

RESEARCH ARTICLE

Eyes on me: Investigating the role and influence of eye-tracking data on user modeling in virtual reality

Dayoung Jeong ¹, Mingon Jeong ¹, Ungyeon Yang ², Kyungsik Han ^{1*}

1 Department of Artificial Intelligence, Hanyang University, Seoul, Republic of Korea, **2** Electronics and Telecommunications Research Institute, Daejeon, Republic of Korea

* kyungsikhan@hanyang.ac.kr



OPEN ACCESS

Citation: Jeong D, Jeong M, Yang U, Han K (2022) Eyes on me: Investigating the role and influence of eye-tracking data on user modeling in virtual reality. PLoS ONE 17(12): e0278970. <https://doi.org/10.1371/journal.pone.0278970>

Editor: Jude Hemanth, Karunya Institute of Technology and Sciences, INDIA

Received: August 31, 2022

Accepted: November 24, 2022

Published: December 29, 2022

Copyright: © 2022 Jeong et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data underlying the results presented in the study are available from: <https://github.com/Hanyang-HCC-Lab/EyesOnMe.git>.

Funding: All authors received the following fundings for this work: Institute of Information & Communications Technology Planning & Evaluation: IITP-2022-2018-0-01431 Institute of Information & Communications Technology Planning & Evaluation: 2020-0-01373 National Research Foundation: 2021M3A9E4080780 The funders had no role in study design, data collection

Abstract

Research has shown that sensor data generated by a user during a VR experience is closely related to the user's behavior or state, meaning that the VR user can be quantitatively understood and modeled. Eye-tracking as a sensor signal has been studied in prior research, but its usefulness in a VR context has been less examined, and most extant studies have dealt with eye-tracking within a single environment. Our goal is to expand the understanding of the relationship between eye-tracking data and user modeling in VR. In this paper, we examined the role and influence of eye-tracking data in predicting a level of cybersickness and types of locomotion. We developed and applied the same structure of a deep learning model to the multi-sensory data collected from two different studies (cybersickness and locomotion) with a total of 50 participants. The experiment results highlight not only a high applicability of our model to sensor data in a VR context, but also a significant relevance of eye-tracking data as a potential supplement to improving the model's performance and the importance of eye-tracking data in learning processes overall. We conclude by discussing the relevance of these results to potential future studies on this topic.

1 Introduction

One method that has attracted significant attention in recent VR research involves using the data collected from the sensor equipment worn by the user as a way to analyze various aspects of their state or behavior during the VR experience. For example, studies have examined changes in the user's eye movement (e.g., focus point, gaze movement, pupil diameter) based on the data collected from an eye-tracking device, those in their head or body movements collected from head-mounted displays (HMD) or trackers, and those in their electrodermal activity (EDA) collected from wearable devices [1–6]. Studies have also attempted to understand the relationships between the collected sensor data and user-driven factors, such as stress, engagement, satisfaction, and conversation quality, through statistical analysis and model development [7–11]. Findings from these studies not only help determine whether the goal of

and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

a given VR system has been achieved but also provide design insights that can be used to improve the user experience and the system itself.

When collecting data during a VR experience, it is especially important to measure and monitor human states or behaviors as unobtrusively as possible. Hence, collecting signals generated from standard devices (e.g., HMD) or from devices that do not cause the user discomfort is essential. It is worth noting that many studies have been investigating eye-tracking primarily because visuals are the first channel of information collection from a human [12, 13]. The use of eye-tracking data in research is also increasing because eye-tracking devices have become more affordable in recent years. Eye-tracking is key to human processing of visual information as well as measuring attention, interest, and arousal. In other words, eye-tracking data can be used to digitize the ways in which people communicate with computers, allowing researchers to identify and analyze patterns of visual attention of individuals as they perform specific tasks (e.g., reading, searching, driving, scanning an image). Perhaps due to such important roles, eye-tracking data has been used for different purposes, including intention recognition [14], image classification [15–18], depression recognition [19], and attention prediction [20].

Another group of research has focused on developing prediction models. Studies have employed machine learning algorithms (e.g., Support Vector Machine (SVM), Logistic Regression, Random Forest, k -Nearest Neighbors, Na ve Bayes), and more recent studies have begun to utilize deep learning methods (Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM)) to extract deep features of the data and construct a more complicated learning models [18, 21, 22]. These studies have utilized various types of features extracted from sensor signals as the data for their prediction models and demonstrated high possibilities of predicting users' states or behaviors through their models.

In this work, we aimed to achieve the same goal as prior studies, which is to measure and model a user's state and behavior based on sensor signals, particularly in terms of eye-tracking data in the context of VR. Despite its demonstrated role in depicting human characteristics in prior studies, the adaption and study of eye-tracking data in VR has been under-studied. Another limitation is a lack of consideration of measuring and analyzing a prediction model from multiple case studies, which possibly fails to suggest more generalizable results or insights. To fill these research gaps, this paper presents two independent case studies of cybersickness [23, 24] and locomotion [25], both of which are closely related to the VR user experience. Our two case studies leverage various sensor signals generated from an HMD, a tracker, and a physiological device (e.g., head, eye, and ankle movement, electrodermal activity (EDA)) for the development of two deep learning models that classify the level of cybersickness and the type of locomotion, respectively. The model was designed to learn not only characteristics of the individual sensor data modalities through an *attention* technique (i.e., calculating different degrees of the importance of data features during training) [26] but also the *temporal sequence* of the sensor data. The same model architecture was applied to both studies as our intention was to examine a possible generalization of the role of eye-tracking data in understanding a user's state and behavior. For the same purpose, we recruited enough participants for each study comparable to prior studies [27, 28]. The cybersickness study involved 27 participants, and the locomotion study was conducted with 23.

The experiment results of two studies were quite similar. Both studies showed that the addition of eye-tracking data increased the performance of the model, which was at its highest when all sensor modalities were used. These results demonstrated the effectiveness of both our model structure and the use of eye-tracking data as a supplemental method to improve the prediction performance.

In summary, this paper makes the following contributions.

- We proposed a deep learning-based architecture for modeling a user's state and behavior (cybersickness and locomotion), based on sensor data in VR.
- We identified a role for eye-tracking data signals in modeling users in VR.

2 Related work

2.1 Research using eye-tracking data

Eye-tracking data is generated based on the movement of the pupils. Some devices that can take these measurements include Tobii (<https://www.tobii.com/>) and Pupil labs (<https://pupil-labs.com/>). Recent HMD devices, such as HTC VIVE Pro Eye (<https://www.vive.com/us/product/vive-pro-eye/overview/>) and FOVE (<https://fove-inc.com/>), provide eye-tracking components at affordable prices, making the utilization of eye-tracking data in a VR environment more practical and relevant for researchers and practitioners.

Research utilizing eye-tracking data has been actively conducted in non-VR contexts. One of the topics that has been widely studied is image classification. To name a few, Ahmed and Noble [15] used eye fixations to classify and acquire the image frames of head, Zhou et al. [29] used fixations and region of interest (ROI), Karessli et al. [16] used gaze points, Saab et al. [18] used gaze data, Roy et al. [17] used eye fixations, fixation duration, pupil diameter, and polar moments to develop a cognitive model for ambiguous image classification. Other classification topics include intention recognition [14], attention prediction [20], depression recognition [19], and reading pattern detection [30]; similar types of eye-tracking features were used.

Eye-tracking in a VR context also provides researchers with opportunities [31] to accurately control the VR environment through ROI settings, which helps enhance the user experience by providing a more thorough understanding of the interactions between the user and the objects inside the VR environment [1–3, 32, 33]. For example, Lahiri et al. [3] developed a VR system that supports social skill improvement for people with autism. Pfeiffer et al. [33] studied the shopping motives of users in a VR environment. While eye-tracking features are identical regardless of whether they were collected and studied in non-VR or VR conditions, differences in the environment from which they were collected are quite different. VR is a virtual world that can be manipulated depending on its goals or tasks; as such, the characteristics and interpretations of sensor signals associated with users' experiences and interactions in such an environment may be different from those in a non-VR environment. In this sense, we aim to examine the role of eye-tracking data in VR contexts and discuss ways to utilize it to enhance user experience.

2.2 Research on prediction model with eye-tracking data

Most studies described in the previous section defined various classification models and their performances, highlighting the role and influence of eye-tracking data [34, 35] on characterizing users and tasks in specific scenarios. Some of the most widely used eye-tracking features include pupil diameter, pupil position, eye fixations, blinking, and gaze point. Various types of classification algorithms, including Support Vector Machine (SVM), *k*-Nearest Neighbors (kNN), Decision Trees, Logistic Regression, and Na ve Bayes have been employed to track this data. SVM appears to be the most promising [34, 35]. Recent studies have begun to employ deep learning algorithms [18, 36], such as CNN and LSTM, but the number of deep learning-based modeling research is smaller than that of machine learning-based one. As deep learning models have shown greater performances than machine learning models in many domains, we expect to see more studies on learning user characteristics through deep

learning algorithms in VR domains. Given that a large volume of sensor data is generated in VR, deep learning algorithms have the potential to play a key role in modeling and understanding users.

The goal of our research is also to utilize eye-tracking data to model a user's state or behavior as we found that the utilization of eye-tracking data in a VR context is quite understudied. In this paper, we present a study that offers additional insights into the way in which eye-tracking data can be collected and used. We present a deep-learning model using *attention*, one of the learning mechanisms employed to calculate different degrees of the importance of data features during training. Our model also considers *temporal sequences* of the data in order to more accurately capture a user's state and behavior. Furthermore, we present findings from the application of the model to two specific case studies (one is cybersickness and the other is locomotion). The demonstration of the use of our model in two different case studies also highlights the robustness of our proposed multimodal, attention-based deep learning model in terms of its potential use and application in other VR/AR scenarios as well.

3 Case study backgrounds

Fig 1 illustrates the overall procedure, which was the same for both studies. Two studies collected different types of data to define their participants' experiences. Eye-tracking, head, and physiological data were collected from the cybersickness study while eye-tracking, head, waist, and ankle data were collected from the locomotion study. The model learning and analysis phases were the same for both studies. Our study was reviewed and approved by the internal institutional review board at the authors' university (Hanyang University, HYUIRB-202209-003). We recruited participants over the age of 19 and obtained written informed consent from all participants in the experiment.

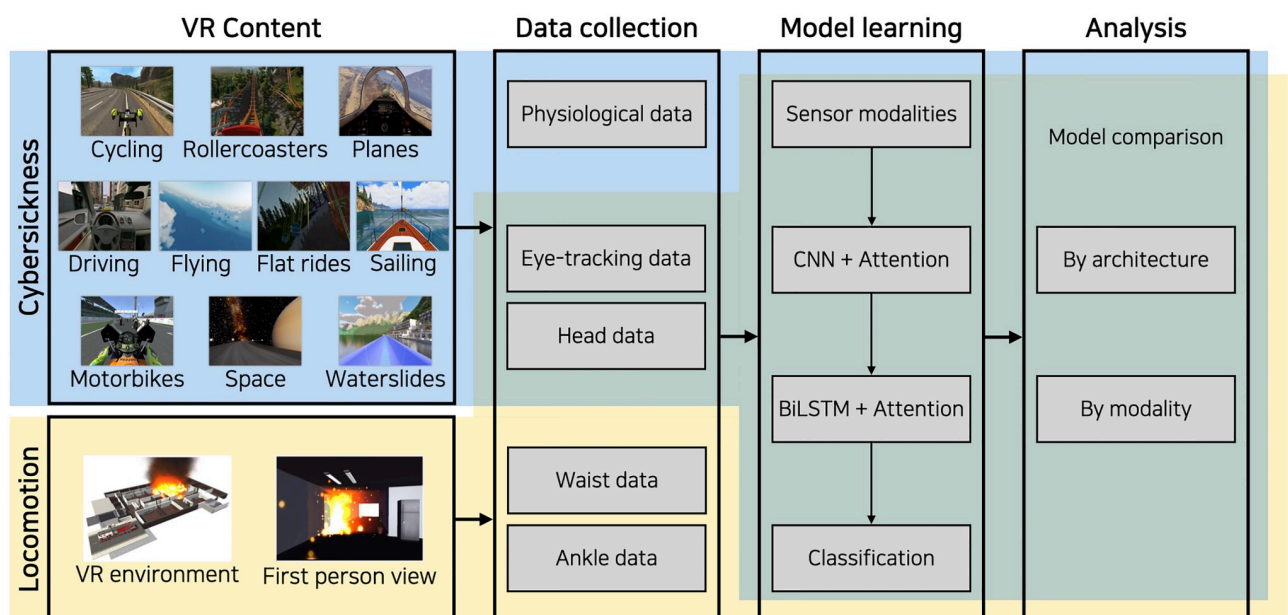


Fig 1. The study procedure. Each study has its own VR content, providing participants with the proper environment for cybersickness or locomotion. We collected eye-tracking, head, and physiological data from the cybersickness study and collected eye-tracking, head, waist, and ankle data from the locomotion study. We used the same model architecture and analysis methods for both studies.

<https://doi.org/10.1371/journal.pone.0278970.g001>

3.1 VR sickness (cybersickness)

3.1.1 Study background. Cybersickness is one of the main degrading factors that of the VR/AR experience. One possible way to alleviate this involves developing a technology that can predict and respond to cybersickness based on data collected from the environment. Various body-related data (e.g., eye tracking, head movement, body movement) from an HMD or a tracker as well as biometric data (e.g., skin conductance, heart rate) from additional wearable sensing devices can be used to determine each factor's association with cybersickness. As more effective and advanced artificial intelligence algorithms that learn the characteristics of time-series sensor data continue to develop, it is expected that sensor signals can be utilized in cybersickness prediction.

3.1.2 Study procedure. As individuals may feel different levels of cybersickness, we prepared videos in various topics. Thus, we prepared 20 VR videos with 10 categories (each category comprised two videos), namely cycling, driving, flat rides, flying, motorbikes, planes, roller coasters, sailing, space travel, and water slides (as shown in Fig 1).

We used the HTC VIVE Pro Eye and the Empatica E4 wristband (<https://www.empatica.com/research/e4/>) for the experiment and played the VR videos in VR environments (made with Unity 3D) in a Windows 10 PC with Intel Core i7 and GeForce RTX 2070.

We recruited 27 participants via university bulletin boards and word-of-mouth. The experiment consisted of two phases: (1) answering a survey and (2) watching VR videos. Before the experiment began, we explained the goal and procedure of our study to each participant and let them know that they could opt-out anytime. We then obtained informed consent from the participants.

First, each participant was informed of the goal and procedure of the study and asked to answer demographic questions (age, gender, and VR experience) and to complete the Motion Sickness Susceptibility Questionnaire (MSSQ) [37] before starting the experiment. The mean age was 26.2 (SD = 3.3) and 16 participants were male and 11 were female. 19 of the participants had previous VR experience, and the VR experience question offers yes or no answer options, following [38, 39]. MSSQ is designed to measure the user's usual motion sickness level. It contains nine questions (with a five-point Likert scale) about whether a subject feels motion sickness from various riding conditions (e.g., cars, trains, swings). The higher the sum of the responses, the more likely the subject is to feel motion sickness. The average MSSQ score of our study participants was 10.0 (SD = 8.8), which is similar to those in [37] (mean: 12.9, SD = 9.9 from 148 participants) and [40] (mean: 9.8, SD = 7.9 from 12 participants). This means that our study participants were not different from those in other studies regarding MSSQ. Lastly, the participants were instructed to describe their condition upon experiencing severe cybersickness and were again assured that they could end their participation at any time. The participants were also provided with enough time to become familiar with the HMD and the E4 wristband.

Second, each participant was asked to sit on a chair and watch a VR video in a position that was comfortable for them (Fig 2-left). The video-watching phase consisted of four sessions. Five VR videos were played during each session, with each video running for 45 seconds, and the order of five videos was randomly chosen from different categories. Given that previous cybersickness studies used the videos that ran for around 30 seconds [41, 42], we believe that the length of the videos in our study was sufficient. The participants were given a 15-second break at the end of each video and a 7-minute break at the end of each session to minimize the effect of their experience in the previous video before watching the next one, following [27, 41]. They were also offered additional time to rest between videos and sessions. It took around 46 minutes per participant to complete the study. The

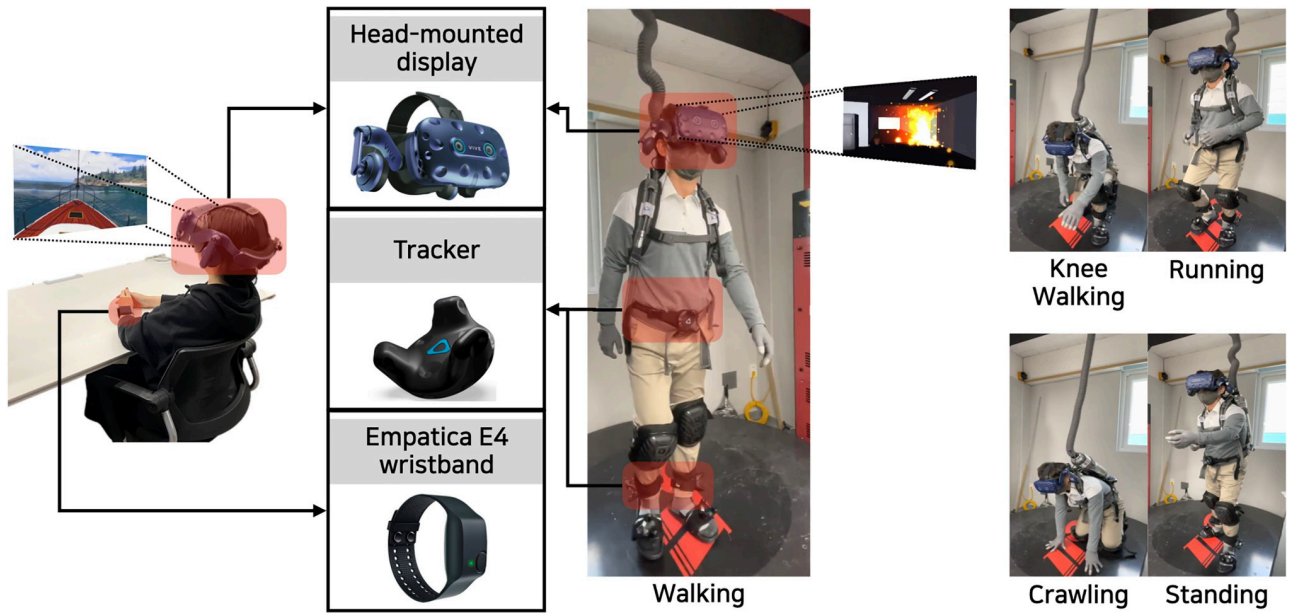


Fig 2. Data collection. Left: cybersickness study, Right: locomotion study.

<https://doi.org/10.1371/journal.pone.0278970.g002>

participants who completed the experiment received a \$10 gift card for their time and participation.

It is important to note that the participants may have had different degrees of inherent motion sickness. The representative scales for measuring cybersickness level are the Simulator Sickness Questionnaire (SSQ) [43] and the Fast Motion Sickness scale (FMS) [44]. The SSQ responds to 16 questions related to cybersickness symptoms, and each symptom is evaluated by dividing it into three factors (oculomotor, disorientation, and nausea). On the other hand, the FMS has a simpler evaluation method than SSQ. The user needs to choose a number between 0 (no sickness at all) and 20 (frank sickness). Prior studies have proven that the responses from the SSQ and from the FMS are highly correlated; thus many studies have used either question method depending on the design and goal of their experiment [27, 41, 45–51]. In this work, we employed the FMS which has the advantage of taking quick responses about each video. The participants were asked to indicate their level of cybersickness after watching each VR video via the FMS. In accordance with prior research [27, 47], we defined the level of cybersickness based on the quartiles which were determined based on the data from all participants. In our data, the first quartile (Q_1) was 1.0, the second quartile (Q_2) was 4.0, and the third quartile (Q_3) was 7.0. The data was labeled as follows.

$$Cybersickness\ level = \begin{cases} \text{None} & \text{if } 0 \leq FMS \leq Q_1 \\ \text{Low} & \text{if } Q_1 < FMS \leq Q_2 \\ \text{Moderate} & \text{if } Q_2 < FMS \leq Q_3 \\ \text{High} & \text{if } Q_3 < FMS \leq 20 \end{cases} \quad (1)$$

Eye-tracking and head movements were collected using the HMD and physiological data was collected using the E4 wristband, both were set to a frequency of 30Hz. The eye-tracking data consisted of 23 features, including gaze direction (x,y,z) with both eyes integrated as well

Table 1. Data features used in each study. Two studies used the same features from the HMD. The physiological sensor was used in the cybersickness study, and the features from the tracker were used in the locomotion study.

| Data device | Cybersickness | Locomotion |
|--------------|--|--|
| HMD | <ul style="list-style-type: none"> • head movement and orientation (x,y,z) • gaze direction (x,y,z) with both eyes integrated • gaze direction and origin (x,y,z) • pupil diameter, pupil position (x,y) • eye openness <p>presented from right and left eyes</p> | |
| Tracker | - | <ul style="list-style-type: none"> • waist movement and orientation • ankle movement and orientation <p>all presented by x,y,z values</p> |
| E4 wristband | electrodermal activity | - |

<https://doi.org/10.1371/journal.pone.0278970.t001>

as gaze direction (x,y,z), gaze origin (x,y,z), pupil diameter, pupil position (x,y) and eye openness for each eye. The head movement data was made up of six features, including position and rotation data (x,y,z). We considered a single EDA feature as a physiological signal that is frequently done in VR research [27, 45, 52] (See Table 1). Overall, we collected a total of 24,300 seconds of data: 27 (number of participants) \times 20 (number of videos) \times 45 (one video time).

3.2 VR locomotion

3.2.1 Study background. Given that locomotion is body reactions, VR environments need to be able to support it if the corresponding VR scenarios involve user movement (e.g., walking, running, moving around). Locomotion research can be broadly classified into walking-, steering-, selection-, and manipulation-based techniques [53]; of these, the walking-based technique has been studied the most. In this technique, Walking-in-Place (WiP), which refers to step-like movement while remaining stationary [25], is best known. The user's limb movements can be tracked, or stepping and treadmill-like input devices can be used, similar to our study environment (Fig 2-right), to register the step-like movements and translate them into VR motion.

3.2.2 Study procedure. VR technology has been widely adopted to support virtual training in many scenarios [54–56]; as such, the context of our study of VR locomotion was firefighting training. This training involves many different types of movements that are expected to be supported by VR training; the locomotion types (walking, running, walking on one's knees, crawling, standing) considered in our study were the ones that firefighters use during their missions. Walking and running were defined based on movement speed and magnitude. Kneeling forward and crawling on one knee are actions that lower one's posture and move the body forward. Standing refers to a posture with the knee of the front leg slightly bent and the hind leg extended to support the weight. Our research team collaborated with local firefighters for the design and implementation of this VR training system, and one of their design requirements was the inclusion of such movements.

We recruited 23 participants (all firefighters) for the user study by sending official invitations to local municipal fire departments because the study was part of firefighting training system development. We believe that data collected by domain professionals (real firefighters) would increase the reproducibility of our data in other training scenarios (e.g., military

training, safety training) in which locomotion data generated from professionals is important. The participants were invited to a living laboratory organized by a research institution and given instructions on how to use the VR devices and what to do in the VR environment. The mean and SD of the participants' age were 24.3 and 2.6, respectively. All participants were male, and twelve of them had previous casual VR experience.

[Fig 1](#) (the locomotion area colored in yellow) shows the virtual environment (a case of a fire disaster in a basement of a building) that we created for data collection, and [Fig 2-right](#) shows five types of locomotion that the participants performed. The participants were asked to wear an HMD, and VIVE trackers were attached to their waist and ankles. We asked the participants to follow the prompt (e.g., "walk forward for ten seconds") on the HMD screen as naturally as possible when it appeared on the screen. After each prompt, the participants were given a five-second break before another message prompting the next locomotion type (e.g., "kneel and move forward for ten seconds") popped up on the screen. One session includes five movement types and lasted about 70 seconds in total. We planned for three sessions during the experiment. Most of the participants (74%; 17 out of 23) completed three session, but some did up to five sessions as they wanted to allow for the collection of a sufficient amount of sensor data (305 seconds per participant on average, 7,020 seconds in total) that corresponded to each movement types. All sensor data was also collected at 30Hz, same as the cybersickness study. The participants who completed the experiment received a \$10 gift card for their time and participation.

We used the x , y , and z coordinates of waist and ankle movement and orientation. Same as the cybersickness study, we used the same features of head movement, orientation, and eye-tracking data from the HMD.

4 Model development

Our proposed model consists of three key components as follows: (1) an attention-based individual subnetwork that considers the relative importance of each data modality to fuse modality-specific features, (2) an attention-based LSTM subnetwork that extracts the importance of timestep and fuse the hidden state of the Bidirectional LSTM (BiLSTM), and (3) an output layer that uses a softmax function to obtain the probabilities for classification. [Fig 3](#) illustrates the detailed architecture of our model.

4.1 Attention-based individual subnetwork

Each data modality has different feature sizes, thus, we first passed the raw features to Principal Component Analysis (PCA) [57] in order to have the same feature size for all data modalities. Individual convolutional subnetwork was used to extract features of each data modality. The frequency representation of the i^{th} sensor at time t , x_{ti} , was passed to the convolutional subnetwork. Two convolutional layers with the activation function ReLU (Rectified Linear Unit) [58] were used in the subnetwork. Then a feature vector v_{ti} was generated and used as the input to the attention fusion subnetwork.

We employed an attention network to prioritize the importance of data modalities [59]. The network takes the feature vectors of data modality [$v_{t1}, v_{t2}, \dots, v_{ti}$] as input and generates an attention weight for each modality. The hidden representation of v_{ti} was computed to get μ_{ti} with a sensor-level context vector w_1 .

$$\mu_{ti} = \tanh(W_1 v_{ti} + b_1) \quad (2)$$

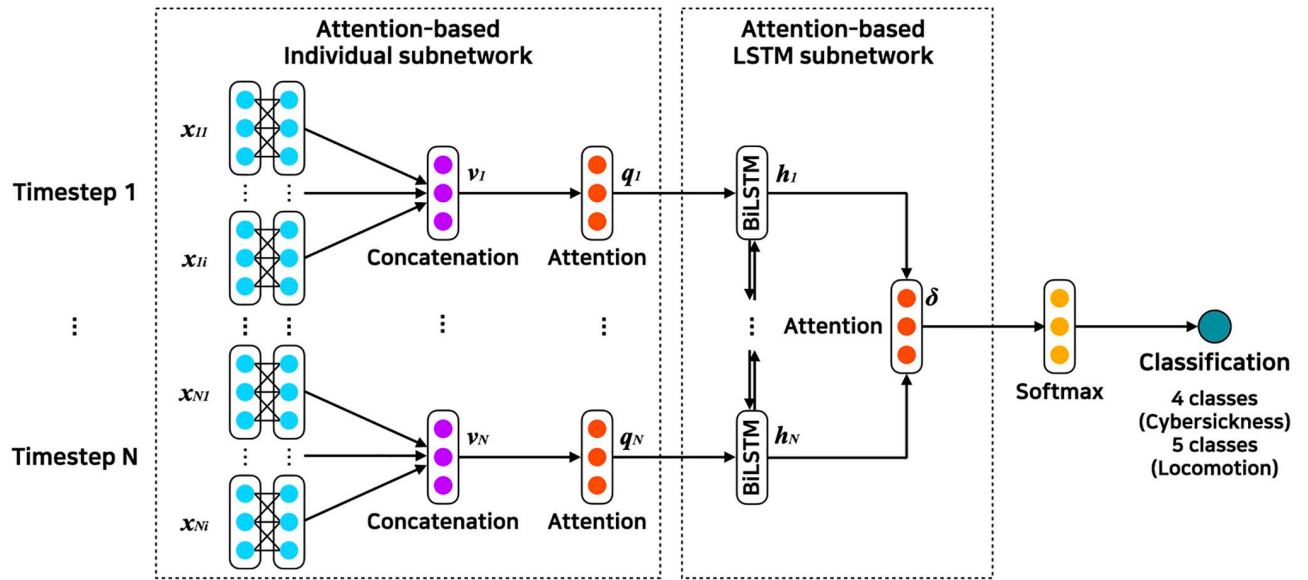


Fig 3. Model architecture used in two studies. The model mainly consists of the attention-based individual network and the attention-based LSTM subnetwork. The cybersickness study has four classes, and the locomotion study has five classes (N means time t , which is $t = 1, 2, \dots, N$).

<https://doi.org/10.1371/journal.pone.0278970.g003>

Then a normalized weight α_{ii} was computed through a softmax function.

$$\alpha_{ii} = \frac{\exp((\mu_{ii})^T w_1)}{\sum_i \exp((\mu_{ii})^T w_1)} \tag{3}$$

where W_1, b_1, w_1 are parameters of the attention network. They are randomly initialized and jointly learned through the training phase. Then the vectors of all data modalities are fused by using their attention scores as weights in order to make a uniform feature representation vector q_t .

$$q_t = \sum_i \alpha_{ii} v_{ii} \tag{4}$$

4.2 Attention-based LSTM subnetwork

The output $[q_1, q_2, \dots, q_N]$ is passed to a stacked LSTM structure [60]. LSTM is a recurrent neural network (RNN) architecture that remembers values over arbitrary intervals and deals with the vanishing gradient problem that can be encountered when training traditional RNNs.

$$f_t = \sigma(W_f \cdot [h_{t-1}, q_t] + b_f) \tag{5}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, q_t] + b_i) \tag{6}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, q_t] + b_C) \tag{7}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{8}$$

$$o_t = \sigma(W_o [h_{t-1}, q_t] + b_o) \tag{9}$$

$$h_t = o_t * \tanh(C_t) \tag{10}$$

where f_t is a forget gate, i_t is an input gate, and o_t is an output gate. C_t is a cell state. A hidden state h_t is generated at each timestep. Standard RNN (including LSTM) uses the last timestep as a single representation for the whole input sequence. This generally leads to less consideration of the front part of the sequence for classification. Because the hidden state at each timestep may show a different level of impact on classification (in our case, occurrence of cybersickness), we applied the attention mechanism again to calculate the weighted average sum of all hidden states.

Given all hidden states $H = [h_1, h_2, \dots, h_N]$ (h_t refers to a hidden state at timestep t), the attention for LSTM can be formalized as follows:

$$\gamma_t = \tanh(W_2 h_t + b_2) \tag{11}$$

$$\beta_t = \frac{\exp((\gamma_t)^T w_2)}{\sum_i \exp((\gamma_i)^T w_2)} \tag{12}$$

$$\delta = \sum_t \beta_t h_t \tag{13}$$

where w_2 is a time-level context vector, β_t is a normalized weight through a softmax function, and δ is the uniform representation of the whole sequence computed based on the sum of all hidden states. Each hidden state is updated by its attention weights. W_2, b_2, w_2 are the parameters of the attention-based LSTM subnetwork which are randomly initialized and jointly learned during the training phase. We constructed a BiLSTM model that better learns the temporal characteristics of the data. BiLSTM has been found to be more efficient than unidirectional LSTM because it considers both past and future data through an interactive network [61].

4.3 Output layer

The output of attention-based LSTM subnetwork is calculated through an output layer using a fully-connected layer and a softmax function to predict cybersickness.

$$prediction = \underset{a \in A}{\operatorname{argmax}}(\operatorname{softmax}(W_3 \cdot \delta + b_3)) \tag{14}$$

where A is the set of all data modalities. δ is transformed to the probability of each modality and the prediction result is determined by searching modality with maximum probability.

5 Experiment setup

For the experiment, we implemented our model in Pytorch and trained it on a server with GeForce RTX 2070.

First, for cybersickness, the data of each participant consists of a set S of different data modalities in the form of time series data $S_t = \{E_t, H_t, P_t\}$, where $E, H,$ and P refer to eye-tracking, head movement, and physiological data, respectively. For locomotion, $S_t = \{E_t, H_t, I_t, A_t\}$ was used, where I and A refer to waist and ankle data, respectively.

Then, each item in S_t is divided into a set of r time windows $W_t^a = \{w_{t1}^a, w_{t2}^a, \dots, w_{tr}^a\}$ of a fixed length of T_w seconds (we set $r = 30$ and $T_w = 1$). S_t is then split into the training set, validation set, and test set with the ratio of 7:1:2 by chronological order.

Second, for model training, we used cross entropy for loss function and Adam for optimizer. The model was trained up to 500 epochs, and an early stop strategy was used with 20 times of patience. The best parameters of the model was selected through parameter tuning with the validation set (batch size = 64 and learning rate = 0.001). We used 5-fold cross

Table 2. The model performance by architecture (cybersickness). Our model yielded the highest performance.

| Model architecture | F1-score |
|--|-------------|
| SVM | 0.57 |
| CNN | 0.52 |
| BiLSTM | 0.71 |
| CNN-BiLSTM | 0.68 |
| A-INV (CNN-Attention-BiLSTM) | 0.81 |
| A-BiLSTM (CNN-BiLSTM-Attention) | 0.78 |
| Ours (CNN-Attention-BiLSTM-Attention) | 0.82 |

<https://doi.org/10.1371/journal.pone.0278970.t002>

validation for model training and used the macro F1-score as the metric for the performance evaluation of the model on the testing dataset. The macro F1-score is preferred over the micro F1-score, because it can evaluate the overall performance of the multi-classifier by giving equal weight to all classes [62, 63].

We measured the effectiveness of our model by comparing it with other models. The models that we considered for the comparison include SVM, CNN, BiLSTM, and CNN-BiLSTM that have been extensively used in cybersickness prediction in prior studies [27, 28, 41, 49, 64]. Our model has three variants including **A-INV** that has the attention network only in the individual subnetwork, **A-BiLSTM** that has the attention network only in the LSTM subnetwork, and our proposed model (**Ours**) that has the attention network both in the individual and LSTM subnetworks.

6 Results

6.1 VR sickness (cybersickness)

Table 2 summarizes performance results of the model for cybersickness. Our proposed model yielded the highest performance (0.82 F1-score) compared with other models. A-INV showed a better performance than A-BiLSTM (3% difference). Overall, the application of the attentions for the sensor modalities and the BiLSTM appeared to be effective.

As a result of the ablation study, the role of eye-tracking data is found to be clear. When it is used together with other data modalities, the performance of the model significantly increased (**Table 3**). Gain indicates the increased ratio of the model performance after the addition of the eye-tracking data (shown in the parenthesis). This result is also well aligned with the result of attention weights placed on each modality (**Fig 4-top**). Eye-tracking data showed the highest result (35.7%), showing its importance during the model training in the context of cybersickness.

6.2 VR locomotion

Table 4 summarizes performance results of the model for the locomotion study. Same as the cybersickness study results, our proposed model yielded the highest performance (0.95

Table 3. Ablation study for cybersickness (addition of eye-tracking). Eye-tracking data already shows a strong relationship with cybersickness, and when it is considered with other data modalities, the performance of the model significantly increases. Gain indicates the increased ratio of the model performance after the addition of the eye-tracking data.

| Feature groups | F1-score | Gain (%) |
|------------------------------------|----------------------|----------|
| Eye | 0.68 | - |
| Head (+ Eye) | 0.56 (0.75) | 33.9 |
| Physiological (+ Eye) | 0.36 (0.78) | 116.0 |
| [All] Head + Physiological (+ Eye) | 0.57 (0.82) | 43.8 |

<https://doi.org/10.1371/journal.pone.0278970.t003>

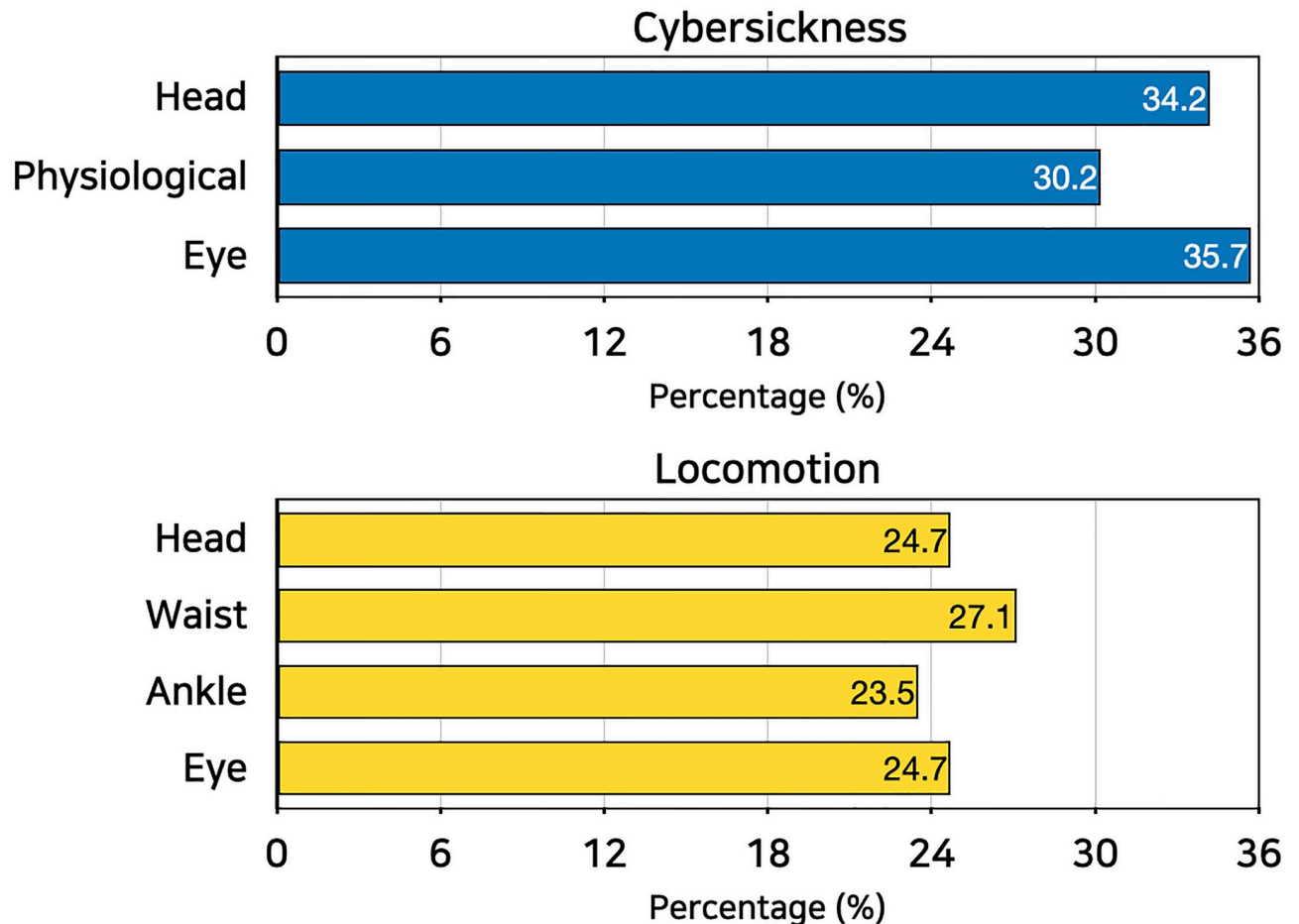


Fig 4. Attention weights of data modalities.

<https://doi.org/10.1371/journal.pone.0278970.g004>

F1-score) compared with other models. The model performances from both studies not only demonstrate the effectiveness of our model architecture (i.e., leveraging attention and temporal characteristics) but also highlight a promising research direction of modeling a user's states and behaviors in VR.

One different point between two studies is a relatively weak role of eye-tracking when only the eye-tracking data was used in modeling. This might be because the second study with locomotion involved a large amount of movements and signals from the participants' head and

Table 4. The model performance by architecture (locomotion). Our model yielded the highest performance.

| Model architecture | F1-score |
|--|-------------|
| SVM | 0.86 |
| CNN | 0.79 |
| BiLSTM | 0.85 |
| CNN-BiLSTM | 0.90 |
| A-INV (CNN-Attention-BiLSTM) | 0.93 |
| A-BiLSTM (CNN-BiLSTM-Attention) | 0.94 |
| Ours (CNN-Attention-BiLSTM-Attention) | 0.95 |

<https://doi.org/10.1371/journal.pone.0278970.t004>

Table 5. Ablation study for locomotion (addition of eye-tracking). Using eye-tracking data only shows little influence on locomotion. However, when eye-tracking data are combined with other data modalities, they served as a strong supplement to improve the performance of the model.

| Feature groups | F1-score | Gain (%) |
|------------------------------------|-------------|----------|
| Eye | 0.15 | - |
| Head (+ Eye) | 0.61 (0.80) | 31.1 |
| Waist (+ Eye) | 0.58 (0.82) | 41.3 |
| Ankle (+ Eye) | 0.67 (0.74) | 10.4 |
| Head + Waist (+ Eye) | 0.87 (0.90) | 3.4 |
| Head + Ankle (+ Eye) | 0.91 (0.89) | 2.1 |
| Waist + Ankle (+ Eye) | 0.85 (0.91) | 7.0 |
| [All] Head + Waist + Ankle (+ Eye) | 0.94 (0.95) | 1.0 |

<https://doi.org/10.1371/journal.pone.0278970.t005>

body. Thus, it appears that eye-tracking data only did show small influence; yet were used as a strong supplement to improve the learning of locomotion types. This insight is quite well presented in the ablation study. Eye-tracking data shows a comparable attention weight when it is trained together with other sensor modalities (Table 5). The results of the attention weight analysis showed a comparable importance of the eye-tracking data in learning (Fig 4-bottom).

7 Discussion

One notable aspect of our findings involves the role of multiple sensor data and especially eye-tracking data features obtained from the HMD. When our model was trained with just the HMD data, it performed reasonably similar to its best result (0.75 F1-score from the HMD data vs 0.82 from all data for cybersickness; 0.80 from the HMD data vs 0.95 from all data for locomotion). The HMD is a basic device within the VR/AR sphere, and the fact that the user's state or behavior can be well understood using only the data from this device is promising but also calls for a more thorough investigation for its extensive application to other VR scenarios. The fact that users can receive support for cybersickness or locomotion issues without wearing additional equipment also emphasizes this data's high usability.

In this work, we have proven the role of eye-tracking data. Since it has a number of features (e.g., gaze direction, gaze origin, pupil diameter, pupil position, eye openness), as the next step, we believe that it would be important to investigate the role of each feature of the eye-tracking data more extensively. Two approaches can be considered. First, we can measure the attention weight of each feature. Similar to the examination of the attention weight for each data modality, we can extract the attention weight for the features and specifically examine the importance of each feature. Second, we can measure the changes of the eye-tracking data features. Selected, important features can be used in the model to reduce the time for model training and to derive faster model results. This is also expected to enable real-time application of the model to the VR system, which might be one of the requirements for a wider application of the model to many VR scenarios and domains.

Although the role of the attention mechanism was seen to be useful in terms of improving our model performance, we found the influence of the individual subnetwork and the LSTM subnetwork were different between two studies. This indicates that assigning different weights on each subnetwork depending on a VR scenario or goal seems necessary to better learn a degree of a user's state or behavior. While this insight is interesting, we propose that it may not be generalizable and requires further verification through more experiments. Applying a different approach, such as the Fast Fourier transform [65] (presenting the changes within the energy content of a signal) may improve the characterization of sensor data and produce a

better representation thereof through attention. The resulting model would also be more specific to each user upon retraining. As the amount of data used for training for each user increases over time, it may take longer to retrain the model each time. This issue could potentially be mitigated by either applying a weighted model that prefers recency or truncating the past data and applying the most recent data. Lastly, in this work, we trained the model using all participants' data; though, it is worth considering developing a more personalized model based on only an individual's data if the amount of the data is sufficient. In summary, there are several potential directions for model development, which is the inspiration for one of our future studies. While the performance of our proposed model has been verified in both scenarios (cybersickness and locomotion), it is also important to examine how good the performance can be in real time. Some VR studies have proved the applicability of the model by comparing the prediction results with the actual values (ground-truth) over time by providing a visualization of the comparison results in real-time [27, 66]. Other studies have reported the performance of the model in a VR environment that is significantly different from the environment in which the data was collected [67]. Although we did not consider the aspect of real-time in this work, we demonstrated high performance of our model in two different VR environments, which also highlights a possible extension of our methodology to learning other types of user states or behaviors that may influence user experience in VR. We believe these are unique contributions compared to other studies, yet will take the real-time aspect of the model into account in our future study.

8 Conclusion

This paper examined the role of sensor signals, with a particular focus on eye-tracking data, in the understanding and learning of users' state or behavior. Although many studies have examined eye-tracking data in the context of user modeling, our study found that the understanding of their role within the virtual environment is somewhat lacking. To fill this gap, we conducted experiments with a total of 50 participants using multimodal sensor data generated during VR experiences in the context of cybersickness (27 participants) and locomotion (23). Our study results highlighted the high feasibility of the methods of using sensor signals to understand users and demonstrated the role that eye-tracking data could play in improving user modeling. The results also showed the effectiveness of our model architecture, which employed the attention mechanism and learned temporal sequences for individual sensor modalities. Important future studies could involve the consideration of more diverse VR situations and users in sensor-based user modeling, the comparison of eye-tracking data to other measurements of users' state or behavior, and the practical validation of the model through its application to VR systems.

Author Contributions

Conceptualization: Dayoung Jeong, Mingon Jeong, Kyungsik Han.

Data curation: Dayoung Jeong, Mingon Jeong, Ungeon Yang.

Formal analysis: Dayoung Jeong, Mingon Jeong, Ungeon Yang.

Funding acquisition: Ungeon Yang, Kyungsik Han.

Investigation: Dayoung Jeong, Mingon Jeong, Ungeon Yang, Kyungsik Han.

Methodology: Dayoung Jeong, Mingon Jeong, Ungeon Yang.

Project administration: Kyungsik Han.

Resources: Dayoung Jeong, Mingon Jeong, Kyungsik Han.

Software: Dayoung Jeong, Mingon Jeong.

Supervision: Kyungsik Han.

Validation: Dayoung Jeong, Mingon Jeong.

Visualization: Dayoung Jeong, Mingon Jeong.

Writing – original draft: Dayoung Jeong, Mingon Jeong.

Writing – review & editing: Ungyeon Yang, Kyungsik Han.

References

1. Bekele E, Zheng Z, Swanson A, Crittendon J, Warren Z, Sarkar N. Understanding how adolescents with autism respond to facial expressions in virtual reality environments. *IEEE transactions on visualization and computer graphics*. 2013; 19(4):711–720. <https://doi.org/10.1109/TVCG.2013.42> PMID: [23428456](https://pubmed.ncbi.nlm.nih.gov/23428456/)
2. Bekele E, Wade J, Bian D, Fan J, Swanson A, Warren Z, et al. Multimodal adaptive social interaction in virtual environment (MASI-VR) for children with Autism spectrum disorders (ASD). In: 2016 IEEE Virtual Reality (VR). IEEE; 2016. p. 121–130.
3. Lahiri U, Bekele E, Dohrmann E, Warren Z, Sarkar N. Design of a virtual reality based adaptive response technology for children with autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2012; 21(1):55–64. <https://doi.org/10.1109/TNSRE.2012.2218618> PMID: [23033333](https://pubmed.ncbi.nlm.nih.gov/23033333/)
4. Ranasinghe N, Jain P, Thi Ngoc Tram N, Koh KCR, Tolley D, Karwita S, et al. Season traveller: Multi-sensory narration for enhancing the virtual reality experience. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems; 2018. p. 1–13.
5. Keighrey C, Flynn R, Murray S, Murray N. A Physiology-based QoE Comparison of Interactive Augmented Reality, Virtual Reality and Tablet-based Applications. *IEEE Transactions on Multimedia*. 2020; 23:333–341. <https://doi.org/10.1109/TMM.2020.2982046>
6. Alcañiz Raya M, Chicchi Giglioli IA, Marín-Morales J, Higuera-Trujillo JL, Olmos E, Minissi ME, et al. Application of Supervised Machine Learning for Behavioral Biomarkers of Autism Spectrum Disorder Based on Electrodermal Activity and Virtual Reality. *Frontiers in human neuroscience*. 2020; 14:90. <https://doi.org/10.3389/fnhum.2020.00090> PMID: [32317949](https://pubmed.ncbi.nlm.nih.gov/32317949/)
7. Yadav M, Sakib MN, Feng K, Chaspari T, Behzadan A. Virtual reality interfaces and population-specific models to mitigate public speaking anxiety. In: 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII). IEEE; 2019. p. 1–7.
8. Salgado DP, Martins FR, Rodrigues TB, Keighrey C, Flynn R, Naves ELM, et al. A QoE assessment method based on EDA, heart rate and EEG of a virtual reality assistive technology system. In: Proceedings of the 9th ACM Multimedia Systems Conference; 2018. p. 517–520.
9. Anderson AP, Mayer MD, Fellows AM, Cowan DR, Hegel MT, Buckley JC. Relaxation with immersive natural scenes presented using virtual reality. *Aerospace medicine and human performance*. 2017; 88(6):520–526. <https://doi.org/10.3357/AMHP.4747.2017> PMID: [28539139](https://pubmed.ncbi.nlm.nih.gov/28539139/)
10. Marín-Morales J, Higuera-Trujillo JL, Greco A, Guixeres J, Llinares C, Scilingo EP, et al. Affective computing in virtual reality: emotion recognition from brain and heartbeat dynamics using wearable sensors. *Scientific reports*. 2018; 8(1):1–15. <https://doi.org/10.1038/s41598-018-32063-4> PMID: [30209261](https://pubmed.ncbi.nlm.nih.gov/30209261/)
11. Cho D, Ham J, Oh J, Park J, Kim S, Lee NK, et al. Detection of stress levels from biosignals measured in virtual reality environments using a kernel-based extreme learning machine. *Sensors*. 2017; 17(10):2435. <https://doi.org/10.3390/s17102435> PMID: [29064457](https://pubmed.ncbi.nlm.nih.gov/29064457/)
12. Krafska K, Khosla A, Kellnhofer P, Kannan H, Bhandarkar S, Matusik W, et al. Eye tracking for everyone. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 2176–2184.
13. Duchowski AT, Duchowski AT. *Eye tracking methodology: Theory and practice*. Springer; 2017.
14. Cao Y, Miura S, Kobayashi Y, Kawamura K, Sugano S, Fujie MG. Pupil variation applied to the eye tracking control of an endoscopic manipulator. *IEEE Robotics and Automation Letters*. 2016; 1(1):531–538. <https://doi.org/10.1109/LRA.2016.2521894>
15. Ahmed M, Noble JA. Fetal ultrasound image classification using a bag-of-words model trained on sonographers' eye movements. *Procedia Computer Science*. 2016; 90:157–162. <https://doi.org/10.1016/j.procs.2016.07.021>

16. Karessli N, Akata Z, Schiele B, Bulling A. Gaze embeddings for zero-shot image classification. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 4525–4534.
17. Roy AK, Akhtar MN, Mahadevappa M, Guha R, Mukherjee J. A novel technique to develop cognitive models for ambiguous image identification using eye tracker. *IEEE Transactions on Affective Computing*. 2017; 11(1):63–77. <https://doi.org/10.1109/TAFFC.2017.2768026>
18. Saab K, Hooper SM, Sohoni NS, Parmar J, Pogatchnik B, Wu S, et al. Observational supervision for medical image classification using gaze data. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2021. p. 603–614.
19. Zhu J, Wang Z, Zeng S, Li X, Hu B, Zhang X, et al. Toward Depression Recognition Using EEG and Eye Tracking: An Ensemble Classification Model CBEM. In: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE; 2019. p. 782–786.
20. Vidyapu S, Vedula VS, Bhattacharya S. Attention prediction on webpage images using multilabel classification. In: 2019 IEEE international conference on systems, man and cybernetics (SMC). IEEE; 2019. p. 1975–1980.
21. Qi L, Xu Y, Shang X, Dong J. Fusing visual saliency for material recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops; 2018. p. 1965–1968.
22. Yoo S, Jeong DK, Jang Y. The study of a classification technique for numeric gaze-writing entry in hands-free interface. *IEEE Access*. 2019; 7:49125–49134. <https://doi.org/10.1109/ACCESS.2019.2909573>
23. Davis S, Nesbitt K, Nalivaiko E. A systematic review of cybersickness. In: Proceedings of the 2014 conference on interactive entertainment; 2014. p. 1–9.
24. Clifton J, Palmisano S. Effects of steering locomotion and teleporting on cybersickness and presence in HMD-based virtual reality. *Virtual Reality*. 2020; 24(3):453–468. <https://doi.org/10.1007/s10055-019-00407-8>
25. Boletsis C. The new era of virtual reality locomotion: A systematic literature review of techniques and a proposed typology. *Multimodal Technologies and Interaction*. 2017; 1(4):24. <https://doi.org/10.3390/mti1040024>
26. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *Advances in neural information processing systems*. 2017; 30.
27. Islam R, Lee Y, Jaloli M, Muhammad I, Zhu D, Rad P, et al. Automatic detection and prediction of cybersickness severity using deep neural networks from user's physiological signals. In: 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). IEEE; 2020. p. 400–411.
28. Wang Y, Chardonnet JR, Merienne F. 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). IEEE; 2019. p. 1874–1881.
29. Zhou X, Gao X, Wang J, Yu H, Wang Z, Chi Z. Eye tracking data guided feature selection for image classification. *Pattern Recognition*. 2017; 63:56–70. <https://doi.org/10.1016/j.patcog.2016.09.007>
30. Singh M, Walia GS, Goswami A. Using Supervised Learning to Guide the Selection of Software Inspectors in Industry. In: 2018 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW). IEEE; 2018. p. 12–17.
31. Clay V, König P, Koenig S. Eye tracking in virtual reality. *Journal of Eye Movement Research*. 2019; 12(1). <https://doi.org/10.16910/jemr.12.1.3> PMID: 33828721
32. Zhu Y, Zhai G, Min X. The prediction of head and eye movement for 360 degree images. *Signal Processing: Image Communication*. 2018; 69:15–25.
33. Pfeiffer J, Pfeiffer T, Meißner M, Weiß E. Eye-tracking-based classification of information search behavior using machine learning: evidence from experiments in physical shops and virtual reality shopping environments. *Information Systems Research*. 2020; 31(3):675–691. <https://doi.org/10.1287/isre.2019.0907>
34. Lim JZ, Mountstephens J, Teo J. Emotion recognition using eye-tracking: taxonomy, review and current challenges. *Sensors*. 2020; 20(8):2384. <https://doi.org/10.3390/s20082384> PMID: 32331327
35. Lim JZ, Mountstephens J, Teo J. Eye-Tracking Feature Extraction for Biometric Machine Learning. *Frontiers in Neurorobotics*. 2021; 15. <https://doi.org/10.3389/fnbot.2021.796895> PMID: 35177973
36. Mou L, Zhou C, Zhao P, Nakisa B, Rastgoo MN, Jain R, et al. Driver stress detection via multimodal fusion using attention-based CNN-LSTM. *Expert Systems with Applications*. 2021; 173:114693. <https://doi.org/10.1016/j.eswa.2021.114693>
37. Golding JF. Motion sickness susceptibility questionnaire revised and its relationship to other forms of sickness. *Brain research bulletin*. 1998; 47(5):507–516. [https://doi.org/10.1016/S0361-9230\(98\)00091-4](https://doi.org/10.1016/S0361-9230(98)00091-4) PMID: 10052582

38. Weech S, Kenny S, Lenizky M, Barnett-Cowan M. Narrative and gaming experience interact to affect presence and cybersickness in virtual reality. *International Journal of Human-Computer Studies*. 2020; 138:102398. <https://doi.org/10.1016/j.ijhcs.2020.102398>
39. Wang Y, Chardonnet JR, Merienne F, Ovtcharova J. Using Fuzzy Logic to Involve Individual Differences for Predicting Cybersickness during VR Navigation. In: 2021 IEEE Virtual Reality and 3D User Interfaces (VR). IEEE; 2021. p. 373–381.
40. Bala P, Dionísio D, Nisi V, Nunes N. Visually induced motion sickness in 360° videos: Comparing and combining visual optimization techniques. In: 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct). IEEE; 2018. p. 244–249.
41. Kim J, Kim W, Oh H, Lee S, Lee S. A deep cybersickness predictor based on brain signal analysis for virtual reality contents. In: Proceedings of the IEEE/CVF International Conference on Computer Vision; 2019. p. 10580–10589.
42. Jin W, Fan J, Gromala D, Pasquier P. Automatic prediction of cybersickness for virtual reality games. In: 2018 IEEE Games, Entertainment, Media Conference (GEM). IEEE; 2018. p. 1–9.
43. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*. 1993; 3(3):203–220. https://doi.org/10.1207/s15327108ijap0303_3
44. Keshavarz B, Hecht H. Validating an efficient method to quantify motion sickness. *Human factors*. 2011; 53(4):415–426. <https://doi.org/10.1177/0018720811403736> PMID: 21901938
45. Dennison MS, Wisti AZ, D’Zmura M. Use of physiological signals to predict cybersickness. *Displays*. 2016; 44:42–52. <https://doi.org/10.1016/j.displa.2016.07.002>
46. Hu P, Sun Q, Didyk P, Wei LY, Kaufman AE. Reducing simulator sickness with perceptual camera control. *ACM Transactions on Graphics (TOG)*. 2019; 38(6):1–12.
47. Islam R, Desai K, Quarles J. 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). IEEE; 2021. p. 31–40.
48. Ang S, Quarles J. You’re in for a Bumpy Ride! Uneven Terrain Increases Cybersickness While Navigating with Head Mounted Displays. In: 2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE; 2022. p. 428–435.
49. Kim HG, Lim HT, Lee S, Ro YM. Vrsa net: Vr sickness assessment considering exceptional motion for 360 vr video. *IEEE transactions on image processing*. 2018; 28(4):1646–1660. <https://doi.org/10.1109/TIP.2018.2880509> PMID: 30418904
50. Islam R, Ang S, Quarles J. CyberSense: A Closed-Loop Framework to Detect Cybersickness Severity and Adaptively apply Reduction Techniques. In: 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). IEEE; 2021. p. 148–155.
51. Li G, McGill M, Brewster S, Chen CP, Anguera J, Gazzaley A, et al. Multimodal Biosensing for Vestibular Network-Based Cybersickness Detection. *IEEE Journal of Biomedical and Health Informatics*. 2021;
52. Martin N, Mathieu N, Pallamin N, Ragot M, Diverrez JM. Virtual reality sickness detection: An approach based on physiological signals and machine learning. In: 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). IEEE; 2020. p. 387–399.
53. Martinez ES, Wu AS, McMahan RP. Research Trends in Virtual Reality Locomotion Techniques. In: 2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE; 2022. p. 270–280.
54. Li C, Liang W, Quigley C, Zhao Y, Yu LF. Earthquake safety training through virtual drills. *IEEE transactions on visualization and computer graphics*. 2017; 23(4):1275–1284. <https://doi.org/10.1109/TVCG.2017.2656958> PMID: 28129163
55. Puel D. An authoring system for VR-based firefighting commanders training. *Electronic Imaging*. 2018; 2018(3):469–1.
56. Jeon SG, Han J, Jo Y, Han K. Being more focused and engaged in firefighting training: Applying user-centered design to vr system development. In: 25th ACM Symposium on Virtual Reality Software and Technology; 2019. p. 1–11.
57. Abdi H, Williams LJ. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*. 2010; 2(4):433–459. <https://doi.org/10.1002/wics.101>
58. Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. In: *Icml*; 2010.
59. Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning. *Neurocomputing*. 2021; 452:48–62. <https://doi.org/10.1016/j.neucom.2021.03.091>
60. Greff K, Srivastava RK, Koutník J, Steunebrink BR, Schmidhuber J. LSTM: A search space odyssey. *IEEE transactions on neural networks and learning systems*. 2016; 28(10):2222–2232. <https://doi.org/10.1109/TNNLS.2016.2582924> PMID: 27411231

61. Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF models for sequence tagging. arXiv preprint arXiv:150801991. 2015;.
62. Grandini M, Bagli E, Visani G. Metrics for multi-class classification: an overview. arXiv preprint arXiv:200805756. 2020;.
63. Rhim S, Lee U, Han K. Tracking and modeling subjective well-being using smartphone-based digital phenotype. In: Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization; 2020. p. 211–220.
64. Jeong D, Yoo S, Yun J. Cybersickness analysis with eeg using deep learning algorithms. In: 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE; 2019. p. 827–835.
65. Alsheikh MA, Selim A, Niyato D, Doyle L, Lin S, Tan HP. Deep activity recognition models with triaxial accelerometers. In: Workshops at the AAAI Conference on Artificial Intelligence; 2016.
66. Wang J, Liang HN, Monteiro D, Xu W, Xiao J. Real-time Prediction of Simulator Sickness in Virtual Reality Games. IEEE Transactions on Games. 2022;.
67. Monteiro D, Liang HN, Tang X, Irani P. Using trajectory compression rate to predict changes in cybersickness in virtual reality games. In: 2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). IEEE; 2021. p. 138–146.