Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 25, MARCH 2026

DOI: **10.37943/25EJPK6829**

Anar Mengdigali
Researcher, Master of Technical Science, School of Artificial Intelligence and Data Science
242886@astanait.edu.kz, orcid.org/0009-0005-0823-5909
Astana IT University, Kazakhstan
Temirlan Karibekov
Director, Doctor of Medical Sciences, Science and Innovation Center "MedTech"
T.Karibekov@astanait.edu.kz, orcid.org/0009-0008-9801-1774
Astana IT University, Kazakhstan
Medet Mukushev
Assistant professor, PhD, School of Creative Industry
m.mukusev@astanait.edu.kz, orcid.org/0000-0002-3655-9928
Astana IT University, Kazakhstan
Manzura Zholdasova
Associate professor, PhD, Department of Biophysics, Biomedicine, and Neuroscience, Brain Institute
manzur777@gmail.com, orcid.org/0000-0002-8186-9650
Al-Farabi Kazakh National University, Kazakhstan
Diana Arman
Assistant Professor, PhD, School of Information Technology and Engineering
d.arman@kbtu.kz, orcid.org/0000-0003-4259-9296
Kazakh-British Technical University, Kazakhstan
Almira Kustubayeva
Professor, Head of the Department, Candidate of Biological Sciences
Department of Biophysics, Biomedicine, and Neuroscience, Brain Institute
almkusto@kaznu.kz, orcid.org/0000-0001-6575-6288
Al-Farabi Kazakh National University, Kazakhstan

CROSS-SUBJECT EEG-BASED FATIGUE CLASSIFICATION USING MACHINE LEARNING, RIEMANNIAN GEOMETRY, AND COMPACT DEEP NEURAL NETWORKS

Abstract: Drowsiness reduces efficiency in perceptual processing, reaction time, and executive control, posing risks in safety-critical domains such as driving and long-duration monitoring tasks. EEG-based fatigue detection has emerged as a powerful approach for quantifying early neurophysiological signs of vigilance decline, yet many proposed algorithms are insufficiently evaluated in strictly subject-independent conditions. To address this gap, we systematically compare classical machine learning models, Riemannian geometry-based classification, and compact deep neural architectures on a publicly available electroencephalography (EEG) dataset containing 11 subjects. We employ a rigorous leave-one-subject-out (LOSO) protocol, ensuring that no individual contributes information simultaneously to the training and test sets.

The study evaluates logistic regression, support vector machines with radial-basis kernels, random forests, a Log-Euclidean Riemannian classifier, EEGNet, a transformer encoder, and a bidirectional long short-term memory (BiLSTM) with temporal attention. Across folds, accuracy and macro-F1 scores were calculated and summarized with mean and standard deviation. The BiLSTM-attention model achieved the highest performance (accuracy $74.00\% \pm 11.31$; macro-F1 $73.03\% \pm 12.25$) but only moderately exceeded EEGNet and the classical baselines. Wilcoxon signed-rank tests revealed no significant difference between EEGNet and BiLSTM ($p = 0.70$), although BiLSTM significantly outperformed the transformer model ($p = 0.039$). Analysis of error structure demonstrated a notable asymmetry with 295 false positives and 184 false negatives aggregated across folds.

Received: 15.02.2025 Accepted: 25.02.2026 Published: 30.03.2026

DOI: 10.37943/25EJPK6829

Band-specific analysis revealed theta activity as the strongest contributor to class separation, followed by delta and alpha rhythms. Channel-importance analysis indicated that posterior and paracentral regions were consistently more informative. These findings highlight that model complexity does not guarantee superior performance in small datasets with large inter-subject variability. The study provides a transparent, fully reproducible baseline for future fatigue-classification research and demonstrates the practical relevance of compact architectures and Riemannian geometry in low-data conditions.

Keywords: electroencephalogram; fatigue detection; drowsiness; deep learning; Riemannian geometry.

Introduction

Electroencephalography has long been recognized as one of the most sensitive modalities for detecting fluctuations in vigilance. Unlike behavioral or ocular indicators, EEG captures the underlying neural dynamics associated with fatigue, including shifts in oscillatory power and synchronization across cortical networks. As individuals transition from alert wakefulness into drowsiness, characteristic increases in theta and delta activity and reductions in alpha power emerge, forming physiologically interpretable markers of declining attentional stability. Identifying these transitions reliably is critical for applications in driver monitoring, aviation, industrial safety, and clinical care.

Although recent developments in machine learning and deep neural networks have produced promising models for EEG classification, many studies rely on within-subject or random-split validation protocols. Such methods allow overlapping characteristics of the same individual to appear in both training and testing data, thereby inflating performance estimates. Subject-independent evaluation, particularly leave-one-subject-out (LOSO) cross-validation, provides a much more stringent test of a model's ability to generalize to unseen individuals. However, strict LOSO evaluation remains challenging because EEG varies substantially across subjects, and small datasets limit the capacity of deep architectures to learn robust spatiotemporal patterns [1, 2].

This work addresses these challenges by presenting a systematic LOSO benchmark comparing classical machine learning models, Riemannian geometry methods, and compact deep neural networks. The goal is not merely to identify the best-performing model, but to analyze the strengths and weaknesses of each methodological family under conditions that reflect realistic deployment scenarios. We incorporate detailed statistical analysis, error pattern assessment, and feature-importance inspection to produce a comprehensive and interpretable characterization of model behavior.

Literature Review and Problem Statement

Recent advances in drowsiness detection research have emphasized the importance of robust subject-independent performance. Studies published between 2020 and 2025 demonstrate a shift from traditional handcrafted features toward hybrid methods that integrate physiological priors, geometric representations, and deep learning models optimized for limited data [3, 4]. Classical approaches often utilize spectral measures such as theta-to-alpha ratios, bandpower changes, or entropy-based features, and they continue to serve as strong baselines due to their interpretability and data efficiency [5].

Riemannian geometry has gained substantial attention as covariance matrices provide reliable representations of EEG activity. Log-Euclidean mapping offers an efficient strategy for embedding covariance matrices into vector space while preserving essential structure [6, 7]. Multiple studies report that such methods outperform deep learning baselines on small datasets, particularly under LOSO evaluation. This robustness arises from the geometric invariance of covariance structures and the relative insensitivity to channel-level noise.

Deep learning approaches have diversified in recent years. Compact Convolutional Neural Networks (CNN) such as EEGNet aim to integrate spatial filtering and temporal convolution in a lightweight architecture, making them compatible with small datasets [8]. Meanwhile, recurrent neural networks, including Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) architectures, model temporal dependencies explicitly [9]. Several recent works apply attention mechanisms to highlight critical temporal or spatial components [10, 11]. Transformers, while demonstrating remarkable performance in natural

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 25, MARCH 2026

language and vision tasks, often overfit when applied to limited EEG datasets due to their high parameter count [12].

Across studies, a consistent observation is that subject-independent performance remains considerably lower than within-subject accuracy, reflecting the challenge of capturing individual variability in neural signatures. LOSO evaluation is increasingly recognized as the most reliable protocol for assessing generalization [13]. The present work situates itself within this movement, aiming to provide a standardized comparison across multiple methodological families using a consistent and transparent pipeline.

Aim and Objectives of the Study

The primary aim of this study is to systematically benchmark multiple model families, classical machine learning, Riemannian geometry-based classifiers, and compact deep neural networks for cross-subject EEG-based fatigue classification under a strictly subject-independent evaluation protocol.

The specific objectives are:

- to implement and evaluate seven model families under leave-one-subject-out cross-validation;
- to perform statistical significance testing of pairwise model differences using the Wilcoxon signed-rank test;
- to analyze error asymmetry and its implications for safety-critical deployment;
- to identify the most informative EEG frequency bands and electrode regions for fatigue discrimination.

Methods and Materials

Dataset

We used the publicly available EEG Driver Drowsiness Dataset [14], which contains EEG recordings from 11 healthy adult participants during a simulated driving or sustained vigilance task. The data are distributed as pre-segmented three-second epochs stored in a MATLAB file (dataset.mat). Each epoch is represented as a multichannel time series of shape (30×384), where 30 denotes the number of EEG channels and 384 the number of samples per epoch. With a sampling rate of 128 Hz, this corresponds to an epoch duration of exactly 3 s.

In total, the dataset contains 2,022 epochs, balanced across classes as 1,011 alert and 1,011 drowsy segments. The distribution across subjects is moderately imbalanced: the eleven participants contribute 188, 132, 150, 148, 224, 166, 102, 264, 314, 108, and 226 epochs, respectively. The MATLAB file provides three variables: an array EEGsample of shape $(2022,30,384)$, a binary label vector substate (0 for alert, 1 for drowsy), and a subject index vector subindex.

Preprocessing

Preprocessing was performed using MNE-Python and SciPy. All signals were re-referenced to the common average reference. A 4th-order Butterworth band-pass filter between 1 and 50 Hz was applied using zero-phase filtering to preserve the canonical EEG bands relevant for vigilance. Residual power-line interference at 50 Hz was attenuated using a 2nd-order IIR notch filter. Finally, subject-wise standardization was performed in the time domain. For each subject s and channel c, the normalized signal is:

$$\tilde{x}_{s,c}(t) = \frac{x_{s,c}(t) - \mu_{s,c}}{\sigma_{s,c} + 10^{-6}}, \qquad (1)$$

where $\mu_{s,c}$ and $\sigma_{s,c}$ are the mean and standard deviation computed across all epochs and time samples for that subject–channel pair. This yielded a time-domain tensor $X_{\text{time}} \in \mathbb{R}^{2022\times30\times384}$ with global mean approximately zero and unit standard deviation. All 2,022 epochs were retained to preserve statistical power.

Feature Extraction

Classical machine learning and Riemannian classifiers operated on spectral band-power and covariance features, respectively, while deep models received the normalized multichannel time series directly.

DOI: 10.37943/25EJPK6829

Spectral band-power features. For each epoch and channel, the power spectral density (PSD) was estimated using Welch's method with a Hamming window of 256 samples, 50% overlap, and sampling frequency 128 Hz. Band-power features were computed by averaging log-PSD values within the delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), and beta (13–30 Hz) bands:

$$\text{Bandpower}(f_1, f_2) = \frac{1}{|\mathcal{F}|} \sum_{f \in \mathcal{F}} \log P(f), \quad \mathcal{F} = \{f : f_1 \le f \le f_2\}. \qquad (2)$$

This produced a 120-dimensional feature vector per epoch ($30 \text{ channels} \times 4 \text{ bands}$).

Riemannian covariance features. For each epoch $X \in \mathbb{R}^{C \times T}$ with $C = 30$ and $T = 384$, the channel-wise covariance matrix was computed with a small diagonal regularization $C_\epsilon = C + 10^{-6} I$. Each matrix was projected onto the Log-Euclidean tangent space via the matrix logarithm:

$$\log_m(C_\epsilon) = V \log(\Lambda)\, V^\top, \qquad (3)$$

where $C_\epsilon = V \Lambda V^\top$ is the eigendecomposition. Vectorizing the upper triangular part (including the diagonal) yields a $C(C+1)/2 = 465$-dimensional feature vector per epoch.

Models

Seven model families were evaluated. Classical models operated on standardized band-power features:

- Logistic Regression $-\ \ell_2$ regularization ($C = 1.0$), lbfgs solver, 500 iterations, class-balanced.
- SVM (RBF) - radial-basis function kernel ($C = 3.0$, $\gamma =$ "scale"), class-balanced.
- Random Forest - 300 trees, unrestricted depth, class-balanced, random seed 42. Feature importances from this model were used for channel and band importance analysis.
- Riemannian LogReg - logistic regression on Log-Euclidean tangent-space features, ℓ_2 regularization, 1,000 iterations.

Three deep learning architectures were implemented in TensorFlow/Keras:

- EEGNet (2,050 parameters) - compact CNN with temporal and depthwise convolutional layers, ELU activations, and average pooling [8].
- BiLSTM-Attention (57,155 parameters) - bidirectional LSTM with 64 units per direction, learned softmax attention over 384-time steps, yielding a 128-dimensional context vector fed to a dense classifier.
- Transformer encoder (22,810 parameters) - 30-dimensional positional embeddings, multi-head self-attention (4 heads), position-wise feed-forward network (128 units), global average pooling.

All deep models were trained with Adam (learning rate 10^{-3}), categorical cross-entropy loss, batch size 32, and up to 60 epochs with early stopping (patience = 10, best-weight restoration). Time-domain data augmentation was applied exclusively to training data: Gaussian noise ($\sigma = 0.02$), random circular shifts (± 10 samples, $\approx \pm 78$ ms), and amplitude scaling in $[0.9, 1.1]$.

Experimental Setup

A leave-one-subject-out (LOSO) cross-validation protocol was adopted. In each of the 11 folds, all epochs from one subject were held out as the test set, and epochs from the remaining ten subjects were used for training and internal validation. For deep learning models, a stratified 80/20 split of the training subjects' epochs formed the training and validation subsets for early stopping; test-subject data were never used during training. For classical and Riemannian models, hyperparameters were fixed a priori, and feature scalers were fitted exclusively on training folds.

Evaluation Metrics

All models were evaluated using accuracy and macro-averaged F1-score. Accuracy is the proportion of correctly classified epochs. For each class $k \in \{\text{alert}, \text{drowsy}\}$, precision and recall are defined in the standard way, and the macro-F1 score is the unweighted average:

$$\text{Macro-F1} = 1/2 \left(\text{F1}_{\text{alert}} + \text{F1}_{\text{drowsy}} \right). \qquad (4)$$

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 25, MARCH 2026

Mean and standard deviation across 11 LOSO folds are reported for all models. For the best-performing model, predictions were aggregated across folds to compute a global confusion matrix and to count false positives and false negatives.

Statistical Analysis

To assess whether performance differences between models were statistically significant, Wilcoxon signed-rank tests were applied to paired per-subject accuracies under the LOSO protocol. This non-parametric procedure does not assume normality, which is appropriate given the modest sample size of eleven participants. Two-sided p-values were computed for several planned pairwise comparisons. No test-subject data were used for model selection or hyperparameter tuning, keeping all comparisons unbiased with respect to the cross-validation scheme.

Results

The performance of all models is summarized in Table 1. The BiLSTM-attention model produced the strongest results overall, with an accuracy of $74.00\% \pm 11.31$ and macro-F1 of $73.03\% \pm 12.25$ across LOSO folds. EEGNet followed closely, while classical models such as Random Forest and the Riemannian classifier demonstrated competitive performance only marginally below the deep models.

Table 1. LOSO accuracy and macro-F1 (mean $\pm$ standard deviation).

Model	Accuracy (%)	Macro-F1 (%)
BiLSTM-Attention	74.00 ± 11.31	73.03 ± 12.25
EEGNet	72.60 ± 11.09	71.34 ± 12.17
Random Forest	71.58 ± 12.52	70.24 ± 15.38
Riemann LogReg	70.10 ± 9.84	68.67 ± 10.81
Logistic Regression	69.86 ± 11.37	68.74 ± 12.09
SVM (RBF)	68.69 ± 9.90	66.10 ± 13.03
Transformer	68.20 ± 10.13	65.47 ± 13.24

Wilcoxon tests revealed no statistically significant difference between EEGNet and BiLSTM-attention ($p = 0.70$). A significant difference was found in favor of BiLSTM-attention compared to the transformer model ($p = 0.039$). EEGNet and the Riemannian classifier did not differ significantly ($p = 0.43$), highlighting the strength of covariance-based methods.

Error analysis showed a pronounced asymmetry: 295 false positives (alert predicted as drowsy) and 184 false negatives (drowsy predicted as alert). This bias toward predicting the drowsy class may be advantageous in safety-critical settings, though it increases false-alarm rates.

Theta-band power emerged as the most discriminative spectral feature across subjects, with delta and alpha also contributing strongly. Random forest feature-importance scores emphasized posterior and paracentral electrodes. Figure 1 shows model performance, Figure 2 presents the aggregated confusion matrix, and Figures 3–4 show channel and band importance results.

DOI: 10.37943/25EJPK6829

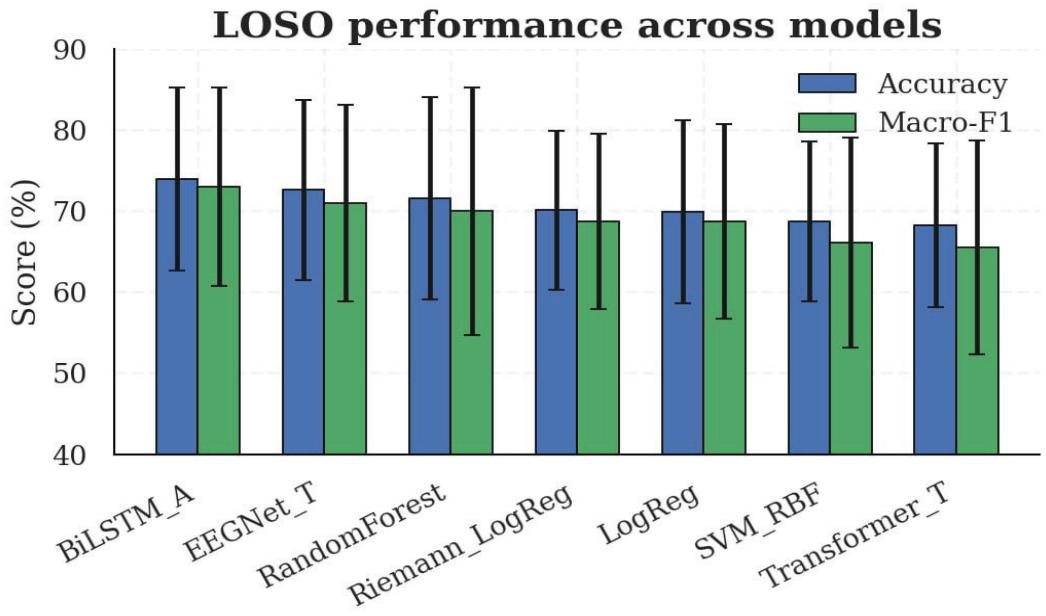

Figure 1. Model performance under LOSO cross-validation. Accuracy and macro-F1 scores for all evaluated models averaged across 11 subjects.

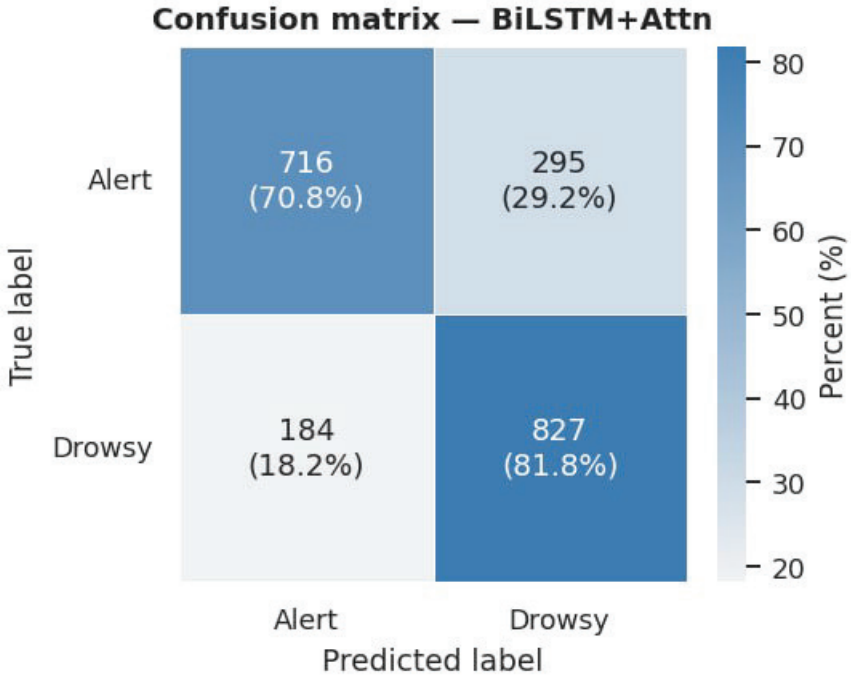

Figure 2. Aggregated confusion matrix of the BiLSTM-Attention model across LOSO folds. False positives (alert → drowsy) are more frequent than false negatives.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 25, MARCH 2026

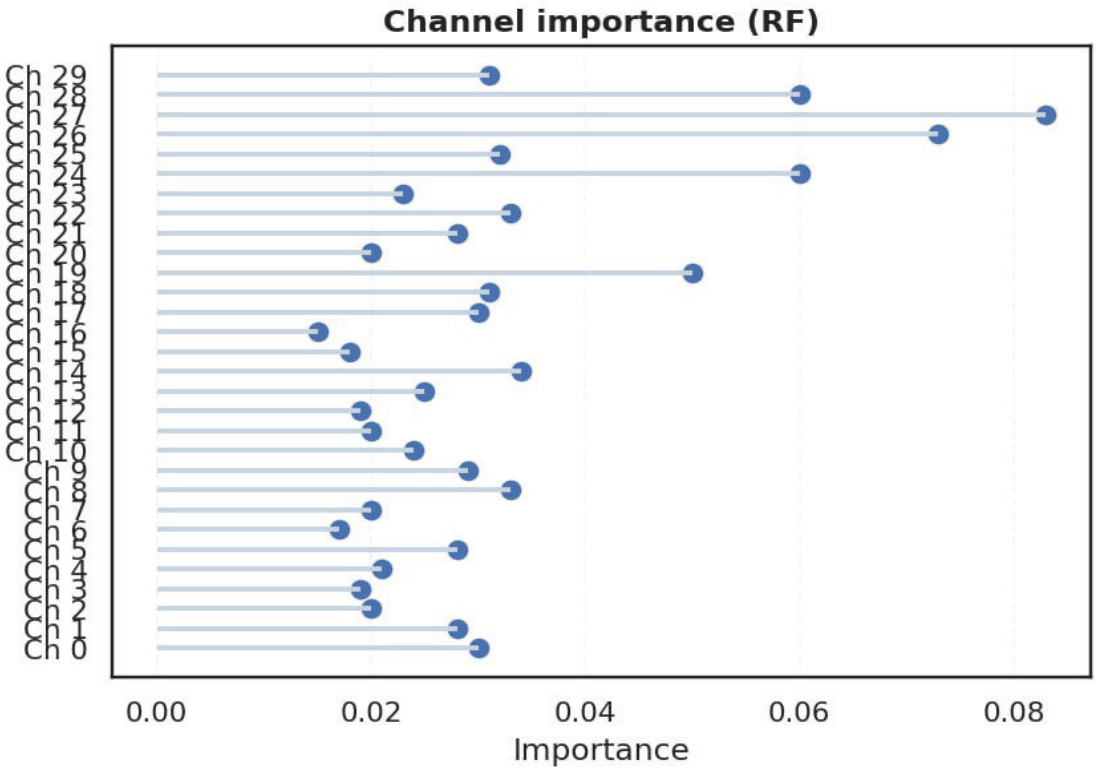

Figure 3. Channel importance derived from the Random Forest classifier. Posterior and paracentral electrodes exhibit the highest contribution to fatigue discrimination.

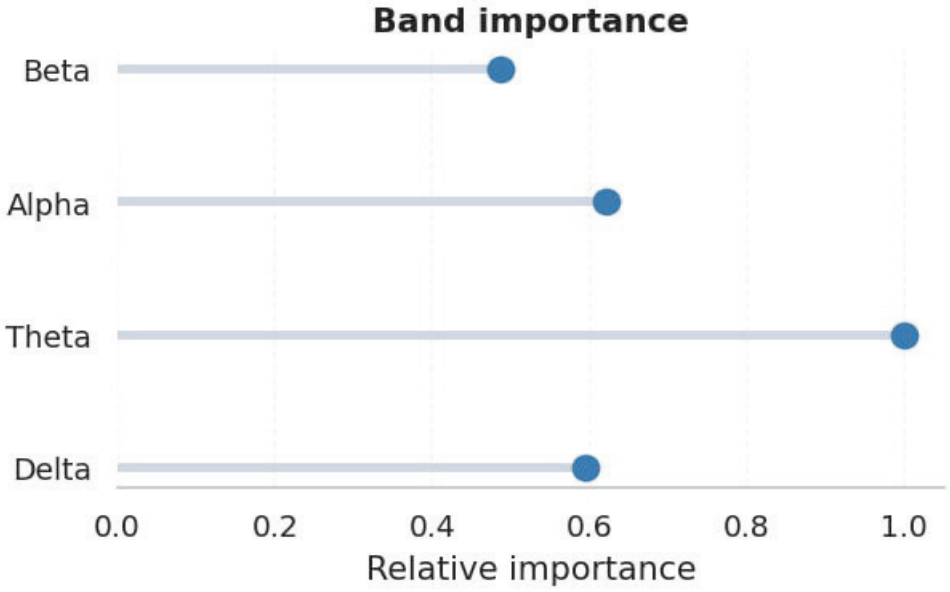

Figure 4. Band importance across all subjects. Theta power is the strongest predictor of drowsiness, followed by delta and alpha rhythms.

Discussion

Comparison with Prior Work on the Same Dataset

The BiLSTM-attention model achieved a mean LOSO accuracy of 74.00%, which situates it competitively within published results on the same 11-subject EEG Driver Drowsiness Dataset. Cui et al. (2022) reported 73.22% accuracy for a compact single-channel interpretable CNN (CompactCNN) on the

same LOSO protocol [14], and their subsequent interpretable CNN published in IEEE TNNLS reached 78.35% using multi-channel spatial filtering with a more complex network architecture [15]. The present BiLSTM-attention result of 74.00% is notably achieved with all 30 channels and a relatively lightweight model (57,155 parameters), which confirms that temporal attention over the full channel set provides competitive generalization without requiring specialized spatial convolution designs.

EEGNet achieved 72.60% in the present study, consistent with its established role as a strong multi-channel baseline. These findings collectively confirm that the performance ceiling for subject-independent classification on this 11-subject dataset using LOSO is approximately 73–78% with current architectures, and that further improvements likely require either larger datasets, domain adaptation, or test-time calibration strategies [16, 17].

Model Complexity and Generalization

A striking observation from Table 1 is the narrow spread of performance across seven architecturally diverse models: accuracy ranged from 68.20% (Transformer) to 74.00% (BiLSTM-attention), a gap of less than 6 percentage points. This compressed range is characteristic of small-N cross-subject EEG evaluations and has been noted in comparable benchmarking studies [13, 18]. The transformer model, despite its theoretical capacity to capture long-range temporal dependencies, performed worst among deep models ($68.20\% \pm 10.13$). This is consistent with recent findings showing that transformers tend to overfit on EEG datasets comprising fewer than 20 subjects when trained from scratch without pre-training [12]. In contrast, EEGNet's inductive biases, temporal convolution followed by depthwise spatial filtering, appear to regularize learning implicitly, yielding robust performance with only 2,050 parameters [8].

Notably, the Transformer was also outperformed by the classical Random Forest baseline ($71.58\% \pm 12.52$), reinforcing that high-capacity architectures often struggle to capture robust cross-subject patterns in small-sample settings without extensive pre-training. Furthermore, the high standard deviation observed in the BiLSTM-Attention results ($\pm 11.31\%$) underscores significant inter-subject variability, where specific "outlier" subjects exhibit neural signatures that deviate significantly from the group mean.

The Riemannian classifier ($70.10\% \pm 9.84$) performed comparably to EEGNet despite operating on a fundamentally different feature representation. Covariance matrices capture the second-order statistics of the multichannel signal and are invariant to orthogonal transformations, making them robust to subject-level differences in electrode impedance and reference choice [19]. The relatively low standard deviation of the Riemannian method (9.84% vs. 11–13% for deep models) suggests more stable generalization across subjects, which may be advantageous in deployment settings where reliability matters more than peak accuracy.

For resource-constrained environments, these results suggest that the Riemannian approach offers a stable and computationally efficient alternative to deep learning, as it maintains competitive accuracy without the high computational overhead.

Error Asymmetry and Safety Implications

The aggregated confusion matrix revealed 295 false positives (alert classified as drowsy) against 184 false negatives (drowsy classified as alert). This asymmetry — a false positive rate exceeding false negatives by approximately 60% - has direct safety implications. In a real-time driver monitoring context, a false negative (missed drowsiness) is the more dangerous error, as it corresponds to a failure to alert a driver who is genuinely falling asleep. From this perspective, the model's bias toward predicting drowsiness is arguably desirable, since it produces fewer missed detections at the cost of more false alarms. However, excessive false alarms may cause alarm fatigue in drivers, reducing system credibility [20]. Calibrating the decision threshold to control the false negative rate explicitly, rather than relying on the default 0.5 threshold, would be a practical step toward deployment [21].

Neurophysiological Interpretation

Band-importance analysis confirmed that theta power (4–8 Hz) is the strongest single predictor of drowsiness, followed by delta and alpha contributions. This is consistent with a large body of neurophysiological evidence documenting frontal and central theta increases during vigilance decline [22, 23]. Alpha suppression in occipital regions, which was captured by the high importance of posterior

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 25, MARCH 2026

electrodes, reflects the withdrawal of attentional resources during drowsiness and has been identified as a reliable drowsiness marker across studies [24].

The relatively low importance of beta activity (13–30 Hz) is consistent with its known association with active cognitive engagement rather than fatigue, and with previous feature selection studies on this dataset [14]. The emphasis on paracentral and parietal electrodes aligns with posterior alpha modulation linked to sensorimotor inhibition during drowsiness. These results support the interpretability of the spectral feature pipeline and suggest that electrode reduction strategies could retain most discriminative information by focusing on a subset of 8–12 posterior and central channels [25].

Limitations

Several limitations should be acknowledged. First, the dataset comprises only 11 subjects, which constrains statistical power and limits the stability of cross-validated performance estimates. Second, the LOSO protocol does not account for session-level variability; within-subject recordings from different days would provide a more stringent test of generalization. Third, the deep models were trained without pre-training or domain adaptation, strategies that have demonstrated substantial accuracy improvements in recent work [16, 17]. Finally, the present study used a fixed 3-second epoch length inherited from the original dataset segmentation; adaptive window strategies may improve temporal resolution of drowsiness onset detection.

Conclusion

This study presents a systematic benchmarking of seven model families for cross-subject EEG-based fatigue detection using a strict leave-one-subject-out evaluation protocol. The BiLSTM-attention architecture achieved the highest mean accuracy ($74.00\% \pm 11.31$) and macro-F1 ($73.03\% \pm 12.25$), placing it competitively relative to prior published results on the same 11-subject dataset. However, the narrow performance range across all seven models (less than 6 percentage points) underscores that architecture alone does not determine cross-subject generalization in small EEG cohorts. The Riemannian classifier demonstrated particularly stable performance with the lowest standard deviation, suggesting geometric methods remain highly relevant for low-data scenarios.

Error structure analysis revealed a bias toward false positives, which — while reducing missed drowsiness detections — highlights the need for threshold calibration in safety-critical applications. Neurophysiological analysis confirmed theta and alpha activity in posterior and paracentral regions as the dominant discriminative features, consistent with established fatigue neuroscience.

Future work should investigate domain adaptation and test-time adaptation strategies, which have shown substantial improvements in recent calibration-free drowsiness detection systems. Extending evaluations to larger and more heterogeneous datasets, incorporating multimodal signals, and exploring interpretable architectures represent the most promising directions for practical deployment of fatigue monitoring systems.

Acknowledgment

This study was funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. BR27198099).

References

[1] Craik, A., He, Y., & Contreras-Vidal, J. (2019). Deep learning for electroencephalogram (EEG): A review. *Journal of Neural Engineering*, *16*(3), 031001. https://doi.org/10.1088/1741-2552/ab0ab5 https://doi.org/10.1088/1741-2552/ab0ab5?urlappend=?utm_source=researchgate.net&utm_medium=article

[2] Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., & Faubert, J. (2019). Deep learning-based electroencephalography analysis: A systematic review. *Journal of Neural Engineering*, *16*(5), 051001. https://doi.org/10.1088/1741-2552/ab260c

[3] Stancin, I., Cifrek, M., & Jovic, A. (2021). A review of EEG signal features and their application in driver drowsiness detection systems. *Sensors*, *21*(11), 3786. https://doi.org/10.3390/s21113786

[4] Cao, Z., Chuang, C.-H., King, J.-K., & Lin, C.-T. (2019). Multi-channel EEG recordings during a sustained-attention driving task. *Scientific Data*, *6*(1), 1–8. https://doi.org/10.1038/s41597-019-0027-4

[5] Paulo, J. R., Pires, G., & Nunes, U. J. (2021). Cross-subject zero calibration driver's drowsiness detection: Exploring spatiotemporal image encoding of EEG signals for convolutional neural network classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *29*, 905–915. https://doi.org/10.1109/TNSRE.2021.3079505

[6] Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., & Yger, F. (2018). A review of classification algorithms for EEG-based brain–computer interfaces: A 10-year update. *Journal of Neural Engineering*, *15*(3), 031005. https://doi.org/10.1088/1741-2552/aab2f2

[7] Barachant, A., Bonnet, S., Congedo, M., & Jutten, C. (2012). Multiclass brain–computer interface classification by Riemannian geometry. *IEEE Transactions on Biomedical Engineering*, *59*(4), 920–928. https://doi.org/10.1109/TBME.2011.2172210

[8] Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., & Lance, B. J. (2018). EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, *15*(5), 056013. https://doi.org/10.1088/1741-2552/aace8c

[9] Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A review of recurrent neural networks: LSTM cells and network architectures. Neural computation, 31(7), 1235-1270. https://doi.org/10.1162/neco_a_01199

[10] Zheng, W.-L., & Lu, B.-L. (2015). Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Transactions on Autonomous Mental Development*, *7*(3), 162–175. https://doi.org/10.1109/TAMD.2015.2431497

[11] Gao, Z., Wang, X., Yang, Y., Mu, C., Cai, Q., Dang, W., & Zuo, S. (2019). EEG-based spatio-temporal convolutional neural network for driver fatigue evaluation. *IEEE Transactions on Neural Networks and Learning Systems*, *30*(9), 2755–2763. https://doi.org/10.1109/TNNLS.2018.2886414

[12] Kostas, D., Aroca-Ouellette, S., & Rudzicz, F. (2021). BENDR: Using transformers and a contrastive self-supervised learning task to learn from massive amounts of EEG data. *Frontiers in Human Neuroscience*, 15, 653659. https://doi.org/10.3389/fnhum.2021.653659

[13] Banville, H., Chehab, O., Hyvarinen, A., Engemann, D.-A., & Gramfort, A. (2021). Uncovering the structure of clinical EEG signals with self-supervised learning. *Journal of Neural Engineering*, *18*(4), 046020. https://doi.org/10.1088/1741-2552/abca18

[14] Cui, J., Lan, Z., Liu, Y., Li, R., Li, F., Sourina, O., & Müller-Wittig, W. (2022). A compact and interpretable convolutional neural network for cross-subject driver drowsiness detection from single-channel EEG. *Methods*, *202*, 173–184. https://doi.org/10.1016/j.ymeth.2021.04.017

[15] Cui, J., Lan, Z., Sourina, O., & Müller-Wittig, W. (2022). EEG-based cross-subject driver drowsiness recognition with an interpretable convolutional neural network. *IEEE Transactions on Neural Networks and Learning Systems*, *34*(10), 7921–7933. https://doi.org/10.1109/TNNLS.2022.3147208

[16] Feng, X., Guo, Z., & Kwong, S. (2025). ID3RSNet: Cross-subject driver drowsiness detection from raw single-channel EEG with an interpretable residual shrinkage network. *Frontiers in Neuroscience*, *18*, 1508747. https://doi.org/10.3389/fnins.2024.1508747

[17] Yuan, L., Zhang, S., Li, R., Zheng, Z., Cui, J., & Siyal, M. Y. (2025). Benchmarking EEG-based cross-dataset driver drowsiness recognition with deep transfer learning. *IEEE Journal of Biomedical and Health Informatics*, *29*(3), 1970–1981. https://doi.org/10.1109/embc40787.2023.10340982

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 25, MARCH 2026

[18] Kwon, O. Y., Lee, M. H., Guan, C., & Lee, S. W. (2020). Subject-independent brain–computer interfaces based on deep convolutional neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, *31*(10), 3839–3852. https://doi.org/10.1109/TNNLS.2019.2946869

[19] Barachant, A., Bonnet, S., Congedo, M., & Jutten, C. (2013). Classification of covariance matrices using a Riemannian-based kernel for BCI applications. *Neurocomputing*, *112*, 172–178. https://doi.org/10.1016/j.neucom.2012.12.039

[20] Tran, Y., Craig, A., Craig, R., Chai, R., & Nguyen, H. (2020). The influence of mental fatigue on brain activity: Evidence from a systematic review with meta-analysis. *Psychophysiology*, *57*(5), e13554. https://doi.org/10.1111/psyp.13554

[21] Schirrmeister, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W., & Ball, T. (2017). Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping*, *38*(11), 5391–5420. https://doi.org/10.1002/hbm.23730

[22] Monteiro, T. G., Skourup, C., & Zhang, H. (2019). Using EEG for mental fatigue assessment: A comprehensive look into the current state of the art. *IEEE Transactions on Human-Machine Systems*, *49*(6), 599-610. https://doi.org/10.1109/THMS.2019.2938156

[23] Arefnezhad, S., Hamet, J., Eichberger, A., Lex, C., Koglbauer, I. V., Scholler, G., & Naderi, A. (2022). Driver drowsiness estimation using EEG signals with a dynamical encoder-decoder modeling framework. *Scientific Reports*, *12*, 2650. https://doi.org/10.1038/s41598-022-05810-x

[24] Chuang, C. H., Cao, Z., King, J. T., Wu, B. S., Wang, Y. K., & Lin, C. T. (2018). Brain electrodynamic and hemodynamic signatures against fatigue during driving. *Frontiers in neuroscience*, *12*, 181. https://doi.org/10.3389/fnins.2018.00181

[25] Othmani, A., Sabri, A. Q. M., Aslan, S., Chaieb, F., Rameh, H., Alfred, R., & Cohen, D. (2023). EEG-based neural networks approaches for fatigue and drowsiness detection: A survey. *Neurocomputing*, *557*, 126709. https://doi.org/10.1016/j.neucom.2023.126709