Journal on Multimodal User Interfaces
https://doi.org/10.1007/s12193-020-00325-z

ORIGINAL PAPER

fNIRS-based classification of mind-wandering with personalized window selection for multimodal learning interfaces

Ruixue Liu[1] · Erin Walker[2] · Leah Friedman[3] · Catherine M. Arrington[4] · Erin T. Solovey[1]

Received: 23 January 2020 / Accepted: 6 May 2020

Abstract

Automatic detection of an individual's mind-wandering state has implications for designing and evaluating engaging and effective learning interfaces. While it is difficult to differentiate whether an individual is mind-wandering or focusing on the task only based on externally observable behavior, brain-based sensing offers unique insights to internal states. To explore the feasibility, we conducted a study using functional near-infrared spectroscopy (fNIRS) and investigated machine learning classifiers to detect mind-wandering episodes based on fNIRS data, both on an individual level and a group level, specifically focusing on automated window selection to improve classification results. For individual-level classification, by using a moving window method combined with a linear discriminant classifier, we found the best windows for classification and achieved a mean F1-score of 74.8%. For group-level classification, we proposed an individual-based time window selection (ITWS) algorithm to incorporate individual differences in window selection. The algorithm first finds the best window for each individual by using embedded individual-level classifiers and then uses these windows from all participants to build the final classifier. The performance of the ITWS algorithm is evaluated when used with eXtreme gradient boosting, convolutional neural networks, and deep neural networks. Our results show that the proposed algorithm achieved significant improvement compared to the previous state of the art in terms of brain-based classification of mind-wandering, with an average F1-score of 73.2%. This builds a foundation for mind-wandering detection for both the evaluation of multimodal learning interfaces and for future attention-aware systems.

Keywords Mind-wandering · Functional near-infrared spectroscopy · fNIRS · Attention-aware systems

1 Introduction

Mind-wandering occurs when an individual is engaging in internal non-task thoughts, instead of processing external task-related information [20]. Even though people may be generally unaware of when it occurs, mind-wandering could occupy 46.9% of daily life [36]. While some studies suggest that mind-wandering may contribute to future planning and creative problem solving, mind-wandering has shown to be disruptive and detrimental to individuals' performance when it happens during cognitively demanding tasks [56]. Therefore, the detection of mind-wandering states is important for many domains, and particularly for learning and training. For example, when a student is engaging in cognitively demanding tasks such as learning, mind-wandering would negatively affect task performance and lead to errors [43].

While technology-enhanced learning such as intelligent tutoring interfaces, serious games and Virtual Reality (VR) environments show promise for enhancing learning and training experiences, research shows not all interface features or virtual environments elements increase the effectiveness [47,66]. As such, identification of a user's mind-wandering episodes and on-task episodes in a learning interface could inform evaluations. Further, detecting mind-wandering states is an important step towards attention-aware systems, which can dynamically update interfaces and content to facilitate user focus on task-related information. For example, when

This work was supported in part by the U.S. National Science Foundation under Grants NCS-1835307, CNS-1711773 and NCS-1835251.

✉ Erin T. Solovey
esolovey@wpi.edu

1 Worcester Polytechnic Institute, Worcester, MA 01609, USA

2 University of Pittsburgh, Pittsburgh, PA 15260, USA

3 Drexel University, Philadelphia, PA 19104, USA

4 Lehigh University, Bethlehem, PA 18015, USA

Published online: 02 June 2020

the system detects an individual is mind-wandering during training sessions, it could change the presentation to help the user focus on important materials [54].

To measure mind-wandering, many researchers use an experience sampling methodology. With this method, researchers ask individuals to self-report when mind-wandering occurs during a task or place thought probes during the task, which periodically ask individuals whether they are mind-wandering. However, these methods have a limitation due to their dependence on participants to be aware of their mind-wandering episodes and respond accurately. Also, the thought probes interrupt both the task and the mind-wandering episodes [43].

One possible solution to address these limitations is to examine an individual's brain activity directly and use the brain data to disentangle focused states from mind-wandering states. Functional magnetic resonance imaging (fMRI) studies show that mind-wandering is associated with activation in the default network [20]. Several default mode network areas have shown consistent activation during mind-wandering, including medial prefrontal cortex, medial temporal lobe, posterior cingulate cortex, and bilateral inferior parietal lobule [14]. Moreover, as non-invasive neuroimaging techniques become less expensive and more portable, we can monitor brain activity during various activities.

Recently, the use of functional near-infrared spectroscopy (fNIRS) has received focus because of its promise for detecting an individual user's cognitive state in more ecologically valid studies. While fMRI has become the gold standard for brain imaging, in real-world environments, fNIRS is a more convenient and more affordable technology than fMRI [17]. fNIRS emits near-infrared light into the brain, and the light returned to the surface is measured and used to calculate oxygenation in the blood. This calculation reflects brain activity in that particular area. Prior work has shown the potential of using fNIRS data to identify brain activation related to mind wandering episodes [19].

In this paper, we aim to build on previous findings and present a data-driven classification framework to improve mind-wandering classification accuracy. Since prior fNIRS studies have shown that the classifier performance can be improved by focusing on a specific window [35,44,58], we utilize a moving window method for the classification of mind-wandering, which can select the best window for classification during a time period. In addition to building models for each individual, we also demonstrate the feasibility of building machine learning models across individuals to differentiate mind-wandering episodes versus on-task episodes.

For individual-level classification, we use the moving window method combined with a shrinkage LDA classifier to find the best window for detecting mind-wandering. For group-level classification, to incorporate individual differences in window selection and hence improve the classification results, we propose a novel individual-based time window selection (ITWS) algorithm. The ITWS algorithm iteratively chooses the best window for each individual through embedded individual-level classifiers, and then uses data from these windows as training data and test data for the group-level classifier. We validate the framework using an fNIRS dataset we collected with mind-wandering episodes and on-task episodes during the Sustained Attention to Response Task (SART). The errors during the SART have been shown to be correlated with mind-wandering [43], and thus form a ground truth for our classification results.

The main contributions of this paper are as follows:

- We propose to use fNIRS brain data for evaluating learning interfaces and for attention-aware systems that can automatically detect mind-wandering state without interrupting the task with experience sampling probes.
- We describe a study in which we collected fNIRS brain data during the SART task. This dataset provides examples of mind-wandering and on-task episodes, defined based on behavioral data, that can be used to investigate robust classification algorithms. We confirm that there are differences in frontal lobe blood oxygenation patterns between mind-wandering episodes and on-task episodes.
- To improve classification accuracy, we investigate window selection when classifying mind-wandering states versus on-task state using fNIRS. We show individual-level classifiers can achieve better classification results when focusing on specific windows rather than those using the entire episodes.
- To further improve model robustness and performance, we extend the window selection method for group-level classification. We propose a novel individual-based time window selection (ITWS) algorithm to incorporate individual differences in window selection when building group-level classifiers. We show that the ITWS algorithm can improve the group-level classification result by comparing with other methods that do not use the ITWS algorithm.

2 Background

2.1 fNIRS-based brain–computer interfaces

Brain–Computer Interfaces (BCIs) have shown promise for improving interactive experiences by offering unique insights into users' underlying cognitive processes. As such, brain-sensing techniques can be used not only as the primary control module for interfaces but can also be employed as a complementary input to traditional control modules for multimodal interfaces.

fNIRS is a brain-imaging tool that is safe, portable, easy to use, and quick to set up characteristics that have led to increasing adoption. It detects hemodynamic changes associated with neural activity in the brain while participants perform cognitive or behavioral tasks [10]. Because fNIRS enables brain activity to be measured continuously during interactive tasks, it has promise for understanding user experience in realistic settings. The fNIRS sensors use light to detect hemodynamic changes. The light sources send two wavelengths of near-infrared light into the forehead, where it continues through the skin and bone 1–3 cm deep into the cortex. Biological tissues are relatively transparent to these wavelengths. Oxygenated and deoxygenated hemoglobin are the main absorbers of this light. After the light scatters in the brain, some reaches the light detector. By determining the amount of light picked up by the detector, the amount of oxygenated and deoxygenated hemoglobin can be calculated in the area, which indicates hemodynamic activity associated with brain activation. Thus, fNIRS measurements can be used to understand changes in a person's cognitive states while performing tasks [50,63].

There are other techniques that can measure the changing state of the brain (e.g., fMRI, EEG, positron emission tomography (PET), and magnetoencephalography (MEG)). These tools are often prohibitively expensive and require restrictions on the study participant that are not reasonable for use in realistic settings. Also, PET requires ingestion of hazardous material, and fMRI exposes individuals to loud noises that may interfere with the study [30]. The strong magnetic field prevents typical computer usage in both fMRI and MEG. EEG is less intrusive, more portable, and less expensive than these other tools. It has been widely used in brain–computer interfaces research, such as interface evaluation and adaptation [13,34,48]. However, it can have a significant setup time and has limited spatial resolution. Electronic devices in the room can also interfere with the signal, and it is susceptible to artifacts in the data due to user movement.

fNIRS avoids many of the restrictions of other techniques and therefore has promise for use in real-world settings and virtual environments. It has been shown to be robust in typical human-computer interaction scenarios, including during typing and mouse clicking [63], and verbalization [52]. Real-time fNIRS brain data has been used to make appropriate adaptations to user interface elements [2] as well as to modulate interruptions [55] and enable attention-aware systems [51]. fNIRS hyperscanning has also shown promises to monitor multiple participants' brain activation simultaneously during their natural interactions [39]. Significant improvements have been made recently in terms of fNIRS hardware to make it wearable and wireless, and we foresee it being increasingly integrated with wearable computing platforms currently being developed [40,41].

2.2 Detection of mind-wandering in multimodal learning interfaces

Technology-enhanced learning is increasingly adopted for providing novel solutions in educational and training activities, such as intelligent tutoring interfaces, serious games, and VR environments. Previous studies have shown the positive effects of such applications in improving students' cognitive states during learning, including motivation and attention [16,18]. Mind-wandering has also shown to play an important role in students' learning performance. Mind-wandering can be detrimental to student learning, where instead of processing external task-related information, students engage in internal non-task thoughts [61]. Therefore, detection of mind-wandering would be valuable for understanding users' attention control mechanisms during these interfaces. Nevertheless, since mind-wandering involves internal thoughts instead of expressive behaviors and the dynamics of mind-wandering remain elusive, detecting mind-wandering is a challenging task [32]. Prior research has investigated using physiological and behavioral metrics, as well as brain data for mind-wandering detection.

2.2.1 Physiological and behavioral metrics of mind-wandering

Probe-caught mind-wandering has been predicted using eye gaze [5,29], physiological sensing [6,9], behavioral indices [21,42], and facial expression [7]. Hutt et al. used eye gaze and contextual cues as features to predict mind-wandering state when participants were interacting with an intelligent tutoring system. Participants were randomly probed to report mind-wandering instances. They achieved a prediction accuracy of about 25% above chance [29]. Physiological features, including heart rate [9] and skin conductance [6], have also been used for mind-wandering detection. Blanchard et al. measured participants' skin conductance and skin temperature to detect mind-wandering during a reading task. They achieved 22% above chance accuracy [6].

Some researchers also used behavioral indices, including reading behaviors and textual features, to detect mind-wandering during reading tasks [21,42]. The resulting accuracy is 20% above chance for a somewhat naturalistic reading paradigm [42]. However, this method is limited to reading tasks.

Another approach is using facial expressions and movements to detect mind-wandering state. Bosch and D'Mello applied this approach in a laboratory study where participants read a text and in a classroom study where high school students learned biology from an intelligent tutoring system. After extracting facial and movement features from the recorded video and applying machine learning classifiers, they achieved 25.4% and 20.9% above-chance accuracy for detecting mind-wandering in the lab and classroom, respectively [7].

For all of these investigations, the models built are ad-hoc and depend on a set of measurable factors that have shown to be related to mind-wandering in the specific task.

2.2.2 Brain-based metrics of mind-wandering

Brain sensing techniques provide an alternative to detect mind-wandering objectively across different domains. Some researchers explored using EEG brain signals to differentiate mind-wandering versus on-task. Kawashima et al. used EEG variables to estimate mind-wandering intensity through support vector machine regression during a sustained attention task [33]. However, the mind-wandering intensity was determined by thought probes, which were placed at a fixed interval. This could lead to individuals anticipating the probe occurrence and becoming more conscious of mind-wandering. Jin et al. trained machine-learning models on EEG markers to determine participants' state as either mind wandering or on-task, and they achieved a mean accuracy of 64% for a sustained attention task [32].

2.2.3 Considerations

In all these studies, probes were used to catch mind wandering by asking participants whether they are mind-wandering. Researchers then focused on an interval of time that precedes the probes (10 s or 30 s). These probes allow researchers to mark the time point when mind wandering is actually happening. However, it interrupts the mind wandering episodes, and can only collect the mind wandering episodes that participants are aware of. Therefore, exploring the detection of mind-wandering episodes without interruption would be an important step toward fully automated attention-aware systems and environments. In this work, we explore fNIRS brain measures of mind wandering and use the SART task to elicit mind-wandering episodes, since the errors during the SART task have been shown to be correlated with mind-wandering [43].

2.3 Mind-wandering classification with fNIRS

2.3.1 Accuracy of mind-wandering detection with fNIRS

As mentioned earlier, activation of the medial prefrontal cortex during mind-wandering has been detected using fNIRS during a sustained attention task [19]. This study showed promise for detecting default network activations related to mind-wandering from fNIRS data. However, this work also highlighted the difficulty of real-time detection of mind-wandering using only fNIRS data. Their machine learning model achieved a mean accuracy of 56% for classifying mind-wandering episodes versus on-task episodes using Linear Discriminant Analysis for each individual separately. For real-world use, this accuracy would need improvement. Therefore, there is a need to explore methods that can achieve higher accuracy. Two approaches that may hold promise are 1) exploring automatic detection of optimal time windows, and 2) exploring both individual and group models.

2.3.2 Optimal time windows for classification

While many studies build and evaluate machine learning classifiers using fNIRS data associated with the entire episodes (e.g., entire mind-wandering episode or on-task episode [19]), other fNIRS studies have shown that we can improve the classifier performance by focusing on a specific window, instead of using the fNIRS data from the overall task period [35,44,58]. Naseer et al. used fNIRS data to classify right- and left-wrist motor imagery task and they analyzed six different temporal windows within an overall 10 s task. They showed that the 2–7 s period after the stimulus was the most critical period and they could enhance the average classification accuracy by around 4% by focusing on this period [44]. Khan et al. used linear discriminant analysis to find the best window size for detecting drowsiness using fNIRS [35]. They analyzed three different time windows (0–3 s, 0–4 s and 0–5 s), and proposed drowsiness detection in 0–4 s window when using fNIRS. These approaches compare a few predefined windows and select the one with the best outcomes.

Researchers have also used the moving window method to explore all windows with a specific size and find the best window for classification using fNIRS data [8,58]. For example, Shin et al. conducted two fNIRS experiments (left vs. right-hand motor imagery; mental arithmetic vs. resting state), and used a 3 s moving window with 1 s step size to find the maximum classification accuracy over time. The classification accuracy achieved by the best window is significantly higher than those for the other windows [58]. Hennrich et al. adopted an n-back task with fNIRS to induce different levels of workload and extracted 10 s windows for workload classification. Their results show that classification accuracy differs between different windows, and peak around 10 s after the trial start [24].

From all these studies, the results show that the optimal windows vary between different participants and different tasks. Moreover, both the window sizes and the offset from start time can affect the accuracy of the classification results. Therefore, in this work, we use the moving window method along with different window sizes to find the best window for mind-wandering classification.

2.3.3 Individual versus group models

In prior work, machine learning models were built for each participant separately (individual-level models) [19]. Considering the small dataset of each participant and the high feature space of the brain data, building models per-participant could lead to overfitting and achieving overly-optimistic results [15]. As such, researchers have shown the need for building models across participants [4].

Building models across participants (i.e., group-level models) can enable researchers to get an adequate amount of data for model training while reducing the time for collecting brain data. Compared to individual-level models, the group-level models are more robust and can achieve more reliable results. However, due to the individual differences in hemodynamic responses, it is a challenge to build robust models across participants based on fNIRS data.

To solve this issue, prior work has investigated optimal feature combinations for each participant [27,49]. For example, Noori et al. used the hybrid genetic algorithm to choose the optimal feature for each participant [49]. Hossein et al. applied a personalized feature normalization approach to standardize the extracted feature values of each participant to improve the performance of group-level models. However, even though prior work shows that the optimal windows vary between different participants, little attention has been paid to the effect of individual differences in window selection on the performance of group-level models. In this paper, we investigate possible methods for incorporating individual differences in window selection for group-level modeling.

3 Toward attention-aware multimodal interfaces with fNIRS

We propose to use fNIRS data for attention-aware interfaces that can automatically detect mind-wandering states without interrupting the task with experience sampling probes. This could be valuable for designing and evaluating learning interfaces, as well as developing adaptive learning systems. For example, identifying mind wandering during an evaluation of a novel educational game could lead to adjustments in the design to keep the users engaged. Alternatively, when a system automatically detects an individual is mind-wandering during online learning, it could change the presentation to help the user focus on important tasks and materials.

To build multimodal interfaces with fNIRS, Solovey et al. pointed out that there are some common high-level phases [64], with calibration phase, modeling phase, and real-time classification phase being the main phases for real-time applications [64]. During the calibration phase, users are asked to perform a set of cognitive benchmark tasks. The cognitive benchmark tasks are experiment tasks from cognitive psychology that can elicit different targeted cognitive states [25]. fNIRS data recorded during the cognitive benchmark tasks is then used to train machine learning classifiers. In the real-time classification phase, the machine learning model continuously classifies the new data coming in. Classification results can then be used for evaluating the interfaces or sent to the system for necessary adaptations. We describe these phases for developing attention-aware multimodal learning interfaces with fNIRS in Fig. 1.

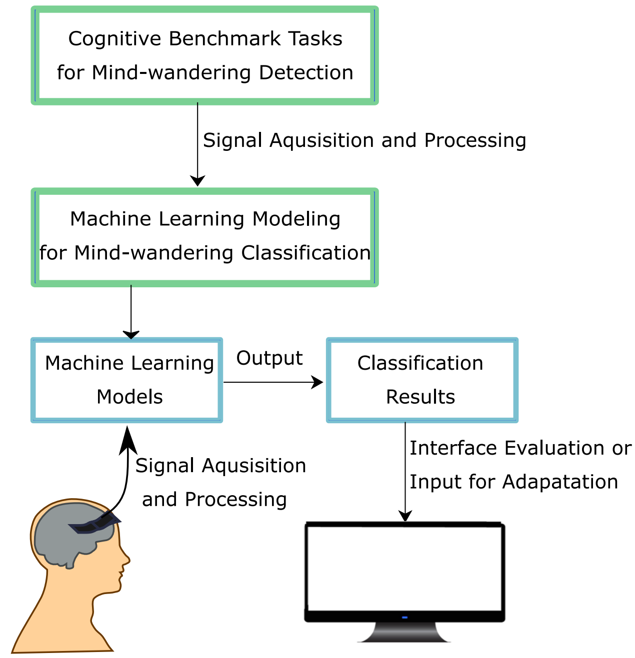

Fig. 1 The workflow of developing attention-aware multimodal interfaces using fNIRS

To move toward this goal, mind-wandering classification accuracy needs to be higher than shown in previous work [19]. To do this, appropriate datasets need to be created for validating algorithms for mind-wandering classification. Further, research needs to explore the impact of individual and group models and appropriate time windows to demonstrate the potential of this approach.

4 Data collection

We set out to build a dataset of fNIRS data associated with mind-wandering episodes without using experience sampling probes and to investigate methods of distinguishing mind-wandering states from on-task states with high accuracy. To do this, we conducted a human-subjects study that was approved by our institutional review board, and informed consent was obtained for all participants.

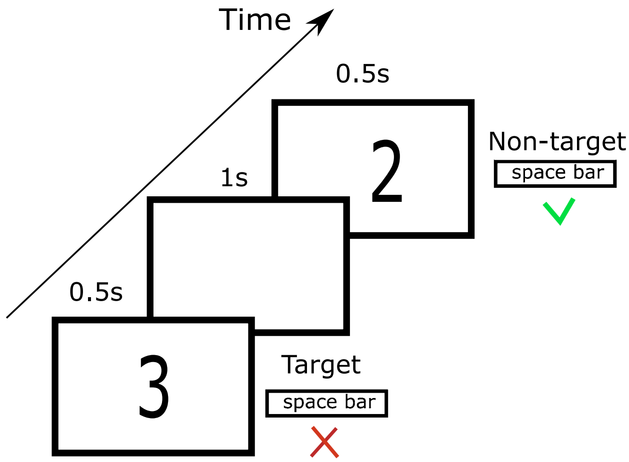

Fig. 2 Time course of the SART protocol. The number was shown on a white screen for 0.5 s, followed by a blank screen for 1.0 s. Participants were asked not to press the space bar for the target number 3 and press the space bar for any other numbers

4.1 Sustained attention to response task

To elicit mind wandering, we used a well-studied paradigm called the Sustained Attention to Response Task (SART) [38]. The SART shows a number (0–9) at the center of a blank white screen for 0.5 s, followed by a blank screen for 1.0 s. Participants were instructed to respond by pressing a key for each stimulus that appears except for the target stimulus, the number 3. When a '3' is shown, the participant is instructed not to press any key and to wait for the next number. For typical SART tasks, the target stimuli occurs around 5–11% of all stimuli [31,38]. Since prior work has shown that a low proportion of target stimuli allows increased mind-wandering during the task [38], we adopted a frequency of 5% for the target stimuli to elicit mind-wandering state from the participants. Also, following previous work [38], targets are presented pseudorandomly throughout all trials and are arranged to ensure that they did not appear immediately next to each other.

Figure 2 shows the time course of the SART protocol. An incorrect keypress for the target stimulus has been associated with mind-wandering, while a correct response indicates on-task behavior [62].

4.2 Procedure

Participants were given an overview and instructions for the task and informed about the brain sensing equipment that would be worn during the study. After providing informed consent, each participant was given instructions about the SART task and the opportunity to ask questions. Participants were equipped with the fNIRS sensors on their forehead. Then participants performed the SART task on a computer.

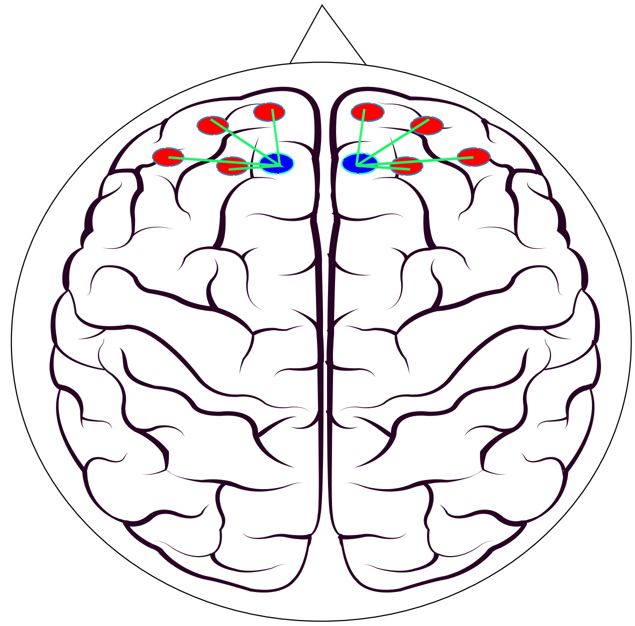

Fig. 3 Placement of fNIRS sources (red circles) and detectors (blue circles). Green solid line indicates fNIRS channels

The experiment consists of 6 sections, with 10 targets and 190 non-targets. In between sections, there was a 10-s break.

At the end of the experiment, individuals were given a questionnaire where they were asked how focused they were throughout the task (scale of 1–7), and if they experienced unrelated thought or drowsiness (from 'never', 'rarely', 'occasionally', 'sometimes', 'frequently' to 'very frequently', later converted to a 6-point scale).

4.3 fNIRS recording

The fNIRS data was acquired using a multichannel frequency domain Imagent from ISS Inc. (Champaign, IL). Two probes were placed on the forehead to measure the two hemispheres of the anterior prefrontal cortex (Fig. 3). The source-detector distances were of 0.8 cm or 3 cm. Each light source emits two light wavelengths (690 nm and 830 nm) to detect and differentiate between oxygenated and deoxygenated hemoglobin. The sampling rate was 6.649 Hz. The sensors were kept in place using headbands, which can also reduce light interference.

4.4 Participants

The study included 11 healthy volunteers (5 males) between the ages of 18 and 41 (average 26.27).

5 Dataset curation

Based on the fNIRS data collected during the experiment, we built the dataset for investigating the classification of *on-task* and *mind-wandering* states. We also analyze participants' performance on the task and compare the results with prior work.

5.1 General dataset description

The dataset consists of fNIRS data of 6 channels, from 11 participants. Since the two short-separation channels (0.8 cm) contain mostly noise, we only analyze fNIRS signals from the six long-separation channels (Fig. 3).

5.2 Dataset preprocessing

The fNIRS signals from the device may contain noise from various sources, including instrumental noise, motion artifact and physiological noise [45]. Following typical preprocessing techniques [53], we used a band-pass filter with a high pass value of 0.02 and a low pass value of 0.5 to remove the physiological noise (e.g., heart rate, respiration) and the instrumental noise. The motion artifacts were removed using a wavelet-based de-noising and correction procedure [45]. Raw light intensity data was then converted to oxygenated and deoxygenated hemoglobin values using the Modified Beer-Lambert Law. All preprocessing was completed in MATLAB using HomER [28].

5.3 Dataset labeling

To prepare the datasets for analysis and classification, following the work of Durantin et al. [19], for each target episode, we extracted fNIRS data from 30 s before the target and 10 s after the target. Target episodes with a correct response were labeled as *on-task* episodes, while target episodes with an incorrect response were labeled as *mind-wandering* episodes. All non-target episodes were ignored for this analysis since they were not indicative of our target classes. Figure 4 shows the number of mind-wandering episodes and the number of on-task episodes from each participant's dataset. Due to the nature of the task, the number of on-task and mind-wandering episodes varied across participants. For each participant, the number of mind-wandering episodes ranged from 8 to 33 out of 60 total targets and the number of on-task episodes varied from 27 to 52 out of 60 total targets (Fig. 4). Across all participants, the dataset contains 239 mind-wandering episodes and 421 on-task episodes in total.

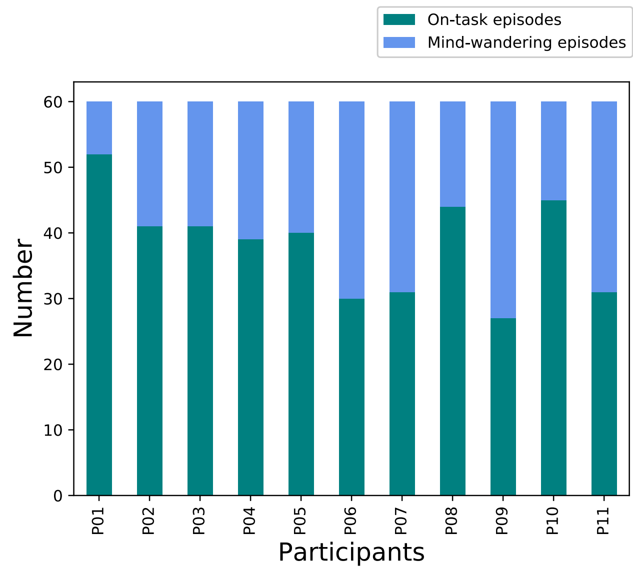

Fig. 4 The number of mind-wandering episodes and the number of on-task episodes from each participant. Each episode consists of 30 s before the target and 10 s after the target

5.4 Behavioral data

For the 60 target episodes of the experiment, the mean accuracy across all participants was 0.63 (SE: 0.044) (Fig. 4). For the non-target episodes, the mean accuracy across all participants was 0.98 (SE: 0.002). These are not used for our classifier. Participants made significantly more errors on the target episodes than on the non-target episodes (the Wilcoxon signed-rank test, $p < 0.05$), which is consistent with prior work [43]. For the post-survey, the mean level of focus participant reported was 4.45 (SE: 0.35, scale of 1–7), and the mean frequency of unrelated thoughts and drowsiness was 4.18 (SE: 49) and 4.18 (SE: 0.45), respectively (converted to scale of 1–6 from 'never' being 1 to 'very frequently' being 6). This shows that participants experienced mind-wandering states during the study.

5.5 Dataset overview

For the overview of the dataset, we calculated the folded average of oxygenated hemoglobin (HbO) and the deoxygenated hemoglobin (HbR) change across all participants for the on-task (correct) and mind-wandering (incorrect) target responses. Specifically, we calculated the folded average of all long-separation channels on the left side of the head and all long separation channels on the right side of the head separately. From Fig. 5, we can see the average change in HbO from the right side of the cortex showed a significant increase during 30–15 s prior to an incorrect response to the target, followed by a decrease before the target. From the left side of the cortex, there was a slight increase in HbO

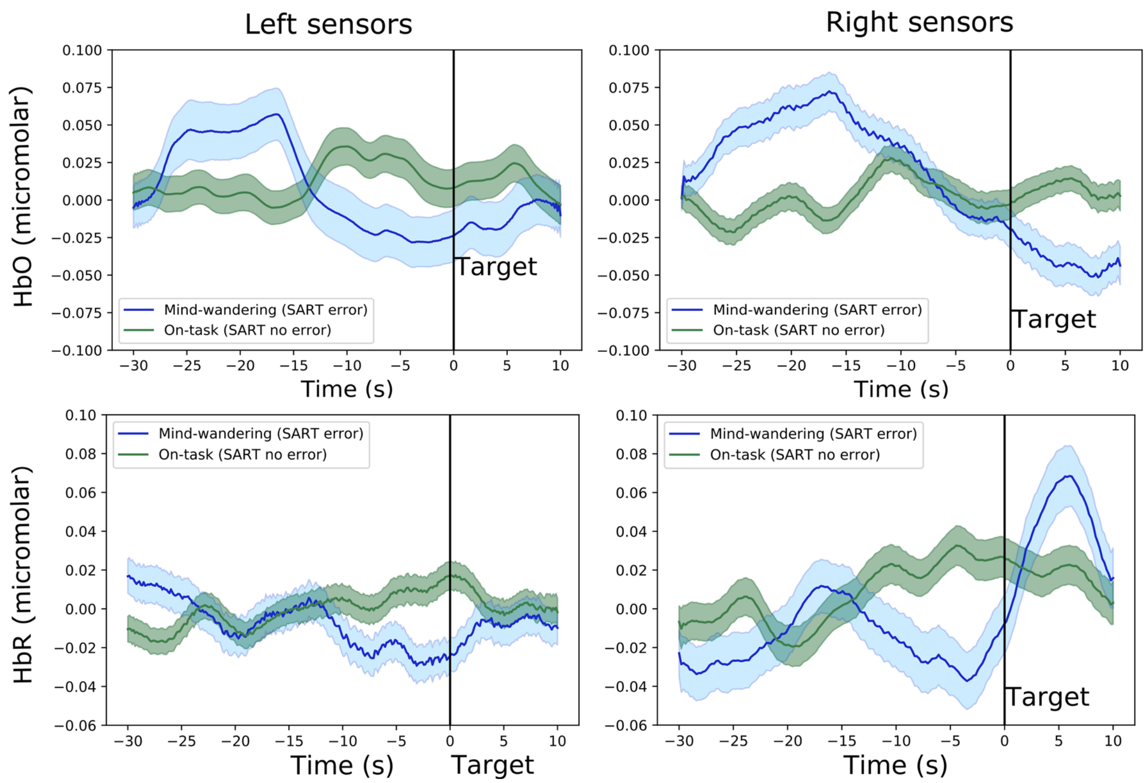

Fig. 5 Variation of the oxygenated hemoglobin (HbO) and deoxygenated hemoglobin (HbR) concentration for the mind-wandering episodes (SART error, in blue) and on-task episodes (SART no error, in green). The figures show the mean (averaged across individuals) and standard error over the 40 s. The figures on the left are data from the sensor on the left side of the head, and the figures on the right are data from the right side of the head. Shaded areas represent the standard error of the mean for each condition

change around 25–15 s before a target error and then return to normal. The average change in HbR on the left side of the cortex showed a slight decrease around 10 s before an incorrect response to the target. The average change in HbR on the right side of the cortex showed a decrease around 15 s before an incorrect response to the target and followed by an increase immediately before the target.

Consistent with prior findings [14,19], our results suggest there are differences in frontal lobe blood oxygenation patterns between mind-wandering episodes and on-task episodes. Also, our results indicate activation in the prefrontal area preceding mind-wandering occurrence, as the level of HbO increases on both sides of the prefrontal cortex before SART errors. This is consistent with the findings of previous investigations, which suggest that the prefrontal area contributes to the switching from an on-task state to mind-wandering [14,19]. In contrast with the previous findings of Durantin et al. [19], where they found no significant variations on the HbR relative to incorrect responses to the target, our results showed a decrease on both sides of the cortex before incorrect responses to the target. Since both a decrease in HbR and an increase of HbO indicate cerebral activation, our results are consistent and suggest activation at the prefrontal area at the beginning of mind-wandering episodes.

Moreover, from Fig. 5, we can see that the time series behaviors of the hemodynamic patterns are different in different windows during the mind-wandering episodes. In the next section, we investigate window selection for detecting mind-wandering and develop a data-driven classification framework.

6 Data-driven classification framework

Using the fNIRS dataset that we built and validated above with mind-wandering episodes and on-task episodes, we develop a data-driven classification framework for detecting mind-wandering.

In this section, we investigate window selection when classifying mind-wandering episodes versus on-task episodes using fNIRS data, with the goal of improving the classification accuracy. In addition to individual-level classification, we also explore the feasibility of building machine learning models across participants for detecting mind-wandering. We evaluate the window selection method by comparing the

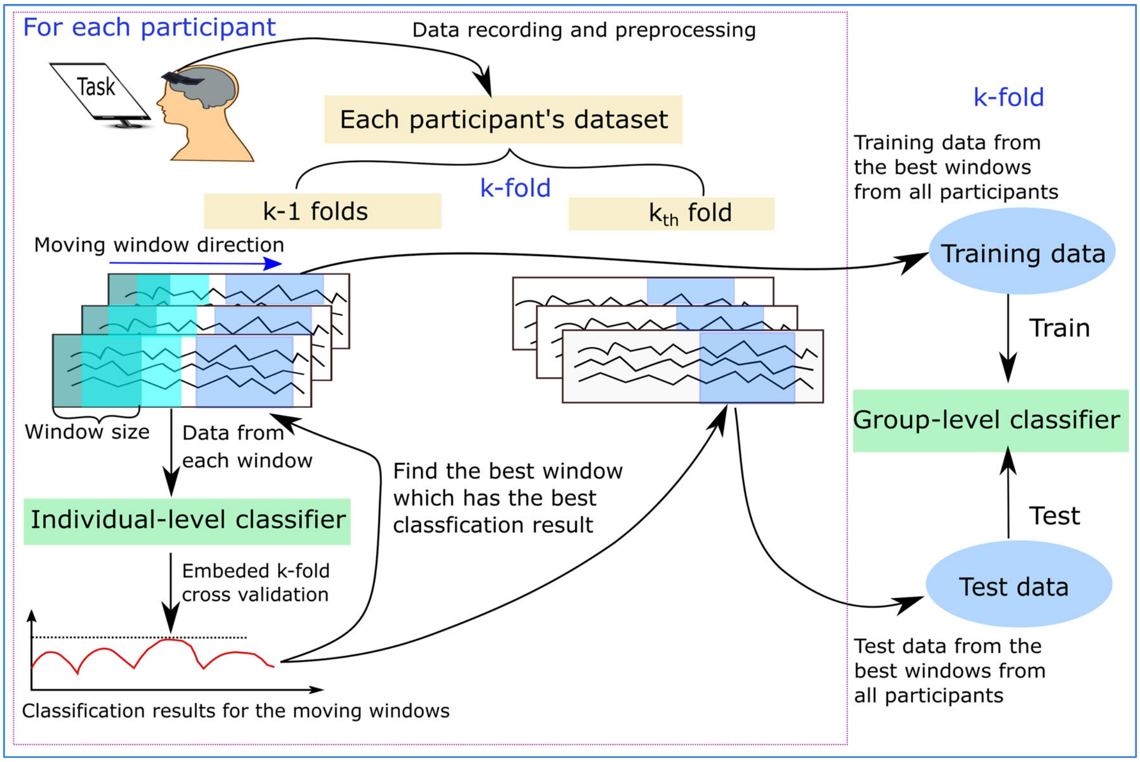

Fig. 6 Structure of the ITWS algorithm. The dataset (episodes of 40 s) of each participant is divided into k folds. In each fold, $k-1$ folds of the dataset are used to find the best window for classification, and the data from this window of these $k-1$ folds are later used as training data for the group-level classifier. The data from the same window of the remaining fold are used as the test data to evaluate the classifier. For each participant, a moving window method combined with an individual-level classifier was used to obtain the best window, which has the best cross-validation results

results with the same classifiers, but without window selection.

6.1 Individual-level classification

We start with building models for each participant to classify mind-wandering episodes versus on-task episodes. The goal is to determine if the window selection method can improve the individual-level classification accuracy.

6.1.1 Moving window method

We use the moving window method to find the best sub-window. The moving window method iterates through all the windows with a specific size during a period, and then all data processing is performed separately on each time window, i.e., feature extraction and classification. Therefore, the moving window method requires a predefined window size and a step size. To investigate the effect of the window size on classification results, we use three different window sizes that are commonly used in previous fNIRS studies [24,44,59], which are 5 s, 10 s, and 15 s. For each of the window sizes, we use a 1 s step size [59]. The best sub-window is defined as the window with the best classification result.

6.1.2 Individual-level classifier

Because of its simplicity and low computational requirements, linear discriminant analysis with shrinkage (shrinkage LDA) is commonly used as the classifier in fNIRS studies [46]. Particularly, shrinkage LDA has shown advantages when dealing with datasets with a small sample size and a large number of features [57]. LDA uses discriminant hyperplane(s) to separate data from different classes [3]. It assumes the class covariance are identical and then models the class conditional distribution of the data for each class. However, with a small sample size, the number of features of each sample could exceed the number of samples in each class. In this case, the empirical sample covariance is a poor estimator [37]. Using a shrinkage estimator of the covariance matrix can help solve this issue [1]. In this study, considering the small sample size for each participant, we use the shrinkage LDA as the individual-level classifier.

From Fig. 4, we can see that for most participants (ten out of eleven participants), the dataset is not balanced between the two classes. Most participants had fewer mind-wandering episodes compared to on-task episodes. To avoid the bias of training the classifier towards one class, we use the synthetic minority oversampling technique (SMOTE) to balance the training data for each participant. SMOTE is an oversampling method that has shown effectiveness in many imbalanced datasets. It can generate new synthetic examples by finding the nearest neighbors of the examples from the minority class [11]. This oversampling method is used only during the training process.

All long-separation channels are used to build the classifier for each participant. The average values of HbR and HbO and the slope over the moving time window are used as features [49]. Each feature is normalized.

Then, in combination with the moving window method, we apply shrinkage LDA to build individual-level classifiers and find the best window for classification.

6.2 Group-level classification

Individual-level classifiers often rely on a relatively small dataset with high feature space, which could lead to model overfitting. Group-level models can solve this issue by training on data collected from all participants. However, it is difficult to achieve high accuracy on group-level models due to individual differences. We propose an individual-based time window selection (ITWS) algorithm to improve the group-level classification results.

6.2.1 ITWS algorithm

When using the window selection method, the best windows could vary between different participants. If we use the standard moving window method to build group-level classifiers across individuals, then data from the same windows from all individuals will be used to build and evaluate the classifier. However, since the best window could vary for different participants, using the standard moving window method could lead to suboptimal classification results.

We propose a novel individual-based time window selection (ITWS) algorithm to select the best window for each individual when building the group-level classifier. Figure 6 shows the structure of the algorithm. The main principle of the ITWS algorithm is to use an embedded individual-level classifier to determine the best window for each participant. The embedded individual-level classifiers are used in combination with the moving window method and are applied on the entire episode. Data from each participant are first separated into two blocks (training data and test data for the group-level classifier). Then, the embedded individual level classifiers are trained and evaluated only on one block (embedded k-fold cross-validation).

Algorithm 1 Individual-based time window selection (ITWS) algorithm

1: **Initialize** Divide the dataset of each participant into k folds. k-1 folds of the dataset are used as the training data and to obtain the best window for this specific participant, and the remaining one fold is used as the test data. Set the group-level training data and group-level test data to empty
2: **for** current k **in** k-fold cross-validation **do**
3:    **for** participant **in** all participants **do**
4:       Generate all moving windows by sliding the window on the data from the k-1 folds
5:       **for** window **in** all moving windows **do**
6:          Use embedded subject-level classifier to obtain the classification score on this window (embedded k-fold cross-validation)
7:       **end for**
8:       Select the best window by finding the maximum cross-validation score from all moving windows
9:       Add the data from the best window from the k-1 folds to the group-level training data
10:       Add the test data from the best window from the remaining fold to the group-level test data
11:    **end for**
12:    Train the group-level classifier on the group-level training data
13:    Obtain the test results by applying the group-level classifier on the group-level test data
14: **end for**
15: Calculate the average test result after k-fold cross-validation.

To effectively assess the performance of the machine learning models, the algorithm can be used together with k-fold cross-validation for the group-level classifier. The flow of the ITWS algorithm is described in Algorithm 1. Specifically, for k-fold cross-validation, we first divide the dataset (episodes of 40 s) from each individual into k folds. Then, during each fold, $k-1$ folds of the dataset are used to find the best window for classification. The data from this window of these folds are later used as the training data for the group-level classifier. The data from the same window of the remaining fold are used as the test data to evaluate the group-level classifier. We repeat this procedure for all individuals. At each fold, training data from all individuals together are used to train the group-level classifier, and test data from all individuals are used to evaluate the classifier. The test result from all folds are then averaged to give the final mean test results.

6.2.2 Embedded individual-level classifier and group-level classifier

We use the shrinkage LDA as the embedded individual-level classifier as described in Sect. 6.1.2 (line 6 in Algorithm 1). Also, similar to individual-level classification, we examine the effect of window sizes by using 5 s, 10 s, and 15 s as the window size for the moving window method.

Comparing to individual-level classification, the group-level classification can be trained on a larger dataset from all participants. Therefore, we aim to use modern machine learning models that take advantage of the larger sample size. Modern machine learning models, including XGBoost and deep learning techniques, have achieved state-of-the-art results on many machine learning problems [12,22,65], and have shown promise for fNIRS data classification [23,26]. Therefore, to evaluate the performance of the ITWS algorithm, we use Deep Neural Networks (DNNs), Convolution Neural Networks (CNNs), and XGBoost as the group-level classifier (line 12 in Algorithm 1).

XGBoost is a gradient tree boosting system that builds trees sequentially, such that each subsequent tree learns from its predecessors to reduce the errors of the previous tree [12]. Specifically, a greedy algorithm is used in the model, which starts from a single leaf, and iteratively adds branches to the tree by evaluating every possible split loss reduction. The ensemble model gives the aggregate output from all trees. To prevent overfitting, we set the learning rate to be 0.01, the maximum depth of a tree to be 4, the number of estimators to be 200, and the subsampling ratio of training instance subsample to be 0.8.

A DNN is a layered organization of connected neurons. Between the input and output layers, there are multiple hidden layers. During each hidden layer, each neuron is associated with a weight that is used to compute the weighted input. The weighted inputs are then summed and transformed by the activation function to determine the output of the neuron. By adjusting the weights of neurons, DNNs can model complex non-linear relationships between the input and output [22]. In this work, we use a network consists of three hidden layers with rectified linear unit (ReLU) activation function. Each hidden layer has 300 units, 100 units, and 40 units, respectively. We implemented an optimizer using RMSprop with a learning rate of 0.01.

CNNs are neural networks that use convolutions over the input layer. The hidden layers of a CNN typically consist of a series of convolutional layers, ReLU layers, and pooling layers. By performing specific functions, each layer learns a useful representation from the input [65]. In this work, our CNN architecture has three convolutional layers, which consist of 32 filters of size 3×1, 64 filters of size 3×1, and 64 filters of size 5×1, respectively. Each of them is followed by a batch normalization layer and a ReLU layer. Then, a max-pooling layer and a dropout layer are utilized to prevent overfitting. Finally, a fully connected layer with 64 input neurons and two output neurons is used for the binary classification. We implemented an optimizer using SGD with a learning rate of 0.01.

Similar to individual-level classification (see Sect. 6.1.2), the samples from the two classes are first balanced using SMOTE [11]. We used the same features for the embedded individual-level classification and group-level classification, which include the average values and slope of HbR and HbO from all long-separation channels. Each feature is normalized as well. All features are then used as the input for XGBoost and DNNs. For the input of CNNs, features of each channel are concatenated into a 2D matrix (*number of channels* $\times$ *number of features*).

Then, following the ITWS algorithm, we iteratively choose the best windows from each individual. Data from these windows are then used as training data and test data for the group-level classifier.

7 Evaluation

7.1 Methodology

We evaluate the effectiveness of our window selection method by comparing the results with the same classifiers, but without window selection.

For individual-level classification, our research questions are whether focusing on a specific window will improve the classification results, and whether the window size of the moving window method can affect the classifier's performance. Therefore, we compare the classification results achieved using the moving window method with 5 s, 10 s, and 15 s as the window size, as well as with the classification results achieved using the entire episodes.

For group-level classification, our research questions are whether the ITWS algorithm can improve the group-level classification results, and whether the choices of window sizes and classifiers can affect its performance. Therefore, when XGBoost, DNNs, and CNNs are used as the group-level classifier, we compare the classification results achieved using the ITWS algorithm with classification results achieved using a standard moving window method, as well as using the entire episodes. Specifically, we compare the results when 5 s, 10 s, and 15 s are used as the window size for the moving window method.

Due to the imbalance in our dataset, the test accuracy of the classifiers could be misleading. Therefore, we report F1-scores of our classifiers. F1-scores are commonly used to account for dataset imbalances. We also use 5-folds cross-validation to assess the performance of the classifiers.

7.2 Results

7.2.1 Individual-level classification using moving window

Figure 7a shows comparative results of maximum F1-score achieved using the moving window method, with the window size of 5 s, 10 s, and 15 s, respectively, and the F1-score achieved using the whole episode. We can see that, for all participants, the maximum F1-scores achieved by using the moving window method with all three window sizes are significantly higher than the F1-score achieved when

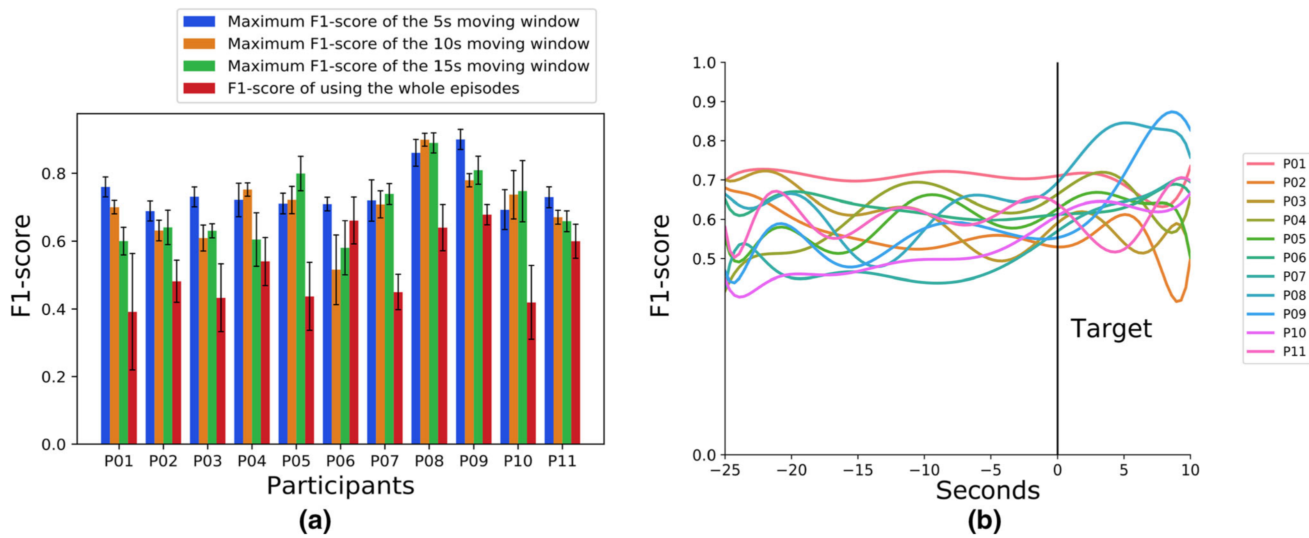

Fig. 7 **a** Comparison results of maximum F1-score achieved using the moving window method (with 5 s, 10 s, and 15 s as the window size) and the F1-score achieved using the whole episodes (5-fold cross-validation). **b** Classification results for 5 s moving windows for each individual over the 40 s time period, the x-axis indicates the right edge of the moving time window. The F1-score represents the mean F1-score of the 5-fold cross-validation on each window

using the whole episodes (Wilcoxon signed-rank test, $p < 0.05$). When using the whole episode, only four participants achieved an F1-score over 60%. The average F1-score for all participants was $52.1 \pm 3.0\%$. When using moving the window method with different window sizes, the window size of 5 s achieved the highest average value ($74.8 \pm 2.0\%$) for all participants' maximum F1-score, while the average values for all individuals' maximum F1-scores are $70.0 \pm 2.8\%$ and $70.2 \pm 3.0\%$ with window sizes of 10 s and 15 s respectively. Particularly, for each individual, six out of eleven participants achieved a maximum mean F1-score when using the window size of 5 s, while two and three participants achieved a maximum mean F1-score when using the window size of 10 s and 15 s, respectively. Furthermore, Fig. 7b shows the F1-score for the moving windows for each participant with the window size of 5 s. For each participant, we can see that the mean cross-validation F1-score varies for different windows.

These results suggest that for each participant, focusing on a specific window can achieve better classification results than using the whole episode. Also, the window size of the moving window method can slightly affect the classification results for different participants.

7.2.2 Group-level classification using the ITWS algorithm

Table 1 represents the F1-score achieved for group-level classification with different classifiers, when using the ITWS algorithm, the standard moving window method, and using the whole episodes as input respectively. We can see that the ITWS algorithm greatly improved the group-level classification results with all three different window sizes (5 s, 10 s, and 15 s), as well as with all three classifiers. Particularly, for different window sizes, applying the ITWS algorithm with the window size of 5 s achieved the highest performance. This is easy to understand since most participants achieved the best individual-level classification results when using the window size of 5 s (see Sect. 7.2.1). Specifically, when using the ITWS algorithm with a window size of 5 s and with XGBoost as the group-level classifier, we achieved the highest average F1-score of $73.2 \pm 2.0\%$, while CNNs and DNNs achieved an average F1-score of $72.8 \pm 0.13\%$ and $69.4 \pm 0.08\%$ respectively. Also, it is worth noting that even with the window size of 10 s and 15 s, for all classifiers, the ITWS algorithm achieved superior performance than the standard moving window method, as well as using the whole episodes as input. Therefore, we can conclude that the proposed ITWS algorithms can improve the classification result for detecting mind-wandering episodes across-participants using fNIRS, and is generally not affected by choice of window sizes and classifiers.

To further investigate the effectiveness of the ITWS algorithm, we analyzed the selected windows for each participant by using the ITWS algorithm with a window size of 5 s. Figure 8 shows the distribution of the right edge of selected time windows for each individual during the 5-fold cross-validation. For each individual, the box shows the quartiles with the inner line indicating the mean value. The whiskers extend to show the rest of the distribution, and the points are the outliers determined as a function of the inter-quartile range. Even though the selected best window for each indi-

Table 1 Comparative results of using the ITWS algorithm, the moving window method, and using the whole episodes for group-level classification (5-fold cross-validation)

	XGBoost	CNNs	DNNs
Whole episodes	45.4 ± 0.80	52.6 ± 0.14	48.8 ± 0.12
Moving window method (5 s window)	57.0 ± 0.31	60.5 ± 0.13	55.2 ± 0.24
Moving window method (10 s window)	52.3 ± 0.42	58.8 ± 0.22	52.5 ± 0.70
Moving window method (10 s window)	51.6 ± 0.36	59.5 ± 0.53	54.6 ± 0.68
ITWS algorithm (5 s window)	$\mathbf{73.2 \pm 0.18}$	$\mathbf{72.8 \pm 0.13}$	$\mathbf{69.4 \pm 0.08}$
ITWS algorithm (10 s window)	70.1 ± 0.10	72.4 ± 0.06	68.7 ± 0.11
ITWS algorithm (15 s window)	71.3 ± 0.21	70.7 ± 0.07	66.3 ± 0.07

The F1-score of the moving window method represents the maximum F1-score
Bold signifies the highest F1-score in comparison of the parameters

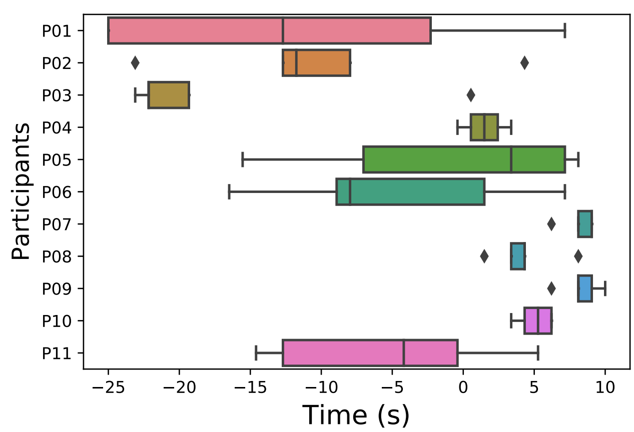

Fig. 8 The distribution of the selected best windows (the right edge) for each individual during the 5-fold cross-validation, when using a window size of 5 s. 0 s represents the timing of the targets

vidual varies during each fold, we can still see there are individual differences related to window selection. For example, the selected best windows for individual P03 concentrate around 20 s before the target, while the selected best window for individual P10 centered around 5 s after the target. Also, while some participants show a broader spread of window selection than the others, the classification results for test data from each participant did not show any differences. These findings further confirm that the proposed ITWS algorithms can incorporate individuals' differences in window selection and ensure the best window for classification for each individual is used to build the final classifier across individuals.

8 Discussion

Our study aimed to build classifiers based on fNIRS data to detect whether an individual is mind-wandering or focusing on-task. To build a dataset for exploration, we conducted a study using fNIRS during the SART task. The errors during the task are correlated with mind-wandering [43]. Consistent with previous findings, we showed individuals made a higher number of errors for the target than non-target trials. We also showed activation in the prefrontal cortex during mind-wandering episodes, as the changes of HbO increase and the changes of HbR decrease before the targets with incorrect responses. All individuals retrospectively reported mind-wandering during the task in the post-survey.

For classification, we labeled the target episodes (30 s before the target and 10 s after the target) with a correct response as the on-task episodes, and we labeled the target episodes with incorrect response as the mind-wandering episodes. Particularly, we investigated window selection during the episodes when building classifiers both on an individual-level and group-level.

Compared to the previous state of the art in terms of brain-based classification of mind-wandering [19,32], our proposed approach achieved significant improvement. Previous work using EEG to predict task-general mind-wandering achieved a mean accuracy of 64% [32], while prior work using fNIRS for mind-wandering classification achieved a mean accuracy of 56% [19]. Our results suggest that focusing on a specific window can improve the classification results for individual-level classifiers. For group-level classification, we proposed a novel algorithm to incorporate individuals' differences in window selection. We show that when using the XGBoost as the group-level classifier and 5 s as the window size, the proposed ITWS algorithm achieved a mean F1-score of 73.2%. Moreover, we show that even though the window size can slightly affect the individual-level classification results for different participants, the performance of the ITWS algorithm is generally not affected by choice of window sizes. Also, our results show that the ITWS algorithm can improve the classification results when used with different classifiers (XGBoost, CNNs, and ANNs).

Our findings have important implications for designing and evaluating engaging and effective learning interfaces, as well as building attention-aware systems that can automatically detect mind-wandering states using fNIRS. For real-time applications, labeled brain data is required to train the classifier, which can then detect the activation at the prefrontal area associated with mind-wandering. However, the classification of mind-wandering is a challenging task. Different windows during mind-wandering episodes exhibit different time series behavior for each individual. As such, machine learning models trained on different windows of the labeled data can have different classification performance. Our classification methods serve the role of finding the best windows of training data for real-world applications. Classifiers trained on these windows can then be used to predict the label of real-time data. To do so, the first step is to collect labeled brain data from individuals. Then, the ITWS algorithm can be used to incorporate individuals' differences in window selection and determine the best windows for building the final classifier.

Our results show that the spread of selected windows varies a lot for some participants during cross-validation while applying the ITWS algorithm. This could be due to the overfitting of the individual-level classifiers since the dataset for each participant is small. Therefore, even though the classification results for test data from each participant did not show any differences in our work, further work that explores methods for more robust window selection can potentially improve the overall group-level classification results.

A limitation of this study is the mind-wandering episodes are inferred from behavioral responses and explicit reports of mind wandering. We aimed to avoid interrupting the mind-wandering episodes and therefore chose to determine mind-wandering episodes by SART errors, instead of using experience sampling probes. While previous research supports that SART errors are linked to mind-wandering [38,43,60], there is also research suggesting that SART errors could be related to impulsivity in individuals' responses [19]. Therefore, further investigation using experience sampling protocols and analyzing the window selection during the mind-wandering periods would be needed to confirm our findings.

9 Conclusion

In this paper, we investigated window selection for classifying mind-wandering episodes and on-task episodes using fNIRS. The proposed classification framework is data-driven and enables a more accurate detection of mind-wandering. The findings from this study also reveal individual differences in window selection for mind-wandering detection. This work could inform further research about the time course aspects of mind-wandering, and it builds a foundation for both evaluation of multimodal learning interfaces and future attention-aware systems based on fNIRS data.

References

1. Friedman JH (1989) Regularized discriminant analysis. J Am Stat Assoc 84(405):165–175
2. Afergan D, Peck EM, Solovey ET, Jenkins A, Hincks SW, Brown ET, Chang R, Jacob RJ (2014) Dynamic difficulty using brain metrics of workload. In: Proceedings of the 32nd annual ACM conference on Human factors in computing systems—CHI'14, pp 3797–3806. https://doi.org/10.1145/2556288.2557230
3. Balakrishnama S, Ganapathiraju A (1998) Linear discriminant analysis—a brief tutorial. Inst Signal Inf Process 18(4):1–8
4. Bandara D, Velipasalar S, Bratt S, Hirshfield L (2018) Building predictive models of emotion with functional near-infrared spectroscopy. Int J Hum–Comput Stud 110:75–85
5. Bixler R, DMello S (2016) Automatic gaze-based user-independent detection of mind wandering during computerized reading. User Model User-Adapt Interact 26(1):33–68. https://doi.org/10.1007/s11257-015-9167-1
6. Blanchard N, Bixler R, Joyce T, D'Mello S (2014) Automated physiological-based detection of mind wandering during learning. In: International conference on intelligent tutoring systems. Springer, Cham, pp 55–60. https://doi.org/10.1007/978-3-319-07221-0_7
7. Bosch N, Dmello S (2019) Automatic detection of mind wandering from video in the lab and in the classroom. IEEE Trans Affect Comput. https://doi.org/10.1109/taffc.2019.2908837
8. Buccino AP, Keles HO, Omurtag A (2016) Hybrid EEG-fNIRS asynchronous brain–computer interface for multiple motor tasks. PLoS ONE 11(1):1–16. https://doi.org/10.1371/journal.pone.0146610
9. Champaign J, McCalla G (2015) AttentiveLearner: improving mobile MOOC learning via implicit heart rate tracking. In: International conference on artificial intelligence in education. Springer, Cham, pp 367–376. https://doi.org/10.1007/978-3-319-19773-9
10. Chance B, Anday E, Nioka S, Zhou S, Hong L, Worden K, Li C, Murray T, Ovetsky Y, Pidikiti D, Thomas R (1998) A novel method for fast imaging of brain function, non-invasively, with light. Opt Express 2(10):411. https://doi.org/10.1364/oe.2.000411
11. Chawla Keven NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: Synthetic Minority Over-sampling Technique Nitesh. J Artif Intell Res 16(1):321–357. https://doi.org/10.1613/jair.953
12. Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. In: Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining 13–17 August, pp 785–794. https://doi.org/10.1145/2939672.2939785
13. Cho BH, Lee JM, Ku J, Jang DP, Kim J, Kim IY, Lee JH, Kim SI (2002) Attention enhancement system using virtual reality and EEG biofeedback. In: Proceedings IEEE virtual reality. IEEE, pp 156–163
14. Christoff K, Gordon AM, Smallwood J, Smith R, Schooler JW (2009) Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. In: Proceedings of the National Academy of Sciences, pp 8719–8724. papers3://publication/uuid/F7FC47FD-5AB1-4FCE-8F30-A99EE1870E01
15. Combrisson E, Jerbi K (2015) Exceeding chance level by chance: the caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. J Neurosci Methods 250:126–136
16. Connolly TM, Boyle EA, MacArthur E, Hainey T, Boyle JM (2012) A systematic literature review of empirical evidence on computer games and serious games. Comput Educ 59(2):661–686
17. Cui X, Bray S, Bryant DM, Glover GH, Reiss AL (2011) A quantitative comparison of nirs and fmri across multiple cognitive tasks. Neuroimage 54(4):2808–2821

18. D'Mello S, Olney A, Williams C, Hays P (2012) Gaze tutor: a gaze-reactive intelligent tutoring system. Int J Hum–Comput Stud 70(5):377–398
19. Durantin G, Dehais F, Delorme A (2015) Characterization of mind wandering using fNIRS. Front Syst Neurosci 9:45
20. Fox KC, Spreng RN, Ellamil M, Andrews-Hanna JR, Christoff K (2015) The wandering brain: meta-analysis of functional neuroimaging studies of mind-wandering and related spontaneous thought processes. NeuroImage 111:611–621. https://doi.org/10.1016/j.neuroimage.2015.02.039
21. Franklin MS, Smallwood J, Schooler JW (2011) Catching the mind in flight: using behavioral indices to detect mindless reading in real time. Psychon Bull Rev 18(5):992–997. https://doi.org/10.3758/s13423-011-0109-6
22. Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew MS (2016) Deep learning for visual understanding: a review. Neurocomputing 187:27–48
23. Harrivel AR, Stephens CL, Milletich RJ, Heinich CM, Last MC, Napoli NJ, Abraham NA, Prinzel LJ, Motter MA, Pope AT (2017) Prediction of cognitive states during flight simulation using multimodal psychophysiological sensing. AIAA Inf Syst AIAA Infotech Aerosp 2017:1–10. https://doi.org/10.2514/6.2017-1135
24. Herff C, Heger D, Fortmann O, Hennrich J, Putze F, Schultz T (2014) Mental workload during n-back task quantified in the prefrontal cortex using fNIRS. Front Hum Neurosci 7:935
25. Hirshfield LM, Solovey ET, Girouard A, Kebinger J, Jacob RJ, Sassaroli A, Fantini S (2009) Brain measurement for usability testing and adaptive interfaces: an example of uncovering syntactic workload with functional near infrared spectroscopy. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, pp 2185–2194
26. Ho TKK, Gwak J, Park CM, Song JI (2019) Discrimination of mental workload levels from multi-channel fNIRS using deep leaning-based approaches. IEEE Access 7:24392–24403. https://doi.org/10.1109/ACCESS.2019.2900127
27. Hosseini R, Walsh B, Tian F, Wang S (2018) An fNIRS-based feature learning and classification framework to distinguish hemodynamic patterns in children who stutter. IEEE Trans Neural Syst Rehabil Eng 26(6):1254–1263. https://doi.org/10.1109/TNSRE.2018.2829083
28. Huppert TJ, Diamond SG, Franceschini MA, Boas DA (2009) HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. Appl Opt 48(10):33. https://doi.org/10.1364/AO.48.00D280
29. Hutt S, Mills C, Bosch N, Krasich K, Brockmole J, D'mello S (2017) Out of the Fr-"Eye"-ing Pan: towards gaze-based models of attention during learning with technology in the classroom. In: Proceedings of the 25th conference on user modeling, adaptation and personalization. ACM, pp 94–103. https://doi.org/10.1145/3079628.3079669
30. Izzetoglu M, Izzetoglu K, Bunce S, Ayaz H, Devaraj A, Onaral B, Pourrezaei K (2005) Functional near-infrared neuroimaging. IEEE Trans Neural Syst Rehabil Eng 13:153–159. https://doi.org/10.1109/TNSRE.2005.847377
31. Jha AP, Morrison AB, Dainer-Best J, Parker S, Rostrup N, Stanley EA (2015) Minds at attention: mindfulness training curbs attentional lapses in military cohorts. PloS ONE 10(2)
32. Jin CY, Borst JP, Vugt MKV (2019) Predicting task-general mind-wandering with EEG. Cogn Affect Behav Neurosci 19:1–15
33. Kawashima I, Kumano H (2017) Prediction of mind-wandering with electroencephalogram and non-linear regression modeling. Front Hum Neurosci 11(July):1–10. https://doi.org/10.3389/fnhum.2017.00365
34. Kerous B, Skola F, Liarokapis F (2018) Eeg-based bci and video games: a progress report. Virtual Real 22(2):119–135
35. Khan MJ, Liu X, Bhutta MR, Hong KS (2016) Drowsiness detection using fNIRS in different time windows for a passive BCI. In: 2016 6th IEEE international conference on biomedical robotics and biomechatronics (BioRob). IEEE, pp 227–231. https://doi.org/10.1109/BIOROB.2016.7523628
36. Killingsworth MA, Gilbert DT (2010) A wandering mind is an unhappy mind. Science 330(6006):932. https://doi.org/10.1126/science.1192439
37. Makantasis K, Doulamis A, Doulamis N, Nikitakis A, Voulodimos A (2018) Tensor-based nonlinear classifier for high-order data analysis. In: 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp 2221–2225
38. Manly T, Robertson IH, Galloway M, Hawkins K (1999) The absent mind: further investigations of sustained attention to response. Neuropsychologia 37(6):661–670
39. Mayseless N, Hawthorne G, Reiss AL (2019) Real-life creative problem solving in teams: fNIRS based hyperscanning study. NeuroImage 203(August):116161. https://doi.org/10.1016/j.neuroimage.2019.116161
40. McKendrick R, Parasuraman R, Ayaz H (2015) Wearable functional near infrared spectroscopy (fNIRS) and transcranial direct current stimulation (tDCS): expanding vistas for neurocognitive augmentation. Front Syst Neurosci. https://doi.org/10.3389/fnsys.2015.00027
41. McKendrick R, Parasuraman R, Murtza R, Formwalt A, Baccus W, Paczynski M, Ayaz H (2016) Into the wild: neuroergonomic differentiation of hand-held and augmented reality wearable displays during outdoor navigation with functional near infrared spectroscopy. Front Hum Neurosci 10(MAY2016):216. https://doi.org/10.3389/fnhum.2016.00216
42. Mills C, Mello SD (2015) Toward a real-time (Day) Dreamcatcher: sensor-free detection of mind wandering during online reading. In: International educational data mining society
43. Mooneyham BW, Schooler JW (2013) The costs and benefits of mind-wandering: a review. Can J Exp Psychol 67(1):11–18. https://doi.org/10.1037/a0031569
44. Naseer N, Hong KS (2013) Classification of functional near-infrared spectroscopy signals corresponding to the right- and left-wrist motor imagery for development of a brain–computer interface. Neurosci Lett 553:84–89. https://doi.org/10.1016/j.neulet.2013.08.021
45. Naseer N, Hong KS (2015) fNIRS-based brain–computer interfaces: a review. Front Hum Neurosci 9:3
46. Naseer N, Hong KS (2015) fNIRS-based brain–computer interfaces: a review. Front Hum Neurosci 9:1–15. https://doi.org/10.3389/fnhum.2015.00172
47. Nelson BC (2007) Exploring the use of individualized, reflective guidance in an educational multi-user virtual environment. J Sci Educ Technol 16(1):83–97
48. Ninaus M, Kober SE, Friedrich EV, Dunwell I, Freitas SD, Arnab S, Ott M, Kravcik M, Lim T, Louchart SJJ et al (2014) Neurophysiological methods for monitoring brain activity in serious games and virtual environments: a review. Int J Technol Enhanc Learn 6(1):78
49. Noori FM, Naseer N, Qureshi NK, Nazeer H, Khan RA (2017) Optimal feature selection from fNIRS signals using genetic algorithms for BCI. Neurosci Lett 647:61–66. https://doi.org/10.1016/j.neulet.2017.03.013
50. Orihuela-Espina F, Leff DR, James DR, Darzi AW, Yang GZ (2010) Quality control and assurance in functional near infrared spectroscopy (fNIRS) experimentation. Phys Med Biol 55(13):3701–3724. https://doi.org/10.1088/0031-9155/55/13/009
51. Peck EM, Carlin E, Jacob R (2015) Designing brain–computer interfaces for attention-aware systems. Computer 48(10):34–42. https://doi.org/10.1109/MC.2015.315

52. Pike MF, Maior HA, Porcheron M, Sharples SC, Wilson ML (2014) Measuring the effect of think aloud protocols on workload using fNIRS. In: Proceedings of the 32nd annual ACM conference on Human factors in computing systems. ACM, pp 3807–3816. https://doi.org/10.1145/2556288.2556974
53. Pinti P, Scholkmann F, Hamilton A, Burgess P, Tachtsidis I (2018) Current status and issues regarding pre-processing of fNIRS neuroimaging data: an investigation of diverse signal filtering methods within a general linear model framework. Front Hum Neurosci 12:505
54. Rapp DN (2006) The value of attention aware systems in educational settings. Comput Hum Behav 22(4):603–614
55. Schmorrow DD, Fidopiastis CM (2015) Phylter: a system for modulating notifications in wearables using physiological sensing. In: International conference on augmented cognition, vol 9183. Springer, Cham, pp 167–177. https://doi.org/10.1007/978-3-319-20816-9
56. Schooler JW, Smallwood J, Christoff K, Handy TC, Reichle ED, Sayette MA (2011) Meta-awareness, perceptual decoupling and the wandering mind. Trends Cogn Sci 15(7):319–326. https://doi.org/10.1016/j.tics.2011.05.006
57. Shin J, Müller KR, Hwang HJ (2016) Near-infrared spectroscopy (NIRS)-based eyes-closed brain–computer interface (BCI) using prefrontal cortex activation due to mental arithmetic. Sci Rep 6(October):1–11. https://doi.org/10.1038/srep36203
58. Shin J, Von Luhmann A, Blankertz B, Kim DW, Jeong J, Hwang HJ, Muller KR (2017) Open access dataset for EEG + NIRS single-trial classification. IEEE Trans Neural Syst Rehabil Eng 25(10):1735–1745. https://doi.org/10.1109/TNSRE.2016.2628057
59. Shin J, Von Lühmann A, Kim DW, Mehnert J, Hwang HJ, Müller KR (2018) Simultaneous acquisition of eeg and nirs during cognitive tasks for an open access dataset. Sci Data 5:180003
60. Smallwood J, Schooler JW (2006) The restless mind. Psychol Bull 132(6):946–958. https://doi.org/10.1037/0033-2909.132.6.946
61. Smallwood J, Fishman DJ, Schooler JW (2007) Counting the cost of an absent mind: mind wandering as an underrecognized influence on educational performance. Psychon Bull Rev 14(2):230–236. https://doi.org/10.3758/BF03194057
62. Smallwood J, Beach E, Schooler JW, Handy TC (2008) Going awol in the brain: mind wandering reduces cortical analysis of external events. J Cogn Neurosci 20(3):458–469
63. Solovey ET, Girouard A, Chauncey K, Hirshfield LM, Sassaroli A, Zheng F, Fantini S, Jacob RJK (2009) Using fNIRS brain sensing in realistic HCI settings: experiments and guidelines. In: Proceedings of the 22nd annual ACM symposium on user interface software and technology. ACM
64. Treacy Solovey E, Afergan D, Peck EM, Hincks SW, Jacob RJ (2015) Designing implicit interfaces for physiological computing: guidelines and lessons learned using fNIRS. ACM Trans Comput–Hum Interact 21(6):35
65. Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E (2018) Deep learning for computer vision: a brief review. Comput Intell Neurosci 2018:7068349. https://doi.org/10.1155/2018/7068349
66. Wouters P, Van der Spek ED, Van Oostendorp H (2009) Current practices in serious game research: a review from a learning outcomes perspective. In: Games-based learning advancements for multi-sensory human computer interfaces: techniques and effective practices. IGI Global, pp 232–250