# Development of classifiers to determine factors associated with older adult's cognitive functions and game user experience in VR using head kinematics

Faraz Ali [1], John Munoz [2], Jennifer Boger [2], Laura Middleton [2], Aysha Basharat [2], Samira Mehrabi [2], Michael Barnett-Cowan [2], and Shi Cao [2]

[1]University of Waterloo
[2]Affiliation not available

October 30, 2023

**Abstract**

Virtual reality (VR) is increasingly being used to promote exercise among older adults. The data captured through VR may be useful indicator of the game user's experience as well as providing insight into functional ability of older adults. This paper presents classifiers to predict game user experience variables using VR data from community-dwelling older adults. Head-kinematic data of the VR headset was collected from 13 participants over a six-week period with three 20-minutes exergame sessions per week (e.g., 360 minutes per participant). Cognitive function was assessed using the Montreal Cognitive Assessment (MoCa) and multisensory response-time (RT). Game user experience was captured through perceived-levels of cybersickness, enjoyment, and exertion after each session. Data was used as references for discrete binary and ternary classification patterns. Combinations of kinematic features were used to train different classifiers: K-nearest-neighbors (KNN), linear discriminant analysis (LDA), support vector machines (SVM), and decision trees. Maximum classification accuracy of 70% was found for MoCa, 68% for perceived exertion, 60% for cybersickness, 59% for multisensory RT, and 53% for perceived enjoyment. Results suggest unobtrusive recording of head kinematics from VR headsets combined with machine learning classifiers could be used to predict cognition, exertion, and game user experience among older adults.

# Development of classifiers to determine factors associated with older adult's cognitive functions and game user experience in VR using head kinematics

John E Muñoz*, Faraz Ali*, Aysha Basharat†, Samira Mehrabi†, Michael Barnett-Cowan†, Shi Cao*, Laura E Middleton†‡ and Jennifer Boger*‡
*Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, CA
†Department of Kinesiology, University of Waterloo, Waterloo, ON, CA
‡Research Institute for Aging, Waterloo, ON, CA

***Abstract*—Virtual reality (VR) is increasingly being used to promote exercise among older adults. The data captured through VR may be useful indicator of the game user's experience as well as providing insight into functional ability of older adults. This paper presents classifiers to predict game user experience variables using VR data from community-dwelling older adults. Head-kinematic data of the VR headset was collected from 13 participants over a six-week period with three 20-minutes exergame sessions per week (e.g., 360 minutes per participant). Cognitive function was assessed using the Montreal Cognitive Assessment (MoCa) and multisensory response-time (RT). Game user experience was captured through perceived-levels of cybersickness, enjoyment, and exertion after each session. Data was used as references for discrete binary and ternary classification patterns. Combinations of kinematic features were used to train different classifiers: K-nearest-neighbors (KNN), linear discriminant analysis (LDA), support vector machines (SVM), and decision trees. Maximum classification accuracy of 70% was found for MoCa, 68% for perceived exertion, 60% for cybersickness, 59% for multisensory RT, and 53% for perceived enjoyment. Results suggest unobtrusive recording of head kinematics from VR headsets combined with machine learning classifiers could be used to predict cognition, exertion, and game user experience among older adults.**

***Index Terms*—Virtual reality, older adults, machine learning, head kinematics, game user experience.**

## I. Introduction

RESEARCH in virtual reality (VR) is increasingly being conducted in a wide variety of situations to promote healthcare and wellbeing [1]. Leveraging the data generated within virtual environments to investigate user experience is one of the most exciting and trending topics in VR user research. The combination of data science methodologies and human-computer interaction paradigms in VR promises to create sophisticated, digitally connected systems in healthcare [2]. In particular, input modalities have the potential of capturing multivariate information from the players when interacting with VR games that can then be used to understand the people playing them.

Conventional immersive VR uses head-mounted-displays (HMDs) to provide a sense of physical presence and interactivity that allows for high levels of engagement and a more natural interaction. Interactivity is achieved by using movement sensors in both the HMD and the controllers of the VR systems [3]. The captured data represents kinematic information that describes features such as displacement and orientation of specific body segments or joints. In the case of commercially available VR systems, this kinematic information is from the head and hands. The trajectories, velocities and accelerations recorded during the interaction within virtual environments contain rich information about player's behaviors and reactions to the virtual experience [4]. To date, kinematic information is used to gain insights about a player's quality of movement, especially when involving applications in rehabilitation and therapy [5].

Kinematic information could be used to predict other important variables that affect the VR experience such as cybersickness (e.g., visually induced motion sickness [6]), perceived enjoyment, and physical exhaustion. Many of these variables can be grouped in an umbrella term called "game user experience", which is a term used to refer to the psychological, behavioral and thinking process of players when interacting with games [7]. This field uses cognitive science principles and the scientific method in the game design and game development processes to evaluate the game experience of players and improve upon findings [7]. The creation of VR systems capable of leveraging computational models for predicting the likelihood of occurrence of game user experience factors could improve the way VR content is created as well as researchers and developers' ability to systematically analyse their content. For instance designers could use the models to better tune or modify the VR content based on the predicted game user experience.

In this paper we used data collected from a group of older adults using a custom-built VR exergame to explore the use of head kinematic data and ML classification models to determine the likelihood having a good or a bad user experience as well as to classify relevant cognitive aspects in aging research such as cognitive impairment or response time.

## II. Related Work

Several studies have explored ML applications to explore human factors using physiological responses collected during VR experiences. Cybersickness is one of the most widely explored factors, which reflects its propensity to shape the user experience [8]. Garcia-Agundez et al. developed ML classifiers for binary (two classes) and ternary (three classes) classifications of cybersickness in VR environments [9]. In a controlled study, they collected electrocardiogram (ECG), electrooculogram (EOG), and kinematic data from the headset from 66 participants playing a VR game in an aviation task. Scores collected from each participant's Simulation Sickness Questionnaire after playing the VR game were used as ground-truth data. Using ML classification algorithms such as decision trees (DTs), support vector machines (SVMs), and K-nearest neighbours (KNNs) as classifiers, the authors explored classification of cybersickness using physiological data. For binary classification, they achieved highest classification accuracy of 82% with the Gaussian SVM classifier and the lowest with 58% for ternary classification using a KNN classifier. Similar results were achieved in a different study using electroencephalography (EEG) data and deep learning models [10].

Diaz-Romero et al. recruited 21 healthy participants who had to first view images associated with standardized emotional outcome scores before playing a Whack-a-Mole VR game while collecting EEG, EOG, accelerometer, and gyroscope data during the sessions as well as participants' self-reported arousal and valence states [11]. After filtering the raw data, the researchers explored combinations of mean, standard deviation, power spectrum, jerk, skewness, and eye acceleration (a total of 281 features), before selecting the top 30 most important features based on a random forest (RF) model. Splitting the Valence and Arousal responses into high and low categories based on sample means, the authors tested SVM, logistic regression (LR), RF, and ensemble learning models for classification in a 10-fold cross-validation scheme. They achieved the best accuracy with the RF model by attaining 75% and 84% for Valence and Arousal response variables, respectively.

While the collection of multivariate physiological signals has been posed to potentially improve the classification accuracy across ML models for specific human factors [12], it also carries with more obtrusive sensing techniques, which can affect the user experience itself. To mitigate obtrusive sensing that could impact user game experience, researchers started exploring the use of pure kinematic data that is unobtrusively collected from the VR hardware, thus removing the need of adding sensors and wires to the data collection process. Moreover, this approach is more implementable in everyday applications as it does not require any additions or retrofits to out-of-the box VR technology.

Mustafa et al. developed a new biometric authentication modality for VR by training ML models on kinematic data collected from accelerometer and gyroscope sensors in participants' VR headsets [13]. They collected the data from the accelerometer and gyroscope sensors from 23 users interacting with a simple VR game of following the path of a ball. First, each sensor's X, Y, Z, and magnitude streams of data were smoothed through a 10th order Butterworth filter with a 5 Hz cut-off frequency, before being split into 12-second windows with 50% overlap. The authors extracted a variety of time and frequency domain features from each of the four streams, resulting in a total of 178 features across both sensors. Principal component analysis (PCA) was applied and only principal components accounting for 95% of the explained variance were kept. The resulting features were used to authenticate the participants via training LR and SVM models to perform user classification. The LR model achieved the lowest classification error of about 7%. Head gesture recognition has been also explored using ML models that used kinematic variables (both linear and angular) using PCA-generated features, random forest, and neural networks as classifiers. The experiment also used kinematic data only collected from a single inertial measurement unit (IMU) located in a VR headset [14], showing the potential of this data to model and classify head gestures.

Zhao et al. also tested various machine learning models with kinematic features extracted from IMU sensors placed on the heads of children with autism spectrum disorder (ASD) [15]. The authors computed the rotation range and amount of rotation per minute in the pitch, raw, and roll directions. They then used linear mixed effects modelling to decide which features out of the six were statistically significant in delineating ASD children from those without ASD. Their experiments revealed the decision tree classifier using the rotation range in the roll direction, and amount of rotation in the raw direction, as most successful with an accuracy of 92%.

While previous work has explored the use of ML techniques for classifying human factors associated with healthcare, biometrics, and user experience among different populations, little is known in regards to the effects of head kinematics on user game experience metrics in older adults. The development of ML classifiers that enable predication of aspects such as motion sickness or perceived fatigue could significantly improve the creation of content that is more tailored to older adults' needs and preferences, potentially avoiding discomfort and improving the overall user experience. Moreover, similar ML algorithms could use unobtrusively and automatically collect kinematic data to flag possible cognitive decline for older adults using VR exergames. To achieve this goal, it is first necessary to determine which kinematic features could be relevant to key aspects such as cybersickness and cognitive performance.

The goal of this study is to investigate what classification algorithm and combination of head kinematic variables from VR headsets would best predict:

- Game user experience variables (e.g., perceived cyber sickness, perceived levels of enjoyment and perceived levels of exertion)
- Cognitive assessment variables (e.g., cognitive screening via MoCa and audiovisual response time)

## III. Methodology

In order to develop classifiers that are able to determine factors associated with older adult's cognitive functions and game user experience, we first developed a data pipeline that describes the methodologies used to acquire, clean, and extract relevant features for the overall processing of the data before using it as input for the ML classifiers (see 1).

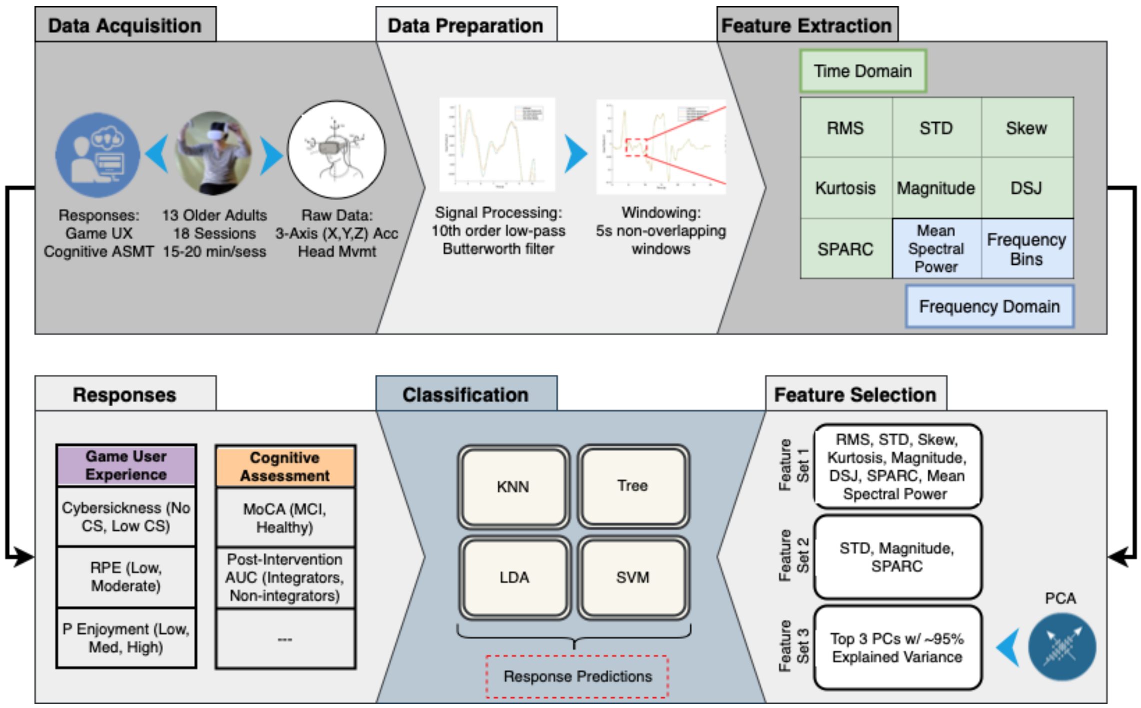


Fig. 1. Machine learning analysis pipeline.

### A. Data Acquisition

Our study used data from our pilot trials of *Seas the Day*, a custom-made VR exergame created to promote physical activity in older adults with and without cognitive impairments [16]. *Seas the Day* runs on portable VR-HMD (Oculus Quest 2) to encourage older adults to play a 15-20 minutes game that is tailored to engage upper-limb and core physical exercises in a seated position. *Seas the Day* has three different successive games: Tai- Chi (stretching, 3-5 minutes), rowing (exercise conditioning stage, 9 minutes) and fishing (cool-down, 3 minutes) and it was created following a user-centered design process involving people living with dementia, mild-cognitive impairment and exercise professionals. Game-play screenshots of the VR exergame are shown in (see 2). Data was collected during a remote intervention that happened during COVID-19 lock-downs in Canada during the summer and fall of 2021. A group of 13 community-dwelling older adults participated in the intervention, which involved a training program a total of 18 sessions of exercise using VR at-home playing *Seas the Day* (6 weeks with 3 game play sessions per week). The study received approval of the ethics research board at the University of Waterloo in Canada (ORE 42908). The game was developed using a data logging system that allows recording of **kinematic data** from the VR headset at 60 samples per second (60 Hz). Kinematic data from the inertial measurement unit embedded in the VR headset consisted of recorded accelerations in the X, Y and Z axis. After completing the session, kinematic data were automatically stored in the internal memory of the VR headset in structured in time series using comma separated values (CSV) files of about 15 MB per session. Raw data were extracted upon the completion of the training program and further processed to extract the kinematic features.

### B. Target variables for classification: responses

Two categories of target variables were collected for the analysis by the ML classifiers: game experience and cognitive function. Regarding **game experience**, we explored three different aspects:

1) *Perceived cybersickness (CS)*: A simplified self-reported tool was used to investigate commonly experienced motion sickness or nausea feelings during the intervention. A custom-made scale was used to collect levels of cybersickness. Half way through the trials, participants were asked to report their perceived levels of CS using a 0 (no nausea) to 10 (extreme nausea) Likert scale. Participants were asked to weight their overall feelings considering the previous sessions they have completed (around 9). Responses were collected via video calls with the participants. The CS variable was categorized into no CS (0 in the CS scale), for 5 participants, and mild CS (1-4 in the scale), for 7 participants, based their self-reported states, with one participant's data being unusable due to errors in data collection. No participant who completed the study reported greater than 4; two participants who are not included in this paper's analysis dropped out early into the study and cited cybersickness as the primary cause.
2) *Perceived levels of enjoyment*: Along with the VR headset, participants were provided with printed materials to complete after each session, which included a pictorial scale with five different faces expressing emotions (not enjoyable at all, slightly enjoyable, neutral, very enjoyable, extremely enjoyable) asking participants to rate how much they enjoyed the *Seas the Day* exergaming session. The perceived enjoyment responses were

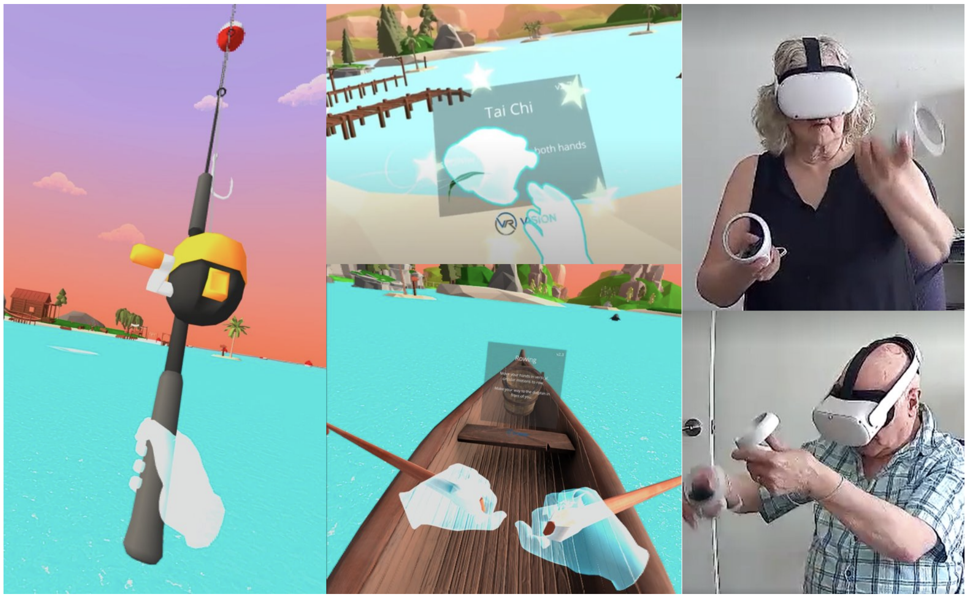


Fig. 2. Seas the Day VR exergame. Left: Fishing stage, Top center: Tai-Chi stage, Bottom center: Rowing stage, Right: Participants playing the game while seated.

categorized into three groups: scores $< 3$ were labelled as low, those equal to 3 were medium, and scores $> 3$ were denoted high.

3) *Rating of perceived exertion (RPE)*: A pictorial self-reported scale for perceived levels of physical exhaustion was used to collect the RPE after the VR exergame session. This pictorial scale used drawings that were used in previous work (e.g., [17]) to illustrate different stages of physical exhaustiveness on a 1-10 scale (e.g., 0=extremely easy, easy, somewhat easy, somewhat hard, hard, 10=extremely hard). RPE was similarly categorized into two levels of low for RPE scores $< 3$ and moderate for scores $> 3$ but $< 6$. There were no reported RPE scores greater than 6, which are considered to be intense.

Regarding **cognitive assessment** variables, data from two categories were collected:

1) *Cognitive assessment using Montreal Cognitive assessment (MoCa)*: A remote version of the MoCa test was administered virtually as screening tool assess each participant's baseline cognition. MoCa is widely used to screen older adults with cognitive impairments. The test was administered remotely during a video call and it lasted around 10-15 minutes per participant. The total possible score is 30 points, assessing different cognitive domains such as attention, concentration, executive functions, memory, language, conceptual thinking, calculations, and orientation [18]. For the purpose of this study, MoCA scores were divided into two groups using a cut-off score of 26 points, resulting in 9 participants with a high MoCA score (e.g., healthy), and 4 with a MoCA score suggesting mild cognitive impairment.
2) *Multisensory response time (RT)*: RT is a task used to evaluate multisensory integration via pressing buttons based on audiovisual stimuli presented on a computer screen. The computerized tasks are used to measure the time participants take to press the response button (e.g., space bar) as soon as they detect target stimuli. Specifically, the area under the curve was used to compute the degree to which participants are more or less likely to integrate multisensory information in order to generate faster response times (see [19], for recent application of this technique). Results from this task were categorized into two as 7 integrators with positive scores and 6 non-integrators for negative scores.

### C. Kinematic Data Preparation

After compiling a unified database with data from the 13 participants and 18 sessions, we used MATLAB to pre-process the data and run the ML classifiers.

Raw data was first filtered using low-pass filtering. Via visual examination of the data, the frequency content of the head kinematic data was identified to be at the 1 Hz or lower range. As such, four different filters were tested on the data: 1) 10th order Butterworth with 1 Hz cut-off, 2) 2nd order Butterworth with 1 Hz cut-off, 3) 2nd order 1D median filter, and 4) 10th order 1D median filter. MATLAB's *filtfilt* command was used to adjust for the noticeable time-delay introduced by applying the Butterworth filters. By visual inspection, it was quickly apparent that the median filters were ineffective compared to the Butterworth filters. Additionally, the 10th order Butterworth filter was able to capture the peaks and troughs of the signal better, thus it was selected as best filter for our data.

After filtering the signal, we created partitions or windows of the data to further extract the relevant features. The data of each axis was partitioned or windowed into 5-second non-overlapping windows, which is similar to what has been done previously for head gesture recognition using motion sensors [27].

TABLE I
EXTRACTED TIME AND FREQUENCY DOMAIN FEATURES.

| Domain | Feature | Explainability | Citations | Included in models? |
|---|---|---|---|---|
| Time | Root Mean Square | Expresses average magnitude of measures in each axis stream | [20] | No |
| | Magnitude (Euclidean Norm) | Expresses the total amount of movement during a game session | [20], [21] | Yes |
| | Standard Deviation | Expresses the variation in movements for a given axis/plane | [20], [22] | Yes |
| | Skew | Expresses comparability of outliers in movement to the mean movement value in each axis/plane | [20], [22] | No |
| | Kurtosis | Expresses amount of outliers in movement along an axis/plane | [20], [22] | No |
| | Dimensionless Jerk | Expresses smoothness in movement | [22], [23] | No |
| | SPARC | Expresses smoothness in movement (with less sensitivity to measurement noise) | [24] | Yes |
| Frequency | Mean Power | Expresses axis containing highest frequency content of movement | [20], [22], [25] | No |
| | Most Active Frequency Bin | Expresses range of frequencies characterizing most movement | [21], [26] | No |

### D. *Feature Extraction*

A feature extraction stage followed the data preparation process. In this stage, we explored some of the most widely used and reported head kinematic features in the literature [20]–[26]. Several time and frequency domain features were extracted from the data, as tabulated in Table 1. Time domain features include the root mean square (RMS), standard deviation (STD), skew, kurtosis, magnitude, dimensionless jerk, and SPARC. Dimensionless jerk and SPARC metrics were calculated using the original formulas provided in [22] and [24], respectively. Frequency domain features were computed from Welch's Power Spectral Density (PSD) estimate and included the mean spectral power and the frequency bin of width 0.5 Hz with the most frequency content. Each of these features offered some level of potential characterization of the user's kinematic movement during game-play, as detailed in the 'Explainability' column of Table 1. All features, apart from magnitude, which was computed as the Euclidean norm, were first calculated for the X, Y, and Z, axes independently. The features were then averaged across these three axes to obtain a scalar value for each. Table 1 presents a summary of both the time and frequency domain kinematic features used for the classification as well as the explanation of each metric. A column that verifies whether or not the features have been used for ML classifiers in the past was also added to the table.

### E. *Feature Selection*

Three feature sets were constructed out of the pool of features extracted from the raw kinematic data. The first feature set included all 8 features that have been widely used in the literature, except for frequency bin containing most frequency content. The frequency bin feature was dropped because the computed values did not show measurable changes across sessions or participants. The second feature set was grouped the STD, magnitude, and SPARC features. The shortlisting of these three features was based on a principal component analysis (PCA) for feature selection (similar to what has been done in [28]). We used a vector plotting approach that allows the identification of redundant variables that could be removed. The analysis revealed that the first two principal components accounted for approximately 75% of variation in the data. In particular, the STD and magnitude features were both among the top two most important features influencing the first principal component were also used often in previous works pertaining to kinematic data feature extraction [20]–[22]. The SPARC metric was chosen since it was a top contributor to the second principal component and also a more refined kinematic jerk measurement as compared to dimensionless jerk, due to its lower susceptibility to accelerometry-based noise artifacts [24]. The third feature set was comprised of the first three principal components as these accounted for 95% of explained variance.

### F. *Classification*

We used the MATLAB's Classification Learner tool for rapid experimentation with standard ML models once the three feature sets had been made. The four ML models tested during classification experiments were: boosted Decision Tree (DT), K-nearest neighbors (KNN), support vector machines (SVM), and linear discriminant analysis (LDA). We define classifier accuracy as the correct classification rate. In terms of parametrization, DT models had 100 maximum splits using the Gini's diversity index as split criterion, KNN models used a Euclidean distance metric with K=1, and SVM models had a linear kernel (all of which were parameters values initially set by the Classification Learner tool). Each experiment used a holdout validation methodology with 80% of data partitioned for training and the remaining 20% used for validation.

## IV. RESULTS

Results are summarized in Table 2. The model names are suffixed with the feature set (FS) they were trained on. Only one of the ML classifiers (perceived enjoyment) used ternary classifications; the rest of the response variables used binary classifiers. Overall, the DT models consistently performed best across domains when used with the first set of features (i.e., all the features) as can be seen in the bold values in Table 2. For game user experience, the classification of the RPE exhibited the best accuracy (67.9%) compared to 59.6% achieved for cybersickness and 53.4% for perceived enjoyment. Similar classification accuracy values were achieved when the second set of features (STD, Magnitude, SPARC) were used with the same DT classifier model. The classification of the MoCa

TABLE II
CLASSIFICATION ACCURACY ON GAME EXPERIENCE AND COGNITIVE ASSESSMENT RESPONSES.

| | Game Experience Response Accuracy (%) | | | Cognitive Assessment Response Accuracy (%) | |
|---|---|---|---|---|---|
| Model | CS | RPE | Perceived Enjoyment | MoCA | Multisensory RT |
| KNN-FS1 | 54 | 58.3 | 42.5 | 60.9 | 55.2 |
| KNN-FS2 | 53.9 | 57.6 | 42.7 | 59 | 53.7 |
| KNN-FS3 | 51 | 56 | 40 | 57.8 | 51 |
| LDA-FS1 | 53 | 66.7 | 51.6 | 68.9 | 53.6 |
| LDA-FS2 | 52.8 | 66.6 | 51.6 | 68.9 | 54.4 |
| LDA-FS3 | 53.1 | 66.7 | 51.6 | 68.9 | 53.6 |
| SVM-FS1 | 53.1 | 66.7 | 51.7 | 68.9 | 53.6 |
| SVM-FS2 | 50.1 | 56.7 | 51.6 | 68.8 | 53.6 |
| SVM-FS3 | 51.4 | 34.9 | 51.6 | 68 | 46.4 |
| DT-FS1 | **59.6** | **67.9** | **53.4** | **69.9** | **59.2** |
| DT-FS2 | 55 | 67.6 | 52.7 | 69.6 | **59.2** |
| DT-FS3 | 58.1 | 67.3 | 51.6 | 68.9 | 54.9 |

scores using head kinematic variables achieved 69.9% accuracy using the first set of features, whereas the maximum value of classification accuracy for the multisensory RT achieved 59.2% using both the first and second set of features.

## V. DISCUSSION

As the application of ML to kinematic data from VR exergames for older adults is an emerging field, the approaches explored in this paper are preliminary ones. Our approach aims to simplify data collection by using existing, embedded kinematic sensors in the HMD (as opposed to external physiological sensors, such as in [9]).

### A. ML for cognitive assessment

ML analysis of head-based kinematic data from custom-made VR showed some potential for predicting cognitive function. Similar approaches have been previously used to develop ' digital biomarkers' for the early detection of mild cognitive impairment in combination with clinical and neuropsychological data [29]. While the use of ML on HMD kinematic data requires greater research and refinement, the ML methods we explored shows reasonable potential in our exploratory analysis for use to predict cognitive function where cognitive status based on the MoCA was predicted with almost 70% accuracy. Adaptive VR games that leverage ML models to assess cognitive function and create real-time adaptations Would improve the value of VR in digital health applications because it would enable a more profound understanding and support of the people who are using it.

### B. ML for game user experience

While previous work has explored the use of ML techniques for classifying human factors associated with cybersickness, little is known with respect to the effects of head kinematics on cybersickness combined with other user game experience metrics such as perceived level of enjoyment and perceived level of exertion in older adults. Our results show that the classification for rating of perceived exertion exhibited the best accuracy (67.9%) compared to 59.6% for cybersickness and 53.4% for perceived enjoyment. Other factors such as the sense of presence should be examined in future work as it has been shown that the sense of presence in virtual reality is negatively related to cybersickness( [30]). Further, it has been shown that an enriched narrative integrated into the gaming experience can also reduce cybersickness, particularly in populations with less gaming experience ( [31]). As *Seas the Day* was specifically designed to increase a sense of presence, minimize cybersickness, and provide an enriched narrative ( [16], it is possible that cybersickness could be better predicted from head kinematics for content that is known to cause cybersickness (i.e., compare interventions that cause different levels of cybersickness with the same group of participants).

### C. Limitations

The results presented in this work are a starting point. Our data was from a small number of older adults who were playing the same game on the same VR system. Though we used a well-structured experimental trial design, the game play was at home. This has the advantage of being closer to a 'real world' setting but likely includes uncontrolled for distractions and changes in the environment that may have influenced the data. Additionally, kinematic data from the hand controllers could not be recorded because of hardware and data logging limitations in our setup, therefore we could not explore its impact on predicting game experience or cognitive function. A larger, more diverse sample (including people who are prone to cybersickness) playing a variety of games on different VR technologies and resulting data sets is needed to better understand which methods align well with different use-cases.

## VI. CONCLUSIONS AND FUTURE WORK

The goal of this research was to investigate ML classification algorithms to predict game experience and cognitive function among older adults using head kinematics collected during play of a VR-based exergame. Our results complement insights from others; specifically, our work contributes to feature engineering stage when using head kinematics. We demonstrated a relatively computationally-inexpensive and unobtrusive sensing technique has potential to be used in predicting cognitive function and perceived exertion among older adults, both of which were just shy of 70% accuracy, even in our small sample size. We are confident these results could be improved by adding more sensors to record kinematic information from the hands or trunk (some of which could be extracted from hand-held controllers), as well as adding more data regarding partcipants' status (e.g., recording cybersickness levels after each game level during each session).

In future studies, we aim to collect data from more participants, adding kinematic variables from other VR hardware (e.g., controllers, trackers) as well as exploring other types of ML algorithms. While our results are not definitive, they suggest that unobtrusive recording of head kinematic data from VR headsets combined with machine learning classifiers is a promising method to determine levels of cognition and user's game experience for older adult gamers. Greater research into this area is required to verify these hypotheses and to improve binary and ternary classification accuracies for real-time classification scenarios.

## References


[1] R. Y. C. Kwan, J. Y. W. Liu, K. N. K. Fong, J. Qin, P. K.-Y. Leung, O. S. K. Sin, P. Y. Hon, L. W. Suen, M.-K. Tse, and C. K. Lai, "Feasibility and effects of virtual reality motor-cognitive training in community-dwelling older people with cognitive frailty: Pilot randomized controlled trial," *JMIR serious games*, vol. 9, no. 3, p. e28400, 2021.

[2] M. Javaid and A. Haleem, "Virtual reality applications toward medical field," *Clinical Epidemiology and Global Health*, vol. 8, no. 2, pp. 600–605, 2020.

[3] P. Fuchs, *Virtual reality headsets-a theoretical and pragmatic approach*. CRC Press, 2017.

[4] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, "Activity recognition from accelerometer data," in *Aaai*, vol. 5, no. 2005. Pittsburgh, PA, 2005, pp. 1541–1546.

[5] A. Schwarz, C. M. Kanzler, O. Lambercy, A. R. Luft, and J. M. Veerbeek, "Systematic review on kinematic assessments of upper limb movements after stroke," *Stroke*, vol. 50, no. 3, pp. 718–727, 2019.

[6] J. Jerald, *The VR book: Human-centered design for virtual reality*. Morgan & Claypool, 2015.

[7] C. Hodent, *The gamer's brain: How neuroscience and UX can impact video game design*. Crc Press, 2017.

[8] D. Saredakis, A. Szpak, B. Birckhead, H. A. Keage, A. Rizzo, and T. Loetscher, "Factors associated with virtual reality sickness in head-mounted displays: a systematic review and meta-analysis," *Frontiers in human neuroscience*, vol. 14, p. 96, 2020.

[9] A. Garcia-Agundez, C. Reuter, H. Becker, R. Konrad, P. Caserman, A. Miede, and S. Göbel, "Development of a classifier to determine factors causing cybersickness in virtual reality environments," *Games for health journal*, vol. 8, no. 6, pp. 439–444, 2019.

[10] C.-Y. Liao, S.-K. Tai, R.-C. Chen, and H. Hendry, "Using eeg and deep learning to predict motion sickness under wearing a virtual reality device," *IEEE Access*, vol. 8, pp. 126 784–126 796, 2020.

[11] D. J. Diaz-Romero, A. M. R. Rincón, A. Miguel-Cruz, N. Yee, and E. Stroulia, "Recognizing emotional states with wearables while playing a serious game," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–12, 2021.

[12] D. Novak, M. Mihelj, and M. Munih, "A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing," *Interacting with computers*, vol. 24, no. 3, pp. 154–172, 2012.

[13] T. Mustafa, R. Matovu, A. Serwadda, and N. Muirhead, "Unsure how to authenticate on your vr headset? come on, use your head!" in *Proceedings of the Fourth ACM International Workshop on Security and Privacy Analytics*, 2018, pp. 23–30.

[14] T. Hachaj and M. Piekarczyk, "Evaluation of pattern recognition methods for head gesture-based interface of a virtual reality helmet equipped with a single imu sensor," *Sensors*, vol. 19, no. 24, p. 5408, 2019.

[15] Z. Zhao, Z. Zhu, X. Zhang, H. Tang, J. Xing, X. Hu, J. Lu, and X. Qu, "Identifying autism with head movement features by implementing machine learning algorithms," *Journal of Autism and Developmental Disorders*, vol. 52, no. 7, pp. 3038–3049, 2022.

[16] J. Muñoz, S. Mehrabi, Y. Li, A. Basharat, L. E. Middleton, S. Cao, M. Barnett-Cowan, J. Boger *et al.*, "Immersive virtual reality exergames for persons living with dementia: User-centered design study as a multistakeholder team during the covid-19 pandemic," *JMIR Serious Games*, vol. 10, no. 1, p. e29987, 2022.

[17] R. F. Gearhart Jr, K. M. Lagally, S. E. Riechman, R. D. Andrews, and R. J. Robertson, "Strength tracking using the omni resistance exercise scale in older men and women," *The journal of strength & conditioning research*, vol. 23, no. 3, pp. 1011–1015, 2009.

[18] Z. S. Nasreddine, N. A. Phillips, V. Bédirian, S. Charbonneau, V. Whitehead, I. Collin, J. L. Cummings, and H. Chertkow, "The montreal cognitive assessment, moca: a brief screening tool for mild cognitive impairment," *Journal of the American Geriatrics Society*, vol. 53, no. 4, pp. 695–699, 2005.

[19] A. Basharat, J. R. Mahoney, and M. Barnett-Cowan, "Temporal metrics of multisensory processing change in the elderly," *Multisensory Research*, vol. 32, no. 8, pp. 715–744, 2019.

[20] T. Mustafa, R. Matovu, A. Serwadda, and N. Muirhead, "Unsure how to authenticate on your vr headset? come on, use your head!" in *Proceedings of the Fourth ACM International Workshop on Security and Privacy Analytics*, ser. IWSPA '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 23–30. [Online]. Available: https://doi.org/10.1145/3180445.3180450

[21] C. Floris, S. Solbiati, F. Landreani, G. Damato, B. Lenzi, V. Megale, and E. G. Caiani, "Feasibility of heart rate and respiratory rate estimation by inertial sensors embedded in a virtual reality headset," *Sensors*, vol. 20, no. 24, p. 7168, 2020.

[22] D. J. Diaz-Romero, A. M. R. Rincón, A. Miguel-Cruz, N. Yee, and E. Stroulia, "Recognizing emotional states with wearables while playing a serious game," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–12, 2021.

[23] N. Hogan and D. Sternad, "Sensitivity of smoothness measures to movement duration, amplitude, and arrests," *Journal of motor behavior*, vol. 41, no. 6, pp. 529–534, 2009.

[24] A. I. Figueiredo, G. Balbinot, F. O. Brauner, A. Schiavo, R. R. Baptista, A. S. Pagnussat, K. Hollands, and R. G. Mestriner, "Sparc metrics provide mobility smoothness assessment in oldest-old with and without a history of falls: A case control study," *Frontiers in physiology*, vol. 11, p. 540, 2020.

[25] S. J. Preece, J. Y. Goulermas, L. P. Kenney, and D. Howard, "A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 3, pp. 871–879, 2008.

[26] D. T. Bundy, N. Szrama, M. Pahwa, and E. C. Leuthardt, "Unilateral, 3d arm movement kinematics are encoded in ipsilateral human cortex," *Journal of Neuroscience*, vol. 38, no. 47, pp. 10 042–10 056, 2018.

[27] N. Rudigkeit, M. Gebhard, and A. Graser, "An analytical approach for head gesture recognition with motion sensors," in *2015 9th International Conference on Sensing Technology (ICST)*. IEEE, 2015, pp. 1–6.

[28] F. Song, Z. Guo, and D. Mei, "Feature selection using principal component analysis," in *2010 international conference on system science, engineering design and manufacturing informatization*, vol. 1. IEEE, 2010, pp. 27–30.

[29] S. Cavedoni, A. Chirico, E. Pedroli, P. Cipresso, and G. Riva, "Digital biomarkers for the early detection of mild cognitive impairment: artificial intelligence meets virtual reality," *Frontiers in Human Neuroscience*, vol. 14, p. 245, 2020.

[30] S. Weech, S. Kenny, and M. Barnett-Cowan, "Presence and cybersickness in virtual reality are negatively related: a review," *Frontiers in psychology*, vol. 10, p. 158, 2019.

[31] S. Weech, S. Kenny, M. Lenizky, and M. Barnett-Cowan, "Narrative and gaming experience interact to affect presence and cybersickness in virtual reality," *International Journal of Human-Computer Studies*, vol. 138, p. 102398, 2020.