Probabilistic Verification of Cybersickness in Virtual Reality Through Bayesian Networks

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

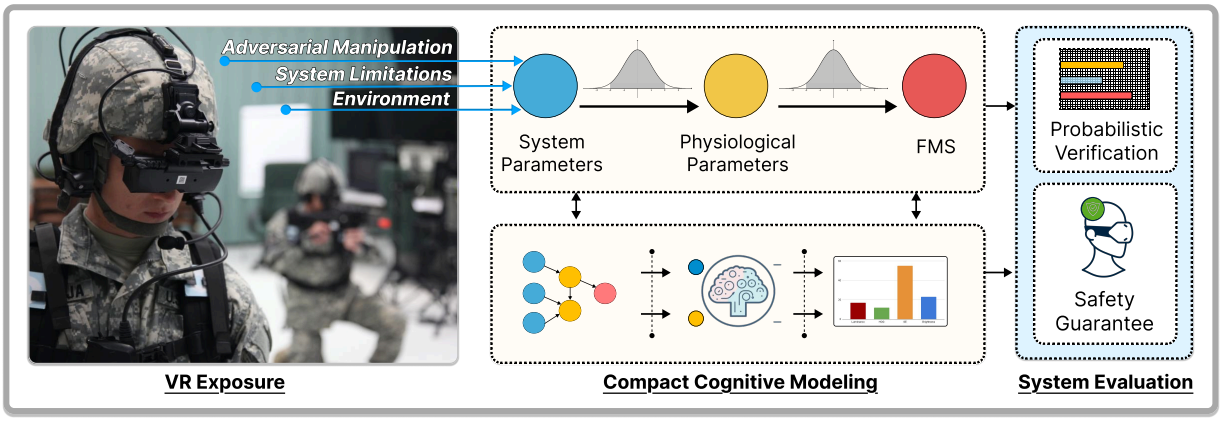

Figure 1: Overview of the proposed verification framework for cybersickness in VR. Bayesian Network, structured hierarchically, models how adversarial, system, and environmental factors influence system parameters, which in turn affect physiological responses and ultimately determine Fast Motion Sickness (FMS). The model enables probabilistic verification of safety constraints under observed or hypothetical conditions.

Abstract

Cybersickness remains a major challenge in virtual and mixed reality (VR/MR), yet existing methods primarily focus on predicting its onset without offering formal guarantees regarding its occurrence or effective mitigation. As VR/MR applications expand into safety-critical domains like healthcare, defense, verifiable safety assurances become essential to protect users from adverse physiological and psychological effects. This paper introduces a probabilistic verification framework leveraging Bayesian Networks (BN) to explicitly model the interactions among system parameters, human physiological responses, and cybersickness severity. Unlike deep learning approaches that lack interpretability and formal verification capabilities, the proposed BN model explicitly captures how environmental and system-level factors (e.g., luminance, spectral entropy, and image gradient complexity via HoG features) influence physiological responses (e.g., heart rate, reaction time, eye tracking), ultimately affecting cybersickness severity. By learning the joint probability distribution of these factors, our approach provides rigorous formal guarantees on cybersickness risk under specified operational conditions. If these guarantees are not met, automated adaptive adjustments are recommended to restore safe conditions. Experimental validation involving physiological and system-level data demonstrates that Bayesian Networks provide an interpretable and efficient framework, uniquely enabling formal probabilistic verification of cybersickness risks. This capability makes the proposed approach particularly suitable for designing and deploying VR/MR systems with explicitly verified safety constraints.

Index Terms: Virtual Reality, Cybersickness, Bayesian Network, Probabilistic Verification

1 Introduction

Virtual Reality (VR) technologies have gained significant traction in fields ranging from education and entertainment to defense and healthcare [12, 40]. With the rise of immersive applications, ensuring user comfort and engagement is a crucial prerequisite. However, a widespread challenge that undermines VR's user experience is cybersickness (CS) [37, 21, 36, 34] , which manifests as a collection of discomforts, such as dizziness, nausea, and headaches. Surveys indicate that between 20% and 80% of VR users report experiencing cybersickness [17], underscoring the pressing need for approaches that can detect and, importantly, mitigate cybersickness.

Most existing efforts to address cybersickness revolve around predicting its onset using physiological signals (e.g., heart rate, galvanic skin response), along with VR system parameters like frame rate, field of view (FoV), or refresh rate [26, 43, 30, 39, 2, 9]. These data-driven techniques—often leveraging deep learning—offer insights into cybersickness triggers, but they typically require substantial training data and tend to function as "black-box" models that provide little interpretability regarding how these various fac-

[*]e-mail: wu.p@northeastern.com
[†]e-mail: nahmed25@students.kennesaw.edu
[‡]e-mail: abv5402@psu.edu
[§]e-mail: kzh529@psu.edu
[¶]e-mail: rislam11@kennesaw.edu
[‖]e-mail: binli@psu.edu
[**]e-mail: tlan@gwu.edu
[††]e-mail: gtan@psu.edu
[‡‡]e-mail: m.imani@northeastern.edu

tors interact. Moreover, they do not offer guarantees that a user will remain free from cybersickness or that a mitigation strategy will reliably protect against it.

This paper investigates the joint influence of system conditions and physiological responses on cybersickness, constructing a model that captures user experience across varying VR conditions and enables formal safety guarantees. Table 1 categorizes three key contributors to perceptual disruption and cybersickness in VR environments: *(i) adversarial manipulations*, where adversaries manipulate system parameters (e.g., brightness, latency, or field of view) to exploit perceptual vulnerabilities, inducing disorientation. *(ii) System limitations*, including tracking drift, sensor noise, rendering instabilities, and network latency, create inconsistencies between virtual stimuli and human perception, amplifying discomfort. *(iii) Environmental factors*, such as overcrowding, rapid motion, abrupt contrast gradients, and depth perception conflicts, which overload sensory processing and exacerbate cybersickness. While some applications may tolerate these risks, the growing use of MR/VR in safety-critical domains necessitates formal guarantees on user safety. As VR/MR are increasingly used in safety-critical domains such as surgery, aviation, and defense, formally verifying cybersickness-related safety constraints becomes imperative.

This paper introduces an offline probabilistic verification framework , trained on real-world dataset, that provides both formal guarantees and mitigation strategies for cybersickness. The approach models the joint impact of system parameters, human physiological responses, and cybersickness severity using a Bayesian Network (BN) model. The schematic diagram of the BN's hierarchical structure is shown in the middle of Figure 2, where system parameters act as external nodes influenced by three key contributing factors: adversarial manipulations, system limitations, and environmental conditions. These variations propagate to human physiological responses, such as eye tracking, heart rate, galvanic skin response (GSR), and pupil dilation, ultimately influencing cybersickness severity. Unlike purely statistical or deep learning models, a Bayesian Network offers an interpretable and structured representation of how system and physiological variables influence cybersickness, encoding the probabilistic interdependencies among them.

The Bayesian Network framework enables formal guarantees on cybersickness risk under varying conditions by reasoning over probabilistic dependencies among system and human factors. We develop a probabilistic verification framework based on Bayesian inference to evaluate whether a given configuration of system and physiological parameters satisfies a predefined cybersickness probability threshold. This allows designers to define safe operational conditions and obtain exact probabilistic assurances that these constraints will be met. If constraints are violated, the framework identifies parameter adjustments that can restore compliance. This verification capability lays the foundation for data-driven, safety-aware design workflows in immersive system development.

The VR-Walking dataset[32], which includes system-level features and physiological responses from 36 participants engaged in a structured VR walking task, is used to develop and verify the probabilistic cybersickness model. Key features include luminance, spectral entropy, and motion smoothness, along with eye tracking, heart rate, and galvanic skin response. The dataset also contains fast motion sickness (FMS) scores[16], a validated proxy for cybersickness severity. A Bayesian Network with a hierarchical structure is constructed to explicitly capture causal relationships among these variables influencing FMS. The model supports the formal verification of safety conditions and identifies parameter settings that can reduce the likelihood of mild to severe cybersickness symptoms.

[1]While factors such as latency or rendering instability can occur due to either benign technical limitations or deliberate adversarial actions, this categorization emphasizes their origin for clarity in safety analysis.

2 Related Works

In this section, we provide a comprehensive overview of prior research on cybersickness, with a focus on diverse modeling approaches. We examine both predictive and verification-oriented methods, outlining their respective contributions to understanding cybersickness in VR/MR environments. This survey also identifies critical gaps in the current literature, which motivate the need for more robust and generalizable models.

Cybersickness Models and Contributing Factors: Cybersickness symptoms closely mirror those of motion sickness and often persist after VR exposure [35, 34, 27]. Among various explanations, sensory conflict theory remains predominant, attributing cybersickness to discrepancies between visual stimuli and vestibular inputs [22, 7]. Researchers have identified key VR system parameters (e.g., display latency, mismatched frame rates, large fields of view, excessive brightness, and high scene complexity) as primary contributors to sensory conflicts [6, 23]. Physiological signals, including heart rate and galvanic skin response, reliably track discomfort levels, serving as proximate indicators for cybersickness [17].

Deep Learning Approaches: Recent studies widely adopt deep learning methods to predict cybersickness from physiological and system-level data. Utilizing complex models like convolutional neural networks and LSTMs trained on eye-tracking, head-tracking, and biosignal datasets, these methods achieve robust prediction performance [41, 11]. Islam et al. focused on forecasting the onset of cybersickness by introducing a multimodal deep fusion network that combines physiological, head-tracking, and eye-tracking data, achieving early prediction of cybersickness severity[11]. Complementarily, Zhu et al. presented a lightweight model for real-time individual cybersickness prediction by creating a video-aware bio-signal representation through cross-modal fusion of video features and bio-signal data (head/eye tracking, physiological signals), achieving high accuracy even with VR video inputs alone, thus addressing computational efficiency for real-time applications[44].

While deep learning models offer high predictive accuracy, they often lack formal safety assurances, rely on large datasets, and suffer from limited interpretability due to their black-box nature. These limitations are particularly critical in VR environments, where cybersickness risks arise under partially observed or uncertain conditions [19]. Recent efforts, such as LiteVR [20], address interpretability using explainable-AI (XAI) techniques like SHAP with LSTM, GRU, and MLP models to enhance transparency and reduce computational overhead. However, a key research gap remains: current approaches focus on post-hoc explanation and do not integrate safety constraints or dynamic uncertainty into the model training process, limiting their effectiveness in real-world, evolving VR scenarios.

Statistical Models: Statistical approaches, including linear regression, logistic regression, and Markov chain models, explicitly identify significant factors influencing cybersickness severity, such as refresh rate and FoV [2, 33]. These models provide clear interpretability and highlight factor importance. However, they typically lack explicit causal modeling capabilities and, importantly, cannot rigorously verify whether cybersickness risks remain within acceptable thresholds under partial or uncertain observational conditions.

Bayesian Networks in Cybersickness Research: Bayesian Networks have emerged as a compelling approach, combining interpretability and probabilistic modeling of causal relationships among different factors. Widely applied in fields like medical diagnosis and reliability engineering [42], BNs incorporate expert knowledge and empirical data to capture complex interdependencies effectively. Prior BN studies on cybersickness have primarily modeled symptom likelihood under clearly defined conditions[33]; notably absent is their utilization for formal probabilistic verification of safety thresholds—an essential requirement in safety-

Table 1: Categorization of factors influencing system dynamics and contributing to cybersickness in MR/VR environments, motivating the need for probabilistic safety verification. *Adversarial manipulations* are defined as deliberate modifications designed to degrade user experience, contrasted with benign system or environmental limitations.[1]

Category	Factor	Impact on Cybersickness
Adversarial Manipulations	Brightness Manipulation	Deliberate, rapid brightness shifts cause visual discomfort, ocular fatigue, and spatial disorientation.
	Latency & Frame Drops	Intentionally induced lag or skipped frames disrupt motion synchronization, significantly exacerbating motion sickness symptoms.
	Field of View Distortion	Purposeful alterations to FoV parameters create perceptual confusion and exacerbate spatial disorientation.
	Scene Complexity Attacks	Intentionally overloading rendering pipelines increases frame pacing instability, causing perceptual confusion and visual discomfort.
System Limitations	Tracking Drift	Inaccurate or jittery tracking causes perceptual mismatches and increased susceptibility to motion sickness.
	Positional Calibration Errors	Misalignment between real-world and virtual depth cues disrupts spatial perception, causing discomfort.
	Rendering Instability	Dynamic fluctuations in frame rate lead to visual instability, perceptual confusion, and potential nausea.
	Network Latency	Delays in cloud-rendered environments produce asynchronous visual-physical feedback, inducing sensory mismatch and motion sickness.
Environmental Scene Factors	Overcrowding & Motion Density	High-density visuals and rapid simultaneous motions overload perceptual processing, causing fatigue and spatial disorientation.
	High Object Gradients	Abrupt texture changes and high contrast gradients result in excessive visual stress and perceptual discomfort.
	Rapid Scene Motion	Sudden, high-speed virtual movements disrupt vestibular balance, inducing dizziness and nausea.
	Depth Perception Conflicts	Incorrect scaling or mismatched depth cues distort spatial awareness, increasing nausea and disorientation risks.

sensitive VR/MR applications where partial data, adversarial manipulations, and uncertain operational conditions frequently arise.

Probabilistic Verification and Cognitive Attacks: Formal verification has been employed to guarantee correctness properties of software and hardware systems [28, 24, 13, 15], recently extending into probabilistic verification for stochastic systems [14]. Within VR, where outcomes are highly sensitive to human cognitive states and adversarial manipulations, probabilistic verification uniquely provides rigorous assurances of user safety under partially known or hypothetical conditions. While adversarial and cybersecurity literature extensively addresses attacks on system integrity [25], existing VR cybersickness research lacks formal verification frameworks to guarantee user comfort and safety against adversarially induced sensory manipulations and environmental uncertainties.

Our Contribution: The current literature predominantly offers predictive cybersickness models—deep learning methods optimize predictive accuracy at the cost of interpretability and formal assurances, while statistical models provide interpretability without causal clarity or verification capabilities. Bayesian Networks, despite their ability to model rich causal relationships and uncertainty, have not yet been integrated with formal probabilistic verification methods to ensure cybersickness risks remain acceptably low under uncertain, partially observable, or adversarially manipulated conditions. This paper addresses this critical gap explicitly. Instead of seeking superior predictive performance or interpretability alone, this paper proposes a comprehensive BN-based probabilistic verification framework that rigorously guarantees cybersickness safety thresholds, even under incomplete information or hypothetical operational scenarios. This distinct focus enables formal assurances of safety-critical constraints, fundamentally advancing cybersickness modeling beyond mere prediction towards robust, formally verified VR/MR system design.

3 METHODOLOGY

In this section, we present our methodology for modeling cybersickness risk. Section 3.1 outlines the Bayesian Network construction and learning process, while Section 3.2 details the formal probabilistic verification of cybersickness safety thresholds.

3.1 Bayesian Networks for Cybersickness

Bayesian Networks provide a structured probabilistic framework for modeling uncertainty, making them particularly suitable for modeling causal relationships among system parameters (e.g., hog_features, luminance, temporal smoothness), human physiological responses (e.g., HR, GSR), and cybersickness severity in virtual reality environments, a detailed explanation is provided in Table 2. In this paper, cybersickness severity, denoted as X_q, is the primary variable of interest, while system parameters (S) and human physiological responses (H) serve as explanatory variables.

To explicitly capture causal relationships, we impose a hierarchical structure within the Bayesian Network. System parameters are treated as exogenous variables without incoming edges, initiating the causal relationships. Human physiological responses act as intermediate variables that reflect reactions to system parameters and potentially interact with each other. Finally, cybersickness severity are the outcome node, influenced solely by physiological responses. Formally, a Bayesian Network is represented by a directed acyclic graph (DAG) $G = (V, E)$, with nodes representing random variables and directed edges encoding conditional dependencies. In our BN:

- **System parameters** $S = \{S_1, S_2, \ldots, S_m\}$ include features such as latency, brightness, spectral entropy, and motion smoothness.
- **Human physiological responses** $H = \{H_1, H_2, \ldots, H_k\}$ encompass heart rate, galvanic skin response, eye tracking, and pupil dilation.
- **Cybersickness severity** X_q represents a probabilistic distribution over intensity levels (e.g., none to severe).

The joint probability distribution over these variables is factorized according to the conditional independence assumptions defined explicitly by the learned hierarchical structure:

$$P(S, H, X_q) = \prod_i P(S_i) \prod_j P(H_j \mid Pa(H_j)) P(X_q \mid Pa(X_q)), \quad (1)$$

where $Pa(H_j)$ and $Pa(X_q)$ indicate subsets of variables (parent nodes) directly influencing each physiological response and cybersickness severity, respectively. This structured factorization ensures computational efficiency and interpretability. The BN facilitates inference about cybersickness severity even under partial observability (e.g., missing physiological signals or unavailable system logs) or uncertain conditions. Given observed values for system and physiological parameters, the posterior over cybersickness severity is inferred as:

$$P(X_q \mid S = s, H = h) \propto P(X_q \mid Pa(X_q)) \prod_j P(H_j \mid Pa(H_j)) \prod_i P(S_i).$$

Here, the terms $P(S_i)$ denote prior distributions of system parameters, which reflect their exogenous nature and assumed independence. Unlike traditional models trained solely to predict cybersickness severity from observed data, our Bayesian Network explicitly models the joint posterior distribution of all relevant variables, providing a rigorous foundation for formal reasoning and probabilistic safety guarantees, as detailed in the subsequent verification framework.

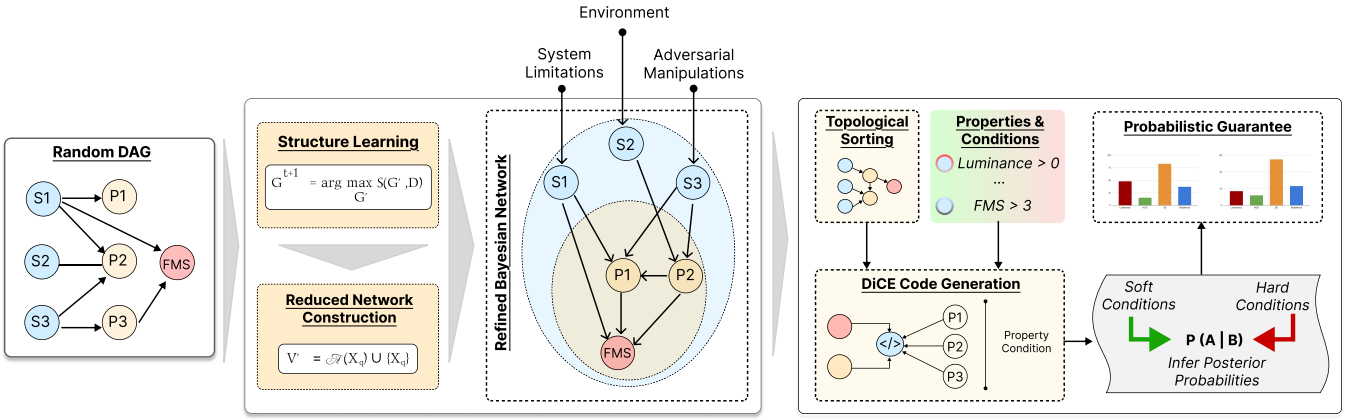

Figure 2: Overview of the hierarchical Bayesian Network modeling and probabilistic verification framework. The process involves constructing an optimized BN structure from system parameters and physiological responses, followed by posterior inference and formal verification of cybersickness safety conditions.

3.1.1 Constructing and Learning the BN Structure

Constructing the BN involves identifying causal dependencies among system parameters, physiological responses, and cybersickness severity. Structure learning seeks the optimal DAG representation, balancing data-driven insights and domain-specific causal assumptions. Specifically, the hierarchical constraints are enforced:

System Parameters $\rightarrow$ *Human Responses* $\rightarrow$ *Cybersickness Severity*

Given the hierarchical constraints, the BN structure is learned using the Hill Climbing (HC) algorithm[38], which iteratively refines the DAG by adding, removing, or reversing edges. The optimization seeks to maximize a scoring function $S(G,D)$, quantifying how well a network structure G explains observed data D. Formally:

$$G^{(t+1)} = \arg\max_{G'} S(G',D), \quad \text{subject to hierarchical constraints.}$$

The hierarchical constraint significantly reduces the search space, making structure learning both efficient and aligned explicitly with causal assumptions underlying cybersickness.

3.1.2 Node Reduction for Efficiency and Clarity

Following structure learning, node reduction is performed to remove variables that do not directly or indirectly affect cybersickness severity. Retaining only nodes in the ancestor set $\mathscr{A}(X_q)$ of X_q, the BN graph reduces to $G' = (V', E')$ with:

$$V' = \mathscr{A}(X_q) \cup \{X_q\}, \quad E' = \{(X_i, X_j) \in E \mid X_i, X_j \in V'\}.$$

This pruning step eliminates irrelevant nodes, ensuring computational efficiency, interpretability, and a clear focus on causal pathways that directly affect cybersickness severity.

Figure 3 provides an example Bayesian Network inference, showing the posterior probability distribution of cybersickness severity given observed conditions. Unlike deterministic predictions, the BN quantifies uncertainty explicitly, providing the probability that a user might experience each cybersickness severity level. Such probabilistic outputs are critical for formally verifying whether operational conditions satisfy predefined safety thresholds, as detailed in Section 3.2.

The left and middle boxes in Figure 2 summarize the hierarchical Bayesian Network modeling process, illustrating how adversarial manipulations, environmental factors, and system limitations influence system-level properties and physiological conditions, leading to variations in cybersickness severity, as measured by FMS. The structure learning refines an initially random DAG to an optimized, reduced network used explicitly to infer posterior probabilities, enabling formal verification of cybersickness safety conditions.

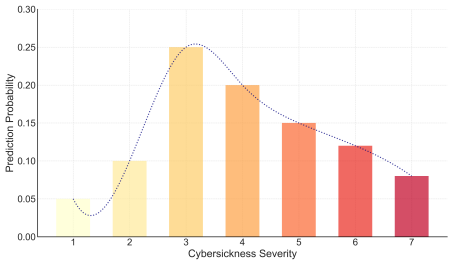

Figure 3: Example of posterior prediction using the learned Bayesian Network. Given observed system and physiological conditions, the model infers the probability distribution over cybersickness severity levels, enabling uncertainty-aware risk assessment for verification.

3.2 Probabilistic Verification

While Bayesian Networks effectively model causal and probabilistic dependencies among VR system parameters, physiological responses, and cybersickness severity, modeling alone does not inherently yield rigorous guarantees about system behavior under uncertainty, partial observations, or adversarial manipulations. Therefore, our verification framework evaluates whether specific configurations of system and human parameters satisfy critical safety constraints (e.g., keep severity below a defined threshold) with quantifiable certainty, which is essential for safety-critical VR applications.

Our verification process builds on the hierarchical Bayesian Network, where system parameters (S), human physiological responses (H), and cybersickness severity (X_q) were introduced. We define the full set of *random variables* as $\mathbf{X} = S \cup H \cup \{X_q\}$, where each variable $X_i \in \mathbf{X}$ is associated with a *conditional probability table* (CPT) of the form $P(X_i \mid \mathrm{Pa}(X_i))$, with $\mathrm{Pa}(X_i)$ denoting the *parent nodes* as defined by our hierarchical structure (i.e., system parameters influence physiological responses, which in turn influence cybersickness severity). The BN thereby defines a joint probability distribution that factorizes as shown in Equation 1.

To formalize verification, *Boolean properties* $\phi(\mathbf{X})$ are defined over the variable set $\mathbf{X}$, representing safety conditions of interest (e.g., FMS > 4 captures severe cybersickness). The objective is to compute the probability $P(\phi(\mathbf{X}) = 1 \mid C)$, where C denotes a set of *evidence conditions* reflecting observed or hypothetical constraints on system or physiological variables. This is achieved through Bayesian conditioning, where the joint distribution specified by the BN is updated in light of the evidence C. The resulting posterior distribution supports the evaluation of whether the specified prop-

erty holds with high (or low) probability, enabling formal reasoning about cybersickness under uncertainty and partial observability.

Hard vs. Soft Conditioning. Evaluating these properties $\phi(\mathbf{X})$ usually requires reasoning under various forms of evidence, which fall into two categories: *hard* and *soft* conditioning. Hard conditioning applies when variable values are known precisely (e.g., $\texttt{luminance} = 5$ or $\texttt{luminance} \geq 3$). This is formalized via indicator function $C(\mathbf{x})$:

$$C(\mathbf{x}) = \begin{cases} 1 & \text{if state } \mathbf{x} \text{ satisfies the condition,} \\ 0 & \text{otherwise.} \end{cases}$$

Only those states consistent with the hard evidence contribute to the conditioned distribution, yielding:

$$P(\phi = 1 \mid C) = \frac{\sum_{\mathbf{x}} \phi(\mathbf{x}) C(\mathbf{x}) P(\mathbf{x})}{\sum_{\mathbf{x}} C(\mathbf{x}) P(\mathbf{x})}.$$

Soft conditioning captures uncertain or noisy information, such as $\texttt{luminance} \sim \mathcal{N}(5,1)$. This is modeled via a likelihood function $\ell(\mathbf{x})$, which reweights each outcome by its likelihood:

$$P_\ell(\mathbf{x}) = \frac{\ell(\mathbf{x}) P(\mathbf{x})}{\sum_{\mathbf{x}} \ell(\mathbf{x}) P(\mathbf{x})}, \quad P(\phi = 1 \mid \ell) = \frac{\sum_{\mathbf{x}} \phi(\mathbf{x}) \ell(\mathbf{x}) P(\mathbf{x})}{\sum_{\mathbf{x}} \ell(\mathbf{x}) P(\mathbf{x})}.$$

In practice, VR scenarios involve both types of evidence. For instance, latency may be fixed (hard condition), while physiological measures (e.g., GSR) are subject to noise (soft condition). The combined evidence is represented using unified weighting function:

$$w(\mathbf{x}) = \left(\prod_j C_j(\mathbf{x}) \right) \cdot \left(\prod_k \ell_k(\mathbf{x}) \right), P(\phi = 1 \mid C, \ell) = \frac{\sum_{\mathbf{x}} \phi(\mathbf{x}) w(\mathbf{x}) P(\mathbf{x})}{\sum_{\mathbf{x}} w(\mathbf{x}) P(\mathbf{x})}.$$

This general formulation supports the rigorous evaluation of safety properties under partial, noisy, or adversarial conditions. Its integration within a structured Bayesian Network trained from VR data enables formal verification of cybersickness risk—a core contribution of this paper. This verification process is depicted in Figure 2, which illustrates the flow from hierarchical BN construction through structure learning, node reduction, and posterior inference under hard and soft evidence. Boolean properties, such as $[\text{FMS} > 4]$, representing encountering cybersickness in this work, are verified by computing posterior probabilities under these conditions, enabling probabilistic safety guarantees to be established.

4 Experiment Settings

In this section, we detail our experimental setup and methodology for evaluating the proposed cybersickness model. Section 4.1 covers the dataset collection and labeling process, Section 4.2 outlines the data preprocessing and feature selection steps, Section 4.3 describes the training process of the BN, and Section 4.4 presents implementation of the verification of the BN using Dice [10].

4.1 Dataset Description

Our Bayesian Network framework requires a dataset containing both system-level parameters and physiological signals to support rigorous probabilistic modeling and verification of cybersickness. To this end, we use the "VRWalking" dataset [32], chosen specifically due to its comprehensive collection of the variables necessary for our analysis. This dataset consists of data collected from 36 participants (23 male, 13 female; mean age = 25.67 years, SD = 7.22) who navigated virtual mazes by physically walking for 15 minutes while simultaneously performing cognitive tasks related to working memory and attention. System parameters—including optical flow and Histogram of Oriented Gradients (HOG)—were captured using an HTC-Vive Pro Eye headset. Physiological signals (e.g., heart rate (HR) and galvanic skin response (GSR)) were measured via external biosensors, a detailed explanation is provided in Table 2. Although subjective self-reported metrics (e.g., cybersickness questionnaires, mental load, physical exertion) were also collected, our Bayesian modeling approach relies primarily on objective measures to ensure consistency and rigor within the verification framework. Not all participants provided complete data across the required dimensions. Therefore, we retained data from 32 participants, whose records spanned the complete range of cybersickness severity. This subset provides a robust basis for training, validating, and formally verifying our BN, effectively capturing both system-level and physiological influences on cybersickness.

Table 2: Categorization of system-level and human-centered features used in the Bayesian Network model for cybersickness prediction.

Category	Features
System Features	**luminance**: Scene brightness level. **hog_features**: Histogram of Oriented Gradients, captures motion-related texture. **spectral_entropy**: Frequency-domain signal complexity in visual input. **temporal_smoothness**: Measures visual jitter or temporal instability across frames [29].
Human Physiological Features	**Left_Openness**, **Right_Openness**: Eyelid openness, linked to fatigue and visual strain. **Left_Diameter**, **Right_Diameter**: Pupil size, reflects cognitive load, attention, or stress. **Combined_Gaze_Error_Angle**: Discrepancy between actual and target gaze direction. **GSR**: Galvanic Skin Response, indicator of emotional and physiological arousal. **ReacTime**: Reaction time during tasks, reflects cognitive load and task engagement.
Outcome Variable	**fms**: Fast Motion Sickness score, the target variable indicating cybersickness severity.

4.2 Data Preprocessing

The preprocessing involved comprehensive feature refinement, dataset balancing, and careful partitioning to ensure accurate causal modeling and rigorous formal verification of cybersickness risk.

Features selection: Initially, the dataset contained 93 features, including objective system-level parameters, physiological signals, and subjective self-reported metrics. Subjective user questionnaire data (e.g., mental load ratings) were explicitly excluded to maintain consistency with an objective verification framework and remove their inherent variability and incomplete coverage across participants. Redundant objective features were identified using correlation analysis (Pearson correlation threshold of 0.8), and domain expertise was applied to guide removal or consolidation. This systematic process reduced the feature set from 93 to 39 variables. Further dimensionality reduction consolidated logically related features, enhancing interpretability and causal plausibility; for instance, gaze error angle was computed from gaze direction and target direction, and optical flow measurements were aggregated into summary motion metrics. Ultimately, the dataset was refined to a concise and interpretable set of 21 relevant features, selected specifically due to their demonstrated relevance to cybersickness severity and suitability for causal inference. These included 10 system-level features (e.g., luminance, spectral entropy, temporal smoothness, optical flow, Histogram of Oriented Gradients, and frame rate) and 11 human-level physiological features (e.g., eye tracking metrics, galvanic skin response, pupil dilation, and heart rate variability).

Class balance enhancement: Following feature refinement, significant dataset imbalance required attention. The original dataset exhibited severe class imbalance; cybersickness severity level 1 comprised 16,900 data points, whereas severity level 6 contained only 660 data points. Additionally, participant-level data were highly non-independent and identically distributed (non-i.i.d.) since each participant contributed multiple correlated samples. To address this, participant-level selection criteria were implemented: from the original 36 participants, data from 32 participants were explicitly selected based on their coverage across a broad spectrum of cybersickness levels (CS levels 1-7). Each selected participant contributed multiple samples spanning at least five cybersickness

severity levels, ensuring balanced representation and reducing bias.

The resulting dataset still presented challenges with class imbalance; thus, the Synthetic Minority Over-sampling Technique (SMOTE) [5] was applied, chosen specifically for its effectiveness in balancing datasets without losing valuable information compared to undersampling methods. Recognizing potential risks of SMOTE generating unrealistic synthetic samples, participant-level 5-fold cross-validation was explicitly implemented, ensuring that samples from the same participant never appeared simultaneously in both training and validation sets, thereby mitigating data leakage and overfitting concerns. Finally, 150 real data points were set aside as an independent test set, specifically drawn from held-out participants excluded entirely from training and SMOTE augmentation procedures. This separation assured rigorous and unbiased evaluation of generalization performance.

4.3 Training and inference Settings

To facilitate the learning process, continuous numerical variables were discretized into categorical bins to enable the Bayesian Network (BN) to operate within a discrete variable framework. This approach simplifies the exact computation of conditional probabilities, which is critical for the precise probabilistic verification of cybersickness conditions. Specifically, K-means clustering [31] was applied to categorize each numerical feature into five distinct bins. The number of bins was empirically chosen to balance the retention of variability with computational efficiency for exact inference.

For example, *"luminance"*, representing scene brightness, was divided into five bins based on the clustering results. Lower bins (e.g., levels 1 and 2) correspond to lower luminance values, while higher bins (e.g., levels 4 and 5) represent higher luminance values. The boundaries of these bins were determined by the clustering algorithm, which groups data points with similar values into the same bin. This ensures that each bin captures a meaningful range of the feature's variability. Throughout this paper, all numerical features are referenced in terms of these discrete levels, with higher bins generally indicating higher values of the corresponding feature. This discretization enables consistent and interpretable analysis while preserving the relationships between features and their impact on cybersickness conditions.

BN structure learning was conducted using the Hill-Climbing algorithm with domain-specific causal constraints. As described in Section 3.1, these constraints enforced a hierarchical structure (System parameters $\rightarrow$ Human physiological responses $\rightarrow$ Cybersickness severity), preserving causal interpretability. Candidate structures were scored using the Bayesian Dirichlet equivalent uniform (BDeu) metric[4], which is well-suited for limited data scenarios. The resulting compact hierarchical structure not only ensures causal interpretability but also yields computational efficiency. While exact inference in general Bayesian Networks is NP-hard [18], our constrained structure limits maximum parents per node to $k \leq 3$ and reduces the network to 12 nodes after pruning. This results in inference complexity of $O(12 \cdot 5^3) \approx O(1,500)$ operations, enabling real-time queries well within VR frame budgets ($<$ 1ms per inference).

The model evaluation used two key metrics. First, *standard accuracy* measured the proportion of correctly classified FMS severity levels. While informative, this metric alone fails to account for inherent subjectivity and uncertainty in FMS labeling. Therefore, we also used *Extended Label Accuracy (± 1)*, which treats a prediction as correct if it falls within one level of the true label.

4.4 Implementation and Verification using Dice

Having established the Bayesian Network training and evaluation procedures, this section details the implementation and formal verification process using the Dice probabilistic programming language [10]. Probabilistic programming languages typically employ either sampling-based inference (e.g., STAN [3]) or exact inference (e.g., Dice, PSI [8]). Given the need for rigorous probabilistic guarantees in formal verification scenarios, Dice was selected as it efficiently performs exact inference, which is critical for obtaining precise posterior distributions required by the verification framework. The Dice implementation involves four primary steps:

Step 1: BN Encoding in Dice. The Bayesian Network is encoded by assigning each node a corresponding conditional probability table. Nodes are defined in topological order to accurately represent causal dependencies established by the learned structure.

Step 2: Encoding Conditions (Evidence). Dice encodes evidence conditions using built-in observational statements, enabling verification under various scenarios:

- *Hard Conditions:* Represent precise constraints on variables. For example, enforcing discretized luminance greater than level 3 (in a 5-bin discretization) is encoded as: *observe*(luminance > 3);
- *Soft Conditions:* Represent uncertain evidence. For instance, "hog_features at least discretized level 4 with 90% probability": *observe*((hog_features $>= 4$) with 90% probability);

Step 3: Property Specification. Verification properties are defined as Boolean nodes in Dice. For example, cybersickness severity exceeding discretized level 4 (indicating significant cybersickness) is represented as: $\phi(\mathbf{X}) = [\text{FMS} > 4]$.

Step 4: Exact Inference and Verification. Dice computes posterior probabilities exactly, as follows:

$$P(\phi_{\text{FMS}} = \text{true} \mid C, \ell) = \frac{\sum_{\mathbf{x}} \phi_{\text{FMS}}(\mathbf{x}) \cdot w(\mathbf{x}) \cdot P(\mathbf{x})}{\sum_{\mathbf{x}} w(\mathbf{x}) \cdot P(\mathbf{x})}, \tag{2}$$

where $\phi_{\text{FMS}}(\mathbf{x})$ is the Boolean condition (FMS > 4), and $w(\mathbf{x})$ represents combined evidence conditions (Section 3.2).

Illustrative Verification Scenario. Consider verifying cybersickness risk under realistic VR conditions with mixed evidence:

- *Hard Condition:* luminance strictly above discretized level 3.
- *Soft Condition:* HOG features (visual complexity) at least discretized level 4 with 90% probability, reflecting uncertainty or adversarial conditions.

Formally, this verification scenario is represented as:

$$P(\text{FMS} > 4 \mid \text{luminance} > 3, \text{hog\_features} \geq 4 \sim \text{Bernoulli}(0.9)). \tag{3}$$

Dice calculates this posterior probability exactly, quantifying cybersickness risk explicitly. This allows direct verification against formal safety thresholds, confirming whether VR scenarios remain within predefined safety bounds.

Mitigation Strategies Based on Verification Outcomes. The explicit quantification of cybersickness risk using Dice's exact inference provides a promising foundation for the design of real-time mitigation strategies. Specifically, when the inferred probability of severe cybersickness surpasses predefined safety thresholds (e.g., $P(\text{FMS} > 4) \geq 0.8$), actionable system interventions can be triggered automatically. For instance, if high luminance is a contributing factor, the system can reduce display brightness to a safer level, thereby lowering visual strain and its impact on cybersickness likelihood. Similarly, when scene complexity (reflected in elevated HOG features) significantly increases risk, the system may dynamically simplify the scene by reducing texture detail or object density. These content-level adjustments help maintain a balance between immersion and safety.

Real-time physiological monitoring also plays a crucial role in such adaptive systems. Galvanic Skin Response (GSR), for example, has been identified as a key physiological correlate of stress. Continuous monitoring of GSR can enable immediate system interventions when user stress exceeds safe thresholds. Such interven-

tions might include visual simplification, short rest recommendations, or a temporary pause in intense VR activity.

Furthermore, user-centric responses can be incorporated through feedback loops. For example, if the risk estimate crosses a warning threshold, the system may provide real-time alerts to the user or suggest adaptive pauses to avoid physiological overload. These mechanisms serve as soft interventional layers that enhance system transparency and user trust while preserving engagement.

A notable extension to this verification workflow is modeling the mitigations themselves within the Dice framework. By encoding interventions as modifications to the Bayesian Network or as changes to evidence parameters, one can re-evaluate the cybersickness risk post-mitigation. This allows formal guarantees not only about the current risk level but also about whether proposed interventions are sufficient to bring risk below a target threshold. Formally, this entails verifying a post-mitigation condition of the form:

$$P(\text{FMS} > 4 \mid \text{mitigation}(C)) \leq \tau,$$

where τ is a safety bound, and mitigation(C) reflects updated conditions due to system adaptation. The following pseudocode illustrates how these concepts translate into a runtime context:

Algorithm 1 Adaptive Mitigation on Verified Cybersickness Risk

```
if infer(P(FMS > 4 | conditions)) > 0.8 then
    reduce(luminance)
    simplify(scene)
    notify("Cybersickness risk elevated. Adjusting for safety.")
end if
```

Although these mitigation strategies are grounded in probabilistic verification, their effectiveness remains to be validated through empirical studies involving diverse user populations. Additionally, the underlying models and thresholds are derived from a specific dataset, potentially introducing dataset-specific biases. Therefore, while the proposed framework provides a rigorous foundation for real-time adaptive safety in VR, future work must emphasize generalization and human-centered validation across broader usage.

5 Results Analysis

In this section, we present the findings from our Bayesian Network–based cybersickness modeling and verification. Section 5.1 evaluates predictive performance using accuracy metrics, and Section 5.2 details how probabilistic verification was conducted to ensure safety constraints under varying VR conditions.

5.1 Bayesian Network Prediction Analysis

To enhance the prediction of high cybersickness levels, we applied a series of data preprocessing steps as described in Section 4.2. The Bayesian Network (BN) was trained on a carefully curated subset of features, selected through a combination of statistical correlation analysis and expert domain knowledge. From an initial pool of 20 candidate features (10 system-level and 10 human-level), the final model retained four critical system-level features (Spectral Entropy, HOG Features, Temporal Smoothness, and Luminance) and seven human-level physiological signals, including galvanic skin response, pupil dilation, heart rate variability, and eye tracking metrics. Figure 4 illustrates the learned BN structure, showing the causal and probabilistic dependencies among these variables.

Model performance was evaluated using participant-level five-fold cross-validation, ensuring that data from each participant appeared in only one fold to avoid data leakage. To account for the inherent subjectivity and variability in cybersickness ratings, we adopted two performance metrics:

1. **Exact Label Accuracy**, which counts only perfectly matched predictions as correct.

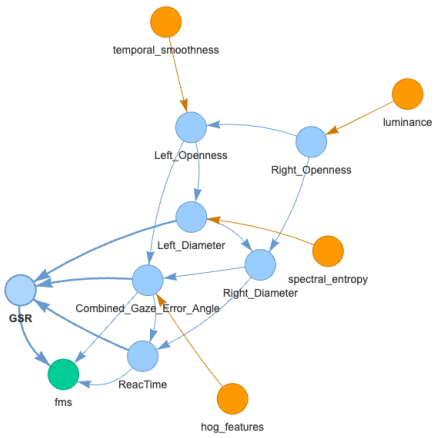

Figure 4: Structure of the learned BN showing system-level features (orange), physiological human-level features (blue), and the cybersickness severity (FMS, green) node. Directed edges represent learned dependencies that capture causal influences among these variables.

2. **Extended Label Accuracy ($\pm$1)**, which treats predictions within one level of the true FMS score as correct, reflecting perceptual ambiguity in cybersickness severity.

The average Exact Label Accuracy across folds was 62.12%, reflecting the difficulty of classifying subjective symptom severity at fine granularity. However, the Extended Label Accuracy ($\pm$1) increased to 83.55%, showing that most errors were minor deviations rather than severe misclassifications. This substantial improvement demonstrates the network's capability to capture meaningful probabilistic distributions over FMS levels.

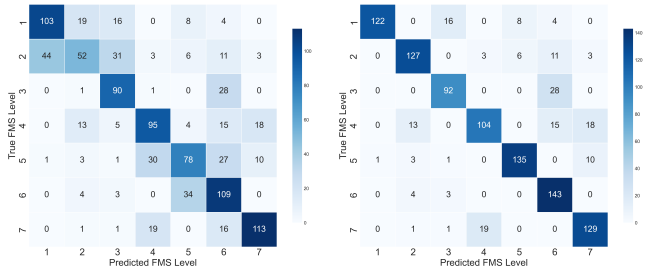

Figure 5: Comparison of confusion matrices for Fold 1 under different accuracy metrics, left figure shows exact Label Accuracy, right figure shows extended label accuracy.

As shown in Figure 5 , confusion matrices from Fold 1 (as an example) illustrate that errors typically occur between adjacent FMS levels. Notably, FMS levels 3 and 4 exhibit the most frequent confusion, which aligns with the variability and overlap expected in moderate cybersickness cases.

Figures 6 show the average per-class accuracy across all folds under both metrics. The model maintains balanced performance across all FMS levels without strong bias toward lower-severity classes. Importantly, accuracy for high-severity levels (FMS 6 and 7) exceeds 70% under Exact Accuracy and surpasses 80% with Extended Accuracy—highlighting the model's reliability in detecting the most safety-critical conditions. These results confirm that the Bayesian Network effectively supports both accurate prediction and robust probabilistic reasoning. By modeling full posterior distributions over cybersickness severity, the model enables formal

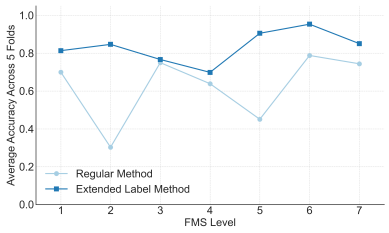

Figure 6: Average accuracy per cybersickness (FMS) level across five cross-validation folds using the Exact Label Accuracy metric (regular) compare with extended label accuracy method.

verification under uncertainty, advancing the design of user-safe, verification-driven VR systems.

5.2 Verification Results Analysis

Our probabilistic verification framework enables formally grounded queries over the Bayesian Network to estimate cybersickness risk under defined evidence conditions. Specifically, this analysis addresses three core verification questions:

- **Q1:** Which parameters have the most significant impact on cybersickness severity?
- **Q2:** How is user cybersickness affected under adverse scenarios where certain parameters are elevated?
- **Q3:** What are the "safe" operational ranges for critical parameters that reliably minimize the risk of cybersickness?

5.2.1 Impact of Individual Parameters

To address Q1, we systematically evaluated individual system parameters and selected human-level signals. This approach enables the simulation of adversarial manipulations (e.g., intentionally increased luminance), supports early risk detection, and guides targeted mitigation strategies. Additionally, examining parameter extremes characterizes worst-case cybersickness scenarios. Each retained system parameter (Spectral Entropy, HOG Features, Temporal Smoothness, and Luminance) and three directly influential human signals were evaluated independently: Galvanic Skin Response, gaze error angle and Reaction Time. These variables were retained due to their empirical relevance and potential causal roles established through Bayesian Network structure learning. Conditions simulating elevated parameter values were applied using Bayesian conditioning, and posterior probabilities $P(\text{FMS} > 4 \mid \text{condition})$ were computed explicitly to estimate cybersickness risk.

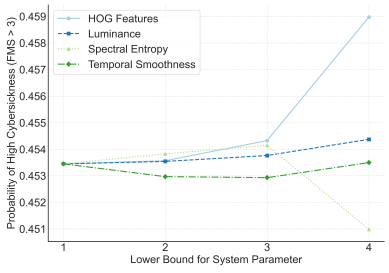

Figure 7: Impact of individual system parameters. While absolute probability shifts are modest, luminance and HOG features consistently exhibit the highest relative influence on FMS elevation.

Figure 7 illustrates how changes in each system parameter affect cybersickness probability. The x-axis denotes parameter lower bounds, while the y-axis shows the posterior probability of elevated FMS levels. Although absolute shifts in probability are modest (typically under 1%), even small probability shifts can be practically meaningful in safety-critical VR scenarios, where subtle parameter adjustments significantly affect user comfort over extended durations. Luminance and particularly, HOG features consistently exhibit the greatest relative influence, highlighting them as critical candidates for real-time monitoring or mitigation. In contrast, temporal smoothness and spectral entropy show negligible correlation, suggesting limited operational utility as standalone indicators of cybersickness.

Figure 8 examines the influence of human physiological signals. Elevated GSR shows a strong correlation with cybersickness: GSR values above 3 correspond to an 80% likelihood of experiencing cybersickness (FMS > 4) compared to a baseline of approximately 45%. Combined gaze error angle also demonstrates significant impact, increasing cybersickness probability from 45% to 60%. These factors directly connect with the cybersickness node (FMS) in our Bayesian Network. In contrast, reaction time, despite its direct connection in the network structure, shows negligible effect on FMS probability. Similarly, other physiological features demonstrate minimal influence on cybersickness probability.

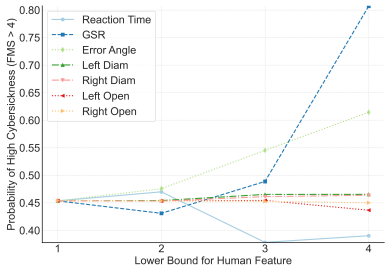

Figure 8: Impact of human physiological signals on cybersickness. Elevated GSR strongly correlates with increased cybersickness risk, while reaction time shows minimal consistent impact.

This analysis isolates the effects of individual parameters, providing insight into their standalone influence on cybersickness. However, features that appear less impactful in isolation may still play a significant role when combined with others. As explored in subsequent experiments, combinatorial effects can amplify cybersickness risk, though exhaustively presenting all possible combinations is infeasible. Instead, we focus on integrating features that show strong individual influence, as these are more likely to yield meaningful compounded effects. These findings underscore the importance of parameter-specific prioritization in operational VR systems. Notably, signals such as luminance and GSR emerge as actionable indicators for real-time monitoring and proactive mitigation. Moreover, identifying such critical parameters and their effective thresholds directly contributes to probabilistic safety guarantees, thereby enhancing formal verification processes.

5.2.2 Probability of Cybersickness Under Adverse Scenario

To address Q2, this analysis combines multiple parameters identified as influential (luminance, HOG features, and GSR). Since these parameters individually showed significant correlations with increased FMS, their combined elevation represents realistic conditions where heightened visual complexity and physiological stress might occur (e.g., demanding task or environmental changes). An exhaustive search was conducted across discretized parameter bounds (levels 1–5), yielding 125 possible combinations. The analysis revealed that cybersickness severity peaks when all three parameters are simultaneously constrained to discrete levels of at least 4. Formally, it corresponds to the posterior probabilistic query:

$$P(\text{FMS} > 4 \mid \text{luminance} \geq 4, \text{HOG} \geq 4, \text{GSR} \geq 4),$$

which quantifies cybersickness risk under these conditions.

Figure 9 illustrates the resulting shift in FMS distribution. The mean FMS level increases notably from 4.21 (baseline) to 6.13 (adverse scenario), and the probability of severe cybersickness (FMS $>$ 4) escalates from 45.34% to 80.65%. Practically, this translates to roughly four out of five users experiencing significant discomfort under these elevated conditions. Such results underscore the need for adaptive mitigation mechanisms—such as dynamic luminance adjustment, scene simplification, or proactive user alerts—to maintain user comfort and operational safety in mixed-reality systems.

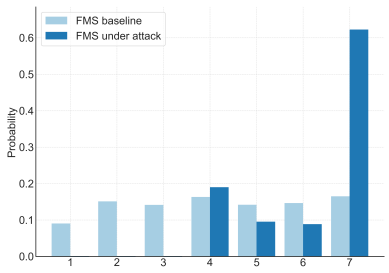

Figure 9: Shift in FMS distribution comparing baseline conditions (light blue) and elevated conditions (dark blue). Under elevated luminance, HOG features, and GSR (levels ≥ 4), cybersickness risk increases significantly, particularly at higher FMS scores.

5.2.3 Establishing Operational Safety Guarantees

To address Q3, this analysis defines and evaluates formal probabilistic safety constraints, such as the following safety constraint: $P(\text{FMS} \leq 4) = 0.95$. This constraint represents a soft probabilistic condition, explicitly requiring that severe cybersickness occurs with no more than a 5% probability under given conditions. Due to its strong correlation and direct impact on FMS, GSR was selected for in-depth posterior analysis under this constraint.

Figure 10 shows the posterior distribution of GSR under imposed safety constraints; a clear shift toward lower GSR. The likelihood of high-risk GSR (levels 4–5) becomes negligible under this condition. Practically, maintaining GSR within levels 1–3 (representing physiologically mild to moderate stress) can serve as a clear operational guideline for real-time monitoring and intervention. This recommendation also aligns with the insights presented in Figure 8.

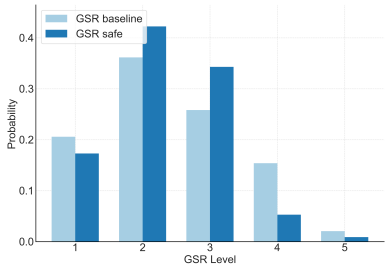

Figure 10: Comparison of Galvanic Skin Response (GSR) distributions under baseline (light blue) and safety-constrained conditions (dark blue). Enforcing the probabilistic safety constraint (FMS ≤ 4 with 95% probability) significantly reduces the likelihood of elevated GSR levels, effectively guiding real-time system management.

These results provide formally grounded safety verification, directly informing mixed-reality system design and adaptive management. By continuously monitoring and maintaining critical physiological signals such as GSR within empirically derived safe bounds, systems can proactively adjust environmental and operational parameters—ensuring immersive experiences remain within safe, comfortable physiological limits.

6 Conclusion and Future Work

This paper introduced a novel probabilistic verification framework using Bayesian Networks (BN) to assess and mitigate cybersickness risks in virtual and mixed reality (VR/MR) environments. By explicitly modeling interactions between system parameters, physiological signals, and cybersickness severity, this approach provides interpretable and formally verifiable insights into cybersickness dynamics. Unlike traditional deep learning approaches, the BN framework offers structured causal transparency, enabling explicit reasoning about probabilistic safety guarantees. Empirical analyses demonstrated the framework's effectiveness in both accurately predicting cybersickness severity and identifying high-impact mitigation strategies. Notably, under adverse scenarios with elevated luminance, HOG features, and galvanic skin response (GSR), the probability of severe cybersickness (FMS $>$ 4) increased substantially from 45.3% to 80.7%. Additionally, GSR was confirmed as a critical physiological indicator, with operational safety requiring its levels to remain within discrete ranges (levels 1–3) to reliably maintain low cybersickness risk. These results underscore the practical utility of probabilistic verification—particularly given the subjective and inherently variable nature of cybersickness—in ensuring robust safety thresholds for immersive environments. By effectively bridging predictive modeling with formal verification, this paper provides a robust operational foundation for designing safer, adaptive, and user-centric VR/MR experiences.

While our framework demonstrates effective cybersickness verification, several limitations warrant discussion. First, generalization across diverse VR/MR platforms remains challenging—different headsets and environments (e.g., forest scenes with unique luminance patterns) may require model fine-tuning or structural updates to capture platform-specific features. Second, our current approach processes visual information through extracted features (luminance, HOG, spectral entropy) and physiological signals; incorporating direct sensory modalities such as visual, audio could impact cybersickness but would necessitate expanded signal processing pipelines, model restructuring, and updated verification procedures. Third, our offline verification analyzes prespecified scenarios, which, while comprehensive, cannot exhaustively cover all possible runtime conditions; developing online verification capabilities that adapt to real-time user responses represents an important future direction. Additionally, alternative verification frameworks such as sampling-based methods could complement our probabilistic approach by enabling verification without explicit probabilistic models, using statistical sampling and confidence intervals to estimate safety properties from black-box simulators or complex user models that are computationally intractable for exact inference. Finally, advanced data augmentation techniques like Safe-Level-SMOTE [1] could better address class imbalance while preserving the natural distribution of cybersickness severity levels. Addressing these limitations will enhance the framework's robustness and practical applicability across diverse VR/MR deployments.

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] C. Bunkhumpornpat, K. Sinapiromsaran, and C. Lursinsap. Safe-level-smote: Safe-level-synthetic minority over-sampling technique for handling the class imbalanced problem. In *Advances in knowledge discovery and data mining: 13th Pacific-Asia conference, PAKDD*

2009 Bangkok, Thailand, April 27-30, 2009 proceedings 13, pp. 475–482. Springer, 2009. doi: 10.1007/978-3-642-01307-2_43 9

[2] F. Campos, M. Campos, T. Silva, and M. Van Gisbergen. User experience in virtual environments: Relationship between cybersickness issues and the optical aspects of the image by contrast levels. In D. Russo, T. Ahram, W. Karwowski, G. Di Bucchianico, and R. Taiar, eds., *Intelligent Human Systems Integration 2021. IHSI 2021*, vol. 1322 of *Advances in Intelligent Systems and Computing*. Springer, Cham, 2021. doi: 10.1007/978-3-030-68017-6_65 1, 2

[3] B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell. Stan: A probabilistic programming language. *Journal of statistical software*, 76:1–32, 2017. 6

[4] A. M. Carvalho and T. Avançados. Scoring functions for learning bayesian networks. 2009. 6

[5] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357, June 2002. doi: 10.1613/jair.953 6

[6] S. V. Cobb, S. Nichols, A. Ramsey, and J. R. Wilson. Virtual reality-induced symptoms and effects (vrise). *Presence: Teleoperators and Virtual Environments*, 8(2):169–186, 1999. doi: 10.1162/105474699566152 2

[7] S. V. G. Cobb, S. Nichols, A. Ramsey, and J. R. Wilson. Virtual reality-induced symptoms and effects (vrise). *Presence*, 8(2):169–186, 1999. doi: 10.1162/105474699566152 2

[8] T. Gehr, S. Misailovic, and M. Vechev. Psi: Exact symbolic inference for probabilistic programs. In *Computer Aided Verification: 28th International Conference, CAV 2016, Toronto, ON, Canada, July 17-23, 2016, Proceedings, Part I 28*, pp. 62–83. Springer, 2016. 6

[9] C. Groth, J.-P. Tauscher, N. Heesen, S. Castillo, and M. Magnor. Visual techniques to reduce cybersickness in virtual reality. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 486–487, 2021. doi: 10.1109/VRW52623.2021.00125 1

[10] S. Holtzen, G. Van den Broeck, and T. Millstein. Scaling exact inference for discrete probabilistic programs. *Proceedings of the ACM on Programming Languages*, 4(OOPSLA):1–31, 2020. 5, 6

[11] 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, 2021. doi: 10.1109/ISMAR52148.2021.00017 2

[12] M. Javaid and A. Haleem. Virtual reality applications toward medical field. *Clinical Epidemiology and Global Health*, 8(2):600–605, 2020. 1

[13] Z. Jiang, M. Pajic, and R. Mangharam. Cyber–physical modeling of implantable cardiac medical devices. *Proceedings of the IEEE*, 100(1):122–137, 2012. doi: 10.1109/JPROC.2011.2161241 3

[14] S. Junges, J.-P. Katoen, N. Jansen, and U. Topcu. Probabilistic verification for cognitive models: Controller synthesis and model evaluation. 2016. 3

[15] C. Kern and M. R. Greenstreet. Formal verification in hardware design: a survey. *ACM Trans. Des. Autom. Electron. Syst.*, 4(2):123–193, Apr. 1999. doi: 10.1145/307988.307989 3

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

[17] H. Kim, D. Kim, W. Chung, K. Park, J. Kim, D. Kim, et al. Clinical predictors of cybersickness in virtual reality (vr) among highly stressed people. sci rep 11: 12139, 2021. 1, 2

[18] U. Kjærulff. Reduction of computational complexity in bayesian networks through removal of weak dependences. In *Uncertainty in Artificial Intelligence*, pp. 374–382. Elsevier, 1994. 6

[19] 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 2

[20] 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 2

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

[22] J. J. LaViola. A discussion of cybersickness in virtual environments. *SIGCHI Bull.*, 32(1):47–56, Jan. 2000. doi: 10.1145/333329.333344 2

[23] J. J. LaViola Jr. A discussion of cybersickness in virtual environments. *ACM SIGCHI Bulletin*, 32(1):47–56, January 2000. doi: 10.1145/333329.333344 2

[24] X. Leroy. Formal verification of a realistic compiler. *Communications of the ACM*, 52(7):107–115, 2009. doi: 10.1145/1538788.1538814 3

[25] L. Li. Comprehensive survey on adversarial examples in cybersecurity: Impacts, challenges, and mitigation strategies. *arXiv preprint arXiv:2412.12217*, 2024. 3

[26] J.-W. Lin, H. Duh, D. Parker, H. Abi-Rached, and T. Furness. Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. In *Proceedings IEEE Virtual Reality 2002*, pp. 164–171, 2002. doi: 10.1109/VR.2002.996519 1

[27] A. Mazloumi Gavgani, F. R. Walker, D. M. Hodgson, and E. Nalivaiko. A comparative study of cybersickness during exposure to virtual reality and "classic" motion sickness: are they different? *Journal of Applied Physiology*, 125(6):1670–1680, 2018. 2

[28] T. Murray, D. Matichuk, M. Brassil, P. Gammie, T. Bourke, S. Seefried, C. Lewis, X. Gao, and G. Klein. sel4: from general purpose to a proof of information flow enforcement. In *2013 IEEE Symposium on Security and Privacy*, pp. 415–429. IEEE, 2013. 3

[29] R. M. Nasiri, Z. Duanmu, and Z. Wang. Temporal motion smoothness and the impact of frame rate variation on video quality. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 1418–1422, 2018. doi: 10.1109/ICIP.2018.8451613 5

[30] S. Palmisano, R. Mursic, and J. Kim. Vection and cybersickness generated by head-and-display motion in the oculus rift. *Displays*, 46:1–8, 2017. doi: 10.1016/j.displa.2016.11.001 1

[31] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. 6

[32] 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. 2, 5

[33] S. P. Smith. Exploring cybersickness experiences via a markov chain model. *IEEE Access*, 12:33595–33604, 2024. doi: 10.1109/ACCESS.2024.3371516 2

[34] 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. 1, 2

[35] K. M. Stanney, R. S. Kennedy, and J. M. Drexler. Cybersickness is not simulator sickness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 41:1138 – 1142, 1997. 2

[36] A. Szpak, S. C. Michalski, D. Saredakis, C. S. Chen, and T. Loetscher. Beyond feeling sick: The visual and cognitive aftereffects of virtual reality. *Ieee Access*, 7:130883–130892, 2019. 1

[37] 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. doi: 10.1007/s10055-022-00638-2 1

[38] 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. 4

[39] J. Wang, R. Shi, W. Zheng, W. Xie, D. Kao, and H.-N. Liang. Effect of frame rate on user experience, performance, and simulator sickness in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2478–2488, 2023. doi: 10.1109/TVCG.2023.3247057 1

[40] M. Wedel, E. Bigné, and J. Zhang. Virtual and augmented reality: Advancing research in consumer marketing. *International Journal of Research in Marketing*, 37(3):443–465, 2020. 1

[41] M. Yalcin, A. Halbig, M. Fischbach, and M. E. Latoschik. Automatic cybersickness detection by deep learning of augmented physiological data from off-the-shelf consumer-grade sensors. *Frontiers in Virtual Reality*, 5:1364207, 2024. doi: 10.3389/frvir.2024.1364207 2

[42] 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. 2

[43] C. Zhang. Investigation on motion sickness in virtual reality environment from the perspective of user experience. In *2020 IEEE 3rd International Conference on Information Systems and Computer Aided Education (ICISCAE)*, pp. 393–396, 2020. doi: 10.1109/ICISCAE51034.2020.9236907 1

[44] Y. Zhu, T. Li, and Y. Wang. Real-time cross-modal cybersickness prediction in virtual reality, 2025. 2