Comparing Autonomic Physiological and Electroencephalography Features for VR Sickness Detection Using Predictive Models

Gang Li
School of Psychology and Neuroscience
University of Glasgow
Glasgow, United Kingdom
Gang.Li@glasgow.ac.uk

Ogechi Onuoha
School of Computing Science
University of Glasgow
Glasgow, United Kingdom
Ogechi.Onuoha@glasgow.ac.uk

Mark McGill
School of Computing Science
University of Glasgow
Glasgow, United Kingdom
Mark.McGill@glasgow.ac.uk

Stephen Brewster
School of Computing Science
University of Glasgow
Glasgow, United Kingdom
Stephen.Brewster@glasgow.ac.uk

Chao Ping Chen
Department of Electronic Engineering
Shanghai Jiao Tong University
Shanghai, China
ccp@sjtu.edu.cn

Frank Pollick
School of Psychology and Neuroscience
University of Glasgow
Glasgow, United Kingdom
Frank.Pollick@glasgow.ac.uk

***Abstract*—How the performance of autonomic physiological, and human vestibular network (HVN)-based brain functional connectivity (BFC) features differ in a virtual reality (VR) sickness classification task is underexplored. Therefore, this paper presents an artificial intelligence (AI)-aided comparative study of the two. Results from different AI models all show that autonomic physiological features represented by the combined heart rate, fingertip temperature and forehead temperature are superior to HVN-based BFC features represented by the phase-locking values of inter-electrode coherence (IEC) of electroencephalogram (EEG) in the same VR sickness condition (that is, as a result of experiencing tunnel travel-induced illusory self-motion (vection) about moving in-depth in this study). Regarding EEG features per se (IEC-BFC vs traditional power spectrum), we did not find much difference across AI models.**

Keywords—Virtual Reality, Cybersickness, EEG, Autonomic Physiological Signals, Machine Learning

I. INTRODUCTION

In 2013, Neuroscape of University of California, San Francisco (UCSF), published a cover story in *Nature*. This cover story, called *Game Changer* by *Nature* editors, described how a driving video game (NeuroRacer) could enhance cognitive control in older adults [1]. In recent years, and with the rise of the consumer-friendly Virtual Reality (VR), research bodies, including UCSF Neuroscape [2], [3] , are increasingly using consumer-grade VR headsets to deliver gamified cognitive and rehabilitation therapies instead of using standard personal computer (PC) monitors. However, a frequently encountered problem, requiring some participants to withdraw from the ongoing therapies is VR-induced illusory self-motion (that is, vection, such as the illusory moving in-depth in VR driving tasks), which results in a general discomfort or even nausea for some simulator sickness-susceptible participants (about 60% of the general population). Therefore, objective and automatic VR sickness assessment is a very important step before developing any targeted intervention measures if porting a well-established cognitive paradigm (e.g., NeuroRacer) from a traditional 2D platform to an immersive VR platform is our final goal.

Since autonomic responses (such as sweating, increased heart rate and nausea) are the most obvious symptoms, there are many autonomic physiological signal-based solutions for objective and automatic VR sickness assessment [4]–[7]. However, the current state-of-the-art approach in motion sickness and simulator sickness assessment (including but not limited to pure VR sickness) is artificial intelligence (AI)-integrated multimodal biosensing including behavioural and autonomic physiological data and neural data [8], [9]. For example, Prof Gargiulo et al. from the motion sickness lab at the Reykjavik University are using EMG (Electromyography, that is, behavioural physiological signals), EEG (electroencephalogram, that is, neural signals) and heart rate (autonomic physiological signals) together to evaluate VR-based simulator sickness. Here, the difference between VR-based simulator sickness and pure VR sickness is that the former involves real physical motions that to some extent match with the motion cues in VR (e.g., in [8], the VR shows a boat surrounded by waves whose frequency and amplitude are synchronized with the simulator platform movement.); while pure VR sickness refers to visually induced motion sickness without the involvement of real physical motions. They concluded that EMG features are the most correlated with self-reported simulator sickness levels, which are followed by EEG features; while the heart rate features showed the poorest performance. In another pure VR sickness study [9], Prof. Marcus et al. from the Technical University of Braunschweig used EEG and a set of autonomic physiological signals (including finger-mounted heart rate, respiration as well as electrodermal activity (EDA) signals). They found that EDA may be a solid indicator of simulator sickness if compared to EEG features. The inconsistent findings in [8] (which shows that EEG data is superior to autonomic physiological data) and [9] (which showed that autonomic physiological data is superior to EEG data) leads to a research question of how autonomic physiological and neural data differs in the context of pure VR sickness assessment. This is especially pertinent given that both [8] and [9] did not explore brain network-based EEG features, which we believe are more correlated with VR sickness levels compared to conventional EEG power spectrum if according to the recently proposed human vestibular network (HVN) [10].

HVN is a widespread brain network including at least the autonomic, sensorimotor and cognitive domains [10]. Given the close relationship between HVN and one of the widely-accepted motion sickness aetiology — visual-vestibular sensory conflict theory (SCT [11]), we hypothesized that HVN-based EEG brain functional connectivity (BFC)

This research is sponsored by European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (No. 835197) and National Natural Science Foundation of China (No. 61901264) (Corresponding Author: Gang Li).

2021 IEEE Symposium Series on Computational Intelligence (SSCI) | 978-1-7281-9048-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/SSCI50451.2021.9660126

TABLE I
SUMMARY OF THE TECHNICAL PARAMETERS OF THE SENSOR SYSTEM

Experimental facilities		Parameters	Value
Fully Commercial Device	VR	Oculus Quest	Ver 2.0
	EEG/EOG	Electrodes	Conventional wet electrodes
		Sampling rate	500 Hz
		Analog-to-digital (ADC)	24 bit
		Communication	WiFi/USB
Arduino-compatible prototype	PPG	Sensing type	reflective
	Contact temperature	Accuracy	±0.5°C from -10°C to +85°C
	Non-contact temperature		±0.5°C from 0°C to +50°C
	Arduino	ADC	10 bit
		Communication	USB at 115200bps
		Sampling rate	50 Hz

features would be superior to traditional EEG power spectrum features. Also, it is worthy to explore whether the HVN-based BFC features could achieve better performance than autonomic physiological features. Therefore, the goal of the present study was to explore the VR sickness classification performance of autonomic physiological and HVN-based neural features with the aid of AI models. The present study does not involve the most advanced AI models. Instead, we employed some basic structures from the commonly used AI models to compare the autonomic physiological and neural data features in the context of the same framework (that is, the moving in-depth-induced VR sickness).

II. SYSTEM ARCHITECTUE AND SETTINGS

Our signal acquisition system was equipped with continuous monitoring of multi-modal signals (e.g. EEG, EOG (electrooculography), PPG (blood volume pulse), contact-type fingertip temperature, and non-contact infrared forehead temperature) of the participant. As shown in Table I, the EEG and EOG data were recorded by a medical-grade device. Other data were collected by Arduino-compatible commercial sensors. More technical information about these experimental facilities can be found in the following subsections *A* to *C*. Regarding the VR device, we used the newly released all-in-one Oculus Quest 2 (72Hz display refresh rate and 89° horizontal field of view).

Fig. 1. Sample experimental environment.

A. Apparatus for Autonomic Physiological Signals

PPG data were recorded from the fingertip of the left index finger using a set of reflective-type photoelectric sensors (including light transmitter and receiver; https://pulsesensor.com/). The more blood volume in the fingertip, the more light can be absorbed, and thus the less the receiver can sense. According to the circuit design, the voltage output of the receiver and the amount of received light is negatively related. Thus, the voltage reaches peak level when heart pumps blood to fingertip (that is the peak of the PPG signal). Therefore, HR can be extracted and analyzed from continuous cardiac cycle-driven PPG signals. Fig. 2 shows the examples of detected crests and troughs from the PPG signals using our custom MATLAB script.

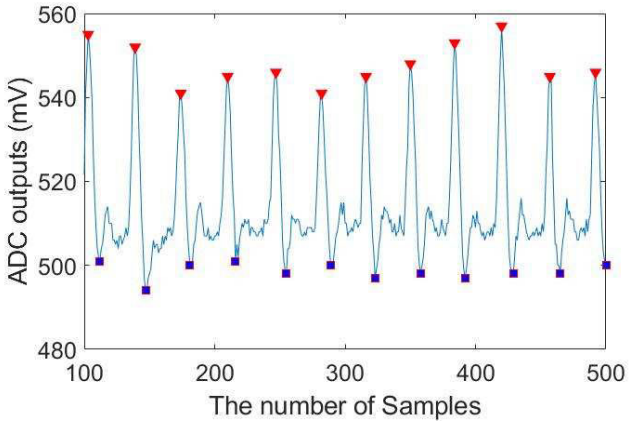

Fig. 2. The examples of detected crests and throughs from the PPG signals using our custom MATLAB script. By this script, HR was calculated.

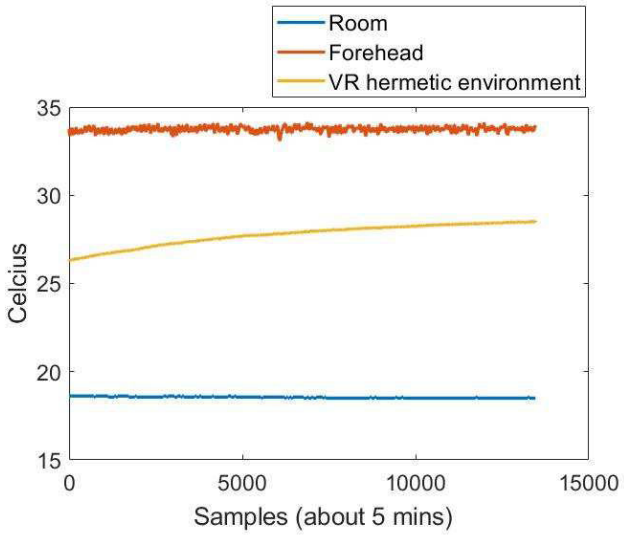

Fig. 3 The forehead temperature data in the initial baseline session (5 mins) in our pilot test where the power of the VR headset had already been turned on for 2 hours at the testing moment. As we can see that the forehead temperature was stable and not affected by the hermetic VR environment when room temperature was set to 18°C.

Forehead temperature data were collected from the participant's forehead skin (see Fig. 3) using a non-contact infrared temperature sensor (MLX90614, Melexis, Inc.). The position of the forehead temperature sensor was just above the Quest 2's built-in proximity sensor). Fingertip

temperature data were collected from the fingertip of the left middle finger using digital thermometer (DS18B20, Maxim Integrated, Inc.). Both fingertip and forehead temperature sensors have a ±0.5℃ measurement accuracy which is acceptable in the context of up to 2-4℃ CS-induced difference [12]. It is important to note that the non-contact infrared temperature could simultaneously detect the target (the forehead) and ambient (VR hermetic environment) temperature. Through our pilot test we did not find any confounding effect of the VR hardware heat (26-27°C) on forehead skin temperature (higher than 30°C, see Fig. 2), where 26-27°C was the steady VR hardware heat we observed after two hours of continuous usage.

B. Apparatus for EEG Signals

EEG data were collected through an eight-channel EEG recording device (NE® StarStim 8™, Neuroelectrics Inc, Barcelona, Spain), which uses a high-resolution, high-speed analog-to-digital converter (24 bit at 500Hz sampling rate). Conventional wet electrodes were used and placed at seven EEG channels (see Fig. 4). The remaining one channel (EXT) was used to collect EOG using a disposable electrode from the left lower eyelid. The ground and reference electrodes were connected and placed on the right earlobe by an ear clip. Note, we did not use medium or high-density EEG, since the reduced seven EEG channels were carefully selected by our previous study for the cognitive domain (Fz, Cz, Pz [3]) and other researchers' studies on the sensorimotor domain (To be more specific, that is, P3 and P4-based parieto-insular vestibular cortex (PIVC) and CP5 and CP6-based temporoparietal junction (TPJ) [13]–[16]).

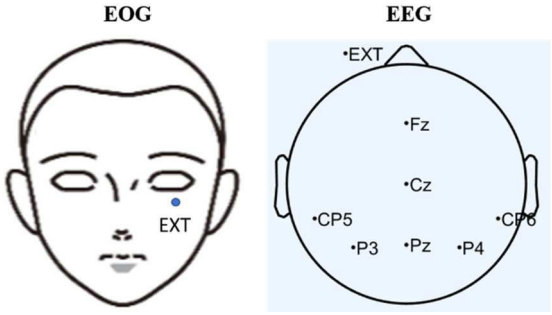

Fig. 4. The channel location for EEG and EOG.

III. METHODS

A. The Estimation of the Ground Truth

The fast motion sickness (FMS) scale was used to estimate the ground truth of VR sickness in the present study. FMS is a single item questionnaire that requires participants to focus on nausea, general discomfort, and stomach problems only and finally give an overall single score ranged from 0 (not at all) to 20 (severe) at 1-minute intervals [17]. Here, a digital version of the FMS was directly implemented inside the VR scene (see Fig. 5) so that participants could report their FMS scores while wearing the VR headset by moving the virtual slider immediately after VR tasks finished. The purpose of the use of digital questionnaires was to minimize the opportunities for subjects to remove the VR headset which could temporarily alleviate their symptoms and create a confound to

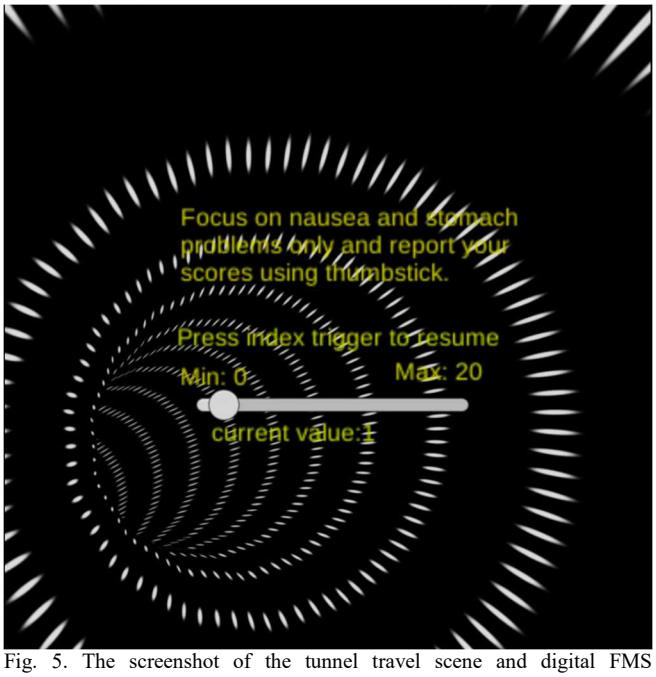

Fig. 5. The screenshot of the tunnel travel scene and digital FMS questionnaire. Note, the FMS questionnaire pops-up every minute. The tunnel travel will stop moving at the moment of the appearance of the FMS questionnaire and resume after participant's rating is finished. The time a participant spend on FMS rating was not counted in the 10-minute tunnel travel task.

the relationship between sickness ratings and objective measures.

B. Experimental Procedure and VR Sickness Induction

The experiment contained a 5-minute baseline measurement and a 10-minute tunnel travel task. During the baseline measurements, participants were required to close their eyes and listen to soft music (slow instrumental featuring traditional Chinese instruments). During the tunnel travel task, participants were required to keep their body and head still, and watch the VR scene and perceive the moving in-depth-induced illusory self-motion. The route of the tunnel was set as a normal driving scenario, including curves, uphill and downhill paths, but without upside down and off-axis paths. The moving speed was uniform without any acceleration but adjustable. To be more specific, it would be increased by 20% in the second time window if the same FMS score was reported in consecutive two-minute periods, in order to match with participant's different susceptibility threshold. The maximum value of FMS score is 20. Unlike [18], which brought participants to the point of retching, in order to minimize the confounding effect of such symptomatic behavior and to comply with ethics requirement (that is, no more than moderate nausea symptoms), we would stop the experiment once the FMS score exceeded 11.

The procedure was approved by the ethics panel of the University of Glasgow (No. 300200009), College of Science and Engineering.

C. Signal Pre-Processing and Feature Extractions

For EEG and EOG, a low-pass filter with a cutoff frequency of 40 Hz and high-pass filter with a cutoff frequency of 0.1 Hz were applied to remove power line noise and DC drift, respectively. The filtered EEG data were then corrected using the mean of each channel and EOG-based independent component analysis. Regarding PPG, a band-pass filter

(0.2Hz-5.6Hz) was used and then a 1st-order differential operation was used to remove the baseline wander (more details can be found in our previous study [19]). For temperature, we used the raw data without any pre-processing. Next, the features shown in Table II were extracted respectively for AI model-based classification. Note, Cz-related features (not shown) are reserved for another ongoing study.

TABLE II
SUMMARY OF THE EXTRACTED FEATURES IN THIS STUDY

	Features	Feature Sources
EEG	Relative band power (RBP)* (δ~β)	CP5, P3, P4, CP6
	Inter-electrode coherence (IEC)-based phase-locking values (δ~β)	1. Frontal cognitive to Left sensorimotor (Fz to averaged CP5 & P3); 2. Frontal cognitive to Right sensorimotor (Fz to averaged CP6 & P4); 3. Frontal cognitive to Central posterior cognitive (Fz to Pz) 4. Right sensorimotor to Left sensorimotor (Averaged CP5 & P3 to averaged CP6 & P4)
Autonomic	Heart rate	Left index fingertip
	Fingertip temperature	Left middle fingertip
	Forehead temperature	Non-hairy forehead

Where, *RBP was calculated by dividing the FFT power of one EEG band by the sum of the FFT power of all four EEG bands (δ, θ, α and β).

The IEC-based PLV is a commonly used measure of phase consistency across electrodes in EEG, which can be used as a measurement of brain functional connectivity [1], [20]. To be more specific, we created 1 frontal (Fz) and 2 parietal composite electrodes of interest (EOI) from the average of the following electrodes: left TPJ and PIVC-based CP5 and P3 as well as right TPJ and PIVC-based CP6 and P4, with PLVs calculated for each frontoparietal EOI combination separately. The specific equations for PLV calculation can be found in Eq.1-3.

$$\varphi_1(f, n)=\text{Hilbert transform}\left(\text{EEG}(f, n)_{Fz}\right) \quad (1)$$

$$\varphi_2(f, n)=\text{Hilbert transform}\left(\text{EEG}(f, n)_{\text{Parietal}}\right) \quad (2)$$

$$\text{PLV}(f,n)=\frac{1}{N}\left|\sum_{i=1}^{N} e^{\varphi_1^i(f,n)-\varphi_2^i(f,n)}\right| \quad (3)$$

where φ_1 is the phase time course of EEG samples (data points) of that epoch (the length = n) at the Fz channel after Hilbert transform; φ_2 is the phase time course of EEG samples (data points) of that epoch (the length = n) at the averaged left or right parietal channel after Hilbert transform; N is the total number of the EEG epoch; φ^i is the phase time course of the i^{th} EEG epoch; $f \in \{\delta\ (0.1-3\text{Hz}), \theta(4-7\text{Hz}), \alpha(8-12\text{Hz}), \beta(13-20\text{Hz})\}$. In this study, the length of the EEG epoch is 2 sec.

D. Baseline Corretion and Normalization

We took the values in the first minute of the tunnel travel task as the baseline, and then all features and FMS scores were normalized (see Eq. 4 and 5). The baseline-corrected and normalized data provided a means to normalize results so that the assessment of sickness severity was not confounded by individual differences.

$$\text{Normalized Features}=\frac{\text{Raw -Baseline}}{\text{Baseline}}\times 100\% \quad (4)$$

$$\text{Normalized Score}=\frac{\text{Raw -Baseline}}{\text{Full Score}}\times 100\% \quad (5)$$

where, the full score used in Eq. 5 was to avoid division errors just in case the baseline score was zero.

E. Participants

A total of twenty-four healthy right-handed young adults attended this study. Two participants were excluded onsite for the following reasons: 1) The PPG signals from one of them were abnormal due to very cold hands; 2) One participant's data were unable to be normalized since their FMS score in the first minute had already exceeded the pre-defined threshold. Therefore, finally twenty-two participants (range 20-31 years; 4 males) took part in the entire study.

F. AI Model-based Prediction

We set up two tasks to predict the FMS score/label through regression and classification based on the input feature sets. The goal is to identify which prediction models work best on each individual feature set. For the regression task, we made predictions against the FMS score; whereas we create FMS class labels for the classification task and classification models to predict the FMS class. In order to select the parameter values for each estimator in both regression and classification, we used the grid search cross validation library provided by sklearn [21] for model selection. The parameters explored for the support vector machine (SVM) using grid search include the kernel (linear, poly and rbf), the polynomial degree (2,3,4), and c (0.1, 1, 10, 100, 1000). In the artificial neural network (ANN) grid search setup, we explored activation functions in the hidden layer ('logistic', 'tanh', 'relu'), hidden layer sizes ((1), (10), (50), (10,10), (50,50), (10,10,10,10,10), (50,50,50,50,50)) and solvers ('lbfgs', 'sgd', 'adam'). The grid search for each estimator was run for a maximum of 500 iterations using cross validation with 10 splits. The results shown in Table III-VI below are all the best results after the grid search-based model selection.

When applying the data to each model selected, we used cross validation with ten splits (k=10). For the regression task, we applied linear regression and ANN to each feature set in the dataset. The accuracy metric applied is the Adjusted R^2 score. The adjusted R^2 instead of R^2, root mean square error as well as mean absolute error was used here as the metric to judge the performance of regression in the context of minimizing the overfitting problem. For the classification task, we grouped the normalized FMS scores (%) into target classes using the following scheme:

- Class I: normalized FMS scores that are <=0 (LOW)
- Class II: normalized FMS scores that are >0 and <=15% (MEDIUM)
- Class III: normalized FMS scores that are >15% (HIGH)

Since we collected a dataset of 203 feature points (extracted from 203 1-minute autonomic physiological and EEG signals), labelling the data using this scheme yields the following class counts:

- LOW: 57
- MEDIUM: 65
- HIGH: 81

Finally, based on the selected models, we investigated whether reduced autonomic physiological and EEG IEC features could achieve better results. To do so, for autonomic physiological feature set, heart rate, fingertip temperature and forehead temperature was input into the AI models separately instead of the combined one. Given that the number of IEC features is larger (N=4 feature sources* 4 EEG bands=16), we did not try it one by one. Instead, we measured the importance of the different features in the IEC feature set using the coefficients of a linear regression model for the regression task and a logistic regression model for classification task. We explored different thresholds when selecting the features and identified that using a threshold of 0.15 and 0.01 on the absolute values of the coefficient produces the best results for regression and classification task, respectively.

Note, we were using Google Colab (colab.research.google.com) as the computing platform for AI modelling, training and validation throughout the whole study.

IV. Results

A. Linear Regression

As can be seen in Table III, the linear regression results show that the autonomic physiological feature set achieved the best performance with adjusted R^2 score = 29%. All EEG features got negative values of adjusted R^2 score, indicating the failure of using EEG features to establish the regression model for VR sickness prediction. The superiority of autonomic physiological signals presented here is consistent with previous study [9].

TABLE III
REGRESSION RESULTS

Feature set	R^2 score	Adjusted R^2 score
Autonomic	0.30	0.29
EEG Power spectrum	0.11	-0.02
EEG IEC	-0.06	-0.14

TABLE IV
NON-LINEAR REGRESSION RESULTS

Feature set	R^2 score	Adjusted R^2 score
Autonomic	0.05	0.05
EEG Power spectrum	0.10	-0.04
EEG IEC	-0.12	-0.21

B. Non-linear Regression (NLNs)

Similarly, the ANNs-based non-linear regression results show that the autonomic physiological feature set achieved the best performance. However, as shown in Table IV, the value of adjusted R^2 was reduced from 29% to only 5%, indicating that the autonomic physiological feature set primarily presented a linear rather than non-linear relationship with VR sickness severity. Regarding EEG features, we still got the negative values of adjusted R^2 for both IEC and power spectrum features. Taking Table III's results together, this suggests that EEG features per se did not contain adequate information for regression modeling in the present study, which has nothing to do with the regression model's structure. However, apparently, the performance of EEG feature set in linear regression model marginally exceeded non-linear regression model, indicating that EEG features also present a linear relationship with VR sickness severities. Here, the architecture selected during grid search was a ANN with 1 hidden layer of 50 units with ReLU activation function and the Adam optimizer.

TABLE V
SVM-BASED CLASSIFICATION RESULTS

Feature set	Accuracy
Autonomic	0.48
EEG Power spectrum	0.46
EEG IEC	0.46

TABLE VI
ANN-BASED CLASSIFICATION RESULTS

Feature set	Accuracy
Autonomic	0.55
EEG Power spectrum	0.49
EEG IEC	0.48

C. Classification

Unlike the obvious difference in feature performance in regression analysis, autonomic physiological and EEG features did not show too much difference in the classification task. However, as shown in Table V and VI, the autonomic physiological feature set still maintains its superiority if compared to EEG features. To be more specific, the autonomic physiological feature set achieved 48% and 55% classification accuracy by using SVM and ANN, respectively; while EEG power spectrum feature set achieved the exact same classification accuracy (46%) with that achieved by EEG IEC in SVM condition. Regarding ANNs, EEG power spectrum and IEC got 49% and 48% accuracy, respectively. These results perhaps indicate that EEG features are comparable to autonomic physiological features in a discrete label prediction-based classification task. For SVM, the best kernel identified using grid search was rbf with the default values of *C*=1 and *gamma* (*g*) set to 'scale' resulting in *g*=1/(n_features * X.var()) as outlined in the sklearn API (https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html), where "n_features" and "X.var()" stand for the number of features and variance of features, respectively. The architecture of the best ANN was 1 hidden layer with 50 units with ReLU and Softmax activation functions at the output layer.

D. Reduced IEC features

1) Regression

The EEG IEC features selected using the aforementioned threshold for regression analysis were as follows:

- *'Frontal_Central_beta';*

- *'Frontal_Central_alpha';*
- *'Frontal_Left_beta';*
- *'Frontal_Right_delta'.*

The importance of each feature is plotted in the figure below (see Fig. 6). We applied the same estimators used for the regression task above to the reduced feature set. Table VII shows that the reduced IEC features achieved better result (adjusted R^2 score=5%) if compared to full IEC feature set (negative value of adjusted R^2 score). This result perhaps indicates that full IEC feature set contains some noisy (irrelevant) information for VR sickness assessment, for example the features at theta band and all features about Right sensorimotor to Left sensorimotor. We are not clear why IEC features at theta band were not selected, but a reasonable explanation for the unselected Right sensorimotor to Left sensorimotor features is very likely that the range of right-left IEC is too short so that the volume conduction issue in EEG become a confounding factor.

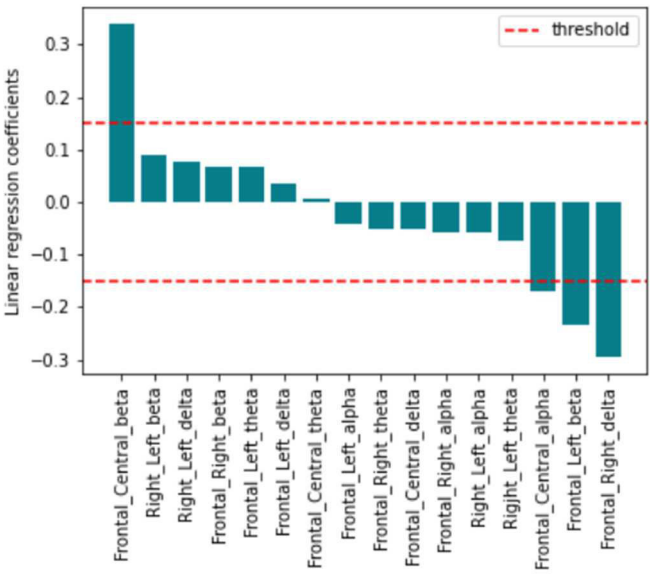

Fig. 6. IEC feature importance obtained from Linear regression coefficients.

TABLE VII
REGRESSION RESULTS USING REDUCED AUTONOMIC PHYSIOLOGICAL AND EEG IEC FEATURES

Feature set	Estimator	R^2 score	Adjusted R^2 score
Fingertip temperature	Linear Regression	0.05	0.06
	ANN	-0.23	-0.25
Forehead temperature	Linear Regression	0.01	0.01
	ANN	-0.56	-0.54
Heart rate	Linear Regression	0.20	0.20
	ANN	-0.09	-0.09
Reduced IEC	Linear Regression	0.06	0.05
	ANN	-0.07	-0.08

However, as can be seen in Table VII, the improved IEC features (whose adjusted R^2 score=5%) are still much inferior to the reduced autonomic physiological feature, like heart rate whose adjusted R^2 score is 20%. Of course, for autonomic feature per se, the reduced feature set leads to reduced regression performance (from 29% of full set to present 20% of heart rate, 6% of fingertip temperature and 1% of forehead temperature), indicating that each feature in the autonomic feature set has their uniqueness, so the combined autonomic feature set should be maintained.

2) Classification

The image below shows the coefficient values when the IEC feature set was fit on a logistic regression model. The features included in the reduced feature set are as follows:

- *'Frontal_Central_alpha'; 'Frontal_Left_beta';*
- *'Frontal_Left_alpha'; 'Frontal_Right_delta';*
- *'Right_Left_theta'; 'Right_Left_alpha';*
- *'Right_Left_delta'; 'Frontal_Central_delta';*
- *'Frontal_Central_beta'; 'Frontal_Left_theta';*
- *'Right_Left_beta'.*

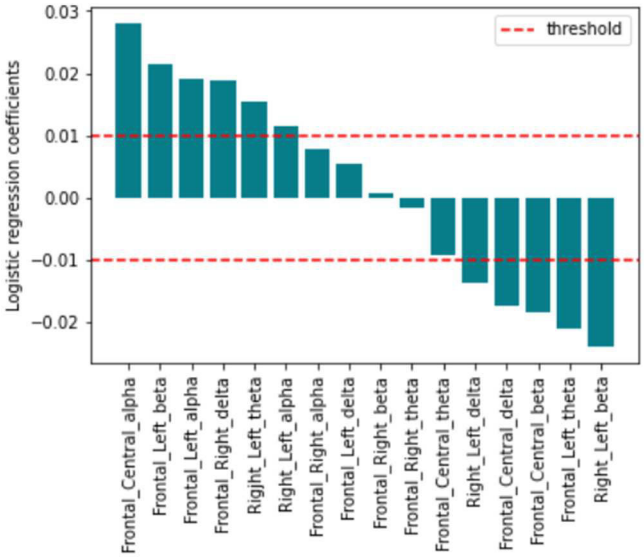

Fig. 7. IEC feature importance obtained from Logistic regression coefficients.

TABLE VIII
CLASSIFICATION RESULTS USING REDUCED AUTONOMIC PHYSIOLOGICAL AND EEG IEC FEATURES

Feature set	Estimator	Accuracy
Fingertip temperature	SVM	0.43
	ANN	0.45
Forehead temperature	SVM	0.35
	ANN	0.35
Heart rate	SVM	0.48
	ANN	0.48
Reduced IEC	SVM	0.48
	ANN	0.52

Table VIII shows the results when applying the same estimators used in the classification task with the full feature set. Interestingly, we found that IEC features achieved better performance (52% accuracy) than autonomic physiological features (heart rate with 48% accuracy) for the first time throughout the whole study. The learning curve for the best IEC result is shown in Fig.8. However, the improved accuracy of IEC is still lower than that when full autonomic physiological feature set was using (accuracy=55%, see Table VI).

V. Discussion

A. Primary Results

Using FMS-based estimation of VR sickness ground truth, the current study used basic AI models to compare the regression and classification performance of autonomic physiological and neural (EEG) feature sets, with an attempt to optimize the sensor inputs in the context of multimodal bio-sensing-centric objective and automatic VR sickness detection. Overall experimental results show that the autonomic physiological feature set is superior to neural ones in both the regression and classification tasks. This result leads to questions of whether 1) autonomic physiological sensors alone are effective enough to detect VR sickness; 2) the superiority of autonomic physiological signals can be maintained if other approaches are used to estimate the ground truth; 3) the superiority of autonomic physiological signals can be maintained if the VR sickness level is not at a moderate level as induced by the tunnel travel (e.g., a more severe level induced by rollercoaster scenes or on the contrary just a mild level induced by relatively neutral VR scenes) .

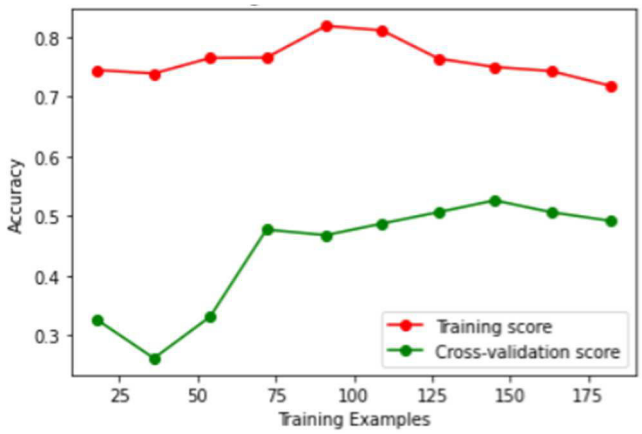

Fig. 8. Cross validation learning curves (NN with reduced EEG IEC features).

B. Regression

Apparently, EEG features perhaps did not capture sufficient granularity if compared to autonomic physiological features to establish a regression relationship with FMS scores in this study. A possible reasoning behind this result is that FMS is only a measurement (estimated ground truth) of autonomic responses during VR sickness induction. Thus, the conclusion about a predominant position of the autonomic domain during VR sickness induction could provide different results if other neural responses-based questionnaires are used. However, to the best of our knowledge, except for FMS, currently no questionnaires are regularly used to evaluate VR sickness with a shorter time window. For example, although the simulator sickness questionnaire is a questionnaire with 16 items including assessments of participant's cognitive aspects, such as fatigue, headache and difficulty concentrating, it is just a pre-post experiment questionnaire, not designed for an automatic assessment-oriented study. Also, although prior motion sickness research has utilized physical continuous inputs (e.g., sliders) to ask participants to indicate the sickness ratings when they feel better or worse in a manner that is very likely faster than using FMS at 1-minute intervals [22], the drawback is that the participants tend to forget to indicate their scores and thus leading to data missing. Therefore, it is imperative to design a proper real-time or near real-time questionnaire (the ground truth) for neurotechnology-centric and AI-aided VR sickness detection.

C. Classification

Since the reduced requirements of the classification task for the granular capture ability (if compared to the regression task), EEG feature set was finally able to achieve a comparable level now with those autonomic physiological feature set. However, an interesting question is that why non-linear rbf kernel-based SVM stood out if both autonomic physiological and EEG features present a linear relationship with FMS scores. Apart from this, we found that the performance of non-linear activation function-based ANN model marginally exceeded SVM in both full and reduced feature set conditions. In general, the success of SVMs depends largely on choosing a suitable kernel for the specific problem to be solved whereas a neural network can approximate (using gradient descent) the non-linearity that best fits the data. Future studies can be planned for better explaining these results.

Acknowledgment

A special thanks to Prof. Ian M. Thornton of Department of Cognitive Science, University of Malta, for his help in the design of tunnel travel task for this study. Source code requests about the AI models should be sent to Ogechi Onuoha via email: ogechi.onuoha@glasgow.ac.uk

References

[1] J. A. Anguera *et al.*, 'Video game training enhances cognitive control in older adults', *Nature*, vol. 501, no. 7465, pp. 97–101, Sep. 2013, doi: 10.1038/nature12486.

[2] P. E. Wais, M. Arioli, R. Anguera-Singla, and A. Gazzaley, 'Virtual reality video game improves high-fidelity memory in older adults', *Sci. Rep.*, vol. 11, no. 1, Art. no. 1, Jan. 2021, doi: 10.1038/s41598-021-82109-3.

[3] G. Li, J. A. Anguera, S. V. Javed, M. A. Khan, G. Wang, and A. Gazzaley, 'Enhanced Attention Using Head-mounted Virtual Reality', *J. Cogn. Neurosci.*, vol. 32, no. 8, pp. 1438–1454, Aug. 2020, doi: 10.1162/jocn_a_01560.

[4] M. S. Dennison, A. Z. Wisti, and M. D'Zmura, 'Use of physiological signals to predict cybersickness', *Displays*, vol. 44, pp. 42–52, Sep. 2016, doi: 10.1016/j.displa.2016.07.002.

[5] R. Islam, Y. Lee, M. Jaloli, I. Muhammad, D. Zhu, and J. Quarles, 'Automatic Detection of Cybersickness from Physiological Signal in a Virtual Roller Coaster Simulation', in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, Mar. 2020, pp. 648–649. doi: 10.1109/VRW50115.2020.00175.

[6] N. Martin, N. Mathieu, N. Pallamin, M. Ragot, and J.-M. D. October 2018, 'Automatic recognition of virtual reality sickness based on physiological signals', *IBC*. https://www.ibc.org/automatic-recognition-of-virtual-reality-sickness-based-on-physiological-signals/3337.article (accessed Oct. 19, 2020).

[7] R. Islam *et al.*, 'Automatic Detection and Prediction of Cybersickness Severity using Deep Neural Networks from user's Physiological Signals', in *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, Nov. 2020, pp. 400–411. doi: 10.1109/ISMAR50242.2020.00066.

[8] M. Recenti *et al.*, 'Toward Predicting Motion Sickness Using Virtual Reality and a Moving Platform Assessing Brain, Muscles, and Heart Signals', *Front. Bioeng. Biotechnol.*, vol. 9, p. 132, 2021, doi: 10.3389/fbioe.2021.635661.

[9] J. Tauscher *et al.*, 'Exploring neural and peripheral physiological correlates of simulator sickness', *Comput. Animat. Virtual Worlds*, vol. 31, no. 4–5, Jul. 2020, doi: 10.1002/cav.1953.

[10] E. R. Ferrè and P. Haggard, 'Vestibular cognition: State-of-the-art and future directions', *Cogn. Neuropsychol.*, pp. 1–8, Mar. 2020, doi: 10.1080/02643294.2020.1736018.

[11] J. T. Reason, 'Motion sickness—some theoretical considerations', *Int. J. Man-Mach. Stud.*, vol. 1, no. 1, pp. 21–38, Jan. 1969, doi: 10.1016/S0020-7373(69)80009-X.

[12] E. Nalivaiko, S. L. Davis, K. L. Blackmore, A. Vakulin, and K. V. Nesbitt, 'Cybersickness provoked by head-mounted display affects cutaneous vascular tone, heart rate and reaction time', *Physiol. Behav.*, vol. 151, pp. 583–590, Nov. 2015, doi: 10.1016/j.physbeh.2015.08.043.

[13] J. Dowsett, C. S. Herrmann, M. Dieterich, and P. C. J. Taylor, 'Shift in lateralization during illusory self-motion: EEG responses to visual flicker at 10 Hz and frequency-specific modulation by tACS', *Eur. J. Neurosci.*, vol. 51, no. 7, pp. 1657–1675, 2020, doi: 10.1111/ejn.14543.

[14] N. Takeuchi, T. Mori, Y. Suzukamo, and S.-I. Izumi, 'Modulation of Excitability in the Temporoparietal Junction Relieves Virtual Reality Sickness', *Cyberpsychology Behav. Soc. Netw.*, vol. 21, no. 6, pp. 381–387, Jun. 2018, doi: 10.1089/cyber.2017.0499.

[15] A. Kyriakareli, S. Cousins, V. E. Pettorossi, and A. M. Bronstein, 'Effect of transcranial direct current stimulation on vestibular-ocular and vestibulo-perceptual thresholds':, *NeuroReport*, vol. 24, no. 14, pp. 808–812, Oct. 2013, doi: 10.1097/WNR.0b013e3283646e65.

[16] Q. Arshad *et al.*, 'Electrocortical therapy for motion sickness', *Neurology*, vol. 85, no. 14, pp. 1257–1259, Oct. 2015, doi: 10.1212/WNL.0000000000001989.

[17] B. Keshavarz and H. Hecht, 'Validating an Efficient Method to Quantify Motion Sickness', *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 53, no. 4, pp. 415–426, Aug. 2011, doi: 10.1177/0018720811403736.

[18] C.-L. Lin, T.-P. Jung, S.-W. Chuang, J.-R. Duann, C.-T. Lin, and T.-W. Chiu, 'Self-adjustments may account for the contradictory correlations between HRV and motion-sickness severity', *Int. J. Psychophysiol.*, vol. 87, no. 1, pp. 70–80, Jan. 2013, doi: 10.1016/j.ijpsycho.2012.11.003.

[19] G. Li and W.-Y. Chung, 'Detection of Driver Drowsiness Using Wavelet Analysis of Heart Rate Variability and a Support Vector Machine Classifier', *Sensors*, vol. 13, no. 12, pp. 16494–16511, Dec. 2013, doi: 10.3390/s131216494.

[20] J. A. Anguera, K. Lyman, T. P. Zanto, J. Bollinger, and A. Gazzaley, 'Reconciling the influence of task-set switching and motor inhibition processes on stop signal after-effects', *Front. Psychol.*, vol. 4, p. 649, 2013, doi: 10.3389/fpsyg.2013.00649.

[21] L. Buitinck *et al.*, 'API design for machine learning software: experiences from the scikit-learn project', *ArXiv13090238 Cs*, Sep. 2013, Accessed: Aug. 20, 2021. [Online]. Available: http://arxiv.org/abs/1309.0238

[22] Y.-C. Chen *et al.*, 'Spatial and temporal EEG dynamics of motion sickness', *NeuroImage*, vol. 49, no. 3, pp. 2862–2870, Feb. 2010, doi: 10.1016/j.neuroimage.2009.10.005.