Not Everyone Gets Sick the Same Way: Personalized Cybersickness Prediction via Bayesian Networks with Individual Susceptibility Modeling

Nasim Ahmed*, Peng Wu*, Kaiming Huang, Bin Li, Tian Lan, Gang Tan, Mahdi Imani, and Rifatul Islam

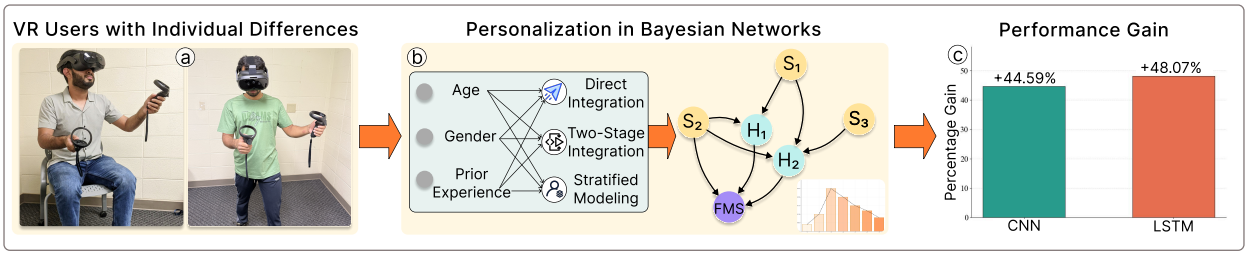

Figure 1: Overview of the personalized cybersickness prediction framework. (a) VR users exhibit different levels of cybersickness susceptibility due to individual characteristics such as age, gender, and prior VR experience. (b) These personalization attributes are incorporated into a Bayesian Network together with system (visual scene) parameters (e.g., optical flow) and physiological signals to model probabilistic dependencies between stimuli, user responses, and cybersickness severity. (c) The resulting personalized models improve prediction performance compared with non-personalized and deep learning baselines, demonstrating the benefit of integrating individual differences into probabilistic cybersickness prediction.

Abstract— Predicting the onset of cybersickness enables proactive mitigation strategies that minimize user discomfort, which is a key requirement for the widespread adoption of virtual reality (VR) in education, healthcare, training, and entertainment. Current prediction models leverage multimodal data such as visual scene parameters and physiological signals, yet often overlook individual differences that strongly influence cybersickness susceptibility, limiting their generalizability across diverse users. To address this limitation, we propose a personalized cybersickness prediction framework based on Bayesian Networks (BN) that incorporates individual factors, including age, gender, and prior VR experience. We introduce three personalization strategies: 1) direct integration of user attributes, 2) a two-stage pipeline that learns a latent susceptibility score, and 3) stratified models trained on homogeneous user groups. We validate our proposed method on two publicly available datasets (Maze and Simulation21). On the Maze dataset, our approach achieves a 14.27% relative improvement over the non-personalized Bayesian baseline and outperforms the state-of-the-art (SOTA) DeepTCN and LSTM models for cybersickness prediction by 44.59% and 48.07%, respectively. Our model also surpasses SOTA personalization-based deep learning methods, improving accuracy by 37.83% over the transfer learning strategy. We observe similar trends on the Simulation21 dataset, where our personalized framework yields a 13.40% accuracy gain over the non-personalized baseline, outperforms DeepTCN and LSTM by 9.16% and 10.82%, and improves accuracy by 3.89% over the transfer learning approach. Our results further demonstrate consistent improvements in F1 score, Cohen's κ, and Spearman correlation. Our method also maintains low inference latency, supporting interpretable and efficient probabilistic predictions in VR settings.

Index Terms— Cybersickness Prediction, Virtual Reality, Personalization, Bayesian Networks

1 Introduction

Cybersickness remains a fundamental barrier to the sustained use of virtual reality (VR), manifesting as nausea, dizziness, disorientation, and visual discomfort during immersive exposure [17, 33, 35, 39]. Critically, both its onset and severity exhibit substantial inter-individual variability driven by factors such as age, gender, and prior VR experience [63], making reliable prediction inherently user-dependent. To support proactive mitigation, prior work has leveraged multimodal sensor streams, including physiological signals, eye tracking, and system-level parameters, together with temporal learning models for real-time cybersickness prediction [13–15, 46, 50]. While these approaches achieve strong predictive performance, they predominantly rely on complex deep learning architectures that lack interpretability and introduce substantial computational overhead, limiting their suitability for real-time and resource-constrained VR systems [6, 32]. In parallel, recent work has explored explainable and probabilistic formulations to improve transparency in cybersickness modeling [26, 27, 64], yet these approaches largely treat users as homogeneous and do not explicitly account for individual susceptibility despite strong empirical evidence of its influence [5, 9, 37]. This gap highlights the need for predictive frameworks that explicitly incorporate user-specific factors while remaining interpretable, computationally efficient, and suitable for VR deployment.

Personalization has been studied in related VR research, including adaptive user interfaces and feedback systems [24, 56]. However, its use in cybersickness prediction is still limited [54]. Existing cybersickness personalization methods commonly use group stratification or model fine-tuning, which may require additional computation and may be

• *Nasim Ahmed and Peng Wu contributed equally to this work.*
• *Nasim Ahmed and Rifatul Islam are with Kennesaw State University. E-mail: {nahmed25@students.kennesaw.edu, rislam11@kennesaw.edu}.*
• *Peng Wu and Mahdi Imani are with Northeastern University. E-mail: {wu.p@northeastern.edu, m.imani@northeastern.edu}.*
• *Kaiming Huang and Gang Tan are with The Pennsylvania State University. E-mail: {kzh529@psu.edu, gtan@psu.edu}.*
• *Tian Lan is with George Washington University. E-mail: tlan@gwu.edu.*
• *Bin Li is with The Pennsylvania State University. E-mail: binli@psu.edu.*

Manuscript received xx xxx. 202x; accepted xx xxx. 202x. Date of Publication xx xxx. 202x; date of current version xx xxx. 202x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.202x.xxxxxxx

difficult to use in real-world VR settings [54]. In addition, there is still limited agreement on how to combine user attributes, such as age, gender, and prior VR experience, with simulation factors, such as VR scenario type, visual motion, scene complexity, and locomotion conditions, within prediction models [22].

To address this gap, we propose a Bayesian Network (BN) framework for personalized cybersickness prediction that explicitly integrates individual factors such as age, gender, and prior VR experience for a personalized cybersickness prediction. We propose three distinct personalization strategies: **(1) Direct Integration,** where personalization features are included as input variables to the model; (2) **Two-Stage Personalization,** where a latent susceptibility score is first learned from personal features and then used as an intermediate variable in prediction; and **(3) Stratified Modeling,** where separate models are trained for different demographic groups. We evaluated our proposed approaches on two publicly available datasets: 1) Maze dataset [50] and 2) Simulation21 dataset [15]. Each dataset includes measurements from external sensors and head-mounted display signals, along with cybersickness self-reports in the form of Fast Motion Sickness (FMS) scores. Our model also supports interpretable graph-based reasoning, modular personalization, and flexible extension across datasets. While BNs are established methods, our main contribution is a VR-focused framework that integrates physiological signals, individual susceptibility factors, and real-time probabilistic risk estimation. Instead of simply adding demographic variables, the model treats personal attributes as probabilistic factors that can be observed, omitted, or marginalized, while still providing continuous risk estimates with low computational cost and 0.34 ms latency. **In summary our contributions includes but are not limited to:**

- We propose a probabilistic personalization mechanism that integrates user-specific attributes, including age, gender, and prior VR experience, directly into the causal structure of the BN. This enables interpretable modeling of how individual characteristics influence physiological responses and cybersickness severity.
- We develop a modular two-stage susceptibility modeling approach that first learns a latent user-specific susceptibility score from personalization features and then incorporates it into the Bayesian Network as an intermediate variable, enabling scalable personalization while controlling model complexity.
- We proposed a stratified Bayesian modeling strategy that trains specialized probabilistic models for homogeneous user groups, capturing heterogeneous cybersickness dynamics across diverse user populations.

Our results show that the proposed personalization method improves prediction performance across both datasets. On the Maze dataset, accuracy increases from 74.20% with the non-personalized baseline to 84.79% with the stratified personalization strategy, while on the Simulation21 dataset, accuracy improves from 68.30% to 77.45%. These improvements are also reflected in F1 score, Cohen's κ, and Spearman correlation. Among the three personalization strategies, the stratified and two-stage approaches achieve the strongest overall performance across both datasets while maintaining low inference latency and efficient execution. We further compare our method with the personalization strategies of Tasnim et al. [54], including transfer learning and demographic data grouping on DeepTCN models [15]. On the Maze dataset, transfer learning and data grouping achieve accuracies of 61.52% and 60.85%, compared with 84.79% for the proposed stratified Bayesian model. On the Simulation21 dataset, transfer learning and data grouping achieve 74.55% and 73.80%, while our method reaches 77.45%.

2 Related Studies

This section describes the previous studies related to the prediction of cybersickness, individual differences in cybersickness susceptibility, and personalization techniques.

2.1 Cybersickness and Measurement Approaches

Cybersickness is the most common and prominent challenge in the VR experience, often described as a collection of symptoms similar to motion sickness, including nausea, dizziness, eye strain, disorientation, and cold sweats [39]. Even though its root cause is yet unknown, several well-established theories have tried to define it in terms of biophysiological and cognitive perspectives [3]. Among them, the most promising one is the sensory-conflict theory, according to which cybersickness typically occurs due to a mismatch between visual and vestibular cues [2, 18]. However, the Postural Instability Theory [34] argues that discomfort arises when users fail to maintain balance in an unfamiliar environment. To measure cybersickness, researchers rely on both subjective and objective methods [14]. The Simulator Sickness Questionnaire (SSQ) is the most widely used tool, consisting of 16 items categorized into nausea, oculomotor strain, and disorientation [19]. Other subjective tools include the Virtual Reality Sickness Questionnaire (VRSQ), which simplifies SSQ to 9 items focusing on two core symptoms [21], and the Fast Motion Sickness Scale (FMS), a quick 0–20 scale self-assessment used during VR exposure [20]. However, since self-reports can be delayed or biased, recent work increasingly incorporates objective physiological signals such as heart rate (HR), electrodermal activity (EDA), eye blink rate, gaze behavior, and pupil dilation, which offer real-time insights into cybersickness [36,47]. Together, these techniques help researchers and developers better understand, predict, and mitigate cybersickness across diverse users and virtual environments.

2.2 The Role of Individual Differences in Cybersickness

Individual differences play an important role in how people experience cybersickness in VR [55]. People respond differently to immersive environments due to a variety of personal characteristics such as age, gender, and even prior VR exposure [40, 42, 44]. For example, older adults might experience less cybersickness than younger adults due to slower sensory or motor processing [7]. Research also suggests that females often report higher levels of cybersickness compared to males, which may be related to biological or perceptual differences [37, 47]. Inexperienced users tend to report more symptoms during their first sessions, while those with more exposure often develop tolerance over time [57, 61]. Certain types of VR simulation or content, such as those with high levels of visual motion, rapid scene changes, or mismatches between visual and vestibular cues, are more likely to induce cybersickness [40]. Recognizing and accounting for these individual differences is essential when designing virtual experiences that are inclusive, comfortable, and effective for a wide range of users. Personalizing content delivery, adjusting system settings, or incorporating adaptive feedback based on user profiles can help reduce the onset of symptoms [45]. As VR continues to expand into everyday applications, considering individual variability will be key to improving user satisfaction and long-term usability.

2.3 Cybersickness Prediction

Recent research has increasingly applied machine learning and probabilistic methods to automate the prediction of cybersickness by analyzing patterns in physiological signals, behavioral data, and system parameters [14,15,49,50,64]. Machine learning models such as Convolutional Neural Networks (CNNs), Long Short-Term Memory networks (LSTMs), and hybrid CNN-LSTM architectures have been widely used to process time-series physiological data like heart rate, galvanic skin response, eye movements, and EEG signals [23, 49, 50]. Although these methods can achieve high accuracy, they are typically seen as black-box models, limiting their transparency [53]. To address this gap, researchers developed explainable AI models for cybersickness prediction, which will not only provide the prediction result, but also provide the concrete explainability [26, 27]. In parallel, probabilistic models such as Bayesian Networks are the most recent integration in this domain, which are especially valuable due to their interpretability and ability to encode causal relationships under uncertainty [4,41,64]. They allow researchers to integrate domain knowledge, such as assuming system-level features as root causes, and explore design choices such as

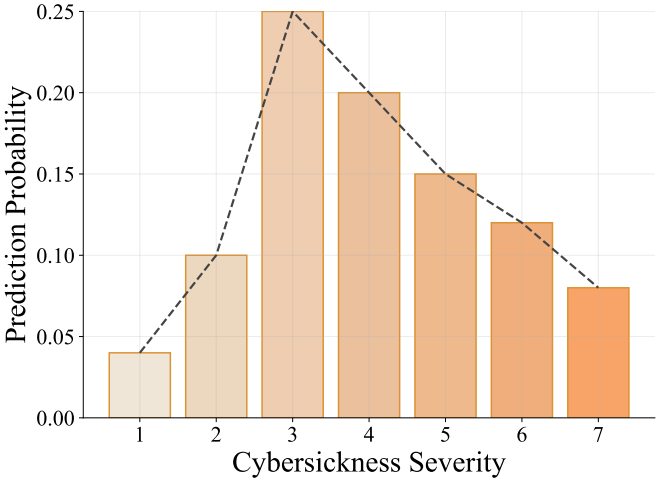

Figure 2: Example of the posterior prediction by the Bayesian Network on the Maze dataset. Based on the observed system and physiological parameters, the model estimates the probability distribution across different levels of cybersickness severity.

including user-specific factors as direct inputs or as latent variables [64]. However, challenges remain in deploying these models for real-time, user-aware applications, especially when accounting for individual variability of cybersickness [48, 52]. As the field progresses, combining probabilistic reasoning with personalization techniques could offer scalable and interpretable solutions for real-time cybersickness prediction and mitigation [63].

2.4 Personalized Prediction Models of Cybersickness

Recent research has explored personalized cybersickness prediction to account for individual variability. Most approaches use machine learning models trained on multimodal physiological and behavioral signals, such as EEG, eye tracking, and motion data [30, 58]. These methods often achieve high classification accuracy by leveraging complex neural architectures or multimodal feature fusion, such as hybrid EEG models or cross-modal neural networks designed to capture temporal and spatial dependencies across signals [8]. Similarly, recent work in consumer VR environments has explored scalable prediction frameworks using non-invasive signals such as head motion and eye tracking, demonstrating that personalized representations can be learned through deep neural models and cross-modal alignment mechanisms [56]. In parallel, other studies have investigated personalization through transfer learning, sample weighting, or demographic grouping strategies that adapt a global model to individual users [54]. While these data-driven approaches have shown promising improvements in predictive accuracy, they typically rely on complex neural architectures that provide limited interpretability and do not explicitly model the probabilistic relationships between user characteristics, physiological responses, and cybersickness outcomes [66]. In contrast, probabilistic frameworks enable the incorporation of domain knowledge and the structured representation of uncertainty [63]. While age, gender, and prior VR experience are often studied, prior work shows mixed findings on how these factors affect cybersickness. This motivates a probabilistic framework. Our Bayesian Network treats demographic factors as uncertain variables and learns their conditional relationships from the data, allowing it to estimate their influence without relying on broad assumptions [59, 65]. Motivated by these observations, our work investigates how personalization can be explicitly incorporated into Bayesian Network models, enabling interpretable reasoning about individual differences while maintaining strong predictive performance across diverse VR conditions.

3 Methodology

This section provides a detailed description of the personalization strategies, datasets, and baseline approach design.

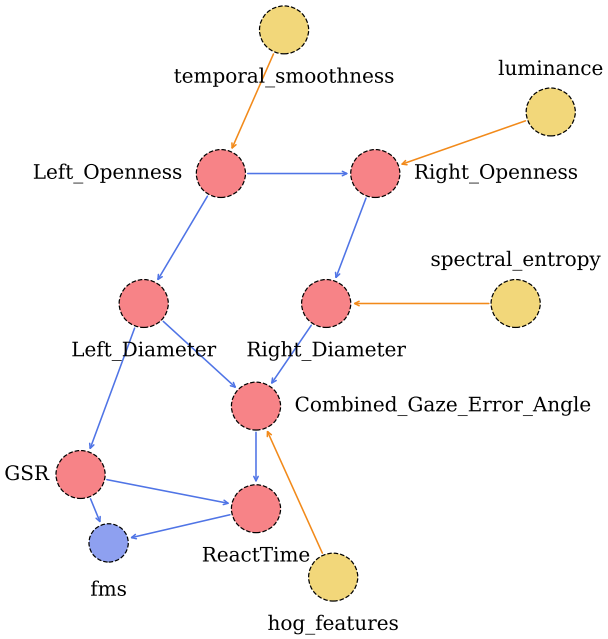

Figure 3: BN structure for cybersickness prediction, showing dependencies among physiological, system, and behavioral features. Red nodes represent physiological attributes, yellow nodes represent system factors, and the blue node represents the FMS target outcome.

3.1 Bayesian Network Model

A Bayesian Network is a directed acyclic graph $G = (V, E)$ where nodes V represent random variables and edges E encode conditional dependencies. The joint probability distribution factorizes as:

$$P(X_1, ..., X_n) = \prod_{i=1}^{n} P(X_i | Pa(X_i)) \quad (1)$$

where $Pa(X_i)$ denotes the parent nodes of X_i. In our cybersickness model, we instantiate a hierarchy informed by domain knowledge to ensure interpretability and compactness (Figure 2). An example BN for cybersickness prediction is shown in Figure 3 where the system parameters (e,g., optical flow, HOG features) act as exogenous drivers, personal features as user-specific moderators, physiological parameters as mediators, and the FMS score as the outcome. Structure learning uses a hill-climbing search with BDeu scoring (ESS = 10) under constraints that prevent edges into system parameters and prevent outgoing edges from the outcome node. After learning, nodes that are neither ancestors nor descendants of the outcome are pruned. Parameters are estimated with a Bayesian estimator and Laplace smoothing; inference uses variable elimination with discretized evidence to produce probabilistic ordinal predictions. We enforce hierarchical constraints:

- **System (Visual Scene) Parameters (e.g., optical flow, HOG features)** $S = \{s_1, ..., s_m\}$: Root nodes (no incoming edges)
- **Personalization Parameters** $I = \{i_1, i_2, i_3\}$: Age, Gender, VR-Experience
- **Physiological Parameters** $H = \{h_1, ..., h_k\}$: Intermediate nodes
- **Output** Y: FMS score (cybersickness severity)

This organization encodes plausible causal directionality (stimulus → response → outcome) while allowing the learning algorithm to discover how personal characteristics modify the pathways between stimuli and outcomes.

3.2 Baseline Model Without Personalization (BN_Strategy0)

This baseline strategy excludes all personalization features and focuses solely on the relationship between system parameters, physiological responses, and cybersickness severity. The Bayesian Network is learned under the previously described hierarchical constraints, where system parameters serve as root nodes and physiological parameters mediate the effect on the FMS outcome. Since no personal features are included, the model captures a generalized mapping from external stimuli to internal responses and ultimately to cybersickness, treating all users uniformly.

3.3 State of the Art Deep Learning Baselines

We implemented several deep learning baseline models for comparison in this study. The model architectures and training strategies were adapted from the work of Tasnim et al. [54], which provides validated temporal learning approaches for physiological signal analysis in immersive environments. To ensure a fair comparison, all deep learning baselines used the same task setting as the proposed Bayesian models, with participant-level 5-fold cross-validation and a 20% inner validation split for hyperparameter tuning.

Long Short-Term Memory Long Short-Term Memory (LSTM) networks are deep learning architectures that are designed to process sequential time series data by learning dependencies across time [11]. These models are particularly effective for predicting cybersickness because physiological responses such as heart activity, skin conductance, and motion behavior evolve gradually over time rather than appearing as isolated events [3, 50]. The internal memory structure of LSTM allows the network to store relevant information from earlier time steps and use it when making predictions at later moments [10]. This capability enables the model to capture long-term symptom progression patterns, which is important when estimating sickness severity during immersive experiences. In addition, LSTM models can handle noise and variability in biological signals because the gating mechanism selectively filters useful information while ignoring irrelevant fluctuations [12, 65]. As a result, these networks have been widely adopted as baseline models for temporal prediction tasks in immersive environments.

Convolutional Neural Network (CNN) with LSTM The hybrid architecture that combines CNN with LSTM networks, shortly CNN-LSTM, is designed to analyze complex structured signals that contain both spatial and temporal information [14, 51]. The convolutional component acts as an automatic feature extractor that identifies local patterns and correlations within the sensor data. For example, it can detect relationships between multiple physiological channels or motion signals that occur simultaneously [14]. After these spatial patterns are extracted, the LSTM component processes the features sequentially to capture how they change over time. This combination allows the model to learn both instantaneous relationships and temporal evolution, which is essential for understanding cybersickness dynamics [15]. Because immersive experience responses often involve interactions between multiple biological systems, the joint architecture provides improved predictive capability compared to using either component alone [10].

Deep Temporal Convolutional Network The Deep Temporal Convolutional Network (Deep-TCN) is a powerful architecture designed specifically for long sequence modeling using convolutional operations rather than recurrent connections [28]. One of its main advantages is the use of dilated convolutions, which allow the network to capture long-range temporal dependencies without significantly increasing computational cost. Dilated convolution expands the receptive field so the model can analyze both short-term variations and long-term trends simultaneously [54]. In addition, residual connections help stabilize training by allowing information to flow directly across layers, which reduces the risk of gradient degradation in deep networks [29]. Weighted normalization techniques further improve learning stability and generalization performance. These design choices enable the network to extract complex temporal patterns from physiological signals and motion data, making it highly suitable for predicting sickness severity in immersive environments [16].

Transfer Learning on Deep Temporal Convolutional Network Transfer learning is a strategy that adapts a general model to individual users by leveraging knowledge learned from a larger population dataset [43]. Initially, the network is trained on data collected from many participants so that it learns general physiological and behavioral patterns. To personalize the model for a new user, the final prediction layer is replaced with a new layer that starts with random parameters [54]. This new layer is then trained using only the data from the target individual while keeping the earlier layers fixed. The earlier layers act as a shared feature extractor that provides meaningful representations for the personalized prediction task [25, 31]. While this approach improves individual accuracy, retraining the model for each user demands significant computation and time, limiting scalability and real-time deployment in immersive systems [62].

Data Grouping on Deep Temporal Convolutional Network We implemented the data grouping personalization strategy following the approach proposed by Tasnim et al. [54]. This method organizes participants into demographic groups before model training instead of using a single global model or performing individual-level retraining. Users are grouped based on shared attributes such as age categories and gender, and a separate Deep-TCN is trained using only the data from each corresponding group. During inference, a new user is assigned to the appropriate demographic group based on their profile, and the corresponding specialized model is used for prediction. This grouping strategy reduces variability within the training data and allows the model to learn patterns that are more specific to homogeneous user populations. The approach provides a practical compromise between generalized population models and fully personalized transfer learning methods, while avoiding the computational overhead associated with retraining models for each user. In this work, we adopt this method as a deep learning personalization baseline to compare against the proposed Bayesian personalization strategies.

3.4 Proposed Personalized Cybersickness Models

We propose and evaluate three methods for incorporating individual features into the BN structure. Figure 4 provides an overview of the proposed personalization strategies and their integration within the BN framework.

3.4.1 Direct Integration for Personalization (BN_Strategy1)

Personalization features, including age, gender, and prior VR experience, are added as root nodes alongside system-level features. This design allows the structure learning algorithm to capture both direct and indirect influences of these features on intermediate physiological signals and the cybersickness outcome:

$$P(S,I,H,Y) = \prod_{s\in S} P(s) \prod_{i\in I} P(i) \prod_{h\in H} P(h \mid Pa(h)) \cdot P(Y \mid Pa(Y)) \quad (2)$$

The parent sets $Pa(h)$ may include both system variables from S and personalization variables from I, enabling explicit modeling of pathways such as Age $\rightarrow$ GSR $\rightarrow$ FMS. This approach preserves the full predictive signal of personalization features without compression and allows the BN to uncover interaction effects between personal and system-level drivers. All personalization and system parameters are declared as root nodes. The same structural constraints used in the baseline model are enforced: no edges are allowed in system parameters, and no edges are allowed out of the outcome variable. To ensure tractability and generalizability, we limit the maximum number of parents per node and apply BDeu scoring with an equivalent sample size of 10 to encourage sparsity in the learned graph. Variables that do not influence the FMS node either directly or indirectly are pruned after learning to simplify the model.

This strategy retains the full predictive signal of each personalization features without collapsing them into a single latent variable. It is well-suited for scenarios where the number of personalization features is moderate and trustworthy. While it offers high interpretability and prediction accuracy, its performance may degrade if the personalization

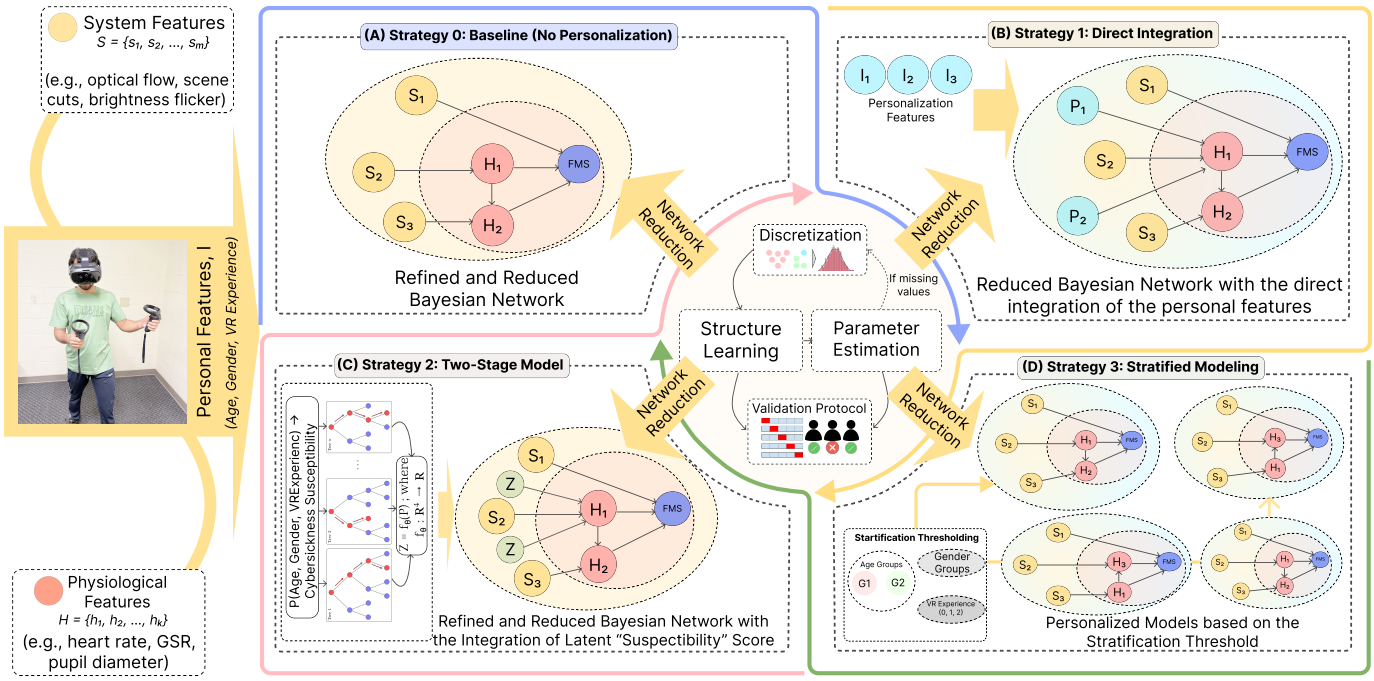

Figure 4: System overview of the proposed personalization techniques to predict cybersickness level (FMS). (A) Strategy-0: No personalization, (B) Strategy-1: Direct Integration of the personalization features (age, gender, VR Experience) as the root nodes, (C) Strategy-2: Two-stage modeling with an embedding learnt from a random forest model, (D) Strategy-3: Stratified learning for each of the personalization features.

space becomes too large, as conditional probability tables (CPTs) can grow exponentially with the number of parents per node. However, it is especially effective in our combined dataset setting, where the type of simulation provides important contextual cues that interact with both system and physiological signals to shape the susceptibility to cybersickness.

3.4.2 Two-Stage Model for Personalization (BN_Strategy2)

A susceptibility score Z is learned from personalization features using a separate model f_θ, then incorporated into the BN:

$$P(S,I,Z,H,Y) = \prod_{s\in S} P(s) \prod_{i\in I} P(i) \cdot P(Z \mid I) \prod_{h\in H} P(h \mid Pa(h)) \cdot P(Y \mid Pa(Y)), \tag{3}$$

where

$$Z = f_\theta(I) \quad \text{with } f_\theta : \mathbb{R}^d \to \mathbb{R}. \tag{4}$$

Here, I includes personalization features such as Age, Gender, and Prior VR Experience.

We use Random Forest regression for f_θ to model the non-linear relationship between personalization variables I and the susceptibility score Z. This design addresses three practical considerations. First, when the number of personalization variables grows, including them directly as parents in the Bayesian Network can significantly increase the size of the conditional probability tables (CPTs), leading to parameter explosion and data sparsity. For a node with r states and d discrete parents with cardinalities $k_1, \ldots, k_d$, the CPT requires $r\prod_{j=1}^{d} k_j$ parameters, which quickly becomes impractical as d increases. Second, the compact scalar representation Z acts as a bottleneck that summarizes the personalization information while preserving its predictive signal, thereby keeping the BN structure smaller and easier to interpret. Third, the learned mapping f_θ enables *cold-start* prediction for new users using only their personalization attributes I, without requiring retraining of the Bayesian Network. Overall, the two-stage approach balances predictive performance and model complexity, avoiding over-parameterization while maintaining a modular and interpretable architecture.

3.4.3 Stratified Modeling for Personalization (BN_Strategy3)

Separate Bayesian Network models are trained for different user groups defined by a stratification variable $k \in I$ (e.g., Age group, Gender, or VR Experience):

$$\begin{aligned} P(S,H,Y \mid k=c) = &\prod_{s\in S} P(s \mid k=c) \\ &\times \prod_{h\in H} P(h \mid Pa(h), k=c) \\ &\times P(Y \mid Pa(Y), k=c) \end{aligned} \tag{5}$$

Final predictions are aggregated using sample-weighted accuracy across strata:

$$\text{Accuracy} = \frac{\sum_c n_c \cdot \text{Acc}_c}{\sum_c n_c} \tag{6}$$

where n_c denotes the number of samples in stratum c, and Acc_c represents the prediction accuracy obtained by the model trained for that stratum.

Stratification acknowledges that population subgroups can exhibit distinct mechanisms (e.g., habituation effects by VR experience). Training separate BNs per group avoids forcing one model to explain heterogeneous dynamics and can improve fairness by tailoring structures to each segment.

We define discrete groups for a chosen attribute (e.g., Young/Older by Age), remove all personal variables from each stratum's training data, and learn a BN with the same constraints as the baseline. During evaluation, we select the appropriate model based on a user's group membership and aggregate performance with sample-weighted accuracy across groups. While Strategy 3 captures subgroup-specific patterns, maintaining separate models for each demographic combination can limit scalability. This motivates Strategy 2, which maps multiple personal attributes to one latent susceptibility score and supports additional traits without requiring separate models.

3.5 Dataset Descriptions

In this study, we used two open-sourced datasets comprising users' physiological, eye-tracking, and head-tracking data, as well as the

system parameters. Both of the datasets are described below in detail.

3.5.1 Maze Dataset

This publicly available Maze dataset comprises participants' data for a virtual navigation task, which was designed to explore the relationships between cybersickness, cognitive load, attention, and working memory during real walking tasks in VR [50]. The dataset includes data from 36 participants (23 male, 13 female) with a mean age of 25.67 years (SD = 7.22). During the experiment, participants navigated randomly generated mazes in VR via physical walking for 15 minutes, while also performing dual tasks involving attention and working memory. The dataset captures a rich set of multimodal features, including heart rate (HR), galvanic skin response (GSR), eye tracking, head tracking, and self-reported measures such as the FMS scores. Each session is timestamped and synchronized across modalities. This dataset enables analysis of cybersickness under natural locomotion conditions, making it a good choice for the investigation of the VR navigation task.

3.5.2 Simulation21 Dataset

The Simulation21 dataset [15] was originally introduced by Islam et al. and contains multimodal data collected from 30 participants (mean age of 30.04 years, and standard deviation: 4.12 years) exposed to passive VR video content while seated. The dataset captures physiological signals, including HR, EDA, respiration, and head movement, along with eye tracking metrics such as gaze direction and pupil dilation. Each participant also reported their cybersickness levels using the FMS scale after viewing various 1-minute video stimuli that varied in intensity. In addition to sensor data, the dataset includes demographic information such as age, gender, and VR experience. The diversity in stimulus types and the multimodal nature of the recordings make this dataset well-suited for evaluating cybersickness prediction models across different user profiles and video conditions.

3.6 Data Preprocessing

The datasets used in this study consist of multimodal time series signals collected during immersive VR environments. Although physiological, eye tracking, and system measurements are temporally synchronized, raw sensor recordings naturally contain noise, transient spikes, and occasional missing samples caused by sensor imperfections and motion artifacts. To ensure data quality and temporal consistency, several preprocessing steps were applied. Outliers were first removed using a z-score criterion that discards observations exceeding three standard deviations from the mean. The remaining signals were smoothed using a rolling average window of five samples to reduce high-frequency fluctuations while preserving the underlying physiological trends. All modalities were then standardized to a common sampling rate of 1 Hz by averaging measurements within each time interval, followed by z score normalization to place features on comparable scales. The synchronized signals were subsequently organized into fixed temporal segments by summarizing consecutive one-second samples within a 30-second window for each participant, producing stable observations that retain the temporal dynamics of the multimodal responses [54]. Samples containing missing values were removed to ensure complete feature vectors before modeling. These steps produce a consistent representation of the multimodal signals suitable for reliable cybersickness prediction.

4 Experimental Setup & Evaluation

This section describes the experimental protocol used to evaluate the proposed personalization strategies. It outlines the computational environment, implementation details of the Bayesian models, and the evaluation procedure used to ensure fair comparison across all methods. The same preprocessing, training configuration, and validation protocol are applied consistently to all models.

4.1 Device Configurations

All experiments were executed on a Dell Precision 3680 workstation running Microsoft Windows 11 Enterprise (Build 26100) in a standalone configuration. The system was equipped with an Intel 13th Gen Intel64 processor (Family 6, Model 183, Stepping 1, base frequency 2.1 GHz), 16 GB physical memory, and an NVIDIA GeForce RTX 4090 GPU (24 GB VRAM) for accelerated computation.

4.2 Implementation Details

The experiments were implemented with the following common settings, applied to all strategies for comparability:

- **Discretization**: Continuous features are normalized and discretized into 10 bins using k-means clustering to preserve ordinal relationships while mitigating the effect of high-variance signals. Clustering is fit only on training folds and applied to held-out participants to avoid leakage. The target FMS labels are discrete ordinal scores, with the score range depending on the dataset.
- **Structure Learning**: We use hill-climbing search with BDeu scoring under hierarchical constraints that (i) forbid edges from personal or physiological parameters into system parameters, and (ii) forbid outgoing edges from the FMS node. A maximum parent limit per node controls the conditional probability table (CPT) size. The equivalent sample size (ESS) for BDeu is set to favor sparse, interpretable structures.
- **Parameter Estimation**: CPT parameters are estimated via maximum likelihood with Laplace smoothing. Missing values are rare; when they occur, training-fold medians are imputed before discretization.
- **Two-Stage Model f_θ (Strategy 2)**: A Random Forest regressor with 200 trees predicts a continuous susceptibility score Z from personal features. The maximum tree depth is tuned via inner cross-validation on the training set. The output Z is discretized into 5 bins before inclusion in the BN.
- **Stratification thresholds (Strategy 3)**:
  - *Age:* Two age groups were defined based on the median age from the datasets, following prior findings that age influences susceptibility to cybersickness. Group 1 included younger participants aged lower than the median, while Group 2 included older participants aged higher than the median. This grouping strategy was adapted from Tasnim et al. [54].
  - *Gender:* Categorical groups based on original encoding
  - *VR Experience:* Ordinal variable representing prior exposure to VR (0 = none, 1 = occasional, 2 = frequent).

4.3 Evaluation Protocol

Model performance is evaluated using 5-fold participant-level cross-validation to ensure that samples from the same participant do not appear in both training and test sets. For each fold, the participant-level training data is further split so that 20% of the training portion is used as a validation set for hyperparameter tuning and model selection, while the remaining data is used for model training. This validation-based tuning procedure prevents information leakage from the test set and follows the evaluation strategy adopted in prior works [1, 50]. For all reported metrics, Accuracy refers to the percentage of held-out participant samples with the correct predicted FMS level. In addition to standard evaluation metrics such as accuracy, F1 score, Cohen's κ, and Spearman correlation, we also report the Posterior Probability Certainty Index (PPCI), which measures the confidence of the Bayesian Network predictions by quantifying how concentrated the posterior probability distribution is around the predicted cybersickness level [38]. Higher PPCI values indicate stronger predictive certainty and provide an additional perspective on model reliability beyond classification accuracy.

5 Results

This section summarizes the experimental results obtained from evaluating the proposed Bayesian personalization strategies and baseline

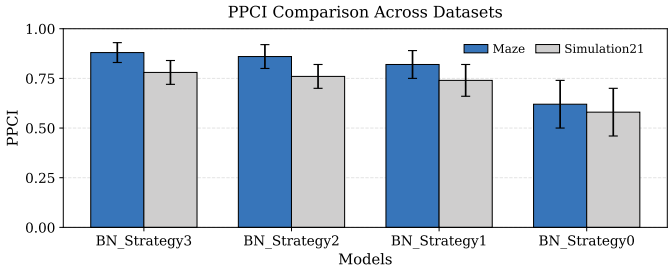

Figure 5: Posterior Probability Certainty Index (PPCI) comparison across Bayesian personalization strategies for the Maze and Simulation21 datasets. PPCI quantifies the confidence in probabilistic predictions by measuring how concentrated the posterior probability distribution is around the predicted cybersickness level. Higher values indicate greater prediction certainty.

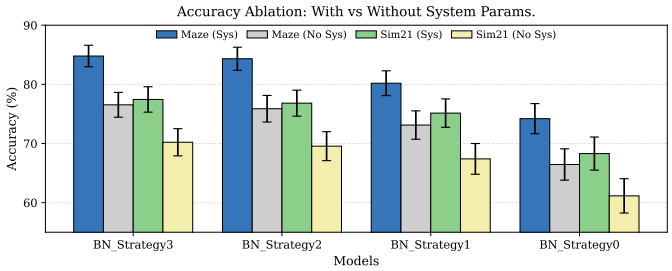

Figure 6: Ablation analysis illustrating the impact of system parameters on prediction accuracy across personalization strategies for the Maze and Simulation21 datasets. Accuracy is reported for models trained with the full feature set and for models trained without system parameters. Including system parameters consistently results in higher prediction accuracy across both datasets and all personalization strategies.

models on the Maze and Simulation21 datasets. Performance is analyzed using prediction accuracy, agreement metrics, and computational efficiency.

5.1 Overall Performance Across Personalization Strategies

Tables 1 and 2 summarize the performance of all evaluated models across the Maze and Simulation21 datasets. On the Maze dataset, BN_Strategy3 achieved the highest accuracy of $84.79 \pm 1.82\%$ with an F1 score of 0.83 ± 0.02, followed closely by BN_Strategy2 with $84.33 \pm 1.95\%$ accuracy and BN_Strategy1 with $80.19 \pm 2.10\%$. The non-personalized baseline model obtained $74.20 \pm 2.55\%$ accuracy with an F1 score of 0.71 ± 0.04. Deep learning models produced lower performance, with Transfer Learning reaching $61.52 \pm 3.15\%$ accuracy, CNN LSTM reaching $59.33 \pm 3.25\%$, DeepTCN reaching $58.64 \pm 3.30\%$, and LSTM reaching $57.26 \pm 3.45\%$. On the Simulation21 dataset, BN_Strategy3 again obtained the highest accuracy of $77.45 \pm 2.15\%$ with an F1 score of 0.75 ± 0.03, followed by BN_Strategy2 with $76.82 \pm 2.20\%$ and BN_Strategy1 with $75.14 \pm 2.40\%$. The baseline Bayesian model achieved $68.30 \pm 2.80\%$ accuracy with an F1 score of 0.65 ± 0.05, while Transfer Learning reached $74.55 \pm 2.90\%$, CNN LSTM reached $71.42 \pm 3.10\%$, DeepTCN reached $70.95 \pm 3.15\%$, and LSTM reached $69.88 \pm 3.20\%$. Some prior studies report higher accuracy for deep learning baselines using randomized frame-level splits. However, these splits may place data from the same participant in both training and testing sets. In our evaluation, the 57% to 70% accuracy range reflects participant-level cross-validation, where test participants are unseen during training.

5.2 Agreement and Correlation Metrics

Agreement-based evaluation metrics (Cohen's κ and Spearman correlation) are summarized in Tables 1 and 2. These metrics show consistent trends across both datasets. On the Maze dataset, BN_Strategy3 achieved the highest Cohen's kappa value of 0.81 ± 0.03 and Spearman correlation of 0.85 ± 0.02, compared to the baseline values of 0.68 ± 0.05 and 0.73 ± 0.04, respectively. BN_Strategy2 produced Cohen's kappa of 0.80 ± 0.03 and Spearman correlation of 0.84 ± 0.02, while BN_Strategy1 produced values of 0.76 ± 0.04 and 0.81 ± 0.03. Deep learning models produced lower agreement scores across all metrics. On the Simulation21 dataset, BN_Strategy3 achieved Cohen's kappa of 0.72 ± 0.04 and Spearman correlation of 0.76 ± 0.03, while the baseline model achieved 0.61 ± 0.06 and 0.66 ± 0.05. BN_Strategy2 and BN_Strategy1 produced slightly lower but comparable values relative to Strategy3, and deep learning models again showed lower agreement metrics across the dataset.

5.3 Accuracy Gain Relative to Baselines

Percentage accuracy improvements relative to baseline models are summarized in Table 3. On the Maze dataset, BN_Strategy3 achieved a gain of 14.27% compared to the non-personalized baseline and gains of 44.59% and 48.07% compared to DeepTCN and LSTM, respectively. BN_Strategy2 produced gains of 13.64% compared to the baseline and gains above 43% compared to temporal deep learning models. BN_Strategy1 achieved an 8.07% improvement relative to the baseline. On the Simulation21 dataset, BN_Strategy3 achieved a gain of 13.40% compared to the baseline and gains of 9.16% and 10.82% compared to DeepTCN and LSTM. BN_Strategy2 produced a gain of 12.48% compared to the baseline, while BN_Strategy1 achieved a gain of 10.01%. Transfer Learning and Data Grouping methods produced smaller gains relative to deep learning baselines.

5.4 Ablation Study: PPCI

The PPCI comparison across personalization strategies is shown in Figure 5. The highest PPCI values were observed for BN_Strategy3 across both Maze and Simulation21 datasets, followed by BN_Strategy2 and BN_Strategy1, while the baseline model showed the lowest PPCI values. The ablation analysis results are presented in Figure 6, where models trained with system parameters consistently achieved higher accuracy compared to models trained without system parameters across both datasets and all personalization strategies. Accuracy reductions were observed when system parameters were removed from the feature set.

5.5 Ablation Study: Impact of System (Visual Scene) Parameters

We evaluate the contribution of system-level visual scene features through an ablation study in which models are trained with and without system parameters. As illustrated in Figure 6, removing system parameters consistently reduces prediction accuracy across all personalization strategies and both datasets. On the Maze dataset, BN_Strategy3 achieves 84.79% accuracy with system parameters, which decreases to 76.54% when these features are removed. Similar reductions are observed for BN_Strategy2 and BN_Strategy1, where accuracy drops from 84.33% to 75.88% and from 80.19% to 73.12%, respectively, while the baseline BN_Strategy0 decreases from 74.20% to 66.45%. The same trend appears in the Simulation21 dataset, where BN_Strategy3 decreases from 77.45% to 70.21%, BN_Strategy2 from 76.82% to 69.55%, BN_Strategy1 from 75.14% to 67.40%, and the baseline from 68.30% to 61.15%. These results indicate that system parameters provide critical contextual information about the VR environment and visual stimuli, enabling the Bayesian Network to better capture the interaction between external conditions, physiological responses, and individual susceptibility to cybersickness.

6 Discussion

This section discusses the implications of the experimental results and examines how personalization improves cybersickness prediction and supports adaptive VR systems.

6.1 Impact of Personalization on Cybersickness Prediction

The results show that personalization improves cybersickness prediction across both datasets and metrics. On the Maze dataset, the baseline

Table 1: Maze Dataset [50]: Performance Comparison for the Models Evaluating Accuracy, Real Time Usability, and Statistical Reliability. All baseline models were evaluated under identical task conditions, utilizing the same participant-level 5-fold cross-validation.

Model	Accuracy (%)	F1 Score	Cohen's κ	Spearman ρ	Mean Latency (ms)	P95 Latency (ms)	Throughput (samples/s)
BN_Strategy3	84.79 ± 1.82	0.83 ± 0.02	0.81 ± 0.03	0.85 ± 0.02	0.34	0.39	2941.18
BN_Strategy2	84.33 ± 1.95	0.82 ± 0.02	0.80 ± 0.03	0.84 ± 0.02	6.12	6.80	163.40
BN_Strategy1	80.19 ± 2.10	0.79 ± 0.03	0.76 ± 0.04	0.81 ± 0.03	8.45	9.20	118.34
BN_Strategy0	74.20 ± 2.55	0.71 ± 0.04	0.68 ± 0.05	0.73 ± 0.04	0.33	0.38	3030.30
TransferLearning	61.52 ± 3.15	0.59 ± 0.05	0.55 ± 0.06	0.62 ± 0.05	1.96	2.09	510.20
DataGrouping	60.85 ± 3.20	0.58 ± 0.05	0.54 ± 0.06	0.61 ± 0.05	1.65	1.82	606.06
CNN LSTM	59.33 ± 3.25	0.57 ± 0.06	0.52 ± 0.07	0.59 ± 0.06	0.42	0.47	2380.95
DeepTCN	58.64 ± 3.30	0.56 ± 0.06	0.51 ± 0.07	0.58 ± 0.06	1.58	1.74	632.91
LSTM	57.26 ± 3.45	0.54 ± 0.07	0.49 ± 0.08	0.56 ± 0.07	0.40	0.43	2500.00

Table 2: Simulation21 Dataset [15]: Performance Comparison for the Models Evaluating Accuracy, Real Time Usability, and Statistical Reliability. All baseline models were evaluated under identical task conditions, utilizing the same participant-level 5-fold cross-validation.

Model	Accuracy (%)	F1 Score	Cohen's κ	Spearman ρ	Mean Latency (ms)	P95 Latency (ms)	Throughput (samples/s)
BN_Strategy3	77.45 ± 2.15	0.75 ± 0.03	0.72 ± 0.04	0.76 ± 0.03	0.35	0.41	2857.14
BN_Strategy2	76.82 ± 2.20	0.74 ± 0.03	0.71 ± 0.04	0.75 ± 0.03	5.95	6.65	168.07
BN_Strategy1	75.14 ± 2.40	0.72 ± 0.04	0.69 ± 0.05	0.73 ± 0.04	8.20	9.05	121.95
BN_Strategy0	68.30 ± 2.80	0.65 ± 0.05	0.61 ± 0.06	0.66 ± 0.05	0.34	0.40	2941.18
TransferLearning	74.55 ± 2.90	0.71 ± 0.04	0.68 ± 0.05	0.72 ± 0.04	1.85	1.98	540.54
DataGrouping	73.80 ± 2.95	0.70 ± 0.04	0.67 ± 0.05	0.71 ± 0.04	1.60	1.75	625.00
CNN LSTM	71.42 ± 3.10	0.68 ± 0.05	0.65 ± 0.06	0.69 ± 0.05	0.44	0.50	2272.73
DeepTCN	70.95 ± 3.15	0.67 ± 0.05	0.64 ± 0.06	0.68 ± 0.05	1.62	1.78	617.28
LSTM	69.88 ± 3.20	0.66 ± 0.05	0.62 ± 0.06	0.67 ± 0.05	0.38	0.42	2631.58

model without personalization achieved an accuracy of 74.20%, while the best personalized approach, BN_Strategy3, reached 84.79%, representing an absolute improvement of 10.59 percentage points and a relative improvement of approximately 14.27%. On the Simulation21 dataset, accuracy increased from 68.30% to 77.45% using BN_Strategy3, corresponding to an absolute improvement of 9.15 percentage points and a relative improvement of approximately 13.40%. Agreement-based metrics show the same trend: on Maze, Cohen's kappa increased from 0.68 to 0.81, while Spearman's correlation increased from 0.73 to 0.85. On Simulation21, Cohen's kappa increased from 0.61 to 0.72, and Spearman's correlation increased from 0.66 to 0.76. Compared with deep learning baselines, BN_Strategy3 also performed better. On Maze, the strongest deep learning model achieved 61.52% accuracy, whereas BN_Strategy3 achieved 84.79%, an improvement of more than 23 percentage points. On Simulation21, deep learning models achieved accuracies between 69.88% and 74.55%, while the personalized Bayesian model achieved 77.45%. Overall, age, gender, and prior VR experience add useful user-specific information beyond physiological and system parameters, enabling the probabilistic model to better capture variability in cybersickness responses. Although BN_Strategy3 achieves the strongest results, its demographic stratification divides the 30 to 36 participants into smaller subgroups, which may affect statistical reliability. Therefore, BN_Strategy3 should be interpreted as an upper bound for personalization performance. For practical deployment, BN_Strategy2 is preferred because it combines multiple personal attributes into a single latent susceptibility score without splitting the training pool.

6.2 Advantages of BNs for Personalized Prediction

The BN formulation provides several practical advantages for personalized cybersickness prediction. First, the probabilistic structure allows the model to represent uncertainty in both physiological responses and user characteristics, which is important in real-world immersive environments where sensor measurements can be noisy and incomplete. This capability highlights a key practical advantage over alternative personalization frameworks, such as the fuzzy logic approach proposed by Wang et al. [60]. While fuzzy inference systems are highly sensitive to missing inputs, as omitted variables can interrupt fuzzification and rule evaluation, our BN naturally handles missing demographic or physiological data by marginalizing over the corresponding prior distributions. Consequently, if a new user chooses not to share specific personal attributes (such as age or gender) due to privacy concerns, the model treats those variables as unobserved and continues to produce continuous risk estimates, successfully balancing predictive performance with practical and ethical deployment constraints. Second, the graphical representation enables interpretable relationships between system parameters, personal attributes, physiological responses, and predicted sickness levels. This transparency is valuable for both researchers and system designers who require insight into why predictions are generated. Third, the modular nature of the model allows different personalization strategies to be implemented without requiring extensive retraining. This flexibility supports deployment scenarios where new users or environments are introduced over time. The model also supports direct interpretability through its probabilistic structure. Unlike deep learning methods that often require post hoc explanation, the Bayesian Network allows learned relationships to be inspected through conditional probability tables (CPTs). For example, these tables can show how prior VR experience relates to physiological changes and sickness risk, helping designers understand how individual traits affect predictions. Together, these properties make Bayesian approaches well-suited for adaptive immersive systems that require both accuracy and interpretability. Finally, this probabilistic conditioning may help explain the stronger personalization effects of our model compared with deep learning baselines. With only 30 to 36 users under participant-level cross-validation, larger neural models may overfit to training participants. In contrast, the Bayesian Network uses a structured graphical model and BDeu regularization with ESS = 10, thereby supporting generalization on smaller participant-level datasets.

6.3 Computational Efficiency of the Personalized Models

In addition to predictive performance, the proposed Bayesian personalization strategies show efficient computational performance. As presented in Tables 1 and 2, BN_Strategy3 achieves a mean inference latency of 0.34 ms with a 95th percentile latency of 0.39 ms on the Maze dataset. On the Simulation21 dataset, it achieves a mean inference latency of 0.35 ms with a 95th percentile latency of 0.41 ms. The corresponding throughputs are 2941.18 and 2857.14 samples per second, respectively. BN_Strategy2 achieves a mean inference latency of 6.12 ms with a 95th percentile latency of 6.80 ms on Maze and 5.95 ms with a 95th percentile latency of 6.65 ms on Simulation21. These results correspond to throughputs of 163.40 and 168.07 samples per second, respectively. BN_Strategy1 achieves mean inference latencies of 8.45 ms on Maze and 8.20 ms on Simulation21, with 95th percentile latencies of 9.20 ms and 9.05 ms, respectively. Its corresponding throughputs

Table 3: Percentage Accuracy Gain Relative to Baselines Across Maze and Simulation21 Datasets

Method	Maze Dataset Gain (%)			Simulation21 Dataset Gain (%)		
	vs BN_Strategy0	vs DeepTCN	vs LSTM	vs BN_Strategy0	vs DeepTCN	vs LSTM
BN_Strategy3	+14.27	+44.59	+48.07	+13.40	+9.16	+10.82
BN_Strategy2	+13.64	+43.81	+47.23	+12.48	+8.27	+9.94
BN_Strategy1	+8.07	+36.76	+40.04	+10.01	+5.90	+7.53
TransferLearning	-17.10	+4.91	+7.44	+9.16	+5.07	+6.68
DataGrouping	-17.99	+3.77	+6.26	+8.07	+4.02	+5.60
CNN-LSTM	-20.03	+1.18	+3.62	+4.56	+0.66	+2.20

are 118.34 and 121.95 samples per second. All proposed Bayesian personalization strategies maintain mean inference latencies below 10 ms, supporting their use in real-time VR systems. Although these strategies incorporate personalization information, their inference costs remain low because predictions are generated through probabilistic inference over compact graphical structures. For comparison, the CNN-LSTM, DeepTCN, and LSTM baselines achieve mean inference latencies of 0.42 ms, 1.58 ms, and 0.40 ms on the Maze dataset, respectively. On the Simulation21 dataset, their corresponding mean inference latencies are 0.44 ms, 1.62 ms, and 0.38 ms. The proposed Bayesian framework, therefore, offers an efficient and interpretable approach while maintaining competitive predictive performance.

6.4 Implications for Adaptive VR Systems

The ability to predict cybersickness using personalized information has broader implications for the design of adaptive VR systems. Rather than serving only as an offline prediction tool, our framework can support real-time VR adaptation through integration with a closed-loop immersive experience. With an inference latency of 0.34 ms, the model combines live physiological indicators with available user attributes to provide continuous risk estimates. These personalized estimates can help the VR system identify increasing discomfort and apply timely adjustments, such as temporary field of view reduction, visual stabilization, or changes to visual motion and scene complexity. This capability can support user comfort, longer sessions, and improved accessibility for users with different levels of cybersickness sensitivity. The probabilistic output also supports risk-aware decision-making by allowing adaptations to be based on predicted likelihoods rather than binary classifications. This adaptive approach can help create immersive experiences that better accommodate diverse users. Overall, the findings highlight the importance of combining personalization with interpretable probabilistic modeling to support safe and user-centered VR technologies.

7 Limitations and Future Work

Although the proposed approach demonstrates strong performance across datasets, several limitations remain. **First,** the datasets used in this study represent controlled experimental environments with specific task conditions and participant populations. While they provide valuable multimodal signals, real-world virtual reality usage may involve more diverse interaction patterns, longer exposure durations, and broader demographic variability. Future work should evaluate the proposed personalization framework in large-scale deployments and in more heterogeneous user populations to confirm generalizability. **Second,** the personalization features considered in this work are limited to demographic attributes and prior experience. Other factors such as motion sickness history, cognitive traits, fatigue levels, or environmental context may also influence cybersickness susceptibility. Incorporating richer user profiles and contextual information could further improve prediction performance and provide more comprehensive personalization. In addition, the current discretization strategy simplifies continuous physiological signals into categorical representations. While this supports efficient Bayesian inference, future work may explore hybrid continuous Bayesian models or probabilistic deep learning approaches to preserve more detailed signal information. **Third,** the study focuses primarily on prediction rather than intervention. Although accurate forecasting is an important step, the ultimate goal is to enable adaptive virtual reality systems that actively reduce discomfort. Future research should integrate the prediction model with real-time adaptation mechanisms, such as dynamic motion scaling or visual stabilization, and evaluate the impact on user comfort and task performance. Longitudinal studies could also investigate how personalization evolves over repeated exposures and whether adaptive systems can support user habituation over time. Additionally, datasets used in this study include 30 to 36 participants. Although this is common for controlled multimodal VR studies, it is still limited for machine learning. This limitation supports our use of Bayesian Networks, which provide a simpler probabilistic structure and may reduce overfitting under participant-level validation. We also use BDeu scoring with an equivalent sample size of 10 to regularize learning and support reliable prediction with a modest participant pool. **Overall,** extending the framework to broader datasets, richer personalization features, and closed-loop adaptive systems represents an important direction for future research.

8 Conclusion

This work presents a personalized approach for cybersickness prediction using Bayesian Networks that integrate physiological signals, system parameters, and individual user attributes. By incorporating personalization features such as age, gender, and prior virtual reality experience, the proposed framework improves prediction accuracy across multiple datasets compared to non-personalized and deep learning baselines. The probabilistic structure enables interpretable relationships between variables while maintaining low computational latency suitable for efficient deployment. The findings demonstrate that personalization plays a critical role in modeling cybersickness susceptibility and that probabilistic methods provide an effective and practical solution for adaptive immersive systems. This research contributes toward the development of user-aware virtual reality technologies that can anticipate discomfort and support safer, more inclusive immersive experiences.

Acknowledgement

This work was supported in part by the Defense Advanced Research Projects Agency (DARPA) under grant HR00112420366. During the preparation of this manuscript, Large Language Models and Grammarly were used solely for language refinement and clarity. All research contributions, including the conceptual framework and methodology, were developed entirely by the authors. All figures are original and were created manually without the use of AI-based image generation tools.

References

[1] N. Ahmed, M. M. Hassan, M. M. Hossain, N. S. Shadin, X. Zhang, and R. Islam. Obscuring the 'who,'preserving the 'what': Targeted eye-tracking feature obfuscation in virtual reality for privacy-utility balance. In *2026 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 250–260. IEEE, 2026. 6

[2] A. R. Allred, A. R. Gopinath, and T. K. Clark. Validating sensory conflict theory and mitigating motion sickness in humans with galvanic vestibular stimulation. *Communications Engineering*, 4(1):1–11, 2025. doi: 10.1038/s44172-025-00417-2 2

[3] N. Biswas, A. Mukherjee, and S. Bhattacharya. "are you feeling sick?" – a systematic literature review of cybersickness in virtual reality. *ACM*

Comput. Surv., 56(11), June 2024. doi: 10.1145/3670008 2, 4

[4] M. E. Borsuk, C. A. Stow, and K. H. Reckhow. A bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. *Ecological Modelling*, 173(2-3):219–239, 2004. 2

[5] B. Cheung and K. Hofer. Lack of gender difference in motion sickness induced by vestibular coriolis cross-coupling. *Journal of Vestibular Research*, 12(4):191–200, 2003. 1

[6] Y. Choi et al. Early prediction of cybersickness in virtual reality using a large language model for multimodal time series data. In *Companion of the 2024 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 1–6. ACM, 2024. 1

[7] A. T. Dilanchian, R. Andringa, and W. R. Boot. A pilot study exploring age differences in presence, workload, and cybersickness in the experience of immersive virtual reality environments. *Frontiers in Virtual Reality*, 2:736793, 2021. doi: 10.3389/frvir.2021.736793 2

[8] D. I. Emegano, D. U. Ozsahin, B. Uzun, and I. Ozsahin. The integration of virtual reality (vr), augmented reality (ar), and artificial intelligence in revolutionizing healthcare: A systematic review. *Augmented Human Research*, 10(1):9, 2025. 3

[9] A. Hadadi, C. Guillet, J. Chardonnet, M. Langovoy, Y. Wang, and J. Ovtcharova. Prediction of cybersickness in virtual environments using topological data analysis and machine learning. *Frontiers in Virtual Reality*, 3:973236, 2022. doi: 10.3389/frvir.2022.973236 1

[10] A. Hag, M. R. C. Qazani, and H. Asadi. Quantifying motion sickness in virtual reality using a multimodal 1cnn–gru–attention approach with gsr data. *IEEE Transactions on Intelligent Transportation Systems*, 2025. 4

[11] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997. 4

[12] C. Hua, L. Chai, Y. Yan, J. Liu, Q. Wang, R. Fu, and Z. Zhou. Assessment of virtual reality motion sickness severity based on eeg via lstm/bilstm. *IEEE Sensors Journal*, 23(20):24839–24848, 2023. 4

[13] M. J. Islam and R. Islam. Towards optimized cybersickness prediction for computationally constrained standalone virtual reality devices. In *2024 IEEE conference on virtual reality and 3D user interfaces abstracts and workshops (VRW)*, pp. 1003–1004. IEEE, 2024. 1

[14] 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. doi: 10.1109/ISMAR52148.2021.00017 1, 2, 4

[15] 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. IEEE, 2022. doi: 10.1109/ISMAR55827.2022.00028 1, 2, 4, 6, 8

[16] 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. 4

[17] J. Josupeit. Cybersickness as the virtual reality sickness questionnaire (vrsq) measures it!?–an environment-specific revision of the vrsq. *Frontiers in Virtual Reality*, 4:1291078, 2023. 1

[18] J. J. L. Jr. A discussion of cybersickness in virtual environments. *ACM SIGCHI Bulletin*, 32(1):47–56, 2000. 2

[19] 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. doi: 10.1207/s15327108ijap0303 2

[20] B. Keshavarz and H. Hecht. Validating an efficient method to quantify motion sickness. *Human Factors*, 53(4):415–426, 2011. doi: 10.1177/0018720811403736 2

[21] H. K. Kim, J. Park, Y. Choi, and M. Choe. Virtual reality sickness questionnaire (vrsq): Motion sickness measurement index in a virtual reality environment. *Applied Ergonomics*, 69:66–73, 2018. doi: 10.1016/j.apergo.2017.12.016 2

[22] J. Kim, J. Kim, T.-H. Kim, 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 (ICCV)*, 2019. 2

[23] 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. 2

[24] J. Kritikos, G. Alevizopoulos, and D. Koutsouris. Personalized virtual reality human-computer interaction for psychiatric and neurological illnesses: a dynamically adaptive virtual reality environment that changes according to real-time feedback from electrophysiological signal responses. *Frontiers in Human Neuroscience*, 15:596980, 2021. 1

[25] R. K. Kundu, I. Ahmed, and K. A. Hoque. Enhancing immersive virtual reality experiences with multiple tasks prediction using pre-trained large foundation models. In *Proceedings of the 2025 31st ACM Symposium on Virtual Reality Software and Technology*, pp. 1–11, 2025. 4

[26] 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 (ISMAR)*, pp. 777–786, 2022. doi: 10.1109/ISMAR55827.2022.00096 1, 2

[27] 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, 2023. doi: 10.1109/VR55154.2023.00076 1, 2

[28] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, and G. D. Hager. Temporal convolutional networks for action segmentation and detection. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 156–165, 2017. 4

[29] C. Lea, R. Vidal, A. Reiter, and G. D. Hager. Temporal convolutional networks: A unified approach to action segmentation. In *European conference on computer vision*, pp. 47–54. Springer, 2016. 4

[30] G. Li, H.-R. Byeon, S.-B. Cho, and F. Pollick. Vr cybersickness classification using machine learning models on open-access eeg datasets from the human vestibular network. In *2025 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR)*, pp. 171–175. IEEE, 2025. 3

[31] R. Li, Y. Wang, H. Yin, J.-R. Chardonnet, and P. Hui. A deep cybersickness predictor through kinematic data with encoded physiological representation. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 1132–1141. IEEE, 2023. 4

[32] T. Li, Y. Zhu, H.-N. Liang, and Y. Wang. Adapting to the user: A systematic review of personalized interaction in vr. *arXiv preprint arXiv:2510.13123*, 2025. 1

[33] Y. Li, I. Shiyanov, and B. Muschalla. Assessing cybersickness in a virtual reality intervention among older adults living in nursing homes: a feasibility study. *Disability and Rehabilitation: Assistive Technology*, pp. 1–9, 2025. 1

[34] S. Litleskare. The relationship between postural stability and cybersickness: it's complicated–an experimental trial assessing practical implications of cybersickness etiology. *Physiology & Behavior*, 236:113422, 2021. doi: 10.1016/j.physbeh.2021.113422 2

[35] J. Lohman and L. Turchet. Evaluating cybersickness of walking on an omnidirectional treadmill in virtual reality. *IEEE Transactions on Human-Machine Systems*, 52(4):613–623, 2022. 1

[36] T. Luong, N. Martin, A. Raison, F. Argelaguet, J.-M. Diverrez, and A. Lécuyer. Towards real-time recognition of users mental workload using integrated physiological sensors into a vr hmd. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 425–437, 2020. doi: 10.1109/ISMAR50242.2020.00068 2

[37] 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*, CHI '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411764.3445701 1, 2

[38] B. G. Marcot. Metrics for evaluating performance and uncertainty of bayesian network models. *Ecological modelling*, 230:50–62, 2012. 6

[39] M. E. McCauley and T. J. Sharkey. Cybersickness: Perception of self-motion in virtual environments. *Presence: Teleoperators and Virtual Environments*, 1(3):311–318, 08 1992. doi: 10.1162/pres.1992.1.3.311 1, 2

[40] M. Melo, J. Vasconcelos-Raposo, and M. Bessa. Presence and cybersickness in immersive content: Effects of content type, exposure time and gender. *Computers & Graphics*, 71:159–165, 2018. 2

[41] A. B. Mrad, V. Delcroix, S. Piechowiak, P. Leicester, and M. Abid. An explication of uncertain evidence in bayesian networks: Likelihood evidence and probabilistic evidence: Uncertain evidence in bayesian networks. *Applied Intelligence*, 43(4):802–824, 2015. 2

[42] H. Oh and W. Son. Cybersickness and its severity arising from virtual reality content: A comprehensive study. *Sensors*, 22(4):1314, 2022. 2

[43] S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009. 4

[44] K. Petri, K. Feuerstein, S. Folster, F. Bariszlovich, and K. Witte. Effects of age, gender, familiarity with the content, and exposure time on cybersickness in immersive head-mounted display based virtual reality. *American Journal of Biomedical Sciences*, 12(2), 2020. 2

[45] T. M. Porcino, E. Clua, D. Trevisan, C. N. Vasconcelos, and L. Valente. Minimizing cyber sickness in head mounted display systems: design guidelines and applications. In *2017 IEEE 5th international conference on serious games and applications for health (SeGAH)*, pp. 1–6. IEEE, 2017. 2

[46] C. Qi, D. Ding, H. Chen, Z. Cao, and W. Zhang. Cpnet: Real-time cybersickness prediction without physiological sensors for cybersickness mitigation. *ACM Transactions on Sensor Networks*, 2025. 1

[47] C. Qu, X. Che, S. Ma, and S. Zhu. Bio-physiological-signals-based vr cybersickness detection. *CCF Transactions on Pervasive Computing and Interaction*, 4(3):268–284, 2022. 2

[48] 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. 3

[49] 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. 2

[50] 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 1, 2, 4, 6, 8

[51] S. Shimada, P. Pannattee, Y. Ikei, N. Nishiuchi, and V. Yem. High-frequency cybersickness prediction using deep learning techniques with eye-related indices. *IEEE Access*, 11:95825–95839, 2023. 4

[52] J.-P. Stauffert, F. Niebling, and M. E. Latoschik. Latency and cybersickness: Impact, causes, and measures. a review. *Frontiers in Virtual Reality*, 1:582204, 2020. 3

[53] N. Subramani, R. Kazemi, J. Park, S. Kim, and S. C. Lee. Integrating deep learning and signal processing for cybersickness classification using electroencephalogram and exploratory factor analysis approach. *International Journal of Human–Computer Interaction*, pp. 1–26, 2025. 2

[54] 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 1, 2, 3, 4, 6

[55] 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. 2

[56] S. Valmorisco, L. Raya, and A. Sanchez. Enabling personalized vr experiences: a framework for real-time adaptation and recommendations in vr environments. *Virtual Reality*, 28(3):128, 2024. 1, 3

[57] S. Vlahovic, L. Skorin-Kapov, M. Suznjevic, and N. Pavlin-Bernardic. Not just cybersickness: Short-term effects of popular vr game mechanics on physical discomfort and reaction time. *Virtual reality*, 28(2):108, 2024. 2

[58] H. Wang, Z. Li, C. Hua, J. Zhu, B. Zhang, Z. Dai, D. Chen, and J. Liu. Stbb-net: A hybrid eeg model for virtual reality motion sickness detection. *IEEE Sensors Journal*, 2026. 3

[59] P. Wang, M. Li, Z. Wang, Y.-J. Liu, and L. Wang. Multimodal contrastive learning for cybersickness recognition using brain connectivity graph representation. *IEEE Transactions on Visualization and Computer Graphics*, 2025. 3

[60] Y. Wang, J.-R. Chardonnet, F. Merienne, and J. Ovtcharova. Using fuzzy logic to involve individual differences for predicting cybersickness during vr navigation. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 373–381. IEEE, 2021. 8

[61] S. Weech, S. Kenny, M. Lenizky, and M. Barnett-Cowan. Narrative and gaming experience interact to affect presence and cybersickness in virtual reality. *International Journal of Human-Computer Studies*, 138:102398, 2020. 2

[62] K. Weiss, T. M. Khoshgoftaar, and D. Wang. A survey of transfer learning. *Journal of Big data*, 3(1):9, 2016. 4

[63] P. Wu, N. Ahmed, K. Huang, T. Lan, and M. Imani. Personalized bayesian networks for cybersickness prediction in virtual reality. In *Proceedings of the 1st Workshop on Enhancing Security, Privacy, and Trust in Extended Reality (XR) Systems (XR Security 2025)*. Houston, TX, USA, Sept. 2025. 1, 3

[64] 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, 07 2025. doi: 10.13140/RG.2.2.34967.48807/1 1, 2, 3

[65] A. H. X. Yang, N. Kasabov, and Y. O. Cakmak. Machine learning methods for the study of cybersickness: a systematic review. *Brain Informatics*, 9(1):24, 2022. 3, 4

[66] Y. Zhang, H. Zhou, Z. Jiang, Z. Tang, T. Luo, and Q. Lei. Exploring viewing modalities in cinematic virtual reality: A systematic review and meta-analysis of challenges in evaluating user experience. *Proceedings of the ACM on Human-Computer Interaction*, 9(2):1–30, 2025. 3