Utilization of Neurophysiological Data to Classify Player Immersion to Distract from Pain

Kellyann Stamp[1[0000-0002-0351-6825]], Chelsea Dobbins[2[0000-0001-9420-2452]] and Stephen Fairclough[3[0000-0002-7850-5688]]

[1] Department of Computer Science, Liverpool John Moores University, Liverpool, L3 3AF, United Kingdom
[2] School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, QLD, 4072, Australia
[3] School of Natural Sciences and Psychology, Liverpool John Moores University, Liverpool, L3 3AF, United Kingdom
K.Stamp@2012.ljmu.ac.uk, C.M.Dobbins@uq.edu.au, S.Fairclough@ljmu.ac.uk

Abstract. Painful experiences during clinical procedures can have detrimental effects on the physical and mental health of a patient. Current pain reduction methods can be effective in reducing pain, however these methods are not without fault. Active distraction via computer games have been proven to effectively reduce the experience of pain. However, the potential of this distraction to effectively alleviate pain is dependent on players' engagement with the game, which is determined by the difficulty of the game and the skill of the player. This paper aims to model and classify immersion through increasingly difficult levels of game play, in the presence of pain, using functional Near Infrared Spectroscopy (fNIRS) and heart rate data. Twenty people participated in a study wherein fNIRS data (4 channels located at the prefrontal cortex, four channels located at the somatosensory cortex) and heart rate data were collected whilst participants were subjected to experimental pain, via the Cold Pressor Test (CPT). Participants played a computer game at varying difficultly levels as a distraction. Data were then pre-processed using an Acceleration Based Movement Artefact Reduction Algorithm (AMARA) and Correlation Based Signal Improvement (CBSI). Classification was subsequently undertaken using Linear Discriminant Analysis (LDA), Support Vector Machine (SVM) and Recursive Partitioning (rPart). The results demonstrate a maximum accuracy of 99.2% for the binary detection of immersion in the presence of pain.

Keywords: Functional Near Infrared Spectroscopy, Machine Learning, Classification, Immersion, Gaming, Pain.

1 Introduction

The experience of pain can have long lasting and detrimental effects on the sufferer, including Post Traumatic Stress Disorder and dissociative experiences [1], [2]. As such, adequately managing pain is crucial to alleviate discomfort. It is common for opioids to be administered as a method of pain relief and anxiety reduction in clinical settings. However, this method of pain control is associated with a number of serious adverse effects, including tolerance and withdrawal symptoms from opioid addiction, such as insomnia, vomiting, diarrhoea and tremors [3]. However, distraction techniques are non-pharmaceutical and can be an effective alternative in reducing the perception of pain, due to selective attention. This is especially prevalent in patients whose tendency to catastrophize can affect their perception of pain [4]. Pain is considered to be a threatening form of stimulation that can interrupt attention to other stimuli in the environment [5]. However, directing attention away from painful stimuli, through goal-orientated tasks, can modulate this interruptive function of pain and deliver analgesic relief from painful stimulation [6].

In order to maximize the effectiveness of distraction, the distracting stimuli or task must require a high level of cognitive effort in order to draw attention away from painful stimuli [7]. There are a wide variety of techniques available to distract people from pain, such as watching television or reading a book, however computer games have been proven to be the most effective approach [8]–[10]. The act of playing a computer game functions as an active distraction, requiring effortful pursuit of game-related goals and focused attention, whereas watching television is relatively passive, in comparison. For a distraction technique to effectively distract from pain, it must require a high degree of attentional focus and mobilization of mental effort. The psychological demands of a computer game can create an immersive experience that actively distracts from pain stimulation [11]. However, achieving this immersive experience is not straightforward. As noted by Fairclough et al. [12], most games are aimed at an 'ideal player', but this ideal player does not exist. Some players will have more or less gaming experience than said 'ideal player' and will therefore find a standard game to be either too easy or too difficult. It is important that the demands or difficulty of the game are optimized and adapted *per* player to engage the individual according to their capabilities. However, adaptive gaming is a method that can be used to enhance immersion by matching game demand to the engagement of the player [13]. In order to create an adaptive game, it is essential to quantify the attentional state of the player.

This paper aims to model and classify immersion of a player using implicit measures of physiological and neurological data, in order to distract from the experience of pain. The platform includes a number of devices to collect various streams of data, including Functional Near Infrared Spectroscopy (fNIRS), which is used to passively record neurovascular data, electrocardiogram (ECG) to collect heart rate and an accelerometer to measure head movement. The hypothesis of this paper is that immersion can be classified using implicit measures, such as fNIRS and heart rate, and that the experience of immersion can reduce the experience of pain. A study was undertaken to determine the effectiveness of using machine learning classifiers to predict the level of immersion. Collecting a subjective measure of immersion and a behavioral measure of pain enables us to understand how the physiology and physicality of a person is affected under such conditions. These data can then be used to verify the effects of pain tolerance, and the effects that a distraction task has on the cortical process.

2 Related Work

Csikszentmihalyi [14] coined the term ‘flow’ in reference to the psychological state experienced by a person when they are totally involved with a task [15]. The flow model focuses on three cognitive states: boredom, flow and anxiety, in order to describe this optimal realm of engagement and immersion.

Flow is an ideal state for an active distraction from pain, whilst both anxiety and boredom will prevent the player from entering the flow state. However, tailoring the level of game demand in order to achieve this state is difficult, due to the individual differences of players. This theory indicates that monitoring the players in-game performance is not sufficient for determining player state, because it would be unknown as to whether a poorer performance was due to anxiety (as the game is too difficult) or boredom (as the game is not difficult enough) [16]. Using game metrics alone would also not allow the level of cognitive effort that a player was exerting to be determined, as there would be no way to identify whether successful game play was due to high effort or low difficulty. This is especially relevant when considering the previous point that there is no ‘ideal player’ for a game. The experience of flow is considered to be incredibly delicate and easily broken. This means that outside stimuli are likely to decrease the chances of the player entering (or remaining in) the flow state [15]. For this reason, it would not be suitable for a player to provide a self-score of engagement, as the player would have to attend to this question rather than to the game. In this instance, adaptive gaming would be an ideal solution to enhance immersion by personalizing the game to the player.

Functional Near Infrared Spectroscopy (fNIRS) is a neuroimaging technique designed to measure neurovascular coupling/neuronal activation in the cortex. Neurovascular coupling is characterized by an increase in oxygenated hemoglobin (HbO) and a decrease in deoxygenated hemoglobin (Hbb). One major benefit of fNIRS over other techniques, such as function Magnetic Resonance Imaging (fMRI), is the absence of any need to confine a participant within an apparatus, which enables more flexible data capture to be undertaken [17]. In addition, fNIRS also has a greater spatial resolution than other techniques, such as electroencephalogram (EEG) [18]. fNIRS requires the placement of a montage of sources and detectors, secured to a cap, on a participant’s head. The sources emit infrared light, which can penetrate the skull and the outermost 10-15 millimeters of intracranial space. fNIRS works because skin, bone and tissue have a low absorbency rate for infrared light, whereas HbO and Hbb have a high absorbency rate. Therefore, the amount of light that is returned to the detectors indicates the changes in HbO and Hbb within the cortex. These data can be used to infer psychological concepts, such as attentional state, based on relative level of neuronal activation provoked by a set of stimuli or a specific task [19].

fNIRS has previously been used to monitor attentional state, using machine learning to classify the resulting data [18], [20]. These studies observed maximum classification accuracies of 80% and 90% respectively, in the distinction between high and low attentional states. fNIRS have also been used to observe pain signals in the brain with the intention of demonstrating the feasibility of using the technology as a measure of pain response [21]. One drawback of fNIRS data is that it is highly affected by movement from the participant, especially movement of the head, and physiological changes. For instance, if an individual becomes excited, their heart will beat faster and more blood will circulate [22]. This is true also of a pain response [23]. Both acceleration and physiological data are required to ensure the clearest possible fNIRS signal [24]. Head movement artefacts contained within fNIRS data can lead to false negatives and false positives, where it would appear that a participant is more or less engaged in a task than they actually are. Using filters based on acceleration allows for the removal of head movement artefacts. Acceleration based filters will also enable genuine responses in the brain to be preserved, and not discounted as movement artefacts.

3 Materials and Methods

Our approach focused on collecting fNIRS, heart rate, and head movement (via acceleration) data under three distinct conditions, whereby participants 1) played a computer game, 2) were exposed to a painful stimulus and 3) played a computer game whilst being exposed to a painful stimulus. The study was carried out to determine whether pain tolerance was increased when a distraction task was used (i.e. playing the game) and was conducted as a precursor to the development of a real-time adaptive gaming system. The purpose of this paper is to classify immersion of the player during increasingly difficult levels of game play in order to determine whether immersion can be quantified using objective physiological and neurophysiological measures.

3.1 Participants

Data were collected from 20 participants, of whom 6 were female. Participants were aged between 19 and 29 (M = 22.75 SD = 3.23). Exclusion criteria included being pregnant, history of cardiovascular disease, fainting, seizures, chronic or current pain, Reynaud’s disease or diabetes, fractures and open cuts or sores on the feet or calves. Participants were required to confirm that they were not currently taking any medication, with the exception of the contraceptive pill. A full review of the ethics of the experiment was undertaken, and approval was granted by the Liverpool John Moores University Research Ethics Committee. All participants were briefed before their experimental session and were provided with a detailed Participant Information Sheet prior to taking part. Full written consent was provided by each participant involved in this study.

3.2 Design

Participants were exposed to four levels of game difficulty (easy, medium, hard and impossible), which were determined during pre-piloting. After each gaming session, participants completed the subjective Immersive Experience Questionnaire (IEQ), which relates to overall feelings of immersion [15]. The study was designed in this way so that these self-scores could be used to label the physiological data for classification.

This would then enable the accuracy of using physiological signals as an indication of immersion to be determined.

Fig. 1.1 illustrates the data collection protocol that was undertaken. To ensure that the results collected during each step of this study were independent, a 90 second baseline period was established between each condition. The entire protocol was repeated a total of four times, to enable participants to play each of the four levels of game difficulty (i.e. easy, medium, hard and impossible). Throughout the course of the experiment, all four levels of the game were played in a randomized order. This ensured that the results collected were not influenced by the participant gaining more experience with the game on Easy and Medium levels before they then played Hard and Impossible levels. As such, the participant’s response to each level of the game is independent of their level of skill or their familiarity with the game.

3.3 Raw Data Collection

The components, depicted in Fig. 1.2 a – c, were utilized to collect raw physiological and neurological data during the experiment. The platform also consisted of components to distract (via a computer game) and induce (via a Cold Pressor Stimulus Tank) pain (see Fig. 1.2 d – e).

fNIRS data were used to determine changes relating both to game difficulty and the presence of pain. Data was collected using an Artinis™ OxyMon Mk III device, which measured cortical activity (see Fig. 1.2 a). A 2x4 cross-channel configuration was used, with a total of 2 sources and 8 detectors. This configuration was created on the fNIRS cap, which was worn on the participants head. Data was collected from 8 channels in total. Four channels were situated at the prefrontal cortex, between Fz and: F1, AFZ, F2 and FCz, and the remaining 4 channels were situated at the somatosensory cortex, between CPZ and: CP1, Cz, CP2 and Pz. This optode layout is illustrated in Fig. 1.3. Source-detector separation was 3cm and source optodes emitted light at 847nm and 761nm wavelengths. The device was configured to record optical density data at a sampling rate of 10Hz. Data was recorded using the Oxysoft data recording software. Zephyr™ BioHarness monitoring system was used to record electrocardiogram (ECG) data at a sampling rate of 250 Hz (see Fig. 1.2 b). The BioHarness was worn around the torso, underneath the participants clothing. A Shimmer3™ inertial measurement unit (IMU) was also used to record accelerometer data at a sampling rate of 512 Hz (see Fig. 1.2 c). The Shimmer3™ unit was worn around the participants head on an elasticated band. Care was taken to ensure that the band of the Shimmer3™ unit did not disturb the optodes on the fNIRS cap. Accelerometer data was used to influence the pre-processing of fNIRS data.

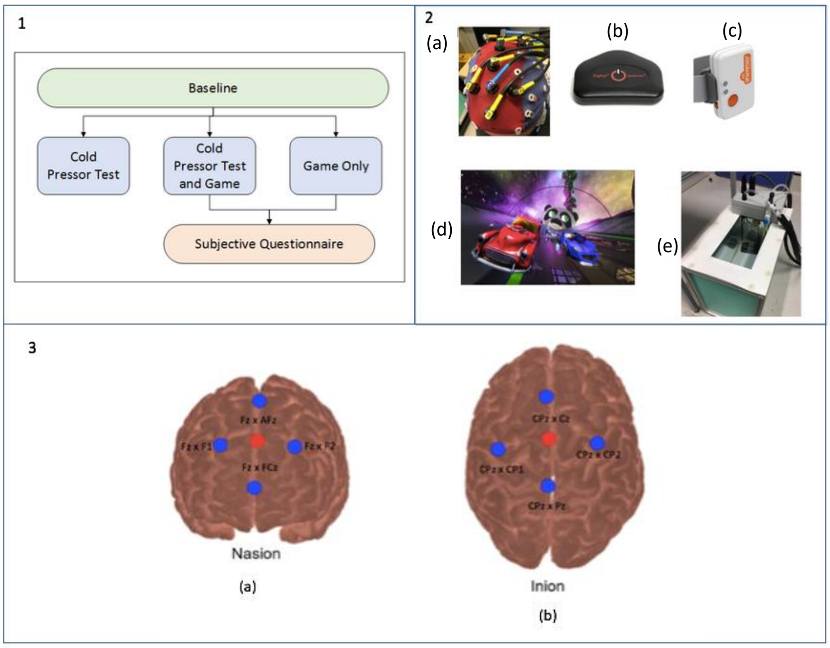

Fig. 1.1) The data collection protocol, which was repeated four times per participant to correspond to each level of game difficulty. During each cycle, the same level of game difficulty was played for both the Cold Pressor and Game and the Game Only conditions. **1.2)** Materials that were used to collect raw data during the experimental protocol. (a) fNIRS Equipment (Dancer Design), (b) Bioharness™ ECG Sensor, (c) Shimmer™ Accelerometer, (d) Computer game – Space Ribbon and (e) Cold Pressor Stimulus Tank. fNIRS Optode Placement. **1.3)** Two light sources (central), and 8 detectors surrounding the light source, with reference to the nasion and inion. (a) Front view. (b) Top view.

A racing game was used during the experimental sessions as the distraction (see Fig. 1.2. d). The racing game was a strategy-based game, wherein the goal was to finish the race in first position. Participants had the option to achieve a first-place position via a number of performance boosters, which could be picked up during gameplay. Participants could collect rockets, shields and invisibility boosters by driving over them on the track. Rockets, an offensive weapon, could be fired by the participant at competing vehicles on the track, which would cause the opponent vehicles to be temporarily disrupted. Shields, both an offensive and a defensive weapon, allowed the player to protect themselves from weapons being fired by opponent vehicles, and caused any opponent vehicle that came into contact with the players vehicle to be temporarily disrupted. Invisibility boosters were a defensive weapon that enabled the player to drive through opponent vehicles without causing collisions.

Through pilot testing, four set levels of game difficulty were established: easy, medium, hard, and impossible. Factors within the game were changed to create these four difficulty levels, including the Artificial Intelligence (AI) of the game, the amount of non-playable cars that were also on the track, gravity and race/maneuver speed. Each factor was represented by a number and were manipulated for each level's settings. For instance, during the easy level, the options chosen were set to the minimum, whilst during the impossible level the options were set to the maximum. The options for the Medium and Hard levels were set in-between the minimum and maximum and were determined via pilot testing.

Experimental pain was induced via the Cold Pressor Test (CPT) [22] (see Fig. 1.2 e). Throughout the duration of the experimental session, water was kept at a consistent temperature of 2 degrees centigrade and was not warmed by the participant's foot. During the stages whereby, participants submerged their foot in the CPT and then also played the game, they were instructed to remove their foot from the cold pressor stimuli tank when they felt that the pain they were experiencing was unbearable. Each game condition lasted for 3 minutes, which is consistent with the recommended maximum duration of the CPT. Over the course of the experiment, participants alternated the foot that was placed in the cold pressor. This was to ensure that repeated immersion of the same foot would not have an effect on pain tolerance. Each CPT was timed, which provided an objective measure of pain tolerance at each difficulty level. This result was kept blind from the participant. Participants played each level of game difficulty with and without the induction of experimental pain. The CPT was used to ensure that the experience of immersion did have an effect on the experience of pain.

In total, 690,028 instances of raw fNIRS data, 16,763,628 instances of raw heart rate data and 26,177,699 instances of raw accelerometer data were collected; thus totalling 43,631,355 instances of raw data overall.

3.4 Data Pre-Processing

A data pre-processing pipeline has been created, using a variety of filters and algorithms, which were applied to the raw data (see Fig. 2). These filters were applied to ensure that the signals, which would be used for classification, were free from artefacts, which could affect the classification results.

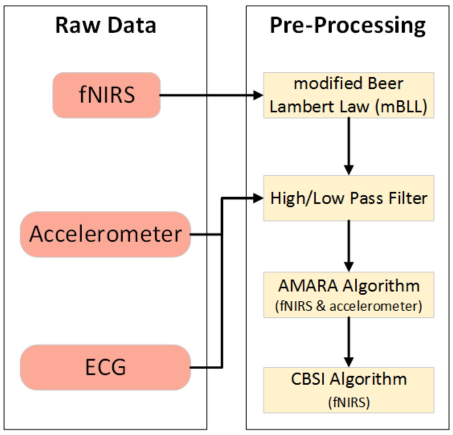

Fig. 2. Data pre-processing pipeline

Firstly, in order to determine activation in the target cortices, the raw Optical Density data that were collected via the fNIRS cap were converted into measures of oxygenated hemoglobin (HbO) and deoxygenated hemoglobin (Hbb), using the modified Beer Lambert Law (mBLL) [25]. In order to ensure that the data related specifically to the condition that the participant was experiencing, and not to unrelated activity within the brain, the baseline and condition data were combined prior to the application of the modified Beer Lambert Law, before the baseline data was again removed. This ensured that the converted Hemoglobin changes reflected only data that was related directly to the condition.

Following this conversion, the fNIRS data was filtered using a 6th Order Chebyshev filter, with passband edge frequencies of 0.5 and 0.1 respectively, for high and low pass filtering. These filters were applied to reduce noise within the signal that related to heart rate and respiration [26], as well as Mayer waves, which occur within the fNIRS signal due to changes in arterial pressure [27]. The same filters were also applied to the ECG and accelerometer data to reduce noise within these signals.

The fNIRS data were then treated with two head movement related filters: Acceleration-Based Movement Artefact Reduction Algorithm (AMARA) and Correlation Based Signal Improvement (CBSI), which were used to ensure the results represented genuine hemodynamic response. Both fNIRS and accelerometer data were firstly processed using the Acceleration-Based Movement Artefact Reduction Algorithm (AMARA) [28]. Artefacts relating to head movement are commonly found in fNIRS signals, which causes a change in blood flow to the brain. These changes can appear to represent changes in activation if they are not removed from the signal. AMARA detects periods of movement within an accelerometer signal and then compares these periods of movement to the fNIRS data. Where it is found that the moving standard deviation (MSD) of the fNIRS signal has changed considerably during the same period of time that movement has been detected within the accelerometer signal, these segments of fNIRS data are marked as ‘artefact’ segments. Segments where the MSD of both the accelerometer and fNIRS data have no significant deviation are marked as ‘acceptable’ segments. Reconstruction of artefact segments uses forward and backward baseline adjustments and interpolation to reconstruct the entire signal, with the movement artefacts corrected [28].

A further consequence of head movement artefacts is that they can cause a positive correlation between HbO and Hbb [29]. Usually, these signals are negatively correlated, as a drop in Hbb is expected when HbO rises, and vice versa [29]. Therefore, if the signals illustrate that there isn’t a strong negative correlation between HbO and Hbb, it could be an indication that there is remaining noise contained within the signal. Therefore, in order to correct the correlation between the HbO and Hbb signals, the final stage included applying the Correlation Based Signal Improvement (CBSI) algorithm to the fNIRS data[30].

3.5 Experimental Measures and Feature Extraction

Following pre-processing, additional measurements were calculated from the HbO and Hbb data. The first was Total Hemoglobin (HbT), which occurs through addition of the

HbO and Hbb signals to provide a signal that details the total cortical activity, as in equation (1). The second was Hemoglobin Difference, which is created by finding the difference between HbO and Hbb Hemoglobin, as in equation (2).

$$HbT = HbO + Hbb \tag{1}$$

$$HD = Hb0 - Hbb \tag{2}$$

As reported by Xu et al. [31], Hemoglobin Difference can achieve better results than HbO, Hbb and HbT when fNIRS data are being used for machine learning classifications. The creation of these new measurements provided 32 different features: 8 HbO, 8 Hbb, 8 HbT and 8 Hemoglobin Difference, each relating to the 8 original channels of interest.

In order to determine the heart rate of the participants, ECG data collected from the Bioharness was converted into beats per minute (BPM), as depicted in equation (3).

$$BPM = 60000/IBI(rPeakECG(n) - rPeakECG(n-1)) \tag{3}$$

This process involved calculating the inter-beat interval (IBI), which corresponds to the time between consecutive R peaks in an ECG signal, which indicates that the heart has beaten. BPM detection was undertaken by identifying the R peaks in the signal and then finding the difference in milliseconds between two successive beats and then dividing by 60,000 (the number of milliseconds in a minute).

The data were then separated into 8 second windows, which corresponds to the Hemodynamic Delay that is present in fNIRS data. This occurs when the response to the onset of stimuli has a delay of several seconds after the stimuli has been introduced, before changes in the signal reflect this [32]. Within each 8 second epoch, standard descriptive statistics were extracted from the fNIRS data, including: mean, median, range, minimum, maximum and standard deviation. These features were created for each of the 32 original measurements discussed above, to create a feature set comprised of 192 features in total. These descriptive statistics were also extracted from the BPM data, thus totaling 198 features. Following the creation of the feature set, each participant's dataset was normalized using a Standardized Z Score [33] and combined into a single dataset. It should be noted that two participant's data were excluded from this dataset due to short CPT immersion times, where none of their immersions in any condition exceeded 25 seconds.

3.6 Data Labelling

Subjective levels of immersion were gathered via the Immersive Experience Questionnaire (IEQ), which consists of 32 questions that are scored on a 7-point Likert scale from 1 (*Not at All*) to 7 (*A Lot*). The resulting scores from the IEQ were used as subjective labels for the neurophysiological data. As described above, participants played each level of the game (easy, medium, hard and impossible) under two conditions: 1) game only and 2) experiencing pain whilst playing the game. The average IEQ score was then calculated for each game level, per condition. The data was then labelled such that any participant whose score for each level/condition was a) greater than or equal to the mean was labelled '*immersed'* or b) any score below the mean was labelled '*not immersed*'. As a result of this labelling process, Fig. 3 illustrates the frequency of immersed/not immersed participants for each level during the a) game only condition and b) game and pain condition.

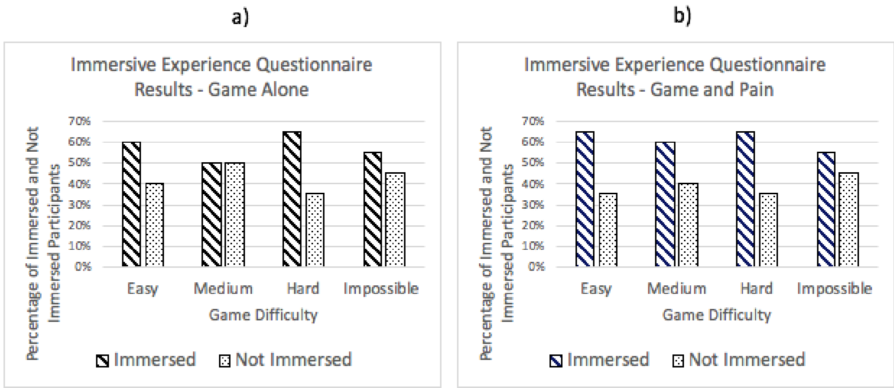

Fig. 3. a) IEQ distribution of *Immersed* and *Non-Immersed* participants during the Game Only condition shown as a percentage and **b)** IEQ distribution of *Immersed* and *Non-Immersed* participants during the Game and Pain condition shown as a percentage.

As it can be seen in Fig. 3, the datasets per level, per condition are unbalanced, apart from the medium level of the game only condition, whereby the balance is 50/50. As such, in order to balance the remaining datasets, the majority classes (i.e. those labelled as *immersed*) were randomly undersampled using the *SpreadSubsample* function in Weka [34] in order to create equally labelled data.

3.7 Feature Selection

Feature selection was undertaken using the RELIEFF algorithm [35]. RELIEFF uses a *k* nearest neighbor approach to weight the estimated quality of features. The *k* value is the value that determines the number of nearest neighbors that should be compared to each data point. This is done in order to determine the nearest values in the same class (hits) and the nearest values in a different class (misses). Each feature is weighted to estimate its quality, based on the amount of hits and misses.

In order to determine how each set of measures (frontal fNIRS sites, central fNIRS sites and heart rate) independently performed during classification, these three sets of measures were separated into individual datasets. The purpose of this separation was to determine the best set of features to use in order to classify immersion. The RELIEFF algorithm was then applied independently to the frontal and central fNIRS data. Due to the heart rate measures only containing six features, feature selection was not performed over this data. The resulting weights were then ordered, from highest to lowest, and plotted on a graph (see Fig. 4). The point in the graph where the "elbow" appears indicates the most relevant features and the feature set is cut down at this point.

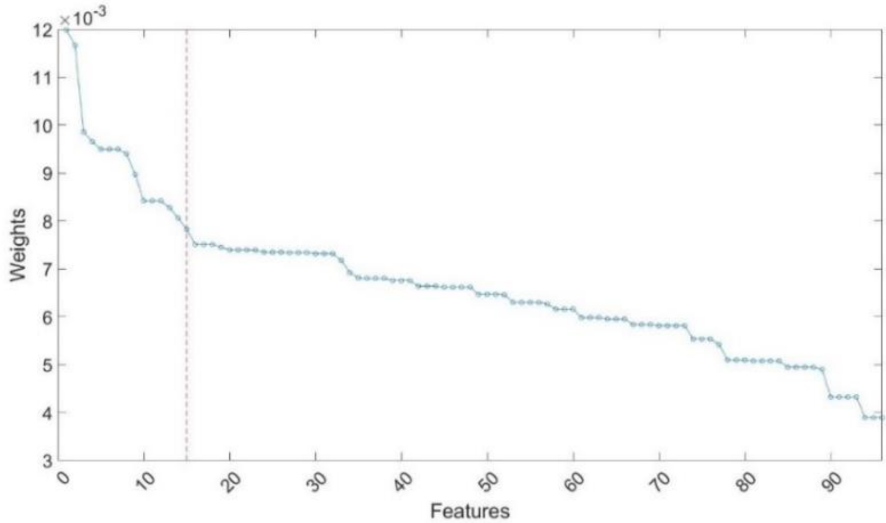

Fig. 4. Example of RELIEFF feature selection of the frontal fNIRS sites. The graph indicates that the elbow, where the relevance of the features becomes obsolete, is at 36 features.

The application of the RELIEFF algorithm determined that a total of 15 Frontal and 12 Central fNIRS features were relevant for classification. The features that were selected can be seen in Table 1.

Table 1. Features selected for classification determined by the RELIEFF Algorithm

Frontal Features	Weight	Central Features	Weight
Fz x AFZ HbT Mean	0.011988036	CPZ x CP2 HbT Mean	0.012807196
Fz x F1 HbT Mean	0.011664447	CPZ x CP2 HbT Median	0.010977391
Fz x F1 HbT Max	0.009856833	CPZ x Cz Oxy Max	0.00930525
Fz x F1 HbT Median	0.00965673	CPZ x Cz Diff Max	0.009305134
Fz x F1 Oxy Max	0.009499116	CPZ x Cz Deoxy Min	0.009305041
Fz x F1 Diff Max	0.009498993	CPZ x CP2 Diff Min	0.009290371
Fz x F1 Deoxy Min	0.009498955	CPZ x CP2 Deoxy Max	0.009290342
Fz x AFZ HbT Median	0.009402018	CPZ x CP2 Oxy Min	0.009290215
Fz x F2 HbT Max	0.008972051	CPZ x CP2 HbT Min	0.009095969
Fz x F2 Oxy Max	0.008418477	CPZ x Cz HbT Mean	0.009075785
Fz x F2 Diff Max	0.008418435	CPZ x Cz HbT Max	0.008980522
Fz x F2 Deoxy Min	0.008418274	CPZ x CP1 HbT Max	0.008715578
Fz x F2 HbT Mean	0.008276955		
Fz x F2 HbT Median	0.008061099		
Fz x FCZ HbT Max	0.007826961		

4 Results

The evaluation uses both parametric and non-parametric machine learning classifiers, including Linear Discriminant Analysis (LDA), Support Vector Machine (SVM) and Recursive Partitioning (rPart), to classify immersion using the selected features in the frontal fNIRS sites, central fNIRS sites and heart rate datasets. These classification algorithms are commonly used in fNIRS studies that involve machine learning [36]. Classification was carried out using RStudio Version 1.1.414 and the Machine Learning in R (MLR) package. The results were validated using k-fold cross validation, where k = 10. Performance measurements that were calculated included:

- *Accuracy* – an overall score of the performance
- *F_1 Score* – the weighted average of the precision and recall of the classifier
- *Balanced Error Rate (BER)* – an average of classification errors that occur in each class

Binary classification was performed for each of the four game levels (Easy, Medium, Hard and Impossible), per condition (i.e. 1) Game and Pain and 2) Game Only). For each condition, data were labelled as '*Immersed'* or '*Not Immersed*' according to the average IEQ score per level (see section 3.6). The aim of this classification was to determine whether immersion could be classified from heart rate and fNIRS data independently, and to identify which type of data was more accurately classified. A combined dataset (fNIRS and heart rate) was also classified, to identify whether this would improve the classification results. Classification was carried out over *Game Only*, and *Game and Pain* conditions independently, to examine the effect of the experience of pain on the classification of immersion.

Each classifier was utilized independently using only the 1) heart rate features, 2) frontal site features, 3) central site features and 4) a merger of 1 – 3 (i.e. heart rate, frontal and central site features together). The purpose of this was to evaluate each set of features independently to establish the set of features that provided the best results. Data were separated into four difficulty levels for the classification of immersion, to determine whether the level of difficulty would have an effect on the accuracy of the classification. Data were also separated by *Game and Pain* and *Game Only* conditions, to determine whether the inclusion of the CPT has an effect on the accuracy of the classifier. The results of the Heart Rate analysis can be seen in Table 2.

Table 2. Heart rate classification results per condition/level

Condition	Performance Measure	Game and Pain			Game Only		
		LDA	SVM	rPart	LDA	SVM	rPart
Easy	Accuracy	85.1%	**85.6%**	84%	86%	***86.6%***	85.6%
	BER	**49.6%**	50%	50.8%	48%	50%	***45.6%***
	F1	91.7%	**92.1%**	91.2%	92.3%	***92.6%***	92%
Medium	Accuracy	***80.7%***	79.8%	71.8%	**70.6%**	68.8%	61.7%

	BER	**50%**	50.5%	53.9%	***42.6%***	45.1%	44.8%
	F1	***89.07%***	88.58%	83.18%	**80.9%**	80.2%	71.6%
Hard	Accuracy	***77.8%***	***77.8%***	68.6%	74.9%	**77.4%**	69.4%
	BER	**48.6%**	50%	50.7%	***44.2%***	50%	50%
	F1	87.1%	***87.4%***	80.5%	85.1%	**86%**	80.4%
Impossible	Accuracy	59.5%	58.9%	***62%***	56%	**58.4%**	49%
	BER	45.5%	47.2%	***39.8%***	**49.5%**	50.5%	52%
	F1	70.4%	***72%***	68.2%	68.7%	***73%***	56.8%

The results in Table 2 report maximum accuracies during both Game and Pain and Game Only conditions during Easy, Medium and Hard games of 86.6%, 80.7% and 77.8%, respectively, whilst the Impossible condition provided adequate results of 62%. The lowest error rates for each level were observed during the Game Only condition, with 45.6% (Easy), 42.6% (Medium) and 44.2% (Hard), whilst the lowest overall BER was observed during the Impossible level at 39.8%. The results illustrate that LDA and SVM outperformed rPart in most cases, with the highest F1 score (92.6%) being observed using SVM during the Game Only/Easy condition. This illustrates that when the game was relatively simple, a linear model using heart rate only features was adequate to distinguish immersion. The results indicate that comparable levels of classification accuracy can be found between the Game and Pain and Game Only conditions of the same difficulty levels. Overall, maximum classification accuracy of 86.6% and BER of 92.6% was found in the Game Only/Easy condition, using SVMs. To evaluate whether these results can be improved upon using fNIRS, the results of the frontal fNIRS sites can be seen in Table 3.

Table 3. Frontal fNIRS classification results per condition/level

Condition	Performance Measure	Game and Pain LDA	Game and Pain SVM	Game and Pain rPart	Game Only LDA	Game Only SVM	Game Only rPart
Easy	Accuracy	**85.6%**	**85.6%**	76.8%	86.1%	***86.6%***	82.1%
	BER	50%	50%	***47.9%***	**48.9%**	50%	52.7%
	F1	92.1%	**92.2%**	86.6%	92.4%	***92.7%***	90%
Medium	Accuracy	***80.7%***	***80.7%***	77.2%	67.6%	**68%**	54%
	BER	50%	50%	***46.7%***	50.5%	**50%**	52.3%
	F1	***89.2%***	89.1%	86.5%	80.2%	**80.8%**	65%
Hard	Accuracy	77.1%	***78%***	63.8%	77%	**77.4%**	69%
	BER	50.5%	***45%***	53.9%	50.3%	50%	**46.7%**
	F1	86.9%	***87.2%***	76.6%	86.8%	**87%**	79.7%
Impossible	Accuracy	56.1%	56.1%	***63.8%***	53.8%	**57.3%**	54.8%
	BER	52.5%	51%	***38%***	54%	51.4%	**46.7%**

	F1	**71.3%**	70.8%	70%	69.4%	***72.3%***	61%

Table 3 illustrates that using the frontal fNIRS sites yielded comparable results as the Heart Rate results seen in Table 2. Maximum accuracies of 86.6% (Easy), 80.7% (Medium), 78% (Hard) and 63.8% (Impossible) were reported for each level, across both Game and Pain and Game Only conditions. The lowest overall BER (38%) was observed during the Game and Pain/Impossible condition using rPart. The highest overall accuracy (86.6%) and F1 score (92.7%) were achieved using an SVM classification during the Easy/Game Only condition, which was similar to the HR results. These results illustrate that the classifiers seem to perform best on the Easy level, using the game on its own and when pain is not present. However, when the game increases in complexity, from the Medium – Impossible levels, the presence of pain does not affect the results and produces higher accuracies. To evaluate whether central fNIRS sites can provide further improvement, the results of the central fNIRS sites can be seen in Table 4.

Table 4. Central fNIRS classification results per condition/level

		Game and Pain			Game Only		
Condition	Performance Measure	LDA	SVM	rPart	LDA	SVM	rPart
Easy	Accuracy	85.7%	85.6%	***88.2%***	86.8%	86.7%	**87.1%**
	BER	50%	50%	***27.7%***	40%	50%	**33.1%**
	F1	92%	92.1%	***93%***	**92.7%**	**92.7%**	**92.7%**
Medium	Accuracy	83.7%	79.3%	***96.9%***	83.3%	80.2%	**90.3%**
	BER	20.9%	40.8%	***7.8%***	19.7%	25.1%	**11%**
	F1	89.3%	87.8%	***98.1%***	88%	86.1%	**92.7%**
Hard	Accuracy	77.8%	77.9%	***99.2%***	74.8%	77.5%	**95.8%**
	BER	50%	50%	***1.3%***	50.9%	50%	**7.5%**
	F1	87.3%	87.4%	***99.4%***	85.3%	87%	**97.2%**
Impossible	Accuracy	54.2%	58%	***95.3%***	60.1%	62%	**88.9%**
	BER	49%	50.3%	***5.7%***	43.3%	44%	**11.5%**
	F1	66.8%	72.3%	***96.2%***	68.4%	72.7%	**90.5%**

The results presented in Table 4 illustrate that using the central fNIRS sites are an improvement to the results achieved using the frontal sites, especially when an rPart classification is implemented. Maximum accuracies of 88.2% (Easy), 96.9% (Medium), 99.2% (Hard) and 95.3 (Impossible) were reported for each level, across both Game and Pain and Game Only conditions. The highest overall accuracy (99.2%) and F1 score (99.4%) and lowest BER (1.3%) were achieved using the rPart classifier during the Hard/Game and Pain condition. In contrast with previous results, the results from the rPart learner are consistently better than that of LDA and SVM for all measures. We were also interested to compare how a combination of the three separate datasets would perform. The results from this combined dataset can be seen in Table 5.

Table 5. Combined dataset (fNIRS and heart rate) classification results per condition/level

		Game and Pain			Game Only		
Condition	Performance Measure	LDA	SVM	rPart	LDA	SVM	rPart
Easy	Accuracy	83.5%	**85.7%**	76.3%	83.1%	***86.6%***	80.6%
	BER	51.3%	**50%**	50.8%	***47.1%***	50%	52.2%
	F1	90.7%	**92.2%**	86.2%	90.5%	***92.7%***	88.8%
Medium	Accuracy	78.5%	***80.7%***	72%	65.4%	**68.4%**	64.1%
	BER	51.3%	**50%**	51.3%	48.8%	45.8%	***43.4%***
	F1	87.8%	***89%***	83.2%	77.4%	**79.9%**	73.9%
Hard	Accuracy	73.2%	***77.9%***	65.5%	74%	**77.4%**	74%
	BER	52.9%	50%	**49.8%**	50.1%	50%	***40.6%***
	F1	84.4%	***87.4%***	77.5%	84.4%	**87.1%**	83.3%
Impossible	Accuracy	51.7%	59.6%	***61%***	47%	**57.6%**	52.5%
	BER	52.5%	49.2%	**40.5%**	58.4%	51.2%	***50.2%***
	F1	63.7%	***73.5%***	65.6%	61.2%	**72.1%**	59.9%

As the results in Table 5 illustrate that the combined dataset does not outperform any individual dataset. Maximum accuracies of 86.6% (Easy), 80.7% (Medium), 77.9% (Hard) and 61% (Impossible) were reported for each level, across both Game and Pain and Game Only conditions. The lowest overall BER (40.5%) was observed during the Game and Pain/Impossible condition using rPart. The highest overall accuracy (86.6%) during the Easy/Game Only condition is comparable to the HR (Table 2) and frontal (Table 3) results. Overall, the results presented in Table 5 indicate that there is no benefit of using the combined data set over the individual datasets. To summarize, the highest accuracy (99.2%) and F1 (99.4%) and lowest BER (1.3%) were achieved using the central dataset, during the Hard/Game and Pain condition and using the rPart classifier. As such, central fNIRS sites, together with rPart, appears to be the most appropriate set of data and classifier to use to detect immersion in the presence of pain.

5 Discussion

It has been established that a computer game is an active distraction task capable of increasing pain tolerance. The rationale of this study was to use fNIRS and heart rate data to differentiate between immersive and non-immersive conditions. Our approach was undertaken in distinct phases, which were utilized to assess the relative contribution of variables derived from fNIRS (frontal and central sites) and heart rate data. Feature selection was undertaken using the RELIEFF algorithm, which indicated that HbT measures were consistently well represented in both the frontal and central sites. It was interesting to note that of the 27 selected features, both from frontal and central sites, 15 of these features (55.5%) were descriptive statistics gathered from the HbT signal. It is also interesting that statistics associated with variation (e.g. min, max) are represented more than measures relating to central tendency (e.g. mean, median). Feature selection has enabled us to determine that measures associated with variation, gathered from the HbT signal, would be the best features derived from fNIRS to utilize for further work. The same pattern, i.e. sensitivity of total hemoglobin in the frontal cortices to the cold pressor test, was observed in earlier studies [3], [5]. Feature selection also indicates that channels Fz x F1, Fz x AFZ, Fz x FCZ, CPZ x CP1, CPZ x Cz and CPZ x CP2 contain the most relevant features. These findings are beneficial for the creation of a real-time system, as the run-time of the real-time data analysis protocol is vital. When a real-time protocol is considered, the removal of channels and measures which did not achieve high weights in this study could improve the run-time of the protocol and enable a more efficient real-time classification.

The classification methodology involved testing heart rate, fNIRS at frontal and central sites independently and then together using LDA, SVM and rPart learners (Table 2 - Table 5). The results achieved indicated that the classification accuracies were lowest during the Impossible condition, which could be due to player frustration. However, they may still have been intent on winning the game, and therefore still exhibiting focus towards the task. This indicates that the subjective measures that were gathered are not effectively measuring immersion at the same rate as the objective measures. However, as a real time adaptive game would still need training data, it may be better to label the data using objective measures rather than subjective – the use of conative probing [37] rather than a subjective score would be one example of this.

The results indicate that the effect of pain does not significantly affect the classification of immersion and that the heart rate features produced comparable results to the frontal fNIRS sites but did not outperform the central fNIRS sites. This is a positive result for our study, as it is important to be able to classify immersion in the presence of pain. The results of the HR classification in comparison to fNIRS classification indicate that, although HR can be classified to the same level as frontal fNIRS sites, central fNIRS sites still provide a more accurate classification. The efficacy of fNIRS classification indicates that the application of fNIRS technology is justifiable even considering the more advanced data collection protocol. Although all participants were subjected to the same pain protocol, some participants may have a naturally lower or higher pain tolerance, and therefore feel more or less pain than anticipated at various stages of the study. This means that the signals in the brain relating to pain could have an effect on the classification when a real-time system is used, depending on the level of pain that a participant is feeling. However, as the results between Game and Pain and Game Only conditions were comparable, we hypothesize that the experience of pain should not affect the accuracy of a classification in a real-time system.

The results have important implications for the future development of a neuroadaptive game. In the first instance, the feature selection process identified the metric (HbT) and specific sites that were most responsible for distinguishing levels of game demand. By focusing on these signals and measures in the design of a neuroadaptive game, it is possible to reduce processing time required for classification during a real-time protocol. Using these results, we intend to adapt the methods that have been used this paper into a real-time system. We believe that adapting a game in this way would enable us to create a more immersive game experience, and therefore reduce the perception of pain.

6 Conclusion and Future Work

The purpose of this work was to identify the foundations of a real-time neuroadaptive system to distract people from painful experiences. The results achieved within this paper have enabled us to identify relevant fNIRS sites and measures that could be used in such a system. A maximum classification accuracy of 99.2% was achieved using the rPart classifier for the detection of immersion in the presence of pain.

Alongside the results shown, we accept the limitations of the current study. For instance, short-distance electrodes were not used during the fNIRS measurements. Short distance electrodes are used to record, and later reduce, the amount of noise within a signal that is not related to neurovascular coupling. Signals such as this are recorded from the extracerebral layers, rather than the cerebral tissue layer, and are task-evoked but not related to neurovascular coupling [36]. Further issues in this study may have arisen due to systemic effects contained within the fNIRS signal that have not been removed [29]. Although filters were applied to the fNIRS signal to reduce noise relating to heart rate, respiration and blood pressure, we cannot be certain that all of these features were removed from the signal. In future studies, steps could be taken to reduce the presence of systemic effects such as the use of heart rate, respiration and blood pressure signals by building personalized filters to be applied to each fNIRS signal [38]. We believe that the future use of personalized filters and short-distance channels could improve on the results that have been gathered in this study.

Acknowledgments

The authors would like to thank Onteca Inc., who provided an adapted version of their title *Space Ribbon* for use in this study. The authors would also like to thank all of the participants for agreeing to take part in the study.

References

[1] R. D. Broad and K. Wheeler, "An adult with childhood medical trauma treated with psychoanalytic psychotherapy and EMDR: A case study," *Perspect. Psychiatr. Care*, vol. 42, no. 2, pp. 95–105, 2006.

[2] T. H. Diseth, "Dissociation following traumatic medical treatment procedures in childhood: A longitudinal follow-up," *Dev. Psychopathol.*, vol. 18, no. 1, pp. 233–251, 2006.

[3] K. J. Anand *et al.*, "Tolerance and withdrawal from prolonged opioid use in critically ill children," *Pediatrics,* vol. 125, no. 5, pp. e1208-25, 2010.

[4] S. Greenbaum *et al.*, "The Impact of Catastrophizing," *Anesthesiology*, no. 6, pp. 1292–1301, 2014.

[5] Y. Hua, R. Qiu, W. Y. Yao, Q. Zhang, and X. L. Chen, “The Effect of Virtual Reality Distraction on Pain Relief During Dressing Changes in Children with Chronic Wounds on Lower Limbs,” *Pain Manag. Nurs.*, vol. 16, no. 5, pp. 685–691, 2015.

[6] C. Eccleston and G. Crombez, “Pain demands attention: a cognitive-affective model of the interruptive function of pain,” *Psychol Bull*, vol. 125, no. 3, pp. 356–366, 1999.

[7] W. K.A., “Interactive versus passive distraction for acute pain management in young children: the role of selective attention and development,” *J. Pediatr. Psychol.*, vol. 38, no. 2, pp. 202–212, 2013.

[8] E. Jameson, J. Trevena, and N. Swain, “Electronic gaming as pain distraction,” *Pain Res. Manag.*, vol. 16, no. 1, pp. 27–32, 2011.

[9] S. J. Bantick, R. G. Wise, A. Ploghaus, S. Clare, S. M. Smith, and I. Tracey, “Imaging how attention modulates pain in humans using functional MRI.,” *Brain*, vol. 125, no. Pt 2, pp. 310–319, 2002.

[10] V. Legrain, S. Van Damme, C. Eccleston, K. D. Davis, D. A. Seminowicz, and G. Crombez, “A neurocognitive model of attention to pain: Behavioral and neuroimaging evidence,” *Pain*, vol. 144, no. 3, pp. 230–232, 2009.

[11] K. E. Weiss, L. M. Dahlquist, and K. Wohlheiter, “The effects of interactive and passive distraction on cold pressor pain in preschool-aged children,” *J. Pediatr. Psychol.*, vol. 36, no. 7, pp. 816–826, 2011.

[12] S. H. Fairclough, K. Gilleade, K. C. Ewing, and J. Roberts, “Capturing user engagement via psychophysiology: measures and mechanisms for biocybernetic adaptation,” *Int. J. Auton. Adapt. Commun. Syst.*, vol. 6, no. 1, p. 63, 2013.

[13] K. C. Ewing, S. H. Fairclough, and K. Gilleade, “Evaluation of an Adaptive Game that Uses EEG Measures Validated during the Design Process as Inputs to a Biocybernetic Loop,” *Front. Hum. Neurosci.*, vol. 10, no. May, pp. 1–13, 2016.

[14] L. Nacke, M. Kalyn, C. Lough, and R. Mandryk, “Biofeedback game design: using direct and indirect physiological control to enhance game interaction,” *Proc. SIGCHI …*, pp. 103–112, 2011.

[15] C. Jennett *et al.*, “Measuring and defining the experience of immersion in games,” *Int. J. Hum. Comput. Stud.*, vol. 66, no. 9, pp. 641–661, 2008.

[16] L. E. Nacke and C. A. Lindley, “Flow and Immersion in First-person Shooters: Measuring the Player’s Gameplay Experience,” *Proc. 2008 Conf. Futur. Play Res. Play. Share*, pp. 81–88, 2008.

[17] A. R. Harrivel, D. H. Weissman, D. C. Noll, and S. J. Peltier, “Monitoring attentional state with fNIRS.,” *Front. Hum. Neurosci.*, vol. 7, no. December, p. 861, 2013.

[18] K. Izzetoglu, S. Bunce, M. Izzetoglu, B. Onaral, and K. Pourrezaei, “fNIR spectroscopy as a measure of cognitive task load,” *Proc. 25th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (IEEE Cat. No.03CH37439)*, vol. 4, pp. 3431–3434, 2003.

[19] E. Olsson, G. Ahlsén, and M. Eriksson, “Skin-to-skin contact reduces near-infrared spectroscopy pain responses in premature infants during blood sampling,” *Acta Paediatr. Int. J. Paediatr.*, vol. 105, no. 4, pp. 376–380, 2016.

[20] C. Jennett, A. Cox, and P. Cairns, “Investigating Computer Game Immersion and the Component Real World Dissociation,” *Proc. 27th Int. Conf. Ext. Abstr. Hum. factors Comput. Syst. CHI EA 09*, no. February 2007, pp. 3407–3412, 2009.

[21] B. D. Kussman *et al.*, “Capturing pain in the cortex during general anesthesia: Near infrared spectroscopy measures in patients undergoing catheter ablation of arrhythmias,” *PLoS One*, vol. 11, no. 7, pp. 1–13, 2016.

[22] C. L. Von Baeyer, T. Piira, C. T. Chambers, M. Trapanotto, and L. K. Zeltzer, “Guidelines for the cold pressor task as an experimental pain stimulus for use with children,” *J. Pain*, vol. 6, no. 4, pp. 218–227, 2005.

[23] W. B. Baker, A. B. Parthasarathy, D. R. Busch, R. C. Mesquita, J. H. Greenberg, and A. G. Yodh, “Modified Beer-Lambert law for blood flow,” *Biomed. Opt. Express*, vol. 5, no. 11, p. 4053, 2014.

[24] N. Naseer, F. M. Noori, N. K. Qureshi, and K.-S. Hong, “Determining Optimal Feature-Combination for LDA Classification of Functional Near-Infrared Spectroscopy Signals in Brain-Computer Interface Application,” *Front. Hum. Neurosci.*, vol. 10, no. May, pp. 1–10, 2016.

[25] L. Kocsis, P. Herman, and A. Eke, “The modified Beer-Lambert law revisited,” *Phys. Med. Biol.*, vol. 51, no. 5, 2006.

[26] D. Bontrager, D. Novak, R. Zimmermann, R. Riener, and L. Marchal-crespo, “Physiological Noise Cancellation in fNIRS

using an Adaptive Filter Based on Mutual Information *," *Present. IEEE Int. Conf. Syst. Man, Cybern. San Diego, CA, USA, Oct. 5-8, 2014*, 2014.

[27] C. Julien, "The enigma of Mayer waves: Facts and models," *Cardiovasc. Res.*, vol. 70, no. 1, pp. 12–21, 2006.

[28] A. J. Metz, M. Wolf, P. Achermann, and F. Scholkmann, "A new approach for automatic removal of movement artifacts in near-infrared spectroscopy time series by means of acceleration data," *Algorithms*, vol. 8, no. 4, pp. 1052–1075, 2015.

[29] M. Caldwell, F. Scholkmann, U. Wolf, M. Wolf, C. Elwell, and I. Tachtsidis, "Modelling confounding effects from extracerebral contamination and systemic factors on functional near-infrared spectroscopy," *Neuroimage*, vol. 143, pp. 91–105, 2016.

[30] X. Cui, S. Bray, and A. L. Reiss, "Functional near infrared spectroscopy (NIRS) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics," *Neuroimage*, vol. 49, no. 4, pp. 3039–3046, 2010.

[31] X. Cui, S. Bray, D. M. Bryant, G. H. Glover, and A. L. Reiss, "A quantitative comparison of NIRS and fMRI across multiple cognitive tasks," *Neuroimage*, vol. 54, no. 4, pp. 2808–2821, 2011.

[32] M. L. Schroeter *et al.*, "Towards a standard analysis for functional near-infrared imaging," *Neuroimage*, vol. 21, no. 1, pp. 283–290, 2004.

[33] H. Tsunashima and K. Yanagisawa, "Measurement of brain function of car driver using functional near-infrared spectroscopy (fNIRS)," *Comput. Intell. Neurosci.*, vol. 2009, 2009.

[34] E. Frank, M. A. Hall, and I. H. Witten, "The WEKA Workbench Data Mining: Practical Machine Learning Tools and Techniques," *Morgan Kaufmann, Fourth Ed.*, p. 128, 2016.

[35] I. Kononenko, E. Šimec, and M. Robnik-Šikonja, "Overcoming the myopia of inductive learning algorithms with RELIEFF," *Appl. Intell.*, vol. 7, no. 1, pp. 39–55, 1997.

[36] N. Naseer and K.-S. Hong, "fNIRS-based brain-computer interfaces: a review," *Front. Hum. Neurosci.*, vol. 9, no. January, pp. 1–15, 2015.

[37] L. R. Krol and T. O. Zander, "Cognitive and Affective Probing for Neuroergonomics Cognitive and Affective Probing for Neuroergonomics," no. June, 2018.

[38] V. Kaiser *et al.*, "Cortical effects of user training in a motor imagery based brain-computer interface measured by fNIRS and EEG," *Neuroimage*, vol. 85, pp. 432–444, 2014.