



Science Arts & Métiers (SAM)

is an open access repository that collects the work of Arts et Métiers Institute of Technology researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: <https://sam.ensam.eu>
Handle ID: <http://hdl.handle.net/10985/22777>



This document is available under CC BY license

To cite this version :

Azadeh HADADI, GUILLET CHRISTOPHE, Mikhail LANGOVOY, Yuyang WANG, Jivka OVTCHAROVA, Jean-Rémy CHARDONNET - Prediction of cybersickness in virtual environments using topological data analysis and machine learning - Frontiers in Virtual Reality - Vol. 3, - 2022

Any correspondence concerning this service should be sent to the repository

Administrator : scienceouverte@ensam.eu



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

Azadeh Hadadi^{1,2,*}, Christophe Guillet³, Jean-Rémy Chardonnet¹, Mikhail Langovoy², Yuyang Wang⁴ and Jivka Ovtcharova²

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

2 ABSTRACT

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

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

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

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

21

22

23

1 INTRODUCTION

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

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

52 There are two methods to evaluate cybersickness: subjective one, using questionnaires and objective
53 one, through physiological and behavioural measurements (Niu et al., 2020). In a subjective evaluation,
54 typically, participants will experience the VR task such as navigation or interaction. After exposure, they
55 complete a survey to assess system comfort. To achieve this aim, various questionnaires, e.g., motion
56 sickness questionnaire (MSQ) (Frank et al., 1983), Simulator Sickness Questionnaire (SSQ) (Kennedy et al.,
57 1993), Fast Motion Sickness Scale (FMS) (Keshavarz and Hecht, 2011), and VR sickness questionnaire
58 (VRSQ) (Kim et al., 2018) were designed to measure the sickness levels in different contexts and they
59 are considered as the cornerstone of the approach. Such methods are however limited as they report a
60 posteriori feedback, which prevents any possibility of acting efficiently to limit cybersickness. When it
61 comes to objective evaluation of this adverse VR side effect, signals like postural sway (Chardonnet et al.,
62 2017)(Lee et al., 2019) galvanic skin response (GSR) (Plouzeau et al., 2018), known as electrodermal
63 activity (EDA) in some literatures, electroencephalograph (EEG) (Kim et al., 2019)(Jeong et al., 2019)(Liao
64 et al., 2020)(Lin et al., 2013), or electrocardiogram (ECG) (Garcia-Agundez et al., 2019) are used to assess
65 physiological response and to complement subjective data from questionnaires. In this type of evaluation,
66 the participant is immersed in a VE to perform a task, while simultaneously physiological indicators are

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

$$f : I \rightarrow \mathbb{R} \quad (1)$$

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

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

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

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

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

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

2 BACKGROUND

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

123 Consider a point cloud $\chi = \{x_1, \dots, x_n\}$, sampled from a space M , it will be mapped into the structures
 124 called simplicial complexes. A k -simplex is a set of $k + 1$ indices from the given set χ . A simplicial
 125 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.
 126 One of the common complexes includes Vietoris-Rips complexes $R(\chi, \varepsilon)$ (Hausmann, 1995) which we use
 127 in this work. This complex is constructed by placing ε -balls (ε that defines the radius of an imaginary ball)
 128 on each vertex, and adding edges whenever they overlap:

$$\chi' \in R(\chi, \varepsilon), \quad \chi' \subset \chi \Leftrightarrow d(x_i, x_j) < \varepsilon, \quad \forall x_i, x_j \in \chi' \quad (2)$$

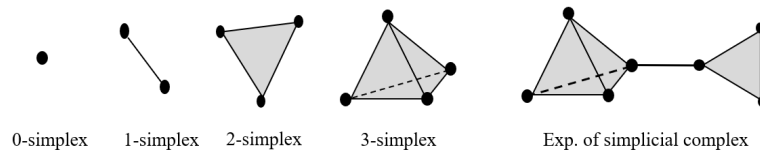


Figure 1. k -simplices in $\mathbb{R}^3, 0 \leq k \leq 3$ and an example of simplicial complex

129

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

$$\varepsilon_1 \leq \varepsilon_2 \implies X(\varepsilon_1) \subset X(\varepsilon_2) \quad (3)$$

133

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

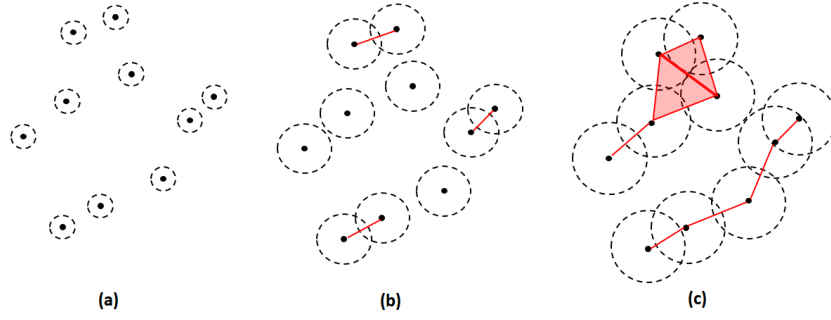


Figure 2. Example of filtration varying the filtration value ε which increased from (a) to (c). The black dot represents the point cloud data that are connected (red line) when the ε -balls around them overlap. The top part of (c) is the union of two adjacent triangles.

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

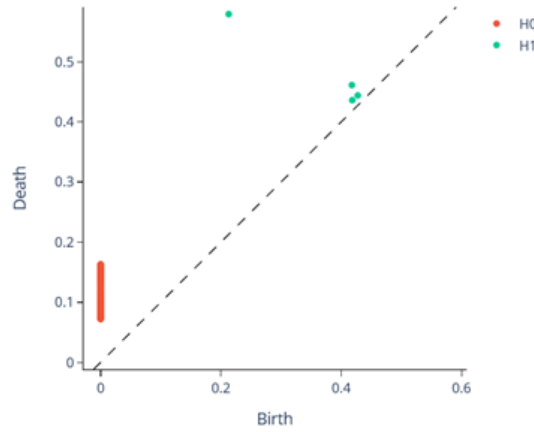


Figure 3. The corresponding persistence diagram with $H_0(X)$ in red and $H_1(X)$ in green, represents the persistence of connected components and holes over the ε -scale of filtration.

141

142 There are too few machine learning or statistical tools that can be applied directly to persistence diagram
 143 space. Hence, a mapping should be done from persistent diagram space to topological vector space which is
 144 appropriate for machine learning tools and further analysis. To achieve this aim by extracting scalar features,
 145 there are different methods like persistent image (Adams et al., 2017), persistent landscape (Bubenik,
 146 2015), and persistent entropy (Rucco et al., 2017) methods. Persistent entropy is defined as the (base 2)
 147 Shannon entropy of the persistence diagram derived from the filtration. (For simplicity of notation, the log
 148 will refer to the log-base-2 function.)

$$E(F) = - \sum_1^n p_i \log(p_i), \text{ where } p_i = \frac{l_i}{S_L}, l_i = y_i - x_i \text{ and } S_L = \sum_{i=1}^n l_i \quad (4)$$

3 RELATED WORKS

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

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

168 Porcino et al. (Porcino et al., 2022) proposed an experimental analysis to estimate the weight of cybersi-
169 ckness causes and not to predict the presence of this phenomenon. These user and context-specific causes
170 were ranked according to their impact using symbolic machine learning in VR games, including a racing
171 game and a flight game. They conducted 6 experimental protocols along with 37 valid samples from a total
172 of 88 volunteers. They used VRSQ to compare the user discomfort level with the verbal feedback collected
173 during the experiment and thereby evaluated the data and discard incompatible samples. They achieved
174 0.79 and 0.95 AUC scores using decision tree and random forest algorithms, respectively. They concluded
175 that exposure time, rotation, and acceleration are most likely the top factors contributing to cybersickness.
176 Since this approach, unlike ours, was not to predict the presence of cybersickness, and the input data of
177 their experiment was not based on the participant's physiological data, it is not possible to compare these
178 results.

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

4 DATA MEASUREMENT AND EXPERIMENT

188 To collect data and showcase our approach, we performed a user experiment. A total of 53 subjects,
189 composed of 26 females and 27 males, having an age distribution with the mean and the standard deviation
190 of 26.3 years and 3.3 years respectively, participated in a VR navigation experiment using an HTC Vive Pro
191 head-mounted display. Participants were asked to repeat the experiment three times on three different days
192 to gather enough samples. In that way, 159 samples were collected and included in the dataset. Upon arrival,
193 participants were asked to sign a consent form and fill out one questionnaire to investigate their health
194 conditions and experience in playing games and using VR devices before participating in the VR task. No
195 issue was reported from this questionnaire. They were then explained the navigation task to achieve, as well
196 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.

197

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

$$SSQ = SSQ_{\text{post}} - SSQ_{\text{pre}} \quad (5)$$

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

210 An Empatica E4 wristband¹ on one participant's arm was used for real-time physiological data acquisition
211 and particularly the electrodermal activity (EDA) of participants during this experiment. This wearable
212 device is equipped with some sensors to gather high-quality data that was sent during navigation to a
213 processing computer through Bluetooth. Galvanic Skin Response (GSR), blood volume pressure (BVP),

¹ <https://www.empatica.com/research/e4/>

214 heart rate (HR), and temperature (TEM) were recorded during their navigation experiment. Moreover, the
 215 longitudinal (LG) and rotational (RT) accelerations were computed from the recorded navigation speed.
 216 An example of the recorded signals is shown in Figure 5. These sensors have different frequencies for
 217 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.

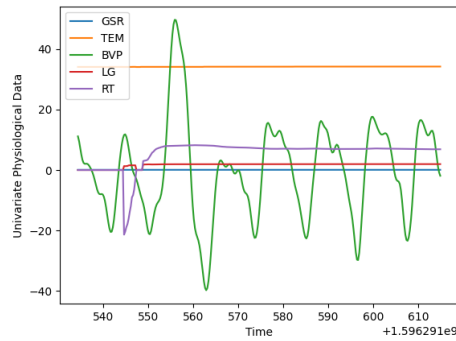


Figure 5. Sample of multivariate physiological data

218

5 METHODOLOGY AND IMPLEMENTATION

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

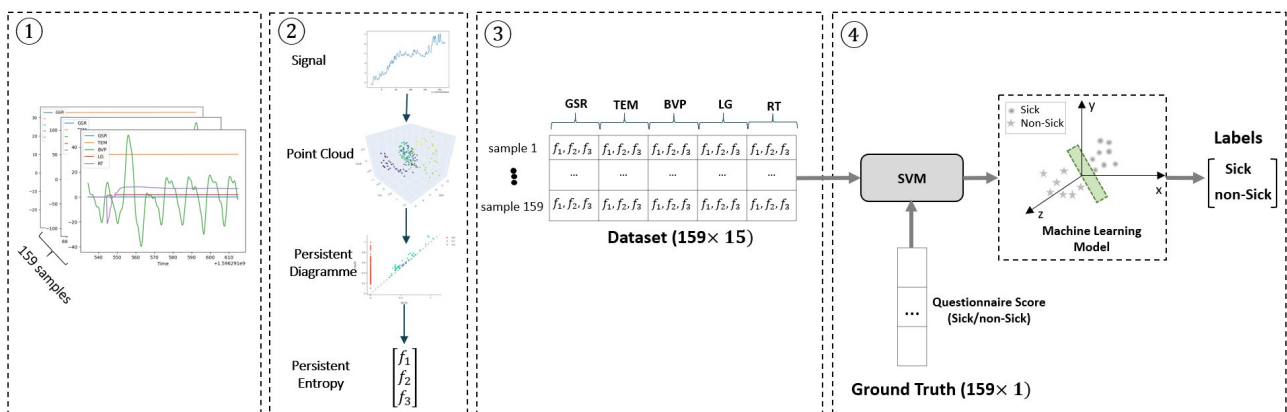


Figure 6. Block diagram of the cybersickness prediction method. 1) Normalization of the time series as a pre-processing step 2) Applying TDA and vectorizing persistent diagram to persistent entropy 3) Making the dataset (159×15 matrix) from the features descriptors that characterize the sample's data 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.

222

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

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

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

238 EMD determines what frequency with what strength in the signal occurs at any given moment. IMFs
 239 can be summed to recover the original signal. Because the first IMF usually carries the most oscillating
 240 (high-frequency) components, it can be rejected to remove high-frequency components (e.g., random noise).
 241 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.

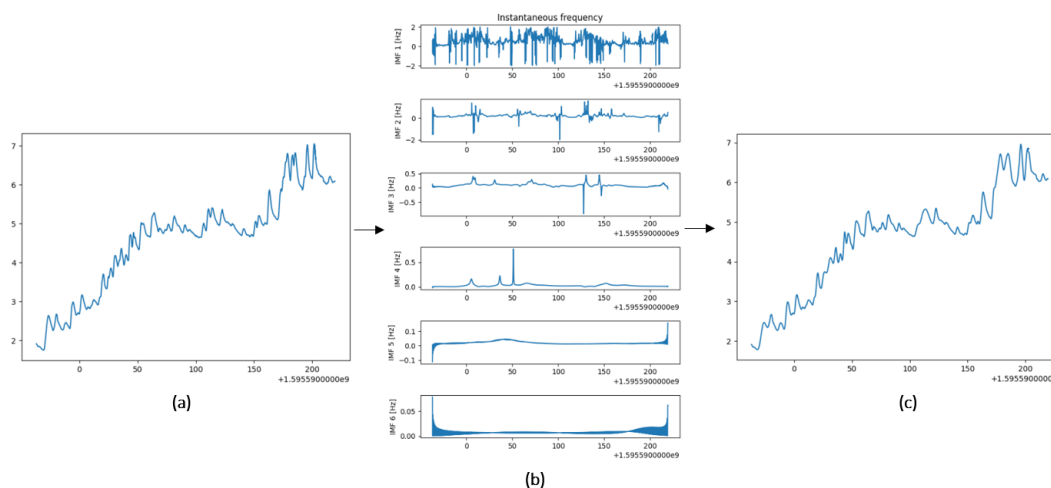


Figure 7. (a) Original GSR time series (b) Six IMFs decomposed by EMD (c) Output Series from EMD on the original signal. It is the sum of the last five IMFs plus residue.

242

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

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

$$\mathbf{f}_i = [f(t_i), f(t_i + 2\tau), \dots, f(t_i + (M - 1)\tau)] \in \mathbb{R}^d \quad (6)$$

249 where $(M-1)$ is the embedding dimension and τ is the time delay. Hence, the window size is the quantity $(M-$
 250 $1)\tau$ and a stride is defined as the difference between t_i and t_{i+1} . In other words, the time delay embedding
 251 of f with parameters (M, τ) is the function:

$$\text{TD}_{M, \tau} f : \mathbb{R} \rightarrow \mathbb{R}^d, t \rightarrow \begin{bmatrix} f(t) \\ f(t + \tau) \\ f(t + 2\tau) \\ \dots \\ f(t + (M - 1)\tau) \end{bmatrix} \quad (7)$$

252 As a result, we have two hyperparameters: M, τ . To determine the time delay automatically, we used the
 253 Mutual Information (MI) technique (Wen and Wan, 2009). MI is used as an analytical measure of the extent
 254 to which the values in the time series can be predicted by earlier values. At first, the Probability Density
 255 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
 256 bin (marginal probability density distribution) and $p_{i,j}$ as the probability that x_t is in the j th bin while $x_{t+\tau}$
 257 is in the i th bin (joint probability density distribution), MI is defined as Kulback-Liebler (KL) divergence
 258 between the $p_{i,j}$ and p_i and p_j i.e.

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

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

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

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

265 By adding one more dimension, the distance will change:

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

266 then, the FNN criterion is defined as:

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

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

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

272 After data structure analysis, the physiological data were labelled using the self-reported sickness
273 questionnaire, collected during the VR experiment from the participants. To achieve this aim, the SSQ
274 score was collected at both pre-and post-exposure, and the score for each participant was calculated using
275 original indexes (see section 4). We considered the SSQ score of 20 as the threshold to define the label of
276 "sick" and "non-sick" (Bimberg et al., 2020). The participants whose SSQ score is equal to or greater than
277 20 are assumed that they suffered from cybersickness and "sick". Conversely, an SSQ score of less than
278 20 is defined as not experiencing cybersickness and labelled "non-sick". Based on this labelling, the 159
279 samples were divided into 87 and 72 "sick" and "non-sick" samples, respectively. This step leads to a table
280 of data of size 159×15 as shown in Figure 6.

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

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

295 After applying EMD, we investigated the TDA performance on both raw and denoised signal reconstructed
296 summing the last five IMFs plus residue. Additional classification algorithms that were implemented and
297 compared to TDA, were the bag-of-patterns (Lin et al., 2012) with sliding window size 16, length of the
298 words 4, and 4 bins to produce without numerosity reduction; ROCKET (Dempster et al., 2020) with 10000
299 kernels in sizes 7,9, and 11; WEASEL (Schäfer and Leser, 2017) with word size 9 and window sizes from
300 10 to 27. In all mentioned studies, SVM (Gaussian RBF) was used as the classifier.

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

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

305

$$\text{precision} = \frac{TP}{TP + FP}, \text{ recall} = \frac{TP}{TP + FN} \quad (13)$$

306 To assess the generalization ability of the classifier and evaluate and test its performance, we used K-fold
 307 Cross-Validation (CV) technique (Berrar, 2019) with K=5. Also, we computed the evaluation metric, i.e.,
 308 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.

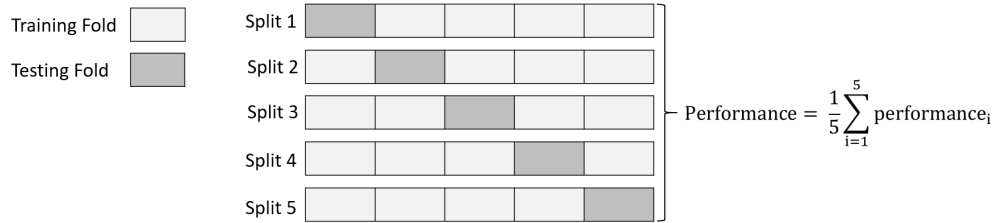


Figure 8. 5-Fold cross-validation in which data is divided into 5-folds composed of 4-folds for training the model and 1-fold for validation. The overall performance is obtained by computing the arithmetic mean.

309

6 RESULTS AND DISCUSSION

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

316 6.1 Comparison of Classification Algorithms

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

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

329 6.2 Comparison of Compound Classifiers

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

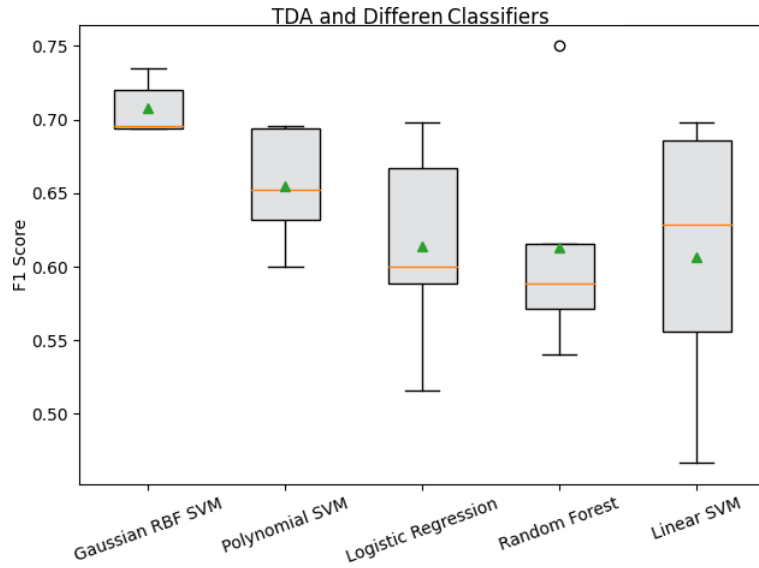


Figure 9. Performance evaluation and the effect of the classifier on sickness detection using the TDA feature descriptor

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

335 and less standard deviation (std) than TDA. However, the difference in the F1 score is only 0.002, with a
 336 0.001 difference in standard deviation, on the dataset of 159-time series, which indicates that the difference
 337 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

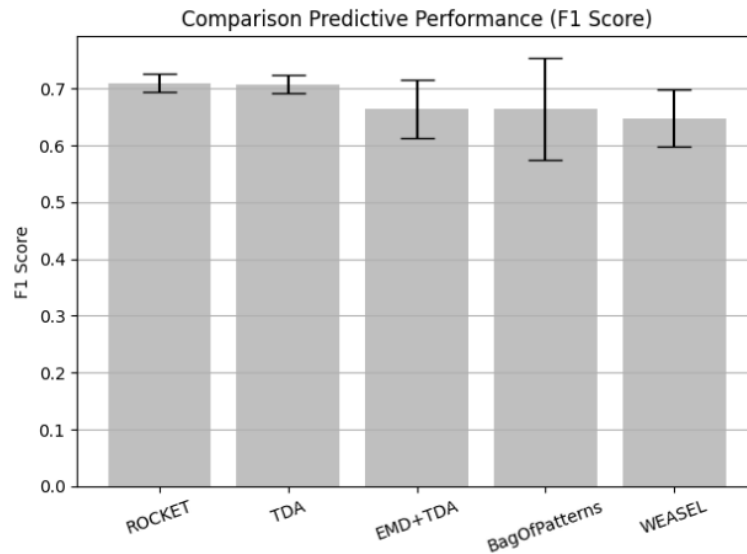


Figure 10. Performance evaluation using different feature extraction algorithms

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

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

353 6.3 Heart Rate and Cybersickness

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

360 6.4 Conclusion

361 In this paper, we proposed a machine learning approach to cybersickness prediction in VR, on the basis of
 362 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

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

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

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

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

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

ACKNOWLEDGMENTS

388 This work was supported in part by a grant from the French-German University (UFA-DFH) No. CDFA
389 03-19.

REFERENCES

390 Adams, H., Emerson, T., Kirby, M., Neville, R., Peterson, C., Shipman, P., et al. (2017). Persistence
391 images: A stable vector representation of persistent homology. *Journal of Machine Learning Research*
392 18

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

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

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

400 Bubenik, P. (2015). Statistical topological data analysis using persistence landscapes. *J. Mach. Learn. Res.*
401 16, 77–102

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

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

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

408 Dempster, A., Petitjean, F., and Webb, G. I. (2020). Rocket: exceptionally fast and accurate time
409 series classification using random convolutional kernels. *Data Mining and Knowledge Discovery* 34,
410 1454–1495

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

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

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

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

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

421 Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., et al. (1998). The empirical
422 mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis.
423 *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences*
424 454, 903–995

425 Jeong, D., Yoo, S., and Yun, J. (2019). Cybersickness analysis with eeg using deep learning algorithms. In
426 *2019 IEEE conference on virtual reality and 3D user interfaces (VR)* (IEEE), 827–835

427 Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G. (1993). Simulator sickness questionnaire:
428 An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*
429 3, 203–220

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

468 Porcino, T., Rodrigues, E. O., Bernardini, F., Trevisan, D., and Clua, E. (2022). Identifying cybersickness
469 causes in virtual reality games using symbolic machine learning algorithms. *Entertainment Computing*
470 41, 100473

471 Porcino, T., Rodrigues, E. O., Silva, A., Clua, E., and Trevisan, D. (2020). Using the gameplay and user
472 data to predict and identify causes of cybersickness manifestation in virtual reality games. In *2020 IEEE*
473 *8th international conference on serious games and applications for health (SeGAH)* (IEEE), 1–8

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

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

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

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

- 482 Sevinc, V. and Berkman, M. I. (2020). Psychometric evaluation of simulator sickness questionnaire and
483 its variants as a measure of cybersickness in consumer virtual environments. *Applied ergonomics* 82,
484 102958
- 485 Strang, G. (2006). *Linear algebra and its applications*. (Belmont, CA: Thomson, Brooks/Cole)
- 486 Takens, F. (1981). Detecting strange attractors in turbulence. In *Dynamical systems and turbulence*,
487 *Warwick 1980* (Springer). 366–381
- 488 Wen, F. and Wan, Q. (2009). Time delay estimation based on mutual information estimation. In *2009 2nd*
489 *International Congress on Image and Signal Processing* (IEEE), 1–5
- 490 West, D. B. (2001). *Introduction to graph theory*, vol. 2 (Prentice hall, Upper Saddle River New Jersey)
- 491 Zomorodian, A. and Carlsson, G. (2005). Computing persistent homology. *Discrete & Computational*
492 *Geometry* 33, 249–274