Eye-Tracking-Based Prediction of User Experience in VR Locomotion Using Machine Learning

Hong Gao[†][1] and Enkelejda Kasneci[2]

[1]University of Tübingen, Germany
[2]Technical University of Munich, Germany

Figure 1: *The virtual environment setup for VR locomotion.*

Abstract

*VR locomotion is one of the most important design features of VR applications and is widely studied. When evaluating locomotion techniques, user experience is usually the first consideration, as it provides direct insights into the usability of the locomotion technique and users' thoughts about it. In the literature, user experience is typically measured with post-hoc questionnaires or surveys, while users' behavioral (i.e., eye-tracking) data during locomotion, which can reveal deeper subconscious thoughts of users, has rarely been considered and thus remains to be explored. To this end, we investigate the feasibility of classifying users experiencing VR locomotion into **L-UE** and **H-UE** (i.e., low- and high-user-experience groups) based on eye-tracking data alone. To collect data, a user study was conducted in which participants navigated a virtual environment using five locomotion techniques and their eye-tracking data was recorded. A standard questionnaire assessing the usability and participants' perception of the locomotion technique was used to establish the ground truth of the user experience. We trained our machine learning models on the eye-tracking features extracted from the time-series data using a sliding window approach. The best random forest model achieved an average accuracy of over 0.7 in 50 runs. Moreover, the SHapley Additive exPlanations (SHAP) approach uncovered the underlying relationships between eye-tracking features and user experience, and these findings were further supported by the statistical results. Our research provides a viable tool for assessing user experience with VR locomotion, which can further drive the improvement of locomotion techniques. Moreover, our research benefits not only VR locomotion, but also VR systems whose design needs to be improved to provide a good user experience.*

CCS Concepts
• *Computing methodologies* → *Classification and regression trees;* • *Human-centered computing* → *Empirical studies in HCI; Virtual reality;*

---

† Corresponding Author: hong.gao@informatik.uni-tuebingen.de

1. Introduction

In recent years, with the proliferation of consumer-grade head-mounted displays (HMDs), virtual reality (VR) has become increasingly integrated into entertainment and education, and thus into people's everyday lives [TGT*21, RMFW20]. VR applications, especially those for entertainment, should not only provide users with an immersive user experience but also enable the exploration of large and unlimited virtual environments in a limited physical space, such as the popular VR theme and amusement parks on Oculus Rift and Steam. This is made possible by VR locomotion techniques. VR locomotion is an essential technique that enables users to move effectively in virtual environments. It is one of the pillars of a great VR experience and has been studied extensively by designers and researchers recently due to the increasing popularity of VR-based entertainment [CMS20]. To date, a wide variety of VR locomotion techniques have been developed for different purposes and scenarios in the VR domain, and can be categorized into controller-based (e.g., joystick [LLS18], teleportation [FMF*19], dash [BMF18]) and motion-based (e.g., walking-in-place (WIP) [TFY*22], arm swing [WKMW16]) locomotion. A locomotion technique that performs the intended locomotion function should also provide a good experience for users, i.e., it should be easy to use and not impose much additional cognitive load on users performing primary tasks in VR. Therefore, when evaluating a VR locomotion technique, it is important to evaluate the user experience.

Previous literature has primarily focused on the development of new locomotion techniques for various use cases in VR applications [BSB11, vWSM*20], but in recent years more emphasis has been placed on the evaluation of locomotion techniques [CBCWS18, FSW17]. Various aspects of locomotion techniques have been evaluated, such as locomotion effectiveness and self-reported post-hoc user experience, presence, motion sickness, etc [CBCWS18, LLS18, PKR*19]. In terms of the locomotion technique itself, user experience is most valued in this research area as it provides direct feedback on the usability of the locomotion technique and how users feel about using it. This provides the designer and researcher with concrete advice on how to improve the locomotion technique. To date, most studies on VR locomotion (or other VR systems design) have assessed user experience using methods such as observational data during VR locomotion (or during interaction with VR systems), post-hoc surveys, and questionnaires [TFY*22]. However, such methods can be time-consuming, especially for a large number of trials, and furthermore may lack deeper insights into users' subconscious thoughts and behaviors during locomotion.

With this in mind, we are thinking about evaluating the user experience with VR locomotion in alternative ways, to which eye tracking can contribute. With the increasing popularity of eye-tracking studies in various research areas such as education [GBH*21], entertainment [HSHK21], daily activities [PPMW20, BGK19], and of course with the development of HMDs that allow easy acquisition of eye-tracking data via integrated eye trackers in the HMDs, eye tracking holds great potential to facilitate VR studies. Eye tracking, widely used as a reliable tool to study human behaviors in real-time (e.g., cognitive process-

ing load and visual attention behavior) [HNA*11, MJR*22], has already demonstrated its informality in classification tasks, using machine learning to predict various human- and task-related goals such as personality traits [BTK*19], learning gains [MMDR20], cognitive load [AGH*21], and task performance [KKT*22]. Although studies on the use of eye-tracking in VR scenarios are limited, there is still literature that uses eye movements to study user behavior during immersion in VR and provide insights to improve VR systems [BOB*19, GBH*21]. However, eye tracking is rarely used for evaluation purposes in VR locomotion studies. Overall, we consider that eye tracking has the potential to be a viable tool for evaluating user experience in VR locomotion research.

Therefore, in this study, we aim to propose a new approach to detecting user experience with VR locomotion. In more detail, the user experience evaluated in this study refers to the usability of the VR locomotion technique and the user's subjective perception and thoughts about the locomotion technique they are using. Specifically, we explore the possibility of classifying participants experiencing VR locomotion techniques into **L-EU** (i.e., low user experience) and **H-UE** (i.e., high user experience) using machine learning methods based on eye-tracking data alone. To this end, we designed a user study involving a navigation task with different locomotion techniques in a virtual environment for data collection. We collected participants' post-hoc user experiences of the usability of VR locomotion techniques and their feelings and enjoyment of using VR locomotion techniques as ground truth with a questionnaire. We applied a sliding window approach [HLMB18, MMDR20] to extract features from time-series data, i.e., eye-tracking data; the extracted features include pupil diameter, fixations, and saccades. We developed classification models using the random forest (RF) algorithm based on the extracted eye-tracking features. The trained RF model achieved an average accuracy of 0.71 in our binary classification task, indicating that the user experience level is predictable from eye-tracking data alone. In addition, we applied the SHapley Additive exPlanations (SHAP) approach to further investigate how eye-tracking features contribute to the model outputs. Our SHAP results show that there are underlying explainable relationships between eye-tracking features and user experience (i.e., **L-UE** and **H-UE**) and that these relationships can be further supported by the statistical results of the eye-tracking metrics.

To our knowledge, this is the first study to predict user experiences with VR locomotion using eye-tracking data, which can reveal human underlying cognitive and visual perceptual behaviors that might be difficult to capture with questionnaires. Our study provides a potential avenue for future studies to detect user experience using eye-tracking data not only in VR locomotion but also in other VR contexts where user experience needs to be evaluated to improve VR systems. Moreover, our results provide deep insights into real-time user experience prediction that may be relevant and important for interactive VR systems or intelligent user interfaces for educational and entertainment purposes. By obtaining real-time feedback from users regarding their experiences with the systems they interact with, it is possible for such systems to provide users with tailored and optimal experiences by adjusting system settings accordingly.

2. Related Work

2.1. VR Locomotion

Previous work has evaluated and compared locomotion techniques from various aspects of the user experience, such as the locomotion usability, users' subjective thoughts about the locomotion technique, preference, etc. For instance, Frommel et al. [FSW17] evaluated the impact of controller-based locomotion methods on user experience during a VR exploration task. Participants navigated a virtual zoo using four locomotion methods, namely free teleport, fix-point teleport, touchpad-based and guided automatic locomotion. User experience was measured using a questionnaire that recorded discomfort and enjoyment. Similarly, Coomer et al. [CBCWS18] compared four commonly used locomotion methods, including joystick, point-tugging, teleportation, and arm-cycling. Post-hoc questionnaires were used to determine the usability of each locomotion method and the participants' opinions about it. Compared to [FSW17], user experience was assessed in more aspects and detail. Questions were asked about the difficulty of understanding and operating the locomotion method, the feeling of being in control, fatigue, and the feeling of enjoyment. Paris et al. [PKR*19] compared two joystick-like (i.e., skiing and magic carpet) and two teleportation-like (i.e., grappling and teleportation) locomotion methods in a navigation task. Results of task performance and post-hoc questionnaires on simulator sickness, presence, and system usability were reported. However, no significant difference was found in system usability, which measures the ease of use of locomotion methods.

In addition to quantitative questionnaires, qualitative surveys were also conducted to measure user experience. Funk et al. [FMF*19] presented and evaluated three point & teleport locomotion methods that differed in the way how their teleportation trajectories were rendered in the virtual environment. Participants reported their qualitative thoughts on these locomotion methods after VR locomotion. The post-hoc survey allows researchers to directly learn how users feel about the locomotion method they used, which gives researchers further insight into how to improve locomotion methods, however, this could also be time-consuming.

In the aforementioned literature on VR locomotion research, user experience has been assessed post-hoc with either quantitative questionnaires or qualitative surveys. However, users' real-time behavioral data during locomotion, which can provide direct insights into their experiences, has rarely been considered and thus remains to be explored. As such, our study aims to propose a novel way to gain a deeper understanding of the user experience using time-series data (i.e., eye-tracking). If the effectiveness of eye tracking in predicting user experiences with VR locomotion can be demonstrated, this will provide deep insights into predicting user experience in other VR applications (e.g., VR for training and education) and thus provide clues for improving VR systems.

2.2. Eye Tracking

Eye tracking has long been used as a tool to study and improve user experience [BS14], such as smartphone app development [QZCD17], marketing [PPMW20], etc. Eye-tracking data contains rich information about how users process visual scenes

and what cognitive processing load is simultaneously triggered during information processing, and such information reveals the user's deep subconscious behavior and thoughts about the system they are interacting with. For this motivation, many researchers are considering incorporating eye-tracking technology into machine learning studies, i.e., training machine learning models based on eye-tracking features to predict various human- or task-related goals. In fact, eye tracking has already made its way into the machine learning field and has proven its effectiveness as an informative feature.

For instance, in the work of Conati et al. [CLRT20], eye movements were identified as informative features in machine learning models for predicting binary labels of different cognitive abilities during a visualization task [LCC17]. Eye-tracking features related to pupil, fixation, saccade, and area-of-interest (AOI) were calculated using various descriptive statistics. Random forest classifiers trained on eye-tracking features alone achieved accuracies over 0.63 for various prediction targets (i.e., cognitive abilities). With respect to cognitive behavior, Appel et al. [AGH*21] trained classification models to predict cognitive load in an emergency simulation game using eye-tracking features of pupils, fixations, blinks, and microsaccades. The trained models achieved accuracies of 0.63 and 0.69 across participants and tasks in the binary classification task. In addition, eye movements also proved informative in predicting personality traits [HLMB18, BTK*19] and expertise [HSHK21]. Apart from these previous works, eye-tracking data can also be used in conjunction with other task performance data or questionnaire data to improve model accuracy [LCC17, KKT*22, ZYdW21]. Kasneci et al. [KKT*22, KKA*21], for example, used eye-tracking and socio-demographic data to predict participants' performance in solving an IQ task. Gradient boosting decision trees (GBDT) models developed on eye movements alone were found to be discriminative with a ROC-AUC of 0.63 and could be improved to 0.65 with socio-demographic features. These previous works have demonstrated the feasibility and effectiveness of eye movements in revealing various underlying human behaviors, and have shown that the underlying features of eye-tracking data can be efficiently learned by machine learning models.

However, eye-tracking research in VR is still limited currently due to hardware limitations and the lack of fine-grained data analysis tools but is increasing with the advent of more and more HMDs with integrated eye trackers. Although no research has yet used eye tracking in VR locomotion for evaluation purposes, let alone to evaluate the user experience with VR locomotion, there are few studies that have used eye tracking to investigate user behavior in VR scenarios under different environmental conditions. For instance, Gao et al. [GBH*21] investigated students' cognitive and visual attention behaviors during a virtual lesson in an immersive VR classroom. Several VR environment design factors were evaluated to improve the design of the VR learning system. An approach for detecting eye movement events, i.e., fixations and saccades, suited for VR was proposed. The results showed that students' eye movements were significantly affected by the environmental factors of the VR classroom, and the underlying meaning of such effects can be interpreted to provide further guidance for improving the VR system. This study provides compelling evidence that eye movements shed light on students' perceptions of different VR environmental configurations. Although VR classroom learn-

ing is different from VR locomotion, this study reinforces our belief that eye movements provide insight into how users perceive different VR locomotion techniques and can further contribute to classification models for predicting the user experience.

3. VR Locomotion User Study

In this section, we provide an overview of our VR locomotion user study designed for data collection. Since our research goal is to assess the feasibility of classifying participants who experienced VR locomotion into **L-UE** and **H-UE** based on eye-tracking data alone, we used five different VR locomotion techniques in our study to obtain as diverse user experiences as possible. Considering that our study is an initial exploration of our research question, we used well-studied and widely used controller-based locomotion techniques as the experimental setup rather than using a niche or introducing new locomotion techniques. The locomotion techniques used in this study are arm swing, dash, grapple, joystick, and teleportation, which reportedly provide different experiences to users [CBCWS18, PKR*19, BMF18]. Our user study involved participants navigating a designed virtual environment using different locomotion techniques. They then provided feedback in a questionnaire about the usability of the locomotion techniques they were currently using, as well as their personal thoughts, such as how they felt and enjoyed the locomotion techniques. Details on the participants, materials, experimental procedure, and data acquisition are given below.

3.1. Participants

Fifteen university students (10 male, 5 female) with an average age of 24.93 ($SD = 2.84$) participated in our experiment as volunteers. All participants have normal or corrected-to-normal (with glasses) vision. They all reported their experience with video games and VR, with five reporting playing video games for more than 5 hours per week, four playing video games for 0 to 5 hours per week, and six reporting no experience with video games. In addition, one participant regularly used VR HMDs, six had some VR experience, and eight had no VR experience. All participants provided informed consent prior to data collection. Our study was approved by the institutional review board (IRB).

3.2. Materials

In order to investigate participants' experiences with different locomotion techniques without the virtual environment exerting additional effects on participants, we designed a very simple virtual environment that simultaneously meets the navigation requirement. The virtual environment consists of a green lawn area bordered by houses and trees. To encourage locomotion, we developed a search and collect task that is typically used in previous studies investigating VR locomotion [CBCWS18, CA17]. Five crystals were placed at different locations in the virtual environment and were easy to find. The overall view of the virtual environment is shown in Figure 1.

The virtual environment was created and rendered using the Unity engine (version 2020.03.23f) on a computer with a 3.5GHz

Core i7 processor and 16GB RAM. The HTC Vive Pro Eye HMD was used to display the virtual environment, which has a resolution of 1440×1600 per eye, a refresh rate of 90 Hz, and a field of view of $110°$. In addition, the HMD is seamlessly integrated with the Tobii eye tracker with a sampling rate of 120 Hz and an accuracy of $0.5° - 1.1°$. The Vive controllers served as the input device for VR locomotion. Two HTC Vive base stations were used to track a $2m \times 2m$ area for the locomotion study. All five selected locomotion techniques are controller-based, meaning participants navigate the virtual environment using Vive controllers that come with the HMD. The locomotion techniques were implemented with SteamVR based on previous work [CBCWS18, PKR*19, BMF18] and the scripts were written in C#.

3.3. Experimental Procedure

Our study used a within-subjects design with five levels of the independent variable of locomotion conditions, namely arm swing, dash, grapple, joystick, and teleportation.

After providing informed consent, all participants completed a demographic questionnaire (e.g., age, gender) and reported their previous experiences with video games and VR. The experimenter then gave instructions about the experiment. All participants took part in a total of five trials (locomotion conditions), with the order counterbalanced using a Latin square to offset order effects. Each trial consisted of three sessions, i.e., the practice session, the actual experiment (recording of eye-tracking data), and the post-hoc user experience questionnaire. In the practice session, participants were asked to practice the current locomotion technique in a clean virtual environment and no task was given. Once they became familiar with the locomotion technique, they were asked to take off the VR headset and take a break until they were ready for the experiment. The pause is compulsory to avoid the practice session exerting additional influence on the actual data collection session. In the actual experimental session, participants entered the task virtual environment after a successful 5-point eye tracker calibration routine. They were asked to search for five crystals placed in different directions of the virtual environment and collect them with controllers. After completing the task, participants were asked to take off the headset and fill out questionnaires about their locomotion experience. This marked the end of one trial. Before starting the next trial, participants were asked to take a break until they felt comfortable for the next trial.

The experiment ended after the participant completed all five trials. The entire experiment lasted approximately 50 minutes for each participant, with each participant wearing the HMD for approximately 20 minutes (including the practice session). All participants were informed before the experiment that they could terminate the experiment at any time if they felt uncomfortable.

3.4. Data Acquisition

All fifteen participants had valid and complete eye-tracking and questionnaire data, so all their data were used in our study. An average of 2.5 minutes of time-series data (i.e., eye-tracking) was collected for each trial. Thus, in total, more than three hours of eye-
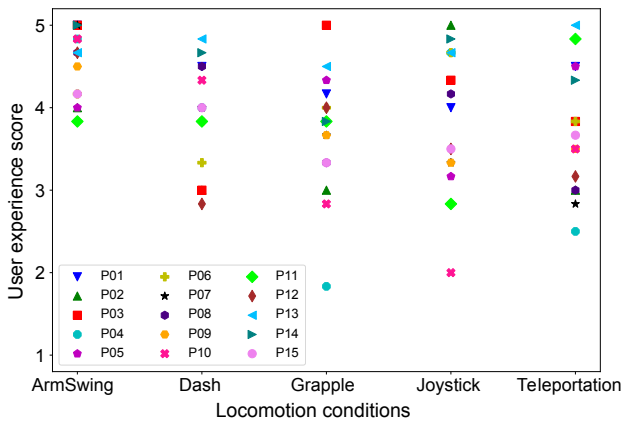
Figure 2: *Distribution of user experience scores for all participants in all locomotion conditions.*

tracking data were collected, including raw pupil and gaze vector data, which can be further used in our machine learning experiment.

In addition to eye-tracking data, participants reported their experiences after each VR locomotion trial. We used a modified version of Loewenthal's [LL01] core elements of the game experience questionnaire. In accordance with our study, six subcategories were selected and measured in our study, including difficulty in understanding and operating the locomotion technique, effort required, feeling of control over the locomotion technique, tiredness, and enjoyment. Responses were on a 5-point Likert scale, with 1 representing "not at all" and 5 representing "very much". For convenience in data analysis, we reversed the scores of some items and then calculated the mean of all six items as the final user experience score, where 1 represents the worst user experience and 5 represents the best. In this case, the best user experience means that participants report that the locomotion technique is easy to use with little effort and no fatigue and that they have a strong sense of control when using it and feel very comfortable with it.

The mean user experience score of all trials was 3.85 ($SD = 0.74$, *Median* = 4.0, ranging from 1 to 5). This indicates that participants had a good experience (above the middle level, i.e., 2.5) with different locomotion techniques. As can be seen in Figure 2, participants showed slightly larger variance in their user experience in some conditions, e.g., grapple and joystick (scores range from 2 to 5). We consider this to be a normal bias between participants, as different individuals perceive the locomotion technique differently. In addition, participants reported different user experiences across the five locomotion conditions, which is what we expected. And this supports our strategy of training our classification models in a trial-dependent manner (see Section 4.3 for details).

4. Machine Learning Method

In this section, we present our machine learning method to investigate our research question: *Is it possible to predict participants' user experience with VR locomotion by building machine learning models based solely on eye-tracking data?*

Table 1: *Thresholds for detection of eye movement events.*

Threshold	Fixation	Saccade
head_velocity	$< 12°/s$	/
gaze_velocity	$< 40°/s$	$> 80°/s$
event_duration	$100 < dur < 500 (ms)$	$30 < dur < 80 (ms)$

In the following, we first describe how we fit the time-series and questionnaire data into machine learning models. This includes the preprocessing of both types of data and the extraction of eye-tracking features. Next, we present details on building classification models, including partitioning training and test datasets in a trial-dependent manner to minimize overfitting, model building and training, and model evaluation and explanation.

4.1. Data Preprocessing

4.1.1. L-UE and H-UE

We obtained the ground truths of user experiences from the post-hoc questionnaires. To label the data samples (75 *trials*) as low level of user experience (i.e., **L-UE**) and high level of user experience (i.e, **H-UE**), we binned the user experience scores in a data-driven manner considering the median value and equal frequency. Therefore, we labeled all user experience scores below the median, i.e., < 4.0, as **L-UE** and all other scores as **H-UE**. This resulted in 33 and 42 data samples in **L-UE** and **H-UE** groups, respectively.

4.1.2. Eye Movements

In the literature, pupil size (e.g., pupil diameter) and eye movements (e.g., fixation, saccade) are commonly used to examine subjects' cognitive processing load and visual perceptual behavior [MJR*22]. We also used these measures in our study. Since only raw sensor data, including raw pupil data and gaze vectors recorded by the eye tracker and head orientations recorded by the HMD tracking system, are available, we should preprocess the data to obtain the above eye-tracking measures [TKRB12]. Pupillometric data can be extremely noisy due to blinks and noisy sensor readings. Therefore, smoothing and normalization are usually performed before calculating pupil diameter. The Savitzky-Golay filter [SG64] was applied to smooth the raw pupil diameters; the divisive baseline correction method [MFVHVdS18] with a baseline duration of ≈ 1.5 seconds was applied for normalization. Detection of eye movement events in VR remains challenging due to head movements and 3D stimulus, and there is no standard method or software to solve this problem. Fixations and saccades had to be detected manually post-experimentally. In our study, a modified velocity-threshold identification (I-VT) algorithm proposed by Gao et al. [GBH*21] that takes head movement into account was used to detect fixations, with parameters adapted to our study. Since saccades are not affected by head movements, the normal I-VT algorithm was used for saccade detection. Before detecting such eye movement events, linear interpolation was performed for the missing gaze vectors. Specifically, fixations were detected with a maximum gaze velocity threshold of $40°/s$ under the condition that head moving velocity was less than $12°/s$; saccades were detected

with a minimum gaze velocity threshold of $80°/s$. Additional duration thresholds were used for filtering. All parameters used for eye movement event detection were listed in Table 1.

4.2. Feature Extraction

Considering that each trial averaged 2.5 minutes of eye-tracking data, we applied the sliding window approach to extract eye-tracking features rather than averaging over the entire trial to avoid eliminating too much temporal information from the data. Since there is no gold standard for determining window size, our study examined time windows ranging from 5s to 30s (with a step of 5s) based on previous literature [HLMB18, MMDR20, HÖH*22]. The window size was considered as a hyperparameter during model training (see Section 4.3), and the best window size was determined based on the training results (see Section 5.1).

For each of the sliding time windows, we calculated and extracted eye-tracking metrics that have been commonly used as features in previous machine learning studies, i.e., metrics related to pupil diameter, fixations (number and duration), and saccades (number, duration, amplitude, and velocity) [CLRT20, KKT*22, KKA*21]. For some of the features, we calculated not only the mean, but also the standard deviation, the minimum, the maximum, and the sum of the values to characterize variables throughout the data [HLMB18, CCC22]. Specifically, the feature of pupil diameter was extracted since pupil diameter has been considered an indicator of cognitive processing load during various tasks in previous literature [CE14]. Fixations have been shown to reveal visual attention behavior and are also an indicator of cognitive processing load [NM20]. We calculated fixation rate and fixation duration and used them as features. In addition, pupil diameter during fixation is an indicator of cognitive processing load (pupil diameter) during visual information processing (fixation) [CBE], so we extracted it as a feature. Saccades are informative eye movements that correlate highly with visual search behavior and also with cognitive processing load [GK99, MKW*90]. Several features can be extracted from saccades, including saccade rate, saccade duration, saccade amplitude, and saccade velocity. We extracted these features and used them for model training. All extracted 33 eye-tracking features were listed in Table 2.

4.3. Model Building

In this work, we developed random forest (RF) models to classify participants into low and high user experience groups. We used the integers 0 and 1 to represent two prediction targets, i.e., class-0 for **H-UE** and class-1 for **L-UE**. For each sliding window, a 1×33 feature vector was generated. Thus, with a window size of *ws* seconds, there are approximate $N = 75 \times 150/ws$ samples for machine learning training, where 75 is the number of trials and 150 seconds (i.e., 2.5 minutes) is the averaged duration of each trial. Min-max normalization was performed for all feature variables. Then, data samples were randomly split into the training set (80%, about $N \times 0.8$ samples) and the test set (20%, about $N \times 0.2$ samples). For example, with a window size of 10s, we have about 1100 samples, of which about 900 samples are used as the training set and about 200 samples are used as the test set. The classification models were trained on the training set and tested on the test set. For

Table 2: *Extracted eye-tracking features for machine learning models.*

Features	Descriptive statistics
Pupil diameter	Mean, Std.dev, Min, Max of the normalized pupil diameter
Pupil diameter (fixation)	Mean, Std.dev, Min, Max of the normalized pupil diameter during fixation
Fixation rate	Number of fixations per minute
Fixation duration	Mean, Std.dev, Min, Max, Sum of the fixation durations
Saccade rate	Number of saccades per minute
Saccade duration	Mean, Std.dev, Min, Max, Sum of the saccade durations
Saccade amplitude	Mean, Std.dev, Min, Max, Sum of the saccade amplitudes
Saccade velocity	Mean, Std.dev, Min, Max of the saccade velocities
Saccade peak velocity	Mean, Std.dev, Min, Max of the saccade peak velocities

note: Std.dev, Min, and Max denote the standard deviation, minimum, and maximum, respectively.

model training, we performed 5-fold cross-validation to tune the hyperparameters of the random forest models on the training set, which means we further split the training set into the sub-training set and the validation set.

To avoid overfitting and to generalize our models to unseen data, all data splits in the training and testing processes were trial-dependent, that is, all feature vectors from the same trial were to remain in the same data sets (i.e., either sub-training, validation, or test set). Here, we split the data in a trial-dependent manner rather than a participant-dependent manner, which may risk overfitting the model because the models might be tested on the seen data (if data samples from a participant exist in both the training and test sets). Actually, such concern can be eliminated as participants had different user experiences across different conditions (See Figure 2) and these differences are reflected in the eye-tracking behavior as well (eye-tracking feature), which means that we can consider each participant's trial as an independent data sample. Thus, we assumed that the models were still tested on the unseen data. Moreover, we performed stratified data split, that is, data samples were split into different sets while maintaining the percentage of samples for each class. These data split policies were performed manually. Notice that participants were randomly assigned during all data splits without regard to identity. Furthermore, to reduce the bias caused by data splitting, we split the data 50 times with different random seeds so that all trials had the opportunity to be used as training and test sets. Thus, in a 5-fold cross-validation, our model was trained and validated on 50×5 different sub-training and validation sets. The best model identified through the training process was then tested 50 times on unseen data, i.e., on the 50 test sets. The final model test results reported in our study are the average of the 50 runs.

4.4. Model Evaluation and Explainability

In this work, the random forest models solved a binary classification problem. Considering that our dataset is not fully balanced (33 **L-UE**: 42 **H-UE**), metrics including accuracy, precision, recall, and F1-score were used to evaluate model performance. These four metrics were calculated based on True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). We report only the model test results using these metrics.

Moreover, post-hoc model explainability becomes significant in the field of machine learning as it provides insights into how a model can be improved. As in our study, a large set of eye-tracking features indicative of different human behaviors were extracted and used. In addition to training the models to achieve the best performance in predicting user experiences, we also emphasize explaining the model at the feature level, i.e., how individual features contribute to the model outputs. To this end, we applied the state-of-the-art method for model explainability, namely SHapley Additive exPlanations (SHAP) [LL17]. First, SHAP can help gain insight into the feature importance and identify the most informative eye-tracking features for the models. Second, SHAP can reveal the underlying relationships between eye-tracking features and model outputs (i.e, **L-UE**, **H-UE**) by demonstrating what impact an individual feature has on model outputs.

5. Results

In this section, we report our results in three parts. First, we report the performance of our RF models on the binary classification task of predicting user experience level. Second, we report the results of SHAP explainability. We identified the most informative eye-tracking features and explained the classification model at the feature level to reveal how these informative eye-tracking features contribute to the model outputs. Third, to further support the SHAP results, we also report the statistical results of the main eye-tracking metrics.

5.1. Model Performance

As can be seen in Table 3, the RF models trained only on eye-tracking features extracted with different window sizes were able to classify participants into **L-UE** and **H-UE**, with average accuracies above 0.62. Of all eight window sizes examined, the RF model

Table 3: *The test performance of RF models trained on features extracted with different window sizes. Best window size and model performance are in bold.*

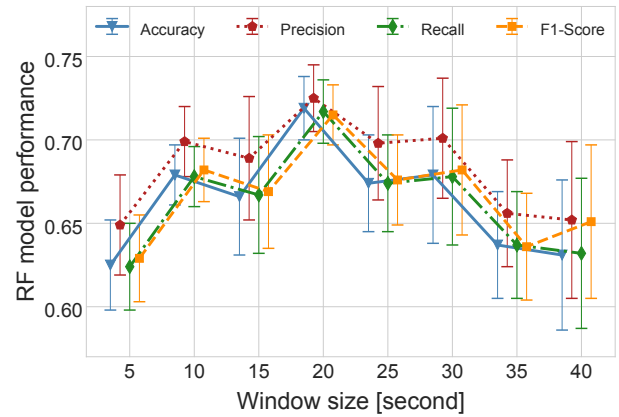| WindowSize[s]	Accuracy	Precision	Recall	F1-score
5	0.62	0.64	0.62	0.63
10	0.68	0.70	0.68	0.69
15	0.67	0.69	0.67	0.67
20	**0.71**	**0.72**	**0.71**	**0.71**
25	0.67	0.69	0.67	0.67
30	0.68	0.70	0.68	0.68
35	0.63	0.65	0.63	0.63
40	0.64	0.65	0.64	0.64

Figure 3: *The test performance of RF models trained on features extracted with different time windows. The best performance was obtained with a 20-second time window.*

trained on features extracted with a 20-second time window performed best, with an average accuracy of 0.71 ($SD = 0.019$), precision of 0.72 ($SD = 0.020$), recall of 0.71 ($SD = 0.019$), and F1-score of 0.71 ($SD = 0.018$), as shown in Figure 3. The best RF model was created with 300 decision trees. The maximum depth of the tree used was 10, the minimum number of samples for splitting an internal node was 4, and the minimum number of samples at a leaf node was 1.

5.2. SHAP Explanation

We applied the SHAP TreeExplainer [LEC*20] for tree-based classifiers, i.e., random forest. The local explanations of the best RF model are shown in a beeswarm-style summary plot of SHAP val-
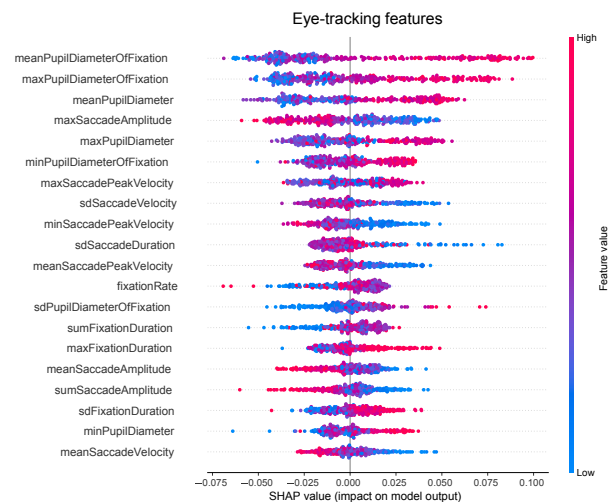

Figure 4: *The top 20 eye-tracking features in the best RF model, ranked by feature importance from top to bottom.*

Table 4: *Statistical comparison of eye-tracking metrics between **L-UE** and **H-UE** groups.*

Eye-tracking metrics	L-UE group Mean (SD)	H-UE group Mean (SD)	$p-$value	Significance
Mean pupil diameter	1.09 (0.11)	1.05 (0.12)	1.79×10^{-14}	****
Mean pupil diameter (fixation)	1.08 (0.16)	1.01 (0.19)	2.03×10^{-17}	****
Fixation rate [$num\_fixation/min$]	55.01 (27.83)	52.91 (29.57)	.096	n.s.
Mean of fixation duration [ms]	216.12 (53.96)	210.53 (62.11)	.079	n.s.
Saccade rate [$num\_saccade/min$]	46.11 (25.14)	49.04 (23.49)	.014	*
Mean of saccade duration [ms]	47.48 (8.82)	48.93 (6.41)	.008	**
Mean of saccade amplitude [$°$]	8.97 (3.05)	10.13 (2.85)	2.95×10^{-10}	****
Mean of saccade velocity [$°/s$]	187.21 (63.59)	206.68 (53.62)	9.98×10^{-9}	****
Mean of saccade peak velocity [$°/s$]	306.49 (96.26)	327.77 (79.30)	4.65×10^{-6}	****

note: *, **, ****, and n.s. represent $p < .05$, $p < .01$, $p < .0001$, and no statistical significance, respectively.

ues, as shown in Figure 4. The 20 most informative features are displayed and sorted by feature importance from top to bottom along the *y*-axis. SHAP values are located along the *x*-axis. Each dot in the summary plot represents the SHAP value of one feature observation. The more spread out the dots of a feature are on the *x*-axis, the greater the impact of the feature on the model. The color of the dots in the summary plot indicates the value of the feature, with blue color indicating a low feature value and red color indicating a high feature value. A color change from blue to red along the *x*-axis from left to right indicates that a feature has a positive (negative) impact on the prediction of class-1 (class 0), i.e., **L-UE** (**H-UE**), in contrast, a color change from red to blue along the *x*-axis from left to right indicates that a feature has a negative (positive) impact on the prediction of class-1 (class-0), i.e., **L-UE** (**H-UE**).

As shown in Figure 4, the feature *meanPupilDiameterOfFixation* followed by the features *maxPupilDiameterOfFixation* and *meanPupilDiameter* are the three most informative features for the RF model in classifying the user experience into low and high level. Moreover, it is worth noting that these three most important features have a positive impact on the prediction output into class-1 (i.e., **L-UE**), which means that the feature value higher than the feature average drives the classification into the prediction output of **L-UE**. The feature *maxSaccadeAmplitude* also contributes greatly to the classification model. Unlike the above three most informative features, *maxSaccadeAmplitude* was observed to have a positive impact on the prediction output into class-0 (i.e., **H-UE**), which means that the feature value higher than the feature average drives the classification into prediction output of **H-UE**. Notably, four of the five most informative features are pupil-related. However, compared to pupil- and saccade-related features, fixation-related features tend to be less informative in classifying user experience levels.

5.3. Statistical Test

To further validate the SHAP explanation results, we applied statistical tests to the eye-tracking metrics typically analyzed in eye-tracking studies, i.e., these features can be interpreted as associated with various human behaviors (e.g., cognitive load, visual attention, and visual search). For this reason, we did not apply statistical tests to these eye-tracking metrics calculated as minimum, maximum,

standard deviation, and sum values as the features of such metrics can be easily learned by machine learning models, but their statistical differences are hard to interpret. The Shapiro-Wilk test was used to test for normality and the Mann-Whitney U test was used as a non-parametric test. The significance level was set at $\alpha = 0.05$ for all the tests. The statistical results for the comparison between **L-UE** and **H-UE** groups are shown in Table 4.

As can be seen, there was no significant difference between the two groups for fixation-related features that were found to be of low informativeness in the classification model. Conversely, significant differences were found between the two groups for pupil diameters, which are highly informative for the classification model. In particular, the mean pupil diameters in the **L-UE** group are significantly larger than the values in the **H-UE** group, with $p < .0001$. Similar to the mean pupil diameter, a significant difference was also found between the groups for the most informative pupil feature, the mean pupil diameter during fixation, with the mean pupil diameters during fixation in the **L-UE** group being significantly larger than in the **H-UE** group, with $p < .0001$. In addition to pupil diameter and fixations, significant differences were also found in saccade metrics. It was found that the saccade rates in the **H-UE** group are higher than in the **L-UE** group, with $p = .014$. Also, the mean saccade duration, amplitude, velocity, and peak velocity were found to be higher in the **H-UE** group than in the **L-UE** group, with $p < .001$, $p < .0001$, $p < .0001$, and $p < .0001$, respectively.

6. Discussion

In this section, the results are discussed according to the structure of the results section.

First, the sliding window approach used for feature extraction proved successful, as our results show that the user experience with VR locomotion is predictable using machine learning models trained solely on eye-tracking features. In particular, the RF model trained on features extracted with a sliding window of 20 seconds performed best with an average accuracy of 0.71 (see Table 3). On the other hand, the models performed worse when the window sizes were too short or too long. This could be because when the sliding window is too short, the features behind the eye movements that could be indicative of the users' perception of VR

locomotion are truncated, whereas when the sliding window is too long, these eye movement features can be easily smoothed out by the averaging effects. Previous literature on the use of the sliding window also shows that the optimal window size is dependent on the task [HLMB18, MMDR20]. Second, although no previous research has predicted user experiences with VR locomotion using eye tracking, the best performance we obtained is comparable to eye-tracking-based classification results for other behavioral targets in other scenarios (see Section 2.2). Our results demonstrate the robustness of our approach to predicting user experiences with VR locomotion based solely on eye-tracking data. This fills a research gap in the field of VR locomotion.

Furthermore, we applied the SHAP approach to explore how eye-tracking features contribute to the classification models and to further uncover the underlying relationships between features and model outputs (see Figure 4). The results show that pupil- and saccade-related features are highly informative for the RF model in predicting user experiences. Fixation-related features, on the other hand, are less informative compared to the other two types of features. In addition, we found that most pupil-related features, such as mean, maximum, and minimum pupil diameter during fixations, as well as mean and maximum pupil diameter, have a positive influence on the classification model into the prediction output of **L-UE** (class-1), which means that a higher value of these features than their respective average value drives the classification into the prediction output of **L-UE**. This may suggest that a higher value of these pupil-related features correlates with a lower user experience, as evidenced also by the statistical test results of pupil diameter (see Table 4). Specifically, mean pupil diameter and mean pupil diameter during fixation were found to be significantly larger in the **L-UE** group than in the **H-UE** group, supporting the SHAP results showing that pupil-related features correlate negatively with user experience level. We found similar consistent results for another type of informative feature, namely saccade-related features. The SHAP results show that features such as maximum saccade amplitude, minimum and mean saccade peak velocity have a negative influence on the classification model into the prediction output of **L-UE**, which means, that a lower value of these features than their respective average value drives the classification into the prediction output of **L-UE**, in contrast, a higher value of these features than their respective average value drives the classification into the prediction output of **H-UE** (class-0). This may suggest that a higher (lower) value of these saccade features correlates with a higher (lower) user experience. It is worth noting that these SHAP results for saccade features are also consistent with the statistical results where saccade-related metrics have a higher value in the **H-UE** group than in the **L-UE** group, further suggesting that these saccade-related features correlate positively with user experience level. However, no significant difference was found in fixation rate and fixation duration in the statistical tests, but we found in the SHAP results that these fixation features affected the classification models differently. This suggests that the model can learn the deeper characteristics of eye movements that are difficult to detect with statistical tests.

Although there is no research investigating correlations between eye movements and user experiences with VR locomotion, our findings from the SHAP approach and statistical tests can still be interpreted based on previous literature. As in our study, participants' user experience was evaluated as to their feelings about using the locomotion methods, i.e., whether the locomotion method was easy to use, whether they felt tired, whether it was fun, that is, a higher user experience indicates that the participant can navigate the virtual environment easily and with little fatigue using the locomotion methods. Thus, in our case, we consider that a higher user experience implies a lower cognitive load, while a lower user experience implies a higher cognitive load. Our results show that there might be a negative relationship between pupil diameter and user experience, which means that participants with high user experience have small pupil diameter. This can be supported by previous literature, stating that pupil diameter correlates positively with cognitive load [HP64]. That is, participants who have a high user experience (i.e., little fatigue, high enjoyment) have low cognitive load as indicated by pupil diameter. Similarly, our results suggest that saccade amplitude and saccade peak velocity are positively correlated with user experience, implying that participants with high user experience have large saccade amplitude and saccade peak velocity. This can be supported by previous literature as well, stating that saccade amplitude and saccade peak velocity are negatively correlated with cognitive load [DSRS*10, MKW*90, KODDM20]. Thus, our model explainability results, revealing the underlying relationships between eye-tracking features and user experience, can be supported by previous literature.

7. Conclusion and Implication

In this work, we investigated the feasibility of predicting user experiences with VR locomotion (i.e., the usability of locomotion methods and users' feelings about them during navigation) based on eye-tracking data alone. We conducted a user study in which participants performed a navigation task in a virtual environment with five different locomotion methods. We collected participants' experience data using a standard questionnaire and binned them into low and high user experience groups as the ground truth. We extracted a variety of eye-tracking features from time-series data using a sliding window approach. We built classification models using the random forest algorithm based solely on the extracted eye-tracking features. Our best model achieved an average accuracy of over 0.7 in 50 runs, demonstrating the feasibility of predicting user experiences with VR locomotion based on eye-tracking data alone and the robustness of our research. By applying the SHAP approach, we identified the most contributed features of the classification model. In addition, the SHAP explainability results revealed some relationships between eye movements and user experience, i.e., **L-UE** and **H-UE**, which can be further supported by the statistical results.

To the best of our knowledge, our study is the first to use eye tracking as a tool to investigate user experience in the research field of VR locomotion. User experience is one of the most important aspects in evaluating or comparing locomotion methods, as it is closely related to the improvement of locomotion methods. As such, our research provides a viable user experience assessment tool for future studies, especially when new locomotion techniques are proposed, and can be extended to other VR research that aims to provide a good experience to users or requires system assessment

and improvement. The ultimate goal would be a system that can detect the user experience level in real-time to offer users a tailored and optimal experience in real-world applications.

References

[AGH*21] APPEL T., GERJETS P., HOFFMAN S., MOELLER K., NIN-AUS M., SCHARINGER C., SEVCENKO N., WORTHA F., KASNECI E.: Cross-task and cross-participant classification of cognitive load in an emergency simulation game. *IEEE Transactions on Affective Computing* (2021), 1–1. doi:10.1109/TAFFC.2021.3098237. 2, 3

[BGK19] BOZKIR E., GEISLER D., KASNECI E.: Assessment of driver attention during a safety critical situation in VR to generate VR-based training. In *ACM Symposium on Applied Perception 2019* (New York, NY, USA, 2019), ACM. doi:10.1145/3343036.3343138. 2

[BMF18] BHANDARI J., MACNEILAGE P., FOLMER E.: Teleportation without spatial disorientation using optical flow cues. In *Proceedings of the 44th Graphics Interface Conference* (Waterloo, CAN, 2018), GI '18, Canadian Human-Computer Communications Society, p. 162–167. doi:10.20380/GI2018.22. 2, 4

[BOB*19] BERTON F., OLIVIER A.-H., BRUNEAU J., HOYET L., PETTRE J.: Studying gaze behaviour during collision avoidance with a virtual walker: Influence of the virtual reality setup. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (2019), pp. 717–725. doi:10.1109/VR.2019.8798204. 2

[BS14] BERGSTROM J. R., SCHALL A.: *Eye tracking in user experience design*. Elsevier, 2014. 3

[BSB11] BOLTE B., STEINICKE F., BRUDER G.: The jumper metaphor: an effective navigation technique for immersive display setups. In *Proceedings of Virtual Reality International Conference* (2011), vol. 1. 2

[BTK*19] BERKOVSKY S., TAIB R., KOPRINSKA I., WANG E., ZENG Y., LI J., KLEITMAN S.: Detecting personality traits using eye-tracking data. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2019), CHI '19, ACM, p. 1–12. doi:10.1145/3290605.3300451. 2, 3

[CA17] CHRISTOU C. G., ARISTIDOU P.: Steering versus teleport locomotion for head mounted displays. In *Augmented Reality, Virtual Reality, and Computer Graphics* (Cham, 2017), Springer International Publishing, pp. 431–446. doi:10.1007/978-3-319-60928-7_37. 4

[CBCWS18] COOMER N., BULLARD S., CLINTON W., WILLIAMS-SANDERS B.: Evaluating the effects of four VR locomotion methods: Joystick, Arm-Cycling, Point-Tugging, and Teleporting. In *Proceedings of the 15th ACM Symposium on Applied Perception* (New York, NY, USA, 2018), SAP '18, ACM. doi:10.1145/3225153.3225175. 2, 3, 4

[CBE] COORS A., BRETELER M. M. B., ETTINGER U.: Processing speed, but not working memory or global cognition, is associated with pupil diameter during fixation. *Psychophysiology*, e14089. doi:https://doi.org/10.1111/psyp.14089. 6

[CCC22] CHEN K.-W., CHANG Y.-J., CHAN L.: Predicting opportune moments to deliver notifications in virtual reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2022), CHI '22, ACM. doi:10.1145/3491102.3517529. 6

[CE14] CHEN S., EPPS J.: Using task-induced pupil diameter and blink rate to infer cognitive load. *Human–Computer Interaction 29*, 4 (2014), 390–413. doi:10.1080/07370024.2014.892428. 6

[CLRT20] CONATI C., LALLÉ S., RAHMAN M. A., TOKER D.: Comparing and combining interaction data and eye-tracking data for the real-time prediction of user cognitive abilities in visualization tasks. *ACM Trans. Interact. Intell. Syst. 10*, 2 (may 2020). doi:10.1145/3301400. 3, 6

[CMS20] CHERNI H., MÉTAYER N., SOULIMAN N.: Literature review of locomotion techniques in virtual reality. *Int. J. Virtual Real. 20* (2020), 1–20. 2

[DSRS*10] DI STASI L. L., RENNER R., STAEHR P., HELMERT J. R., VELICHKOVSKY B. M., CAÑAS J. J., CATENA A., PANNASCH S.: Saccadic peak velocity sensitivity to variations in mental workload. *Aviation, space, and environmental medicine 81*, 4 (2010), 413–417. doi:https://doi.org/10.3357/ASEM.2579.2010. 9

[FMF*19] FUNK M., MÜLLER F., FENDRICH M., SHENE M., KOLVENBACH M., DOBBERTIN N., GÜNTHER S., MÜHLHÄUSER M.: Assessing the accuracy of Point & Teleport locomotion with orientation indication for virtual reality using curved trajectories. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2019), CHI '19, ACM, p. 1–12. doi:10.1145/3290605.3300377. 2, 3

[FSW17] FROMMEL J., SONNTAG S., WEBER M.: Effects of controller-based locomotion on player experience in a virtual reality exploration game. In *Proceedings of the 12th International Conference on the Foundations of Digital Games* (New York, NY, USA, 2017), FDG '17, ACM. doi:10.1145/3102071.3102082. 2, 3

[GBH*21] GAO H., BOZKIR E., HASENBEIN L., HAHN J.-U., GÖLLNER R., KASNECI E.: Digital transformations of classrooms in virtual reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2021), CHI '21, ACM. doi:10.1145/3411764.3445596. 2, 3, 5

[GK99] GOLDBERG J. H., KOTVAL X. P.: Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics 24*, 6 (1999), 631–645. doi:https://doi.org/10.1016/S0169-8141(98)00068-7. 6

[HLMB18] HOPPE S., LOETSCHER T., MOREY S. A., BULLING A.: Eye movements during everyday behavior predict personality traits. *Frontiers in Human Neuroscience 12* (2018), 105. doi:10.3389/fnhum.2018.00105. 2, 3, 6, 9

[HNA*11] HOLMQVIST K., NYSTRÖM M., ANDERSSON R., DEWHURST R., JARODZKA H., VAN DE WEIJER J.: *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford, 2011. 2

[HÖH*22] HOLMQVIST K., ÖRBOM S. L., HOOGE I. T., NIEHORSTER D. C., ALEXANDER R. G., ANDERSSON R., BENJAMINS J. S., BLIGNAUT P., BROUWER A.-M., CHUANG L. L., ET AL.: Eye tracking: empirical foundations for a minimal reporting guideline. *Behavior research methods* (2022), 1–53. doi:https://doi.org/10.3758/s13428-021-01762-8. 6

[HP64] HESS E. H., POLT J. M.: Pupil size in relation to mental activity during simple problem-solving. *Science 143*, 3611 (1964), 1190–1192. doi:10.1126/science.143.3611.1190. 9

[HSHK21] HOSP B. W., SCHULTZ F., HÖNER O., KASNECI E.: Soccer goalkeeper expertise identification based on eye movements. *PLOS ONE 16*, 5 (05 2021), 1–22. doi:10.1371/journal.pone.0251070. 2, 3

[KKA*21] KASNECI E., KASNECI G., APPEL T., HAUG J., WORTHA F., TIBUS M., TRAUTWEIN U., GERJETS P.: Tüeyeq, a rich iq test performance data set with eye movement, educational and socio-demographic information. *Scientific Data 8*, 1 (2021), 1–14. doi:https://doi.org/10.1038/s41597-021-00938-3. 3, 6

[KKT*22] KASNECI E., KASNECI G., TRAUTWEIN U., APPEL T., TIBUS M., JAEGGI S. M., GERJETS P.: Do your eye movements reveal your performance on an iq test? A study linking eye movements and socio-demographic information to fluid intelligence. *PLOS ONE 17*, 3 (03 2022), 1–23. doi:10.1371/journal.pone.0264316. 2, 3, 6

[KODDM20] KESKIN M., OOMS K., DOGRU A. O., DE MAEYER P.: Exploring the cognitive load of expert and novice map users using EEG and eye tracking. *ISPRS International Journal of Geo-Information 9*, 7 (2020). doi:10.3390/ijgi9070429. 9

[LCC17] LALLÉ S., CONATI C., CARENINI G.: Impact of individual differences on user experience with a real-world visualization interface for public engagement. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization* (New York, NY, USA,

2017), UMAP '17, ACM, p. 369–370. doi:10.1145/3079628.3079634. 3

[LEC*20] LUNDBERG S. M., ERION G., CHEN H., DEGRAVE A., PRUTKIN J. M., NAIR B., KATZ R., HIMMELFARB J., BANSAL N., LEE S.-I.: From local explanations to global understanding with explainable AI for trees. *Nature machine intelligence 2*, 1 (2020), 56–67. doi:https://doi.org/10.1038/s42256-019-0138-9. 7

[LL01] LOEWENTHAL K., LEWIS C. A.: *An introduction to psychological tests and scales*. Psychology press, 2001. doi:https://doi.org/10.4324/9781315782980. 5

[LL17] LUNDBERG S. M., LEE S.-I.: A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Red Hook, NY, USA, 2017), NIPS'17, Curran Associates Inc., p. 4768–4777. doi:10.5555/3295222.3295230. 7

[LLS18] LANGBEHN E., LUBOS P., STEINICKE F.: Evaluation of locomotion techniques for room-scale VR: Joystick, Teleportation, and Redirected Walking. In *Proceedings of the Virtual Reality International Conference - Laval Virtual* (New York, NY, USA, 2018), VRIC '18, ACM. doi:10.1145/3234253.3234291. 2

[MFVHVdS18] MATHÔT S., FABIUS J., VAN HEUSDEN E., VAN DER STIGCHEL S.: Safe and sensible preprocessing and baseline correction of pupil-size data. *Behavior Research Methods 50*, 1 (2018), 94–106. doi:10.3758/s13428-017-1007-2. 5

[MJR*22] MAHANAMA B., JAYAWARDANA Y., RENGARAJAN S., JAYAWARDENA G., CHUKOSKIE L., SNIDER J., JAYARATHNA S.: Eye movement and pupil measures: A review. *Frontiers in Computer Science 3* (2022). doi:10.3389/fcomp.2021.733531. 2, 5

[MKW*90] MAY J. G., KENNEDY R. S., WILLIAMS M. C., DUNLAP W. P., BRANNAN J. R.: Eye movement indices of mental workload. *Acta Psychologica 75*, 1 (1990), 75–89. doi:https://doi.org/10.1016/0001-6918(90)90067-P. 6, 9

[MMDR20] MOORE A. G., MCMAHAN R. P., DONG H., RUOZZI N.: Extracting velocity-based user-tracking features to predict learning gains in a virtual reality training application. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (2020), pp. 694–703. doi:10.1109/ISMAR50242.2020.00099. 2, 6, 9

[NM20] NEGI S., MITRA R.: Fixation duration and the learning process: an eye tracking study with subtitled videos. *Journal of Eye Movement Research 13*, 6 (2020). 6

[PKR*19] PARIS R., KLAG J., RAJAN P., BUCK L., MCNAMARA T. P., BODENHEIMER B.: How video game locomotion methods affect navigation in virtual environments. In *ACM Symposium on Applied Perception 2019* (New York, NY, USA, 2019), SAP '19, ACM. doi:10.1145/3343036.3343131. 2, 3, 4

[PPMW20] 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 31*, 3 (2020), 675–691. doi:10.1287/isre.2019.0907. 2, 3

[QZCD17] QU Q.-X., ZHANG L., CHAO W.-Y., DUFFY V.: User experience design based on eye-tracking technology: A case study on smartphone apps. In *Advances in Applied Digital Human Modeling and Simulation* (Cham, 2017), Duffy V. G., (Ed.), Springer International Publishing, pp. 303–315. doi:10.1007/978-3-319-41627-4_27. 3

[RMFW20] RADIANTI J., MAJCHRZAK T. A., FROMM J., WOHLGENANNT I.: A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education 147* (2020), 103778. doi:https://doi.org/10.1016/j.compedu.2019.103778. 2

[SG64] SAVITZKY A., GOLAY M. J. E.: Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry 36* (1964), 1627–1639. doi:10.1021/ac60214a047. 5

[TFY*22] TAN C. T., FOO L. C., YEO A., LEE J. S. A., WAN E., KOK X.-F. K., RAJENDRAN M.: Understanding user experiences across VR Walking-in-Place locomotion methods. In *CHI* (New York, NY, USA, 2022), CHI '22, ACM. doi:10.1145/3491102.3501975. 2

[TGT*21] TAO G., GARRETT B., TAVERNER T., CORDINGLEY E., SUN C.: Immersive virtual reality health games: A narrative review of game design. *Journal of Neuro Engineering and Rehabilitation 18*, 1 (2021), 1–21. doi:https://doi.org/10.1186/s12984-020-00801-3. 2

[TKRB12] TAFAJ E., KASNECI G., ROSENSTIEL W., BOGDAN M.: Bayesian online clustering of eye movement data. In *Proceedings of the symposium on eye tracking research and applications* (New York, NY, USA, 2012), ETRA '12, ACM, pp. 285–288. doi:10.1145/2168556.2168617. 5

[vWSM*20] VON WILLICH J., SCHMITZ M., MÜLLER F., SCHMITT D., MÜHLHÄUSER M.: Podoportation: Foot-based locomotion in virtual reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2020), CHI '20, ACM, p. 1–14. doi:10.1145/3313831.3376626. 2

[WKMW16] WILSON P. T., KALESCKY W., MACLAUGHLIN A., WILLIAMS B.: VR locomotion: Walking & Walking in place & Arm Swinging. In *Proceedings of the 15th ACM SIGGRAPH Conference on Virtual-Reality Continuum and Its Applications in Industry - Volume 1* (New York, NY, USA, 2016), VRCAI '16, ACM, p. 243–249. doi:10.1145/3013971.3014010. 2

[ZYdW21] ZHOU F., YANG X. J., DE WINTER J. C. F.: Using eye-tracking data to predict situation awareness in real time during takeover transitions in conditionally automated driving. *IEEE Transactions on Intelligent Transportation Systems* (2021), 1–12. doi:10.1109/TITS.2021.3069776. 3