Decoding Neural Dissonance: From Sensory Mismatch to Model Neural Recalibration for Cybersickness Prediction

Mengyu Qiao[*,†,1], **Hongyang Xie**[†,1] **& Fang Wang**[2]
[1]Beijing Key Laboratory of Key Technologies for AI+ Domain Applications,
North China University of Technology, Beijing, China
[2]Department of Computer Science, Brunel University of London, London, UK

Abstract

Cybersickness is a major barrier to the widespread adoption of virtual reality, arising from neural dissonance between visual and vestibular stimuli. Although kinematics-based deep learning enables non-intrusive detection, existing models often lack neurophysiological grounding and fail to capture dynamic sensory recalibration and continuous symptom accumulation. To address these limitations, we propose the Neural Sensory Conflict Network (NSCNet), a biologically inspired framework grounded in Sensory Conflict Theory. Specifically, NSCNet incorporates a Conflict Alignment Embedding to model the integration of discordant sensory inputs, a State-Space Re-entrant Experts module that combines selective state-space modeling with Re-entrant Experts to capture the temporal dynamics of neural dissonance and functional modularity, and a Dynamic Sensory Reweighting mechanism that approximates adaptive gain control in the central nervous system. Experiments on the MSCVR and VR Cybersickness datasets demonstrate state-of-the-art performance, suggesting that explicitly modeling neurocognitive mechanisms improves the predictive fidelity of cybersickness detection.

Keywords: Cybersickness; Sensory Conflict Theory; State Space Models; Multimodal Fusion; Bio-inspired Computing; Virtual Reality

Introduction

Virtual reality (VR) has demonstrated transformative potential across diverse domains, including education, healthcare, and entertainment (LaViola, 2000; Stanney et al., 2003). However, its widespread adoption remains hindered by cybersickness, a condition characterized by nausea, dizziness, and disorientation (Reason & Brand, 1975). Beyond subjective discomfort, cybersickness can impair task performance, shorten exposure duration, and increase dropout rates, thereby limiting the feasibility of long-duration training, rehabilitation, and safety-critical VR applications. Sensory Conflict Theory posits that cybersickness arises from a mismatch between visually implied self-motion and vestibular cues signaling relative stationarity, requiring the central nervous system to continuously recalibrate sensory weighting (Nashner & Berthoz, 1978; Reason & Brand, 1975). In practice, this mismatch is highly dynamic: it varies with scene motion, head movements, and oculomotor behavior, and may accumulate as the brain repeatedly attempts to resolve inconsistent sensory evidence. Traditional assessment relies on post-hoc subjective questionnaires such as the SSQ (Kennedy et al., 1993), which cannot capture the real-time temporal dynamics of symptom onset and progression, thereby motivating the need for objective detection methods.

To address this challenge, early studies focused on physiological signals (e.g., EEG and EDA) to quantify autonomic responses. Although approaches based on explainable AI for multimodal biosignals (Sameri et al., 2024) and EEG-based multitaper spectral estimation (Demirel et al., 2025) have achieved high fidelity, they typically depend on intrusive, medical-grade sensors that are impractical for consumer VR. Such systems also introduce setup overhead and motion constraints that may alter natural user behavior, while inter-subject variability and sensor noise further complicate deployment outside controlled laboratory settings. Consequently, recent research has shifted toward non-intrusive approaches that leverage kinematic data readily available from Head-Mounted Displays (HMDs), such as head tracking and eye movements. This sensing pathway is attractive because it is ubiquitous in modern HMDs, inexpensive to acquire, and continuously observable during VR exposure without burdening the user. Deep learning has become the dominant modeling paradigm for this modality. For example, Islam et al. (2021) showed that deep fusion of eye- and head-tracking signals can approach the accuracy of physiology-based detection, whereas Li et al. (2023) introduced kinematic encoders to learn physiology-relevant representations implicitly.

Despite these advances, current kinematics-based deep learning models still have important limitations in capturing the neurophysiological mechanisms underlying cybersickness. First, state-of-the-art architectures such as Transformers (e.g., PRECYSE (Jeong & Han, 2024)) and attention-based LSTMs (e.g., MAC (Jeong et al., 2023)) often treat multimodal fusion as a largely static integration problem. Consequently, they do not explicitly model dynamic sensory recalibration, that is, how the brain continuously resolves conflicts between discordant visual and vestibular inputs over time. Second, cybersickness is inherently cumulative: neural dissonance may build progressively throughout an exposure session. Traditional recurrent neural networks (RNNs) often struggle to capture such long-range dependencies, whereas standard Transformers incur quadratic computational costs for long sequences (Vaswani et al., 2017). These limitations hinder precise modeling of the continuous temporal integration of sensory conflict required for accurate severity prediction.

* Corresponding author. E-mail: myuqiao@ncut.edu.cn
† These authors contributed equally to this work.

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

To bridge this gap between biological theory and computational modeling, we propose the Neural Sensory Conflict Network (NSCNet), a biologically inspired multimodal framework grounded in Sensory Conflict Theory. Unlike prior purely data-driven approaches, NSCNet is designed to model sensory mismatch and symptom accumulation more explicitly. To address the limitations of static fusion, we introduce a Conflict Alignment Embedding that uses cross-modal attention to model sensory recalibration. To capture long-range temporal dependencies efficiently, we further propose a State-Space Re-entrant Experts module, which combines the linear-time sequence modeling capability of Selective State Space Models (Gu & Dao, 2024), recently applied to physiological signal processing (Luo et al., 2024), with the functional specialization enabled by Re-entrant Experts based on the Chain of Experts (CoE) (Z. Wang et al., 2025). Finally, we incorporate a Dynamic Sensory Reweighting mechanism to approximate adaptive gain control in the central nervous system.

The main contributions of this paper are summarized as follows:

- We propose NSCNet, a biologically inspired framework that integrates Sensory Conflict Theory into deep learning to model sensory recalibration and the accumulation of neural dissonance.
- We introduce State-Space Re-entrant Experts, a hybrid architecture that combines a state space model with Re-entrant Experts to capture the long-range temporal progression of cybersickness while alleviating the computational and memory limitations of Transformers and RNNs.
- Extensive experiments on the MSCVR and VR Cybersickness datasets show that NSCNet achieves state-of-the-art performance, providing an accurate and interpretable solution for non-intrusive cybersickness prediction.

Methodology

In this study, we propose NSCNet, a biologically inspired multimodal framework for cybersickness prediction. Grounded in Sensory Conflict Theory, NSCNet addresses two key challenges: cybersickness emerges from neural dissonance between visual inputs (eye movements) and vestibular signals (head movements), requiring effective cross-modal integration, while these heterogeneous signals exhibit distinct temporal dynamics and spatial distributions that complicate the extraction of symptom-related representations. We hypothesize that explicitly modeling the interaction and alignment between visual and vestibular streams improves the characterization of sensory conflict relevant to cybersickness. The overall architecture of NSCNet is shown in Figure 1.

Conflict Alignment Embedding

Cybersickness arises from a mismatch between visual and vestibular inputs. To model this process, we introduce the Conflict Alignment Embedding (Figure 1(a)), which aligns eye- and head-motion signals before temporal modeling. Because these modalities exhibit different dynamics, simple concatenation is insufficient to capture their cross-modal dependencies.

We therefore use cross-modal attention, allowing each modality to update its representation using the other as context. This produces a soft, time-varying alignment map and turns fusion into an explicit interaction-modeling step rather than simple feature stacking.

As shown in Figure 1(a), a learnable prompt is appended to the query stream ($1 \times D$). Rather than implying a literal neural counterpart, it serves as a trainable reference token that stabilizes attention and provides a consistent baseline for conflict estimation across the sequence.

To model sensory recalibration, we use bidirectional cross-modal attention, in which each modality is updated using the other as context. Let $X_{\text{eye}} \in \mathbb{R}^{T\times d}$ and $X_{\text{head}} \in \mathbb{R}^{T\times d}$ denote the eye and head token sequences. For eye-updated cross-attention, the eye stream provides queries and the head stream provides keys and values through learned projections:

$$\mathbf{Q}_{\text{eye}} = X_{\text{eye}}W_Q, \quad \mathbf{K}_{\text{head}} = X_{\text{head}}W_K, \quad \mathbf{V}_{\text{head}} = X_{\text{head}}W_V \tag{1}$$

The corresponding output is

$$\mathbf{A}_{\text{eye}\leftarrow\text{head}} = \text{Softmax}\left(\frac{\mathbf{Q}_{\text{eye}}\mathbf{K}_{\text{head}}^{\text{T}}}{\sqrt{d_k}}\right), \quad \widetilde{X}_{\text{eye}} = \mathbf{A}_{\text{eye}\leftarrow\text{head}}\mathbf{V}_{\text{head}} \tag{2}$$

Head-updated cross-attention is defined symmetrically, yielding $\widetilde{X}_{\text{head}}$. The attention matrix encodes cross-modal dependency weights rather than mismatch itself; sensory conflict is captured implicitly by the updated features relative to the unimodal representations. When $\mathbf{Q}$ and $(\mathbf{K}, \mathbf{V})$ are derived from the same modality, the operation reduces to self-attention.

The updated features are integrated through residual connections:

$$X'_{\text{eye}} = X_{\text{eye}} + \widetilde{X}_{\text{eye}}, \quad X'_{\text{head}} = X_{\text{head}} + \widetilde{X}_{\text{head}} \tag{3}$$

The fused embedding used in subsequent layers is then formed from X'_{eye} and X'_{head}.

A sinusoidal positional encoding is then added to preserve temporal order, which helps distinguish transient discrepancies from accumulated symptoms. More generally, this choice reflects the importance of temporal structure in sequential neural processing without claiming a direct correspondence to specific biological cell types.

Overall, the embedding stage captures complementary visual and postural cues while providing a compact representation of cross-modal conflict for subsequent prediction.

State-Space Re-entrant Experts

The progression of cybersickness involves the long-horizon accumulation of sensory discrepancies. Modeling this process requires (i) a backbone that can integrate conflict signals over long sequences with computational efficiency and stability and (ii) a specialization mechanism that can capture heterogeneous symptom patterns across different conflict regimes. To this end, we propose State-Space Re-entrant Experts (SSR-Experts), which combines a Selective State Space Model (Gu

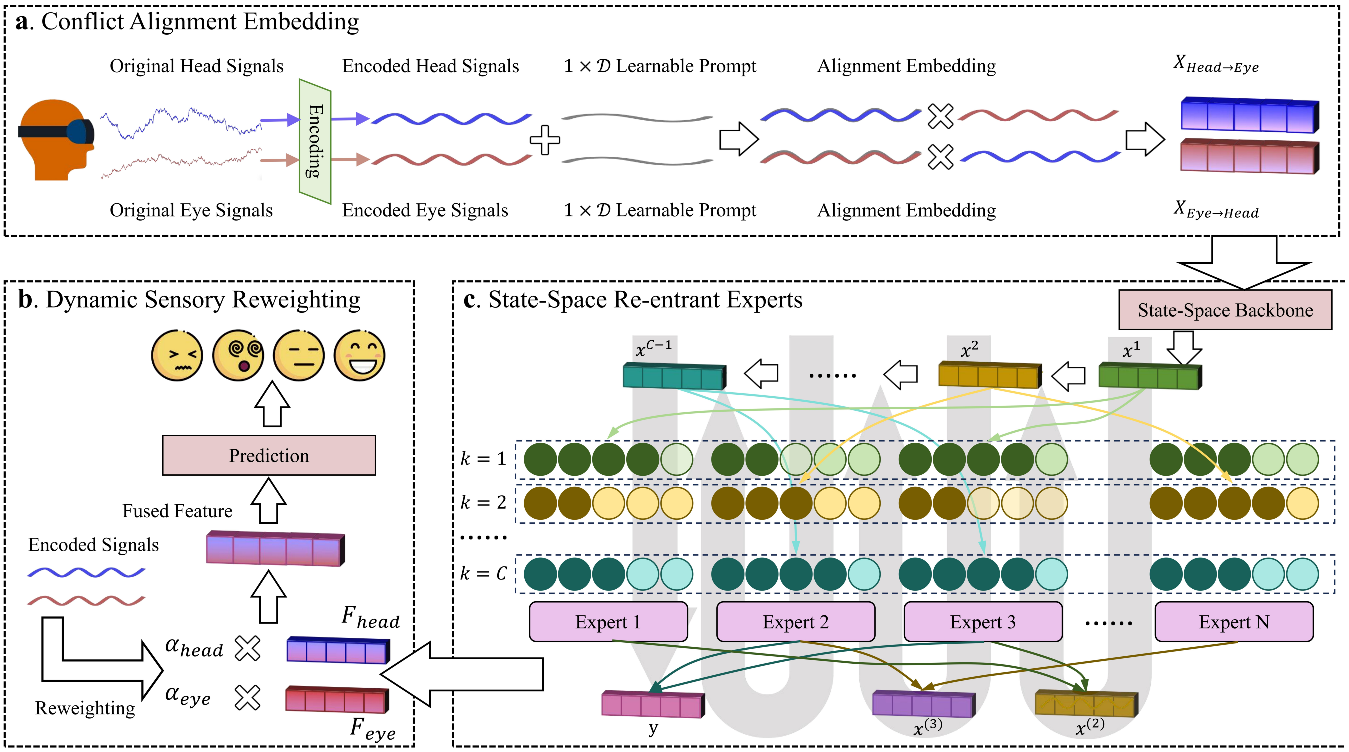

Figure 1: Overview of the NSCNet framework. (a) Conflict Alignment Embedding: aligns encoded head- and eye-motion signals through cross-modal attention with learnable prompts to extract conflict-aware features. (b) Dynamic Sensory Reweighting: computes modality weights ($\alpha_{head}, \alpha_{eye}$) to fuse features based on sequence-level reliability. (c) State-Space Re-entrant Experts: models temporal dynamics with a state-space backbone and iteratively refines symptom representations through gated experts; darker shading indicates higher selection probability.

& Dao, 2024) with Re-entrant Experts (Z. Wang et al., 2025), as illustrated in Figure 1(c).

State Space Backbone As shown in Figure 1(c), the conflict-aligned features first pass through a state-space backbone for continuous temporal integration. Given an input sequence $X \in \mathbb{R}^{B \times L \times d}$, we employ a Selective State Space Model that discretizes continuous dynamics into a stable recurrence, thereby compressing long-range sensory history into a finite latent state. Compared with RNN-style updates, which may forget distant context, and Transformer attention, which scales quadratically with sequence length, the state-space formulation supports linear-time modeling of long sequences while maintaining stable optimization over extended exposure sessions.

Intuitively, discretization enables the model to accumulate persistent conflict patterns while attenuating transient fluctuations. As a result, the backbone preserves conflict signals that are predictive of sickness buildup and suppresses short-lived artifacts or benign motion. The selective mechanism further provides an adaptive memory policy by modulating updates according to the current context, which is well suited to cybersickness because symptoms typically reflect sustained disagreement rather than isolated events.

Specifically, the state-space backbone uses a latent state dimension of $N = 8$ and an expansion factor of $E = 2$, projecting the input dimension $D = 128$ to an inner dimension of 256. We adopt the Zero-Order Hold (ZOH) discretization scheme to transform the continuous parameters $(\mathbf{A}, \mathbf{B})$ into discrete parameters $(\overline{\mathbf{A}}, \overline{\mathbf{B}})$. In the selective scan mechanism, the step size $\boldsymbol{\Delta}$ is input-dependent, making the discretization dynamic:

$$\begin{aligned} \overline{\mathbf{A}}_t &= \exp(\boldsymbol{\Delta}_t \mathbf{A}) \\ \overline{\mathbf{B}}_t &= (\boldsymbol{\Delta}_t \mathbf{A})^{-1} \left(\exp(\boldsymbol{\Delta}_t \mathbf{A}) - \mathbf{I} \right) \cdot (\boldsymbol{\Delta}_t \mathbf{B}) \end{aligned} \tag{4}$$

where $\mathbf{A} \in \mathbb{R}^{N \times N}$ is the structured diagonal state transition matrix initialized with HiPPO-LegS, $\mathbf{B} \in \mathbb{R}^{N \times D}$ is the input projection matrix, and $\mathbf{I}$ is the identity matrix.

Re-entrant Experts While the backbone provides global temporal continuity, manifestations of cybersickness can vary substantially across time and subjects, requiring repeated, context-dependent decoding of latent conflict states. We therefore introduce a Re-entrant Experts module, preceded by LayerNorm to stabilize latent activity and, crucially, to prevent representation drift over long sequences. In practice, this normalization reduces scale drift and makes routing decisions less sensitive to subject-specific amplitude differences.

As shown in Figure 1(c), the re-entrant mechanism performs C routing-and-refinement steps ($\mathbf{x}^0 \rightarrow \mathbf{x}^1 \rightarrow \cdots \rightarrow \mathbf{x}^C$) at each time step. This mirrors re-entrant cortical processing: information is iteratively revisited and reshaped, rather than decoded in a single feed-forward pass. Consequently, early steps can establish a coarse interpretation of the current conflict regime, while later steps progressively refine subtle symptom cues.

Operationally, we implement the Re-entrant Experts as a Chain of Experts (CoE). At each iteration k (where $k = 1, \ldots, C$), a Router produces a sparse gating vector over the experts ($E_1, \ldots, E_N$), allowing different experts to specialize in distinct conflict patterns (e.g., visual lag vs. rotational mismatch). For a given time step t, the routing is computed as:

$$\mathbf{r}_t^k = \text{Softmax}(\mathbf{W}_r \mathbf{x}_t^{k-1}), \quad \mathcal{G}_t^k = \text{TopK}(\mathbf{r}_t^k, K) \tag{5}$$

Here, $\mathbf{W}_r$ denotes the learnable routing weights, and TopK enforces sparse activation by selecting the indices of the top-2 most relevant experts ($K = 2$). This sparsity supports efficient inference and promotes functional specialization by repeatedly exposing each expert to a consistent subset of re-entrant contexts.

Each expert E_i is implemented as a Gated Feed-Forward Network (GFFN) with a SiLU nonlinearity to model gated, nonlinear transformations. The gating acts as an attention-like control of information flow, allowing the expert to amplify symptom-relevant components while suppressing irrelevant nuisance variation. The iterative update is defined as:

$$\mathbf{x}_t^k = \mathbf{x}_t^{k-1} + \sum_{i \in \mathcal{G}_t^k} r_{t,i}^k \cdot E_i(\mathbf{x}_t^{k-1}) \tag{6}$$

where $r_{t,i}^k$ represents the routing weight for expert i. The residual connection preserves the accumulated state from the backbone and earlier re-entrant steps, preventing the overwriting of long-term evidence and enabling stable refinement across iterations. In our implementation, we set the number of re-entrant steps to $C = 8$.

To prevent expert collapse and encourage balanced utilization, we additionally introduce an auxiliary load-balancing loss. This regularizer discourages the Router from assigning most inputs to a small subset of experts, thereby maintaining diversity and improving generalization across subjects and VR scenarios.

In summary, the state-space backbone enables efficient long-range accumulation of sensory conflict, while the Re-entrant Experts module provides iterative, context-dependent specialization for heterogeneous symptom patterns. By revisiting the latent state through multiple routing-and-refinement steps, the model can progressively disambiguate transient motion artifacts from sustained conflict evidence. Together, these components support both stable long-horizon integration and flexible re-entrant specialization, enabling precise decoding of how accumulated sensory mismatches evolve into cybersickness severity.

Dynamic Sensory Reweighting

During cybersickness progression, the specific contribution of visual and vestibular inputs to symptom accumulation varies significantly across different exposure scenarios and physiological states. The central nervous system addresses this via sensory reweighting, dynamically adjusting the gain of sensory channels based on their reliability. For instance, in scenarios with intense visual conflict, the brain may down-regulate visual cues to mitigate distress. From a modeling perspective, a static summation of features fails to capture this variability, potentially allowing a noisy modality to contaminate the fused representation.

To address this, we introduce the Dynamic Sensory Reweighting module, as shown in Figure 1(b). Unlike frame-by-frame gating, which can be sensitive to transient jitter, our approach determines the optimal modality dominance at the sequence level. We project the concatenated modality features into a latent space and apply Global Average Pooling to extract a holistic context vector. This ensures that the synaptic gains (modality weights) α_{eye} and α_{head} reflect the overall informativeness of the entire sequence rather than momentary fluctuations. These weights are normalized via a Softmax function, and the final fused representation is obtained by applying these two weights to the eye- and head-feature representations and then combining them.

This design enables the model to perform sensory-specific reweighting, effectively up-weighting the sensory pathway that provides the most consistent diagnostic value for the current trial while attenuating modality-specific noise.

Optimization Objectives

To decode cybersickness severity from latent neural representations, we train the model via a joint objective inspired by predictive coding, encouraging accurate estimation while avoiding inefficient routing. We optimize two losses:

Symptom Decoding Loss : We use Mean Squared Error (MSE) to align the prediction $\hat{y}_i$ with the ground-truth severity y_i:

$$\mathcal{L}_{\text{main}} = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2 \tag{7}$$

Homeostatic Regularization : To prevent expert collapse and encourage balanced CoE routing, we regularize the gating with an auxiliary loss:

$$\mathcal{L}_{aux} = \sum_{i=1}^{N} \bar{s}_i \cdot f_i \tag{8}$$

where $N = 8$ is the number of experts, $\bar{s}_i$ is the batch-averaged gating score for expert i, and f_i is its token routing fraction.

Therefore, the overall optimization objective is:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{main}} + \lambda_{aux} \cdot \mathcal{L}_{\text{aux}} \tag{9}$$

Here, $\lambda_{aux} = 0.01$ is a fixed hyperparameter that controls the trade-off between symptom-decoding accuracy and the network's homeostatic stability.

Table 1: Modalities and Extracted Features

Modalities	Extracted Features
Eye Movement	[From each of the left and right eyes] Eye openness Gaze direction (x, y, z) Gaze origin (x, y, z) Pupil diameter
Head Transform	Head position (x, y, z) Head rotation (yaw, pitch, roll)

Experiments and Results

Datasets

To evaluate the model's performance, we used two publicly available multimodal datasets: the MSCVR dataset released by Jeong and Han (2024) and the VR Cybersickness dataset proposed by Demirel et al. (2025). Although these datasets differ in acquisition devices and experimental objectives, both include relevant head- and eye-movement data. To ensure consistency and comparability across datasets, we selected 22 shared features: 16 eye-movement features and 6 head-movement features as the input to our model.

MSCVR Dataset MSCVR contains 2,700 labeled segments from 45 participants (aged 20–36 years; mean age 24.4; 24 male, 21 female). Participants had an average Motion Sickness Susceptibility Questionnaire (MSSQ) score of 10.0 (SD = 8.71), indicating typical motion-sickness sensitivity. In the experiment, participants watched 20 different 360° VR videos spanning 10 themes (45 s each), during which eye- and head-movement data were recorded. Participants verbally reported their Fast Motion Sickness Scale (FMS) scores every 15 s, yielding three labeled segments per video.

VR Cybersickness dataset VR Cybersickness includes 16 participants (aged 26–37 years; 10 male, 6 female). Unlike the passive viewing in MSCVR, participants performed an active VR riding task and continuously reported sickness levels (0–1) via a joystick. Although the dataset was not originally designed for eye-movement analysis, relevant eye- and head-movement data were collected. To match the MSCVR feature space, we excluded other modalities and extracted the same 16 eye-movement features and 6 head-movement features for modeling. All participants had prior VR experience and no history of severe motion sickness.

This unified feature space reduces inter-dataset modality differences and supports different datasets evaluation.

Table 2: Performance Comparison on MSCVR Dataset

Work	Method	MAE	RMSE
DeepLSTM	LSTM	0.79	1.26
Informer	Transformer	0.68	1.14
Mamba	StateSpaceModel	0.65	1.15
PhysioDNN	CNN-LSTM	0.54	0.38
HMDPrediction	CNN-LSTM	0.20	0.62
MAC	CNN-LSTM	0.17	0.28
MS-STTN	Transformer	0.12	0.25
AutoDetectionCS	cGAN-LSTM	0.56	1.05
NSCNet	**SSM-CoE**	**0.10**	**0.21**

Table 3: Performance on VR Cybersickness Dataset

Work	Method	MAE	RMSE
DeepLSTM	LSTM	0.80	1.10
Informer	Transformer	0.95	1.31
Mamba	StateSpaceModel	0.70	1.07
KinematicModel	Conv-LSTM	0.86	**0.40**
BeyondSubjectivity	Conv-LSTM	0.96	0.44
AutoDetectionCS	cGAN-LSTM	0.61	0.97
NSCNet	**SSM-CoE**	**0.29**	0.65

Implementation Details

Data processing and model training strictly followed each dataset's original protocol to ensure methodological fidelity and fair comparison. For MSCVR, raw sensor streams sampled at 90 Hz were downsampled to 25 Hz, and each 45 s trial was segmented into three non-overlapping 15 s windows; the original FMS scores (0–20) were used as regression targets. For VR Cybersickness, continuous joystick ratings (0–1) were retained as labels without linear remapping to preserve the original response granularity. Across both datasets, the shared 22-dimensional feature set (16 eye, 6 head) was normalized via min–max scaling. The model—comprising Conflict Alignment Embedding, SSRExperts, and Dynamic Sensory Reweighting—was trained with Adam (initial learning rate 1×10^{-4}) and a cosine annealing schedule for 128 epochs. To address target imbalance (and the differing label scales, 0–20 vs. 0–1), we applied inverse-frequency bin weights during training. Training runs on an NVIDIA RTX 4090 GPU with 24GB memory. Evaluation followed the standard benchmarks for each dataset: 5-fold subject-independent cross-validation for MSCVR and leave-one-subject-out (LOSO) cross-validation for VR Cybersickness.

Comparison with Current Methods

To validate the proposed method, we compared our approach with ten state-of-the-art methods trained on the same bimodal inputs (16 eye-tracking and 6 head-motion features) and evaluated under identical preprocessing and evaluation protocols. Specifically, Y. Wang et al. (2019) used a deep LSTM autoencoder (DeepLSTM), Zhou et al. (2021) proposed Informer, Gu and Dao (2024) proposed Mamba, Islam et al. (2020) developed a CNN–BiLSTM hybrid (PhysioDNN), Islam et al. (2021) used a 3D-CNN with a temporal CNN–LSTM (HMD-Prediction), Jeong et al. (2023) introduced MAC, Jeong and Han (2024) proposed MS-STTN, Yalcin et al. (2024) used a cGAN-augmented BiLSTM (AutoDetectionCS), Li et al. (2023) proposed KinematicModel, and Demirel et al. (2025) proposed BeyondSubjectivity.

Table 2 presents the results on the MSCVR dataset. Our proposed NSCNet (SSM-CoE) achieves the lowest error rates, with an MAE of 0.10 and an RMSE of 0.21. Notably, compared with the vanilla Mamba model (MAE 0.65), our method achieves a substantial performance gain. Moreover, our model outperforms traditional hybrid methods (e.g., PhysioDNN and MAC), supporting our hypothesis that global temporal integration, together with functional modularity via the Chain of Experts, is crucial for resolving the complex nonlinear dynamics of sensory mismatches.

Table 3 presents the results on the VR Cybersickness dataset, further demonstrating the generalizability of our bio-inspired approach: our model achieves a SOTA MAE of 0.29, substantially outperforming KinematicModel (MAE 0.86) and DeepLSTM (MAE 0.80). While some baselines show competitive RMSE values, our model maintains a favorable balance between average prediction accuracy and stability, effectively reducing large deviations in symptom estimation.

In summary, NSCNet's consistent superiority across both datasets confirms that modeling cybersickness as a process of accumulated neural dissonance, captured via continuous temporal integration and specialized cortical routing, yields a more accurate physiological representation than purely data-driven deep learning paradigms.

Ablation Study

To assess the biological plausibility of our framework, we conducted a systematic ablation study on the MSCVR and VR Cybersickness datasets. The results show that disabling any bio-inspired component leads to a significant degradation in symptom decoding fidelity. Specifically, removing the visual or vestibular inputs causes a sharp performance drop (MAE rising to 1.39 and 0.76, respectively), consistent with Sensory Conflict Theory, which posits that cybersickness emerges from dissonance between these streams. Furthermore, excluding the Conflict Alignment Embedding impairs the model's ability to simulate sensory recalibration, whereas deactivating the State-Space Re-entrant Experts (Mamba or Chain of Experts) disrupts continuous temporal integration and functional modularity, thereby preventing accurate tracking of accumulated symptoms. Finally, the inferior performance of static fusion strategies ('w/o Reweighting') supports the necessity of Dynamic Sensory Reweighting to mimic the central nervous system's adaptive gain control. Overall, the superior performance of the full architecture (MAE 0.10 and 0.29) confirms that synergizing sensory alignment, temporal accumulation, and cortical specialization is essential for accurately decoding the physiological mechanisms of cybersickness.

Table 4: Ablation Study Results

Model Variants	MSCVR Dataset		VR Cybersickness Dataset	
	MAE	RMSE	MAE	RMSE
w/o Head feat	0.24	0.74	0.76	1.28
w/o Eye feat	0.47	0.96	1.39	1.49
w/o Embedding	0.23	0.65	0.51	0.94
w/o Backbone	0.18	0.51	0.37	0.86
w/o CoE	0.25	0.73	0.63	1.02
w/o Reweighting	0.17	0.46	0.32	0.75
Full NSCNet	**0.10**	**0.21**	**0.29**	**0.65**

Conclusion

In this paper, we proposed NSCNet, a biologically inspired multimodal framework for cybersickness prediction, which explicitly models the accumulation of neural dissonance between visual and vestibular streams. By integrating the Conflict Alignment Embedding for sensory recalibration, State-Space Re-entrant Experts for continuous temporal integration and cortical modularity, and Dynamic Sensory Reweighting for adaptive neural gain control, our approach achieves state-of-the-art performance on the MSCVR and VR Cybersickness datasets (MAE of 0.10 and 0.29). These results confirm that incorporating biologically motivated constraints yields superior accuracy and interpretability compared with purely data-driven models. Future work will focus on validating the learned expert modules against neural correlates from neuroimaging data and extending the framework to closed-loop neurofeedback systems to proactively mitigate sensory conflict before symptoms manifest.

Acknowledgment

This work is supported by Beijing High-level Overseas Talent Returning Grant, NCUT Research Startup Fund, NCUT Education Research and Reform Project, and Undergraduate Training Program for Innovation and Entrepreneurship.

References

Demirel, B. U., Dogan, A. H., Rossie, J., Möbus, M., & Holz, C. (2025). Beyond Subjectivity: Continuous Cybersickness Detection Using EEG-based Multitaper Spectrum Estimation. *IEEE Transactions on Visualization and Computer Graphics*, *31*(5), 2525–2534. https://doi.org/10.1109/TVCG.2025.3549132

Gu, A., & Dao, T. (2024, May 31). *Mamba: Linear-Time Sequence Modeling with Selective State Spaces*. arXiv: 2312.00752 [cs]. https://doi.org/10.48550/arXiv.2312.00752

Islam, R., Desai, K., & Quarles, J. (2021). Cybersickness Prediction from Integrated HMD's Sensors: A Multimodal Deep Fusion Approach using Eye-tracking and Head-tracking Data. https://doi.org/10.1109/ISMAR52148.2021.00017

Islam, R., Lee, Y., Jaloli, M., Muhammad, I., Zhu, D., Rad, P., Huang, Y., & Quarles, J. (2020). Automatic Detection and Prediction of Cybersickness Severity using Deep Neural Networks from user's Physiological Signals. *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 400–411. https://doi.org/10.1109/ISMAR50242.2020.00066

Jeong, D., & Han, K. (2024). PRECYSE: Predicting Cybersickness using Transformer for Multimodal Time-Series Sensor Data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *8*(2), 1–24. https://doi.org/10.1145/3659594

Jeong, D., Paik, S., Noh, Y., & Han, K. (2023). MAC: Multimodal, attention-based cybersickness prediction modeling in virtual reality. *Virtual Real.*, *27*(3), 2315–2330. https://doi.org/10.1007/s10055-023-00804-0

Jin, M., Wang, S., Ma, L., Chu, Z., Zhang, J. Y., Shi, X., Chen, P.-Y., Liang, Y., Li, Y.-F., Pan, S., & Wen, Q. (2024, January 29). *Time-LLM: Time Series Forecasting by Reprogramming Large Language Models*. arXiv: 2310.01728 [cs]. https://doi.org/10.48550/arXiv.2310.01728

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, *3*(3), 203–220. https://doi.org/10.1207/s15327108ijap0303_3

Keshavarz, B., Peck, K., Rezaei, S., & Taati, B. (2022). Detecting and predicting visually induced motion sickness with physiological measures in combination with machine learning techniques. *International Journal of Psychophysiology*, *176*, 14–26. https://doi.org/10.1016/j.ijpsycho.2022.03.006

LaViola, J. J. (2000). A discussion of cybersickness in virtual environments. *SIGCHI Bull.*, *32*(1), 47–56. https://doi.org/10.1145/333329.333344

Li, R., Wang, Y., Yin, H., Chardonnet, J.-R., & Hui, P. (2023). A Deep Cybersickness Predictor through Kinematic Data with Encoded Physiological Representation. *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 1132–1141. https://doi.org/10.1109/ISMAR59233.2023.00130

Lukacova, I., Keshavarz, B., & Golding, J. F. (2023). Measuring the susceptibility to visually induced motion sickness and its relationship with vertigo, dizziness, migraine, syncope and personality traits. *Experimental Brain Research*, *241*(5), 1381–1391. https://doi.org/10.1007/s00221-023-06603-y

Luo, C., Xie, Y., & Yu, Z. (2024, September 18). *PhysMamba: Efficient Remote Physiological Measurement with SlowFast Temporal Difference Mamba*. arXiv: 2409.12031 [cs]. https://doi.org/10.48550/arXiv.2409.12031

Nashner, L., & Berthoz, A. (1978). Visual contribution to rapid motor responses during postural control. *Brain Research*, *150*(2), 403–407. https://doi.org/10.1016/0006-8993(78)90291-3

Reason, J. T., & Brand, J. J. (1975). *Motion sickness*. Academic Press.

Riccio, G., & Stoffregen, T. (1991). An Ecological Theory of Motion Sickness and Postural Instability. *Ecological Psychology - ECOL PSYCHOL*, *3*, 195–240. https://doi.org/10.1207/s15326969eco0303_2

Sameri, J., Coenegracht, H., Van Damme, S., De Turck, F., & Torres Vega, M. (2024). Physiology-driven cybersickness detection in virtual reality: A machine learning and explainable AI approach. *Virtual Real.*, *28*(4). https://doi.org/10.1007/s10055-024-01067-z

Stanney, K. M., Hale, K. S., Nahmens, I., & Kennedy, R. S. (2003). What to Expect from Immersive Virtual Environment Exposure: Influences of Gender, Body Mass Index, and Past Experience. *Human Factors*, *45*(3), 504–520. https://doi.org/10.1518/hfes.45.3.504.27254

Tian, N., & Boulic, R. (2024). Who says you are so sick? An investigation on individual susceptibility to cybersickness triggers using EEG, EGG and ECG. *IEEE transactions on visualization and computer graphics*, *30*(5), 2379–2389. https://doi.org/10.1109/TVCG.2024.3372066

Tsai, Y.-H. H., Bai, S., Liang, P. P., Kolter, J. Z., Morency, L.-P., & Salakhutdinov, R. (2019). Multimodal Transformer for Unaligned Multimodal Language Sequences. In A. Korhonen, D. Traum, & L. Màrquez (Eds.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 6558–6569). Association for Computational Linguistics. https://doi.org/10.18653/v1/P19-1656

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All You Need*. arXiv: 1706.03762 [cs]. https://doi.org/10.48550/arXiv.1706.03762

Wang, Y., Chardonnet, J.-R., & Merienne, F. (2019). VR Sickness Prediction for Navigation in Immersive Virtual Environments using a Deep Long Short Term Memory Model. *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 1874–1881. https://doi.org/10.1109/VR.2019.8798213

Wang, Z., Pan, R., Yao, J., Csordas, R., Li, L., Yin, L., Wu, J., Zhang, T., Li, M., & Liu, S. (2025, June 23). *Chain-of-Experts: Unlocking the Communication Power of Mixture-of-Experts Models*. arXiv: 2506.18945 [cs]. https://doi.org/10.48550/arXiv.2506.18945

Yalcin, M., Halbig, A., Fischbach, M., & Latoschik, M. E. (2024). Automatic cybersickness detection by deep learning of augmented physiological data from off-the-shelf consumer-grade sensors. *Frontiers in Virtual Reality*, *5*. https://doi.org/10.3389/frvir.2024.1364207

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(12), 11106–11115. https://doi.org/10.1609/aaai.v35i12.17325