Prediction of cybersickness in virtual environments using topological data analysis and machine learning

Azadeh Hadadi, Christophe Guillet, Jean-Rémy Chardonnet, Mikhail Langovoy, Yuyang Wang, Jivka Ovtcharova

Prediction of Cybersickness in Virtual Environments using Topological Data Analysis and Machine Learning

Azadeh Hadadi [1,2,*], **Christophe Guillet** [3], **Jean-Rémy Chardonnet** [1], **Mikhail Langovoy** [2], **Yuyang Wang** [4] **and Jivka Ovtcharova** [2]

[1] *Arts et Metiers Institute of Technology, LISPEN, HESAM Université, UBFC, France*
[2] *Institute for Information Management in Engineering (IMI), Karlsruhe Institute of Technology, Germany*
[3] *Université de Bourgogne, LISPEN, UBFC, France*
[4] *Computational Media and Arts Thrust, Hong Kong University of Science and Technology, China*

Correspondence*:
Azadeh Hadadi
azadeh.hadadi@ensam.eu

ABSTRACT

Recent significant progress in Virtual Reality (VR) applications and environments raised several challenges and proved to have side effects on certain users, thus reducing the usability of the VR technology in some critical domains such as flight and car simulators. One of the common side effects is cybersickness. Some of the major cybersickness symptoms are nausea, oculomotor discomfort, and disorientation. To mitigate these symptoms and consequently improve the usability of VR systems, it is necessary to predict the incidence of cybersickness. In this paper, we propose a machine learning approach to cybersickness prediction in VR, on the basis of both physiological and subjective data. We investigated combinations of topological data analysis with a range of classifier algorithms and assessed the performance of classification. The highest performance of Topological Data Analysis (TDA)-based methods was achieved in combination with SVMs with Gaussian RBF kernel, indicating that Gaussian RBF kernels provide embeddings of physiological time series data into spaces that are rich enough to capture the important geometric features of this type of data.

Comparing several combinations with feature descriptors for physiological time series, the performance of the TDA+SVM combination is in the top group, statistically being on par or outperforming more complex and less interpretable methods.

Surprisingly, our results show that heart rate does not seem to correlate with cybersickness.

Keywords: Virtual Reality, Cybersickness, Navigation, TDA, Persistent Homology, Machine Learning

1 INTRODUCTION

Virtual Reality (VR) is one of the main focuses of the emerging technologies and research domain. The achievement in this domain opens a new horizon into the 3D world to explore which was not possible a few decades ago. The development of VR technology includes both software and hardware aspects. One of the major hardware developments of VR technology was to make scaled displays such as Head-Mounted Displays (HMD) and scaled-1 (real scale) displays like CAVE feasible. Since HMD with its open-source Software Development Kit (SDK) is now publicly available and considered a cost-effective VR, most of the current research is focusing on this type of VR technology. The environment developed for a VR platform is substantially different from games or 2D screen apps. The VR platform is essentially designed to immerse the user in the environment partially or completely while it is not always true for a game application. This is because on a VR platform, users are partially or completely immersed (Merienne, 2017) in the environment and they can experience physical effects similar to real environments but with slightly different sensations. Therefore, substantial efforts are made to minimize the difference that a user feels in VR with respect to the real environment.

Generally, navigation in a Virtual Environment (VE) is defined as the movement between two points, either to execute a task or purely to explore the environment. This basic human capability is considered one of the fundamental features of VR (Diersch and Wolbers, 2019). In a virtual-navigation task, the user usually moves in an environment confined to a physical area, i.e., the VR platform's physical border. A navigation task often involves hand-centric devices (e.g., joysticks). Besides, it is impossible to map directly real walking to virtual walking, even using for instance travel devices such as omnidirectional treadmills. This is because some sensory feedback is missing. Furthermore, adaptation in VR does not take place as in a real environment and always, there is a mismatching. The mismatching and the missing feedbacks lead to some adverse effects and sensory conflict at the onset or session of a sensory rearrangement (Chardonnet et al., 2021). The sensory conflict literary is interpreted as "cybersickness". Cybersickness also called simulator sickness or Virtual Reality Induced Sickness Effects (VRISE), is a kind of motion sickness (Mazloumi Gavgani et al., 2018). It is considered one of the serious challenges of virtual navigation which poses a severe impact on the usability of VR applications. It emerges as discomfort, nausea, headache, and vomiting, in severe cases and is associated with the discrepancies perceived between real and virtual worlds during motion.

There are two methods to evaluate cybersickness: subjective one, using questionnaires and objective one, through physiological and behavioural measurements (Niu et al., 2020). In a subjective evaluation, typically, participants will experience the VR task such as navigation or interaction. After exposure, they complete a survey to assess system comfort. To achieve this aim, various questionnaires, e.g., motion sickness questionnaire (MSQ) (Frank et al., 1983), Simulator Sickness Questionnaire (SSQ) (Kennedy et al., 1993), Fast Motion Sickness Scale (FMS) (Keshavarz and Hecht, 2011), and VR sickness questionnaire (VRSQ) (Kim et al., 2018) were designed to measure the sickness levels in different contexts and they are considered as the cornerstone of the approach. Such methods are however limited as they report a posteriori feedback, which prevents any possibility of acting efficiently to limit cybersickness. When it comes to objective evaluation of this adverse VR side effect, signals like postural sway (Chardonnet et al., 2017)(Lee et al., 2019) galvanic skin response (GSR) (Plouzeau et al., 2018), known as electrodermal activity (EDA) in some literatures, electroencephalograph (EEG) (Kim et al., 2019)(Jeong et al., 2019)(Liao et al., 2020)(Lin et al., 2013), or electrocardiogram (ECG) (Garcia-Agundez et al., 2019) are used to assess physiological response and to complement subjective data from questionnaires. In this type of evaluation, the participant is immersed in a VE to perform a task, while simultaneously physiological indicators are

monitored, and instantaneous signals are recorded within the exposure time. The signals are processed and analysed to identify the extent of cybersickness during the exposure and to determine the impact of the VR task on participants. Though, when indicators of cybersickness are detected in these signals, the onset of cybersickness has already passed, which limits the possibility to prevent users from avoiding cybersickness effects. The need to better control the evolution of these signals becomes prevalent to ensure cybersickness will not rise, thus justifying the interest in predicting and interpreting cybersickness. Since each physiological feedback changes over time, they can be represented as a time series signal (Pincus and Goldberger, 1994). Time series is a real-valued function over a bounded time domain I and defined as:

$$f : I \to \mathbb{R} \tag{1}$$

When it comes to classifying time series data using machine learning algorithms, because of the temporal nature of the input data, many of these algorithms would not be directly applicable to raw time series. Therefore, additional preprocessing might be needed before using learning algorithms on time series data. This preprocessing could also sometimes improve predictive performance. There is a wide range of methods to analyze time series, ranging from bag-of-words models to deriving new metrics to imaging time series to artificial neural networks. Random Convolutional Kernel Transform (ROCKET) algorithm (Dempster et al., 2020) extracts the maximum and the proportion of positive values as two features from time series using a large number of random convolutional kernels. Bag-of-patterns algorithm (Lin et al., 2012) extracts sub-sequences from a time series, discretizes each real-valued subsequence into a discrete-valued word (a sequence of symbols over a fixed alphabet), and builds a histogram (feature vector) from word counts. The Word Extraction for Time Series Classification (WEASEL) algorithm (Schäfer and Leser, 2017) relies on discretizing Fourier coefficients and using a sliding-window approach applied to the time series, then extracts discrete features per window. Here, mathematics plays a role like the mentioned dedicated methods to analyze time series.

Topology is a mathematical theory that emerged to study the data from the perspective of geometrical structures, e.g., loops or holes (Zomorodian and Carlsson, 2005). Traditionally, data belonging to spaces equipped with a similarity measurement or metric spaces, are analyzed using a similarity metric such as Euclidean distance or Manhattan distance. While this approach is convenient and already well-developed, it ignores valuable information about the problem: the shape and the connectivity properties of the data. In complex multidimensional problems, the data additionally has a topological (geometric) structure that can be used to improve the analysis. We see that it would be beneficial to link the topology theory to computational methods.

TDA is a mathematical apparatus to bridge these two fields. TDA was initially popularized by Carlsson (Carlsson, 2009) and has its roots in the fields of topology (Hatcher, 2005), linear algebra (Strang, 2006), and graph theory (West, 2001). It provides a means to infer cluster-like geometrical structures in order to better understand the shape of data and discover patterns of all dimensions and elucidate even weak connections between them. Topological features do not rely on a specific coordinate system and can compare data derived from different platforms. Also, they are invariant under small deformations. Furthermore, TDA helps to create tools to represent the data in a compressed way. These properties allow TDA to take advantage of the topological information to further process the data and to do various machine learning tasks, e.g., classification, clustering, etc. (Moroni and Pascali, 2021).

In this paper, we used TDA as the feature extractor to classify the multivariate physiological time series of participants during a virtual-navigation experiment. We employed a Gaussian Radial Basis Function

(RBF) kernel Support Vector Machine (SVM) (Schölkopf et al., 2002) classifier to classify the time series windows into "sick" and "non-sick" occasions based on the difference in total sickness score extracted from SSQ before and after exposure. In the literature, this topic was studied using different classifiers and features (see (Garcia-Agundez et al., 2019), (Padmanaban et al., 2018), (Porcino et al., 2020)). We will compare the accuracy of our approaches with existing methods. As an important addition, we will explore the effect of different types of physiological data on detection accuracy.

To this end, our paper is organized as follows: first, we provide a short recapitulation of the basic concepts of TDA. We will present the state of research in this field in section 3, where we describe the previous approaches to the problem. In section 4 we present the database of physiological data used in our study, as well as the data structure and the recorded signals. In section 5, we will demonstrate the application of TDA and other approaches. The TDA-SVM (Gaussian RBF) classification result will be presented in section 6. It will be compared with other approaches, and the effect of physiological data on performance will be studied. Our paper will end up with a conclusion.

2 BACKGROUND

As discussed in the previous section, TDA uses some computational algorithms to keep track of the topological features and discover patterns of all dimensions in a point cloud.

Consider a point cloud $\chi = \{x_1, \ldots, x_n\}$, sampled from a space *M*, it will be mapped into the structures called simplicial complexes. A *k*-simplex is a set of *k + 1* indices from the given set χ. A simplicial complex Υ is a set of simplices such that for any $\sigma \in \Upsilon$ and any $\sigma' \subset \sigma, \sigma' \in \Upsilon$ as is shown in Figure 1. One of the common complexes includes Vietoris-Rips complexes $R(\chi, \varepsilon)$ (Hausmann, 1995) which we use in this work. This complex is constructed by placing ε-balls (ε that defines the radius of an imaginary ball) on each vertex, and adding edges whenever they overlap:

$$\chi' \in R(\chi, \varepsilon), \quad \chi' \subset \chi \Leftrightarrow d\left(x_i, x_j\right) < \varepsilon, \quad \forall x_i, x_j \in \chi' \tag{2}$$

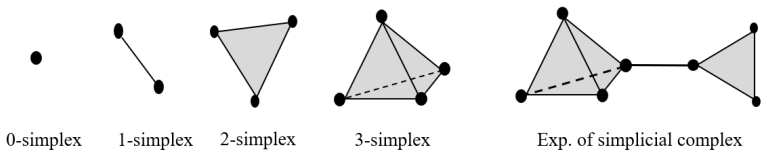

Figure 1. k-simplices in $\mathbb{R}^3, 0 \leq \mathrm{k} \leq 3$ and an example of simplicial complex

For a sample *X*, an interval over the scale ε can be found, for which the constructed simplicial complexes belong to the same class of topological invariants as *M*. By increasing ε a sequence of such complexes will be created which is called a filtration (Figure 2) with the property:

$$\varepsilon_1 \leq \varepsilon_2 \Longrightarrow X\left(\varepsilon_1\right) \subset X\left(\varepsilon_2\right) \tag{3}$$

During filtration, the classes of *n*-dimensional topological features -connected components (0-dimension), holes (1-dimension), cavities (2-dimension),... - appear at b_n^i and disappear at d_n^i using the values of ε

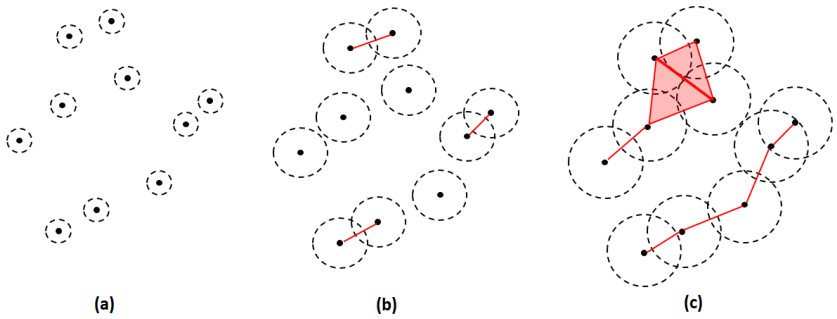

Figure 2. Example of filtration varying the filtration value ε which increased from (a) to (c). The black dot represents the point cloud data that are connected (red line) when the ε-balls around them overlap. The top part of (c) is the union of two adjacent triangles.

which be computed by persistent homology (Zomorodian and Carlsson, 2005). b_n^i and d_n^i referred to the birth and death values of the *i*-th class in dimension n, respectively. This information is represented by a collection of points $\left(b_n^i, d_n^i\right)$ which is drawn in the Cartesian plane $\mathbb{R}^2$ and called Persistent Diagram (Figure 3). A persistent diagram is a great tool for presenting the robustness and stability of features since the points near the diagonal are often considered noise while those further from the diagonal represent more robust topological features.

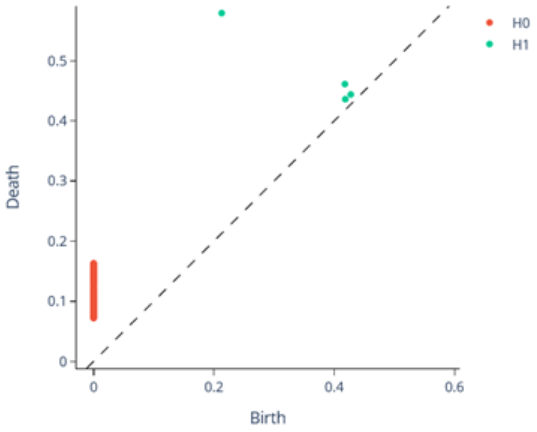

Figure 3. The corresponding persistence diagram with $H_0(\mathrm{X})$ in red and $H_1(\mathrm{X})$ in green, represents the persistence of connected components and holes over the ε-scale of filtration.

There are too few machine learning or statistical tools that can be applied directly to persistence diagram space. Hence, a mapping should be done from persistent diagram space to topological vector space which is appropriate for machine learning tools and further analysis. To achieve this aim by extracting scalar features, there are different methods like persistent image (Adams et al., 2017), persistent landscape (Bubenik, 2015), and persistent entropy (Rucco et al., 2017) methods. Persistent entropy is defined as the (base 2) Shannon entropy of the persistence diagram derived from the filtration. (For simplicity of notation, the log will refer to the log-base-2 function.)

$$\mathrm{E(F)} = -\sum_{1}^{\mathrm{n}} \mathrm{p_i} \log\left(\mathrm{p_i}\right), \text{ where } \mathrm{p_i} = \frac{\mathrm{l_i}}{\mathrm{S_L}}, \mathrm{l_i} = \mathrm{y_i} - \mathrm{x_i} \text{ and } \mathrm{S_L} = \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{l_i} \tag{4}$$

3 RELATED WORKS

There were past attempts to propose cybersickness prediction based on machine learning. The data used can be either based on stereoscopic 3D videos ((Padmanaban et al., 2018)), profile attributes ((Porcino et al., 2020)), or physiological signals like electrocardiographic ((Garcia-Agundez et al., 2019)). In almost all cases, questionnaires were mixed with objective data ((Padmanaban et al., 2018), (Porcino et al., 2020), (Garcia-Agundez et al., 2019)). Padmanabhan et al. (Padmanaban et al., 2018) presented a cybersickness prediction algorithm for desktop applications based on a symbolic machine learning model, such as bagged decision trees classifier (Rao and Potts, 1997) using optical flow as a feature. No physiological signal was recorded during the experiment. Only the combination of two sickness questionnaires: MSSQ and SSQ, was used to find a single sickness value. The precision of their method varied from 26% to 65% depending on the use case. There are classifiers that outperform this (see section 6).

Porcino et al. (Porcino et al., 2020), as Padmanabhan (Padmanaban et al., 2018), presented some classification results without measuring any physiological feedback. Instead, they worked based on profile attributes and concluded that the most relevant features were the exposure time, the z-axis rotation and profile attributes of the individual (gender, age, and VR experience). Moreover, the VRSQ was used to validate inconsistencies between subjective and objective data captured. As some details such as the correlation of the features with SSQ are absent, we could not reproduce their experiment. While very high precision was reported for some classifiers such as Random Forest (96.6% for binary classification of the data from both racing game and flight game scenarios), it is hard to compare and evaluate these results because no physiological feedback was acquired in that experiment.

Porcino et al. (Porcino et al., 2022) proposed an experimental analysis to estimate the weight of cybersickness causes and not to predict the presence of this phenomenon. These user and context-specific causes were ranked according to their impact using symbolic machine learning in VR games, including a racing game and a flight game. They conducted 6 experimental protocols along with 37 valid samples from a total of 88 volunteers. They used VRSQ to compare the user discomfort level with the verbal feedback collected during the experiment and thereby evaluated the data and discard incompatible samples. They achieved 0.79 and 0.95 AUC scores using decision tree and random forest algorithms, respectively. They concluded that exposure time, rotation, and acceleration are most likely the top factors contributing to cybersickness. Since this approach, unlike ours, was not to predict the presence of cybersickness, and the input data of their experiment was not based on the participant's physiological data, it is not possible to compare these results.

The experimental setting of this paper is closer to Garcia et al. (Garcia-Agundez et al., 2019). They collected electrocardiographic, electrooculographic, respiratory, and skin conductivity data from a total of 66 participants given a 10min experiment. They presented two classifiers to classify cybersickness, i.e., Binary and Ternary, based on KNN and SVM classifiers and achieved 82% and 56% of accuracy for cybersickness classification, respectively. Some approaches (see section 6) outperform the ternary classifier. In view of the relatively small number of observations, the occasional 82% accuracy of the binary classifier requires investigation with more data. The result of the binary classifier highly depends on several thresholds that need to be selected by the user to define sick people. In our future work, we plan to develop a multi-threshold version of the methods of the present paper.

4 DATA MEASUREMENT AND EXPERIMENT

To collect data and showcase our approach, we performed a user experiment. A total of 53 subjects, composed of 26 females and 27 males, having an age distribution with the mean and the standard deviation of 26.3 years and 3.3 years respectively, participated in a VR navigation experiment using an HTC Vive Pro head-mounted display. Participants were asked to repeat the experiment three times on three different days to gather enough samples. In that way, 159 samples were collected and included in the dataset. Upon arrival, participants were asked to sign a consent form and fill out one questionnaire to investigate their health conditions and experience in playing games and using VR devices before participating in the VR task. No issue was reported from this questionnaire. They were then explained the navigation task to achieve, as well as indications on how to navigate using the HTC Vive Pro hand controllers. Participants had to navigate in a virtual forest following a gravel path including curves and straight lines, which is depicted in Figure 4.

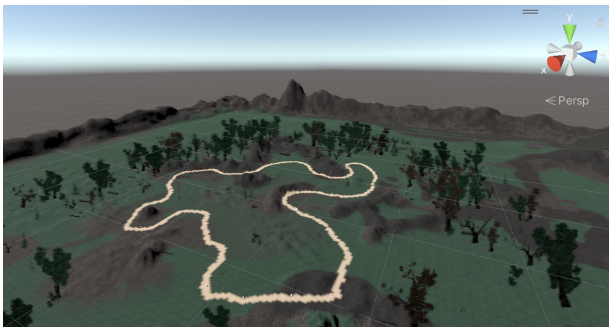

Figure 4. A virtual navigation environment used through the experiment. The highlighted line shows the navigation path.

Completing the navigation task took approximately four minutes. A Simulator Sickness Questionnaire (SSQ) was deployed in the experiment in which three different categories (nausea, oculomotor, disorientation) were measured in the form of 16 questions to quantify the degree of severity of each possible symptom of cybersickness. Participants filled one SSQ before and after the experiment to measure the psychological impact of the VR task on them. The difference between the pre-exposure and the post-exposure scores (called SSQ score) was included in the dataset:

$$\mathrm{SSQ} = SSQ_{\mathrm{post}} - \mathrm{SSQ}_{\mathrm{pre}} \tag{5}$$

It is worth noting that we used the SSQ as a subjective tool, as being very predominant and largely administered in most VR studies, despite the existence of strong debates about its validity and reliability in VR studies (e.g., (Sevinc and Berkman, 2020) or (Bouchard et al., 2021)) and recommendation to use VR-more-dedicated questionnaires, such as the VRSQ. The focus of this paper is on a methodology to grasp and predict cybersickness from any sickness-related data, we leave the use of different subjective means and their incidence on our method for future work.

An Empatica E4 wristband [1] on one participant's arm was used for real-time physiological data acquisition and particularly the electrodermal activity (EDA) of participants during this experiment. This wearable device is equipped with some sensors to gather high-quality data that was sent during navigation to a processing computer through Bluetooth. Galvanic Skin Response (GSR), blood volume pressure (BVP),

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

heart rate (HR), and temperature (TEM) were recorded during their navigation experiment. Moreover, the longitudinal (LG) and rotational (RT) accelerations were computed from the recorded navigation speed. An example of the recorded signals is shown in Figure 5. These sensors have different frequencies for measurement data sampling: EDA sensor 4Hz, PPG Sensor 64Hz (BVP), Infrared Thermopile 4Hz (TEM), 3-axis accelerometer sensor 32Hz, and average heart rate values are computed in spans of 10 seconds.

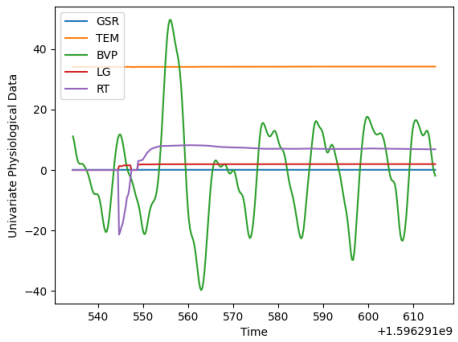

Figure 5. Sample of multivariate physiological data

5 METHODOLOGY AND IMPLEMENTATION

We propose an overall workflow in 4 steps for the complete prediction process as shown in Figure 6. The workflow represented in this figure has two main sub-processes: pre-processing, data structure analysis, labelling the data (steps 1-3), and classification (step 4).

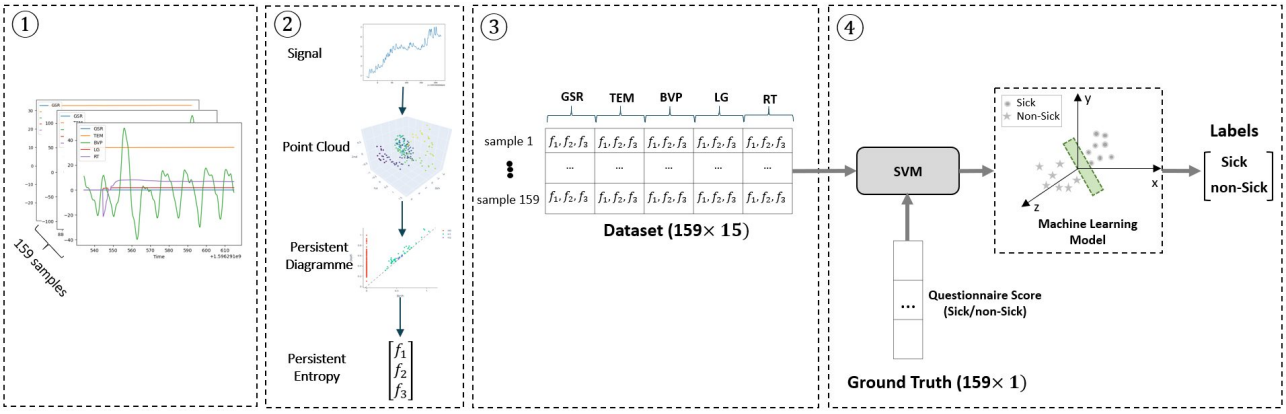

Figure 6. Block diagram of the cybersickness prediction method. 1) Normalization of the time series as a pre-processing step 2) Applying TDA and vectorizing persistent diagram to persistent entropy 3) Making the dataset (159 ×15 matrix) from the features descriptors that characterize the sample's data 4) Feeding the dataset into the SVM classifier along with ground truth labels (i.e., sick:1 and non-sick:0) provided by the questionnaire scores. Finally, SVM finds a suitable hyperplane (the green surface shown in the figure) that can cleanly separate the samples into two groups (i.e., sick and non-sick). The output of the method is a set of predicted labels based on the decision made by the SVM classifier.

Recorded data for each subject includes five sensor output variables as discussed in the previous section. Each variable corresponds to one physiological sensor data. Applying the persistent entropy, three features are obtained per each variable, namely, birth, death, and dimension. We obtain the dataset with 159 rows and 5×3 = 15 columns, where each row is related to each sample.

As data recording was performed with different frequencies, discussed in section 4, the timesteps of various time series were different. Therefore, as a pre-processing step, we normalized the data based on minimum timesteps throughout the dataset.

We improved our pre-processing in the workflow by adding a denoising approach. We applied Empirical Mode Decomposition (EMD) to the physiological data on every variable before applying the TDA. EMD is used to decompose the time series into a finite and often small number of components which is named its Intrinsic Mode Functions (IMFs) and residue series (Pereira and de Mello, 2015). To decompose a signal and get the IMFs, lower and upper envelopes are obtained by connecting all the local maxima/minima using a cubic spline. Subsequently, a low-frequency component is calculated using the mean of these envelopes. This component is subtracted from the original signal. Eventually, based on two specific criteria which are detailed in (Huang et al., 1998), the output signal is calculated as an IMF.

EMD determines what frequency with what strength in the signal occurs at any given moment. IMFs can be summed to recover the original signal. Because the first IMF usually carries the most oscillating (high-frequency) components, it can be rejected to remove high-frequency components (e.g., random noise). Figure 7 shows one example in which EMD was applied to the GSR time series. Original, decomposed and reconstructed signals are shown on the left, middle and right sides, respectively.

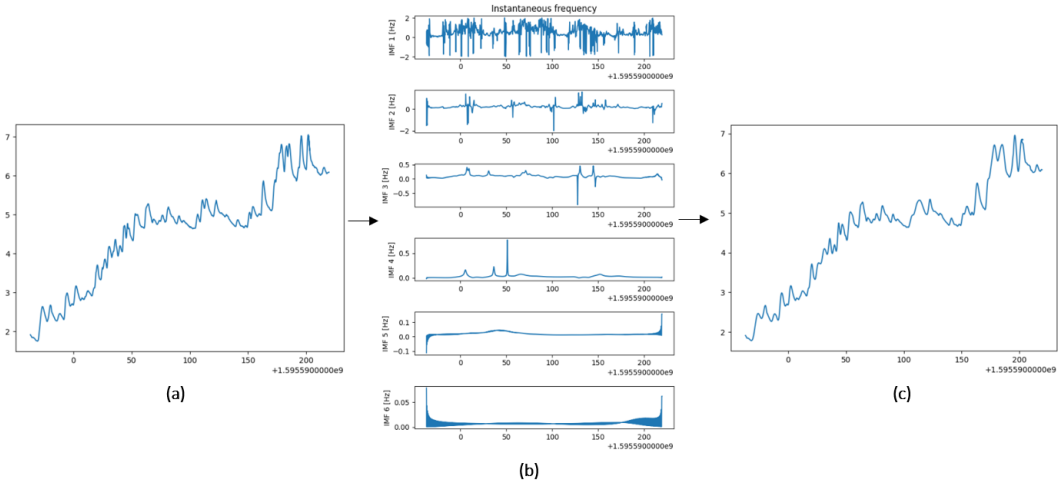

Figure 7. (a) Original GSR time series (b) Six IMFs decomposed by EMD (c) Output Series from EMD on the original signal. It is the sum of the last five IMFs plus residue.

After pre-processing, we additionally do qualitative data structure analysis. We detected the 0 (connected components), 1 (loop), and 2 (void) dimensional persistent topological features across multiple scales and used the time delay embedding method, based on the results of Taken's embedding theorem (Takens, 1981), which can be thought of as sliding a window of certain size over the signal. Each window is represented as a point in a high-dimensional space.

More formally, given a time series *f(t)*, a sequence of vectors extracted has the form:

$$\mathrm{f_i} = \left[\mathrm{f}\left(\mathrm{t_i}\right), \mathrm{f}\left(\mathrm{t_i} + 2\tau\right), \ldots, \mathrm{f}\left(\mathrm{t_i} + (\mathrm{M}-1)\tau\right)\right] \in \mathbb{R}^{\mathrm{d}} \tag{6}$$

where *(M-1)* is the embedding dimension and τ is the time delay. Hence, the window size is the quantity *(M-1)*τ and a stride is defined as the difference between t_i and t_{i+1}. In other words, the time delay embedding of f with parameters *(M,*τ*)* is the function:

$$\mathrm{TD_{M,\tau}f} : \mathbb{R} \to \mathbb{R}^d, \mathrm{t} \to \begin{bmatrix} \mathrm{f(t)} \\ \mathrm{f(t+\tau)} \\ \mathrm{f(t+2\tau)} \\ \cdots \\ \mathrm{f(t+(M-1)\tau)} \end{bmatrix} \tag{7}$$

As a result, we have two hyperparameters: *M*,τ. To determine the time delay automatically, we used the Mutual Information (MI) technique (Wen and Wan, 2009). MI is used as an analytical measure of the extent to which the values in the time series can be predicted by earlier values. At first, the Probability Density Function (PDF) of the time series is calculated with *n* bins. Given p_i as the probability that x_t is in the *i*th bin (marginal probability density distribution) and $p_{i,j}$ as the probability that x_t is in the *j*th bin while $x_{t+\tau}$ is in the *i*th bin (joint probability density distribution), MI is defined as Kulback-Liebler (KL) divergence between the $p_{i,j}$ and p_i and p_j i.e.

$$I(\tau) = -\sum_{i=1}^{n_{\text{bins}}} \sum_{j=1}^{n_{\text{bins}}} p_{i,j}(\tau) \log \frac{p_{i,j}(\tau)}{p_i p_j} \tag{8}$$

According to the MI technique, the optimal time delay can be computed as the minimum value of $I(\tau)$.

The False Nearest Neighbor (FNN) technique is used to get the optimal value for embedding dimension *M*. According to this algorithm, points lying close together due to projection, are separated in higher embedding dimensions, and conversely, nearest neighbour points which are close in one embedding dimension should be close in a higher one. Suppose that p_j is the nearest neighbor of p_i in m dimensional space. The Euclidean distance between p_i and p_j is:

$$R_m^2(i,j) = \sum_{k=0}^{m-1} [x(i+k\tau) - x(j+k\tau)]^2 \tag{9}$$

By adding one more dimension, the distance will change:

$$R_{m+1}^2(i,j) = R_m^2(i,j) + [x(i+m\tau) - x(j+m\tau)]^2 \tag{10}$$

then, the FNN criterion is defined as:

$$R_i = \left(\frac{R_{m+1}^2(i,j) - R_m^2(i,j)}{R_m^2(i,j)} \right)^{\frac{1}{2}} = \frac{|x(i+m\tau) - x(j+m\tau)|}{R_d(i,j)} > R_{th} \tag{11}$$

More formally if we have a point p_i and neighbor p_j , we check if the normalised distance R_i for the next dimension is greater than some threshold R_{th}. If $R_i > R_{th}$ then we have a false nearest neighbour, and the optimal embedding dimension is obtained by minimizing the total number of such neighbours.

In a nutshell, time delay embeddings translate a 1-dimensional time series to a d-dimensional time series in which the current value at each time with $(d-1)$ lag coordinates.

After data structure analysis, the physiological data were labelled using the self-reported sickness questionnaire, collected during the VR experiment from the participants. To achieve this aim, the SSQ score was collected at both pre-and post-exposure, and the score for each participant was calculated using original indexes (see section 4). We considered the SSQ score of 20 as the threshold to define the label of "sick" and "non-sick" (Bimberg et al., 2020). The participants whose SSQ score is equal to or greater than 20 are assumed that they suffered from cybersickness and "sick". Conversely, an SSQ score of less than 20 is defined as not experiencing cybersickness and labelled "non-sick". Based on this labelling, the 159 samples were divided into 87 and 72 "sick" and "non-sick" samples, respectively. This step leads to a table of data of size 159×15 as shown in Figure 6.

The next step of the workflow is the classification of the data. This includes selecting a proper classifier, training the classifier with the above data, and then evaluating the performance of the classifier on some test data. The data extracted in step 3 is used as an input to the machine learning classifier with two classes, i.e., "sick" and "non-sick".

To investigate the effect of the machine learning classification algorithms on the overall performance and the accuracy of the detection process, we selected several classifiers of different types and implement them in the workflow. First, we applied SVM classifiers with linear, polynomial (second degree) and Gaussian RBF (gamma $= 10^{-5}, \mathrm{C} = 1$, regularization parameter = 1) kernels. Second, we used Random Forest as an ensemble method considering two features when looking for the best split, 100 Decision Trees in the forest, Gini as the criteria with which to split on each node, minimum of 2 samples to split an internal node, and minimum 1 sample to be at a leaf node. As the last classifier algorithm, Logistic Regression with "lbfgs" as the solver which was used for the optimization problem, a maximum of 500 iterations was taken to converge the solvers, "l2" as the penalty, and C=1 to control the penalty strength was compared. In total, five classifiers were implemented and tested.

After applying EMD, we investigated the TDA performance on both raw and denoised signal reconstructed summing the last five IMFs plus residue. Additional classification algorithms that were implemented and compared to TDA, were the bag-of-patterns (Lin et al., 2012) with sliding window size 16, length of the words 4, and 4 bins to produce without numerosity reduction; ROCKET (Dempster et al., 2020) with 10000 kernels in sizes 7,9, and 11; WEASEL (Schäfer and Leser, 2017) with word size 9 and window sizes from 10 to 27. In all mentioned studies, SVM (Gaussian RBF) was used as the classifier.

Another point that we have investigated was which kind of physiological data has more influence on cybersickness prediction. We applied the compound approach of TDA with SVM (Gaussian RBF) to different combinations of variables. The performance and accuracy of each classifier were evaluated using the F1 score metric because we had an imbalanced classification problem.

$$\mathrm{F}^1 = \left(\frac{2}{\text{recall}^{-1} + \text{precision}^{-1}} \right) = 2. \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \tag{12}$$

$$\text{precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \ \text{recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{13}$$

To assess the generalization ability of the classifier and evaluate and test its performance, we used K-fold Cross-Validation (CV) technique (Berrar, 2019) with K=5. Also, we computed the evaluation metric, i.e., the F1 score and its mean and standard deviation (std) in every fold. Finally, we summarized the efficiency of the model using the averaging of model evaluation scores as demonstrated in Figure 8.

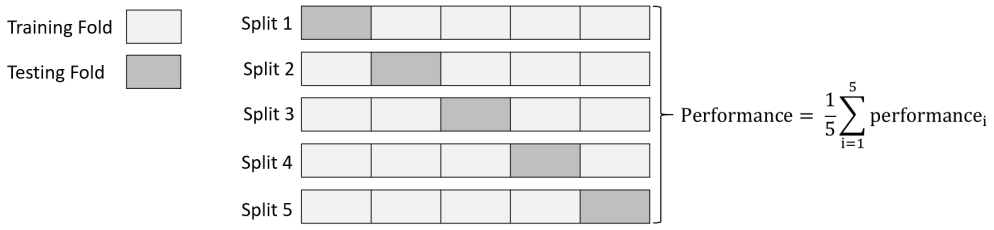

Figure 8. 5-Fold cross-validation in which data is divided into 5-folds composed of 4-folds for training the model and 1-fold for validation. The overall performance is obtained by computing the arithmetic mean.

6 RESULTS AND DISCUSSION

As discussed in section 5, first, we have presented the pre-processing, the TDA-based feature descriptor (in three steps) and machine learning classifiers to classify sensor data into two classes. Then, we analyzed the accuracy of TDA in combination with various classification algorithms. In the second investigation, we applied other types of feature descriptors which are dedicated to tackling time series classification and compared the results with TDA. In the third investigation, we studied the effect of the heat rate signal on the classification.

6.1 Comparison of Classification Algorithms

The effect of the classification algorithm on the correct detection of the affected subjects is demonstrated in Table 1. Figure 9 visually demonstrates these outcomes. As seen, the SVM (Gaussian RBF) presents a higher mean and lower standard deviation (std) than other classifiers in terms of the F1 score metric, which means a more accurate and more stable classification, with around 71% of accuracy. Interestingly, while the mean of the F1 metric decreases from SVM (Gaussian RBF) to SVM (linear kernel) in Table 1, the standard deviation (std) consistently increases conversely. The worst classification result was achieved by SVM (linear kernel), with an average precision of 61%.

Therefore, we consider SVM (with Gaussian RBF) to be the stronger classifier algorithm for this type of data, and we will be using it below to analyze the effect of feature descriptors on cybersickness detection. Our explanation for this performance is that Gaussian RBF kernels provide embeddings of physiological time series data into spaces that are rich enough to capture the important geometric features of these time series.

6.2 Comparison of Compound Classifiers

Five feature descriptors were selected: TDA, EMD+TDA, bag-of-patterns, ROCKET, as well as WEASEL, and applied to the data to extract features. The features were classified using SVM (Gaussian RBF) with the configuration detailed in section 5. The mean and the standard deviation of the F1 score are presented in Table 2 and visualized similar to the classifier effect as shown in Figure 10. The first finding is that the ROCKET feature descriptor achieved a little bit better precision, 71%, in turn, a higher mean F1 score

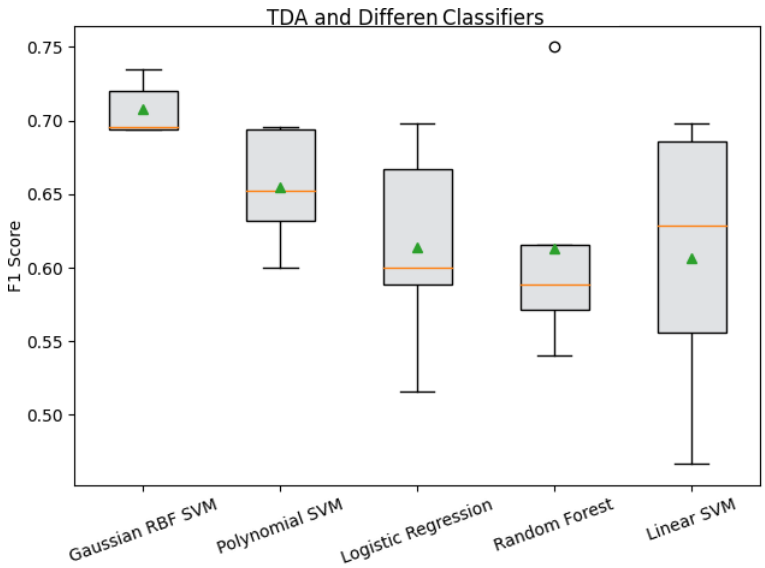

Figure 9. Performance evaluation and the effect of the classifier on sickness detection using the TDA feature descriptor

Table 1. Comparison between various classification algorithms in combination with TDA

Classifier Algorithm	F^1 **Score mean**	F^1 **Score std**
SVM (Gaussian RBF)	0.708	0.017
SVM (polynomial kernel)	0.655	0.037
Logistic Regression	0.614	0.064
Random Forest	0.613	0.073
SVM (linear kernel)	0.607	0.086

and less standard deviation (std) than TDA. However, the difference in the F1 score is only 0.002, with a 0.001 difference in standard deviation, on the dataset of 159-time series, which indicates that the difference between ROCKET and TDA is not statistically significant.

Table 2. Comparison between TDA and four other methods and their effect on performance

Feature Extraction	F^1 **Score mean**	F^1 **Score std**
ROCKET	0.710	0.016
TDA	0.708	0.017
EMD+TDA	0.664	0.051
bag-of-patterns	0.664	0.090
WEASEL	0.648	0.050

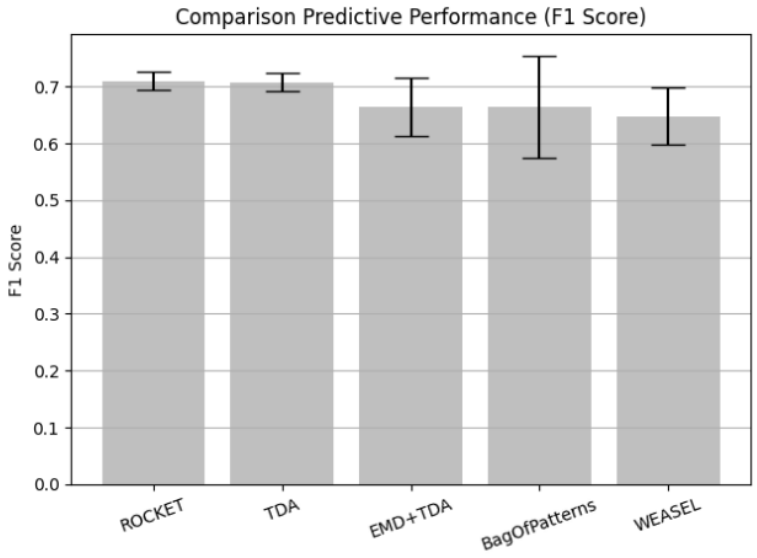

Figure 10. Performance evaluation using different feature extraction algorithms

As ROCKET is using a very large number of random convolutional kernels (10000 in this case), it is a more computationally demanding method than the TDA. Due to the technical similarities with convolutional neural networks, we expect that a rigorous mathematical or statistical performance analysis, as well as explicit interpretation of solutions, would be difficult for ROCKET. On the other hand, given that TDA is based on basic concepts of mathematical topology, and that SVMs are already proven to be universally consistent, combinations of TDA and SVMs are expected to be much easier for both rigorous analyses as well as for explainability.

It turned out that the mean of the evaluated metric of the TDA alone is higher than that of a more complex combination of TDA with EMD denoising and conversely, its standard deviation (std) is less. This implies that important details of the sensor data are removed by EMD during the denoising process. It can be concluded the EMD parameter shall be set more precisely taking into account the sampling frequency of each signal otherwise some useful high-frequency information, where they are close to the noise, can be easily eliminated with the improper setting of filtering parameters. A mathematical framework accounting for this effect was developed in (Langovoy, 2007). In all other cases, TDA has a higher mean F1 score and a smaller standard deviation (std).

6.3 Heart Rate and Cybersickness

The result of the third investigation is shown in Table 3. The inclusion of the heart rate features leads to lower precision and high stability and a significant standard deviation (std) increase. Therefore, this sensor data was excluded from the above analysis. Surprisingly, heart rate did not show to be a relevant predictor for cybersickness. Since the sampling rate of heart rate is less than the other signals, thereby a large portion of the signals and subsequently the details are removed during the normalization phase, leading to a decrease in prediction accuracy.

6.4 Conclusion

In this paper, we proposed a machine learning approach to cybersickness prediction in VR, on the basis of both physiological and subjective data. We investigated combinations of dynamic topological data analysis

Table 3. The effect of heart rate (HR) on cybersickness prediction

Physiological Data	F^1 **Score mean**	F^1 **Score std**
Exclude Heart Rate	0.708	0.017
Include Heart Rate	0.660	0.050

with a range of classifier algorithms and assessed the performance of classification using F1 score. The highest performance of TDA-based methods was achieved in combination with an SVM with a Gaussian RBF kernel. Our explanation for this performance is that Gaussian RBF kernels provide embeddings of physiological time series data into spaces that are rich enough to capture the important geometric features of these time series.

A comparison of TDA with other feature descriptors for physiological time series classification showed that the performance of TDA+SVM is at the top of the list: whilst it is slightly lower than ROCKET+SVM, the difference is not significant, and the accuracy is higher than combinations of SVMs with bag-of-patterns and WEASEL.

As ROCKET is using a very large number of random convolutional kernels (10000 in this case), it is a more computationally demanding method than the TDA. Due to the technical similarities with convolutional neural networks, we expect that a rigorous mathematical or statistical performance analysis, as well as explicit interpretation of solutions, would be difficult for ROCKET. Similar reasoning can be applied to complex methods such as bag-of-patterns and WEASEL. On the other hand, given that TDA is based on basic concepts of mathematical topology, and that SVMs are already proven to be universally consistent, combinations of TDA and SVMs are expected to be much easier for both rigorous analyses as well as for explainability. We noticed that performance of TDA-based methods on time series data got worse after adding the standard EMD reconstruction. This serves as a warning against the noncritical application of data smoothing and data transformation techniques.

Surprisingly, our results show that heart rate does not have any effect on cybersickness prediction. It still remains to investigate whether this would be the case for other types of VR experiments as well.

In addition, just a few machine learning or statistical tools can be applied directly to the persistence diagram space. We will attempt to create such machine learning tools by proposing visual perception-based metrics for persistence diagram spaces, similar to (Langovoy et al., 2014). This will allow the building of more direct and advanced combinations of machine learning methods with the TDA.

ACKNOWLEDGMENTS

This work was supported in part by a grant from the French-German University (UFA-DFH) No. CDFA 03-19.

REFERENCES

Adams, H., Emerson, T., Kirby, M., Neville, R., Peterson, C., Shipman, P., et al. (2017). Persistence images: A stable vector representation of persistent homology. *Journal of Machine Learning Research* 18

[Dataset] Berrar, D. (2019). Cross-validation.

Bimberg, P., Weissker, T., and Kulik, A. (2020). On the usage of the simulator sickness questionnaire for virtual reality research. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (IEEE), 464–467

Bouchard, S., Berthiaume, M., Robillard, G., Forget, H., Daudelin-Peltier, C., Renaud, P., et al. (2021). Arguing in favor of revising the simulator sickness questionnaire factor structure when assessing side effects induced by immersions in virtual reality. *Frontiers in psychiatry* 12

Bubenik, P. (2015). Statistical topological data analysis using persistence landscapes. *J. Mach. Learn. Res.* 16, 77–102

Carlsson, G. (2009). Topology and data. *Bulletin of the American Mathematical Society* 46, 255–308

Chardonnet, J.-R., Mirzaei, M. A., and Mérienne, F. (2017). Features of the postural sway signal as indicators to estimate and predict visually induced motion sickness in virtual reality. *International Journal of Human–Computer Interaction* 33, 771–785

Chardonnet, J.-R., Mirzaei, M. A., and Merienne, F. (2021). Influence of navigation parameters on cybersickness in virtual reality. *Virtual Reality* 25, 565–574

Dempster, A., Petitjean, F., and Webb, G. I. (2020). Rocket: exceptionally fast and accurate time series classification using random convolutional kernels. *Data Mining and Knowledge Discovery* 34, 1454–1495

Diersch, N. and Wolbers, T. (2019). The potential of virtual reality for spatial navigation research across the adult lifespan. *Journal of Experimental Biology* 222, jeb187252

Frank, L., Kennedy, R. S., Kellogg, R. S., and McCauley, M. E. (1983). *Simulator sickness: A reaction to a transformed perceptual world. 1. scope of the problem*. Tech. rep., ORLANDO Florida

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

Hatcher, A. (2005). *Algebraic topology*

Hausmann, J.-C. (1995). On the vietoris-rips complexes and a cohomology theory for metric spaces. *Annals of Mathematics Studies* 138, 175–188

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., et al. (1998). The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences* 454, 903–995

Jeong, D., Yoo, S., and Yun, J. (2019). Cybersickness analysis with eeg using deep learning algorithms. In *2019 IEEE conference on virtual reality and 3D user interfaces (VR)* (IEEE), 827–835

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

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

Kim, H. K., Park, J., Choi, Y., and Choe, M. (2018). Virtual reality sickness questionnaire (vrsq): Motion sickness measurement index in a virtual reality environment. *Applied ergonomics* 69, 66–73

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

Langovoy, M. (2007). Data-driven goodness-of-fit tests. *arXiv preprint arXiv:0708.0169*

[Dataset] Langovoy, M., Wübbeler, G., and Elster, C. (2014). Novel metric for analysis, interpretation and visualization of brdf data

Lee, S., Kim, S., Kim, H. G., Kim, M. S., Yun, S., Jeong, B., et al. (2019). Physiological fusion net: Quantifying individual vr sickness with content stimulus and physiological response. In *2019 IEEE International Conference on Image Processing (ICIP)* (IEEE), 440–444

Liao, C.-Y., Tai, S.-K., Chen, R.-C., and Hendry, H. (2020). Using eeg and deep learning to predict motion sickness under wearing a virtual reality device. *IEEE Access* 8, 126784–126796

Lin, C.-T., Tsai, S.-F., and Ko, L.-W. (2013). Eeg-based learning system for online motion sickness level estimation in a dynamic vehicle environment. *IEEE transactions on neural networks and learning systems* 24, 1689–1700

Lin, J., Khade, R., and Li, Y. (2012). Rotation-invariant similarity in time series using bag-of-patterns representation. *Journal of Intelligent Information Systems* 39, 287–315

Mazloumi Gavgani, A., Walker, F. R., Hodgson, D. M., and Nalivaiko, E. (2018). A comparative study of cybersickness during exposure to virtual reality and "classic" motion sickness: are they different? *Journal of Applied Physiology* 125, 1670–1680

[Dataset] Merienne, F. (2017). Virtual reality: Principles and applications

Moroni, D. and Pascali, M. A. (2021). Learning topology: bridging computational topology and machine learning. *Pattern Recognition and Image Analysis* 31, 443–453

Niu, Y., Wang, D., Wang, Z., Sun, F., Yue, K., and Zheng, N. (2020). User experience evaluation in virtual reality based on subjective feelings and physiological signals. *Electronic Imaging* 2020, 60413–1

Padmanaban, N., Ruban, T., Sitzmann, V., Norcia, A. M., and Wetzstein, G. (2018). Towards a machine-learning approach for sickness prediction in 360 stereoscopic videos. *IEEE transactions on visualization and computer graphics* 24, 1594–1603

Pereira, C. M. and de Mello, R. F. (2015). Persistent homology for time series and spatial data clustering. *Expert Systems with Applications* 42, 6026–6038

Pincus, S. M. and Goldberger, A. L. (1994). Physiological time-series analysis: what does regularity quantify? *American Journal of Physiology-Heart and Circulatory Physiology* 266, H1643–H1656

Plouzeau, J., Chardonnet, J.-R., and Merienne, F. (2018). Using cybersickness indicators to adapt navigation in virtual reality: a pre-study.ieee vr. In *2018 IEEE conference on virtual reality and 3D user interfaces (VR)* (IEEE), 661–662

Porcino, T., Rodrigues, E. O., Bernardini, F., Trevisan, D., and Clua, E. (2022). Identifying cybersickness causes in virtual reality games using symbolic machine learning algorithms. *Entertainment Computing* 41, 100473

Porcino, T., Rodrigues, E. O., Silva, A., Clua, E., and Trevisan, D. (2020). Using the gameplay and user data to predict and identify causes of cybersickness manifestation in virtual reality games. In *2020 IEEE 8th international conference on serious games and applications for health (SeGAH)* (IEEE), 1–8

Rao, J. S. and Potts, W. J. (1997). Visualizing bagged decision trees. In *KDD*. 243–246

Rucco, M., Gonzalez-Diaz, R., Jimenez, M.-J., Atienza, N., Cristalli, C., Concettoni, E., et al. (2017). A new topological entropy-based approach for measuring similarities among piecewise linear functions. *Signal Processing* 134, 130–138

Schäfer, P. and Leser, U. (2017). Fast and accurate time series classification with weasel. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. 637–646

Schölkopf, B., Smola, A. J., and Bach, F. (2002). *Learning with kernels: support vector machines, regularization, optimization, and beyond* (MIT press)

Sevinc, V. and Berkman, M. I. (2020). Psychometric evaluation of simulator sickness questionnaire and its variants as a measure of cybersickness in consumer virtual environments. *Applied ergonomics* 82, 102958

Strang, G. (2006). *Linear algebra and its applications.* (Belmont, CA: Thomson, Brooks/Cole)

Takens, F. (1981). Detecting strange attractors in turbulence. In *Dynamical systems and turbulence, Warwick 1980* (Springer). 366–381

Wen, F. and Wan, Q. (2009). Time delay estimation based on mutual information estimation. In *2009 2nd International Congress on Image and Signal Processing* (IEEE), 1–5

West, D. B. (2001). *Introduction to graph theory*, vol. 2 (Prentice hall, Upper Saddle River New Jersey)

Zomorodian, A. and Carlsson, G. (2005). Computing persistent homology. *Discrete & Computational Geometry* 33, 249–274