A Deep Cybersickness Predictor through Kinematic Data with Encoded Physiological Representation

Ruichen Li*
The Hong Kong University of
Science and Technology (Guangzhou)

Yuyang Wang†
The Hong Kong University of
Science and Technology (Guangzhou)

Handi Yin‡
The Hong Kong University of
Science and Technology (Guangzhou)

Jean-Rémy Chardonnet§
Arts et Metiers Institute of Technology
LISPEN, HESAM Université

Pan Hui¶
The Hong Kong University of
Science and Technology (Guangzhou)

ABSTRACT

Users would experience individually different sickness symptoms during or after navigating through an immersive virtual environment, generally known as cybersickness. Previous studies have predicted the severity of cybersickness based on physiological and/or kinematic data. However, compared with kinematic data, physiological data rely heavily on biosensors during the collection, which is inconvenient and limited to a few affordable VR devices. In this work, we proposed a deep neural network to predict cybersickness through kinematic data. We introduced the encoded physiological representation to characterize the individual susceptibility; therefore, the predictor could predict cybersickness only based on a user's kinematic data without counting on biosensors. Fifty-three participants were recruited to attend the user study to collect multimodal data, including kinematic data (navigation speed, head tracking), physiological signals (e.g., electrodermal activity, heart rate), and Simulator Sickness Questionnaire (SSQ). The predictor achieved an accuracy of 97.8% for cybersickness prediction by involving the pre-computed physiological representation to characterize individual differences, providing much convenience for the current cybersickness measurement.

Index Terms: Cybersickness Prediction, VR, Kinematic data, Physiological Representation, Deep Neural Classifiers

1 INTRODUCTION

Virtual reality (VR) technology has made significant progress in recent years, gaining popularity among many fields such as education, gaming, and healthcare [54]. It can create realistic virtual scenes, providing an immersive environment through multimodal feedback for users. However, when users are exposed to VR applications, they may experience cybersickness (CS), which is a common side effect of VR and an unavoidable problem. CS symptoms include headache, eyestrain, stomach awareness, and disorientation [35], and these symptoms can last up to five hours after VR immersion [49]. Studies report that more than 60 % of VR users experience CS, making it a primary impediment to the expansion of the VR market [36]. To alleviate the influence of CS on VR users, it is vital to develop some strategies to predict it. Predicting the early onset of CS can help us understand the possible causes and potential risk factors so that we can take corresponding strategies for the prevention of severe CS symptoms.

---

*e-mail: rli965@connect.hkust-gz.edu.cn
†e-mail: yuyangwang@ust.hk (Corresponding author)
‡e-mail: hyin335@connect.hkust-gz.edu.cn
§e-mail: jean-remy.chardonnet@ensam.eu
¶e-mail: panhui@ust.hk (Corresponding author)

Previous studies have investigated CS prediction utilizing data from integrated sensors in head-mounted displays (HMDs) [1, 7, 24, 32]. These sensors can acquire stereoscopic videos/images and kinematic data, which are then used as input for machine learning algorithms to predict the severity of CS. Kinematic data include head movement, eye movements, and operations [38], which have been commonly used as objective cybersickness indicators and have been utilized for predictions [7, 15]. However, kinematic data from different users may demonstrate similar patterns and trajectories in the same VR environment [57], making it challenging to provide deeper insights into individual differences. Similar to kinematic data, stereoscopic videos/images also struggle to accurately reflect individual differences. This is because even with the same visual content, people can have varying interpretations.

To involve individual differences, CS prediction using physiological data and biometrics has received increasing attention [52]. Physiological data are highly valued because of their reliability and objectivity in characterizing individual differences from both physiological and psychological perspectives. Prior studies report that electrodermal activity (EDA) on the forehead has a higher correlation with CS and can be used for the prediction [18, 60]. In addition to EDA, other physiological data are widely used for CS prediction, such as heart rate (HR), heart rate variability (HRV), electroencephalography (EEG), and electrogastrogram (EGG) [11, 26, 31]. However, collecting physiological signals requires external biosensors, which presents more challenges in deployment among current affordable consumer-level HMDs such as Oculus Quest 2 and HTC Vive. In addition, the data collection process for physiological signals requires users to limit their interaction experience to prevent noisy information, further complicating the use of external biosensors. To fully unleash the potential of VR, the data collection method should be less intrusive and remain cognitively effortless during the interaction.

To overcome the above-mentioned limitations, we aim to predict the severity of CS only using kinematic data from integrated sensors in HMDs without employing additional biosensors. In addition, we will involve individual differences in the predictive model with EDA-related encoded physiological representation. Kim *et al.* have used stereoscopic videos to estimate EEG-related cognitive state, achieving an accuracy of 90.48% for CS prediction [31]. However, using stereoscopic videos for prediction has several drawbacks, such as the need for large transmission bandwidth and storage space, privacy concerns, difficulty in analysis, and sophisticated neural network models. Moreover, relying only on video data leads to a narrow perspective, which may lose some key information associated with the CS, leading to worse prediction accuracy. EEG data are recorded from electrodes on the scalp, requiring specialized and expensive equipment, and participants need to prepare by washing their hair and avoiding stimulants before recording. Thus, in this work, we use kinematic data including motion and head-tracking data, which are much convenient for deployment and avoid many privacy concerns. Moreover, instead of using EEG whose collection and processing are

more complex and time-consuming, we opt for EDA and propose an EDA-enhanced kinematic model. Specifically, we first encode EDA data into physiological representation. Subsequently, we train an embedding layer to learn physiological embedding from kinematic data, gradually approaching the representation. Finally, our model can interpret individual differences with physiological embedding, without the need for the encoded physiological representation from EDA data. Using the physiological representation to enhance kinematic data provides an in-depth understanding of CS since it captures both inter and intra-individual differences of the CS. In summary, we have made the following three key contributions:

- Achieve improved performance with only EDA data using temporal features and numerical representation formulated from statistical and topological features.

- Propose a novel deep learning model to predict CS, which considers kinematic and physiological features using kinematic data only.

- Insight into the influence of time span and exposure time on CS.

2 RELATED WORK

Cybersickness is a common phenomenon often appearing to VR users during or after an immersive experience. This phenomenon has become an important area of research due to the increasing popularity of immersive technologies and the potential adverse effects on user experience and well-being. In related work, we will review the current understanding of CS, including its causes, measurements, and predicting approaches.

2.1 Causes

Cybersickness symptoms are similar to motion sickness, but the underlying causes are different [53]. Researchers have proposed various theories to explain the mechanisms of CS. One of the most popular theories is the sensory conflict theory, which suggests that CS occurs due to conflicts between visual and vestibular information [48]. The subjective vertical theory proposes that CS occurs due to a conflict between the sensory information from the body and the brain's expectations [3]. To better understand CS, researchers have recommended integrating these two theories into a comprehensive model [4].

Factors associated with CS can be classified into three categories: hardware, content, and human factors [6]. Hardware factors such as field of view (FOV) and latency have been widely studied. Adaptive FOV has been proposed as a way to reduce CS, with FOV representing the maximum visual angle of a display device. Kim *et al.* employed a dynamic FOV system that adjusts the FOV using electrophysiological signals. Their results indicate that reducing FOV effectively alleviates discomfort, especially during acceleration and rotational movements [33]. Another hardware factor that affects CS is latency. DiZio *et al.* demonstrated that increasing latency increases the severity of VR sickness [12]. Content factors, such as optical flow, graphic realism, and controllability, have also been studied. Optical flow is the VR scene's movement that allows individuals to experience illusory self-motion. Faster locomotion speed in the VR scene is associated with more severe CS [8, 39]. Human factors such as age, ethnicity, and gaming experience significantly correlate with CS levels [58].

2.2 Measurements

Several methods have been developed to measure CS, including subjective and objective measures.

2.2.1 Subjective measurements of CS

First introduced by Kellogg *et al.*, the Motion Sickness Questionnaire (MSQ) remains a subjective tool for evaluating CS across various dimensions [28]. Later, Golding *et al.* proposed a shortened version of the Motion Sickness Susceptibility Questionnaire (MSSQ) [20]. However, the Simulator Sickness Questionnaire (SSQ), developed by Kennedy *et al.*, is widely recognized and universally used as a subjective measurement tool for CS [29]. The SSQ comprises 16 questions divided into three categories (i.e., nausea, oculomotor, disorientation) to assess the severity of each potential manifestation of CS. The scores from these subcategories are combined to determine the total SSQ score, which indicates the overall severity of the sickness. Although the SSQ is a comprehensive and widely used tool, it may lead to an extra workload in answering 16 questions. Therefore, researchers have developed relatively short and quick-to-report questionnaires. For example, the Fast Motion Sickness Scale (FMS) [30] and the Misery Scale (MISC) [50] are well-known unidimensional questionnaires. For example, FMS comes with one question asking participants to rate their CS level on a scale ranging from 0 (no sickness) to 20 (severe sickness). The severity of CS experienced by the participant is reported as a score between 0 and 10 by the MISC.

2.2.2 Objective measurements of CS

In addition to subjective measurements, researchers have explored objective measurements on top of self-reported evaluation. Two commonly measured indicators are postural sway and electrophysiological changes. Participants are asked to stand on a motion platform to measure postural sway while their VR position is recorded [9, 55]. Multiple studies have demonstrated that these measurements can effectively predict the onset of VR sickness [13, 46]. Real-time monitoring of participants' psychophysical state changes during the VR experience is also conducted. A significant study by Kim *et al.* examined the physiological indicators of VR sickness, recording various electrophysiological measures before, during, and after the VR experience [34]. Among these indicators, electrogastrogram (EGG), eye blink, heart period, and the delta and beta power bands of electroencephalogram (EEG) exhibit CS-specific responses. Additionally, phasic skin conductance at the forehead, as reported by Golding and John [19], has also shown a strong correlation with motion sickness onset. In a subsequent study, Gavgani *et al.* [18] confirmed that skin conductance on the forehead is a reliable measure of nausea during immersion. Other studies have reported that CS can cause a $4°C$ increase in body temperature and a 20-50ms increase in reaction time [43]. Furthermore, researchers have examined blood pressure, fMRI, theta/total parameter at the frontal and parietal lobes, and hormonal levels as potential measures of VR sickness.

2.3 Predicting Approaches

Previous studies have explored using Machine Learning (ML) and Deep Learning (DL) models to predict CS based on individual factors, stereoscopic 3D videos, kinematic data, and physiological signals. According to the data source, these approaches can be divided into two categories.

2.3.1 CS Predicting with integrated sensors in HMDs

Integrating various sensors within VR HMDs allows for collecting multimodal data, including videos, images, and user kinematic data. These sensors track the user's head movement, operation, and body position, which are then processed by the system to create a seamless VR experience.

Regarding videos, Padmanaban *et al.* [44] created a dataset of 19 stereoscopic videos and corresponding sickness levels to predict simulator sickness. They considered FOV, velocity, and stimulus depth as features and evaluated them with a decision tree model.

The prediction performance was later improved with a 3D CNN model that used saliency, disparity, and optical flow features [37]. Du *et al.* built a CNN model that automatically learnt and integrated multi-level stereoscopic visual sickness descriptors for simulator sickness prediction [14].

In addition to utilizing video data for prediction, researchers have also leveraged user kinematic data, including eye-tracking [21], motion [15, 22], head-tracking [27] and postural data [1, 51, 59]. Gazing activity has been found to help reduce CS severity by Chang *et al.* [7]. Lopes *et al.* [40] found that the CS level can be affected by both the position of the pupil and the type of blink. Wang *et al.* utilized normal-state postural data collected from participants to train a long short-term memory (LSTM) model, which was then used to predict CS levels [56]. Islam *et al.* proposed a novel approach for assessing CS severity using a multimodal deep fusion network that integrates data from various modalities, including eye-tracking, head-tracking, and stereo-image data [24].

2.3.2 CS Predicting with external biosensors

Previous research has suggested various approaches to predict CS by utilizing physiological signals, which have been found to be significantly correlated with CS. Garcia *et al.* conducted an experiment with 66 participants, collecting electrocardiographic, electrooculographic, respiratory, and skin conductivity data to classify cybersickness using KNN and SVM classifiers [17]. The accuracy achieved by the Binary and Ternary classifiers were 82% and 56%, respectively. Kim *et al.* achieved an accuracy of 89.16% in predicting CS by using EEG data [31]. Their study involved collecting EEG data from 200 participants who were immersed in 44 virtual reality (VR) simulations using an 8-channel EEG. The authors estimated the cognitive state and its association with CS levels using CNN and LSTM models based on brain signals. Jeong *et al.* proposed a data processing technique for DNN models and achieved an impressive accuracy rate of 98.02% using EEG data. Additionally, they observed a distinct pattern of CS occurrence during VR simulations.

Besides EEG, electrodermal activity (EDA) is a widely used measure of emotional arousal. It includes Skin Conductance Level (SCL) and Skin Conductance Response (SCR) [23]. The SCL changes slightly on a time scale of tens of seconds to minutes and can vary between individuals. The SCR, on the other hand, is the phasic component that shows rapid changes and is sensitive to emotionally arousing events, such as event-related SCRs (ER-SCRs). These variations in the phasic component are visible as EDA bursts or peaks, which can occur within 1-5 seconds after the onset of emotional stimuli. Islam *et al.* achieved an 87.38% accuracy in predicting CS by analyzing heart rate (HR) and EDA data from 22 participants using deep neural networks [26]. They also proposed a multimodal fusion network involving eye tracking, HR, and EDA to forecast CS onset 60 seconds in advance with a root-mean-square error of 0.49 [25].
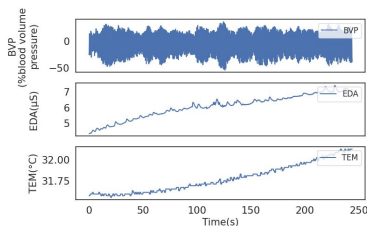
3 USER STUDY

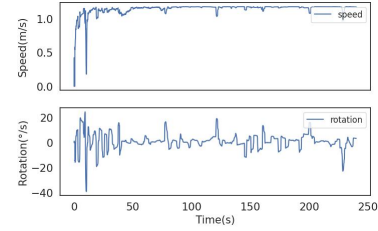Figure 1: Physiological data of one sample.

Figure 2: Motion data of one sample.

To train a high-quality model to make CS predictions based on physiological and kinematic data, we conducted a user study to collect the data as ground truth. The task was performed in the HTC Vive Pro Eye. Participant information and task design are presented in Sect. 3.1, and we further introduce the data collection process shown in Sect. 3.2.

3.1 Participants and Task Design

In this study, 53 participants ($M_{age} = 26.3$, $SD_{age} = 3.3$, females: 26) completed a navigation task in an immersive environment. Most participants were young individuals as VR applications are more popular among this group. Each participant completed the task three times over three days, resulting in 159 samples. However, two samples were discarded due to device defects, leaving 157 samples for data analysis.

Upon arrival, participants signed a consent form indicating potential risks and their agreement for the experiment. They completed a questionnaire assessing their health conditions and experience with gaming and VR devices. All participants confirmed their eligibility, reporting no pre-existing medical issues that would affect their performance. They were compensated with various gifts upon completion of the study.

Participants received clear instructions on the navigation task and how to use HTC Vive Pro hand controllers. They had to navigate a virtual forest along a winding gravel path. The VR immersion lasted four minutes to induce moderate cybersickness without posing health risks. Fig. 3 shows a visual representation of the path.

Figure 3: Virtual scenario in which the participants navigate along the highlighted path.

3.2 Data Collection Procedure

The general experimental procedure is presented in Fig. 4.

Figure 4: Experiment Procedure for the data collection

1. Participants received instructions on using HTC Vive Pro hand controllers and were informed they could terminate the experiment if experiencing CS symptoms.

2. Participants wore an HTC Vive Pro headset and an Empatica E4 wristband on one arm. Real-time physiological data (EDA, BVP, TEM) were collected using the Empatica E4 wristband [1] with sensors transmitting data via Bluetooth to a processing computer with sampling frequencies: EDA (4Hz), BVP (64Hz), TEM (4Hz). One physiological sample is shown in Fig. 1.

3. Participants engaged in the virtual environment (Fig. 3), directing movements using the HTC Vive Pro hand controller touchpad. Physiological and kinematic data, including head-tracking and motion data, were collected synchronously during four-minute navigation. After completion, participants removed the head-mounted display. One sample of motion data is illustrated in Fig. 2.

4. The Simulator Sickness Questionnaire (SSQ) assessed CS severity, measuring nausea, oculomotor, and disorientation symptoms through 16 questions. Participants completed pre- and post-experiment SSQ to report CS degree. SSQ score calculated by subtracting post-exposure from pre-exposure score, given as,

$$SSQ = SSQ_{post} - SSQ_{pre} \qquad (1)$$

We gathered three distinct sets of data, each of which shared the same starting and ending times: head-tracking, motion, and physiological data. The data collected from the integrated sensors of the HMD and Empatica E4 wristband are summarized in Table 1.

Table 1: List of temporal data collected during the navigation task

Data Type	Data
Head-Tracking Data	Head Position (i.e., x, y and z)
(90Hz)	Head Rotation (i.e., x, y and z)
Motion Data	Speed
(90Hz)	Rotation
	Electrodermal Activity (4Hz)
Physiological Data	Temperature (4Hz)
	Blood Volume Pulse (64Hz)

4 DATA PROCESSING

It is necessary to pre-process the raw data from the experiment to construct the input and output for building deep learning models. First, we explain how to establish the ground truth in Sect. 4.1, then define the CS prediction task in Sect. 4.2. Next, we show the data processing details in Sect. 4.3. By carefully handling the data, we can ensure that the deep learning model is trained on high-quality inputs and outputs, which will improve its performance and accuracy.

4.1 Ground Truth Construction

SSQ covers a wide range of symptoms that are commonly associated with CS and it is a standardized tool that has been used as ground truth in many studies [22, 27]. So our work used SSQ scores to construct the ground truth. Stanney *et al.* stated that cybersickness could be classified using the following criteria: SSQ scores below 5 indicate negligible symptoms, while scores above 20 suggest a bad intervention [53]. However, a recent meta-analysis challenges using scores above 20 as an indicator of a bad intervention, as it found that approximately one third of users discontinued with a weighted

---
[1] https://www.empatica.com/research/e4/

SSQ score of 40 or higher [5]. Therefore, we employ a revised classification rule for cybersickness level, which includes four categories: negligible ($0 \leq SSQ \leq 5$), low ($5 < SSQ \leq 20$), moderate ($20 < SSQ \leq 40$), and high ($SSQ > 40$). This updated classification standard provides a more accurate reflection of cybersickness level. Also, the samples were uniformly distributed among such classification, with 30, 41, 53, and 33 samples for each class. The data was labeled as follows:

$$CS = \begin{cases} Negligible & ,if \quad 0 \leq SSQ \leq 5 \\ Low & ,if \quad 5 < SSQ \leq 20 \\ Moderate & ,if \quad 20 < SSQ \leq 40 \\ High & ,if \quad SSQ > 40 \end{cases} \qquad (2)$$
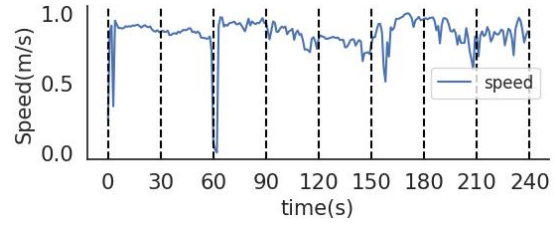
4.2 Problem Definition

Figure 5: The original four minutes are divided into eight 30s segments.

Previous studies indicated that the exposure time does not affect the severity of the VR experience. For example, Melo *et al.* [42] found no significant difference between shorter and longer (5 or 7 min) experiences concerning participants' sense of presence and CS. Katharina *et al.* [45] also reported no CS differences between 10-minute and 20-minute virtual immersion for karatekas aged 18-60. Thus, if a participant experiences sickness after the four-minute VR experience, we assume the CS level (CS) remains constant for all observed data within that time. We will investigate the exposure time-CS relationship further in Sect. 6.3. The entire four-minute VR experience is labeled as CS_{240}.

$$[D^1, D^2, D^3, D^4, ..., D^{239}, D^{240}] \rightarrow CS_{240}$$

D^t represents observed data at time t, comprising 4 data points from EDA and Temperature (4Hz), 64 data points from BVP (64Hz), and 90 data points from head-tracking and motion (90Hz). The original data was segmented based on time span T_s. For instance, using $T_s = 30s$, the four-minute data was divided into eight segments (see Fig. 5). Each time segment was labeled with CS level CS_{240}.

The CS level prediction task can be defined as giving the observed data (e.g., physiological, head-tracking, and motion data) in the time segment with a time span of T_s from a person, determining the user's CS level.

4.3 Data Processing

Table 2: An independent t-test of the physiological data between higher CS and lower CS group (df = 157), p-value < 0.05 represents significant properties

	Mean	SD	t-value	p-value
Low-CS BVP	-0.013569	0.281273	-1.4065	0.1616
High-CS BVP	-0.075292	0.270550		
Low-CS EDA	2.753912	3.253526	2.6919	**0.0079**
High-CS EDA	4.373491	4.158797		
Low-CS TEM	34.429848	1.298672	0.7111	0.4781
High-CS TEM	34.587766	1.467705		

Table 3: Feature set of one sample

Feature Type	Data Type (Dimension)	Features
Temporal	Head-tracking (12)	Head Position First difference of Head Position Head Rotation First difference of Head Rotation (i.e., x, y and z)
	Motion (4)	Speed First difference of Speed Rotation First difference of Rotation
	EDA (15)	EDA, SCR, SCL EDA_{min}, EDA_{max}, EDA_{mean}, EDA_{std} SCR_{min}, SCR_{max}, SCR_{mean}, SCR_{std} SCL_{min}, SCL_{max}, SCL_{mean}, SCL_{std}
Numerical	EDA (38)	mean_scl, std_scl, std_scr, corr, num_responses, ...

The rendering frequency of Unity3D fluctuates around 90Hz due to computer performance, posing a challenge for data collection at a constant rate. Hence, all head-tracking and motion data were resampled to 90Hz during preprocessing. Subsequently, the following steps were carried out:

- **Correlation Analysis.** We combined the first and last two CS level categories, resulting in the higher CS group ($SSQ > 20$) and lower CS group ($0 \leq SSQ \leq 20$). We conducted independent sample t-tests on the physiological data of these two groups. The results showed a significant difference in EDA signals, as presented in Table 2. However, there was no significant difference in BVP and TEM data between the higher and lower CS groups. Therefore, we excluded BVP and TEM signals from further data processing.

- **Data Downsampling.** To account for varying sampling rates, all data was downsampled to 1Hz. The mean value of the samples was used to retain information during downsampling. For example, head-tracking and motion data were aggregated using the mean value over 90 data points per second, while EDA signals were aggregated using the mean value over 4 data points.

- **Data Segmentation.** In order to build our prediction task, we divided the original four-minute data into several windows according to the time span T_s. Since T_s is temporarily set to 30s, we got a total of $8 * 157 = 1256$ samples, as presented in Fig. 5.

- **Feature Selection.** We extracted features from head-tracking, motion, and EDA data. The features extracted from one sample are listed in Table 3.

  - For speed and rotation data, we computed the first difference between consecutive frames to calculate the difference between adjacent seconds. For head-tracking vectors (x, y, and z), we computed the first difference of consecutive vectors.
  - For temporal features of EDA data, courtney *et al.* suggested using a 3-second window to capture changes in EDA data [10]. So we employed a trailing moving average with a 3-second window to smooth the EDA data and account for noise. Properties of physiological signals within the previous 3-second window, including mean, std, maximum, and minimum, were considered indicators of physiological changes. D_{max}, D_{min}, D_{mean}, and D_{std} captured abrupt fluctuations in physiological

signals over a 3-second period. As the EDA signals' phasic component (SCR) is associated with arousing stimulus events and the SCL differs significantly across different individuals, we should link the CS level to them. Using Neurokit2 [41], a Python toolbox designed to process physiological signals, we could extract the SCR and SCL from the original EDA. Then we did the same processing as the original EDA data on the extracted two components.

  - For EDA data, in addition to temporal features, we also analyzed each time segment and extracted numerical representation formulated from statistical and topological features. The meanings of 38-dimension numerical representation are shown in our supplemental material [2].

- **Data Normalization.** To mitigate the impact of individual differences that could potentially hinder the model's training due to varying signal magnitudes, we opted to normalize all temporal features using min-max normalization, thereby scaling the values to fall within the range of $(0, 1)$.
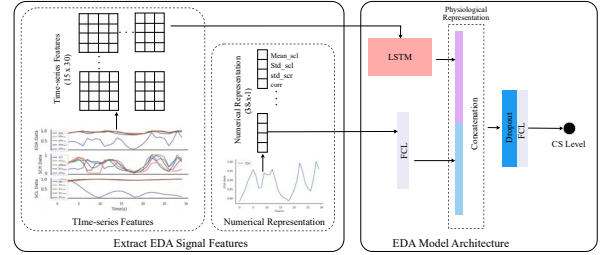
5 MODEL ARCHITECTURE

Figure 6: EDA Model. The model aims to encode EDA data as a low-dimensional physiological representation.
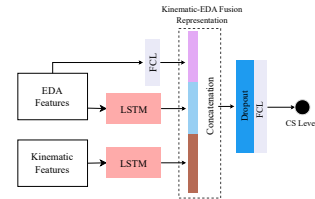

Figure 7: Kinematic-EDA Fusion model. The model learns features from physiological and kinematic data.

After preparing input and output data, we constructed and evaluated multiple deep learning models for different input modalities, identifying relevant inputs for accurate predictions. Analyzing model performance provided valuable insights into the influence of input features on outcomes, leading to the development of precise and efficient deep learning models for diverse applications. This section describes the environment setup (see Sect. 5.1) before exploring specific model architectures for different input combinations in subsequent sections.

---

[2]Available on OSF at osf.io, released under a CC BY 4.0 license. They include (1) Excel containing the meanings of 38-dimension numerical features, (2) an example of fine-tuning results, (3) datasets, and (4) source code.
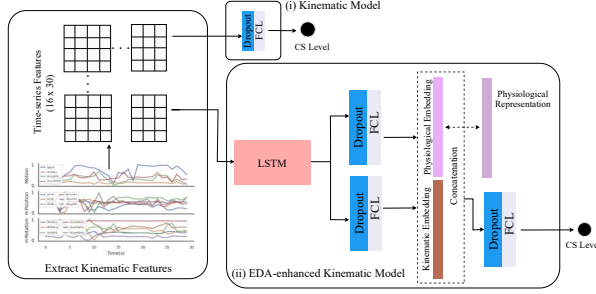
Figure 8: (i) Kinematic Model. The model learns features from kinematic data, including head-tracking and motion data. (ii) EDA-enhanced Kinematic Model.The model learns the physiological and kinematic embedding from Kinematic data.

5.1 Setup

Our deep learning models were trained and evaluated using TensorFlow. Execution took place on a system running Ubuntu 18.04, equipped with two NVIDIA GeForce RTX 3090 Ti GPUs (each with 24GB memory). The dataset underwent 5-fold cross-validation, with 4 groups for training and the remaining group for testing. Mean accuracy served as the primary evaluation metric. To optimize parameters and improve performance, we employed the Optuna [3] Python library, which efficiently explores and exploits the parameter space. This significantly enhanced our hyperparameter tuning process and resulted in improved model performance. Additionally, we utilized the Adam optimizer with an epoch of 100 and a batch size of 64.

5.2 EDA Model

The EDA model encodes EDA data as a low-dimensional physiological representation. Fig. 6 shows the model's schematic architecture. The time span T_s is 30s, resulting in a temporal EDA feature size of $(15 * 30)$ and a numerical representation size of $(1 * 38)$. An LSTM layer with ($LSTM\_size$) units learns the temporal representation, while a fully-connected layer with ($Dense\_size_1$) neurons and activation function ($Acti$) learns the numerical representation. These representations are combined using a concatenate layer, forming the physiological representation used as the ground truth in the EDA-enhanced kinematic model discussed later. To prevent overfitting, a dropout layer with a dropout rate of $Rate$ is applied, followed by a fully-connected Dense layer with ($Dense\_size_2$) neurons and activation function ($Acti$).

5.3 Kinematic Model

The kinematic model learns features from kinematic data, including head-tracking and motion data. The detailed architecture is depicted in Fig. 8(i). The time span T_s is set to 30s, so the temporal kinematic feature is $(16 * 30)$. We use an LSTM layer with ($LSTM\_size$) units to learn the temporal representation. The output is then passed to a dropout layer with a specified dropout rate $Rate$ to reduce overfitting. This is followed by a fully-connected Dense layer with a defined number of neurons ($Dense\_size$) and an activation function ($Acti$).

5.4 Kinematic-EDA Fusion Model

The kinematic-EDA fusion model combines physiological and kinematic data. See Fig. 7 for the detailed architecture. With a time span of 30s, the temporal input EDA feature size is (30 * 15), the temporal input kinematic feature size is (30 * 16), and the numerical input EDA feature size is (1 * 38). Two LSTM layers, with ($LSTM\_size_1$) units for temporal EDA representation and ($LSTM\_size_2$) units for

[3]https://optuna.org

kinematic representation, are employed. A fully-connected Dense layer with ($Dense\_size_1$) neurons and activation function ($Acti_1$) encodes numerical EDA features. The outputs of these three layers are concatenated and passed through a dropout layer with a specified dropout rate. They are then fed into a fully-connected Dense layer with ($Dense\_size_2$) neurons and activation function ($Acti_2$).

5.5 Proposed EDA-enhanced Kinematic Model

The proposed EDA-enhanced kinematic model aimed to learn the physiological embedding from Kinematic data with encoded physiological representation learned from the EDA model mentioned in Sect. 5.2. The overall architecture is illustrated in Fig. 8(ii). Here we use an LSTM network with ($LSTM\_size$) units as a sequential network to process the kinematic features. Then the dropout layer with the $Rate_1$ and fully-connected Dense layer with ($Dense\_size_1$) and activation function ($Acti_1$) output physiological embedding. Next, the dropout layer with the $Rate_2$ and fully-connected Dense layer with ($Dense\_size_2$) and activation function ($Acti_2$) are used to extract kinematic representation. Then the physiological embedding is concatenated with the kinematic representation to produce the final representation vector. The output is forwarded into the dropout layer with the $Rate_3$ and then passed to a fully connected Dense layer with ($Dense\_size_3$) neurons and activation function ($Acti_3$). The model uses the Adam optimizer with the Lr. Note that in order to make the learned embedding closer to the encoded physiological representation, we design a custom loss function, which is formed by two terms: the prediction loss and the regression loss,

$$L = L_{pre} + \beta * L_{reg}$$

Where L_{pre} represents the standard cross-entropy loss between ground truth and the output unit of the last FCL, L_{reg} denotes the mean squared error (MSE) or mean absolute error (MAE) between the physiological features and the physiological representation learned by EDA model, and a constant β is used to adjust the balance between the two terms. These two parameters are fine-tuned with Optuna.

6 Experiments and Results

Table 4: Results of models with different time spans before downsampling

| Time span T_s | 10 | 15 | 20 | 30 | 40 |
	$mean \pm std$	$mean \pm std$	$mean \pm std$	$mean \pm std$	$mean \pm std$
Kinematic Model	0.968 ± 0.006	0.968 ± 0.008	0.959 ± 0.010	0.921 ± 0.021	0.871 ± 0.031
EDA Model	0.697 ± 0.015	0.625 ± 0.027	0.593 ± 0.034	0.502 ± 0.013	0.480 ± 0.024
Kinematic-EDA Fusion Model	0.976 ± 0.005	0.974 ± 0.009	0.963 ± 0.010	0.924 ± 0.026	0.880 ± 0.015
EDA-enhanced Kinematic Model	$\mathbf{0.978} \pm 0.003$	$\mathbf{0.976} \pm 0.005$	$\mathbf{0.968} \pm 0.005$	$\mathbf{0.927} \pm 0.005$	$\mathbf{0.906} \pm 0.005$

Table 5: Results of models with different time spans after downsampling

| Time span T_s | 10 | 15 | 20 | 30 | 40 |
	$mean \pm std$	$mean \pm std$	$mean \pm std$	$mean \pm std$	$mean \pm std$
Kinematic Model	0.836 ± 0.032	0.838 ± 0.012	0.847 ± 0.042	0.870 ± 0.012	0.871 ± 0.031
EDA Model	0.494 ± 0.019	0.489 ± 0.039	0.489 ± 0.035	0.497 ± 0.031	0.480 ± 0.024
Kinematic-EDA Fusion Model	0.845 ± 0.034	0.843 ± 0.020	0.869 ± 0.028	0.884 ± 0.016	0.880 ± 0.015
EDA-enhanced Kinematic Model	$\mathbf{0.867} \pm 0.026$	$\mathbf{0.855} \pm 0.029$	$\mathbf{0.874} \pm 0.028$	$\mathbf{0.888} \pm 0.020$	$\mathbf{0.906} \pm 0.015$

Results are presented in the following subsections. After model setup, Optuna was used for parameter optimization, and an example of fine-tuning results with a time span of 30s before downsampling

Table 6: Results of models with different exposure times. *n* represents the number of segments removed

	n=1	n=2	n=3	n=4	n=5
	mean ± std	*mean ± std*	*mean ± std*	*mean ± std*	*mean ± std*
Kinematic Model	0.908 ± 0.029	0.909 ± 0.025	0.917 ± 0.011	0.922 ± 0.030	0.925 ± 0.031
EDA Model	0.525 ± 0.034	0.505 ± 0.019	0.532 ± 0.016	0.510 ± 0.033	0.518 ± 0.040

is shown in our supplemental material [4]. Firstly, the impact of different time spans on the model was investigated in Sect. 6.1. Next, the performance of different models was analyzed and compared in Sect. 6.2. The proposed EDA-enhanced Kinematic Model achieved the highest accuracy using only kinematic data. Lastly, the relationship between different exposure times and CS was studied in Sect. 6.3.

6.1 Performance on different time span

Time span T_s refers to the length of recently recorded data, including physiological and kinematic data. We aim to study how the models can make predictions based on historical data from a certain period of time,. The time spans included 10s, 15s, 20s, 30s, and 40s, dividing the original 4-minute data into 24, 16, 12, 8, and 6 segments respectively. Retraining the model revealed that as the time span increased, the model's performance gradually decreased, as shown in Table 4. For instance, in the EDA-enhanced Kinematic Model, the accuracy decreased from 0.978 when the time span was 10s to 0.906 when the time span was 40s.

Results Analysis. Surprisingly, the accuracy change contradicted our intuition that a longer time span should improve prediction performance by providing more information. We suspected the limited training sample size to be the bottleneck affecting model performance. At a 40s time span, the sample size is $6 * 157 = 942$, while at a 20s time span, it is $12 * 157 = 1884$. To ensure consistent training sample sizes, we randomly downsampled them to $6 * 157 = 942$ across all time spans. After retraining, results are shown in Table 5. The Pearson correlation p-values between accuracy and time span were as follows: Kinematic Model ($r = .955, p = .011$), EDA Model ($r = -.520, p = .368$), Kinematic-EDA Fusion Model ($r = .875, p = .051$), EDA-enhanced Kinematic Model ($r = .933, p = .020$). Three p-values, close to or less than 0.05, suggested a correlation between accuracy and time span, aligning with our intuition. Comparing experiments before and after downsampling, we observed significantly reduced model performance under the same time span due to downsampling. For example, with a 20-second time span, the sample size decreased from $12 * 157 = 1884$ to 942, resulting in EDA-enhanced Kinematic model performance dropping from 0.968 to 0.874. Hence, training sample size became the model bottleneck.

In addition, we plotted the model performance curves before downsampling shown in Fig. 9 and found that the three blue curves had a turning point at a time span of the 20s. Therefore, 20 is a value that balances both the training sample size and the amount of information carried.

6.2 Model Comparison

The accuracy of different models is presented in Table 4. The results demonstrate that the proposed EDA-enhanced Kinematic Model can achieve the highest accuracy in predicting CS. Take the example when the time span is the 30s; specifically, the EDA model achieves an accuracy of 0.502 using only the EDA signals. When only using the kinematic data, the kinematic model achieves 0.921. The fusion model, which integrates physiological and kinematic data, has a slightly higher accuracy of 0.924 than the kinematic model. The proposed EDA-enhanced Kinematic Model achieves the best
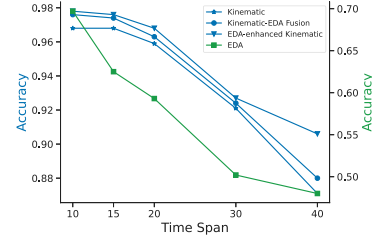
Figure 9: Performance on different time spans before downsampling.

accuracy of 0.927, where kinematic and physiological embeddings were encoded using kinematic data. Hence, prediction is only based on kinematic data without utilizing physiological data.

Results Analysis. As shown in Table 4, the EDA-enhanced Kinematic Model has the highest accuracy across the five time spans. After a χ^2 test, we found a significant correlation ($p < .01$), indicating the effectiveness of the EDA-enhanced Kinematic Model in CS prediction. Fusion data outperforms single-modality models, highlighting the importance of leveraging multiple data modalities. Physiological and kinematic data contain valuable information for CS, enhancing integrated performance. Notably, the kinematic model outperforms the EDA model by 27% when the time span is 10s before downsampling. The inferior EDA performance may be due to noise interference from motion artifacts, muscle activity, and environmental factors, impacting the accuracy and reliability of physiological signals. In contrast, kinematic data provides reliable information on body movement, less affected by noise, improving CS prediction.

6.3 Performance on different exposure time

Exposure time refers to the duration of VR immersion. In Sect. 4.2, we assumed a constant CS level throughout the four-minute VR immersion. However, users may gradually experience CS during the immersion. To analyze the relationship between exposure time and CS, we truncated the four-minute experimental time, removed initial data, and retrained the Kinematic model and EDA model. As mentioned in Sect. 6.1, using a 20s time span balances information and training data. We used it to study the exposure time-CS relationship. Truncating the first *n* time segments (20s each) yields data from *n*20 seconds to 240 seconds. For instance, for $n = 3$, we obtained data from 60s to 240s per sample, with a total size of $9 * 157 = 1413$. To ensure consistent training sample sizes, we randomly sampled to make them all $7 * 157 = 1099$. Table 6 presents prediction results. Increasing *n* improved the Kinematic model's performance. Accuracy increased from 0.908 (when $n = 1$) to 0.925 (when $n = 5$).

Results Analysis. From Table 6, interestingly, after conducting Pearson correlation tests between *n* and accuracy, we found that the Kinematic model showed a significant correlation ($r = .978, p < .01$) and its performance gradually improved as *n* increased. In contrast, the EDA model showed no significant correlation ($r = -.130, p = .834$) and its performance fluctuated as *n* increased. High accuracy and correlation coefficient suggest that kinematic data maintains a high level of correlation with the CS throughout the entire four minutes of VR exposure. This correlation gradually increases as the exposure time lengthens. In contrast, EDA signals showed a weak correlation with the CS, possibly due to signal noise. Overall, there is a certain correlation between exposure time and CS. Users are likely to experience a low CS level at the beginning of exposure, with the CS level gradually increasing as the exposure time lengthens. The experiment is designed as a simple case for navigation and further research is needed to reveal the relationship between exposure time and CS.

7 DISCUSSION

In this section, we discuss the feasibility of our proposed prediction approaches. We will start by examining the prediction performance with EDA data, which is discussed in Sect. 7.1. Then, in Sect. 7.2, we demonstrate the benefits of our EDA-enhanced kinematic model, such as its reliance only on integrated sensors within HMDs and its ability to take into account physiological features. Finally, we discuss the limitations of our work in Sect. 7.3.

7.1 Prediction performance with EDA data

Our EDA feature extraction significantly improved prediction performance using EDA data. While some studies have used multimodal data, including EDA, few have focused solely on EDA for prediction. Compared to these studies, our EDA model achieved higher accuracy. For instance, Hadadi *et al.* employed topological data analysis (TDA) techniques on physiological data, including EDA, TEM, and BVP, achieving an SVM classifier $F1$ score of 0.71 on binary classification [22]. Similarly, we used TDA to extract EDA features, which are included in our EDA numerical features. Our EDA-only model achieved a maximum accuracy of 0.697 and an $F1$ score of 0.690, which is impressive given the four-class classification task. Islam *et al.* collected physiological data, including HR, EDA, and breathing rate, and calculated time series properties (e.g., max, min, avg, percentage change), achieving a high accuracy of 97.44% for CS detection using a convolutional long short-term memory three-class classifier [26]. Their high accuracy can be attributed, in part, to their different data processing methods. They divided the data into 120-second time spans with a rolling window of 1 second, resulting in multiple overlapping time series data, which differs from our dataset.

7.2 Benefits of EDA-enhanced Kinematic Model

Usage of integrated sensors within HMDs. Our motivation is to use fewer data dimensions collected from limited sensors in the headset to achieve higher accuracy. While many studies for CS prediction involving physiological signals often required external biosensors, our EDA-enhanced kinematic model only counted on integrated sensors within HMDs in a non-intrusive manner. Islam *et al.* proposed a multimodal deep fusion approach for forecasting CS onset using the user's physiological, head-tracking, and eye-tracking data, which had to use eye-tracker and biosensors [25]. Similarly, Qu *et al.* utilized real-time physiological signals such as EDA and electrocardiogram and position and bone rotation data of users' virtual avatars to train an LSTM Attention neural network model [47]. Although these models demonstrated remarkable prediction performance, the reliance on external sensors limited user freedom and practical usage. In contrast, our model only needed kinematic data from integrated sensors in HMDs, making it easier to be deployed in existing HMDs.

Reflect individual difference. Despite utilizing only the integrated sensors within HMDs like other studies, our EDA-enhanced kinematic model learned the physiological embedding from kinematic data to reflect individual differences. This gave our model a more comprehensive range of perspectives, improving its performance to predict CS. Islam *et al.* proposed a deep fusion method that uses eye-tracking and head-tracking data to predict CS with an accuracy of 87.77% and an RMSE of 0.51 [24]. In addition, Du *et al.* extracted saliency, optical flow, and disparity features from videos to determine the factors that contribute to simulator sickness. Their 3D CNN model showed improved performance in terms of Root Mean Square Error (RMSE) and Pearson Linear Correlation Coefficient. [14]. In contrast, our model specifically learns individual differences only from kinematic data, setting it apart from the above studies.

Real-time Performance and applicable scenario. Our EDA-enhanced Kinematic model stands out by demonstrating remarkable real-time performance, achieving an impressive accuracy of 97.8% using only 10 seconds of historical kinematic data, as shown in Table 4. In comparison, Islam *et al.* proposed a model that can forecast cybersickness onset 60 seconds in advance [25]. Jin *et al.* presented a machine learning approach to predict cybersickness level with the last 30-second data [27].

Once the user wears the VR headset, our EDA-enhanced Kinematic model utilizes kinematic data collected from the headset to predict the occurrence of cybersickness. This prediction is essential as it enables the implementation of appropriate prevention strategies. By identifying whether a person is experiencing cybersickness, interventions such as reducing the field of view [16,61] or adjusting navigation speed [2] can be promptly employed.

7.3 Limitations

There are three main limitations in our study.

Data Collection and Processing. For data collection, we note that data from integrated sensors in HMDs are not limited to motion and head-trackng but also include video and eye-tracking data, providing opportunities for further research in a multimodal manner [14,27]. We also find that our EDA model did not perform well during the cross-validation compared with the kinematic model, as the physiological signal is subject to noise and artifacts. Compared to motion and head-tracking data, EDA data contains more noisy information; thus, high-quality EDA data for training are encouraged.

In terms of data processing, although our EDA-enhanced kinematic model achieved the best performance by using head-tracking and motion data, the accuracy of our model was limited by the ground-truth construction, which assumes that the SSQ score collected after immersion denotes the CS level during the entire process. Therefore, we suggest that future studies use Fast Motion Scale (FMS) to collect more granular subjective measures of CS during immersion.

Model Architecture. In comparison to previous studies [24, 25, 31], our proposed approach achieves significant performance improvement using only LSTM and Dense layers. However, we notice that more complex networks, such as AttnLSTM and DeepTCN, have been employed by other researchers to classify the severity of CS. Therefore, we suggest further exploring such models to improve the performance of CS prediction.

Generality. According to the sensory conflict theory, cybersickness is aroused by conflicting senses, which are aroused from anything that can stimulate visual-vestibular mismatch. So to enhance generality in our experiments, we enabled users to set their speed freely, thus providing different level of conflict and a way to generalize the findings beyond constant speed settings. However, we acknowledge that the navigation task employed in our study was a rudimentary simulation environment and that more complex simulations would be a necessary. Future research will contribute to a more comprehensive understanding of the effectiveness of our proposed method in predicting CS across different simulation environments.

8 CONCLUSION

In this work, we propose several prediction models for different data modalities. Among these models, the novel EDA-enhanced kinematic model achieves the highest accuracy of 97.8% in CS prediction using only kinematic data, which learns physiological embedding with encoded physiological representation from the EDA model. To the best of our knowledge, this is the state-of-the-art performance for predicting CS using motion and head-tracking data. Note that our model can be practically applied to affordable consumer-level HMDs because it only relies on integrated sensors in HMDs. Our prediction model will allow future research to develop adaptive CS preventive measures to mitigate CS before the onset.

REFERENCES

[1] B. Arcioni, S. Palmisano, D. Apthorp, and J. Kim. Postural stability predicts the likelihood of cybersickness in active hmd-based virtual reality. *Displays*, 58:3–11, 2019.

[2] F. Argelaguet and M. Maignant. Giant: stereoscopic-compliant multi-scale navigation in ves. In *Proceedings of the 22nd acm conference on virtual reality software and technology*, pp. 269–277, 2016.

[3] W. Bles, J. E. Bos, B. De Graaf, E. Groen, and A. H. Wertheim. Motion sickness: only one provocative conflict? *Brain research bulletin*, 47(5):481–487, 1998.

[4] J. E. Bos, W. Bles, and E. L. Groen. A theory on visually induced motion sickness. *Displays*, 29(2):47–57, 2008.

[5] P. Caserman, A. Garcia-Agundez, A. Gámez Zerban, and S. Göbel. Cybersickness in current-generation virtual reality head-mounted displays: systematic review and outlook. *Virtual Reality*, 25(4):1153–1170, 2021.

[6] E. Chang, H. T. Kim, and B. Yoo. Virtual reality sickness: a review of causes and measurements. *International Journal of Human–Computer Interaction*, 36(17):1658–1682, 2020.

[7] E. Chang, H. T. Kim, and B. Yoo. Predicting cybersickness based on user's gaze behaviors in hmd-based virtual reality. *Journal of Computational Design and Engineering*, 8(2):728–739, 2021.

[8] J.-R. Chardonnet, M. A. Mirzaei, and F. Mérienne. Visually induced motion sickness estimation and prediction in virtual reality using frequency components analysis of postural sway signal. In *International conference on artificial reality and telexistence eurographics symposium on virtual environments*, pp. 9–16, 2015.

[9] Y.-C. Chen, X. Dong, J. Hagstrom, and T. A. Stoffregen. Control of a virtual ambulation influences body movement and motion sickness. In *BIO Web of Conferences*, vol. 1, p. 00016. EDP Sciences, 2011.

[10] C. G. Courtney, M. E. Dawson, A. M. Schell, A. Iyer, and T. D. Parsons. Better than the real thing: Eliciting fear with moving and static computer-generated stimuli. *International Journal of Psychophysiology*, 78(2):107–114, 2010.

[11] M. S. Dennison, A. Z. Wisti, and M. D'Zmura. Use of physiological signals to predict cybersickness. *Displays*, 44:42–52, 2016.

[12] P. DiZio and J. R. Lackner. Circumventing side effects of immersive virtual environments. *Advances in human factors/ergonomics*, 21:893–896, 1997.

[13] X. Dong and T. A. Stoffregen. Postural activity and motion sickness among drivers and passengers in a console video game. In *Proceedings of the human factors and ergonomics society annual meeting*, vol. 54, pp. 1340–1344. SAGE Publications Sage CA: Los Angeles, CA, 2010.

[14] M. Du, H. Cui, Y. Wang, and H. B.-L. Duh. Learning from deep stereoscopic attention for simulator sickness prediction. *IEEE Transactions on Visualization and Computer Graphics*, 29(2):1415–1423, 2021.

[15] T. Feigl, D. Roth, S. Gradl, M. Wirth, M. E. Latoschik, B. M. Eskofier, M. Philippsen, and C. Mutschler. Sick moves! motion parameters as indicators of simulator sickness. *IEEE transactions on visualization and computer graphics*, 25(11):3146–3157, 2019.

[16] A. S. Fernandes and S. K. Feiner. Combating vr sickness through subtle dynamic field-of-view modification. In *2016 IEEE symposium on 3D user interfaces (3DUI)*, pp. 201–210. IEEE, 2016.

[17] A. Garcia-Agundez, C. Reuter, H. Becker, R. Konrad, P. Caserman, A. Miede, and S. Göbel. Development of a classifier to determine factors causing cybersickness in virtual reality environments. *Games for health journal*, 8(6):439–444, 2019.

[18] A. M. Gavgani, K. V. Nesbitt, K. L. Blackmore, and E. Nalivaiko. Profiling subjective symptoms and autonomic changes associated with cybersickness. *Autonomic Neuroscience*, 203:41–50, 2017.

[19] J. F. Golding. Phasic skin conductance activity and motion sickness. *Aviation, space, and environmental medicine*, 1992.

[20] J. F. Golding. Motion sickness susceptibility questionnaire revised and its relationship to other forms of sickness. *Brain research bulletin*, 47(5):507–516, 1998.

[21] C. Guo, J. Ji, and R. So. Could okan be an objective indicator of the susceptibility to visually induced motion sickness? In *2011 IEEE Virtual Reality Conference*, pp. 87–90. IEEE, 2011.

[22] A. Hadadi, C. Guillet, J.-R. Chardonnet, M. Langovoy, Y. Wang, and J. Ovtcharova. Prediction of cybersickness in virtual environments using topological data analysis and machine learning. 2022. doi: 10.3389/frvir.2022.973236

[23] IMotions. Galvanic skin response—the complete pocket guide, 2017.

[24] R. Islam, K. Desai, and J. Quarles. Cybersickness prediction from integrated hmd's sensors: A multimodal deep fusion approach using eye-tracking and head-tracking data. In *2021 IEEE international symposium on mixed and augmented reality (ISMAR)*, pp. 31–40. IEEE, 2021.

[25] R. Islam, K. Desai, and J. Quarles. Towards forecasting the onset of cybersickness by fusing physiological, head-tracking and eye-tracking with multimodal deep fusion network. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 121–130. IEEE, 2022.

[26] R. Islam, Y. Lee, M. Jaloli, I. Muhammad, D. Zhu, P. Rad, Y. Huang, and J. Quarles. 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)*, pp. 400–411. IEEE, 2020.

[27] W. Jin, J. Fan, D. Gromala, and P. Pasquier. Automatic prediction of cybersickness for virtual reality games. In *2018 IEEE Games, Entertainment, Media Conference (GEM)*, pp. 1–9. IEEE, 2018.

[28] R. S. Kellogg, R. S. Kennedy, and A. Graybiel. Motion sickness symptomatology of labyrinthine defective and normal subjects during zero gravity maneuvers. Technical report, AEROSPACE MEDICAL RESEARCH LABS WRIGHT-PATTERSON AFB OHIO, 1964.

[29] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993.

[30] B. Keshavarz and H. Hecht. Validating an efficient method to quantify motion sickness. *Human factors*, 53(4):415–426, 2011.

[31] J. Kim, W. Kim, H. Oh, S. Lee, and S. Lee. A deep cybersickness predictor based on brain signal analysis for virtual reality contents. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10580–10589, 2019.

[32] J. Kim and S. Lee. Deep learning of human visual sensitivity in image quality assessment framework. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1676–1684, 2017.

[33] Y. Y. Kim, E. N. Kim, M. J. Park, K. S. Park, H. D. Ko, and H. T. Kim. The application of biosignal feedback for reducing cybersickness from exposure to a virtual environment. *Presence: Teleoperators and Virtual Environments*, 17(1):1–16, 2008.

[34] Y. Y. Kim, H. J. Kim, E. N. Kim, H. D. Ko, and H. T. Kim. Characteristic changes in the physiological components of cybersickness. *Psychophysiology*, 42(5):616–625, 2005.

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

[36] B. D. Lawson et al. Motion sickness symptomatology and origins., 2014.

[37] T. M. Lee, J.-C. Yoon, and I.-K. Lee. Motion sickness prediction in stereoscopic videos using 3d convolutional neural networks. *IEEE transactions on visualization and computer graphics*, 25(5):1919–1927, 2019.

[38] T. Lencioni, I. Carpinella, M. Rabuffetti, A. Marzegan, and M. Ferrarin. Human kinematic, kinetic and emg data during different walking and stair ascending and descending tasks. *Scientific data*, 6(1):309, 2019.

[39] C.-L. Liu and S.-T. Uang. A study of sickness induced within a 3d virtual store and combated with fuzzy control in the elderly. In *2012 9th International Conference on Fuzzy Systems and Knowledge Discovery*, pp. 334–338. IEEE, 2012.

[40] P. Lopes, N. Tian, and R. Boulic. Eye thought you were sick! exploring eye behaviors for cybersickness detection in vr. In *Proceedings of the 13th ACM SIGGRAPH Conference on Motion, Interaction and Games*, pp. 1–10, 2020.

[41] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, and S. H. A. Chen. NeuroKit2: A python toolbox for neurophysiological signal processing. *Behavior Research Methods*, 53(4):1689–1696, feb 2021. doi: 10.3758/s13428-020-01516-y

[42] M. Melo, J. Vasconcelos-Raposo, and M. Bessa. Presence and cyber-

sickness in immersive content: Effects of content type, exposure time and gender. *Computers & Graphics*, 71:159–165, 2018.

[43] E. Nalivaiko, S. L. Davis, K. L. Blackmore, A. Vakulin, and K. V. Nesbitt. Cybersickness provoked by head-mounted display affects cutaneous vascular tone, heart rate and reaction time. *Physiology & behavior*, 151:583–590, 2015.

[44] N. Padmanaban, T. Ruban, V. Sitzmann, A. M. Norcia, and G. 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.

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

[46] J. Plouzeau, D. Paillot, J.-R. Chardonnet, and F. Merienne. Effect of proprioceptive vibrations on simulator sickness during navigation task in virtual environment. 2015.

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

[48] J. T. Reason and J. J. Brand. *Motion sickness.* Academic press, 1975.

[49] E. Regan and A. Ramsey. Some side-effects of immersion virtual reality: The results of four immersions. *APRE report 94R012*, 1994.

[50] A. Reuten, S. Nooij, J. Bos, and J. Smeets. How feelings of unpleasantness develop during the progression of motion sickness symptoms. *Experimental Brain Research*, 239(12):3615–3624, 2021.

[51] D. Risi and S. Palmisano. Effects of postural stability, active control, exposure duration and repeated exposures on hmd induced cybersickness. *Displays*, 60:9–17, 2019.

[52] B. M. Silva and P. Fernando. Early prediction of cybersickness in virtual, augmented & mixed reality applications: a review. In *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, pp. 1–6. IEEE, 2019.

[53] K. M. Stanney, R. S. Kennedy, and J. M. Drexler. Cybersickness is not simulator sickness. In *Proceedings of the Human Factors and Ergonomics Society annual meeting*, vol. 41, pp. 1138–1142. SAGE Publications Sage CA: Los Angeles, CA, 1997.

[54] D. Van Krevelen and R. Poelman. A survey of augmented reality technologies, applications and limitations. *International journal of virtual reality*, 9(2):1–20, 2010.

[55] S. J. Villard, M. B. Flanagan, G. M. Albanese, and T. A. Stoffregen. Postural instability and motion sickness in a virtual moving room. *Human factors*, 50(2):332–345, 2008.

[56] Y. Wang, J.-R. Chardonnet, and F. Merienne. Vr sickness prediction for navigation in immersive virtual environments using a deep long short term memory model. In *2019 IEEE conference on virtual reality and 3D user interfaces (VR)*, pp. 1874–1881. IEEE, 2019. doi: 10.1109/VR .2019.8798213

[57] Y. Wang, J.-R. Chardonnet, and F. Merienne. Development of a speed protector to optimize user experience in 3d virtual environments. *International Journal of Human-Computer Studies*, 147:102578, 2021. doi: 10.1016/j.ijhcs.2020.102578

[58] Y. Wang, J.-R. Chardonnet, F. Merienne, and J. Ovtcharova. Using fuzzy logic to involve individual differences for predicting cybersickness during vr navigation. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 373–381. IEEE, 2021. doi: 10.1109/VR50410. 2021.00060

[59] C. Widdowson, I. Becerra, C. Merrill, R. F. Wang, and S. LaValle. Assessing postural instability and cybersickness through linear and angular displacement. *Human factors*, 63(2):296–311, 2021.

[60] Y. Yokota, M. Aoki, K. Mizuta, Y. Ito, and N. Isu. Motion sickness susceptibility associated with visually induced postural instability and cardiac autonomic responses in healthy subjects. *Acta oto-laryngologica*, 125(3):280–285, 2005.

[61] D. Zielasko, A. Meißner, S. Freitag, B. Weyers, and T. W. Kuhlen. Dynamic field of view reduction related to subjective sickness measures in an hmd-based data analysis task. In *Proc. of IEEE VR Workshop on Everyday Virtual Reality*, vol. 2, 2018.