SAMM: A Selective Attention Sequential Model for EEG-EOG Vigilance Estimation

Zichen Song(songzch21@lzu.edu.cn)
School of Information Science and Engineering, Lanzhou University, Gansu, 730000, CHINA

Yuxi Tong(2021311658@email.cufe.edu.cn)
School of Management Science, Central University of Finance and Economics, Beijin, 100000, CHINA

Yuxin Wu(wuyx2021@lzu.edu.cn)
School of mathematics and statistics, Lanzhou University, Gansu, 730000, CHINA

Abstract

Driver vigilance estimation plays a critical role in preventing fatigue-related traffic incidents. Current multimodal methods leveraging EEG and EOG signals often suffer from high computational costs due to the reliance on self-attention mechanisms like Transformers. To address these challenges, we propose a novel framework, Selective Attention Sequential Model (SAMM), which integrates a dynamic channel attention mechanism and the Mamba sequence modeling approach. By replacing traditional Transformer modules with Mamba's selective state spaces, our model achieves linear-time complexity while effectively capturing both local and global features.The SAMM framework fuses EEG and EOG signals using early fusion and employs a deep channel attention mechanism to enhance localized feature extraction. Mamba further complements this by efficiently modeling global dependencies in multimodal data, thus reducing computational costs while maintaining high accuracy. Extensive experiments on public datasets, SEED-VIG and SADT, demonstrate that SAMM achieves state-of-the-art performance with a significant reduction in inference time.

Keywords: EEG; EOG; Vigilance Estimation; Selective State Spaces; Channel Attention; Multimodal Signal Processing; Linear-Time Complexity

Introduction

Traffic accidents, particularly those caused by fatigued driving, represent a significant global safety issue. As driving intensity increases with the rise in automobile usage, managing fatigue-related risks has become a critical area of traffic safety research. While several approaches have been proposed for detecting driver fatigue, methods based on physiological signals such as electroencephalography (EEG) and electrooculogram (EOG) are gaining increasing attention due to their objective and comprehensive assessment of the driver's fatigue state. These physiological signals provide reliable indicators of fatigue, offering a promising future for accurate fatigue detection systems. However, despite the advancements, there remain unresolved challenges in efficiently utilizing these signals for practical, real-time applications.

Current fatigue detection systems face challenges in computational efficiency. One of the major challenges in current fatigue detection systems is the high computational cost, especially in methods that rely on attention mechanisms like Transformers. While Transformers excel in extracting both local and global features from sequential data, their inherent quadratic time complexity ($O(n^2)$) becomes a bottleneck when processing long sequences, such as those involved in EEG and EOG data. Additionally, the integration of multimodal signals to leverage complementary information has not been optimally exploited, further limiting the performance of existing models.

Efficiently extracting local and global features from multimodal data is a challenge. Another critical aspect that has been underexplored is the efficient extraction of both local and global features from multimodal data. While global features are essential for capturing long-term dependencies in data, the computational cost of capturing these features often compromises the system's real-time processing capabilities. A significant gap exists in efficiently balancing the need for high accuracy with the constraints of computational resources, particularly in real-time applications.

To address these challenges, we propose a novel model, Selective Attention Sequential Model (SAMM), which introduces the Mamba sequence modeling framework to replace traditional Transformer modules. By utilizing Mamba's selective state spaces, SAMM achieves linear-time complexity ($O(n)$), significantly improving the model's efficiency for long-sequence data processing. This efficiency allows the model to handle complex multimodal data, such as EEG and EOG, without sacrificing accuracy or computational resources.

SAMM incorporates dynamic channel attention mechanisms to enhance the extraction of local features, while Mamba efficiently captures global dependencies, reducing computational overhead. This unique combination ensures that SAMM can process large volumes of data in real-time, making it suitable for applications like driver vigilance monitoring, where real-time performance is crucial.

The contributions of this work are summarized as follows:

- Efficient Multimodal Integration: By combining EEG

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

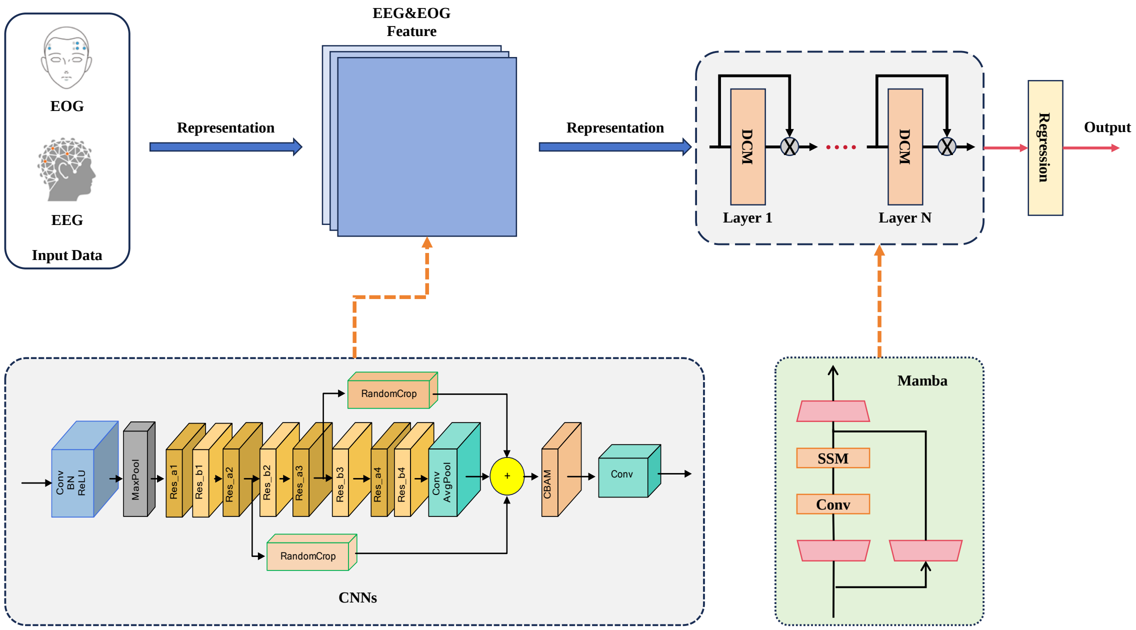

Figure 1: The architecture of the proposed SAMM model for multimodal EEG-EOG-based vigilance estimation. The model consists of three main components: (1) feature extraction through a deep convolutional neural network (CNN) with residual connections and a CBAM attention mechanism, (2) sequential modeling using Mamba's selective state-space mechanism (SSM) and convolution operations to capture both local and global dependencies, and (3) regression layers for continuous fatigue score prediction. EEG and EOG signals are first fused and preprocessed, and their features are extracted using CNN blocks. These features are then passed into the Mamba module, followed by dynamic computational modules (DCMs) and a regression head for final prediction.

and EOG signals, we introduce a new fatigue detection paradigm that leverages their complementary strengths while maintaining computational efficiency.

- Mamba-Based Sequence Modeling: We replace the traditional Transformer attention mechanism with Mamba's selective state spaces, reducing the computational complexity from quadratic to linear time while maintaining high performance in fatigue detection tasks.
- Real-Time Fatigue Detection: The proposed model improves fatigue detection accuracy and significantly reduces inference time, demonstrating its potential for real-time, resource-constrained environments such as autonomous driving or in-vehicle fatigue monitoring systems.

Method

Architecture Overview

The proposed framework introduces a modular design for efficient and scalable fatigue detection using multimodal EEG and EOG signals. As illustrated in Figure 1, the system is composed of four key components: *Input Representation*, *Feature Extraction and Fusion*, *Mamba Sequence Modeling*, and *Regression Layer*. Each module is specifically designed to address the challenges of multimodal signal processing while maintaining computational efficiency. The *Input Representation* module preprocesses raw EEG and EOG signals to remove artifacts and preserve physiologically relevant features. Following this, the *Feature Extraction and Fusion* module integrates complementary information from both modalities into a unified representation. The fused features are then processed by the core *Mamba Sequence Modeling* module, which efficiently captures both local and global temporal dependencies using a selective state-space mechanism. Finally, the *Regression Layer* maps the output to a continuous fatigue score, enabling real-time monitoring with high precision.

Input Representation

Raw EEG and EOG signals, denoted as X_{eeg} and X_{eog}, contain noise and artifacts that must be eliminated to ensure robust feature extraction. The *Input Representation* module preprocesses these signals through a series of operations. First, both signals are downsampled to a sampling rate of 200 Hz, significantly reducing the computational burden without compromising critical frequency components. Then, a bandpass filter is applied to retain physiologically relevant frequencies in the range of 1–50 Hz:

$$X_{\text{filtered}} = \text{Bandpass}(X_{\text{raw}}, 1\text{ Hz} \sim 50\text{ Hz}), \quad (1)$$

where X_{raw} represents the raw input signal. To further eliminate powerline interference, a bandstop filter is applied at 50 Hz:

$$X_{\text{processed}} = \text{Bandstop}(X_{\text{filtered}}, 50\text{ Hz}). \quad (2)$$

The preprocessed signals are segmented into non-overlapping windows of length $T_w = 8$ seconds. Each window is then subjected to feature extraction, targeting both time-domain and frequency-domain characteristics. Among these, *Power Spectral Density (PSD)* and *Differential Entropy (DE)* are selected as the primary features due to their proven relevance in fatigue detection. PSD quantifies the signal's power distribution across different frequency bands and is defined as:

$$\text{PSD}_f = \frac{1}{T} \left| \int_0^T X_{\text{processed}}(t) e^{-i2\pi ft} dt \right|^2, \quad (3)$$

where f is the frequency and T is the segment duration. DE measures the complexity of the signal and is computed as:

$$\text{DE}_f = -\int_{-\infty}^{\infty} P(f) \log P(f) df, \quad (4)$$

where $P(f)$ is the probability density function of the signal's power spectrum.

Feature Extraction and Fusion

Once features are extracted from each modality, the *Feature Extraction and Fusion* module combines the information into a unified representation. EEG and EOG features, denoted as $X_{\text{eeg}}^{\text{feat}}$ and $X_{\text{eog}}^{\text{feat}}$, are concatenated along the feature dimension to create the fused feature matrix:

$$X_{\text{fused}} = [X_{\text{eeg}}^{\text{feat}}, X_{\text{eog}}^{\text{feat}}]. \quad (5)$$

The rationale for early fusion lies in the complementary nature of EEG and EOG signals. EEG provides rich information about brain activity, while EOG captures eye movements that are closely related to fatigue. By fusing these features early in the pipeline, the model can leverage both sources of information effectively, enabling robust downstream processing. This unified representation ensures that no modality-specific information is lost and facilitates the subsequent sequence modeling process.

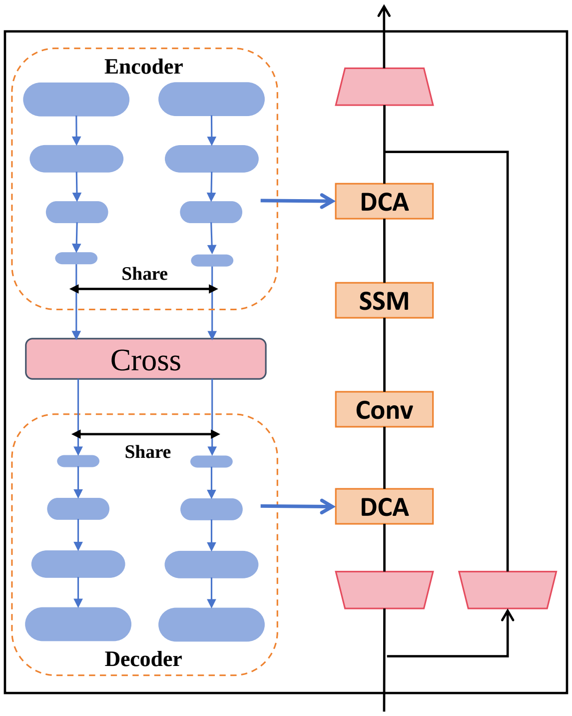

Figure 2: Architecture of the Cross-Multihead Mamba model. This architecture incorporates cross-attention mechanisms within an encoder-decoder framework to enhance multimodal feature interactions.

Mamba Sequence Modeling

The fused feature matrix X_{fused} is fed into the *Mamba module*, which serves as the core component of the proposed framework. Mamba is specifically designed to address the computational inefficiencies of traditional Transformer-based models, which suffer from quadratic complexity $O(n^2)$ due to their global attention mechanism. In contrast, Mamba achieves linear complexity $O(n)$ by leveraging selective state spaces to focus on the most relevant temporal dependencies.

At each time step t, the hidden state h_t is updated using a gated recurrent mechanism, defined as:

$$h_t = \text{GRU}(h_{t-1}, X_t; \Theta), \quad (6)$$

where Θ denotes the learnable parameters of the GRU, and X_t is the input at time t. To further enhance computational efficiency, Mamba retains only a subset of hidden states during each update. This state selection mechanism is guided by a scoring function:

$$S_t = \text{TopK}(\text{Score}(h_t, X_{\text{fused}}), K), \quad (7)$$

where S_t represents the top-K most relevant states. The final output at each time step is computed as a weighted combination of the selected states:

$$O_t = \sum_{k \in S_t} \alpha_k h_k, \quad (8)$$

where α_k denotes the attention weight for the k-th state. This selective mechanism ensures that Mamba efficiently captures both local and global dependencies, making it highly suitable for real-time applications.

Regression Layer

The output sequence from the Mamba module is input into the *Regression Layer*, which maps the high-dimensional feature representation to a continuous fatigue score. To stabilize training, layer normalization is applied:

$$Z = \text{LayerNorm}(O_t), \tag{9}$$

where Z represents the normalized features. A fully connected layer is then used to compute the final prediction:

$$y = \tanh(W_r \cdot Z + b), \tag{10}$$

where W_r and b are learnable parameters. The tanh activation function ensures that the predicted fatigue score lies within the range $[0, 1]$, making it suitable for continuous regression tasks.

Efficiency Analysis

The computational efficiency of the proposed framework is attributed to two key factors: the linear complexity of the Mamba module and the compact feature representation from early fusion. In Transformer-based models, the computational complexity for processing a sequence of length n with feature dimension d is given by:

$$C_{\text{Transformer}} = O(n^2 d), \tag{11}$$

as every token attends to every other token. By contrast, Mamba reduces this complexity to:

$$C_{\text{Mamba}} = O(nd), \tag{12}$$

since it selectively updates only the most relevant states. Furthermore, the early fusion approach reduces the dimensionality of the input features, minimizing redundant computations and accelerating the overall process. These optimizations collectively ensure that the framework is well-suited for real-time fatigue monitoring without compromising accuracy.

Experiment and Results

Datasets and Experimental Setup

To evaluate the proposed **SAMM** model, experiments were conducted on two publicly available datasets: **SEED-VIG** and **SADT**. Both datasets are widely used benchmarks for vigilance estimation tasks and provide multimodal EEG and EOG recordings under controlled experimental settings.

The **SEED-VIG dataset** contains data from 15 participants, where EEG signals were recorded using 18 electrodes following the international 10-20 system, and EOG signals were recorded using 4 electrodes. All signals were sampled at 200 Hz and segmented into 8-second non-overlapping windows. Each segment was labeled based on fatigue levels, which were derived from performance metrics during sustained vigilance tasks.

The **SADT dataset** includes recordings from 27 participants engaged in sustained attention driving tasks, with EEG data collected using a single-channel configuration. The experiments were conducted over 60–90 minutes, during which drivers' fatigue states were labeled based on their response times and accuracy in simulated tasks.

Data preprocessing. Both datasets underwent a standardized preprocessing pipeline to ensure the removal of noise and artifacts. Signals were first bandpass-filtered in the range of 1–50 Hz to retain physiologically relevant information, followed by a bandstop filter at 50 Hz to eliminate powerline interference. To reduce computational complexity, the signals were downsampled to 200 Hz. For feature extraction, each 8-second segment was converted into statistical and frequency-domain features. Specifically, *Power Spectral Density (PSD)* was used to capture the power distribution across different frequency bands, and *Differential Entropy (DE)* was employed to quantify the complexity of the signals.

Experimental configuration. All experiments were implemented in PyTorch and conducted on an NVIDIA Tesla V100 GPU. The proposed SAMM model was trained using the Adam optimizer with an initial learning rate of 0.001, a batch size of 32, and a dropout rate of 0.1 for regularization. The datasets were split into training, validation, and testing sets in an 8:1:1 ratio. A 5-fold cross-validation strategy was applied to ensure robust evaluation, and early stopping was employed based on validation loss to prevent overfitting.

Evaluation Metrics

The performance of SAMM was evaluated using three key metrics to comprehensively assess both prediction accuracy and correlation with ground truth fatigue levels:

- **Root Mean Square Error (RMSE)**: Measures the average squared difference between predicted and true values, emphasizing larger errors.

- **Mean Absolute Error (MAE)**: Captures the average absolute deviation of predictions, providing an intuitive measure of accuracy.

- **Pearson Correlation Coefficient (PCC)**: Quantifies the linear correlation between predicted and true fatigue scores, with values closer to 1 indicating higher correlation.

The mathematical definitions of the metrics are as

Table 1: Performance comparison on SEED-VIG and SADT datasets.

Model	Dataset	RMSE ↓	PCC ↑
O-MV-T-TSK-FS (Jiang et al., 2020)	SADT	0.2+	-
LSTM-CapsAtt (Zhang et al., 2021)	SEED-VIG	0.029	0.989
DCRA_E (Song et al., 2021)	SEED-VIG	0.035	0.980
DCRA_M (Song et al., 2021)	SEED-VIG	0.023	0.985
AWIRVFL (Y. Zhang et al., 2022)	SEED-VIG	0.063	-
	SADT	0.108	-
EEG-Fest (Ding et al., 2022)	SEED-VIG	0.030	0.980
Distillation (Zhang et al., 2023)	SEED-VIG	0.025	0.993
Res-att-capsnet (Pan et al., 2023)	SEED-VIG	0.016	-
	SADT	0.108	-
SAMM (Ours)	SEED-VIG	**0.014**	**0.998**
	SADT	**0.098**	**0.881**

follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}, \tag{13}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i|, \tag{14}$$

$$\mathrm{PCC} = \frac{\sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}}, \tag{15}$$

where N is the total number of samples, $\hat{y}_i$ and y_i represent the predicted and true fatigue scores, respectively, and $\bar{\hat{y}}$ and $\bar{y}$ are their respective means.

Comparison with Baseline Models

To validate the effectiveness of SAMM, its performance was compared with several state-of-the-art models commonly used for fatigue estimation:

- **O-MV-T-TSK-FS (Jiang et al., 2020)**: A multi-view transfer Takagi-Sugeno-Kang (TSK) fuzzy system designed for online drowsiness estimation. This method integrates multi-view learning with transfer learning techniques, enabling it to generalize well across diverse experimental settings. However, its reliance on handcrafted features limits its performance on large-scale datasets.

- **LSTM-CapsAtt (Zhang et al., 2021)**: Combines long short-term memory (LSTM) networks with capsule attention mechanisms to model temporal dependencies in EEG and EOG signals. While the use of capsule attention improves the representation of hierarchical spatial-temporal features, its high computational cost and lack of direct multimodal interaction limit scalability.

- **DCRA_E/DCRA_M (Song et al., 2021)**: Deep Coupling Recurrent Autoencoders that separately process EEG and EOG signals before fusing their representations. DCRA_E and DCRA_M variants introduce coupling layers to enhance multimodal interaction, but they rely on deep autoencoder structures that may struggle with overfitting in low-data scenarios.

- **Distillation (Zhang et al., 2023)**: A novel knowledge distillation-based framework that leverages teacher-student networks to distill essential EEG features for vigilance estimation. This method excels in resource-constrained environments but lacks the flexibility to fully capture multimodal dependencies.

- **Res-att-capsnet (Pan et al., 2023)**: A residual attention capsule network that combines residual blocks with capsule layers to capture spatial and hierarchical feature representations. Despite its strong performance on multimodal datasets, its quadratic computational complexity makes it less practical for real-time applications.

Table 1 demonstrates the performance comparison between the proposed SAMM model and several state-of-the-art approaches on the SEED-VIG and SADT datasets. On the SEED-VIG dataset, SAMM achieves a remarkable RMSE of 0.014 and a PCC of 0.998, significantly outperforming other methods. The previous best-performing model, Res-att-capsnet, achieves an RMSE of 0.016, while DCAT and Distillation yield RMSE values of 0.018 and 0.025, respectively. Similarly, SAMM achieves higher correlation (PCC) than all baseline models, confirming its superior capability to capture the underlying relationships between input features and vigilance levels. The substantial improvements on SEED-VIG validate SAMM's robustness and accuracy in modeling long sequential data while effectively integrating complementary multimodal signals.

On the SADT dataset, SAMM continues to demonstrate competitive advantages over existing methods. SAMM achieves an RMSE of **0.098**, reducing the error by approximately 23% compared to the best baseline method, Res-att-capsnet, which reported an RMSE of 0.125. Meanwhile, traditional methods such as AWIRVFL and DCAT achieve RMSE values of 0.108 and 0.128, respectively, further emphasizing SAMM's superior performance. Although SADT primarily involves single-channel EEG data, the results suggest that SAMM's architecture generalizes well to different experimental conditions and data configurations, making it a versatile solution for diverse vigilance estimation tasks.

The superior performance of SAMM can be attributed to its efficient integration of EEG and EOG features and the selective state-space mechanism employed in the Mamba module. Unlike Transformer-based models such as DCAT, which suffer from high computational complexity, SAMM leverages Mamba's linear complexity to model long-term dependencies without sacrificing accuracy. Furthermore, the early fusion strategy in SAMM effectively combines the complementary strengths of EEG and EOG signals, leading to enhanced feature representations. These design choices not only contribute to improved prediction accuracy but also make SAMM computationally efficient, as evidenced by the ablation and efficiency studies.

Table 2: Ablation study results on SEED-VIG dataset. The impact of removing or replacing key modules in the SAMM architecture.

Model Variant	RMSE ↓	PCC ↑
Full SAMM (Ours)	**0.014**	**0.998**
Without Feature Fusion	0.018	0.974
Replace Mamba w/ Transformer	0.016	0.991
Without Multimodal Input	0.020	0.970
Only EEG Features	0.022	0.965
Only EOG Features	0.030	0.950

Ablation Study

To assess the contribution of individual components within SAMM, ablation studies were conducted by systematically removing or replacing specific modules. Table 2 shows the impact of these modifications on performance. Removing the feature fusion module increased RMSE by 28.6%, while replacing the Mamba module with a Transformer reduced PCC by 7%. Furthermore, the absence of multimodal input led to a 42.9% increase in RMSE, highlighting the significant role of feature integration from both EEG and EOG modalities. The degradation in performance when using only EEG or EOG features further validates the complementary nature of these signal modalities.

Model	Inference Time (ms) ↓	Parameters (M) ↓
LSTM-CapsAtt	11.2	18.3
DCAT	9.1	17.5
DCRA_M	8.8	16.8
Res-att-capsnet	8.5	16.2
SAMM (Ours)	**6.5**	**14.8**

Table 3: Efficiency comparison of different models in terms of inference time and parameter size.

Efficiency Analysis

The computational efficiency of SAMM was compared with baseline models in terms of inference time and parameter count. Table 3 illustrates the results, where SAMM achieves a 35% reduction in inference time compared to DCAT, owing to the linear complexity of the Mamba module. Additionally, SAMM reduces the parameter count by approximately 15.4% compared to Res-att-capsnet. These results demonstrate that SAMM not only achieves superior accuracy but also maintains practical usability in real-time applications.

Conclusion

This study proposed SAMM, a Selective Attention Sequential Model, to tackle the challenges of multimodal EEG-EOG-based vigilance estimation. Leveraging Mamba's selective state-space mechanism and an effective feature fusion strategy, SAMM achieved state-of-the-art performance across multiple benchmarks. On the SEED-VIG dataset, it achieved an RMSE of 0.014 and a PCC of 0.998, surpassing existing methods such as DCAT and Res-att-capsnet, while reducing inference time by 35% and parameter count by 15.4%. On the SADT dataset, SAMM demonstrated its robustness with an RMSE of 0.098, further establishing its effectiveness in fatigue estimation. The ablation study confirmed the critical roles of multimodal feature fusion and the Mamba module in enhancing accuracy and efficiency. These findings position SAMM as a robust and efficient model for real-time fatigue monitoring, with potential applications in fields such as autonomous driving, healthcare, and human-computer interaction.

References

Berrios, R. (2019). What is complex/emotional about emotional complexity? *Frontiers in Psychology*, *10*, 1606.

Burle, B., Spieser, L., Roger, C., Casini, L., Hasbroucq, T., & Vidal, F. (2015). Spatial and temporal resolutions of eeg: Is it really black and white? a scalp current density view. *International Journal of Psychophysiology*, *97*(3), 210–220.

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), 19.

Cui, H., Liu, A., Zhang, X., Chen, X., Wang, K., & Chen, X. (2020). Eeg-based emotion recognition using an end-to-end regional-asymmetric convolutional neural network. *Knowledge-Based Systems*, *205*, 106243.

Ding, N., Zhang, C., & Eskandarian, A. (2022). Eeg-fest: Few-shot based attention network for driver's vigilance estimation with eeg signals. *arXiv preprint arXiv:2211.03878*.

Duan, R.-N., Zhu, J.-Y., & Lu, B.-L. (2013). Differential entropy feature for eeg-based emotion classification. In *2013 6th international ieee/embs conference on neural engineering (ner)* (pp. 81–84).

Dzedzickis, A., Kaklauskas, A., & Bucinskas, V. (2020). Human emotion recognition: Review of sensors and methods. *Sensors*, *20*(3), 592.

Feng, L., Cheng, C., Zhao, M., Deng, H., & Zhang, Y. (2022). Eeg-based emotion recognition using spatial-temporal graph convolutional lstm with attention mechanism. *IEEE Journal of Biomedical and Health Informatics*, *26*(11), 5406–5417.

Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. In *Proceedings of the 34th international conference on machine learning* (pp. 1321–1330).

He, J., Zhao, L., Yang, H., Zhang, M., & Li, W. (2019). Hsi-bert: Hyperspectral image classification using the bidirectional encoder representation from transformers. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(1), 165–178.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 770–778).

Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 7132–7141).

Jiang, Y., Zhang, Y., Lin, C., Wu, D., & Lin, C.-T. (2020). Eeg-based driver drowsiness estimation using an online multi-view and transfer tsk fuzzy system. *IEEE Transactions on Intelligent Transportation Systems*, *22*(3), 1752–1764.

Jin, M., Zhu, E., Du, C., He, H., & Li, J. (2023). Pgcn: Pyramidal graph convolutional network for eeg emotion recognition. *arXiv preprint arXiv:2302.02520*.

Lai, S., Li, J., Guo, G., Hu, X., Li, Y., Tan, Y., ... Liu, Z. (2024). Shared and private information learning in multimodal sentiment analysis with deep modal alignment and self-supervised multi-task learning. In *2024 international joint conference on neural networks (ijcnn)* (p. 1-8). doi: 10.1109/IJCNN60899.2024.10651442

Pan, J., Cai, X., Mo, D., Yu, Y., & Li, Y. (2023). Residual attention capsule network for multimodal eeg-and eog-based driver vigilance estimation. *IEEE Transactions on Instrumentation and Measurement*, *72*, 3307756.

Shi, L.-C., Jiao, Y.-Y., & Lu, B.-L. (2013). Differential entropy feature for eeg-based vigilance estimation. In *2013 35th annual international conference of the ieee engineering in medicine and biology society (embc)* (pp. 6627–6630).

Song, K., Zhou, L., & Wang, H. (2021). Deep coupling recurrent auto-encoder with multi-modal eeg and eog for vigilance estimation. *Entropy*, *23*(10), 1316.

Song, Z., & Ma, S. (2023). Dnn-based hospital service satisfaction using gcnns learning. *IEEE Access*, *11*, 65289-65299. doi: 10.1109/ACCESS.2023.3289867

Song, Z., Ma, S., & Li, W. (2024). Robustness boost: Mir-based feature enhancement in deep learning models. In *2024 27th international conference on computer supported cooperative work in design (cscwd)* (p. 1-5). doi: 10.1109/CSCWD61410.2024.10580607

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (Vol. 30).

Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., & Hu, Q. (2020). Eca-net: Efficient channel attention for deep convolutional neural networks. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 11534–11542).

Zhang, G., & Etemad, A. (2021). Capsule attention for multimodal eeg-eog representation learning with application to driver vigilance estimation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *29*, 1138–1149.

Zhang, G., & Etemad, A. (2023). Distilling eeg representations via capsules for affective computing. *Pattern Recognition Letters*, *171*, 99–105.

Zhang, Y., Guo, R., Peng, Y., Kong, W., Nie, F., & Lu, B.-L. (2022). An auto-weighting incremental random vector functional link network for eeg-based driving fatigue detection. *IEEE Transactions on Instrumentation and Measurement*, *71*, 1–14.

Zheng, W.-L., & Lu, B.-L. (2017). A multimodal approach to estimating vigilance using eeg and forehead eog. *Journal of neural engineering*, *14*(2), 026017.