technically appealing for embedded or near-real-time human-state monitoring scenarios. However, practical aviation deployment would require substantially broader validation, including assessment of sensor robustness, artifact tolerance, synchronization reliability, latency constraints, human-factors acceptance, and the operational cost of false alarms. Thus, this work should be regarded a technically grounded step toward operational cognitive-state monitoring, rather than as a finished deployment-ready solution.

IV. CONCLUSION

This paper presented a multimodal cluster-aware learning framework for pilot cognitive-state recognition using EEG, ECG, GSR, and respiration signals. The proposed approach combines modality-aware preprocessing, hybrid feature compression through PCA and autoencoder-based representation learning, latent-space partitioning with k-means, and cluster-specific classification using LGBM. Under the current benchmark protocol on a public aviation dataset [4], the framework achieved a best overall accuracy of 0.956 and consistently outperformed XGBoost, MLP, and several representative reconstructed baselines.

Beyond the reported accuracy, the main value of this work lies in showing that multimodal physiological fusion and cluster-aware classification can work together in a practical and computationally efficient way. This combination helped the model handle heterogeneous psychophysiological patterns more effectively than a single global classification strategy. The additional class-wise and clustering-oriented analyses further support the view that the gain is not simply a consequence of classifier choice but is related to how the multimodal feature space is represented, organized, and used for local decision making. This is, in our view, the most meaningful contribution of the study.

At the same time, the results should be interpreted with appropriate caution. The evaluation was conducted under a stratified 5-fold benchmark setting on a dataset containing recordings from 18 pilots and therefore does not yet provide a fully subject-independent assessment of generalization. Thus, these findings should be regarded as strong benchmark-level evidence rather than as a final demonstration of deployment readiness in operational cockpit environments. In addition, although the revised analysis strengthens reporting through class-wise evaluation and clustering-quality assessment, further work is still needed to establish subject-independent robustness, statistical reliability, and real-time operational viability.

Overall, this study suggests that a structured multimodal learning pipeline can offer a realistic and technically grounded path for cognitive-state monitoring in aviation. The framework is not presented here as a finished operational solution, but rather as a solid step toward more reliable human-centered monitoring systems. With stronger subject-independent validation, richer statistical analysis, and broader testing under operationally realistic conditions, the proposed approach could provide a useful foundation for next-generation aviation safety support systems.

APPENDIX

The nonlinear EEG representation was learned using a fully connected autoencoder with a symmetric encoder-decoder structure. As summarized in Table IX, the encoder compresses the 20-dimensional EEG input into an 8-dimensional latent representation through an intermediate hidden layer of size 16, while the decoder reconstructs the original signal through reverse mapping. The model was trained using mean squared reconstruction loss with the Adam optimizer, a learning rate of 0.001, and a batch size of 256. Early stopping with a patience of 10 epochs was applied based on validation loss to improve training stability and reduce overfitting.

TABLE IX. HYPERPARAMETER CONFIGURATION OF THE PROPOSED AUTOENCODER

Hyperparameter	**Value**
Input dimension	20
Encoder architecture	20–16–8
Latent dimension	8
Decoder architecture	8–16–20
Hidden-layer activation	ReLU
Output activation	Linear
Loss function	Mean Squared Error (MSE)
Optimizer	Adam
Learning rate	0.001
Batch size	256
Maximum epochs	100
Early stopping	Yes
Early stopping patience	10 epochs
Validation criterion	Validation loss
Weight initialization	Default Xavier/Glorot initialization
Regularization	None/default implementation
Training objective	Reconstruction of EEG input

FUNDING

The authors received no external funding for this study.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENT

Not applicable to this work.

DATA AVAILABILITY

The dataset used in this study is publicly available in [4].

REFERENCES

[1] F. L. Da Silva, "EEG: Origin and Measurement," in *EEG - fMRI*, C. Mulert and L. Lemieux, Eds. Springer Berlin Heidelberg, 2009, pp. 19–38.

[2] M. Martinussen and D. R. Hunter, *Aviation Psychology and Human Factors*, 2nd ed. Taylor & Francis, CRC Press, 2017.

[3] D. Kelly and M. Efthymiou, "An analysis of human factors in fifty controlled flight into terrain aviation accidents from 2007 to 2017," *Journal of Safety Research*, vol. 69, pp. 155–165, June 2019, https://doi.org/10.1016/j.jsr.2019.03.009.

[4] B. A. Hamilton, "Reducing Commercial Aviation Fatalities." https://kaggle.com/reducing-commercial-aviation-fatalities.

[5] E. Van Weelden, M. Alimardani, T. J. Wiltshire, and M. M. Louwerse, "Aviation and neurophysiology: A systematic review," *Applied*

Ergonomics, vol. 105, Nov. 2022, Art. no. 103838, https://doi.org/10.1016/j.apergo.2022.103838.

[6] I. Alreshidi, I. Moulitsas, and K. W. Jenkins, "Multimodal Approach for Pilot Mental State Detection Based on EEG," *Sensors*, vol. 23, no. 17, Aug. 2023, Art. no. 7350, https://doi.org/10.3390/s23177350.

[7] I. Alreshidi, D. Bisandu, and I. Moulitsas, "Illuminating the Neural Landscape of Pilot Mental States: A Convolutional Neural Network Approach with Shapley Additive Explanations Interpretability," *Sensors*, vol. 23, no. 22, Nov. 2023, Art. no. 9052, https://doi.org/10.3390/s23229052.

[8] A. R. Harrivel, C. Liles, C. L. Stephens, K. K. Ellis, L. J. Prinzel, and A. T. Pope, "Psychophysiological Sensing and State Classification for Attention Management in Commercial Aviation," in *AIAA Infotech @ Aerospace*, Jan. 2016, https://doi.org/10.2514/6.2016-1490.

[9] E. Q. Wu *et al.*, "Detecting Fatigue Status of Pilots Based on Deep Learning Network Using EEG Signals," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 13, no. 3, pp. 575–585, Sept. 2021, https://doi.org/10.1109/TCDS.2019.2963476.

[10] D. H. Lee, J. H. Jeong, B. W. Yu, T. E. Kam, and S. W. Lee, "Autonomous System for EEG-Based Multiple Abnormal Mental States Classification Using Hybrid Deep Neural Networks Under Flight Environment," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 10, pp. 6426–6437, Oct. 2023, https://doi.org/10.1109/TSMC.2023.3282635.

[11] X. Wang, G. Gong, N. Li, L. Ding, and Y. Ma, "Decoding pilot behavior consciousness of EEG, ECG, eye movements via an SVM machine learning model," *International Journal of Modeling, Simulation, and Scientific Computing*, vol. 11, no. 04, Aug. 2020, Art. no. 2050028, https://doi.org/10.1142/S1793962320500282.

[12] Q. Wang, Z. Wang, R. Xiong, X. Liao, and X. Tan, "A Method for Classification and Evaluation of Pilot's Mental States Based on CNN," *Computer Systems Science and Engineering*, vol. 46, no. 2, pp. 1999–2020, 2023, https://doi.org/10.32604/csse.2023.034183.

[13] M. J. Hasan, D. Shon, K. Im, H.-K. Choi, D.-S. Yoo, and J.-M. Kim, "Sleep State Classification Using Power Spectral Density and Residual Neural Network with Multichannel EEG Signals," *Applied Sciences*, vol. 10, no. 21, Oct. 2020, Art. no. 7639, https://doi.org/10.3390/app10217639.

[14] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, Apr. 2016, Art. no. 20150202, https://doi.org/10.1098/rsta.2015.0202.

[15] S. R. Jeevakala and H. Ramasangu, "Classification of Cognitive States using Task-Specific Connectivity Features," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10675–10679, June 2023, https://doi.org/10.48084/etasr.5836.

[16] Y. Jia *et al.*, "Preprocessing and Denoising Techniques for Electrocardiography and Magnetocardiography: A Review," *Bioengineering*, vol. 11, no. 11, Nov. 2024, Art. no. 1109, https://doi.org/10.3390/bioengineering11111109.

[17] K. Berahmand, F. Daneshfar, E. S. Salehi, Y. Li, and Y. Xu, "Autoencoders and their applications in machine learning: a survey," *Artificial Intelligence Review*, vol. 57, no. 2, Feb. 2024, Art. no. 28, https://doi.org/10.1007/s10462-023-10662-6.

[18] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982, https://doi.org/10.1109/TIT.1982.1056489.

[19] G. Ke *et al.*, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," in *Advances in Neural Information Processing Systems*, 2017, vol. 30.