One Size Does Not Fit All: Personalized Cybersickness Forecasting with Asymmetric Few-Shot Meta-Learning

Nasim Ahmed* Kennesaw State University
Peng Wu† Northeastern University
Kaiming Huang‡ The Pennsylvania State University
Tian Lan§ George Washington University
Bin Li¶ The Pennsylvania State University
Gang Tan‖ The Pennsylvania State University
Mahdi Imani** Northeastern University
Rifatul Islam†† Kennesaw State University

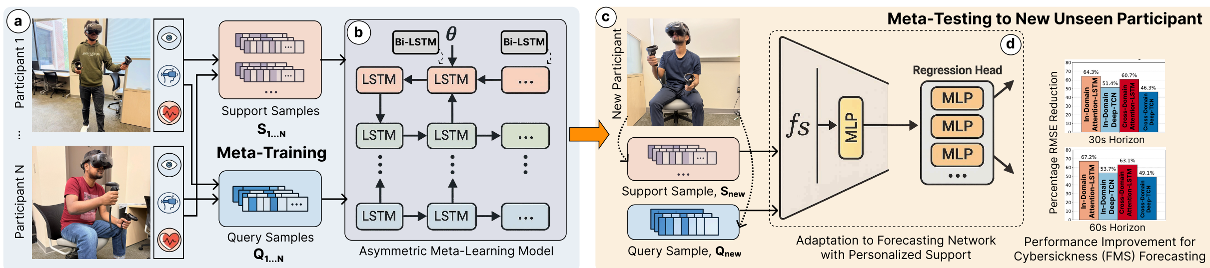

Figure 1: Overview of the proposed few-shot meta learning pipeline for personalized cybersickness forecasting. (a) The model is first trained using data from a set of participants. (b) An asymmetric temporal architecture learns general patterns from the training data. (c) The model then adapts to a new participant using only a few support samples. (d) The adapted model forecasts the user's future cybersickness level from multi-modal sensor data.

Abstract

Cybersickness remains a critical barrier to the widespread adoption of virtual reality(VR), making accurate predictions of its onset a key research priority to address the comfortable use of VR systems. However, the onset of cybersickness is inherently individualized, causing existing models trained on population-level data to generalize poorly to unseen and new users. Moreover, most current approaches detect or classify cybersickness only after symptoms manifest, rather than forecasting how discomfort will evolve, which is essential for applying timely mitigation strategies. To address these research gaps, we propose an asymmetric few-shot meta-learning approach that employs asymmetric temporal encoding to forecast cybersickness severity 30 and 60 seconds into the future using multimodal data (heart rate, eye tracking, and head tracking). Our framework supports efficient personalization using as little as one 30-second calibration window. We validated our approach on two public datasets: a real-walking VR task (Maze, 36 participants) and seated VR tasks (Simulation21, 30 participants). We evaluated our proposed models under in-domain performance and cross-domain transfer between walking and seated postures. Compared to a state-of-the-art DeepTCN baseline model, our approach achieves a 53.7% RMSE reduction in the in-domain setting on the Maze dataset and a 49.1% RMSE reduction under posture transfer to seated environments at the 60 second forecast horizon. We also performed memory and computational profiling of the proposed models, which shows an inference latency of 2.5 milliseconds and a memory footprint under 1.0 MB, demonstrating strong potential for deployment on standalone VR headsets. In summary, our results establish a practical foundation for personalized cybersickness forecasting that can support proactive mitigation strategies in future VR systems and improve user comfort and accessibility across diverse immersive experiences.

Index Terms: Cybersickness, Personalization, Few-shot Adaptation, Limited Data, Meta Learning

1 Introduction

One of the major barriers to the widespread usability of virtual reality (VR) experiences is cybersickness, a condition marked by nausea, dizziness, oculomotor strain, and disorientation [10, 16, 21]. Traditional measurement techniques rely on subjective questionnaires administered after a VR session, such as the simulator sickness questionnaire (SSQ) [28], which cannot capture the real-time onset of symptoms [14]. Moreover, most current machine learning approaches detect or classify cybersickness only after symptoms have manifested, rather than forecasting how discomfort will evolve, which is essential for applying timely mitigation strategies and preventing early session termination [22, 41]. However, forecasting cybersickness onset is inherently challenging because individual susceptibility and sensory conflict thresholds vary widely across users [43, 44]. While recent supervised learning methods attempt user-level forecasting, they rely on offline training paradigms that introduce a severe data scarcity problem, as personalizing these models demands substantial labeled, user-specific data for retraining [22, 60, 63, 65]. This requirement is infeasible in consumer VR settings where extensive calibration cannot be expected of every new user. Moreover, existing personalized approaches, such as those proposed by Tasnim et al. [63], require full model retraining and are typically evaluated within a single domain, overlooking cross-domain variability. Different VR interaction paradigms and locomotion styles, such as seated experiences versus natural walking, fundamentally alter sensory conflict dynamics and user susceptibility [20, 73].

To address these gaps, we propose an asymmetric few-shot meta-learning model that personalizes cybersickness forecasting using

*e-mail: nahmed25@students.kennesaw.edu
†e-mail: wu.p@northeastern.edu
‡e-mail: kzh529@psu.edu
§e-mail: tlan@gwu.edu
¶e-mail: binli@psu.edu
‖e-mail: gtan@psu.edu
**e-mail: m.imani@northeastern.edu
††e-mail: rislam11@kennesaw.edu

30-second windows of multimodal data, including eye tracking, head pose, pupillometry, and physiological signals such as heart rate and galvanic skin response. In our continuous time series forecasting setting, few-shot learning refers to rapid user adaptation from a small number of temporal support windows, rather than its conventional image classification meaning. Furthermore, meta-learning describes our episodic training procedure, which learns a generalized physiological baseline from which the model can rapidly adapt to a novel user with minimal data. Following prior work [23, 30], we adopt 30- and 60-second forecasting horizons, as this range provides sufficient lead time for comfort-preserving interventions, such as reducing visual motion speed or narrowing the field of view, before severe discomfort develops. Our proposed episodic meta-optimization strategy learns a generalized representation of the physiological precursors to discomfort, enabling the model to generalize across entirely different VR interaction paradigms while utilizing only a minimal support set of recent sensory history to instantly personalize to a novel user's specific conflict threshold. We validate this approach on two open-source datasets representing distinct VR paradigms: the Maze dataset [57], featuring time-synchronized multimodal data from 36 participants in real-walking VR navigation, and the Simulation21 dataset [23], containing multimodal data from 30 participants in seated VR simulations. **The key contributions of this work are, but not limited to:**

1. We introduce an asymmetric few-shot meta-learning framework for personalized cybersickness forecasting that adapts to new users using only 30 seconds of calibration data, eliminating the need for per-user retraining.

2. We propose an asymmetric temporal encoder that separately processes historical calibration signals and live input streams, learning a personalized physiological baseline while preserving temporal structure for a 30-60s forecasting window from multimodal data.

3. We establish a cross-domain evaluation protocol across two publicly available datasets with distinct VR interaction paradigms, providing evidence that few-shot meta-learning can generalize cybersickness forecasting across locomotion styles without offline retraining.

4. We conduct computational profiling to assess deployment feasibility on standalone VR headsets, bridging the gap between research-stage forecasting models and practical consumer VR integration.

In summary, our proposed model achieves a 53.7% reduction in RMSE for in-domain forecasting on the Maze dataset and a 49.1% reduction under cross-domain transfer between the Maze and Simulation21 datasets at the 60-second forecasting horizon compared to the state-of-the-art DeepTCN baseline, while maintaining an inference latency of 2.5 milliseconds and a memory footprint below 1.0 MB. These findings indicate that personalized modeling can enable earlier detection of cybersickness risk and support more responsive comfort management in immersive VR experiences.

2 Related Work

In this section, we review related work on cybersickness prediction and forecasting, individual variability, and cross-domain generalization in VR. We also discuss recent advances in few-shot and meta-learning for subjective state forecasting that motivate our approach.

2.1 Cybersickness

Cybersickness is a motion-sickness-like condition experienced in VR environments, characterized by nausea, dizziness, oculomotor strain, and disorientation [37]. Although its exact cause remains debated, the most widely cited explanation is the sensory conflict theory, which states that cybersickness arises from mismatches between visual motion cues and vestibular input [26]. Another explanation, the postural instability theory, attributes it to difficulty maintaining balance under unfamiliar visual conditions [61]. Beyond these theoretical accounts, cybersickness severity is influenced by system-level factors such as latency, low frame rate, and tracking errors, content and interaction design choices such as artificial locomotion and rapid viewpoint changes, and individual variability across users [6, 49]. This complex interplay of hardware, content, and user-specific factors makes cybersickness particularly difficult to predict and mitigate, and underscores the need for adaptive approaches that can account for these diverse sources of variability.

2.2 Individual Factors of Cybersickness

Cybersickness susceptibility varies considerably across individuals due to demographic and physiological factors [15, 43, 44, 56]. For instance, studies involving diverse racial groups found that Black participants generally reported lower cybersickness levels than White participants, while no significant difference was observed between Asian and White groups [43]. Gender also plays a role; several studies report higher susceptibility among females, potentially linked to differences in interpupillary distance and motion sickness history, although findings remain inconsistent [33, 40, 44]. Similarly, prior VR experience tends to reduce symptoms, while the effect of age remains inconclusive across studies [2, 4, 17, 33]. Beyond demographics, individual differences in sensory processing, visual perception, and postural control further modulate how users respond to motion cues in immersive environments [66]. Variations in head movement patterns, gaze behavior, and physiological responses have also been shown to contribute to differences in cybersickness severity [3]. Moreover, these individual factors interact with environmental and system-level variables such as display characteristics and interaction methods, compounding the variability in user experience [59]. Collectively, these findings underscore the necessity of personalized approaches for cybersickness forecasting.

2.3 Cross-Domain Generalization in Cybersickness

Most cybersickness prediction and forecasting models are developed and evaluated within a single VR environment, limiting their generalizability to new settings [22, 23, 57, 68]. In such single-domain settings, models tend to overfit to environment-specific motion patterns, interaction dynamics, and stimulus characteristics that do not transfer across VR applications. Although Tasnim et al. explored personalization strategies to improve individual prediction performance, their evaluation remained confined to a single domain [63], failing to capture the broader variability of cybersickness responses across different VR contexts [73]. Cybersickness severity can vary substantially depending on visual motion intensity, scene complexity, and locomotion technique. For instance, natural walking, controller-based movement, and teleportation each introduce different levels of sensory conflict and perceptual load, leading to divergent cybersickness outcomes [65]. Furthermore, differences in hardware configuration, display properties, and user posture, such as seated versus standing or walking, contribute additional domain shifts between VR applications [74]. These factors highlight the need to evaluate forecasting models under heterogeneous VR conditions rather than relying on single-domain training data [54]. To address this limitation, we evaluate cross-domain generalization by testing cybersickness forecasting across both seated VR experiences and active walking-based VR navigation scenarios.

2.4 Cybersickness Measurement and Forecasting

Cybersickness is commonly assessed via subjective self-reports, such as the Simulator Sickness Questionnaire (SSQ) [28] and the Fast Motion Sickness (FMS) scale [29]. These measures summarize symptom severity [18, 32]. However, because they are collected at discrete time points during or after exposure, they cannot fully represent the continuous onset and dynamic fluctuations of symptoms during VR experiences [7, 14], motivating efforts to complement questionnaires with continuous sensing [52]. To address this, recent research employs data-driven forecasting with multimodal behavioral and physiological signals to estimate cybersickness as the experience unfolds [22, 23]. Prior studies show that eye tracking, head motion, and physiological responses contain predictive information [5, 8], leading to multimodal sensor fusion strategies that improve prediction robustness across diverse VR scenarios [42, 57]. For example, Islam et al. combined these modalities with deep neural networks to forecast cybersickness at short horizons [22]. Subsequent systems such as LiteVR [36] and TruVR [35] further moved this line of work toward lightweight, interpretable, and deployment-oriented prediction pipelines suitable for real-time applications [34]. More recent efforts explore probabilistic frameworks, such as Bayesian Networks, to model relationships among environmental factors, physiological responses, and FMS outcomes [69, 70]. Although these studies show strong progress, most existing approaches remain population-level and fail to adapt to individual variability in cybersickness onset, progression, and tolerance [1, 11, 19]. This motivates research on personalized, data-efficient forecasting strategies for immersive systems [63].

2.5 Few-Shot Meta-Learning for Time-Series Forecasting

Few-shot meta-learning addresses classification and regression problems where only a limited number of labeled examples are available [58]. Instead of relying on large datasets, these methods learn to rapidly adapt to new tasks using a small set of labeled support samples, making them particularly suited to domains where data collection is costly [48]. In the context of cybersickness forecasting, each user represents a new task with only a few labeled temporal segments available for adaptation, which naturally motivates a few-shot formulation. Although most few-shot research has focused on computer vision and recognition tasks, there is growing interest in applying these techniques to time-series forecasting [71, 72], where models must adapt from limited data while capturing temporal dependencies. Applications to subjective state forecasting remain scarce, as these tasks typically involve multimodal time-series data from physiological, behavioral, and environmental sources, adding considerable complexity. Nevertheless, early results in related domains [39, 45] show that few shot meta learning generalizes across individuals and contexts without requiring massive datasets.

3 Methodology

This section presents the proposed asymmetric few-shot meta-learning framework for personalized cybersickness forecasting, covering the datasets, preprocessing pipeline, model architecture, training strategy, and evaluation protocol.

3.1 Datasets

We evaluate our framework on two open-source, multimodal VR datasets representing distinct interaction paradigms: Maze [57], which involves real-walking navigation, and Simulation21 [23], which consists of seated VR experiences. The contrast between these locomotion-based and seated conditions supports the evaluation of both in-domain personalization and cross-posture domain transfer to new participants[1], addressing a key challenge in cybersickness forecasting where differences in locomotion style and interaction modality are known to significantly influence symptom severity [55, 65, 67, 73]. Figure 2 shows example visual scenes from the evaluated VR environments, including the real walking Maze navigation task and the seated Simulation21 tasks.

[1]The authors of both datasets confirmed that there is no participant overlap between the two datasets.

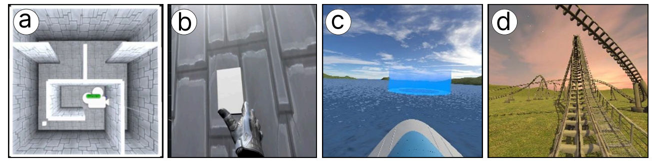

Figure 2: Example VR scenes from the two datasets used in this study. Panels (a) and (b) show the Maze dataset, including a top-down view of a randomly generated virtual maze and a first-person view where a participant grabs the trophy to begin another maze task. Panels (c) and (d) show the Sea and Roller seated VR environments from the Simulation21 dataset, respectively.

3.1.1 Maze Dataset

The Maze dataset, introduced by Nag et al. [57], comprises time-synchronized multimodal data collected from 36 participants (mean age: 25.2 years, SD: 4.9) during a natural walking VR task. Participants navigated a narrow virtual maze while completing attention and working memory tasks under three difficulty conditions, using an HTC Vive Pro Eye headset and a Shimmer physiological device. The dataset includes physiological signals such as heart rate(HR) and galvanic skin response(GSR), eye-tracking, pupil size, gaze direction, openness, and head-tracking data (rotations and position vectors). Cybersickness severity was recorded using FMS scores ranging from 1 to 7, reported every 30 seconds throughout each session.

3.1.2 Simulation21 Dataset

The Simulation21 dataset [23] contains multimodal data from 30 seated participants (mean age: 30.04 years, SD: 4.12). In this study, we selected two seated simulations: Sea Ride and Roller Coaster. In Sea Ride, participants navigated a virtual boat around an island using a handheld controller, steering through a series of checkpoints with full control over the boat's movement. In Roller Coaster, participants passively experienced repeated 57-second ride cycles followed by 10-second pauses, with no control over the motion of the ride. These seated simulations provide a clear contrast with the walking-based Maze dataset, enabling evaluation of cross-domain adaptation between locomotion-based and seated VR experiences. The dataset includes synchronized physiological signals (HR, GSR) and eye-tracking data(pupil size, gaze direction, openness), with cybersickness measured using the FMS scale, ranging from 0 to 10, reported every 30 seconds throughout each session.

3.2 Dataset Pre-processing

We applied several steps to prepare the Maze and Simulation21 datasets for few-shot cybersickness forecasting to ensure signal quality, cross-dataset compatibility, and well-structured input-output sequences for model training and evaluation.

Noise Reduction. The multimodal signals (HR, GSR, eye tracking, and head tracking) contained temporal noise from motion artifacts and sensor drift. We first removed outliers using z-score filtering, excluding values beyond three standard deviations [51], then applied a rolling average with a window size of 5 [22, 63]. To smooth residual noise without distorting short-term trends critical for forecasting, we further applied a Savitzky-Golay filter [53], which fits a low-degree polynomial to a moving window of the data:

$$\hat{x}_t = \mathrm{SGFilter}(x_{t-w:t+w}) \tag{1}$$

where x_t is the raw signal at time t and w is the half-window size for polynomial fitting (set to $w = 7$). Specifically, the Savitzky-Golay filter utilized a polynomial degree of 3. We set $w = 7$, corresponding to a 15-s Savitzky–Golay window after 1 Hz resampling, to provide a practical trade-off between noise suppression and temporal fidelity. This window is long enough to attenuate short-term sensor jitter and motion artifacts, yet remains short relative to our 30-s input window and 30/60-s forecasting horizons, thereby preserving the local dynamics that are informative for cybersickness forecasting [47, 53].

Normalization. To account for inter-participant variation in scale and baseline, we applied z-score normalization independently to each participant's data:

$$x_t^{(norm)} = \frac{x_t - \mu}{\sigma} \tag{2}$$

where μ and σ are the per-feature mean and standard deviation computed for each participant.

Consistent Sampling. To align differing recording rates across datasets ranging from 10 Hz to 60 Hz, we downsampled all signals to 1 Hz using one-second interval averages, preserving dominant temporal patterns while enabling consistent multimodal analysis.

FMS Scale Alignment. Subjective cybersickness ratings are prone to individual rating biases and differ in scale boundaries across datasets (1–7 in Maze vs. 0–10 in Simulation21), making simple linear scaling insufficient for cross-domain evaluation. To address this, we applied a psychometric scaling approach using the Subjective Quality Evaluation SUREAL framework [46]. First, raw scores were mapped to a unit interval. Second, the SUREAL maximum likelihood estimation model resolved each rating into a latent user sickness state and an individual scaling bias parameter. Third, we applied pooled quantile mapping to align these bias-corrected scores onto a unified reference distribution, yielding a shared 0 to 10 severity scale. This mapping ensures that identical numeric values across both datasets represent equivalent levels of underlying physiological discomfort. A Mann-Whitney U test comparing the unified severity distributions of Maze and Simulation21 confirmed no statistically significant difference ($p > 0.05$, as reported in Table 1), supporting the validity of our cross-domain evaluation. We applied this SUREAL scaling as an isolated preprocessing step before any model training. Target test labels were not used for training, tuning model selection, or alignment optimization, ensuring no target domain leakage.

3.3 Ground Truth Construction

We transform both datasets into structured input-output sequences using a sliding window approach. Each input sequence spans 30 consecutive time steps of sensor data, capturing physiological signals (heart rate, GSR), eye-tracking features (gaze direction, pupil diameter), and head-tracking data (quaternions and Euler angles). A 30-second window was selected to balance capturing meaningful temporal dynamics with processing efficiency, consistent with prior work showing that ultra-short windows ($\leq$ 30s) deliver competitive performance while preserving responsiveness in streaming scenarios [22, 64]. This choice also aligns with the FMS reporting interval used in both datasets. Each window is paired with two future labels corresponding to 30-second and 60-second forecasting horizons:

$$y_i^{(30)} = \text{FMS}(t_i + 30), \quad y_i^{(60)} = \text{FMS}(t_i + 60) \tag{3}$$

These horizons provide sufficient lead time for comfort-preserving interventions, such as reducing visual motion speed or narrowing the field of view, before severe discomfort develops [23, 30]. Shorter horizons leave insufficient time for system response, while longer horizons lose reliability as user state and motion dynamics shift rapidly during VR interaction [30]. All sequences are non-overlapping in terms of timestamps and participants. For the Simulation21 dataset, sequences were generated separately for each simulation (Roller Coaster and Sea Ride) to ensure fair comparison and rigorous evaluation across distinct VR environments.

3.4 Asymmetric Few-Shot Learning for Cybersickness

We adapt established sequence models and episodic meta-learning formulations to the practical constraints of real-time VR forecasting. Our architectural novelty lies not in introducing a generic replacement for standard sequence networks, but in designing a specific asymmetric temporal configuration, as illustrated in Figure 3. We process historical support windows bidirectionally to estimate a user-specific physiological baseline, while preserving causal forward only processing for live query windows to guarantee low delay streaming inference. In real-world VR deployments, the available data exists in two fundamentally different temporal states: offline historical data (support set) and live-streaming causal data (query set). Given a support set $\mathscr{S} = \{(x_n, y_n)\}_{n=1}^{N_s}$ of fully observed eye-tracking, head-tracking, and physiological sequences from a small number of prior VR exposures, our goal is to forecast future cybersickness levels for new live-streaming query sequences $x^* = [x_1^*, \ldots, x_T^*]$ from the same user, where $x_t^* \in \mathbb{R}^d$ represents the d-dimensional multimodal feature vector at timestep t.

Each sequence $x_n = [x_{n,1}, \ldots, x_{n,T_n}]$ in the support set has length T_n and is associated with a target cybersickness score y_n measured on the FMS scale. Our asymmetric model learns to construct a forecasting function $f_{\mathscr{S}} : \mathbb{R}^{d \times T} \to \mathbb{R}$ that outputs forecast cybersickness levels $\hat{y}_{T+h}^*$ at forecasting horizons $h \in \{30, 60\}$ seconds. These prediction horizons provide enough advance time for systems to apply comfort-oriented adjustments, as evident by the previous research [23, 30]. These adjustments include reducing visual motion or gradually narrowing the field of view, before noticeable discomfort emerges [27, 65]. Shorter horizons provide limited time for the system to respond, while longer horizons tend to be less stable because user state and motion patterns may change during ongoing VR interaction [12].

3.4.1 Bidirectional Support Encoding

The support set consists of fully observed, historical sequences from a user that are provided to the model for rapid adaptation. Since these sequences are available in their entirety, we encode each support sequence x_n using a bidirectional LSTM to capture both forward and backward temporal dependencies within the window, extracting the richest possible personalized context from limited samples:

$$\overrightarrow{h}_{n,t} = \overrightarrow{f}(\overrightarrow{h}_{n,t-1}, x_{n,t}; \theta_f) \tag{4}$$

$$\overleftarrow{h}_{n,t} = \overleftarrow{f}(\overleftarrow{h}_{n,t+1}, x_{n,t}; \theta_b) \tag{5}$$

$$h_{n,t} = [\overrightarrow{h}_{n,t}; \overleftarrow{h}_{n,t}] \tag{6}$$

where $\overrightarrow{f}$ and $\overleftarrow{f}$ are the forward and backward LSTM functions with parameters θ_f and θ_b, respectively. The concatenated hidden state $h_{n,t} \in \mathbb{R}^{2K_h}$ represents the n-th support sequence at timestep t, with $K_h = 32$ as the hidden dimension per direction.

3.4.2 Causal Query Encoding

Query sequences represent live-streaming sensor data from which the model must forecast future cybersickness. Since future observations are unavailable at inference time, we employ a unidirectional LSTM encoder to preserve the causal structure of the temporal sequence, ensuring that only past observations influence the current representation:

$$z_t = g(z_{t-1}, x_t^*; \theta_q) \tag{7}$$

$$z = z_T \tag{8}$$

where g is the query LSTM function with parameters θ_q, and $z \in \mathbb{R}^{K_z}$ is the final query encoding with $K_z = 32$ hidden units.

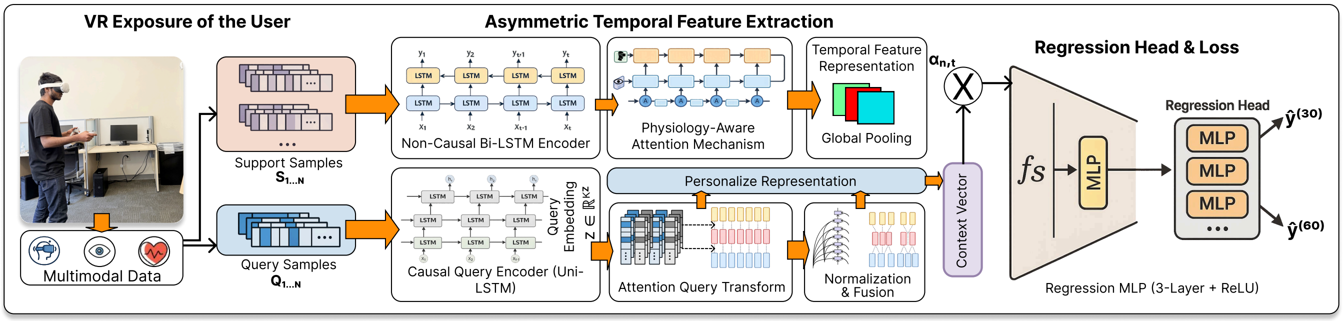

Figure 3: Architecture of the proposed asymmetric few-shot meta learning framework for personalized cybersickness forecasting. Multimodal VR signals, including eye tracking, head tracking, and physiological data, are segmented into support and query sequences. Support samples are encoded using a bidirectional LSTM to capture temporal dependencies, while query samples are processed using a causal LSTM for inference. A physiology gaze attention module extracts a personalized context representation that conditions the query embedding. The combined representation is passed to a regression head to forecast cybersickness at 30-second and 60-second horizons.

3.4.3 Physiology-Aware Attention

Cybersickness triggers vary substantially across individuals and VR domains. To dynamically adapt to a user's physiological baseline, we design a physiology-aware attention mechanism that computes similarity weightings between the user's current sensor state (the query embedding) and their historical physiological patterns (the support representations), rather than treating all historical samples equally:

$$e_{n,t} = v^T \tanh(W_s h_{n,t} + W_q z + b) \tag{9}$$

$$\alpha_{n,t} = \frac{\exp(e_{n,t})}{\sum_{n'=1}^{N_s} \sum_{t'=1}^{T_{n'}} \exp(e_{n',t'})} \tag{10}$$

$$c = \sum_{n=1}^{N_s} \sum_{t=1}^{T_n} \alpha_{n,t} h_{n,t} \tag{11}$$

where $W_s \in \mathbb{R}^{K_a \times 2K_h}$, $W_q \in \mathbb{R}^{K_a \times K_z}$, $v \in \mathbb{R}^{K_a}$, and $b \in \mathbb{R}^{K_a}$ are learnable parameters with attention dimension $K_a = 32$. The resulting context vector $c \in \mathbb{R}^{2K_h}$ serves as a personalized physiological profile, encapsulating the historical sensory patterns most relevant to the user's current trajectory toward discomfort.

3.4.4 Forecasting Head

The final prediction is produced by a three-layer feed-forward network, following the empirical design of Sung et al. [62], that combines the context vector and query representation:

$$\mathbf{r} = [\mathbf{c}; \mathbf{z}] \tag{12}$$

$$\hat{y} = \mathrm{MLP}(\mathbf{r}; \theta_{mlp}) \tag{13}$$

where $\mathbf{r} \in \mathbb{R}^{2K_h + K_z}$ is the concatenated representation, and the MLP consists of three fully connected layers with 32 hidden units each, ReLU activations, and dropout regularization.

3.5 Episodic Meta-Optimization

To address the substantial inter-participant variability in cybersickness susceptibility, we adopt an episodic meta-optimization strategy rather than conventional global batch training. Grounded in the Model-Agnostic Meta-Learning (MAML) paradigm [13], our training procedure explicitly simulates the constraints of live VR deployment by treating each participant's VR exposure as an independent learning episode. Rather than learning a static, global representation of cybersickness that often fails on underrepresented users, this procedure optimizes the network's initialization to learn a generalized physiological baseline from which the model can rapidly adapt to a novel user's sensory conflict threshold with only a few gradient updates. The episodic meta-objective minimizes the expected forecasting error across all simulated episodes:

$$\mathcal{L}(\mathcal{S}, \mathcal{Q}; \Phi) = \frac{1}{N_q} \sum_{m=1}^{N_q} (f_{\mathcal{S}}(\mathbf{x}_m^*; \Phi) - y_m^*)^2 \tag{14}$$

where $\mathcal{S}$ is the support set, $\mathcal{Q}$ is the query set, Φ represents the model parameters, $f_{\mathcal{S}}$ is the adapted model, and $(\mathbf{x}_m^*, y_m^*)$ are query samples.

3.6 Experimental Setup and Evaluation Protocol

This section describes the experimental setup for evaluating the proposed few-shot meta-learning framework, including model configuration, baselines, training protocol, cross-validation strategy, and evaluation across in-domain and cross-domain VR scenarios.

3.6.1 Model Configuration

Our model employs bidirectional LSTMs ($\overrightarrow{f}$, $\overleftarrow{f}$) with $K_h = 32$ hidden units for support set encoding, and a unidirectional LSTM (g) with $K_z = 32$ hidden units for query encoding. The attention mechanism operates with dimension $K_a = 32$, and the forecasting head consists of a three-layer feed-forward network with 32 hidden units per layer. All networks use ReLU activations. For the MAML episodic training and test-time adaptation, the inner-loop updates use Stochastic Gradient Descent (SGD) with a learning rate of $\alpha = 0.01$ for 5 gradient adaptation steps. The outer-loop meta-optimization uses the Adam optimizer [31] with a meta-learning rate of $\eta = 10^{-3}$, dropout rate $p = 0.1$, and gradient clipping with maximum norm 2.0. All models were implemented in PyTorch and trained on an NVIDIA RTX 4090 GPU with mixed-precision training on a Dell Precision 3680 workstation (Intel 64-bit processor at 2.10 GHz, 16 GB RAM, Windows 11 Enterprise).

3.6.2 Baseline Models

To isolate the benefit of meta-learning personalization, we compared our approach against zero-shot and support-adapted baselines sharing identical multimodal inputs, preprocessing pipelines, and sliding windows. Zero-shot baselines (Attn LSTM and DeepTCN) are trained globally across participants and evaluated directly on novel users without adaptation. Support adapted baselines (Fine-Tuned Attn-LSTM and Fine-Tuned DeepTCN) take these globally trained networks and fine-tune their parameters directly on the user-specific support set before query evaluation. This comparison isolates whether performance gains arise from episodic meta initialization or simple test time optimization.

Zero-Shot Attn-LSTM. Recurrent architectures, particularly LSTMs and CNN-LSTMs, are well-established for modeling the temporal progression of cybersickness from physiological and sensor data [22, 23, 38]. We implemented a zero-shot Attn-LSTM that mirrors the encoder architecture of our proposed model without the meta-learning framework, using a bidirectional LSTM for input encoding, a unidirectional LSTM for query representation, and an attention mechanism for context extraction. By sharing our encoder design, this baseline serves as a direct ablation, ensuring that the observed performance gap is due to the few-shot personalization mechanism rather than the underlying architecture.

Zero-Shot DeepTCN. The second baseline is a Deep Temporal Convolutional Network (DeepTCN), which has been established as a strong architecture for continuous cybersickness forecasting due to its ability to capture long-range dependencies through dilated convolutions [23, 63]. Following the architecture validated by Islam et al. [23], our implementation uses five temporal convolutional layers with exponentially increasing dilations, residual connections, ReLU activations, skip connections, weighted batch normalization, and dropout (0.4). Like the Attn-LSTM, the DeepTCN is optimized globally using MSE loss and evaluated directly on novel participants without user-specific adaptation, providing a literature-backed benchmark for non-personalized forecasting.

Support Adapted Baselines. To isolate the benefits of personalization, we also implemented support-adapted counterparts for both base architectures, denoting them as Fine-Tuned Attn-LSTM and Fine-Tuned DeepTCN. These baselines utilize the exact network structures and global training procedures described in the preceding paragraphs. However, during evaluation on a novel participant, their parameters are fine-tuned directly on the user-specific support samples before live query forecasting. Comparing against these benchmarks determines whether predictive improvements stem from our episodic meta initialization or simply from having test-time access to calibration data.

3.6.3 Training Protocol

We employed 5-fold cross-validation across all experiments for robust evaluation. For in-domain experiments, cross-validation was performed at the participant level to prevent data leakage, ensuring no participant appeared in both training and testing sets. Each fold allocates 70% of participants for meta-training, 10% for validation, and 20% for testing. Training runs for a maximum of 200 episodes with early stopping based on validation loss (patience = 15 episodes). For cross-domain evaluation, models are trained on 90% of the source dataset, validated on the remaining 10%, and tested on the full target dataset, which is used exclusively for evaluation. This protocol assesses the model's ability to generalize across different VR environments and hardware configurations.

3.6.4 Evaluation Protocol

We report forecasting performance using Root Mean Squared Error (RMSE) to capture fold-wise variance and penalize larger prediction deviations. To provide a robust measure of central tendency against outlier ratings, we pair this metric with Mean Absolute Error (MAE) and 95% confidence intervals (CI), ensuring a complete evaluation of model accuracy and prediction stability. For each support and query configuration, RMSE is computed over the query predictions after adaptation as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N_q}\sum_{m=1}^{N_q}\left(f_{\mathscr{S}}(\mathbf{x}_m^q;\Phi) - y_m^q\right)^2}, \tag{15}$$

where $f_{\mathscr{S}}(\mathbf{x}_m^q;\Phi)$ is the predicted FMS score for the mth query input obtained using the model adapted to support set $\mathscr{S}$, and y_m^q is the corresponding ground truth label. The final score is averaged across all query samples and test participants. We report mean $\pm$ standard deviation across five folds and use paired t tests against baselines with significance at $p < .05$.

Table 1: Cross-domain consistency of FMS distributions before and after SUREAL alignment for the Maze and Simulation21 datasets using participant-level aggregation. Significant results ($p < 0.05$) are shown in bold.

Condition	N	Scale	Median [IQR]	MWU p	KS p
Maze (raw FMS)	36	1–7	1.143 [1.000, 2.661]	> 0.05	< 0.05
Sim21 (raw FMS)	30	0–10	1.021 [0.352, 1.285]		
Maze (SUREAL MLE)	36	0–10	1.604 [1.604, 1.604]	< 0.05	< 0.05
Sim21 (SUREAL MLE)	30	0–10	1.042 [1.008, 1.060]		
Maze (Unified; SUREAL+Calib.)	36	0–10	1.235 [1.220, 1.247]	> 0.05	> 0.05
Sim21 (Unified; SUREAL+Calib.)	30	0–10	1.244 [1.209, 1.250]		

Overall, our framework is designed to operate with minimal calibration and requires only a single 30-second support segment to initialize personalization. In the few-shot time-series forecasting literature, however, evaluation protocols typically use episodic settings with multiple support observations to stabilize adaptation and reduce sensitivity to short-term noise [24, 71]. Our experiments adopt a Support=10 and Query=15 configuration to stabilize baseline comparisons against short-term noise. However, our architecture natively supports single window personalization corresponding to 30 seconds of calibration or Support=1 without structural modifications, as Eq. 10 and Eq.11 compute the context vector over arbitrary support sizes.

4 Results

This section presents the experimental results of the proposed few-shot forecasting framework. We evaluate the model in in-domain and cross-domain settings against the baselines.

4.1 In-Domain and Cross-Domain Per-FMS-Level Performance

All values in the following analysis are computed on the psychometrically unified FMS scale (validated in Section 3.2, Table 1), ensuring that reported errors reflect genuine differences in predictive accuracy rather than artifacts of dataset-specific scale inconsistencies.

Figure 4 presents the mean RMSE at each FMS severity level for both in-domain and cross-domain settings using the 10/15 support/query configuration, which demonstrated stable and strong performance across all evaluation conditions. The figure shows in-domain results for the Maze, Sea, and Roller environments, along with cross-domain transfer results for Maze→Sea, Maze→Roller, and Sea+Roller→Maze. Across all conditions, a consistent trend emerges: RMSE increases with FMS severity level, indicating that forecasting becomes progressively more challenging as cybersickness intensifies. In the in-domain settings, prediction errors remain low at mild severity levels and gradually rise toward higher levels for both forecasting horizons, with the 30-second horizon consistently yielding smaller errors than the 60-second horizon. This pattern holds across all three environments, suggesting stable model behavior regardless of simulation type. In the cross-domain settings, RMSE values are generally elevated compared to in-domain results, reflecting the distributional shift between training and testing environments. This increase is particularly pronounced at higher FMS levels and for the 60-second horizon, where the compounding effects of domain mismatch and longer prediction windows amplify forecasting difficulty. These results suggest that forecasting accuracy varies with prediction horizon and domain similarity. Shorter horizons are generally more reliable, while cross-domain transfer adds uncertainty, particularly at higher cybersickness levels.

4.2 Improvement over Zero-Shot and Support-Adapted Baselines

As shown in Table 2, zero-shot models consistently produce higher forecasting errors across all evaluated conditions. This performance gap highlights a fundamental limitation of static global models, which fail to account for the substantial inter-user variability in

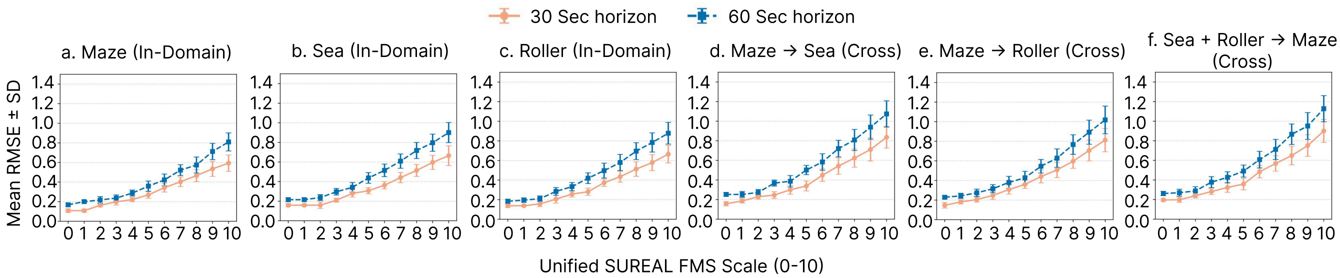

Figure 4: Per-FMS-level mean RMSE performance across different experimental scenarios for 30-second and 60-second forecasting horizons. In-domain performance on (a) the Maze dataset, (b) the Sea environment, and (c) the Roller environment from the Simulation21 dataset.Cross-domain scenarios: (d) Maze → Sea, (e) Maze → Roller, and (f) Sea + Roller → Maze. All RMSE values are reported using the psychometrically unified SUREAL [46] FMS scale (0–10).

Table 2: Comparison of forecasting performance using RMSE ± SD and MAE [95% CI] between zero-shot baselines, support-adapted baselines, and our few-shot meta-learning model. RMSE ± SD captures fold-wise variation, while MAE [95% CI] provides complementary error bounds. The * denotes statistical significance ($p < .05$) from a paired t-test compared to the best baseline.

Evaluation Setting	**Horizon**	**Zero-Shot Baselines**		**Support-Adapted Baselines**		**Our Approach**
		Attn-LSTM	**DeepTCN**	**Fine-Tuned Attn-LSTM**	**Fine-Tuned DeepTCN**	**Few-Shot Meta-Learning**
Maze (In-Domain)	30s	0.98 ± 0.08 0.78 [0.72, 0.84]	0.72 ± 0.06 0.58 [0.54, 0.62]	0.82 ± 0.10 0.66 [0.60, 0.72]	0.60 ± 0.09 0.48 [0.43, 0.53]	0.35 ± 0.03* 0.28 [0.26, 0.30]
	60s	1.34 ± 0.10 1.09 [1.00, 1.18]	0.95 ± 0.07 0.78 [0.72, 0.84]	1.12 ± 0.14 0.91 [0.82, 1.00]	0.82 ± 0.10 0.66 [0.60, 0.72]	0.44 ± 0.03* 0.36 [0.34, 0.38]
Sim21: Sea (In-Domain)	30s	1.05 ± 0.09 0.85 [0.78, 0.92]	0.78 ± 0.06 0.62 [0.58, 0.66]	0.87 ± 0.11 0.70 [0.62, 0.78]	0.65 ± 0.09 0.53 [0.47, 0.59]	0.39 ± 0.03* 0.31 [0.29, 0.33]
	60s	1.42 ± 0.11 1.16 [1.06, 1.26]	1.02 ± 0.08 0.83 [0.77, 0.89]	1.16 ± 0.14 0.94 [0.84, 1.04]	0.88 ± 0.11 0.71 [0.64, 0.78]	0.52 ± 0.04* 0.42 [0.39, 0.45]
Sim21: Roller (In-Domain)	30s	0.95 ± 0.08 0.76 [0.70, 0.82]	0.68 ± 0.05 0.55 [0.52, 0.58]	0.77 ± 0.11 0.63 [0.56, 0.70]	0.57 ± 0.08 0.46 [0.41, 0.51]	0.37 ± 0.03* 0.30 [0.28, 0.32]
	60s	1.28 ± 0.09 1.04 [0.96, 1.12]	0.88 ± 0.06 0.70 [0.65, 0.75]	1.06 ± 0.13 0.85 [0.77, 0.93]	0.76 ± 0.10 0.62 [0.56, 0.68]	0.49 ± 0.04* 0.40 [0.37, 0.43]
Maze → Sea (Cross)	30s	1.18 ± 0.10 0.96 [0.88, 1.04]	0.88 ± 0.07 0.71 [0.66, 0.76]	1.01 ± 0.13 0.81 [0.73, 0.89]	0.76 ± 0.10 0.62 [0.56, 0.68]	0.45 ± 0.04* 0.36 [0.33, 0.39]
	60s	1.58 ± 0.12 1.30 [1.19, 1.41]	1.15 ± 0.09 0.93 [0.86, 1.00]	1.32 ± 0.16 1.07 [0.95, 1.19]	0.98 ± 0.13 0.79 [0.71, 0.87]	0.60 ± 0.05* 0.49 [0.45, 0.53]
Maze → Roller (Cross)	30s	1.12 ± 0.09 0.90 [0.83, 0.97]	0.82 ± 0.07 0.66 [0.61, 0.71]	0.93 ± 0.13 0.75 [0.66, 0.84]	0.69 ± 0.09 0.56 [0.50, 0.62]	0.44 ± 0.03* 0.35 [0.33, 0.37]
	60s	1.49 ± 0.11 1.21 [1.11, 1.31]	1.08 ± 0.08 0.86 [0.80, 0.92]	1.24 ± 0.15 1.01 [0.90, 1.12]	0.92 ± 0.11 0.75 [0.68, 0.82]	0.55 ± 0.05* 0.45 [0.42, 0.48]
Sea + Roller → Maze	30s	1.20 ± 0.10 0.96 [0.88, 1.04]	0.92 ± 0.08 0.75 [0.70, 0.80]	1.03 ± 0.13 0.83 [0.74, 0.92]	0.79 ± 0.11 0.64 [0.57, 0.71]	0.48 ± 0.04* 0.39 [0.36, 0.42]
	60s	1.52 ± 0.12 1.23 [1.12, 1.34]	1.20 ± 0.09 0.98 [0.91, 1.05]	1.28 ± 0.16 1.03 [0.92, 1.14]	1.04 ± 0.13 0.84 [0.76, 0.92]	0.62 ± 0.06* 0.51 [0.47, 0.55]

sensory conflict thresholds. In the Maze in-domain setting at the 60-second horizon, zero-shot Attn-LSTM and DeepTCN yield RMSE values of 1.34 ± 0.10 (MAE 1.09) and 0.95 ± 0.07 (MAE 0.78), respectively. Support adaptation via Fine-Tuned Attn-LSTM and Fine-Tuned DeepTCN reduces these errors to 1.12 ± 0.14 (MAE 0.91) and 0.82 ± 0.10 (MAE 0.66), respectively. In contrast, our few-shot model achieves a substantially lower RMSE of 0.44 ± 0.03 (MAE 0.36). This improvement demonstrates that our meta-learned initialization calibrates the network to each user's susceptibility profile more effectively than standard test-time fine-tuning. The performance gap increases further under cross-domain evaluation. When transferring from a walking-based environment to a seated simulation (Maze → Roller) at the 60-second horizon, zero-shot baselines yield RMSE values of 1.49 ± 0.11 (MAE 1.21) and 1.08 ± 0.08 (MAE 0.86), while fine-tuning yields 1.24 ± 0.15 (MAE 1.01) and 0.92 ± 0.11 (MAE 0.75), respectively. Our model maintains a stable RMSE of 0.55 ± 0.05 (MAE 0.45). Similarly, in the reverse transfer from Sea and Roller to Maze, our model achieves an RMSE of 0.62 ± 0.06 (MAE 0.51), outperforming both the zero-shot DeepTCN at 1.20 ± 0.09 (MAE 0.98) and the Fine-Tuned DeepTCN at 1.04 ± 0.13 (MAE 0.84).

4.3 Modality Ablation

To assess the contribution of individual and combined sensor modalities, we conducted an ablation study across the Maze, Sea, and Roller environments for both 30-second and 60-second forecasting horizons (Figure 5). For all the following experiments, we clarify that the reported errors reflect models that were both trained and tested exclusively on the specified singular or dual input modalities.

Among single-modality inputs, eye-tracking (EYE) consistently yielded the lowest forecasting errors across all environments. In the Maze environment at the 60-second horizon, EYE achieved a mean RMSE of 0.62 ± 0.05, followed by physiological signals (PHYSIO) at 0.66 ± 0.06 and head-tracking (HEAD) at 0.69 ± 0.06. This ranking remained consistent across both the Sea and Roller datasets, suggesting that oculomotor features carry the strongest predictive signal for cybersickness onset. Theoretically, oculomotor features like pupil diameter and gaze instability serve as early indicators of sensory conflict. In contrast, cardiovascular and electrodermal responses exhibit a delayed autonomic onset and naturally contain a higher amount of biological noise. Dual-modality combinations consistently reduced both mean RMSE and variance compared to single-modality baselines. Among dual pairs, EYE+PHYSIO achieved the lowest error, recording 0.51 ± 0.04 in the Maze environment at the 60-second horizon, outperforming EYE+HEAD

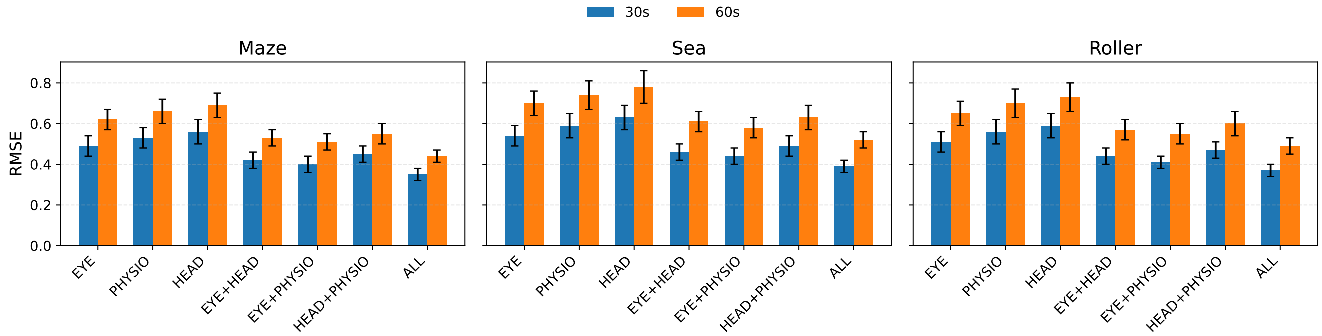

Figure 5: Feature importance and modality ablation results showing RMSE comparison across Maze, Sea, and Roller environments for different sensor combinations (evaluated at Support=10, Query=15). Values are reported as Mean ± Standard Deviation across 5-fold cross-validation for both 30s and 60s prediction horizons on the SUREAL-calibrated FMS scale.

Table 3: Forecasting performance under limited calibration using support sizes of 1, 3, and 5.

Setting	Support	30s Horizon		60s Horizon	
		RMSE ± SD	MAE [95% CI]	RMSE ± SD	MAE [95% CI]
In Domain Maze	1	0.56 ± 0.07	0.45 [0.40, 0.50]	0.69 ± 0.08	0.56 [0.50, 0.62]
	3	0.47 ± 0.05	0.38 [0.34, 0.42]	0.58 ± 0.06	0.47 [0.42, 0.52]
	5	0.40 ± 0.04	0.32 [0.29, 0.35]	0.51 ± 0.05	0.41 [0.37, 0.45]
In Domain Sea	1	0.62 ± 0.07	0.50 [0.45, 0.55]	0.78 ± 0.09	0.63 [0.56, 0.70]
	3	0.52 ± 0.05	0.42 [0.38, 0.46]	0.66 ± 0.07	0.53 [0.48, 0.58]
	5	0.45 ± 0.04	0.36 [0.33, 0.39]	0.58 ± 0.06	0.47 [0.43, 0.51]
In Domain Roller	1	0.59 ± 0.07	0.48 [0.43, 0.53]	0.75 ± 0.09	0.61 [0.55, 0.67]
	3	0.50 ± 0.05	0.40 [0.36, 0.44]	0.61 ± 0.07	0.49 [0.44, 0.54]
	5	0.42 ± 0.04	0.34 [0.31, 0.37]	0.55 ± 0.05	0.44 [0.40, 0.48]
Maze → Sea	1	0.73 ± 0.08	0.59 [0.53, 0.65]	0.87 ± 0.10	0.71 [0.63, 0.79]
	3	0.60 ± 0.06	0.48 [0.44, 0.52]	0.75 ± 0.08	0.61 [0.55, 0.67]
	5	0.52 ± 0.05	0.42 [0.39, 0.45]	0.67 ± 0.07	0.54 [0.49, 0.59]
Maze → Roller	1	0.69 ± 0.08	0.56 [0.50, 0.62]	0.81 ± 0.09	0.66 [0.59, 0.73]
	3	0.58 ± 0.06	0.47 [0.42, 0.52]	0.68 ± 0.07	0.55 [0.50, 0.60]
	5	0.50 ± 0.05	0.40 [0.37, 0.43]	0.61 ± 0.06	0.49 [0.45, 0.53]
Sea + Roller → Maze	1	0.74 ± 0.09	0.60 [0.54, 0.66]	0.92 ± 0.11	0.75 [0.67, 0.83]
	3	0.63 ± 0.07	0.51 [0.46, 0.56]	0.78 ± 0.09	0.63 [0.57, 0.69]
	5	0.55 ± 0.05	0.44 [0.40, 0.48]	0.69 ± 0.07	0.56 [0.51, 0.61]

(0.53 ± 0.04) and HEAD+PHYSIO (0.55 ± 0.05). The fusion of all three modalities (ALL) yielded the lowest forecasting errors and tightest variance across every tested condition. In the Maze setting, the full multimodal configuration achieved an RMSE of 0.35 ± 0.03 at the 30-second horizon and 0.44 ± 0.03 at the 60-second horizon. This pattern held in the Sea (0.39 ± 0.03 for 30s; 0.52 ± 0.04 for 60s) and Roller (0.37 ± 0.03 for 30s; 0.49 ± 0.04 for 60s) environments, confirming that complementary information across modalities consistently improves forecasting accuracy and stability.

4.4 Ablation: Limited Calibration Forecasting

To evaluate data efficiency under realistic calibration limits, we tested support sizes of 1, 3, and 5 samples. As shown in Table 3, our model personalizes effectively with only one 30-second support window. At the 60-second horizon, Support=1 achieves RMSE values of 0.69 ± 0.08 (MAE 0.56) in Maze and 0.81 ± 0.09 (MAE 0.66) for Maze → Roller transfer. With 3 support samples, errors decrease to 0.58 ± 0.06 (MAE 0.47) and 0.68 ± 0.07 (MAE 0.55), respectively. With 5 support samples, performance further improves to 0.51 ± 0.05 (MAE 0.41) in Maze and 0.61 ± 0.06 (MAE 0.49) in Maze → Roller, approaching the Support=10 baseline. These results show that the meta-learned initialization enables rapid personalization with limited calibration.

5 Discussion

In this section, we interpret the experimental findings and discuss their implications for personalized cybersickness forecasting, focusing on how the proposed framework improves forecasting accuracy with limited user data and generalizes across VR domains.

Table 4: Computational Overhead and Deployment Feasibility Metrics for the Proposed Few-Shot Meta-Learning Model.

Metric	Measurement
Total Trainable Parameters	∼125,000
Model Memory Footprint	< 1.0 MB
Average Inference Latency (per query)	∼2.5 ms
Standard VR Frame Budget (90 Hz)	11.1 ms

5.1 Personalized Forecasting Performance

The results consistently demonstrate that test-time personalization is critical for reliable cybersickness forecasting. Using only a brief physiological baseline from the support set, our few-shot meta-learning model substantially outperforms both zero-shot baselines across all evaluation settings. In the Maze in-domain setting at the 60-second horizon, the proposed model achieves an RMSE of 0.44 ± 0.03, compared to 1.34 ± 0.10 for Attn-LSTM and 0.95 ± 0.07 for DeepTCN, representing 67.2% and 53.7% reductions in forecasting error, respectively. These results suggest that globally trained models are unable to adequately capture the inter-user variability inherent in sensory conflict thresholds and cybersickness susceptibility. Our episodic meta-optimization strategy addresses this limitation by adapting to each user's physiological profile from minimal labeled data, achieving strong personalization without requiring full model retraining.

5.2 Cross-Domain Generalization

Beyond in-domain accuracy, our framework demonstrates strong robustness under domain shifts, a property essential for consumer VR deployment, where users encounter varied locomotion and visual styles. We used the SUREAL framework to establish a statistically validated, unified severity scale (0–10) across datasets and evaluated how well the model transfers from a walking-based task (Maze) to seated, visually driven simulations (Roller). In this Maze→Roller cross-domain setting, the zero-shot Attn-LSTM degraded to an RMSE of 1.49 ± 0.11 at the 60-second horizon, while our few-shot model maintained an RMSE of 0.55 ± 0.05, achieving a 63.1% reduction in forecasting error. This performance gap highlights a key design rationale. While visual factors such as optical flow, luminance, and contrast certainly contribute to cybersickness variance across domains, attempting to model them directly can cause models to overfit to specific scenes. Instead, our model relies purely on physiological and kinematic responses. Because these signals are consistently available across both datasets, they support a lightweight forecasting pipeline that is independent of rendering-specific feature extraction. Consequently, the model learns generalizable representations of physiological precursors to discomfort rather than overfitting to environment-specific visual or motion patterns, enabling effective transfer across fundamentally different VR interaction paradigms.

5.3 Multimodal Contribution and Fusion

The modality ablation study reveals the relative predictive value of individual sensor streams. Among single-modality inputs, eye-tracking (EYE) emerged as the most reliable standalone predictor across all environments, outperforming both physiological signals (PHYSIO) and head-tracking (HEAD). However, the results clearly support the value of a multimodal approach. Dual-modality combinations consistently reduced forecasting error compared to single-modality baselines, and fusing all three modalities (EYE+HEAD+PHYSIO) achieved the lowest error and tightest variance across all conditions (e.g., 0.35 ± 0.03 for Maze at 30s). These findings indicate that gaze behavior and pupillometric features serve as strong early indicators of sensory conflict, while cardiovascular, electrodermal, and head motion dynamics provide complementary physiological context that stabilizes predictions across diverse users.

5.4 Forecasting Uncertainty at High Severity Levels

Our model exhibits a notable behavior pattern at elevated cybersickness severity. As the per-FMS-level analysis shows, prediction error and variance increase monotonically as ground-truth FMS scores approach maximum severity, reflecting the inherent heteroscedasticity of right-skewed cybersickness distributions. The datasets contain relatively few severe sickness events, and physiological signals become increasingly variable under intense sensory conflict, which naturally widens the model's confidence intervals. Nevertheless, our few-shot adaptation maintains high-severity errors that remain substantially lower and more tightly bounded than those of the zero-shot baselines. For safety-critical VR applications, these findings suggest that designers should set intervention thresholds conservatively, triggering mitigation strategies before the estimated peak severity to account for the increased variance at extreme levels.

5.5 Computational Efficiency

Our asymmetric architectural design directly addresses the computational constraints of real-time cybersickness forecasting. We encode historical support sequences bidirectionally during an offline preprocessing stage, while the query pathway uses a lightweight unidirectional causal LSTM that preserves temporal causality and minimizes inference overhead. Hardware profiling shows that the model contains 125,000 parameters requiring 0.25 MegaFLOPs, using under 1.0 MB of memory with 2.5 milliseconds of latency on standard workstations (Table 4). This small footprint is compatible with the thermal and compute constraints of standalone VR chipsets such as Snapdragon XR2, where memory bandwidth and processor cycles are limited. This efficiency supports continuous background operation with limited impact on performance, making the model suitable for real-time comfort-aware VR systems.

6 Limitations and Future Work

Despite the strong performance our framework demonstrates, several limitations warrant discussion. **First,** although episodic training reduces the need for large offline datasets, our model achieves optimal performance with a support set of 10 samples used for user-specific calibration. In consumer VR, this initial requirement may present a usability barrier. In future work, we plan to explore rolling window strategies that incrementally build and update the support set in the background during interaction, reducing upfront calibration time while also accommodating natural physiological adaptation over longer sessions. **Second,** while we tested cross-domain generalization between walking-based and seated VR experiences, real-world VR applications span a much broader range of interaction modalities. Environments such as fast-paced games, social platforms, or collaborative training scenarios may introduce distinct visual motion patterns and cognitive demands that influence cybersickness differently. Future studies should evaluate the approach across a wider variety of applications and more demographically diverse populations. **Third,** although the 30- and 60-second forecasting horizons provide actionable lead time for interventions, our per-FMS-level analysis revealed that prediction uncertainty grows as users approach high severity levels, since physiological responses become less stable under intense sensory conflict. Future iterations of the system should incorporate uncertainty-aware decision thresholds to maintain reliable safety margins at these extremes. **Fourth,** the model does not include direct visual scene features such as optical flow, luminance, and contrast. This may limit its response to sudden visual stimuli before their effects appear in personal sensor signals. Future work could incorporate lightweight visual motion descriptors to complement these signals. **Finally,** we plan to integrate this lightweight forecasting model into a closed-loop mitigation framework, where real-time predictions automatically trigger adaptive visual adjustments such as field-of-view reduction or motion smoothing to proactively support user comfort and safety during extended VR use. Evaluating this complete pipeline on standalone hardware under active rendering, sensing, and mitigation workloads remains an important direction for future work.

7 Conclusion

We presented an asymmetric few-shot meta-learning framework for personalized cybersickness forecasting in VR. Our approach addresses the core limitation of conventional globally trained models, which fail to generalize across users due to substantial inter-individual variability in cybersickness susceptibility [9, 22, 25, 50]. The framework employs an asymmetric temporal encoding design that processes historical support sequences bidirectionally to capture personalized physiological baselines, while encoding live observations through a forward causal pathway that preserves temporal order during deployment. We evaluated the approach on two public datasets with distinct VR paradigms: the real walking Maze task and the seated Simulation21 tasks, unified on a validated 0 to 10 SUREAL scale. At the 60 second horizon, our model achieved an in-domain RMSE of 0.44 ± 0.03 on Maze and maintained cross-domain performance with an RMSE of 0.55 ± 0.05 when transferring from walking to seated VR. Although both datasets use similar HMD sensing configurations, these results suggest that few-shot learning can adapt across different postures and locomotion styles. Modality ablation confirmed that fusing eye-tracking, head-tracking, and physiological signals is necessary to achieve optimal accuracy and stability. With an inference latency of approximately 2.5 milliseconds and a memory footprint under 1.0 MB, the framework meets the computational constraints of standalone VR headsets for continuous background operation. These results establish a practical foundation for proactive, personalized cybersickness mitigation and support the development of safer and more accessible VR experiences.

Acknowledgement

This work was supported in part by the Defense Advanced Research Projects Agency (DARPA) under grant HR00112420366. The authors used Large Language Models and Grammarly only to refine the writing and improve clarity. All research contributions and the conceptual design were developed solely by the authors. All figures are original and were manually created without the use of AI-based image generation tools.

References

[1] W. Al-Ashwal, H. Asadi, S. Alsanwy, M. C. Qazani, S. Mohamed, and S. Nahavandi. Objective assessment of cybersickness in a stationary helicopter simulator using pupil diameter as a quantitative indicator. In

2024 IEEE International Systems Conference (SysCon)*, pp. 1–7. IEEE, 2024.
[2] E. Al Khalifah, R. Hammady, M. Abdelrahman, O. Al-Shamaileh, M. Marghany, H. El-Jarn, A. Darwish, and Y. Kurt. Technology anxiety in virtual reality adoption: Examining the impact of age, past experience, and cybersickness. *IEEE Access*, 13:71858–71879, 2025. doi: 10.1109/ACCESS.2025.3562383
[3] M. Alaeifard and M. Safaei. Head movement patterns as predictors of cybersickness in virtual reality games. *International Journal of Advanced Human Computer Interaction*, 2(2):1–10, 2024.
[4] I. M. Arafat, S. M. S. Ferdous, and J. Quarles. The effects of cybersickness on persons with multiple sclerosis. In *Proceedings of the 22nd ACM conference on virtual reality software and technology*, pp. 51–59, 2016.
[5] G. S. Bailey, D. G. Arruda, and T. A. Stoffregen. Using quantitative data on postural activity to develop methods to predict and prevent cybersickness. *Frontiers in Virtual Reality*, 3:1001080, 2022.
[6] P. Bockelman and D. Lingum. Factors of cybersickness. In *international conference on human-computer interaction*, pp. 3–8. Springer, 2017.
[7] P. Caserman, A. Garcia-Agundez, A. Gámez Zerban, and S. Göbel. Cybersickness in current-generation virtual reality head-mounted displays: systematic review and outlook. *Virtual Reality*, 25(4):1153–1170, 2021.
[8] E. Chang, H. T. Kim, and B. Yoo. Predicting cybersickness based on user's gaze behaviors in hmd-based virtual reality. *Journal of Computational Design and Engineering*, 8(2):728–739, 2021.
[9] Y. Choi, D. Jeong, B. Kim, and K. Han. Early prediction of cybersickness in virtual reality using a large language model for multimodal time series data. In *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 25–29, 2024.
[10] S. Cossio, S. Chiappinotto, S. Dentice, C. Moreal, G. Magro, G. Dussi, A. Palese, and A. Galazzi. Cybersickness and discomfort from head-mounted displays delivering fully immersive virtual reality: A systematic review. *Nurse Education in Practice*, p. 104376, 2025.
[11] W. N. da Silva, T. Porcino, C. D. Castanho, and R. P. Jacobi. Prediction and analysis of cybersickness in vr games using symbolic machine learning. *Journal on Interactive Systems*, 17(1):80–98, 2026.
[12] A. da Silva Marinho, U. Terton, and C. M. Jones. Cybersickness and postural stability of first time vr users playing vr videogames. *Applied Ergonomics*, 101:103698, 2022.
[13] C. Finn, P. Abbeel, and S. Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, p. 1126–1135. JMLR.org, 2017.
[14] J. P. Freiwald, Y. Göbel, F. Mostajeran, and F. Steinicke. The cybersickness susceptibility questionnaire: Predicting virtual reality tolerance. In *Proceedings of Mensch und Computer 2020*, pp. 115–118. ACM, 2020.
[15] J. M. Fulvio, M. Ji, and B. Rokers. Variations in visual sensitivity predict motion sickness in virtual reality. *Entertainment Computing*, 38:100423, 2021. doi: 10.1016/j.entcom.2021.100423
[16] M. Fusaro, M. P. Lisi, V. Era, et al. The transformative power of virtual reality: Redefining interactions in virtual platforms, education, healthcare, and workplaces. *Topoi*, 2025. doi: 10.1007/s11245-025-10216-1
[17] L. E. Garrido, M. Frías-Hiciano, M. Moreno-Jiménez, G. N. Cruz, Z. E. García-Batista, K. Guerra-Peña, and L. A. Medrano. Focusing on cybersickness: pervasiveness, latent trajectories, susceptibility, and effects on the virtual reality experience. *Virtual reality*, 26(4):1347–1371, 2022.
[18] G. Gonçalves, M. Melo, C. Serôdio, R. Silva, and M. Bessa. Adaptation and validation of the simulator sickness questionnaire to portuguese (ssqp) based on immersive virtual reality exposure. *IEEE Access*, 12:92708–92717, 2024.
[19] J. Hildebrandt, P. Schmitz, A. Calero Valdez, L. Kobbelt, and M. Ziefle. Get well soon! human factors' influence on cybersickness after redirected walking exposure in virtual reality. In *International conference on virtual, augmented and mixed reality*, pp. 82–101. Springer, 2018.
[20] P. Hořejší, A. Lochmannová, V. Jezl, and M. Dvořák. Virtual reality locomotion methods differentially affect spatial orientation and cybersickness during maze navigation. *Scientific Reports*, 15(1):1–14, 2025. doi: 10.1038/s41598-025-12143-y
[21] P. Hu, Q. Sun, P. Didyk, L.-Y. Wei, and A. E. Kaufman. Reducing simulator sickness with perceptual camera control. *ACM Transactions on Graphics (TOG)*, 38(6):1–12, 2019.
[22] R. Islam, K. Desai, and J. Quarles. Cybersickness prediction from integrated hmd's sensors: A multimodal deep fusion approach using eye-tracking and head-tracking data. In *2021 IEEE international symposium on mixed and augmented reality (ISMAR)*, pp. 31–40. IEEE, 2021.
[23] R. Islam, K. Desai, and J. Quarles. Towards forecasting the onset of cybersickness by fusing physiological, head-tracking and eye-tracking with multimodal deep fusion network. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 121–130, 2022. doi: 10.1109/ISMAR55827.2022.00026
[24] T. Iwata and A. Kumagai. Few-shot learning for time-series forecasting. *arXiv preprint arXiv:2009.14379*, 2020. doi: 10.48550/arXiv.2009.14379
[25] D. Jeong and K. Han. 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, 2024.
[26] J. J. L. Jr. A discussion of cybersickness in virtual environments. *ACM SIGCHI Bulletin*, 32(1):47–56, 2000.
[27] A. Kemeny, J.-R. Chardonnet, and F. Colombet. Getting rid of cybersickness. *Virtual reality, augmented reality, and simulators*, 2, 2020.
[28] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993.
[29] B. Keshavarz and H. Hecht. Validating an efficient method to quantify motion sickness. *Human Factors*, 53(4):415–426, 2011.
[30] J. Kim, W. Kim, H. Oh, S. Lee, and S. Lee. A deep cybersickness predictor based on brain signal analysis for virtual reality contents. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 10580–10589, 2019.
[31] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.
[32] T. Kojić, R. Spang, M. Vergari, L. Meier, S. Möller, and J.-N. Voigt-Antons. Effects of user factors on user experience in virtual reality: age, gender, and vr experience as influencing factors for vr exergames. *Quality and User Experience*, 8(1):3, 2023.
[33] P. Kourtesis, R. Amir, J. Linnell, F. Argelaguet, and S. E. MacPherson. Cybersickness, cognition, & motor skills: The effects of music, gender, and gaming experience. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2326–2336, 2023.
[34] R. K. Kundu, O. Y. Elsaid, P. Calyam, and K. A. Hoque. Vr-lens: Super learning-based cybersickness detection and explainable ai-guided deployment in virtual reality. In *Proceedings of the 28th international conference on intelligent user interfaces*, pp. 819–834, 2023.
[35] R. K. Kundu, R. Islam, P. Calyam, and K. A. Hoque. Truvr: Trustworthy cybersickness detection using explainable machine learning. In *2022 IEEE International Symposium on Mixed and Augmented Reality*, pp. 777–786. IEEE, 2022. doi: 10.1109/ISMAR55827.2022.00099
[36] R. K. Kundu, R. Islam, J. Quarles, and K. A. Hoque. Litevr: Interpretable and lightweight cybersickness detection using explainable ai. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 609–619. IEEE, 2023.
[37] U. Laessoe, S. Abrahamsen, S. Zepernick, A. Raunsbaek, and C. Stensen. Motion sickness and cybersickness – sensory mismatch. *Physiology & Behavior*, 258:114015, 2023. doi: 10.1016/j.physbeh.2022.114015
[38] T. M. Lee, J.-C. Yoon, and I.-K. Lee. Motion sickness prediction in stereoscopic videos using 3d convolutional neural networks. *IEEE Transactions on Visualization and Computer Graphics*, 26(5):2258–2267, 2020. doi: 10.1109/TVCG.2020.2973470
[39] Q. Liu, Y. Tian, T. Zhou, K. Lyu, R. Xin, Y. Shang, Y. Liu, J. Ren, and J. Li. A few-shot disease diagnosis decision making model based on meta-learning for general practice. *Artificial Intelligence in Medicine*, 147:102718, 2024.

[40] C. MacArthur, A. Grinberg, D. Harley, and M. Hancock. You're making me sick: A systematic review of how virtual reality research considers gender & cybersickness. In *Proceedings of the 2021 CHI conference on human factors in computing systems*, pp. 1–15, 2021.

[41] A. Makani, R. Saryazdi, S. Givetash, and B. Keshavarz. The presence of an avatar can reduce cybersickness in virtual reality. *Virtual Reality*, 28(4):163, 2024.

[42] N. Martin, N. Mathieu, N. Pallamin, M. Ragot, and J.-M. Diverrez. Virtual reality sickness detection: an approach based on physiological signals and machine learning. In *2020 IEEE international symposium on mixed and augmented reality (ISMAR)*, pp. 387–399. IEEE, 2020.

[43] A. J. Martingano, E. Brown, S. H. Telaak, A. P. Dolwick, and S. Persky. Cybersickness variability by race: Findings from 6 studies and a mini meta-analysis. *Journal of Medical Internet Research*, 24(6):e36843, 2022. doi: 10.2196/36843

[44] M. Melo, G. Gonçalves, D. Narciso, and M. Bessa. Impact of different role types and gender on presence and cybersickness in immersive virtual reality setups. In *2021 International Conference on Graphics and Interaction (ICGI)*, pp. 1–8. IEEE, 2021. doi: 10.1109/ICGI53575.2021.00007

[45] J. Nazari, A. Motie Nasrabadi, M. B. Menhaj, and S. Raiesdana. Epilepsy seizure prediction with few-shot learning method. *Brain Informatics*, 9(1):1–9, 2022. doi: 10.1186/s40708-022-00170-8

[46] Netflix. SUREAL: Subjective assessment of multimedia video quality. `https://github.com/Netflix/sureal`, 2020.

[47] R. Olivier et al. To pre-process or not to pre-process? on the role of eeg enhancement for cybersickness characterization and the importance of amplitude modulation features. *Frontiers in Virtual Reality*, 6:1468971, 2025.

[48] I. H. Rather, S. Kumar, and A. H. Gandomi. Breaking the data barrier: a review of deep learning techniques for democratizing ai with small datasets. *Artificial Intelligence Review*, 57(9):226, 2024.

[49] L. Rebenitsch and C. Owen. Individual variation in susceptibility to cybersickness. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pp. 309–317, 2014.

[50] G. R. Reddy, C. A. Spencer, K. Durkee, B. Cox, O. Fox Cotton, S. Galbreath, S. Meyer, M. Natali, T. Seech, G. Severe-Valsaint, et al. Estimating cognitive load and cybersickness of pilots in vr simulations via unobtrusive physiological sensors. In *International Conference on Human-Computer Interaction*, pp. 251–269. Springer, 2022.

[51] S. Saleem, M. Aslam, and M. R. Shaukat. A review and empirical comparison of univariate outlier detection methods. *Pakistan Journal of Statistics*, 37(4), 2021.

[52] J. Sameri, H. Coenegracht, S. Van Damme, F. De Turck, and M. Torres Vega. Physiology-driven cybersickness detection in virtual reality: a machine learning and explainable ai approach. *Virtual Reality*, 28(4):174, 2024.

[53] A. Savitzky and M. J. E. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8):1627–1639, 1964. doi: 10.1021/ac60214a047

[54] F. Schöberl, A. Zwergal, and T. Brandt. Testing navigation in real space: contributions to understanding the physiology and pathology of human navigation control. *Frontiers in Neural Circuits*, 14:6, 2020.

[55] N. Sendhilnathan, T. Zhang, D. Bethge, M. Nebeling, T. Grossman, and T. R. Jonker. A multimodal approach for targeting error detection in virtual reality using implicit user behavior. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25. Association for Computing Machinery, New York, NY, USA, 2025. doi: 10.1145/3706598.3713777

[56] J.-C. Servotte, M. Goosse, S. H. Campbell, N. Dardenne, B. Pilote, I. L. Simoneau, M. Guillaume, I. Bragard, and A. Ghuysen. Virtual reality experience: Immersion, sense of presence, and cybersickness. *Clinical Simulation in Nursing*, 38:35–43, 2020.

[57] J. N. Setu, J. M. Le, R. K. Kundu, B. Giesbrecht, T. Höllerer, K. A. Hoque, K. Desai, and J. Quarles. Mazed and confused: A dataset of cybersickness, working memory, mental load, physical load, and attention during a real walking task in vr. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 1048–1057, 2024. doi: 10.1109/ISMAR62088.2024.00121

[58] Y. Song, T. Wang, P. Cai, S. K. Mondal, and J. P. Sahoo. A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities. *ACM Comput. Surv.*, 55(13s), July 2023. doi: 10.1145/3582688

[59] B. Spittle, M. Frutos-Pascual, C. Creed, and I. Williams. A review of interaction techniques for immersive environments. *IEEE Transactions on Visualization and Computer Graphics*, 29(9):3900–3921, 2022.

[60] K. Stanney, B. D. Lawson, B. Rokers, M. Dennison, C. Fidopiastis, T. Stoffregen, S. Weech, and J. M. Fulvio. Identifying causes of and solutions for cybersickness in immersive technology: Reformulation of a research and development agenda. *International Journal of Human–Computer Interaction*, 36(19):1783–1803, 2020. doi: 10.1080/10447318.2020.1828535

[61] T. A. Stoffregen and L. Smart. Postural instability precedes motion sickness. *Brain Research Bulletin*, 47(5):437–448, 1998. doi: 10.1016/S0361-9230(98)00102-6

[62] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. Torr, and T. M. Hospedales. Learning to compare: Relation network for few-shot learning. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1199–1208, 2018. doi: 10.1109/CVPR.2018.00131

[63] U. Tasnim, R. Islam, K. Desai, and J. Quarles. Investigating personalization techniques for improved cybersickness prediction in virtual reality environments. *IEEE Transactions on Visualization and Computer Graphics*, 30(5):2368–2378, 2024. doi: 10.1109/TVCG.2024.3372122

[64] J. Tervonen, K. Pettersson, and J. Mäntyjärvi. Ultra-short window length and feature importance analysis for cognitive load detection from wearable sensors. *Electronics*, 10(5):1–19, Mar. 2021. Publisher Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. Copyright: Copyright 2021 Elsevier B.V., All rights reserved. doi: 10.3390/electronics10050613

[65] N. Tian, P. Lopes, and R. Boulic. A review of cybersickness in head-mounted displays: Raising attention to individual susceptibility. *Virtual Reality*, 26(4):1409–1441, 2022. doi: 10.1007/s10055-021-00569-9

[66] I. Valori, P. E. McKenna-Plumley, R. Bayramova, and T. Farroni. Perception and motion in real and virtual environments: a narrative review of autism spectrum disorders. *Frontiers in psychology*, 12:708229, 2021.

[67] J. Wang, H. Li, S. Pan, and X. Xie. A tutorial on domain generalization. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, WSDM '23, p. 1236–1239. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3539597.3572722

[68] J. Wang, H.-N. Liang, D. Monteiro, W. Xu, and J. Xiao. Real-time prediction of simulator sickness in virtual reality games. *IEEE Transactions on Games*, 15(2):252–261, 2022.

[69] P. Wu, N. Ahmed, K. Huang, R. Islam, T. Lan, G. Tan, and M. Imani. Personalized bayesian networks for cybersickness prediction in virtual reality. In *Proceedings of the Twenty-sixth International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing*, pp. 491–496, 2025.

[70] P. Wu, N. Ahmed, A. Sarma, K. Huang, R. Islam, B. Li, T. Lan, G. Tan, and M. Imani. Probabilistic verification of cybersickness in virtual reality through Bayesian Networks. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, 2025.

[71] F. Xiao, L. Liu, J. Han, D. Guo, S. Wang, H. Cui, and T. Peng. Meta-learning for few-shot time series forecasting. *J. Intell. Fuzzy Syst.*, 43(1):325–341, Jan. 2022. doi: 10.3233/JIFS-212228

[72] J. Xu, K. Li, and D. Li. An automated few-shot learning for time-series forecasting in smart grid under data scarcity. *IEEE Transactions on Artificial Intelligence*, 5(6):2482–2492, 2024. doi: 10.1109/TAI.2024.3358795

[73] W. Xu, H. Liang, Q. He, X. Li, K. Yu, and Y. Chen. Results and guidelines from a repeated-measures design experiment comparing standing and seated full-body gesture-based immersive virtual reality exergames: Within-subjects evaluation. *JMIR Serious Games*, 8(3):e17972, 2020. doi: 10.2196/17972

[74] D. Zielasko and B. E. Riecke. Sitting or standing in vr: about comfort, conflicts, and hazards. *IEEE Computer Graphics and Applications*, 44(2):81–88, 2024.