VectionSense: Multimodal Inertial-Physiological Cybersickness Detection in Consumer VR

Kompal Layal[1,†], Ashish Kumar[1,†], Garvit Chugh[1] and Suchetana Chakraborty[1]

[1]*Indian Institute of Technology Jodhpur, India*

Abstract

Cybersickness remains a major barrier to prolonged VR use, yet most current systems react only after users report discomfort. We present VectionSense, a lightweight multimodal pipeline that detects cybersickness from behavioural and physiological signals in real time. Our system fuses head and controller Simultaneous Localization and Mapping (SLAM) signals from a Meta Quest 3, i.e., 6-DoF visual-inertial tracking outputs from the headset's internal pipeline (position, orientation, and linear/angular velocities) with heart-rate data from an Apple Watch Ultra, aligning them through spline-based resampling and enriching them with kinematic, sway, and HRV-like features. A compact Transformer classifier operates on short multimodal sequences and is evaluated on data from a one-week user study involving 10 participants who experienced a roller-coaster VR scenario. The IMU-only model already achieves high performance (accuracy/F1 = 0.95/0.95), while HR alone is substantially weaker (0.59/0.43); combining both modalities yields the best trade-off, with accuracy = 0.94 and F1 = 0.92, generalising well in leave-one-subject-out evaluation and correlating with post-exposure SSQ scores. The final model has only 77k parameters, latency below 1 ms per window, and fits into a 62 MB VR application, making it practical for continuous in-headset cybersickness monitoring.

Keywords

Cybersickness Detection, Multimodal Sensing, IMU, Heart Rate Variability (HRV), Machine Learning

1. Introduction

Cybersickness is a motion-induced discomfort that arises during immersive virtual reality (VR) experiences, characterised by nausea, oculomotor strain, dizziness, disorientation, and general physical unease [1]. It stems from sensory conflicts where rich visual motion cues are not matched by corresponding vestibular and proprioceptive feedback. *Vection*, the illusory sensation of self-motion induced purely by visual cues, is a key driver of this conflict: strong vection in the absence of matching vestibular feedback is closely linked to the onset and severity of cybersickness [2]. Modern consumer headsets, with wide field-of-view displays, low-latency rendering, and visually intense content such as roller coasters and racing games [3], can amplify these conflicts. As VR spreads across gaming, simulation, rehabilitation, education, and industrial training, cybersickness remains a central barrier to accessibility and long-term use [4].

In practice, cybersickness is still mostly assessed via self-report instruments such as the Simulator Sickness Questionnaire (SSQ) and post-session ratings [5]. While useful for overall analysis, these measures are retrospective, coarse, and blind to the moment-to-moment progression of discomfort within a session, making them unsuitable for real-time adaptation of motion intensity or interaction mechanics [6]. Lab-based approaches using camera systems [7], motion platforms [8], or dense physiological rigs [9] can capture finer-grained responses but are difficult to scale to home settings and do not reflect typical consumer use [10].

Prior work has explored pervasive sensing solutions that use both physiological and behavioural markers to predict or detect cybersickness. Physiological approaches utilize wearable-based heart rate (HR), heart-rate variability (HRV), and electrodermal activity [11], demonstrating that HRV can

SAXR'26: 1st ACM CHI 2026 Workshop on Shaping Future Human Connection: Social Augmentation through XR Technologies, 13-17 April, 2026, Barcelona, Spain

†These authors contributed equally.

m24air004@iitj.ac.in (K. Layal); m24air003@iitj.ac.in (A. Kumar); chugh.2@iitj.ac.in (G. Chugh); suchetana@iitj.ac.in (S. Chakraborty)

support low-latency estimates, albeit with modest person-independent performance. Behavioural approaches use head-motion kinematics as a strong predictor: Salehi et al. [12] predict cybersickness from IMU-based head features, and Li et al. [13] report high accuracy using short windows of headset kinematics while noting that EDA-only models are comparatively noisy. More recent work moves towards multimodal fusion, combining motion with physiology or eye tracking [14, 15], but often relies on non-commodity hardware (e.g., EEG, dedicated biosensors) and lab-style setups. Across these studies, inertial measurement unit (IMU) based predictors emerge as strong baselines, while HR/HRV provides complementary context but suffers from cross-user variability when used alone. To the best of our knowledge, there is no prior work that explicitly fuses headset/controller IMU with heart-rate-derived features for fine-grained cybersickness detection in a consumer VR setting and quantifies the marginal benefit of physiology over an IMU-only baseline.

At the same time, current VR ecosystems already provide sensing channels that could support practical cybersickness monitoring. Headsets and hand controllers continuously provide 6-DoF visual-inertial tracking outputs (position, orientation, and linear/angular velocities) from their internal SLAM pipeline, which implicitly encode motion signatures tied to vection and visual-vestibular conflict. For brevity, we refer to this modality as *IMU* to denote these *IMU-informed, SLAM-fused motion states* provided by the headset (i.e., visual-inertial pose/velocity outputs), rather than raw accelerometer/gyroscope streams. In parallel, many users routinely wear smartwatches or fitness bands that provide heart rate data with little additional burden. Heart rate and simple variability-derived features offer a complementary view of the user's internal state, capturing arousal and stress responses that may not be visible in motion alone. Together, this "head-hand-heart" triad forms an attractive sensing substrate for cybersickness detection using only commodity hardware.

Challenges: However, exploiting this VR-smartwatch configuration for cybersickness monitoring is not straightforward. Characterizing cybersickness from head and hand IMUs, plus wrist-based heart rate, introduces three practical challenges. *First*, motion-induced interference is substantial: the same high-jerk or oscillatory patterns can correspond to either engaging gameplay or uncomfortable vection, and controller gestures inject additional hand dynamics that can mask sickness-related changes. *Second*, the VR and smartwatch channels operate at different sampling rates with independent clocks and intermittent packet loss, making second-level alignment and multimodal fusion of IMU and heart-rate streams non-trivial in practice. *Third*, both motion behaviour and cardiovascular responses vary substantially across individuals and sessions, so models must generalise under subject-independent splits without overfitting to idiosyncratic baselines while still producing outputs that align with subjective SSQ scores.

Contributions: To address these challenges, we present VectionSense, a novel "head-hand-heart" framework that combines headset and controller IMUs with smartwatch-based heart-rate signals for practical, fine-grained cybersickness detection in consumer VR. The contributions of this work are as follows: **(i)** We design and implement VectionSense, a multimodal inertial-physiological pipeline that fuses head and hand kinematics with smartwatch-based heart-rate-derived features in a unified transformer-based classifier, using only sensors available in commodity VR ecosystems. **(ii)** We conduct a validation study with ten participants in a visually intense VR roller coaster scenario, collecting synchronised headset and controller IMUs, heart rate, and manually annotated per-second cybersickness labels obtained via video and sensor-assisted review. **(iii)** We systematically compare three configurations under a subject-independent split, *first,* an inertial-only model, *second,* a heart-rate-only model, and *third,* a fused multimodal model. We show that while IMU-only models already provide strong detection performance, the multimodal model maintains high accuracy and produces cybersickness exposure profiles that align with SSQ scores, suggesting that heart-rate signals are most valuable for contextualising and validating inertial-based estimates. Overall, our results highlight design implications for future VR platforms that aim to adapt content based on user comfort using sensors already embedded in consumer devices.

2. Preliminary Study

Before designing our full pipeline, we conducted a small observational study to understand how cybersickness manifests in naturalistic VR use and whether characteristic body reactions are reflected in head-hand kinematics and heart-rate signals. We recruited six participants (3F, 3M, age 18-30, median 24) from our institute. Each participant wore an Apple Watch Ultra paired with an iPhone SE 2 on the wrist and used a Meta Quest 3 HMD with motion controllers. The study was conducted in a lab environment but with minimal constraints on behaviour. Participants were placed in a visually intense roller-coaster experience and instructed to "interact with it as you normally would". They were free to look around, move their hands, and explore the scene without any scripted tasks or posture restrictions.

Based on prior work, our initial hypothesis was that participants experiencing cybersickness would exhibit a combination of subjective discomfort (e.g., nausea, dizziness, stomach unease), subtle compensatory actions (e.g., adjusting the HMD, pressing the head), and autonomic changes (e.g., elevated or more variable heart rate). During the sessions, we informally noted typical reactions such as burping, reporting stomach ache or nausea, pausing interaction, and adjusting the headset. In addition, we observed several spontaneous behaviours: some participants scratched their nose or face, some tried to fixate on a distant point in the scene to stabilise their perception, others wrapped an arm around their abdomen, pressed their forehead or temples, showed heavier breathing, repeatedly looked around to orient themselves, and ultimately removed the headset and controllers when the discomfort became intolerable.

We recorded head and controller IMU signals from the Quest 3 and heart-rate data from the Apple Watch for all six participants, then visualised the synchronised streams around reported discomfort episodes (Fig. 1). In one representative case, a participant reported burping and feeling unwell; the head IMU showed increased sway, the controller IMU captured hand-to-face movements (e.g., nose-scratching/head-pressing), and heart rate showed a sharp peak relative to baseline. Similar co-occurring head-hand motion and elevated heart rate appeared in other self-reported sickness intervals.

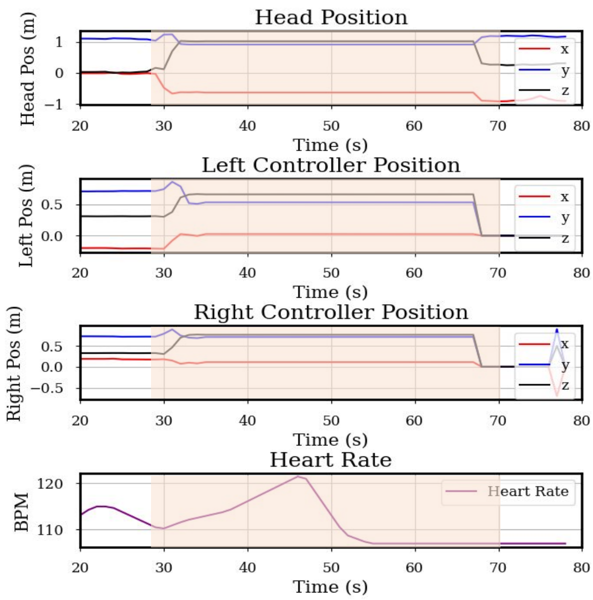


Figure 1: Representative visualisation of IMU and HRV during cybersickness (signs of cybersickness highlighted in orange).

These preliminary observations support our "head-hand-heart" hypothesis: cybersickness in consumer VR is accompanied by a combination of distinctive motion behaviours and physiological changes that are visible in readily available IMU and heart-rate signals. This motivates our formal problem statement: *Can we leverage headset and controller IMUs together with heart-rate features to detect cybersickness at fine temporal granularity and distinguish genuine discomfort from mere enjoyment of visually intense VR content?*

3. Methodology

VectionSense framework has three stages (Fig. 2): (i) signal acquisition from headset, controllers, and smartwatch; (ii) preprocessing, temporal alignment, and feature extraction; and (iii) transformer-based binary classification for presence or absence of cybersickness.

3.1. Signal Acquisition and Preprocessing

We collect two time series per participant: (i) a combined IMU stream from head and controller sensors and (ii) a heart-rate (HR) stream from the smartwatch, each with a binary cybersickness label. Because the IMU and HR streams have different lengths, we align the HR stream to the IMU stream via index-based resampling. If the original HR series has length N_{HR} and the IMU series length N_{IMU}, we define

$$s_i = i \cdot \frac{N_{\mathrm{HR}} - 1}{N_{\mathrm{IMU}} - 1}, \quad i = 0, 1, \ldots, N_{\mathrm{IMU}} - 1,$$

and interpolate HR onto $\{s_i\}$ using cubic-spline interpolation (with linear and nearest-neighbor variants for ablations), yielding an aligned HR sequence $\tilde{h}[i]$ of length N_{IMU}. Labels on the HR stream are not interpolated; for each s_i we assign the nearest original label, i.e., if $k^\star = \arg\min_k |k - s_i|$, then $\tilde{y}[i] = y_{\mathrm{HR}}[k^\star]$. After alignment, IMU and HR channels are concatenated (for the combined modality) and segmented into fixed-length, non-overlapping windows of $T = W \cdot f_s$ samples, where W is the window length in seconds and f_s the effective sampling rate. Each window receives a binary label by majority vote over its `Cybersickness` samples, and these window-level samples form per-participant feature matrices. Finally, we concatenate all participants into global arrays X_{all} and y_{all}, together with a `groups` vector encoding participant identity for subject-disjoint train/test splits and participant-wise cross-validation.

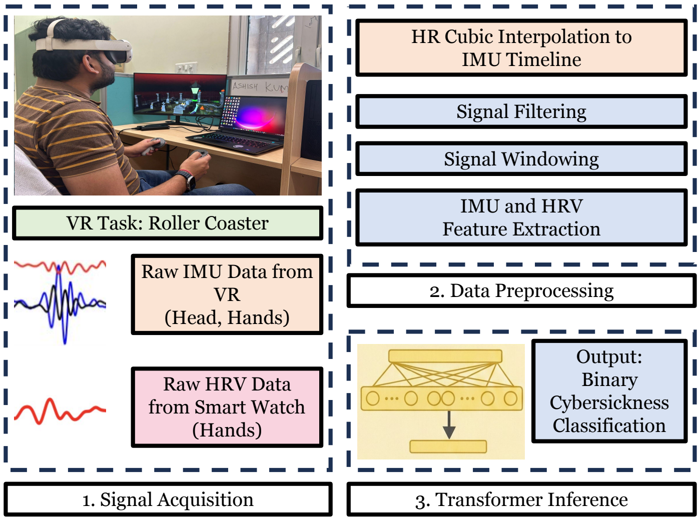


Figure 2: System Architecture of VectionSense.

3.2. Feature Engineering

We derive kinematic and sway-based features from the IMU channels, as well as HRV-like features from the heart-rate stream. Let $\mathrm{IMU}_c[t]$ denote the SLAM-fused motion signal at time t for channel c (e.g., x, y, z axes of head or controller motion). Using a first-order finite difference along time, we obtain velocity, acceleration, and jerk per channel and then summarize them into higher-level measures. Specifically, we compute a sway-variance feature $\mathrm{sway\_var}[t]$ as the variance of $\mathrm{IMU}_c[t]$ across all spatial channels, a trajectory-smoothness term $\mathrm{trajectory\_smoothness}[t] = \|\mathbf{jerk}(t)\|_2$ capturing the overall magnitude of abrupt motion, and a movement-energy feature $\mathrm{movement\_energy}[t]$ given by the sum of squared accelerations across channels. Together, these features capture both slow drift and rapid corrections in head and controller motion that are known to correlate with cybersickness.

For the heart-rate modality, we start from a cleaned beats-per-minute series $\mathrm{hr}[t]$ obtained after forward/backward imputation of missing values. From this, we extract simple time-domain HRV proxies: the first difference $\mathrm{hr\_diff}[t] = \mathrm{hr}[t] - \mathrm{hr}[t-1]$, a quadratic energy term $\mathrm{hr\_sq}[t] = \mathrm{hr}[t]^2$, an approximate RR-interval $\mathrm{rr}[t] = 60/\mathrm{hr}[t]$ (with basic handling of zeros), and its first difference $\mathrm{rr\_diff}[t]$. For classical window-based modelling, all numeric IMU and HR features within a window of T samples are averaged in time to yield a single feature vector $\bar{\mathbf{x}}_w$. Finally, each feature dimension is standardized using the training set's mean and standard deviation, so that all models operate on zero-mean, unit-variance inputs, and the same scaling is consistently applied to the test set.

3.3. Transformer Inference

The preprocessed and aligned feature streams are segmented into non-overlapping sequences of length T time steps, where each sequence is represented as a matrix $X \in \mathbb{R}^{T \times D}$ (time × features). The sequence label y_{seq} is again defined by majority vote over the `Cybersickness` values within that segment,

$$y_{\text{seq}} = \begin{cases} 1, & \text{if } \frac{1}{T}\sum_{t=1}^{T} \texttt{Cybersickness}[t] \geq 0.5, \\ 0, & \text{otherwise.} \end{cases}$$

The classifier operates on sequences X by first applying a linear input projection to each time step. Let $\mathbf{x}_t \in \mathbb{R}^D$ denote the feature vector at time t; we map it into a latent space of dimension $d_{\text{model}} = 64$ via

$$\mathbf{z}_t = W_{\text{in}}\mathbf{x}_t + \mathbf{b}_{\text{in}}, \quad W_{\text{in}} \in \mathbb{R}^{d_{\text{model}} \times D}.$$

A fixed sinusoidal positional encoding $\mathbf{p}_t$ is then added to encode temporal order. The resulting sequence $\{\mathbf{h}_t^{(0)}\}_{t=1}^{T}$ is processed by a stack of $L = 2$ Transformer encoder layers with $H = 4$ attention heads, feed-forward dimension $d_{\text{ff}} = 128$, and dropout probability 0.1, implemented with `batch_first = True`. We denote the encoder output by $F(X) \in \mathbb{R}^{T \times d_{\text{model}}}$. To obtain a fixed-dimensional representation of the entire sequence, we apply global average pooling over time:

$$\mathbf{s} = \frac{1}{T}\sum_{t=1}^{T} F(X)_t,$$

where $F(X)_t$ is the t-th row (time step) of the encoder output. This pooled embedding is passed to a two-layer feed-forward classification head with ReLU nonlinearity and dropout,

$$\mathbf{u} = \text{ReLU}\big(W_1\mathbf{s} + \mathbf{b}_1\big), \quad \mathbf{u'} = \text{Dropout}(\mathbf{u}), \quad \mathbf{o} = W_2\mathbf{u'} + \mathbf{b}_2,$$

where $\mathbf{o} \in \mathbb{R}^2$ are the logits for the two classes (not cybersick vs cybersick). Class probabilities are obtained via a softmax,

$$p_k = \frac{\exp(o_k)}{\exp(o_0) + \exp(o_1)}, \quad k \in \{0, 1\},$$

and the predicted label is $\hat{y} = \arg\max_k p_k$.

Before training, we apply feature scaling per modality by flattening all time steps from all training sequences into a large matrix and fitting a `StandardScaler` exactly as in the window-based case. The same scaling parameters are then applied to every time step of both training and test sequences. Training is performed using the Adam optimizer with learning rate 10^{-3}, mini-batch size 64, and cross-entropy loss

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k\in\{0,1\}} \mathbf{1}[y_i = k] \log p_{ik},$$

4. Experiments and Evaluation

We conducted a one-week lab study with 10 participants (4 female, 6 male), aged 18-32 years (median = 24). Each participant wore a Meta Quest 3 headset and an Apple Watch Ultra, paired with an iPhone SE 2, which were used to record head/controller IMU streams and heart-rate signals during VR exposure. Participants experienced a roller coaster scenario and were instructed to behave naturally, following the same protocol as in our pilot study. Each session lasted for a maximum of 30 minutes or until the participant reported the onset of cybersickness, whichever occurred earlier.

Throughout the session, participants were continuously monitored and recorded from a frontal camera view, which served as the primary source of ground-truth labels for cybersickness episodes. Once participants reported the symptoms, the VR experience was paused, and after some time, they completed the Simulator Sickness Questionnaire (SSQ). After the session, they were also asked to report specific symptoms and the approximate moment they first felt uncomfortable. All participants were provided refreshments and were free to withdraw from the study at any point.

Two independent annotators manually labelled cybersickness episodes from the frontal video and sensor traces. Inter-rater reliability, computed using Cohen's kappa, was $\kappa = 0.77$, indicating substantial agreement. Disagreements were resolved through discussion to obtain the final ground-truth labels. The overall experimental setup, including device placement and data flow, is illustrated in the system architecture diagram (Figure 2). The study protocol was approved by the institutional ethics committee, and all participants provided written informed consent.

4.1. Experimental Conditions and Metrics

Using the collected dataset, we evaluated our multimodal Transformer-based cybersickness detection pipeline under three primary model configurations: (i) *IMU-only*, using features derived solely from head/controller motion; (ii) *HR-only*, using heart-rate and HRV-like features; and (iii) *Combined IMU+HR*, where both behavioural and physiological features are jointly modelled. In addition, we implemented a simple decision-level fusion baseline (IMU+HR "Fused"), which is used as a reference in our analysis.

For each configuration, we trained and evaluated models using participant-disjoint splits. We report standard classification metrics: macro accuracy, precision, recall, and F1-score at the window level. To assess user-level generalization, we further perform a leave-one-subject-out (LOSO) evaluation and report performance per participant. Beyond aggregate metrics, we analyse confusion matrices, permutation-based feature importance, and the relationship between model predictions and post-exposure SSQ sickness scores.

To study temporal stability, we conducted a longitudinal experiment over three consecutive days for a subset of participants (N=2) who repeated the same VR task. For each day, we computed the total number of cybersickness detections, enabling us to visualise day-wise variability in predicted sickness. Finally, we profiled end-to-end system performance in terms of inference latency, memory usage, and battery consumption on the Apple Watch Ultra and iPhone SE 2; these measurements characterize the feasibility of deploying our approach on commodity devices.

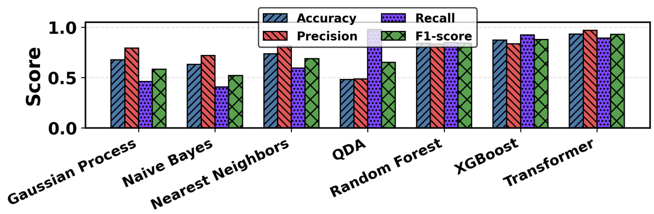


Figure 3: Comparison with Baseline Models

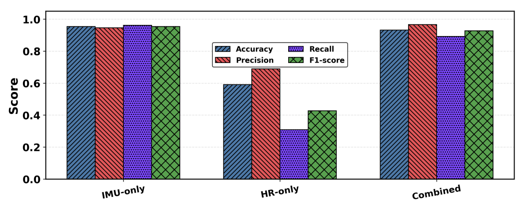


Figure 4: Model Comparison

4.2. Results

Figure 3 compares classical window-based classifiers with our sequential Transformer. Tree ensembles such as Random Forest and XGBoost already outperform shallow models, but they still operate on temporally averaged features and cannot capture the fine-grained evolution of sway and HRV within a window. The Transformer, in contrast, attends over the full IMU+HR sequence and learns cross-channel dependencies and onset patterns that are diagnostic of cybersickness, yielding the highest accuracy and F1-score despite having a comparable parameter budget. This suggests that even a compact attention-based architecture is better suited than static feature models for multimodal VR sickness detection.

4.2.1. Overall Model Performance

Figure 4 summarises the accuracy, precision, recall, and F1-score for the three model variants. The IMU-only model achieves strong performance across all metrics (accuracy = 0.95, F1 = 0.95), indicating that head and controller motion patterns contain substantial behavioural cues related to cybersickness. The HR-only model, in contrast, performs considerably worse (accuracy = 0.59, F1 = 0.43), suggesting that physiological signals alone are insufficiently discriminative at the chosen temporal resolution. The

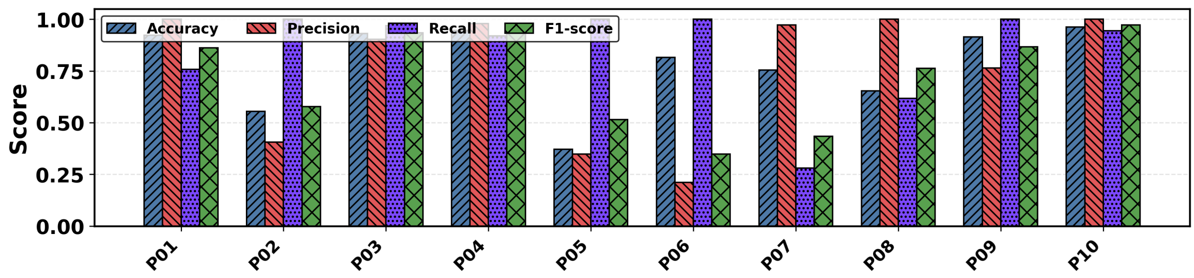


Figure 5: LOSO Evaluation

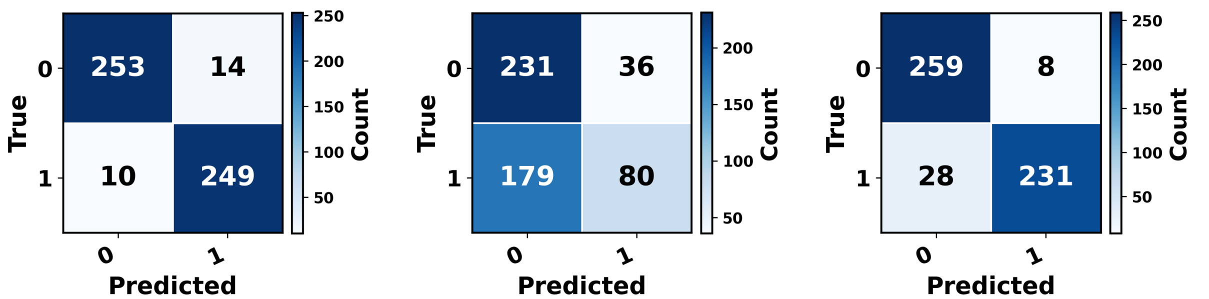
Combined model achieves the highest performance across all metrics, with accuracy = 0.94, precision = 0.97, recall = 0.90, and F1 = 0.92. This reinforces the hypothesis that behavioural and physiological modalities provide complementary information, thereby improving overall predictive robustness.

A leave-one-subject-out (LOSO) evaluation was also conducted to assess model generalisation across users. The per-participant performance, shown in Figure 5, indicates that the multimodal classifier maintains high accuracy for the majority of participants, with several achieving near-perfect scores. Lower-scoring cases reflect expected inter-individual variability in motion behaviour and physiological responses to VR stimuli. Overall, it demonstrates that the proposed model generalises well to unseen users and remains robust despite differences in individual susceptibility to cybersickness.

4.2.2. Confusion Matrix Analysis

Confusion matrices for the three model variants (Figures 6a, 6b, 6c) provide deeper insight into class-specific behaviour. The IMU-only model misclassifies 14 non-sick samples and 10 sick samples, indicating only a slight tendency to under-report sickness, but overall achieves a highly balanced performance. This is consistent with motion-based predictors, which may fail to detect sickness onset during periods of relatively stable head motion.

The HR-only model displays significantly larger confusion, with 179 false negatives and moderate false positives. This highlights the limited discriminative capacity of heart rate alone. The Combined model exhibits the best balance, with only 8 false positives and 28 false negatives. This strong performance indicates that the Transformer effectively integrates the complementary temporal signatures of behavioural instability and physiological change, resulting in highly reliable cybersickness detection across participants.

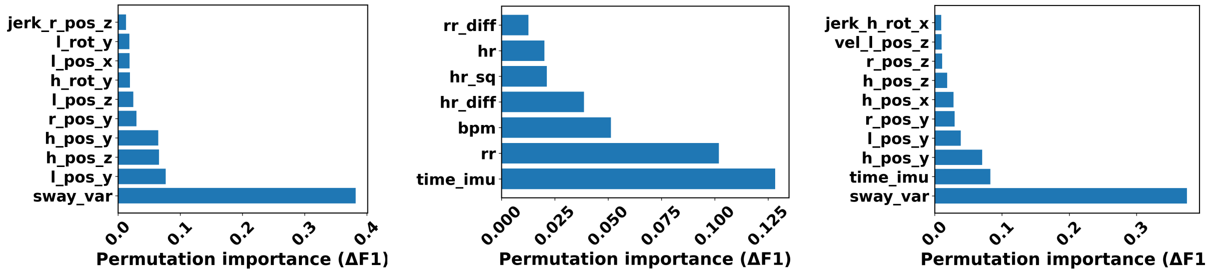


(a) IMU　(b) HRV　(c) IMU + HRV

Figure 6: Confusion Matrix

4.2.3. Feature Importance

To interpret learned representations, permutation-based feature importance was computed for the top 10 features in each modality-specific model (Figures 7a, 7b, 7c). Across all IMU-based models, sway_var

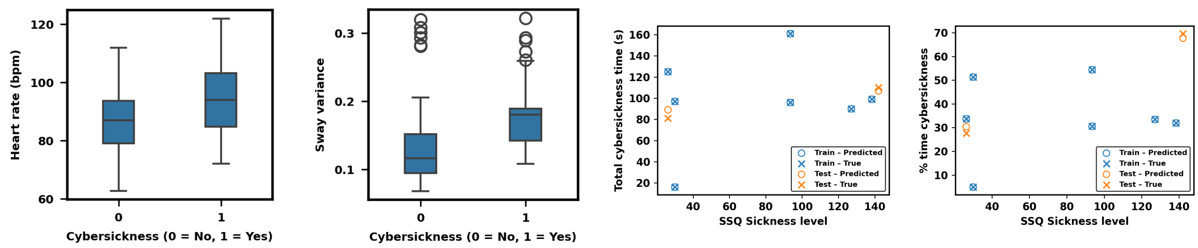


(a) IMU (b) HRV (c) IMU + HRV

Figure 7: Top 10 Feature Importance

emerges as the most important feature, with a dramatic drop in F1-score upon permutation. This aligns with established findings that postural instability is strongly correlated with cybersickness. Additional head-position and controller-position derivatives reflect lateral and vertical micro-adjustments during discomfort. For HR-based models, `rr` and `bpm` contribute most strongly. However, their magnitude of importance remains substantially lower than that of IMU features, reinforcing the weaker standalone performance of HR-only detection. In the Combined model, IMU-derived features still dominate, but HR features contribute non-trivial gains in F1-score.

The exploratory boxplots reveal clear physiological and behavioural differences between cybersick and non-cybersick windows. Heart-rate distributions in Figure 8a display a noticeable upward shift in both median and interquartile range for cybersick segments, indicating increased autonomic arousal during discomfort. Similarly, sway variance extracted from the IMU signals is substantially higher during cybersickness as seen in Figure 8b, consistent with the postural-instability theory and prior findings that body sway becomes more erratic as sickness severity increases.



(a) (b) (c) (d)

Figure 8: (a) Heart rate vs cybersickness detection. (b) Sway variance vs cybersickness detection. (c) SSQ level vs total cybersickness time. (d) SSQ level vs percentage time cybersick.

4.2.4. SSQ Correlation and Behavioural Consistency

To assess consistency with subjective reports, we examined the relationship between SSQ sickness level and two behavioural metrics derived from model predictions: (1) total cybersickness duration (in seconds) and (2) percentage of time participants experienced cybersickness. Figures 8c and 8d show scatter plots comparing predicted and true values for both training and testing participants. Across both plots, the predicted values closely follow the true labels for most data points, indicating that the model is able to capture the underlying trends between subjective SSQ sickness scores and objective cybersickness duration metrics.

4.2.5. System Metrics

The combined Transformer remains lightweight, with only 77,378 parameters and a serialized size of $\approx$ 0.30, MB, which is negligible compared to the 62.1, MB VR application bundle. On our reference hardware, end-to-end training completes in $\sim 3.3, \mathrm{s}$ and full test-time evaluation in $\sim 0.01, \mathrm{s}$. Single-window inference takes under $1, \mathrm{ms}$, indicating that cybersickness detection introduces virtually no additional latency and is suitable for real-time deployment.

4.2.6. Longitudinal Study

Figure 9 illustrates two representative users: Participant 1 shows an increase in detected cybersickness episodes from Day 1 to Day 3, while average heart rate changes only modestly, indicating that the model is not simply reacting to elevated BPM but to richer multimodal patterns. Participant 2 exhibits consistently lower and more stable detection counts, accompanied by a gradual decrease in average heart rate, consistent with partial physiological adaptation to repeated exposure. Overall, the per-day trajectories remain within a narrow range and preserve a coherent within-subject ranking (days that appear more uncomfortable behaviourally/physiologically receive more detections), suggesting that the classifier is robust to session-to-session variability and suitable for repeated-use VR settings.

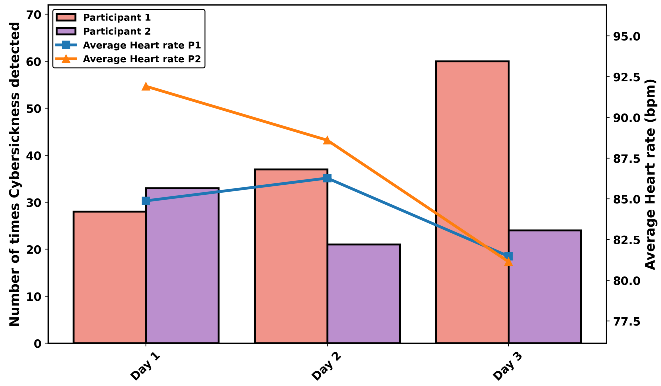


Figure 9: Longitudinal Study

5. Conclusion and Future Work

In this work, we presented a multimodal Transformer-based framework for cybersickness detection that fuses behavioural and physiological signals from commodity XR hardware. Using head and controller IMU data from a Meta Quest 3 and heart-rate signals from an Apple Watch Ultra, collected in a controlled roller-coaster experience with 10 participants, we showed that motion-based features alone can achieve high detection performance, while the addition of heart-rate and HRV-like features yields a robust multimodal model that generalises across users. The Combined IMU+HR model consistently outperforms HR-only baselines and matches or exceeds IMU-only performance, while LOSO evaluation demonstrates strong cross-participant generalisation. Correlations between model-derived cybersickness duration, percentage of time spent sick, and post-exposure SSQ scores further indicate that the learned representations are behaviourally and perceptually meaningful. Future work will focus on expanding the participant pool and evaluating the model across a broader range of VR applications to improve generalizability. Automating parts of the labeling process and incorporating additional physiological signals, such as eye movement, may further enhance accuracy. Ultimately, integrating this model into real-time VR systems could enable adaptive comfort mechanisms that mitigate cybersickness as it emerges, improving accessibility and user experience in immersive environments.

Acknowledgements: We thank all study participants for their time and valuable feedback. This work is supported by the Prime Minister's Research Fellowship (PMRF) awarded by the Government of India.

Declaration on Generative AI

The authors acknowledge the use of OpenAI's ChatGPT[1] to assist with grammar correction and stylistic refinement. No sections of the manuscript were wholly generated by AI tools. All content was critically reviewed and verified by the authors to ensure that it accurately represents the research findings and scholarly intent.

[1]ChatGPT-5 from OpenAI (https://openai.com/index/introducing-gpt-5/), last access: March 3, 2026

References

[1] A. Kemeny, J.-R. Chardonnet, F. Colombet, Getting rid of cybersickness, Virtual reality, augmented reality, and simulators (2020).

[2] J. Teixeira, The Unexpected Vection Hypothesis: Exploring New Relationships Between Vection, Postural Instability, and Cybersickness in Virtual Reality, Ph.D. thesis, University of Wollongong, 2025.

[3] K. Khan, Advancements and challenges in 360 augmented reality video streaming: A comprehensive review, IJC 13 (2024) 1–20.

[4] E. Al Khalifah, R. Hammady, M. Abdelrahman, O. Al-Shamaileh, M. Marghany, H. El-Jarn, A. Darwish, Y. Kurt, Technology anxiety in virtual reality adoption: Examining the impact of age, past experience, and cybersickness, IEEE Access (2025).

[5] V. Sevinc, M. I. Berkman, Psychometric evaluation of simulator sickness questionnaire and its variants as a measure of cybersickness in consumer virtual environments, Applied ergonomics 82 (2020).

[6] K. N. de Winkel, T. M. Talsma, R. Happee, A meta-analysis of simulator sickness as a function of simulator fidelity, Experimental Brain Research 240 (2022) 3089–3105.

[7] T. M. Porcino, E. Clua, D. Trevisan, C. N. Vasconcelos, L. Valente, Minimizing cyber sickness in head mounted display systems: design guidelines and applications, in: SeGAH, IEEE, 2017, pp. 1–6.

[8] Z. Lin, X. Gu, S. Li, Z. Hu, G. Wang, Intentional head-motion assisted locomotion for reducing cybersickness, TVCG (2022).

[9] E. Chang, M. Billinghurst, B. Yoo, Brain activity during cybersickness: a scoping review, Virtual reality 27 (2023).

[10] Y. Zhu, Z. Liang, Y. Wu, T. Li, Y. Wang, Towards consumer-grade cybersickness prediction: Multi-model alignment for real-time vision-only inference, in: ICMM, 2025, pp. 6859–6867.

[11] P. Reyero Lobo, P. Perez, Heart rate variability for non-intrusive cybersickness detection, in: ICIME, 2022, pp. 221–228.

[12] M. Salehi, N. Javadpour, B. Beisner, M. Sanaei, S. B. Gilbert, Cybersickness detection through head movement patterns: A promising approach, in: ICHCI, Springer, 2024, pp. 239–254.

[13] R. Li, Y. Wang, H. Yin, J.-R. Chardonnet, P. Hui, A deep cybersickness predictor through kinematic data with encoded physiological representation, in: ISMAR, IEEE, 2023, pp. 1132–1141.

[14] C. Qu, X. Che, S. Ma, S. Zhu, Bio-physiological-signals-based vr cybersickness detection, CCF TPCI 4 (2022) 268–284.

[15] B. U. Demirel, A. H. Dogan, J. Rossie, M. Möbus, C. Holz, Beyond subjectivity: Continuous cybersickness detection using eeg-based multitaper spectrum estimation, TVCG 31 (2025). doi:10.1109/TVCG.2025.3549132.