POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

CASSANDRA: An Interpretable Real-time System for Cybersickness Detection and Adaptive Decision Support

Tesi di Laurea Magistrale in
Computer Science and Engineering - Ingegneria Informatica

Teka Kimbi Ntimanputu, 225158

Advisor:
Prof. Marco Domenico Santambrogio

Co-advisors:
Marco Gribaudo
Susanna Bardini
Marco Domenico Buttiglione

Academic year:
2024-2025

Abstract: Cybersickness is a complex and dynamically evolving condition arising from the interaction between sensory stimulation, individual susceptibility, and system-level design choices in immersive Virtual Reality (VR). Although physiological monitoring has increasingly been explored as an objective assessment tool, most existing approaches model cybersickness as a single predictive task, implicitly merging symptom onset and intensity. This simplification limits interpretability and reduces practical applicability, particularly for adaptive systems. This thesis introduces a modular framework that explicitly separates these dimensions by modelling cybersickness through two related tasks: binary detection of symptom onset and estimation of symptom severity. The approach integrates consumer-grade electrocardiography (ECG) with environmental descriptors capturing visual motion and scene dynamics, combining physiological responses with technical characteristics of the virtual environment known to act as triggers. Physiological features are extracted over short temporal windows to capture rapid autonomic variations, while environmental variables encode measurable scene properties. Both tasks share a common preprocessing and evaluation pipeline, differing only in the learning objective, enabling a controlled comparison of the factors associated with onset and severity. Evaluation on both public and independent validation datasets demonstrates that the proposed two-layer formulation outperforms direct regression, reducing severity estimation error from mean absolute error values of approximately 0.18–0.60 to below 0.1 for physiological models, while achieving onset detection accuracy up to 0.95. Environmental features provide informative contributions but yield lower predictive accuracy than physiological signals, particularly for onset detection. Explainability analyses indicate that onset detection is driven by a limited set of dominant features, whereas severity estimation depends on more distributed physiological and environmental effects, supporting the hypothesis that the two processes are partially distinct. Overall, the framework offers an interpretable and principled basis for real-time monitoring and adaptive VR system design.

Key-words: Cybersickness, Virtual Reality, Continuous Physiological monitoring, Environmental and interaction telemetry, Machine learning, Explainability

Abstract in lingua italiana

La cybersickness, conosciuta anche come cinetosi da realtà virtuale immersiva, rappresenta un fenomeno che deriva da diversi fattori ed è il risultato dell'interazione tra stimolazioni sensoriali, suscettibilità individuale e impostazioni dell'ambiente virtuale. Il monitoraggio fisiologico, sempre più impiegato come strumento oggettivo di valutazione, viene spesso utilizzato negli studi sulla cybersickness considerando il fenomeno come un indicatore unico, senza distinguere chiaramente tra l'insorgenza dei sintomi e la loro intensità. Questo approccio riduce la possibilità di interpretare i risultati e limita l'applicazione pratica dei modelli, soprattutto negli scenari adattivi in tempo reale. In questa tesi si propone una struttura modulare che affronta separatamente queste due dimensioni, modellando la cybersickness tramite due compiti distinti, ma correlati tra loro, ossia il rilevamento dell'insorgenza dei sintomi e la stima della loro gravità. L'approccio integra segnali fisiologici acquisiti tramite dispositivi commerciali in grado di registrare l'elettrocardiogramma (ECG), fornendo informazioni sulle variazioni del battito cardiaco e sull'attività del sistema autonomo, insieme a descrittori dell'ambiente virtuale che catturano le dinamiche visive della scena, come spostamenti degli oggetti, fluidità della prospettiva e cambiamenti di inquadratura, mettendo in relazione le risposte dell'utente con le caratteristiche tecniche dell'esperienza immersiva. Le informazioni fisiologiche vengono analizzate su brevi finestre temporali, in modo da cogliere variazioni rapide del battito cardiaco e dell'attività del sistema autonomo, mentre le variabili ambientali codificano aspetti della scena virtuale noti in letteratura come possibili fattori scatenanti dei sintomi. Entrambi i compiti vengono elaborati all'interno della stessa procedura di preparazione e analisi dei dati, che gestisce la normalizzazione, l'estrazione delle caratteristiche e la predisposizione dei dati per l'addestramento dei modelli, differenziandosi solo per l'obiettivo di apprendimento. Questa impostazione permette di confrontare direttamente i meccanismi responsabili dell'insorgenza dei sintomi con quelli che ne determinano la gravità. La validazione sperimentale, condotta sia su dataset pubblici sia su un dataset indipendente raccolto in condizioni eterogenee, mostra come la formulazione a due livelli migliori le prestazioni rispetto a un approccio di regressione diretta. In particolare, l'errore medio assoluto nella stima della gravità, inizialmente compreso tra 0,18 e 0,60, si riduce a valori inferiori a 0,1 nei modelli basati su segnali fisiologici, mentre il rilevamento dell'insorgenza raggiunge un'accuratezza fino a 0,95. Sebbene le caratteristiche dell'ambiente virtuale abbiano un ruolo significativo, i segnali fisiologici si rivelano più predittivi, soprattutto nelle fasi iniziali dell'insorgenza dei sintomi. Le analisi di explainability indicano che l'insorgenza dei sintomi è guidata da un numero limitato di variabili predominanti, mentre la gravità risulta influenzata da contributi più distribuiti, sia tra le componenti fisiologiche sia tra quelle ambientali dell'ambiente virtuale. In conclusione, la struttura ed i risultati proposti rappresentano una base metodologica solida e interpretabile per il monitoraggio oggettivo della cybersickness e per lo sviluppo di sistemi VR in grado di adattarsi in tempo reale alle esigenze dell'utente.

Parole chiave: Cybersickness, Realtà Virtuale, Monitoraggio fisiologico continuo, Telemetria ambientale , Machine learning, Explainability

1. Introduction

This introduction provides the background and context necessary to understand the factors contributing to cybersickness in immersive Virtual Reality (VR) systems, including both physiological responses and environmental triggers, as well as the limitations of existing predictive models. In Section 1.1, the general context and motivation for this work are presented. Section 1.2 outlines the main challenges and knowledge gaps identified in previous studies. Section 1.3 defines the research objectives and key contributions. Finally, Section 1.4 provides an overview of the general structure of the thesis.

1.1. Context

Virtual Reality has reached a technological maturity that enables high-fidelity immersion across sectors such as aerospace, medicine, entertainment, and industrial design. However, its large-scale deployment remains fundamentally constrained by **cybersickness**. With prevalence rates between 60% and 80%, this systemic issue often manifests within minutes, forcing early session termination and limiting the feasibility of applications requiring sustained immersion [37]. The consequences extend beyond physical discomfort to significant operational and economic impacts. In high-stakes training, such as surgical or flight simulations, users may adopt compensatory strategies to mitigate nausea. This risks reinforcing negative training and the acquisition of incorrect habits that can be catastrophic in real-world scenarios [44]. Furthermore, post-immersion recovery times and safety risks contribute to high user dropout, reducing the return on investment for immersive systems. Traditional retrospective questionnaires are insufficient for managing these risks because they do not capture the real-time progression of symptoms or their specific environmental triggers. While **wearable sensing** and **machine learning** now allow for continuous physiological monitoring, detection alone is a passive solution. To achieve active prevention, there is a need for explainable and interpretable tools capable of linking physiological distress to specific environmental factors. Establishing this causal link is essential for VR to progress toward environment-aware design, where simulations adapt dynamically to safeguard user well-being and support the sustainable use of VR as a human-centred technology.

1.2. Motivation

Despite being recognized as a major limitation for VR, current approaches to assessing and mitigating cybersickness remain inadequate for real-world applications. Most research still relies on **post-exposure subjective questionnaires**, which provide only aggregated retrospective data and lack the temporal resolution required to track when symptoms appear and how they evolve in response to specific virtual stimuli [26]. This static assessment method limits the ability to track how environmental factors affect users' physiological responses. Although advances in wearable sensors now enable continuous monitoring of signals such as heart rate and electrodermal activity, most current models treat cybersickness either as a simple binary classification or a single regression problem. These **black-box** approaches prioritize predictive accuracy over interpretability, providing little insight into why a user experiences discomfort [28]. This lack of transparency is especially concerning in safety-critical applications, where system behaviour must be actionable. Simply detecting discomfort is of limited value if the system cannot pinpoint environmental triggers or differentiate between the mechanisms that drive symptom onset and those that influence symptom severity. Addressing these challenges requires a framework that views cybersickness as a time-dependent process, links physiological responses to simultaneous virtual-environment events via interpretable models, and enables a shift from passive detection to proactive, environment-aware VR systems that prioritise user well-being.

1.3. Research Questions and Contributions

Building on the gaps highlighted in Section 1.2, this thesis examines cybersickness as a dynamic, system-level phenomenon arising from the ongoing interplay between human physiology and synthetic environments. To tackle these challenges, four interconnected research questions provide the guiding framework. The first question investigates whether cybersickness onset can be detected in real time using time-resolved features from a single commercially available ECG sensor, addressing the limitations of retrospective, session-level labels. The associated objective is to develop and evaluate an online detection framework capable of identifying early physiological changes that signal cybersickness during immersive VR exposure. The second question investigates whether separating cybersickness into distinct stages of onset detection and severity estimation produces more accurate and temporally precise models than single-stage approaches. This is addressed by comparing the predictive performance and explanatory power of single-stage and multi-stage models using both physiological and environmental data. The third research question focuses on explainability. It asks how well ECG-based models can show which parts of the virtual environment are causing a user's discomfort, while separating these effects from normal physiological changes. The objective is to use explainable machine learning methods to interpret the model's predictions and link observed stress signals to measurable environmental conditions at the same time. The fourth research question looks at which features of the virtual environment are consistently linked to physiological signs of cybersickness onset and the progression of symptoms. The goal is to identify patterns that can support proactive, environment-aware strategies to reduce user discomfort. In response to these questions, the thesis makes several contributions. First, it introduces a structured, modular taxonomy for cybersickness analysis, which organizes environmental parameters and mitigation components into a unified framework for sensing, modelling, and intervention in immersive systems. Second, it proposes a time-resolved framework for physiological and environmental monitoring, providing a computational pipeline capable of continuous, real-time assessment of user states during VR exposure using a single commercially available ECG sensor alongside concurrent environmental data. Third, by separating cybersickness into distinct stages of onset detection and severity estimation, the thesis presents a multi-stage phenomenological model that distinguishes triggering mechanisms from intensity modulation, enhancing both temporal resolution and explainability. Finally, it introduces an environment-oriented, interpretable modelling approach, in which ECG-derived and environmental features are used to link model predictions to controllable and modifiable aspects of the virtual environment, offering actionable insights for adaptive VR systems.

1.4. Outline

As introduced in Section 1, the thesis begins by presenting the research motivation, questions, and scope. Section 2 reviews the literature on cybersickness and physiological monitoring, emphasising the need for interpretable, time-resolved models. Section 3 reviews the state of the art, covering previous studies on physiological assessment, VR stimuli, modelling approaches, and reported findings in cybersickness research. Building on this background, Section 4 introduces a multilayered taxonomy, offering a structured framework to organise physiological responses, environmental factors, and mitigation strategies. Section 5 describes the methodology, including experimental protocols, data collection, and modelling techniques. Section 6 presents the main results, analysing the relationships between physiological signals, VR stimuli, and model predictions. These findings are further discussed in Section 7. Finally, Section 8 concludes the thesis by summarising the key contributions, discussing limitations, and outlining directions for future research.

2. Theoretical Background

This section presents the theoretical foundations underlying cybersickness and reviews computational approaches for its assessment. Section 2.1 introduces the dominant theoretical frameworks and discusses how they motivate the use of physiological signals as objective indicators of user discomfort. Section 2.2 focuses on subjective metrics and psychometric evaluation, reviewing the principal self-report instruments adopted in the literature. Section 2.3 surveys predictive models, covering contemporary machine learning techniques, while Section 2.4 outlines the methodological framework of nested cross-validation, detailing its role in robust model selection, hyper-parameter optimization, and unbiased performance estimation. Section 2.5 presents the evaluation metrics used to assess classification and regression performance. Finally, Section 2.6 addresses explainability and feature attribution, describing general frameworks that support the interpretation and assessment of model predictions.

2.1. Theoretical Models for Cybersickness

Cybersickness is primarily classified as a form of **visually induced motion sickness**, sharing both symptoms and underlying theoretical foundations with traditional motion sickness. As shown in Figure 1, which illustrates a taxonomy of motion sickness types, it can be understood as a specific subset within the broader motion sickness domain. Unlike vehicular motion sickness, however, cybersickness arises in the absence of physical movement, with visual stimuli acting as the main source of sensory incongruence. Within this context, two theoretical frameworks have been most widely adopted to explain the phenomenon, namely **Sensory Conflict Theory** and **Postural Instability Theory**. The **Sensory Conflict Theory** explains motion sickness as the result of inconsistencies between sensory signals, particularly between visual motion cues and vestibular feedback [31, 35]. This phenomenon is commonly experienced during car travel, where a passenger reading a book perceives a stable visual scene while the vestibular system simultaneously detects motion. The resulting discrepancy disrupts the brain's ability to form a coherent perception of self-motion and can lead to nausea and discomfort. A similar process occurs in virtual reality, although in a different form. Rather than sensing motion without corresponding visual input, users often experience the opposite condition, perceiving movement through optic flow while the body remains physically stationary. This reversal still produces a mismatch between expected and received sensory information, placing strain on multisensory integration processes and ultimately giving rise to the characteristic symptoms of cybersickness. [27]. Conversely, the **Postural Instability Theory** proposes that sickness arises when individuals are unable to maintain stable postural control over time [40]. From this perspective, the human body continuously adjusts posture to remain balanced based on visual, vestibular, and proprioceptive inputs. When these cues become unreliable or inconsistent, as often occurs in virtual environments, the control strategies used to maintain stability become less effective. Rather than a single moment of mismatch, the theory emphasises a sustained period of instability, during which the user must continuously correct their posture without achieving a stable state. This prolonged effort increases the demand on the sensorimotor system and, over time, leads to discomfort and the onset of cybersickness symptoms. These theories converge on the view that cybersickness is a physiological process emerging from multisensory integration within the central nervous system (CNS). In particular, discrepancies in visual, vestibular, and proprioceptive inputs are processed in cortical and subcortical regions, giving rise to a state of neural conflict and increased cognitive load. This central processing drives subsequent responses in the autonomic nervous system (ANS), which mediates the body's physiological adaptation to perceived sensory inconsistency. The ANS regulates arousal through the interaction of sympathetic and parasympathetic activity, resulting in measurable changes in cardiovascular and electrodermal dynamics. Within this framework physiological signals can be interpreted hierarchically. Electroencephalography (EEG) captures cortical activity and provides a direct window into neural processing associated with sensory conflict, attention, and cognitive load. In contrast, electrodermal activity (EDA) reflects purely sympathetic activation and is therefore a proxy from arousal and stress-related autonomic responses. Cardiovascular measures derived from electrocardiography (ECG), including heart rate (HR) and heart rate variability (HRV), capture the balance between sympathetic and parasympathetic regulation, providing insight into autonomic control and physiological adaptation over time. Additionally, oculomotor signals, such as vergence and pupil dilation, offer an intermediate layer, linking central processing to sensorimotor behaviour and reflecting visual-vestibular coordination demands. While laboratory-grade sensing modalities provide high-fidelity measurements across these levels, their intrusiveness and setup complexity limit scalability in real-world and immersive environments. Consequently, there is growing interest in **consumer-grade wearable sensors**, which offer a practical trade-off between signal fidelity and real-world deployability in immersive settings. However, physiological responses alone do not fully characterise cybersickness, since they emerge as subsequent manifestations of a complex interaction between the user and the virtual environment. Environmental factors such as locomotion speed, field of view, scene complexity, and system latency shape the sensory stimuli presented to the user and define the conditions under which perceptual inconsistencies may arise. These factors do not just induce physiological responses but also modulate the degree of sensory conflict and cognitive load processed by the CNS, which in turn drives autonomic activation. From this point of view, physiological signals should be interpreted as indirect observable consequences of underlying perceptual and neural processes. A comprehensive understanding of cybersickness therefore requires a joint modelling approach that integrates both the **causal factors on the VR side** and the resulting physiological responses, capturing the full chain from environmental stimulus to embodied reaction.

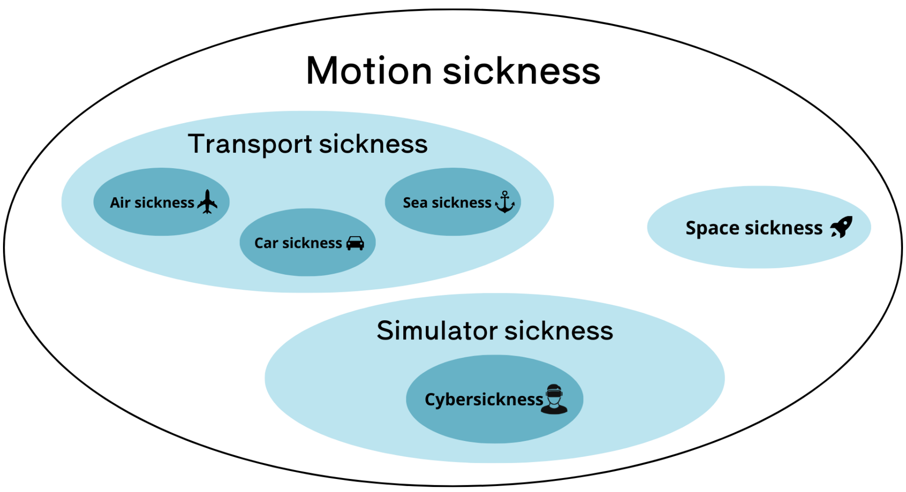


Figure 1: Taxonomic classification of motion sickness, including transport-induced forms such as sea, air, and automobile sickness, space motion sickness which is mechanistically distinct from terrestrial transport, and simulator-induced syndromes, with cybersickness as the focus of the present study.

2.2. Subjective Metrics and Psychometric Evaluation

Subjective evaluation remains the foundation for assessing cybersickness, as it directly reflects the user's conscious perception of discomfort and its qualitative characteristics. These measures are designed to capture how symptoms are experienced, integrating physical sensations, visual strain, and perceptual disorientation into a unified self-report. In practice, they take the form of **multi-dimensional questionnaires** or **single-item rating scales**, each targeting different aspects of the phenomenon [23, 26]. The most widely adopted instrument is the Simulator Sickness Questionnaire (SSQ), which aims to characterise the overall symptoms profile after the exposure. It decomposes the experience into three domains that reflect distinct but related processes. Nausea captures autonomic and gastrointestinal responses such as sweating and stomach awareness. Oculomotor symptoms reflect visual fatigue and strain associated with prolonged visual engagement. Disorientation captures vestibular and spatial disturbances such as dizziness and vertigo. By aggregating these components into a total severity score, the SSQ provided a comprehensive summary of the user's condition. However, because it is administered only after the experience, it inherently treats cybersickness as a static outcome and does not capture how symptoms evolve over time. To address this limitation, more temporally sensitive tools such as the Fast Motion Sickness Scale (FMS) have been introduced. The FMS is designed to track the **moment-to-moment evolution** of discomfort, allowing users to report their current state during immersion on a scale from 0, corresponding to no symptoms, to 20, corresponding to sever sickness. This enables the observation of both the initial onset of symptoms and their subsequent progression, offering a more dynamic view of the experience. Despite their relevance, subjective measures present several limitations that constrain their use in real-time systems. Reporting bias may arise when users consciously or unconsciously alter their responses, for instance to prolong or terminate an experience. In addition, there is often a delay between the physiological onset of discomfort and its conscious perception, which limits temporal precision. Frequent reporting may disrupt immersion and confound the variables under observation. Taken together, these limitations highlight that subjective reports, while essential for capturing perceived experience, provide only a partial view of cybersickness. This motivates the integration of objective, sensor-based measurements, enabling a more continuous and time-resolved characterization of the underlying physiological processes.

2.3. Predictive Models for Cybersickness

While subjective reports provide direct access to perceived discomfort and physiological signals offer continuous, objective observations of the underlying processes, neither modality alone is sufficient to fully characterise or anticipate cybersickness. Subjective measures are temporally sparse and prone to bias, whereas physiological signals are continuous but indirect and often noisy. This complementarity motivates the adoption of **predictive modelling** where both sources of information can be integrated to infer the user's state over time. In supervised settings, predictive models learn the relationship between input features and target outcomes from labelled data, where labels are typically derived from subjective assessments and feature are extracted from physiological and environmental signals. In this context, the task is not only to detect the presence of discomfort but also to capture its temporal evolution and underlying drives. Different model families encode distinct assumptions and inductive biases, which influence their ability to represent non-linear dynamics, handle noisy bio-signals and generalise across high-dimensional VR telemetry. These differences become particularly relevant when modelling cybersickness, which emerges from complex interactions between perception, physiology, and environment over time. The following section provides a concise overview of widely adopted model families, outlining their underlying learning principles and summarising typical strengths and limitations in supervised predictive settings.

2.3.1 Logistic Regression

Logistic Regression (LR) is a **linear probabilistic classifier** that models the likelihood of a binary outcome. By applying the logistic function to a linear combination of input features, LR produces probabilities in the [0, 1] interval, enabling class assignment based on posterior estimates [10]. Parameters are typically estimated through maximum likelihood, often with L1 or L2 regularization to mitigate overfitting in high-dimensional spaces [10]. While LR offers transparency in feature contributions through interpretable coefficients, it is limited in modelling complex, non-linear interactions inherent in multifaceted datasets, as it assumes a linear relationship between the independent variables and the log-odds of the outcome [10].

2.3.2 Generalized Additive Models

Generalized Additive Models (GAMs) extend the Generalized Linear Model (GLM) framework by replacing linear coefficients with smooth, non-parametric functions of the predictors [14, 49]. The central assumption of a

GAM is additivity, where the expected response is modelled as a sum of univariate functions of each predictor:

$$g\big(E[Y]\big) = \beta_0 + f_1(X_1) + f_2(X_2) + \cdots + f_p(X_p), \tag{1}$$

where $g(\cdot)$ is a link function connecting the expected value $E[Y]$ to the additive predictor, β_0 is the intercept, and $f_j(\cdot)$ are smooth functions capturing the non-linear effects of each predictor X_j [14, 49].
In the simple Gaussian case with an identity link, the model reduces to:

$$Y = \beta_0 + f_1(X_1) + f_2(X_2) + \cdots + f_p(X_p) + \varepsilon, \tag{2}$$

where $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ represents the residual error [15].
The additivity assumption allows the contribution of each predictor to be **isolated**, providing both flexibility and interpretability [14, 29]. Each smooth function $f_j(X_j)$ can be visualized independently to examine the specific relationship between predictor and response. Estimation of the smooth functions is typically performed using penalized regression splines [48, 49]:

$$\hat{f}_j = \arg\min_{f_j} \sum_{i=1}^{n} \Big(y_i - \beta_0 - \sum_{j=1}^{p} f_j(x_{ij})\Big)^2 + \lambda_j \int \big(f_j''(x)\big)^2 dx, \tag{3}$$

where the second term imposes a smoothness penalty controlled by λ_j, balancing model fit and function smoothness.
While GAMs provide **interpretability and flexibility**, they are limited in representing higher-order interactions between predictors unless interaction terms are explicitly included [14, 49]:

$$g\big(E[Y]\big) = \beta_0 + \sum_{j=1}^{p} f_j(X_j) + \sum_{k<l} f_{kl}(X_k, X_l), \tag{4}$$

where $f_{kl}(\cdot, \cdot)$ captures pairwise interactions.
This formulation demonstrates why GAMs are particularly valuable in domains that require accurate predictions alongside clear, **feature-specific interpretability**. [15, 29, 49].

2.3.3 Random Forest

Random Forest (RF) is an ensemble learning method based on the principle of bagging, designed to build a robust predictive model from a **collection of de-correlated decision trees** [6]. Each tree is grown on a different bootstrap sample, which is a subset of the training data obtained through random sampling with replacement. To reduce correlation between trees, the algorithm evaluates only a random subset of features at each node split. The final prediction is obtained by aggregating the outputs of all trees through majority voting for classification or averaging for regression [6]. The strength of Random Forest lies in its ability to **reduce model variance without substantially increasing bias**, addressing the instability often seen in individual decision trees. This makes the method resilient to high-dimensional noise and outliers. RF models also provide **intrinsic measures of variable importance**, which quantify the contribution of each predictor to the reduction of node impurity across the forest. Despite these advantages, the aggregation of many trees reduces transparency, as the combined decision logic is not immediately interpretable and requires post-hoc analysis. In cases of severe class imbalance, the standard bootstrap process can bias predictions toward the majority class, requiring cost-sensitive learning or specialized resampling strategies [8].

2.3.4 Balanced Random Forest

Balanced Random Forest (BRF) is an extension of the Random Forest method designed to **address extreme class imbalance** [17]. In standard ensembles, bootstrap samples are drawn uniformly from the dataset, so if one class dominates, the trees tend to bias toward the majority class. BRF mitigates this by modifying the sampling process. For each tree, the algorithm downsamples the majority class or upsamples the minority class, ensuring that each learner is trained on a balanced subset of the data [8, 17]. The strength of BRF is its ability to **improve recall for the minority class without causing a large loss of precision** that often occurs with global resampling. By balancing data at the level of individual trees rather than the whole forest, the method preserves ensemble diversity while forcing each split to focus on features that distinguish the minority class. This makes BRF particularly effective for detecting rare but important events. However, the balanced approach can increase false positives because the model is more sensitive to patterns that are infrequent in the overall population. BRF also retains the variable importance measures of standard Random Forest, but these reflect a balanced distribution rather than the true class distribution.

2.3.5 Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) is a scalable and highly optimized implementation of the gradient boosted decision tree (GBDT) paradigm [9]. Theoretically, the model follows an additive functional approach, where a strong learner is constructed by sequentially integrating shallow decision trees. In each iteration, a new tree is trained to approximate the negative gradient of a differentiable loss function, thereby minimizing the residual error of the preceding ensemble [9]. Unlike traditional boosting, XGBoost incorporates a formal regularization term (L_1 and L_2) directly into the objective function to penalize model complexity and prevent overfitting. The architectural strength of XGBoost lies in its ability to **handle sparse data and heterogeneous feature sets efficiently through a sparsity-aware split-finding algorithm**. Mathematically, it utilizes second-order Taylor expansion to approximate the loss function, providing more precise information regarding the direction of the gradient compared to first-order methods [9]. However, the model's sensitivity to hyper-parameter configurations, such as the learning rate and tree depth, requires rigorous validation, particularly when the input data contains high-frequency noise. Furthermore, while it provides robust feature importance scores, the sequential nature of the boosting process renders the final ensemble a black-box necessitating specialized attribution techniques for structural interpretability.

2.3.6 1D Convolutional Neural Networks

One-dimensional Convolutional Neural Networks (1D CNNs) are a class of deep learning architectures designed to **process sequential data by performing sliding-window convolutions across the temporal dimension** [25]. Theoretically, these models utilize a set of learnable kernels to perform feature extraction, effectively acting as automated pattern recognizers that identify local stationarity and translation-invariant features within a signal. By stacking multiple convolutional layers, the network can develop a hierarchical representation of the data, moving from simple edge-like transients in lower layers to complex structural motifs in deeper layers [25]. The primary advantage of 1D CNNs is their ability to capture local dependencies and high-frequency variations with significantly fewer parameters than fully connected networks, thanks to the principle of weight sharing. This makes them highly efficient for detecting rapid, transient deviations within high-resolution time-series. However, because the receptive field of a standard convolution is restricted by the kernel size, these models are inherently limited in their capacity to capture long-range global dependencies [25]. To address this, they are often integrated with pooling layers to down-sample features or combined with recurrent structures to maintain a broader temporal context.

2.3.7 Long Short-Term Memory Networks

Long Short-Term Memory (LSTM) networks are an advanced variant of Recurrent Neural Networks (RNNs) specifically engineered to overcome the vanishing gradient problem, which typically prevents standard recurrent models from learning long-term dependencies [16]. The core of the LSTM architecture is the **cell state**, a continuous highway that transports information across time steps, and a series of sophisticated gating mechanisms [16]. These gates (input, forget, and output) utilize sigmoid activation functions to regulate the flow of information, enabling the network to selectively retain, discard, or update its internal memory based on the relevance of incoming data. Theoretically, LSTMs are uniquely **suited for phenomena that are inherently cumulative or where the current state is heavily contingent on a history of prior exposures**. By maintaining a hidden state that serves as a dynamic memory, the model can identify patterns in data that manifest over extended temporal horizons [16]. While LSTMs offer superior expressive power for sequential modelling, their complex internal structure increases the computational cost of training and requires substantial datasets to avoid overfitting. Furthermore, as with other deep architectures, the latent representations within the gates are not directly interpretable, often requiring post-hoc attribution methods to understand the factors driving the network's state transitions.

2.4. Nested Cross-Validation

In supervised learning, the estimation of model performance must reflect the ability to generalise to unseen data. This is typically achieved through Cross-Validation (CV), a resampling strategy in which the dataset is partitioned into multiple folds, allowing each subset to be used for both training and evaluation in a rotating manner. Under the assumption of a fixed model configuration, CV provides an effective and widely adopted estimate of predictive performance [4]. However, this assumption no longer holds when model selection is introduced. In practical settings, models are rarely evaluated in isolation, but rather after an optimisation process in which hyper-parameters are tuned to maximise performance. If the same data partitions are used both to select the optimal configuration and to estimate performance, the evaluation becomes biased. This occurs because the model is indirectly exposed to the evaluation data during the tuning process, leading to overly optimistic estimates and reduced reproducibility [45]. To preserve the statistical integrity of the evaluation, Nested Cross-Validation is employed as a principled evaluation framework that separates model selection from performance estimation [7]. The approach is structured as a two-level resampling procedure. An outer loop is used to estimate generalisation performance by splitting the dataset into training and test folds. Within each outer training fold, an inner cross-validation loop is performed to identify the optimal hyper-parameter configuration. The selected model is then re-trained on the entire outer training set and evaluated on the corresponding outer test fold, which remains completely independent from the optimisation process. This hierarchical structure ensures that the evaluation protocol faithfully represents the performance of the full modelling pipeline, including hyper-parameter tuning. By preventing any information from the test folds from influencing model selection, Nested Cross-Validation effectively mitigates information leakage, which is particularly critical in high-dimensional and noisy datasets. The final performance estimate is obtained by aggregating the results across all outer folds, yielding a robust and unbiased assessment of how the model is expected to perform in real-world scenarios [4, 7, 45].

2.5. Evaluation Metrics

The quantitative assessment of predictive models depends on both the nature of the learning task and the statistical characteristics of the data. In supervised learning, evaluation metrics provide a principled framework to quantify how accurately a model captures the relationship between input features and the target variable. These metrics differ according to whether the objective is classification or regression, reflecting distinct notions of error, uncertainty, and predictive performance.

2.5.1 Classification Performance

For binary and multi-class classification tasks, model predictions can be summarised through a **confusion matrix**, which decomposes outcomes into True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). This representation provides the basis for deriving a range of performance metrics, each capturing a different aspect of predictive behaviour.
Accuracy measures the proportion of correctly classified samples over the total number of observations:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

Despite its widespread use, it can provide misleading estimates in the presence of class imbalance, as it does not account for the distribution of errors across classes.
Balanced Accuracy addresses class imbalance by averaging recall across classes, ensuring that performance on both positive and negative samples is equally weighted:

$$\text{Balanced Accuracy} = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right) \tag{6}$$

Sensitivity and Specificity, also referred to as recall, quantifies the proportion of correctly identified positive instances. Specificity measures the proportion of correctly identified negative instances:

$$\text{Sensitivity} = \frac{TP}{TP + FN}, \quad \text{Specificity} = \frac{TN}{TN + FP} \tag{7}$$

Precision and F1-score evaluates the reliability of positive predictions by measuring how many predicted positives are correct. The F1-score combines precision and recall into a single measure through their harmonic mean:

$$F1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{8}$$

Additional metrics such as ROC-AUC quantify the trade-off between true positive and false positive rates across decision thresholds, while PR-AUC provides a more informative evaluation in imbalanced settings by focusing on the precision–recall relationship.

2.5.2 Regression Performance

For continuous targets, model performance is assessed by comparing predicted values ($\hat{y}$) with observed outcomes (y), typically through residual-based metrics that quantify different aspects of prediction error.

Mean Absolute Error (MAE) measures the average magnitude of deviations between predictions and observations:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{9}$$

It provides a direct and interpretable measure of error that is robust to outliers.
Mean Squared Error (MSE) computes the average of squared residuals:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{10}$$

By squaring the errors, it penalises larger deviations more strongly, making it sensitive to outliers.
Root Mean Squared Error (RMSE) is defined as the square root of the MSE:

$$\text{RMSE} = \sqrt{\text{MSE}} \tag{11}$$

It expresses the error in the same units as the target variable while retaining sensitivity to large deviations.
Coefficient of Determination (R^2) quantifies the proportion of variance in the target explained by the model:

$$R^2 = 1 - \frac{\sum_i(y_i - \hat{y}_i)^2}{\sum_i(y_i - \bar{y})^2} \tag{12}$$

It provides a normalised measure of fit relative to a baseline model that predicts the mean.

2.6. Explainability and Feature Attribution

In predictive modelling, it is increasingly important to understand how input features influence model outputs. Explainability provides a framework to assess these contributions, allowing the behaviour of a model to be interpreted, validated, and trusted. Feature attribution techniques enable the quantification of the impact of each input on the predictions, offering insights into both local effects for individual instances and global patterns across the dataset. The following sections describe two widely adopted approaches that exemplify complementary strategies for feature assessment.

2.6.1 Permutation Feature Importance (PFI)

Permutation Feature Importance is a model-agnostic technique that **quantifies the global contribution of a variable by observing the degradation in predictive accuracy when its information is removed**[2, 41]. Unlike intrinsic importance metrics, which are often tied to specific architectural properties like tree-splits, PFI evaluates the model's objective reliance on a feature.The procedure involves randomly shuffling the values of a single feature j within the test set, thereby breaking its association with the target variable y while preserving the underlying distribution of the other predictors [2, 41]. If the model's error increases significantly after the permutation, the feature is deemed highly influential. Mathematically, the importance I of feature j is expressed as the difference between the baseline error e_{orig} and the error on the permuted data e_{perm}:

$$I_j = e_{perm} - e_{orig} \tag{13}$$

A key advantage of PFI is its ability to provide a **model-blind ranking of features**. However, it assumes feature independence; in the presence of strong multicollinearity, the importance of a single feature may be underestimated if the model can still extract redundant information from its correlated counterparts [2].

2.6.2 SHapley Additive Explanations (SHAP)

SHapley Additive Explanations (SHAP) is a unified framework for interpreting model predictions based on cooperative game theory [42]. Each feature is treated as a player in a coalition, where the model's prediction represents the payout. The goal is to fairly distribute this payout among features according to their marginal contribution to the final output. For a specific instance x, the SHAP value ϕ_j of feature j is defined as

$$\phi_j = \sum_{S \subseteq \{x_1,\ldots,x_p\}\setminus\{x_j\}} \frac{|S|!(p-|S|-1)!}{p!}\big[f(S\cup\{x_j\}) - f(S)\big], \tag{14}$$

where p is the total number of features, S is a subset of features not containing x_j, and $f(S)$ is the expected model output given the features in subset S.

Unlike Permutation Feature Importance, which provides a single global ranking, SHAP satisfies the property of local accuracy, meaning that for any instance x, the sum of all SHAP values equals the difference between the model's actual prediction and the base value (mean prediction):

$$f(x) = \phi_0 + \sum_{j=1}^{p} \phi_j. \tag{15}$$

This property enables both granular and instance-level explanations, allowing the contribution of each feature to be interpreted in the context of a specific prediction.

3. State of the Art in Cybersickness Detection

Recent research on cybersickness has shifted from static, post-hoc assessments toward real-time monitoring systems, driven by the need to detect the precise onset of symptoms and enable proactive interventions before discomfort escalates. Physiological monitoring has become the cornerstone of objective assessment, with particular emphasis on autonomic nervous system responses. Early work relied on complex laboratory-grade setups, whereas more recent studies favour consumer-grade wearable sensors, which facilitate experiments that better reflect real-world VR use and increase accessibility [24, 36]. Heart rate (HR) and heart rate variability (HRV) metrics derived from ECG, when combined with electrodermal activity, photoplethysmography (PPG), VR telemetry, and eye-tracking features, can improve predictive robustness relative to single-sensor approaches, particularly when the modalities capture complementary physiological and behavioural processes. Datasets such as VR.net [47], which combine rich VR experiences with physiological recordings, represent a valuable step toward understanding the environmental contributors to cybersickness, even though these data are not publicly available. Computational approaches have evolved from classical ensemble methods such as Random Forest and XGBoost toward deep learning architectures including LSTM and CNN models. Hybrid CNN-LSTM networks are particularly effective at capturing both local signal features and longer-term temporal dynamics. Several studies illustrate both the potential and limitations of combining wearable sensors with deep learning. Islam et al. (2020) employed HR, HRV, galvanic skin response, and breathing rate with an LSTM model, but the electrode-based setup restricted hand movement and subjective SSQ feedback did not consistently align with physiological measurements [20]. In a subsequent study, Islam et al. (2021) integrated stereoscopic video, eye-tracking, and movement data with a CNN-LSTM model, achieving 52 percent accuracy with video alone and 87 percent when adding PPG and EDA, although class imbalance constrained performance [21]. Similarly, Garcia-Agundez et al. (2019) combined ECG, EOG, skin conductance, and respiratory signals with classical classifiers including SVM, kNN, and neural networks, reaching 82 percent accuracy for binary detection [11]. More recently, Wang et al. (2023) demonstrated real-time prediction in VR games by combining eye movements with in-game character trajectories using an LSTM model, achieving 83.4 percent accuracy [46]. A major emerging direction is the development of explainable and adaptive systems. Feature attribution techniques, such as SHAP, are increasingly used to link physiological fluctuations to specific environmental triggers, including optical flow or rapid rotations, providing insight into the mechanisms underlying predicted discomfort. Several research prototypes are being developed to adapt the VR environment in real time, for example by reducing the field of view or inserting rest frames to maintain user comfort. These prototypes, however, remain limited in both scope and architectural sophistication. Many focus solely on physiological sensing while neglecting environmental parameters, which reduces their ability to disentangle the causes of observed user responses [19]. Several implementations rely on medical-grade devices such as the *Empatica* wristband, which, although offering high-fidelity measurements, are not widely accessible and lack continuous streaming necessary for real-time adaptation [19]. Smartwatch-grade sensors offer a more practical and consumer-accessible alternative, enabling scalable deployment and continuous monitoring. Beyond sensor choice, most prototypes are not modular and cannot easily integrate additional physiological modalities or environmental variables without substantial changes to the system [12]. This limits their extensibility and potential for integrated adaptive interventions. Collectively, these constraints underscore the need for fully modular frameworks that combine physiological and environmental sensing with explainable modelling and adaptive mitigation. In this context, the ability to explain model decisions is essential because feature attribution methods can quantify the contribution of individual inputs, providing interpretable insights into the mechanisms driving user responses and supporting targeted, data-informed adaptation strategies.

4. Modular Taxonomy for Cybersickness Analysis

Understanding cybersickness requires not only measuring user responses but also systematically characterising the virtual context in which it occurs and the design elements that can trigger discomfort. The lack of a unified framework limits reproducibility, complicates cross-study comparisons, and constrains designers' ability to anticipate or mitigate symptoms effectively. To address this gap, a multi-layered taxonomy of VR design factors associated with cybersickness is proposed, grounded in a structured review of prior research [3, 5, 38, 39]. This framework synthesizes design dimensions repeatedly linked to cybersickness and organises VR experiences across locomotion, visual appearance, interaction design, and mitigation strategies, providing a structured basis both for controlled experimentation and for mapping design choices to outcomes. The first layer, locomotion design, defines how movement is generated and controlled, including the movement paradigm (body- or vehicle-centric), type (physical, simulated, or hybrid), control (user-driven or externally guided), degrees of freedom, and navigation parameters such as speed, acceleration, jerk, and turn rate. High cybersickness levels are generally associated with high simulated speeds, abrupt acceleration, elevated jerk, continuous smooth rotation, and externally driven motion, whereas techniques such as teleportation or snap-turning can mitigate discomfort by limiting sustained optic flow and vestibular conflicts [5, 38]. The second layer, **appearance and visual features**, captures perceptual and rendering properties including stable orientation cues, colour and luminance control, field of view (FOV), optical flow distribution, and scene complexity. Large FOV during artificial locomotion, strong peripheral optical flow, unstable luminance, and highly dynamic content increase symptom severity, whereas static visual anchors help reduce discomfort [38]. **Interaction design**, the third layer, encompasses user agency and engagement, including interactivity level, body posture, and input modality. Reduced or externally driven interaction often heightens discomfort due to lower predictability and limited motor feedback, while seated configurations generally improve tolerance during artificial locomotion. Finally, the **mitigation layer** defines strategies aimed at reducing cybersickness, balancing comfort with immersion. Visual adaptations such as motion blur, depth-of-field adjustments, and visual anchors, together with movement-based strategies including teleportation, speed regulation, snap-turning, and acceleration smoothing, can lower sensory conflict, though often at the cost of reduced presence or engagement [3]. This taxonomy supports VR design by guiding the minimisation of sensory conflict and enables controlled, reproducible manipulation of variables for studying physiological and behavioural responses. A compressed visual representation of the taxonomy is shown in Figure 2, while Table 1 summarises the main associations between layer, environmental factor, direction (positive for high cybersickness, negative for low), underlying mechanism (e.g., sensory conflict, postural), and supporting evidence from the literature.

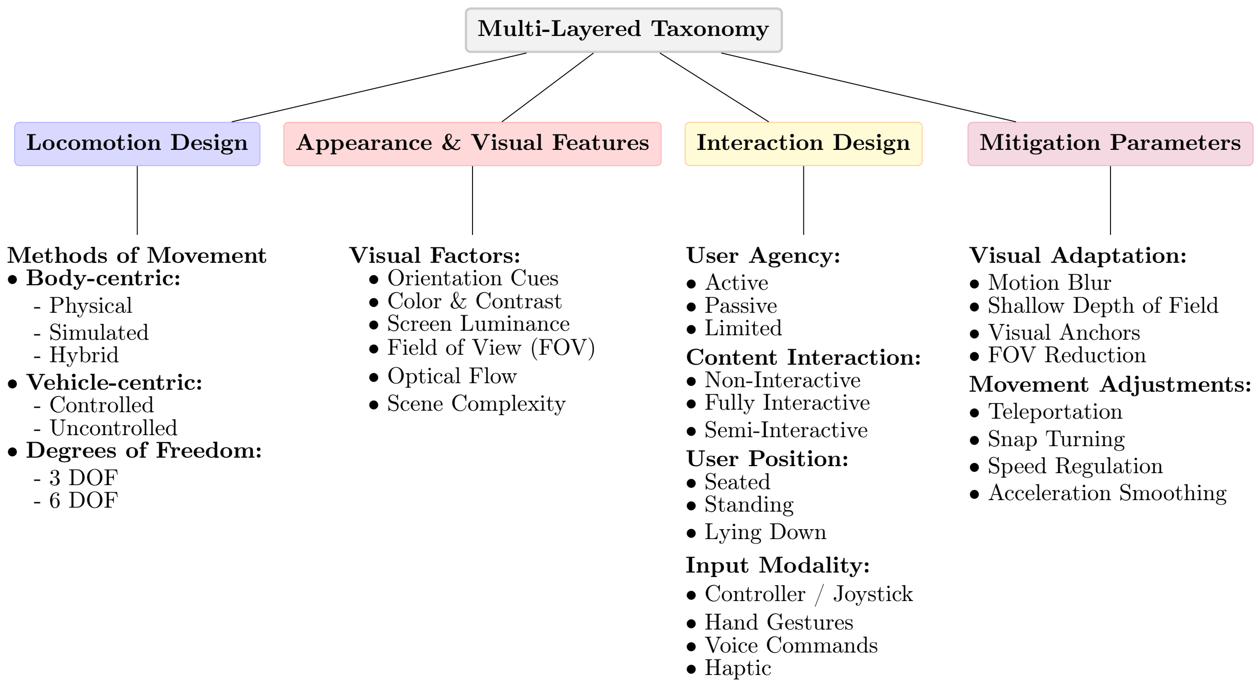


Figure 2: Multi-layered taxonomy of factors influencing user experience in virtual environments. The design space is organised into four primary dimensions: locomotion design, appearance and visual features, interaction design, and mitigation parameters. Key elements within each dimension summarise movement characteristics, perceptual cues, user interaction modalities, and mitigation strategies.

Table 1: Summary of the main associations in the multi-layered VR cybersickness taxonomy. Each row lists: Layer (taxonomic dimension), Factor (environmental element), Direction (positive: high cybersickness; negative: low cybersickness), Mechanism (e.g., sensory conflict, postural), and Evidence from the literature.

Layer	Factor	Direction	Mechanism	Evidence
Locomotion	High simulated speed	Positive	Sensory conflict	Strong
	Abrupt acceleration	Positive	Vestibular mismatch	Strong
	High jerk	Positive	Motion transients	Strong
	Continuous smooth rotation	Positive	Sensory conflict	Strong
	Externally driven motion	Positive	Reduced agency	Moderate
	Teleportation	Negative	Reduced optic flow	Strong
	Snap-turning	Negative	Reduced rotational flow	Strong
Visual	Large FOV (artificial motion)	Positive	Peripheral flow overload	Strong
	Strong peripheral optical flow	Positive	Visual-vestibular conflict	Strong
	Unstable luminance	Positive	Visual instability	Moderate
	High scene complexity	Positive	Cognitive overload	Moderate
	Absence of visual anchors	Positive	Lack of reference frame	Moderate
	Stable horizon / anchors	Negative	Visual stabilization	Moderate
Interaction	Passive experience	Positive	Reduced predictability	Moderate
	Limited user agency	Positive	Motor-sensory mismatch	Moderate
	Seated posture	Negative	Reduced vestibular load	Moderate
	Standing posture	Context-dependent	Postural instability	Emerging
Mitigation	Dynamic FOV reduction	Negative	Flow attenuation	Moderate
	Motion blur	Negative	Reduced visual detail	Emerging
	Depth-of-field modulation	Negative	Peripheral dampening	Emerging
	Speed regulation	Negative	Lower motion intensity	Strong
	Acceleration smoothing	Negative	Reduced motion transients	Emerging

5. Methodology

This section describes the computational framework developed for the analysis and modelling of cybersickness within the proposed multi-layered taxonomy. The methodology integrates synchronized physiological signals, environment-level telemetry, and subjective annotations into a unified processing pipeline. In Subsection 5.1, the datasets used for both exploratory analysis and controlled validation are described, highlighting public sources and the custom environment-linked physiological dataset. Subsection 5.2 introduces the hierarchical modelling framework designed to separately capture cybersickness onset and severity, detailing model selection, training strategy, hyper-parameter optimization, and interpretability techniques. Subsection 5.3 presents the system architecture, including the modular publisher-subscriber design that integrates data acquisition, preprocessing, and AI inference. Subsection 5.4 outlines the validation strategy and experimental protocol, describing the taxonomy-aligned VR environments, data synchronization, and time-resolved subjective ratings. Finally, Subsection 5.5 details feature extraction procedures and the baseline-relative representation used to normalize inter-subject variability and highlight stimulus-induced physiological modulation.

5.1. Dataset Selection

To ensure methodological robustness, comparability with prior work, and principled system design, two independent publicly available datasets were selected and initially employed as analytical instruments to structure the problem space rather than as final validation benchmarks. Their use was exploratory, enabling controlled evaluation of environment-level and physiological predictors while avoiding artificial correlations that could arise from synchronized multimodal acquisition. The Gameplay Dataset *(UFFCSData)* [34] contains 22 environment- and design-related features extracted from virtual racing and flight simulation scenarios, including field-of-view, motion speed, gameplay dynamics, and other navigation variables. The dataset comprises 88 participants, of which 37 include valid symptom annotations. It primarily captures environment-level and locomotion variables and allows analysis of how design configurations influence cybersickness onset and severity independently of physiological signals. Exploratory analysis revealed that the dataset is class imbalanced, with fewer samples at higher severity levels. Player speed tends to increase discomfort, while automatic camera movement and adjustable field-of-view appear to reduce discomfort, acting as comfort factors. Static frame settings are closely linked to camera configurations, suggesting that these features are linearly dependent. Horizontal player position also influences discomfort, indicating that spatial layout affects user experience. The Physiological Dataset *(CyberSicknessClassification)* [18] comprises multimodal recordings from 31 participants exposed to a roller-coaster simulation, including heart rate, heart rate variability, breathing rate, and galvanic skin response. Each signal is represented using four derived metrics. Percentage change from baseline normalizes individual differences, rolling minimum and maximum capture short-term extremes, and a three-second moving average smooths fluctuations while preserving temporal dynamics. Exploratory analysis indicated that heart rate variability decreases as discomfort rises, reflecting the link between low HRV and stress or focus, while galvanic skin response tends to increase with discomfort, reflecting autonomic arousal. Metrics expressed relative to baseline are generally more predictive than raw signals. The dataset is imbalanced, with fewer samples corresponding to higher discomfort levels, and together these features enable detection of both abrupt physiological reactions and sustained autonomic patterns associated with cybersickness.

5.2. Hierarchical Modelling for Detection and Severity Estimation

Exploratory analysis of the public datasets revealed that cybersickness in influenced by multiple interacting physiological and environmental factors and that the distribution of severity levels is inherently imbalanced, with relatively few samples corresponding to higher discomfort states. Feature importance analyses further indicated that the predictors differentiating absence versus presence of symptoms were not identical to those explaining variability among symptomatic states. Certain physiological and environmental variables showed stronger discriminative power for onset detection, whereas others primarily contributed to modelling progressive change in severity. This heterogeneity suggests that a single end-to-end predictive model could mix partially distinct mechanisms, blending the processes that trigger discomfort with those that modulate its intensity. To reflect this structure, a hierarchical two-stage modelling framework was designed. The first stage performs binary onset detection within short time windows, focusing on identifying the emergence of cybersickness. The second stage estimates symptom intensity conditioned on detected onset, modelling gradual physiological responses and VR-environment dynamics associated with symptom progression. This decomposition aims to enhance interpretability by separating triggers from modulators and provides a principled way to mitigate class imbalance, as severity estimation is restricted to symptomatic instances rather than being dominated by the majority non-symptomatic class. Model selection was guided by the need to balance interpretability and flexibility. GAMs, implemented through **Explainable Boosting Machines (EBMs)**, provide transparent additive mappings of feature contributions, while tree-based ensembles such as **XGBoost** and **LightGBM** capture complex interactions and non-linear dependencies expected in psychophysiological responses to immersive environments. Linear models and shallow decision trees were evaluated as baselines, but the combination of GAMs and gradient-boosted trees offered the most favourable trade-off between predictive performance and explanatory clarity across both onset and severity stages. Deep learning approaches were not pursued due to limited dataset size, risk of overfitting, computational demands, and reduced interpretability. To ensure rigorous evaluation, datasets were partitioned at the participant level to prevent data leakage, with most participants allocated to training and inner-loop validation and a held-out subset reserved for final testing. Class imbalance was also tackled through weighting schemes that highlighted underrepresented onset events and more severe levels. Hyper-parameter optimisation was conducted using the **Optuna** framework with nested cross-validation, employing the outer loop for unbiased performance estimation and an inner loop for tuning [1]. The F1-score guided optimisation in the onset detection stage to balance false positives and false negatives, while MAE was used to evaluate severity regression, normalised to the full severity range for comparability. Explainability was assessed through both intrinsic and post-hoc techniques. GAMs allow direct visualization of additive feature effects, whereas tree-based models were analysed using global, local feature importance measures and SHAP values. Global SHAP summaries identity the primary drivers of onset and severity across the dataset, while local SHAP explanations clarify how individual predictors influence specific time windows, supporting potential real-time monitoring and adaptive intervention strategies.

5.3. System Architecture

Figure 3 illustrates the modular, distributed architecture of the system, designed to support both real-time monitoring and offline analysis. Functional components are decoupled and communicate through a publish–subscribe paradigm, enabling independent management of physiological sensing, VR telemetry acquisition, data processing, and predictive modelling. Physiological acquisition and VR telemetry modules act as data publishers, sending time-stamped streams to dedicated MQTT topics [43]. A centralized MQTT broker, implemented using *Eclipse Mosquitto*, mediates message exchange, ensuring reliable delivery and low-latency communication. On the receiving side, a data subscriber collects the relevant streams, performing buffering, synchronisation, and aggregation into fixed-duration windows to maintain temporal alignment. Physiological signals are filtered, assessed for quality, and transformed into window-level features alongside environment variables, which are then forwarded to predictive models for time-resolved cybersickness onset and severity estimation. This design allows sensing and telemetry modules to operate independently of downstream processing, supports seamless integration of additional modalities by publishing to new topics, and enables both real-time monitoring and post-hoc analysis on recorded data, maintaining flexibility, scalability, and modularity across the system.

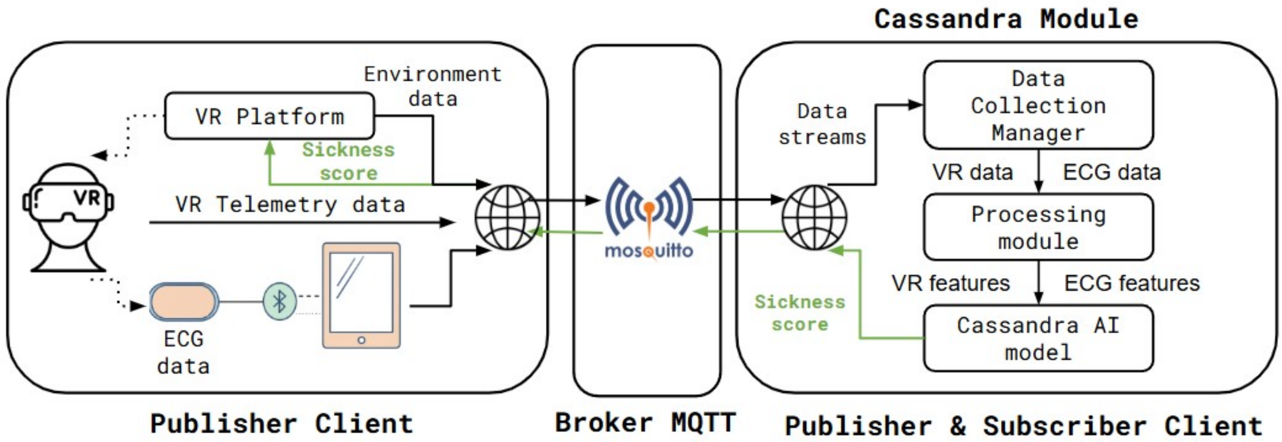


Figure 3: Overview of the proposed modular system architecture and experimental pipeline. ECG and VR/environment telemetry are acquired as time-stamped streams and exchanged via an MQTT-based publish–subscribe infrastructure. Synchronized windowed data support downstream storage ,modelling, and interpretable analysis for time-resolved cybersickness detection and severity estimation.

5.4. Validation Strategy

To complement existing publicly available datasets and address their inherent limitations, a new validation dataset was collected. The two public datasets provide valuable insights into environment-level variables and physiological responses, but they originate from independent experiments and are not temporally synchronized. This restricts the ability to directly link virtual environment characteristics with physiological dynamics and limits causal interpretability. The newly collected dataset overcomes these constraints by integrating controlled VR experiences with synchronized physiological recordings and subjective distress measures, enabling evaluation of short-term responses to specific locomotion, interaction, and visual configurations within the proposed multi-layered taxonomy. Section 5.4.1 presents the experimental protocol, detailing participant recruitment, session design, and overall study flow. Section 5.4.2 describes the VR environments, highlighting how each scenario varies across locomotion, visual, and interaction dimensions. Section 5.4.3 outlines the physiological and telemetry acquisition, including the recording of cardiovascular signals, virtual environment telemetry, and subjective assessments.

5.4.1 Experimental Protocol

The study involved 15 healthy adults, of whom three were female, with ages ranging from 20 to 26 years and one participant over 40. At the start of each session, demographic information and data on visual conditions or impairments were collected to ensure suitability for VR exposure. Each participant wore a chest-worn ECG sensor (*Polar H10*) and a head-mounted display (*Meta Quest 3s*). Sessions comprised four sequential VR exposures, progressing from a baseline condition to increasingly dynamic motion scenarios, with total duration ranging from 15 to 30 minutes depending on individual tolerance and performance. During VR exposure, participants provided repeated subjective ratings of motion-related discomfort using the FMS scale, sampled every 30 seconds, focusing exclusively on motion-related symptoms such as nausea and dizziness. Post-session, participants completed the SSQ for a multidimensional assessment of cybersickness. Continuous ECG data were recorded throughout, and physiological features were computed over fixed-duration windows, temporally aligned with both the active environment condition and corresponding FMS ratings (Figure 4). This design enables analysis of the relationships among locomotion characteristics, progression of subjective discomfort, short-term cardiovascular modulation, and individual task performance.

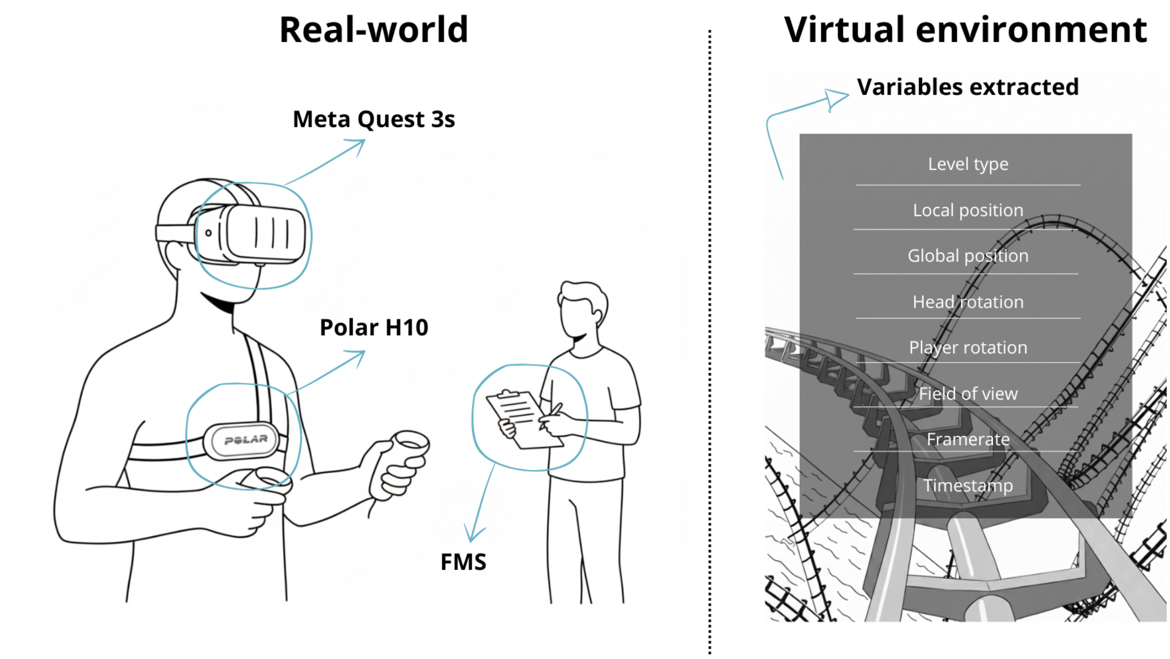
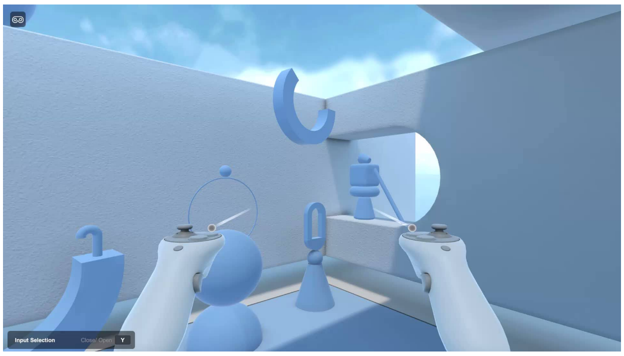


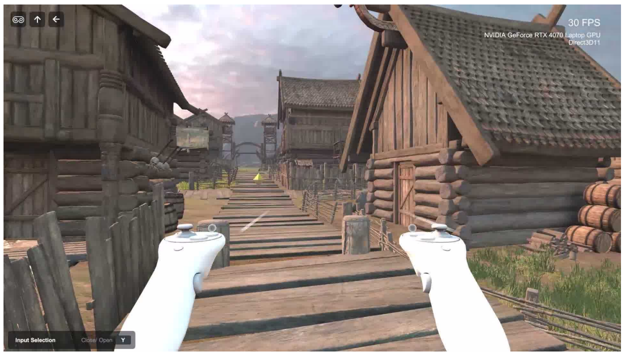
Figure 4: Schematic representation of the validation dataset acquisition. The dataset integrates physiological signals recorded using a Polar H10 chest strap, virtual immersion delivered via Meta Quest 3, and subjective distress assessed through FMS questionnaires. Simultaneously, environment-level and design variables are extracted from the virtual scenes, enabling combined analysis of physiological, self-reported, and contextual factors influencing cybersickness.

5.4.2 VR Environments

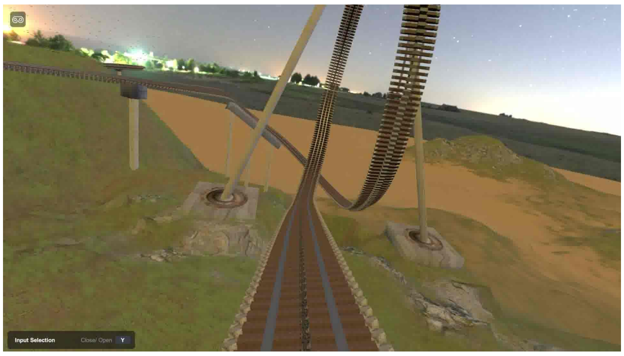
The controlled experiment included four distinct VR experiences, selected to span different regions of the multi-layered taxonomy of cybersickness factors (Figure 2, Table 1) and illustrated in Figure 5. These environments were chosen to systematically vary locomotion intensity, interactivity, optical flow, and scene complexity, enabling evaluation of predictive patterns across heterogeneous VR designs rather than dataset-specific configurations. The **Baseline Level** condition represents a low-intensity configuration, featuring a static frame, body-centric simulated movement, controlled motion, 6-DOF interaction, low speed and acceleration, low scene complexity, adjustable field-of-view, and moderate optical flow. The experience is fully interactive, controller-based, and conducted in a standing posture. This environment serves as a low-risk reference within the taxonomy, capturing minimal sensory conflict and postural challenges. The **Viking Village** scenario maintains body-centric simulated locomotion and 6-DOF interaction while introducing higher speed and acceleration, increased scene complexity, and a fixed field-of-view. Optical flow remains moderate due to controlled navigation, and the experience remains fully interactive. This environment isolates the effect of increased motion intensity and visual complexity while preserving user agency, illustrating how specific factors contribute to cybersickness within the taxonomy. The Rollercoaster Simulation presents a vehicle-centric perspective with automatic movement along a predefined track. Although the user remains physically fixed, the system supports full 6-DOF tracking, allowing head rotations and simulating translations through the virtual environment. The experience is entirely passive and non-interactive, exemplifying strong vestibular-visual conflict in a high-motion, high-optic-flow scenario. Finally, the **360° Rollercoaster Video** shares the automatic, vehicle-centric motion and passive interaction of the previous condition but uses a 3-DOF pre-recorded video [33]. While speed and acceleration remain high, scene complexity is reduced and optical flow coherence is lower compared to the fully rendered simulation. This condition isolates the effect of passive motion with limited interactive rendering dynamics, allowing assessment of cybersickness drivers related specifically to reduced user control and rendering complexity.



(a) Baseline Level.



(b) Viking Village.



(c) Rollercoaster Simulation.

(d) 360° Rollercoaster Video.

Figure 5: Schematic representation of the four VR environments used in the controlled study. The Baseline Level (top-left) provides a low-intensity reference configuration. Viking Village (top-right) increases motion intensity and scene complexity while preserving user agency. Rollercoaster Simulation (bottom-left) combines automatic vehicle-centric motion, high speed, and strong optical flow, representing a high-conflict environment. The 360° Rollercoaster Video (bottom-right) isolates passive motion with reduced rendering dynamics, allowing evaluation of motion-induced discomfort under pre-recorded visual conditions.

5.4.3 Physiological and Telemetry Acquisition

Physiological and environment-level data were recorded in real-time using a dedicated application that interfaces with the chest-worn ECG sensor, with 130 Hz nominal sampling, and streams VR telemetry aligned with the proposed taxonomy. ECG samples were transmitted in batches of 73 samples on average (mean = 72.16, std = 6.85), while the VR environment published telemetry including local and global player positions, head rotation, field of view (FOV), and framerate (FPS) in batches of 10 samples at approximately 1.23 s intervals (mean = 10, std = 0). Incoming physiological and VR samples were aggregated into 3-second sliding windows with 1 s stride, balancing temporal resolution with stability. Timestamp analysis revealed a mean inter-batch interval of approximately 0.554 s for ECG and 1.230 s for VR telemetry, corresponding to effective sampling frequencies of approximately 133 Hz and 8 Hz, respectively. Feature extraction was performed at each stride, yielding an effective feature output frequency of approximately 1 Hz. On average, each window contains approximately 400 ECG samples and 24 VR samples. Due to window overlap, individual samples contribute to multiple feature computations. For VR motion, both linear and angular descriptors were derived from positional and rotational data. Linear motion features were computed from position's data within each window. Net displacement was defined as the difference between the player's position at the end and at the beginning of the window. Mean speed was obtained by diving the magnitude of this displacement by the window duration. Instantaneous speed was defined as the magnitude of the instantaneous velocity vector. Net acceleration was estimated as the change in velocity between at the start and end of the window divided by the same interval, while instantaneous acceleration was computed from consecutive velocity differences. In addition, root-mean-square (RMS) acceleration and positional variance were calculated to characterise motion intensity and variability. Rotational motion metrics were derived from head orientation quaternions. The net camera rotation over each window was computed as:

$$\mathbf{R}_{\text{net}} = \mathbf{R}_{\text{final}}\mathbf{R}_{\text{initial}}^{-1} \tag{16}$$

The resulting quaternion was converted into a rotation vector representation, whose magnitude corresponds to the angular displacement and whose direction defines the rotation axis. Frame-to-frame quaternion differences were further used to estimate instantaneous angular velocity, from which RMS angular velocity and rotational variability metrics were extracted. Additional contextual features included the mean FOV and mean FPS within each window, capturing rendering performance and visual conditions.

Raw ECG signals were processed through a structured pipeline to preserve short-term dynamics while ensuring signal quality. **Low-frequency drift** caused by respiration and posture changes was removed using a **causal band-pass filter** that preserves QRS morphology [13, 22]. **Power-line interference** at 50/60 Hz was attenuated through a **causal notch filter** to prevent contamination of derived features such as HR and HRV [22]. **Baseline wander** induced by motion, breathing, or postural shifts was mitigated to ensure that autonomic dynamics remained intact. Residual high-frequency artifacts were further suppressed using **wavelet-based denoising**, retaining transient cardiac components such as **R-peaks** and **T-wave variations** [22]. R-peaks were then identified in real time with an **adaptive Pan–Tompkins detector**, which accounts for inter-subject variability and amplitude changes across the session, generating reliable **RR intervals** for downstream computation of HR and HRV descriptors [13, 22]. Signal quality was assessed for each window using metrics such as **signal-to-noise ratio (SNR)**, the proportion of physiologically valid **RR intervals**, and **artifact fraction** [32]. Windows failing predefined thresholds were excluded from further modelling. For clarity and completeness, all extracted metrics are summarised in Table 2, where features are grouped by modality and reported with their corresponding aggregation strategy.

Table 2: Summary of extracted features from physiological and VR data.

Category	Feature Group	Features
Physiological	Time-Domain	RMS, RMA, max, min, mean, std, percentage change
	Heart Rate	HR, HR_max, HR_min, HR_std, mean_RR
	HRV	RMSSD, SDNN, pNN50
	Quality Indicators	SNR, signal quality index, valid RR percentage
	Baseline-Relative Metrics	ΔHR, ΔRMS, ΔRMSSD
Environment	Window-Level Kinematics	NetPlayerSpeed, NetAcceleration, NetCameraRotation
	Instantaneous-Derived	MeanInstantSpeed, RMSAcceleration, AngularVelocities
	Spatial Variability	PositionVariance, RotationVariance
	System-Level	MeanFPS, MeanFOV
	Static Context Variables	CameraAutoMovement, StaticFrameCondition

Static variables are transmitted once per session as contextual descriptors. To account for inter-subject variability, physiological features are expressed relative to the baseline VR level:

$$\Delta f_t = f_t - f_{\text{baseline}} \tag{17}$$

This baseline-relative representation improves comparability, highlights stimulus-induced physiological shifts, and enhances sensitivity to cybersickness onset [22]. Both absolute and delta features are supplied to predictive models per window, preserving temporal resolution and supporting hierarchical modelling of onset and severity.

6. Results

This section presents the empirical evaluation of the proposed two-stage framework for cybersickness prediction, with results reported separately for onset detection (binary classification) and severity estimation (regression). Section 6.1 details the predictive performance of the models, including metrics for both tasks, while Section 6.2 focuses on interpretability, analysing feature contributions and their relative importance.

6.1. Predictive Performance

For each task, models were trained independently on the two features domains described in Section 5.1, namely physiological features and VR-environment features. The two domains are intentionally evaluated separately, as they represent distinct information sources. For each feature domain and prediction task, model selection and hyper-parameter optimization were performed using the nested cross-validation procedure. The optimal model configuration was determined based solely on the inner–outer validation process conducted on the training portion of the data. Final performance results are reported on the held-out test set comprising of the participants, which was never used during model selection or hyper-parameter tuning. This strict separation ensures that the reported metrics provide an unbiased estimate of the model's generalization capability to unseen subjects and VR conditions.

6.1.1 Physiological-Based Prediction

Model performance using physiological features is reported in Table 3 for both the public and validation datasets. Three model families were evaluated: EBM, XGBoost, and LightGBM. Each model was assessed in a single-layer configuration, directly predicting severity via regression, as well as in the proposed two-stage hierarchical architecture, in which onset detection and severity estimation are decoupled. The single-layer MAE represents the error when predicting severity without distinguishing onset from intensity. In the two-stage architecture, onset detection is evaluated using accuracy and F1-score, while severity estimation is assessed via MAE. Across both datasets, LightGBM delivers the strongest overall performance, achieving an onset detection accuracy of **0.9577** and F1-score of **0.9800**, with a severity MAE of **0.0548** on the validation set. Its confusion matrix is shown in Figure 6. XGBoost performs slightly below LightGBM, while EBM offers competitive results along with the advantage of intrinsic interpretability. Validation performance is lower than on the public dataset, likely due to differences in experimental conditions, participant variability, and signal quality. Offline processing in the public dataset allowed cleaner filtering and artifact removal, while real-time acquisition in the validation dataset introduced additional noise and minor data loss.These factors increase heterogeneity and challenge generalization, emphasizing the need for robust features and hierarchical modelling. The consistency of performance across the public and validation datasets further supports the generalizability of physiological-based prediction models.

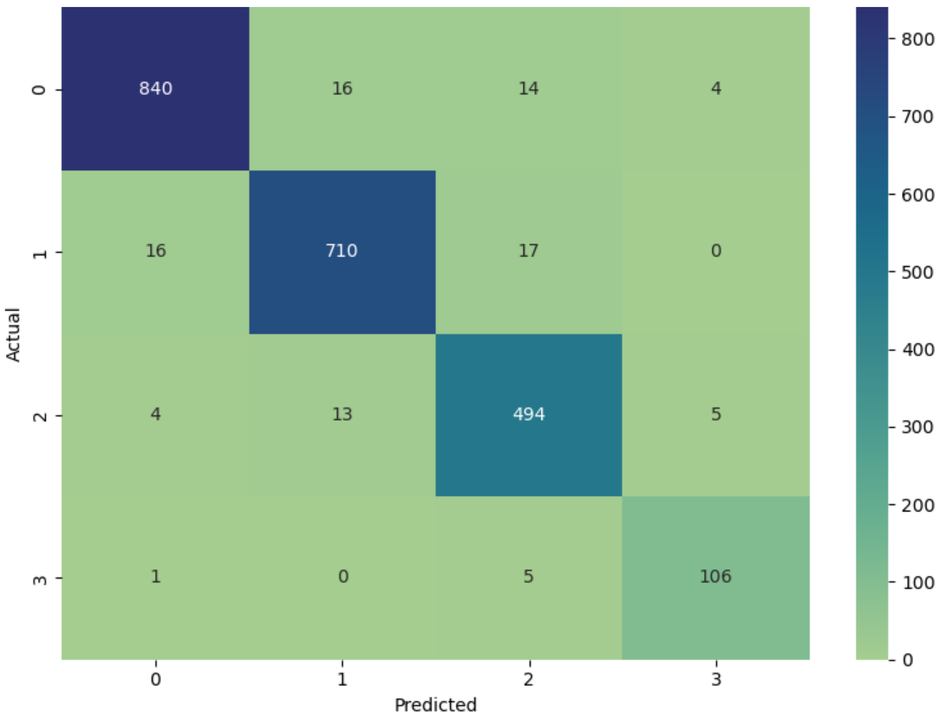


Figure 6: The confusion matrix for the LightGBM hierarchical model on the physiological dataset shows a pronounced diagonal, indicating high predictive accuracy across all four sickness levels (0–3). Off-diagonal entries are minimal, with misclassifications mostly occurring between neighbouring severity classes, underscoring the model's ability to reliably capture gradual increases in cybersickness intensity.

Model	Single Layer	Two-Layer Architecture	
	MAE	**Binary (Acc. / F1)**	**Regression (MAE)**
Public dataset			
EBM	0.275	0.904/0.921	0.0899
XGBoost	0.204	0.911 / 0.939	0.0578
LightGBM	**0.189**	**0.947/0.973**	**0.0538**
Validation dataset			
EBM	0.298	0.9112 /0.9223	0.1045
XGBoost	0.235	0.9470 / 0.9758	0.0668
LightGBM	**0.211**	**0.9577 / 0.9800**	**0.0548**

Table 3: Evaluation of model architectures using physiological. Metrics include Mean Absolute Error (MAE) for single-layer and regression tasks, alongside Accuracy and F1-score for two-layer binary classification.

6.1.2 VR Environment-Based Prediction

Model performance using VR-environment features is reported in Table 4 for both the public and validation datasets. The same model families as in the physiological domain were evaluated: EBM, XGBoost, and LightGBM. Models were trained independently on environment-level features derived from locomotion, visual complexity, camera movement, and other environmental parameters described in Section 5.2.2. Single-layer regression models predict severity directly, while the two-stage hierarchical architecture separates onset detection and severity estimation. In the two-stage framework, onset detection is evaluated using accuracy and F1-score, whereas severity estimation is assessed with normalized MAE, scaled to the [0,1] range for comparability across datasets and feature domains. LightGBM consistently achieves the best performance, with onset detection accuracy of **0.8995** and F1-score of **0.911** on the public dataset, alongside a severity MAE of **0.0781** on the validation set. Its confusion matrix is shown in Figure 7. XGBoost and EBM follow closely, though they show higher error rates in the single-layer configuration. As in the previous scenario, performance on the validation dataset is lower than on the public dataset, primarily due to greater variability in environmental conditions, differences in scene complexity, and the impact of online versus offline data acquisition on temporal alignment and signal quality. While environment-level features provide useful predictive information, their performance is generally lower than that of physiological datasets. Physiological signals capture objective internal responses to distress, whereas VR-environment features reflect external stimuli, which are influenced by individual sensitivity and may not directly indicate the user's physical state.

Model	Single Layer	Two-Layer Architecture	
	MAE	**Binary (Acc. / F1)**	**Regression (MAE)**
Public dataset			
EBM	0.430	0.917 / 0.898	0.1744
XGBoost	0.554	0.9132 / 0.9500	0.1337
LightGBM	**0.397**	**0.9457 / 0.9694**	**0.0847**
Validation dataset			
EBM	0.312	0.829/0.854	0.0912
XGBoost	0.259	0.882/0.868	0.0805
LightGBM	**0.242**	**0.8995/0.911**	**0.0781**

Table 4: Evaluation of model architectures using VR. Metrics include Mean Absolute Error (MAE) for single-layer and regression tasks, alongside Accuracy and F1-score for two-layer binary classification.

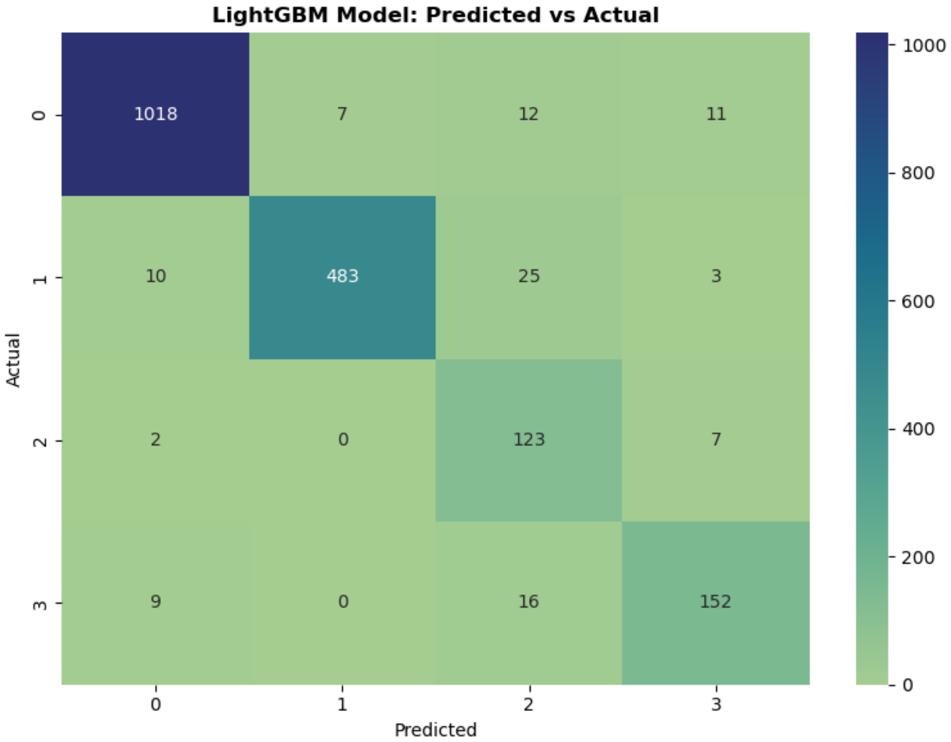
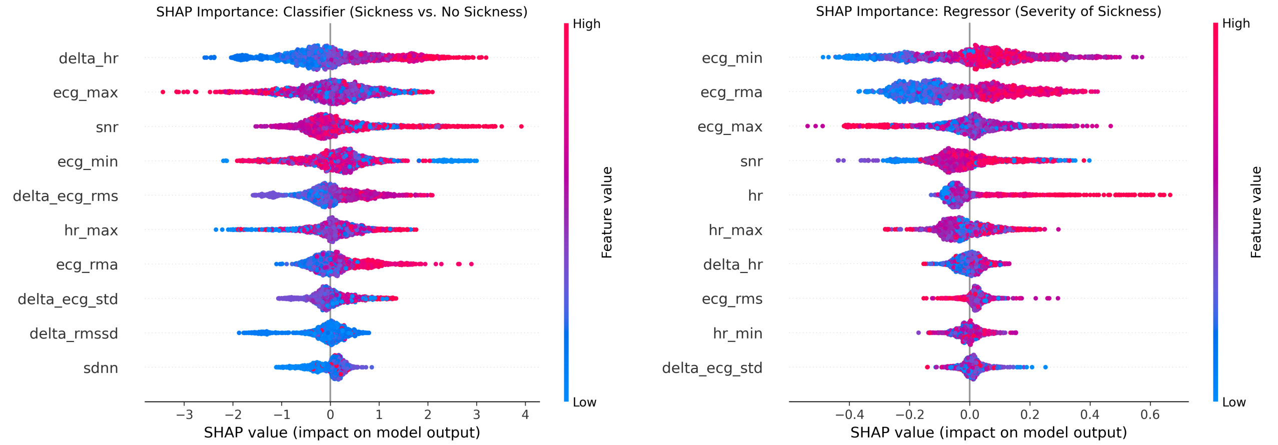


Figure 7: The confusion matrix for the LightGBM hierarchical model on the VR environment dataset exhibits a clear diagonal, reflecting high predictive accuracy across all four sickness levels (0–3). Off-diagonal values are minimal, with most misclassifications occurring between neighbouring severity classes, highlighting the model's ability to distinguish gradual increases in cybersickness intensity.

6.2. Model Explainability

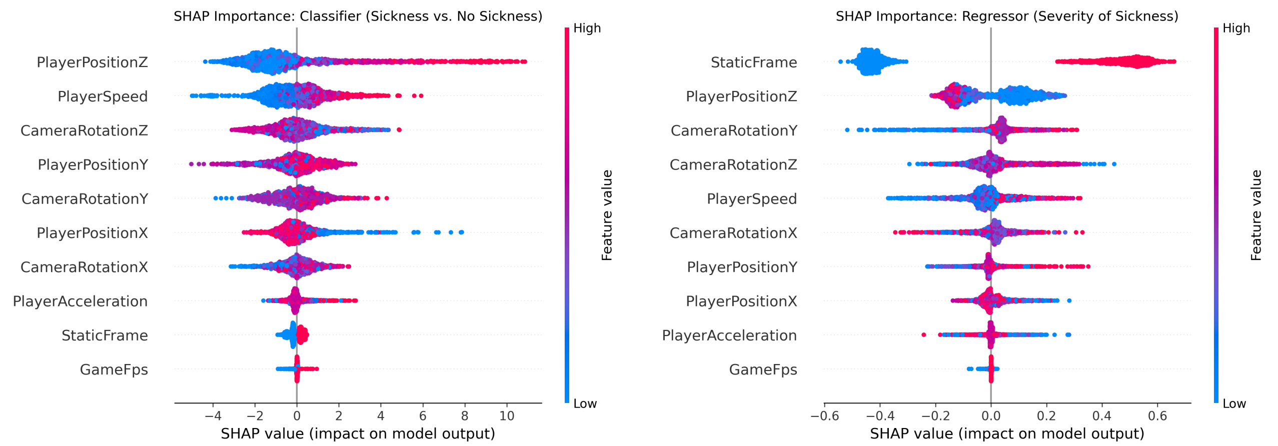
Explainability of the models was assessed using intrinsic explanations from EBMs and post-hoc SHAP analyses (Figure 8) for gradient-boosting models, computed at the level of time windows for both binary onset detection and continuous severity estimation. In the multi-layer architecture evaluated on the public dataset, ΔHRV emerged as the most influential predictor for onset detection, whereas in single-layer models electrodermal activity (EDA) had previously dominated. Rather than indicating a fixed physiological hierarchy, this shift suggests that feature importance can be influenced by model structure and feature interactions. In particular, it highlights the possibility that appropriately designed models can extract sufficiently informative representations from ECG alone, potentially reducing the need for more complex or intrusive sensing modalities such as EDA. Guided by this observation, the validation stage was restricted to ECG-derived features only. SHAP analyses indicate that heart rate dynamics, such as Δhr, hr_{max}, and ECG amplitude-related descriptors, like ecg_{max}, ecg_{min}, and ecg_{rma}, are consistently among the most influential features for both onset and severity. However, their relative importance and directional contributions differ between classification (onset) and regression (severity), suggesting that the physiological patterns underlying sickness onset are not identical to those governing its severity. Environmental and kinematic features exhibited distinct roles across the model's layers (Figure 9). During onset detection, movement-related dynamics, notably *PlayerPositionZ* and *PlayerSpeed*, emerged as primary drivers, indicating that initial cybersickness responses are predominantly triggered by users' motion within the virtual environment. Environmental variables such as *StaticFrame* and *camera auto-movement* contributed to a lesser extent, suggesting that visual factors play a secondary role in the initial onset of discomfort. In the severity estimation stage, the model's attention shifts toward visual stability, with *StaticFrame* becoming the most influential predictor of sickness intensity. Kinematic features, including *CameraRotation* and *PlayerPosition*, remain relevant but exert more distributed effects. This separation underscores the effectiveness of the multi-layer approach, as each stage can specialise in the features most pertinent to its task, with onset detection capturing motion-driven triggers and severity estimation reflecting the impact of visual disruptions.



(a) Shap values for classification task.

(b) Shap values for regression task.

Figure 8: SHAP summary plots in the classification (left) and regression (right) tasks. Each point corresponds to a sample, with colour indicating the feature value (blue: low, red: high) and horizontal position representing its contribution to the model output. Features are ranked by mean absolute SHAP value. Heart rate dynamics (e.g., Δhr, hr_{max}) and ECG amplitude-related features (e.g., ecg_{max}, ecg_{min}, ecg_{rma}) appear prominently in both plots, with a wide spread of SHAP values indicating heterogeneous effects across samples. The colour gradients suggest consistent trends where higher and lower values of these variables tend to contribute in opposite directions.



(a) Shap values for classification task.

(b) Shap values for regression task.

Figure 9: SHAP summary plots for the classification (left) and regression (right) tasks. Each point represents a sample, with colour indicating the feature value (blue: low, red: high) and horizontal position its impact on the model output. In the classification task, kinematic features like *PlayerPositionZ* and *PlayerSpeed* are the primary drivers of sickness onset. In the regression task, importance shifts toward *StaticFrame*, which becomes the dominant predictor for sickness severity. Other spatial dynamics, such as *CameraRotation*, maintain a distributed influence across both tasks, highlighting the distinct roles of motion and visual stability in the cybersickness experience.

7. Discussion

Across feature domains, electrodermal activity and other autonomic signals remain among the most informative indicators of cybersickness. However, the structured, hierarchical representation of ECG-derived features demonstrates that ECG alone can serve as a reliable predictor for both binary onset and continuous severity, as shown in the previous section. Metrics related to heart rate dynamics and ECG amplitude capture the interaction between sympathetic and parasympathetic activity, reflecting the autonomic instability induced by sensory conflict in virtual environments. At the same time, feature importance should be interpreted with caution, as it depends on model structure and feature interactions. The results therefore do not imply that ECG is intrinsically superior to other modalities, but rather that appropriately designed models can extract sufficiently informative representations from ECG alone. This suggests a viable pathway towards reducing reliance on more complex or intrusive sensing modalities. Environmental features, by contrast, provide indirect information, primarily indicating potential triggers rather than immediate physiological responses. This distinction is supported by user feedback collected during data acquisition, where participants consistently reported increased discomfort in conditions involving high-speed motion, static frames, automatic camera movement, and prolonged exposure. These subjective reports align with both the identified environmental triggers and the observed physiological responses, confirming that such features effectively elicit measurable autonomic reactions. These insights directly informed the system design. A modular architecture based on a publisher–subscriber paradigm enables independent yet temporally aligned processing of heterogeneous data streams, including ECG and environmental telemetry. By decoupling acquisition, preprocessing, modelling, and visualisation, the framework supports real-time operation, clear separation between sensing and inference layers, and straightforward extensibility to additional modalities. This design also enables controlled comparisons between physiological-only, environment-only, and multimodal configurations under consistent temporal constraints. Furthermore, the integration of a multi-layered taxonomy linking technical triggers, physiological responses, and mitigation strategies allows predictive outputs to be translated into actionable design insights, moving beyond detection towards interpretation of underlying causes. Methodological rigour was ensured through a fully nested cross-validation pipeline, in which preprocessing, feature selection, and hyper-parameter tuning were strictly confined to training folds. Final performance estimates were obtained from unseen outer folds, reducing the risk of information leakage and providing reliable generalisation metrics. The consistently high F1 scores in the binary classification task indicate that the model effectively detects the presence of cybersickness. In contrast, greater variability in severity estimation can be attributed to the subjective nature of the target labels and class imbalance, which introduce ambiguity between adjacent severity levels. Decomposing the problem into binary onset detection followed by severity estimation further reflects the staged nature of cybersickness. The observed reduction in mean absolute error compared to direct regression suggests that this formulation better aligns with the underlying physiological dynamics. Finally, the ECG feature space explored in this work is not exhaustive. Additional time-domain and non-linear descriptors were not included, leaving room for further improvements in modelling performance and robustness across environments. The current results should therefore be interpreted as a structured and interpretable baseline rather than an upper bound, providing a foundation for future extensions in both feature engineering and multimodal integration.

8. Conclusions

This work introduces a structured framework for the objective assessment of cybersickness through the integration of physiological signals and virtual environment telemetry. Its primary contribution lies in the joint formalisation of physiological responses and environmental stimuli within a unified, time-resolved modelling pipeline. ECG-derived features are shown to support reliable onset detection and provide a meaningful signal for severity estimation, indicating that compact and interpretable representations of autonomic activity can enable practical monitoring without reliance on complex sensing configurations. In parallel, environmental factors are explicitly modelled as measurable variables, allowing controllable scene parameters to be directly linked to physiological responses and enabling a more actionable understanding of cybersickness triggers. Methodologically, the adoption of a fully nested validation pipeline ensures that performance reflects genuine generalisation rather than artefacts of data leakage. Within this framework, strong performance in onset detection suggests clear separability between non-symptomatic and symptomatic states, while greater uncertainty in severity estimation reflects the limitations of subjective ground truth. The decomposition of the problem into onset detection followed by severity estimation aligns with the progressive nature of cybersickness and improves stability compared to direct regression. More broadly, time-resolved modelling enables the alignment of physiological responses with specific environmental conditions, overcoming the limitations of retrospective evaluation. Finally, the modular system architecture supports synchronised yet decoupled processing of heterogeneous data streams and facilitates real-time operation and extensibility. The integration of a multi-layered taxonomy further connects sensing, modelling, and design, allowing predictions to be interpreted in terms of underlying causes. Overall, the framework demonstrates that **interpretable**, **scalable**, and **temporally grounded** assessment of cybersickness is feasible using consumer-grade sensing and structured environment representation, providing a foundation for adaptive virtual environments.

8.1. Limitations and Future Work

Despite the adoption of nested cross-validation and strict preprocessing separation, the dataset size remains a limiting factor, constraining statistical power and generalisation. In addition, the **participant pool was not demographically diverse**, consisting predominantly of male subjects aged between 20 and 26 years, most of whom had limited prior experience with virtual reality. This restricts external validity and may bias the observed physiological responses. Future work should therefore prioritise larger and more heterogeneous datasets, enabling a more robust analysis of inter-individual variability and its interaction with physiological markers. The current temporal modelling relies on sliding windows, which are inherently limited in their ability to capture long-term dependencies and cumulative effects. This assumption is not trivial. Cybersickness is known to develop progressively, and short temporal contexts may fail to encode delayed or sustained responses. Memory-aware approaches, such as Long Short-Term Memory networks or interpretable recurrent tree-based models with memory [30], could provide a richer representation of symptom dynamics. However, these methods introduce a trade-off between temporal expressiveness and interpretability, which must be carefully considered. Finally, the present framework is limited to passive detection and does not address intervention. A key direction for future research is the development of closed-loop adaptive systems, in which predicted physiological strain is used to dynamically adjust virtual environment parameters in real time. Potential strategies include field-of-view restriction, motion smoothing, or the introduction of stabilising visual cues. Existing frameworks such as GingerVR [3] demonstrate the feasibility of such adaptive approaches, but their effectiveness in reducing cybersickness requires systematic experimental validation. These adaptive VR systems can range from predefined rules to algorithmic and learning-based interventions.

References

[1] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery Data Mining*, pages 2623–2631, 2019.

[2] André Altmann, Laura Toloşi, Oliver Sander, and Thomas Lengauer. Permutation importance: a corrected feature importance measure. *Bioinformatics*, 26(10):1340–1347, 2010.

[3] Samuel Ang, Jiacheng Chen, Anthony L. Abbott, and Doug A. Bowman. Gingervr: An open source repository of cybersickness reduction techniques for unity. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 314–315. IEEE, 2020.

[4] Stephen Bates, Trevor Hastie, and Robert Tibshirani. Cross-validation: what does it estimate and how well does it do it? *Journal of the American Statistical Association*, 119(546):1434–1445, 2024.

[5] Benjamin Stephanus Botha and Lizette De Wet. CyPVICS: A framework to prevent or minimise cybersickness in immersive virtual clinical simulation. *Journal of Biomedical Informatics*, 144:104443, 2023.

[6] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[7] Gavin C. Cawley and Nicola L. C. Talbot. On over-fitting in model selection and subsequent selection bias in performance evaluation. *The Journal of Machine Learning Research*, 11:2079–2107, 2010.

[8] Chao Chen, Andy Liaw, and Leo Breiman. Using random forest to learn imbalanced data. *University of California, Berkeley, Technical Report*, 666(1):1–12, 2004.

[9] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1(1):785–794, 2016.

[10] David R. Cox. The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B (Methodological)*, 20(2):215–232, 1958.

[11] A. Garcia-Agundez, C. Reuter, H. Becker, R. A. Konrad, P. Caserman, A. Miede, et al. Development of a classifier to determine factors causing cybersickness in virtual reality environments. *Games for Health Journal*, 8(6):439–444, 2019.

[12] Azadeh Hadadi, Jean-Rémy Chardonnet, Christophe Guillet, and Jivka Ovtcharova. SmartSimVR: An architecture integrating machine learning and virtual environment for real-time simulation adaptation. In *2024 10th International Conference on Automation, Robotics and Applications (ICARA)*, pages 531–535. IEEE, 2024.

[13] Patrick S. Hamilton and Willis J. Tompkins. Quantitative investigation of qrs detection rules. *IEEE Transactions on Biomedical Engineering*, BME-33(12):1157–1165, 1986.

[14] Trevor Hastie and Robert Tibshirani. Generalized additive models. *Statistical Science*, 1(3):297–310, 1986.

[15] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer New York, 2nd edition, 2009.

[16] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.

[17] imbalanced-learn developers. Balancedrandomforestclassifier — api reference. *imbalanced-learn documentation*, 2025.

[18] R. Islam. Physiological dataset for vr cybersickness severity classification. `https://github.com/shovonis/CyberSicknessClassification/tree/master/severity_classification/data`, 2022. Accessed: 2022-05-26.

[19] Rifatul Islam, Samuel Ang, and John Quarles. CyberSense: A closed-loop framework to detect cybersickness severity and adaptively apply reduction techniques. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 149–152. IEEE, 2021.

[20] Rifatul Islam, Youngho Lee, Maryam Jaloli, Ibrahim Muhammad, Dan Zhu, and John Quarles. Automatic detection of cybersickness from physiological signal in a virtual roller coaster simulation. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 648–649. IEEE, 2020.

[21] Rifatul Islam, Youngho Lee, Maryam Jaloli, Ibrahim Muhammad, Dan Zhu, Paul Rad, et al. Automatic detection and prediction of cybersickness severity using deep neural networks from user's physiological signals. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 400–411. IEEE, 2020.

[22] J. Richard Jennings, Wendy K. Berg, J. Steven Hutcheson, Kathleen M. Gill, Clyde F. Williams, and Roscoe A. Dykman. Publication guidelines for heart rate studies in man. *Psychophysiology*, 18(3):226–231, 1981.

[23] Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum, and Michael G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, 3(3):203–220, 1993.

[24] Young-Yoon Kim, Han-Jun Kim, E-Nae Kim, Hee-Dong Ko, and Hyung-Tae Kim. Characteristic changes in the physiological components of cybersickness. *Psychophysiology*, 42(5):616–625, 2005.

[25] Serkan Kiranyaz, Onur Avci, Osama Abdeljaber, Turker Ince, Moncef Gabbouj, and Daniel J. Inman. 1d convolutional neural networks and applications: A survey. *Mechanical Systems and Signal Processing*, 151(107398):1–41, 2021.

[26] Eugenia M. Kolasinski. Simulator sickness in virtual environments. Technical Report 1027, U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA, 1995.

[27] Joseph J. LaViola. A discussion of cybersickness in virtual environments. *ACM SIGCHI Bulletin*, 32(1):47–56, 2000.

[28] Scott M. Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. From local explanations to global understanding with explainable ai for trees. *Nature Machine Intelligence*, 2(1):56–67, 2020.

[29] Giampiero Marra and Simon N. Wood. Coverage properties of confidence intervals for generalized additive model components. *Scandinavian Journal of Statistics*, 39(1):53–74, 2012.

[30] Sascha Marton and Moritz Schneider. Remede: Learning recurrent decision trees with memory. *arXiv preprint arXiv:2502.04052*, 2025.

[31] Charles M. Oman. Motion sickness: a synthesis and evaluation of the sensory conflict theory. *Canadian Journal of Physiology and Pharmacology*, 68(2):294–303, 1990.

[32] Christina Orphanidou, Timothy Bonnici, Peter Charlton, Jordi Alastruey, Dmitri Terentjev, and Lionel Tarassenko. Signal-quality indices for the real-time estimation of the arterial blood pressure and electrocardiogram. *IEEE Journal of Biomedical and Health Informatics*, 19(3):830–838, 2015.

[33] Nitish Padmanaban, Robert Ruban, Vincent Sitzmann, Anthony M. Norcia, and Gordon Wetzstein. Towards a machine-learning approach for sickness prediction in 360° stereoscopic videos. *IEEE Transactions on Visualization and Computer Graphics*, 24(4):1594–1603, 2018.

[34] Thiago Malheiros Porcino. Uffcsdata: Vr cybersickness user responses dataset. `https://github.com/tmp1986/UFFCSData`, 2020. Dataset accessed: 2026-02-22.

[35] James T. Reason and J. J. Brand. *Motion Sickness*. Academic Press, London; New York, 1975.

[36] Lisa Rebenitsch and Charles Owen. Review on cybersickness in applications and visual displays. *Virtual Reality*, 20(2):101–125, 2016.

[37] Lisa Rebenitsch and Charles Owen. Estimating cybersickness from virtual reality applications. *Virtual Reality*, 25(1):165–174, 2021.

[38] Lisa Rebenitsch and Francisco R. Quesada. Prioritizing cybersickness factors through a literature review. *Displays*, 44(1):1–12, 2016.

[39] Lisa Rebenitsch and Francisco R. Quesada. A review of cybersickness in virtual reality and comparison with simulator sickness. *Frontiers in Virtual Reality*, 2(705180):1–22, 2021.

[40] Gary E. Riccio and Thomas A. Stoffregen. An ecological theory of motion sickness and postural instability. *Ecological Psychology*, 3(3):195–240, 1991.

[41] scikit-learn Developers. Permutation feature importance — scikit-learn 1.6.1 documentation. *scikit-learn*, 2023.

[42] Lloyd S. Shapley. A value for n-person games. *Contributions to the Theory of Games*, 2(28):307–317, 1953.

[43] Andy Stanford-Clark and Arlen Nipper. The message queuing telemetry transport (mqtt) protocol. Technical report, IBM, 1999.

[44] Kay M. Stanney, Robert S. Kennedy, and Julie M. Drexler. Cybersickness is not simulator sickness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 41(2):1138–1142, 1997.

[45] Sudhir Varma and Richard Simon. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*, 7(1):91, 2006.

[46] Jingshu Wang, Hai-Ning Liang, Diego Vilela Monteiro, Wenge Xu, and Junhua Xiao. Real-time prediction of simulator sickness in virtual reality games. *IEEE Transactions on Games*, 15(2):252–261, 2023.

[47] Elliott Wen, Chitralekha Gupta, Prasanth Sasikumar, Mark Billinghurst, James Wilmott, Emily Skow, Arindam Dey, and Suranga C. Nanayakkara. Vr.net: A real-world dataset for virtual reality motion sickness research. *IEEE Transactions on Visualization and Computer Graphics*, 30(5):2128–2138, 2024.

[48] Simon N. Wood. Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics*, 62(4):1025–1036, 2006.

[49] Simon N. Wood. *Generalized Additive Models: An Introduction with R.* Chapman & Hall/CRC, 2nd edition, 2017.