Personalized Bayesian Networks for Cybersickness Prediction in Virtual Reality

Peng Wu
wu.p@@northeastern.edu
Northeastern University
Boston, MA, USA

Nasim Ahmed
nahmed25@students.kennesaw.edu
Kennesaw State University
Kennesaw, GA, USA

Kaiming Huang
kzh529@psu.edu
Pennsylvania State University
University Park, PA, USA

Rifatul Islam
rislam11@kennesaw.edu
Kennesaw State University
Kennesaw, GA, USA

Tian Lan
tlan@gwu.edu
George Washington University
Washington, DC, USA

Gang Tan
gtan@psu.edu
Pennsylvania State University
University Park, PA, USA

Mahdi Imani
m.imani@@northeastern.edu
Northeastern University
Boston, MA, USA

Abstract

Personal characteristics fundamentally shape virtual reality (VR) experiences, yet their integration into predictive models remains underexplored. This paper studies how to incorporate personal attributes (age, gender, prior VR experience) into Bayesian networks for cybersickness prediction via: (i) direct inclusion as root nodes, (ii) a two-stage model that learns a susceptibility score from personal attributes, and (iii) a stratified model. Using 26,040 samples from VR maze-navigation experiments, direct inclusion attains 82.53% accuracy (+14.02 percentage points over a 68.51% no-personal baseline). The two-stage approach reaches 77.32% while supporting cold-start prediction for unseen users, and stratified models achieve 73.62%. Using participant-level cross-validation to avoid subject leakage, we find that personalization consistently improves cybersickness prediction. These results argue that personal attributes should be treated as first-class signals in cybersickness models, with clear design trade-offs between maximal accuracy and deployability for unseen users, informing personalized VR systems and adaptive content delivery.

CCS Concepts

• Human-centered computing → Virtual reality; **• Computing methodologies → Bayesian network models**.

Keywords

Virtual Reality, Cybersickness, Bayesian Networks, Personal Information, Predictive Modeling

MobiHoc '25, Houston, TX, USA

ACM ISBN 979-8-4007-1353-8/2025/10.
https://doi.org/10.1145/3704413.3765310

ACM Reference Format:
Peng Wu, Nasim Ahmed, Kaiming Huang, Rifatul Islam, Tian Lan, Gang Tan, and Mahdi Imani. 2025. Personalized Bayesian Networks for Cybersickness Prediction in Virtual Reality . In . ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Virtual Reality (VR) adoption continues to accelerate across domains from entertainment to healthcare, yet cybersickness remains a persistent barrier affecting 20–80% of users [1]. While prior work has largely emphasized content/kinematics and system parameters, as well as physiological responses [9, 10], the critical role of personal attributes in shaping individual susceptibility to cybersickness has received insufficient attention. Users' age, gender, and prior VR experience introduce substantial variability in susceptibility that generic, population-level models fail to capture.

Cybersickness is inherently personal: the same visual scene can elicit markedly different responses across users depending on prior exposure, sensory integration, and cognitive strategy [8]. Treating all users uniformly forces a one-size-fits-all model to explain heterogeneous mechanisms, often yielding conservative predictions that underperform for both highly susceptible and highly resilient users. Incorporating personal attributes offers a principled path toward individualized inference.

Bayesian Networks (BNs) have been successfully applied to predict cybersickness in the VR domain [2], offering probabilistic rigor and interpretability. However, most BN-based approaches to date treat users uniformly and overlook how individual differences modulate cybersickness. In [2], for instance, a BN framework modeled how system parameters (e.g., luminance, optical flow) and physiological responses (e.g., pupil dilation, galvanic skin response) jointly shape Fast Motion Sickness (FMS) scores, but it did not leverage static personal attributes that may fundamentally alter how users perceive and respond to VR stimuli.

In this paper, we extend that framework by treating personal attributes as first-class variables within the BN structure. Our goal is not only to assess whether such attributes improve predictive performance, but to systematically investigate *how* they should be incorporated to balance accuracy, model compactness, and deployability. We hypothesize that personal attributes are not mere demographic labels, but critical moderators of the relationships between system/physiological factors and subjective outcomes, enabling more accurate predictions and genuinely personalized VR experiences.

To explore this, we introduce and compare three personalization strategies: (1) direct integration of personal attributes as BN root nodes, (2) a two-stage approach where attributes are mapped to a latent "susceptibility" score before integration (supporting cold-start), and (3) stratified models trained separately for user subgroups. We analyze trade-offs in terms of original (exact) and ±1 "extended" accuracy, and discuss BN compactness (parent limits/CPT size) and deployability (including cold-start for unseen users). We evaluate these strategies on a VR maze-navigation dataset comprising system parameters, physiological measures, and personal attributes from 32 participants.

We summarize our contributions as follows:

- **Modeling strategies for personalization**: We formalize and compare three BN-based strategies for incorporating personal features, from direct integration to stratified modeling.
- **Design rationale and practicality**: We motivate a two-stage approach that maps personal profiles to a compact susceptibility score, enabling cold-start prediction and controlling BN complexity.
- **Empirical evidence**: Using 26,040 samples from a real-walking VR dataset with 32 participants, we show that personal information yields substantial and consistent gains—direct integration improves accuracy by 14.02% over a strong baseline.

2 Related Work

2.1 Cybersickness and Individual Differences

Cybersickness manifests through symptoms including nausea, disorientation, and visual fatigue, primarily attributed to sensory conflict between visual and vestibular inputs [3]. Research has identified significant individual variability in susceptibility, with factors including age (older adults showing higher sensitivity), gender (conflicting findings on differences), and prior VR experience (habituation effects) [4, 5]. Beyond these factors, users differ in motion sensitivity, visual dependence, and cognitive load tolerance, all of which shape adaptation and recovery processes during VR exposure. The literature increasingly points to personalization as a necessary condition for robust cybersickness mitigation in real-world deployments.

2.2 Bayesian Networks in VR Research

Bayesian Networks provide interpretable probabilistic models ideal for capturing causal relationships in complex systems [6]. Prior research has demonstrated the effectiveness of BNs for verifying cybersickness under uncertainty [2]. However, existing BN applications in VR have not systematically explored personal feature integration strategies, limiting their practical applicability for personalized systems. Compared with black-box deep models [5], BNs offer transparency, modularity, and a natural way to inject domain constraints (e.g., system features as roots). Personalization within BNs raises new design questions: whether to include personal variables directly, compress them into latent representations, or partition the population and train specialized structures. We address these questions empirically and highlight trade-offs among accuracy, model size, and deployment practicality.

3 Methodology

3.1 Bayesian Network Framework

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:

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

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. System features act as exogenous drivers, personal features as user-specific moderators, physiological features as mediators, and the FMS score as the final outcome. Structure learning uses a hill-climbing search with BDeu scoring (ESS = 10) under constraints that prevent edges into system variables 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 features** $S = \{s_1, ..., s_m\}$: Root nodes (no incoming edges)
- **Personal features** $P = \{p_1, p_2, p_3\}$: Age, Gender, VRExperience
- **Physiological features** $H = \{h_1, ..., h_k\}$: 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 Personal Information Integration Strategies

We propose and evaluate three strategies for incorporating personal features into the BN structure:

3.2.1 Strategy 1 — Direct Integration. Personal features are added as root nodes alongside system features, allowing the structure learning algorithm to discover their direct and indirect influences on physiological and outcome variables:

$$P(S, P, H, Y) = \prod_{s \in S} P(s) \prod_{p \in P} P(p) \prod_{h \in H} P(h|Pa(h)) \cdot P(Y|Pa(Y)) \tag{2}$$

The parent sets $Pa(h)$ may include both system and personal features, enabling explicit modeling of pathways such as Age → GSR

→ FMS. This approach preserves the full predictive signal of personal variables without compression and allows the BN to uncover interaction effects between personal and system-level drivers.

We declare personal and system variables as roots, apply the same hierarchical constraints as in the baseline (no edges into system variables, no edges out of the outcome), cap the maximum number of parents per node to control CPT size, and use BDeu scoring to favor sparse graphs. After learning, variables unconnected to FMS are pruned to keep the model compact. This strategy is most appropriate when the number of personal attributes is modest and reliable, as it maximizes interpretability and accuracy, but may be less suitable if the personal feature space grows large due to potential CPT explosion.

3.2.2 Strategy 2 — Two-Stage Model. A susceptibility score Z is learned from personal features using a separate model f_θ, then incorporated into the BN:

$$P(S, Z, H, Y) = \prod_{s \in S} P(s) \cdot P(Z) \prod_{h \in H} P(h|Pa(h)) \cdot P(Y|Pa(Y)), \quad (3)$$

where

$$Z = f_\theta(P) \quad \text{where } f_\theta : \mathbb{R}^3 \to \mathbb{R}. \quad (4)$$

We use Random Forest regression for f_θ, capturing non-linear relationships between personal features and susceptibility. This design addresses three practical considerations. First, when the number of personal attributes grows, adding them all as parents in the BN can dramatically increase conditional probability table (CPT) size and lead to data sparsity. For a node with r states and d discrete parents with cardinalities $k_1, \ldots, k_d$, the CPT has $r \prod_{j=1}^{d} k_j$ parameters, which quickly becomes unwieldy. Second, a compact scalar Z acts as a bottleneck representation that preserves predictive signal while keeping the BN small and interpretable. Third, the learned mapping f_θ enables *cold-start* prediction for new users using demographics alone, without requiring user-specific BN retraining. In short, the two-stage approach balances accuracy and parsimony, avoiding over-parameterization and maintaining a deployable, modular architecture.

3.2.3 Strategy 3 — Stratified Modeling. Separate BN models are trained for different user groups based on a stratification feature $k \in P$:

$$P(S, H, Y|k = c) = \prod_{s \in S} P(s|k = c) \quad (5)$$

$$\prod_{h \in H} P(h|Pa(h), k = c) \cdot P(Y|Pa(Y), k = c) \quad (6)$$

Final predictions use sample-weighted accuracy across strata:

$$\text{Accuracy} = \frac{\sum_c n_c \cdot \text{Acc}_c}{\sum_c n_c} \quad (7)$$

where n_c is the sample size and Acc_c is the accuracy for stratum c.

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/Middle/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 by a user's group membership and aggregate performance with sample-weighted accuracy across groups.

4 Experimental Setup

4.1 Dataset

We use the VR-Walking dataset [7], which contains 26,040 samples collected from 32 participants navigating virtual mazes. Each sample is described by four feature categories:

- **System features** (10): optical flow, HOG features, spectral entropy, luminance, contrast, etc.
- **Physiological features** (10): pupil diameter, GSR, reaction time, gaze error, etc.
- **Personal features** (3): Age (18-40), Gender (0/1/2), VRExperience (0/1/2)
- **Target**: FMS scores (1-7) indicating cybersickness severity

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.
- **Structure Learning**: We use hill-climbing search with BDeu scoring under hierarchical constraints that (i) forbid edges from personal or physiological variables into system variables, and (ii) forbid outgoing edges from the FMS node. A maximum parent limit per node controls 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 prior to 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.
- **Validation protocol**: Performance is measured with 5-fold participant-level cross-validation to ensure no subject overlap between training and test sets. Hyperparameters are tuned using only the training folds in each split.
- **Stratification thresholds (Strategy 3)**:
  - Age: Young (≤22), Middle (23–29), Older (≥30)
  - Gender: Categorical groups based on original encoding
  - VRExperience: Ordinal levels 0, 1, 2

4.3 Extended Label Accuracy

Given the subjective and ordinal nature of FMS ratings, we also evaluate models using an *extended label accuracy* metric. This measure counts a prediction as correct if it falls within ±1 level of the ground truth on the 1–7 FMS scale:

$$\text{ExtAcc} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}\big[\, |y_i - \hat{y}_i| \le 1 \,\big] \tag{8}$$

A tolerance margin is important in VR contexts because small deviations on a discomfort scale rarely change practical outcomes, such as whether a user needs a rest break or a mitigation action. For example, predicting a score of 4 when the true value is 5 is functionally equivalent for most intervention strategies. By reporting both exact accuracy and extended accuracy, we capture complementary perspectives: the former reflects strict ordinal matching, while the latter reflects practical usability in real-world deployments.

5 Results and Discussion

5.1 Strategy Performance Comparison

Table 1 summarizes the performance of all personalization strategies relative to the baseline model without personal features. Exact accuracy shows the percentage of perfect ordinal matches, while extended accuracy reflects predictions within one FMS point of the truth.

5.2 Analysis of Personal Feature Impact

Strategy 1, which directly integrates personal attributes into the BN, produced the largest performance gain, improving exact accuracy by over 14 percentage points compared to the baseline and achieving the highest extended accuracy as well. Examination of the learned structure reveals multiple influence pathways:

- Age connects both directly to the FMS node and indirectly via GSR, suggesting age-related differences in physiological stress reactivity;
- VRExperience impacts reaction time and gaze behavior, potentially reflecting perceptual-motor adaptation;
- Gender is linked to pupil dynamics and other arousal markers, pointing to subtle but measurable differences in visual-physiological responses.

Strategy 2 condenses the three personal features into a latent susceptibility score before integration. This compact representation retains much of the predictive benefit (exact accuracy of 77.32%, extended accuracy of 90.01%) while keeping the BN small and supporting *cold-start* scenarios where only demographics are available at inference time. Such a design is particularly attractive for deployments where collecting physiological signals upfront is impractical.

Strategy 3 applies stratified modeling, learning separate BNs for each subgroup of a chosen personal attribute. Among the tested groupings, age-based stratification delivered the best results (exact accuracy 73.62%, extended accuracy 85.90%), outperforming gender- and VRExperience-based splits. This suggests that age captures more meaningful variation in cybersickness susceptibility than the other attributes in this dataset, although the reduced sample size per group limits absolute accuracy.

5.3 Network Structure Insights

Figure 1 visualizes the learned BN for Strategy 1, where personal attributes open additional pathways to cybersickness beyond purely stimulus–response links.

Across folds, we consistently observe motifs in which *Age* connects both directly to FMS and indirectly via physiological intermediates (e.g., GSR), *VRExperience* links to reaction time and gaze behavior, and *Gender* associates with pupil dynamics. These patterns align with the intuition that user attributes modulate how system stimuli are integrated sensorimotorically before influencing subjective discomfort.

We emphasize that these edges encode conditional dependencies under our modeling assumptions and constraints; they are not, on their own, evidence of causal effects. Edge orientations that were unstable across folds are treated conservatively in our interpretation (i.e., reported as associations or pathways rather than asserted causal directions). To focus the analysis on prediction-relevant structure, we prune nodes not ancestral or descendant to FMS, yielding compact graphs with bounded in-degree and manageable CPTs. This combination of constraints, pruning, and post-hoc inspection explains why adding personal attributes can improve accuracy: they introduce additional, plausible routes through which individual differences shape both physiological responses and reported cybersickness.

Deployment considerations. The three strategies present distinct operational trade-offs. **Strategy 1 (Direct)** maximizes accuracy and interpretability when a small, trusted set of personal attributes is available; graph inspection can surface subgroup pathways useful for debugging and design. **Strategy 2 (Two-stage)** compresses attributes into a single susceptibility score, reducing CPT growth and enabling *cold-start* predictions from demographics alone—useful when physiological data are unavailable at onboarding or when privacy policies limit per-user feature storage. **Strategy 3 (Stratified)** fits separate BNs per subgroup, which can reveal segment-specific structures and mitigate averaging over heterogeneous mechanisms; its main cost is fewer samples per model and thus potentially higher variance. In practice, Strategy 2 is attractive for privacy-aware or resource-constrained deployments, Strategy 1 for controlled or enterprise settings with richer profiles, and Strategy 3 when segments are well-balanced and clearly distinct.

5.4 Implications for VR System Design

Our results suggest several design directives for VR systems:

(1) **Personalization is not optional**: omitting personal attributes leads to substantial performance loss (14 percentage points in exact accuracy relative to Strategy 1).
(2) **Match the personalization strategy to the context**: two-stage modeling enables immediate, low-friction adaptation for new users and can be refined later as richer data accrue; full integration is preferable when maximal accuracy and path-level interpretability are required; stratification helps when populations are clearly segmented and you can afford per-segment tuning.
(3) **Age-aware design** appears especially promising in our dataset: age-informed adjustments to locomotion speed, visual flow, or contrast could reduce discomfort for susceptible groups without penalizing resilient users.

Table 1: Cybersickness prediction across personalization strategies on the VR maze-navigation dataset (32 participants). Values are 5-fold *participant-level* CV means.

Strategy	Exact Accuracy (%)	Extended (±1) Accuracy (%)	Δ vs. Baseline (pp)
Baseline (No Personal)	68.51	80.72	–
Strategy 1: Direct	**82.53**	**92.85**	**+14.02**
Strategy 2: Two-Stage	77.32	90.01	+8.81
Strategy 3: Age-based	73.62	85.90	+5.11
Strategy 3: Gender-based	71.82	84.62	+3.31
Strategy 3: VRExp-based	68.75	80.73	+0.24

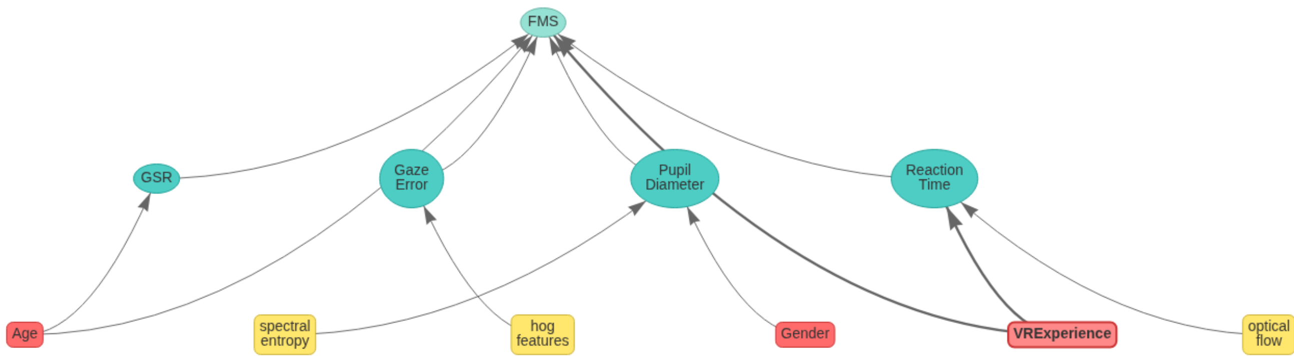

Figure 1: Consensus BN with personal attributes (Strategy 1). Edges shown appear in ≥3/5 folds; orientations for Markov-equivalent edges are illustrative.

(4) **Toward real-time adaptation**: since personal attributes modulate both physiological signals and FMS directly, systems can combine static profiles with live indicators (e.g., elevated GSR or atypical gaze error) to trigger mitigations such as temporary FoV reduction, luminance tweaks, or timed rest prompts before discomfort escalates.

6 Conclusion

This paper provides strong evidence that incorporating personal attributes into Bayesian networks materially improves cybersickness prediction in VR. Direct integration achieves 82.53% exact accuracy, substantially outperforming a user-agnostic baseline, while a two-stage susceptibility representation offers a practical compromise that supports cold-start deployment with competitive performance. Rather than pursuing universal, population-averaged models, we argue for a *personalization-first* approach in which individual characteristics shape both model structure and predictions. As VR adoption broadens, this shift is essential to deliver safe, comfortable, and engaging experiences across diverse users. Our study employed a single VR maze-navigation dataset with 32 participants and ordinal FMS labels; although we employ participant-level cross-validation and train-time discretization to avoid leakage, broader validation across tasks and devices is warranted.

Future directions include expanding the attribute set (e.g., gaming history, motion sensitivity), evaluating ordinal-aware metrics beyond accuracy (e.g., quadratic-weighted kappa, Brier score/calibration), exploring online adaptation policies that jointly use static profiles and streaming physiology, and assessing fairness and privacy impacts when modeling sensitive attributes. These steps will help translate personalization-first modeling into robust, ethical, and widely deployable VR comfort systems.

Acknowledgments

This work was supported in part by the Defense Advanced Research Projects Agency (DARPA) under grant HR00112420366, and the National Science Foundation (NSF) under award 2311969.

References

[1] N. Tian, P. Lopes, and R. Boulic. A review of cybersickness in head-mounted displays: raising attention to individual susceptibility. *Virtual Reality*, 26:1409–1441, 2022.
[2] 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 *Proc. IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2025.
[3] J. J. LaViola Jr. A discussion of cybersickness in virtual environments. *ACM SIGCHI Bulletin*, 32(1):47–56, 2000.
[4] 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.
[5] 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 *Proc. IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 31–40, 2021.
[6] I. Tsamardinos, L. E. Brown, and C. F. Aliferis. The max-min hill-climbing Bayesian network structure learning algorithm. *Machine Learning*, 65:31–78, 2006.
[7] 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. 2024.

[8] 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.

[9] 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)*, pages 121–130. IEEE, 2022. doi:10.1109/ISMAR55827.2022.00026.

[10] 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)*, pages 387–399. IEEE, 2020.