



OPEN ACCESS

EDITED BY

Dajiang Zhu,
University of Texas at Arlington,
United States

REVIEWED BY

Dalin Yang,
Washington University in St. Louis,
United States
Xiaowei Yu,
University of Texas at Arlington,
United States
Pardis Zarifkar,
Copenhagen University Hospital,
Denmark
Sergio Leonardo Mendes,
Federal University of ABC, Brazil

*CORRESPONDENCE

Yang Liu
liuyang@zust.edu.cn

SPECIALTY SECTION

This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

RECEIVED 30 September 2022

ACCEPTED 17 November 2022

PUBLISHED 01 December 2022

CITATION

Guo X, Liu Y, Zhang Y and Wu C
(2022) Programming ability
prediction: Applying an
attention-based convolutional neural
network to functional near-infrared
spectroscopy analyses of working
memory.
Front. Neurosci. 16:1058609.
doi: 10.3389/fnins.2022.1058609

COPYRIGHT

© 2022 Guo, Liu, Zhang and Wu. This
is an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction in
other forums is permitted, provided
the original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

Programming ability prediction: Applying an attention-based convolutional neural network to functional near-infrared spectroscopy analyses of working memory

Xiang Guo¹, Yang Liu^{1*}, Yuzhong Zhang² and Chennan Wu¹

¹School of Information and Electronic Engineering, Zhejiang University of Science and Technology, Hangzhou, China, ²School of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada

Although theoretical studies have suggested that working-memory capacity is crucial for academic achievement, few empirical studies have directly investigated the relationship between working-memory capacity and programming ability, and no direct neural evidence has been reported to support this relationship. The present study aimed to fill this gap in the literature. Using a between-subject design, 17 programming novices and 18 advanced students performed an n-back working-memory task. During the experiment, their prefrontal hemodynamic responses were measured using a 48-channel functional near-infrared spectroscopy (fNIRS) device. The results indicated that the advanced students had a higher working-memory capacity than the novice students, validating the relationship between programming ability and working memory. The analysis results also showed that the hemodynamic responses in the prefrontal cortex can be used to discriminate between novices and advanced students. Additionally, we utilized an attention-based convolutional neural network to analyze the spatial domains of the fNIRS signals and demonstrated that the left prefrontal cortex was more important than other brain regions for programming ability prediction. This result was consistent with the results of statistical analysis, which in turn improved the interpretability of neural networks.

KEYWORDS

programming ability, fNIRS, working memory, convolutional neural network, attention mechanism

Introduction

In the past decade, computer science and programming have been applied in many fields, such as engineering, social sciences music, art, and biology (Buitrago Flórez et al., 2017). Consequently, programming ability has become a basic skill that students may need to master. Several studies have suggested that students with better programming ability have better problem-solving skills and logical reasoning ability (Tu and Johnson, 1990; Shute, 1995; Wing, 2008; Werner et al., 2012; Ivanova et al., 2020; Relkin et al., 2021).

Programming requires memorization of a wide range of information and the ability to manipulate the information at the same time. Students process and hold this information in their working memory, the mode of information storage in the human brain as proposed by cognitive psychology (Baddeley and Hitch, 1974). Working memory is used to store task-relevant information for further application in the process of performing cognitive tasks. It is a memory system with limited capacity for temporary processing and storage of information that supports human thought processes by providing an interface for perception, long-term memory, and action (Baddeley, 2003). Working memory is not only the core of human cognition but also an important component of learning, reasoning, problem-solving, and intellectual activity (Barrouillet and Lépine, 2005; Baddeley, 2010). Working memory plays a critical role in learning. Extensive research has demonstrated a significant relationship between working-memory capacity and academic achievement (Swanson and Alloway, 2012; Anmarkrud et al., 2019). Studies have shown that high performance in math and readings are linked to high working-memory performance (Purpura and Ganley, 2014; Cantin et al., 2016). However, to the best of our knowledge, none of the prior studies showed that programming ability was related to working memory. Nevertheless, since code comprehension involves diverse cognitive domains, including math, logic, and language (Ivanova et al., 2020), programming ability may also be assumed to be related to students' working memory.

The n-back task is one of the most popular experimental paradigms for measuring working memory. The n-back paradigm is a continuous task paradigm (Cohen et al., 1997). In the n-back experiment, participants are asked to monitor a series of verbal/non-verbal stimuli and indicate whether the stimuli currently presented are the same as those that appeared in n trials previously (Braver et al., 1997). The traditional n-back experimental measurements include evaluation of accuracy and reaction time.

In recent years, many researchers have combined functional magnetic resonance imaging (fMRI), electroencephalography (EEG), and functional near-infrared spectroscopy (fNIRS) to measure physiological signals in task-evoked experimental processes to obtain the underlying neuroscientific mechanism

of working memory (Ragland et al., 2002; Herff et al., 2013; Lv et al., 2014, 2015; Yeung et al., 2021).

In comparison with fMRI, fNIRS requires a small volume and is lightweight and portable while yielding images with a higher temporal resolution. fNIRS also shows a faster spatial response speed than EEG (Ferrari and Quaresima, 2012; Hong and Yaqub, 2019; Quaresima and Ferrari, 2019; Yang et al., 2019).

Functional near-infrared spectroscopy is a neuroimaging technique for measurement of hemodynamic processes in the brain. In this technique, the absorption of infrared light with a wavelength of 650–950 nm passing through the brain tissue is evaluated to monitor the changes in blood oxygen concentration in different brain tissue regions (Pinti et al., 2020) and obtain insights into the same activation patterns as fMRI. Matthes and Gross (1938) first demonstrated the spectroscopic determination of oxygenated hemoglobin (HbO) and deoxygenated hemoglobin (HbR) in human tissue in the red and near-infrared regions. In 1993, some research groups proved that fNIRS could be used to investigate brain activity non-invasively (Chance et al., 1993; Hoshi and Tamura, 1993; Villringer et al., 1993). Wolf et al. (2002) first used near-infrared spectroscopy and detected significant changes in the local concentrations of HbO and HbR during brain activity. When the brain executes a task, the increased metabolic demands for oxygen and glucose result in an oversupply of local cerebral blood flow (CBF) to satisfy the increased metabolic demand. CBF is regulated by several neurovascular coupling mechanisms. Therefore, the excessive supply of local CBF leads to an increase in HbO concentration and a decrease in HbR concentration (Pinti et al., 2020). Some previous studies have shown that fNIRS is sensitive to load-dependent working-memory changes in activation (Herff et al., 2013; Meidenbauer et al., 2021) and have demonstrated linear increments in HbO concentrations in frontal activation based on n-back levels (Ayaz et al., 2012; Yeung et al., 2021). The results of a meta-analysis of brain imaging data acquired during the n-back task showed that the participants' prefrontal cortex was activated consistently (Nystrom et al., 2000; Owen et al., 2005). Therefore, in this study, we mainly focused on the concentration changes in HbO and HbR in the prefrontal cortex.

Functional near-infrared spectroscopy is also an effective approach to explore the temporal and spatial states of the human brain (Maki et al., 1995). It provides a balance between temporal and spatial resolution in comparison with other neurophysiological modalities, making it a viable option for mental workload estimation (Isbilir et al., 2019). In the present study, we focused on investigating the channel-wise analysis of fNIRS spatial features to explore the most important brain regions for predicting programming ability.

A recurrent neural network (RNN) is usually considered the best neural network structure for time series prediction, but recent studies have shown that a convolutional neural

network (CNN) can perform these predictions comparably not only with greater accuracy but also more easily and clearly (Bai et al., 2018), particularly when there are many similar time series to learn from (Chen et al., 2019). Dilated convolutions can make one-dimensional CNNs effectively learn time series dependencies (Yu and Koltun, 2016; Borovykh et al., 2018). In RNNs, the prediction of subsequent time steps must occur after the previous time step has been completed. In contrast, convolutions can be performed in parallel because the filters used in each layer are the same. Therefore, in training and evaluation, CNNs can process long input sequences simultaneously rather than sequentially as with RNNs (Bai et al., 2018). Since fNIRS signals in the n-back task show multiple similar time series, and we aimed to capture features over the global theoretical receptive fields, a CNN was the best choice for the backbone network in the present study. CNNs have been widely used for automatic fNIRS signal analysis (Trakoolwilaiwan et al., 2017; Janani et al., 2020) and have been used to investigate mental workload levels using multichannel fNIRS signals (Ho et al., 2019; Saadati et al., 2020). One previous study employed a CNN to analyze fNIRS features during an n-back task and proved that CNNs can learn features automatically and obtain accurate results (Yang et al., 2020).

False discovery rate (FDR) measurements (Singh and Dan, 2006) and statistical parametric mapping (SPM) (Koh et al., 2007) have been applied for channel-wise analysis for fNIRS signals. However, these statistical analysis methods corrupt the temporal domain information in fNIRS signals. Among deep neural networks, squeeze-and excitation network (SENet).

SENet, NIRSIT, PET. represents the pioneering concept of channel attention (Guo et al., 2022). The traditional pooling layer reduces the feature map, resulting in damage to channel important information. In contrast, a squeeze-and-excitation (SE) block is a type of attention layer that can collect channel important weight in train processing. The SE block can be used to collect global information, capture channel-wise relationships, and incorporate spatial attention into the structure of the CNN (Hu et al., 2020), thereby improving the interpretability of neural networks. Moreover, in comparison with the application of convolution in feature mapping, the computational cost of SE and weighted summation is very low (Guo et al., 2022). However, SENet, which is an advanced, novel channel-attention network, has not been reported for use in fNIRS signal analysis.

To the best of our knowledge, no previous study has directly explored the brain mechanisms underlying the relationship between working memory and programming ability by using fNIRS data. Thus, the first aim of the current study was to validate the relationship between programming ability and working memory by using an n-back task. The second aim was to investigate whether the fNIRS signals detected during the performance of n-back tasks can predict the participant's programming ability. The third aim was to explore the capability

of the attention-based CNN method to analyze the spatial information of the fNIRS signals to identify the optimal brain regions to predict programming ability. Thus, we aimed to explore whether general psychological experiments could be used to predict learners' programming ability, and to provide neuroscience evidence for the findings.

Materials and methods

Participants

Thirty-five participants (17 novices and 18 advanced students) were recruited from the School of Information and Electronic Engineering, Zhejiang University of Science and Technology in China. The novice group included 13 male and four female participants, while the advanced group included 14 male and four female participants. All participants were over 18 years of age [mean \pm standard deviation (SD), 20.61 \pm 1.23 years].

The novices were freshmen from C++ courses who had not undergone programming-intensive training previously. On the other hand, the advanced students were from the programming competition team who had at least 2 years of programming-intensive training and had at least received an award in the international collegiate programming contest (We did not investigate the effect of programming training on working memory in the present study, and merely used this approach to select participants). Before the experiment, the participants were asked to complete a programming level test, which consisted of ten items with ten points for each completely correct answer and deductions for incomplete results. The maximum total score in the programming level test was 100. The advanced students had higher scores on the programming level test than the novices (mean \pm SD, 83.9 \pm 5.96 vs. 50.0 \pm 7.71). An independent-sample *t*-test revealed that the scores on the programming level test were significantly different between the two groups [$F(1, 34) = 214.07, p < 0.001, \eta_p^2 = 0.87$].

All participants signed the informed consent form before the experiment and received a small gift at the end of the study to thank them for their time and effort.

Experimental setup and tasks

The participants were assigned to two groups: novice and advanced students. The experimental procedures were conducted through computer programs based on E-prime, a general psychological experiment software. Each participant was individually tested in a laboratory environment for approximately 30 min. Before completing the task, participants learned about the experimental procedure. The participants were asked to relax and do nothing as the baseline task, and

measurements obtained during this baseline task were used as the baseline for comparison of fNIRS signals.

The present study employed an n-back task to investigate the participants' working memory. The participants monitored a series of character stimuli and responded whenever a stimulus presented was the same as the one presented n trials previously (Owen et al., 2005). The main n-back task involved 30 blocks, with 10 blocks of each n-back level presented pseudorandomly (Braver et al., 1997). Figure 1 shows the trial schematic of the n-back task conditions in the present study. For example, in the 2-back task, the third C did not match the first A. However, the fourth B matched the second B, so the participants were required respond positively whenever the character they saw was the same as the one they viewed two characters earlier. fNIRS data were recorded continuously during the entire session.

Functional near-infrared spectroscopy data acquisition

In the present study, we focused on the HbO and HbR concentration changes in the prefrontal cortex. The hemodynamic responses measured in the prefrontal cortex were consistent enough to distinguish three levels of n-back workloads (Owen et al., 2005; Herff et al., 2013).

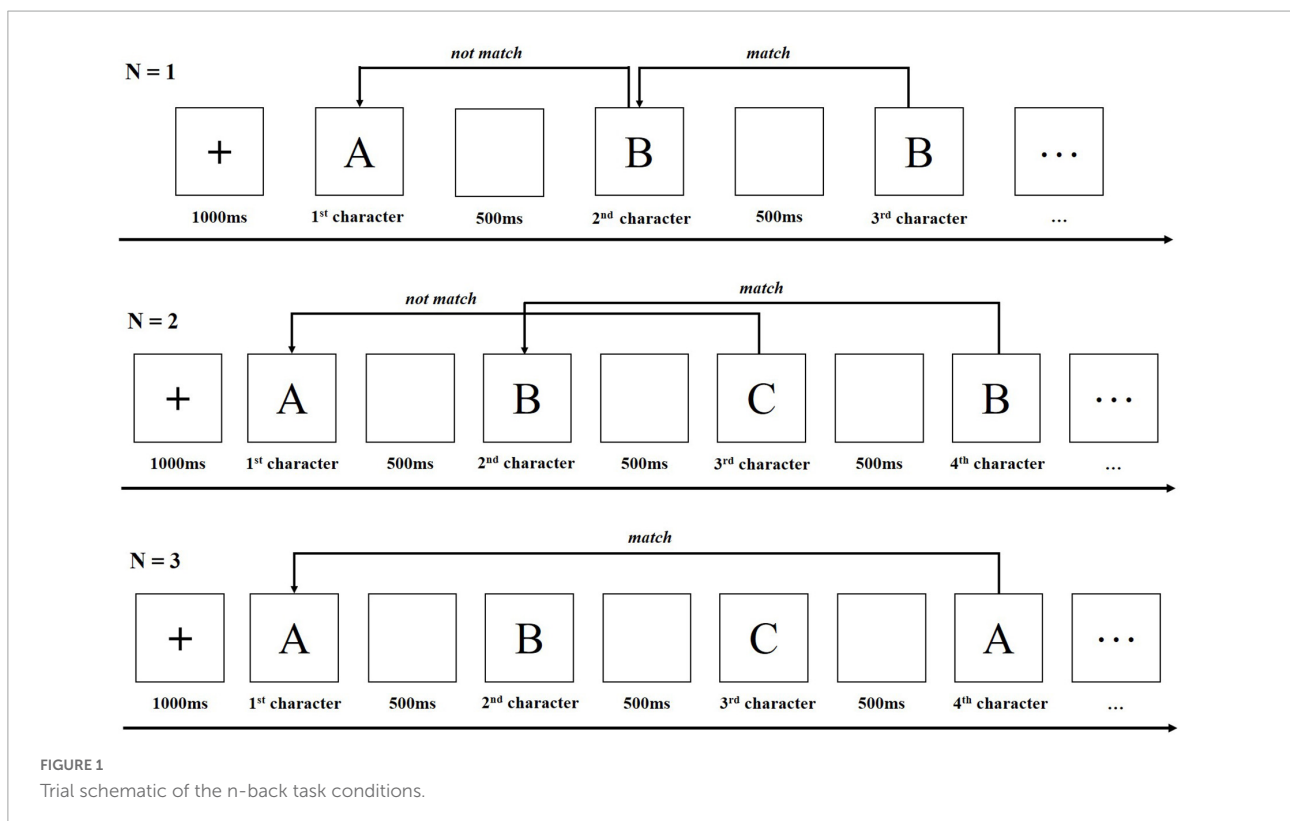
Functional near-infrared spectroscopy (fNIRS) data were recorded at a sampling rate of 8.13 Hz using a wearable NIRS

device, the NIRSIT model from OBELAB (Korea). The NIRSIT device has a comprehensive 48-channel system and can capture depth-dependent hemodynamic changes in the prefrontal cortex. This system utilizes 24 laser sources (780/850 nm; maximum power under 1 mw) and 32 photodetectors (Choi et al., 2016). The grouping of NIRSIT channels is shown in Figure 2 and includes the right (#1–16), center (#17–32), and the left (#33–48) regions.

Functional near-infrared spectroscopy data pre-processing

Figure 3 presents the flow diagram for fNIRS data pre-processing. We first loaded fNIRS data into the NIRSIT Analysis Tool for visual inspection, segmentation of the main n-back trials from practice trials, and division of the prefrontal cortex into three regions—right, center, and left—as shown in Figure 2. We then performed visual inspection at the participant level to examine overall data quality and to evaluate the quality of the data obtained from both sides of the prefrontal cortex, which showed a much lower signal-to-noise ratio (SNR) than the data from the center of the prefrontal cortex.

Since each participant had a different completion time, we tailored the data with the shortest completion time. Thus, the three n-back tasks had different data lengths. However, in



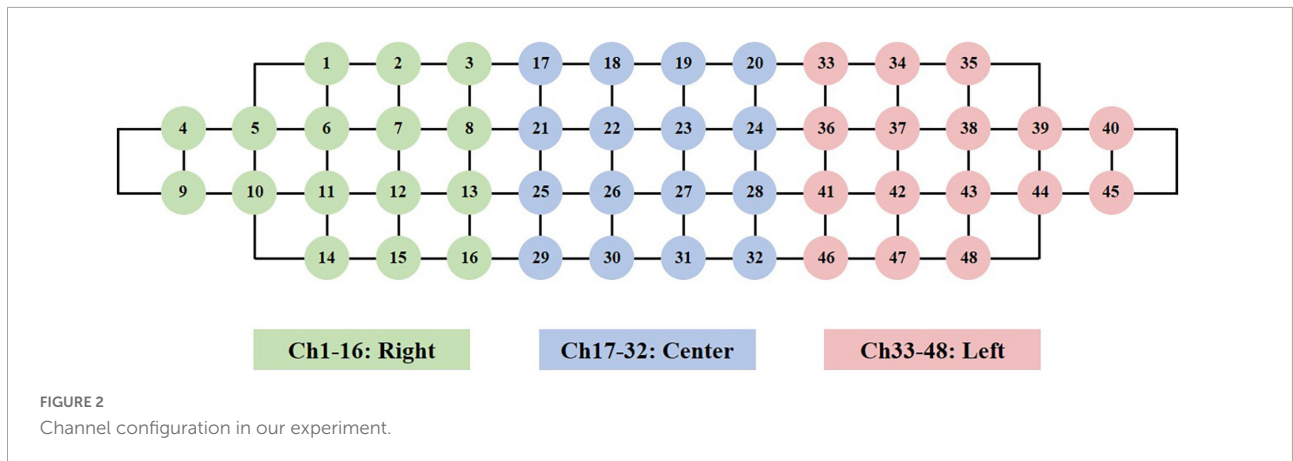


FIGURE 2 Channel configuration in our experiment.

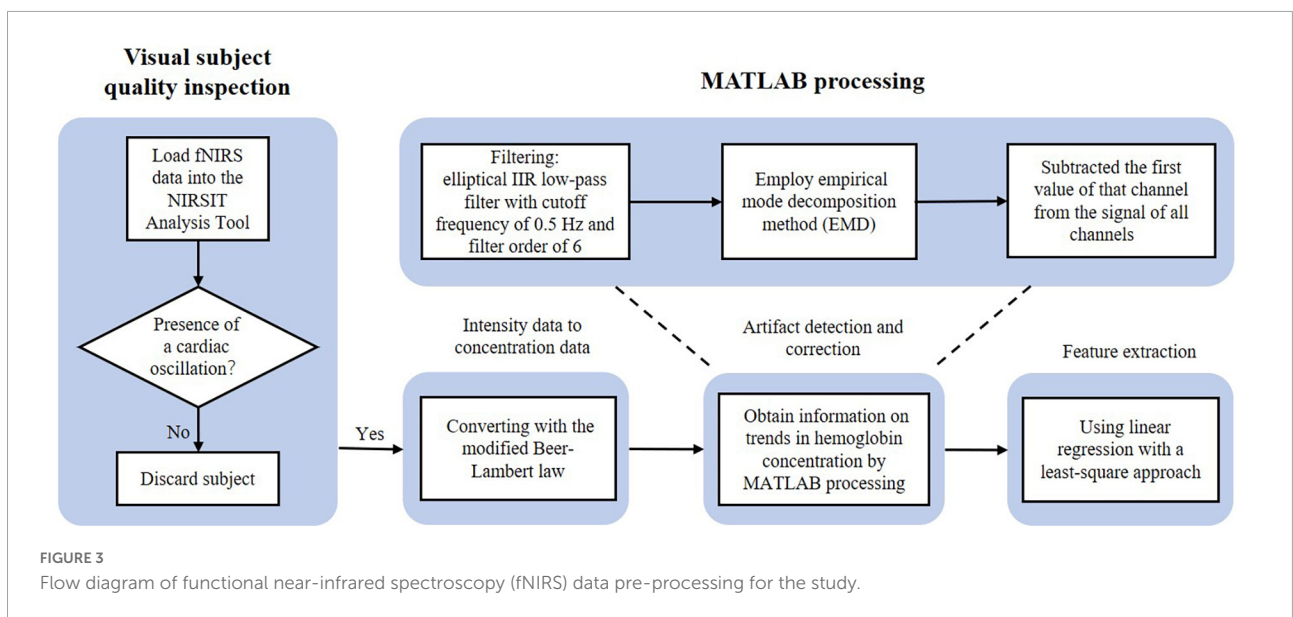


FIGURE 3 Flow diagram of functional near-infrared spectroscopy (fNIRS) data pre-processing for the study.

the subsequent data processing, we focused on the slope and statistical features that are less affected by data length.

Visual inspection was performed by examining the spectrogram of every channel to identify the presence of cardiac oscillations, which are typically approximately 1 Hz (Tong et al., 2011). The presence of this cardiac signal is a good indicator that the optical density signals are successfully coupled with a physiological hemodynamic response (Hocke et al., 2018). This method was employed for preliminary selection of participants. In this visual inspection, one participant with unusable data, which was defined by the presence of more than seven unclean channels in one area, was identified and excluded from further analyses. Thus, the novice and advanced student groups included data from 17 participants each.

Then, we used the NIRSIT Analysis Tool to convert the raw light intensity data into HbO and HbR concentrations by means of the modified Beer–Lambert law (Sassaroli and Fantini, 2004). However, the signals still contained biological

and technical artifacts. Several cardiovascular phenomena, such as heart beats, respiration, and blood pressure (Mayer waves), influenced the recorded data. Movement artifacts such as high-frequency spikes, shifts from baseline intensity, and low-frequency variations, which are present in most fNIRS datasets, can severely affect the quality of recorded data (Franceschini et al., 2006; Huppert et al., 2009).

Therefore, we conducted further data processing in Python-SciPy. To attenuate heartbeat and other biological signals, we used an elliptical Infinite Impulse Response (IIR) low-pass filter with a cutoff frequency of 0.5 Hz and a filter order of 6, which robustly removed biological artifacts in the data (Herff et al., 2013). Then, we tried to use the wavelet artifact removal method to reduce the effect of movement artifacts. Since the signals showed channel- and participant-wise variations and the wavelet basis function had limited adaptability, we found it difficult to identify a suitable wavelet basis function to remove the movement artifacts effectively.

Therefore, we used the empirical mode decomposition (EMD) method, which can decompose signals without any additional parameters and therefore robustly reduced the influence of movement artifacts and Mayer wave-like effects in the data. Some of the spontaneous physiological information, such as breathing rate (~ 0.3 Hz) and very low-frequency oscillations (< 0.01 Hz), were still reflected in the data obtained for post-processing analysis. Therefore, we used the EMD method to deal with this noise after applying a low-pass filter, since direct application of a high-pass filter would have destroyed other useful signal components. Furthermore, due to various factors, the amplitude and intensities of the acquired hemodynamic signals differed significantly among participants. To attenuate the influence of a single participant's data on the grand average of acquired hemodynamic signals, we subtracted the first value of that channel from the signal of all channels and rejected channels that may have significantly influenced the grand average. After these treatments, we obtained information on trends in HbO and HbR concentrations.

The slope of the trend data (Herff et al., 2013) is often used as a simple but effective feature. To obtain this slope, we fitted a straight line to the data using linear regression with a least-squares approach.

Attention-based convolutional neural network for functional near-infrared spectroscopy spatial feature analysis

The structure of the attention-based CNN that we introduced to analyze the fNIRS signals is depicted in Figure 4, with the channel-attention blocks showing the spatial importance of fNIRS channels. After data pre-processing as described in Section "Functional near-infrared spectroscopy data pre-processing," each channel of data was resampled and rescaled to a uniform length $L = 256$. Then, we stacked all channels to build a 48×256 feature matrix and used direct resampling and rescaling. The fNIRS signals did not contain periodic frequency information, which may have been corrupted by those processes.

In our training procedure, the mean-squared-error (MSE) method was chosen for the loss function.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}_i)^2$$

Here, we choose Adam as our optimizer and set the initial learning rate to 0.01.

Three SE blocks were inserted into three normal convolution layers. After the global average pooling operator, each channel data point was consolidated into one data point. In the first SE block, the fully connected layer transforms the 1×48 vector to 1×24 ; this process is also called the "squeeze." The squeezing function also serves to embed the

global distribution of feature responses over all channels. This operator is followed by an excitation operator, which consists of a fully connected layer and a sigmoid activation layer. Excitation is a self-gating mechanism (Hu et al., 2020) that produces a mask representing the per-channel modulation weights. These weights are then applied to the original feature map to generate the new output. This series operation is also known as the self-attention operation (Vaswani et al., 2017). The feature vector needs to be squeezed small and then return to the origin scale *via* excitation because we aimed to improve the training pressure and to prevent overfitting. The reduction factor needs to be carefully adjusted within the training process. A special classified task was chosen. After proper training, we opened the SE block to observe the channel-wise weights mask.

The backbone network of the attention-based CNN we used (as shown in Figure 4) had a traditional CNN structure. Here, FC refers to the fully connected layer; BN refers to the batch-normalization layer; and the two round circles indicate the dot-product operator. However, to prevent mixing of the channel information, a generic convolution (GC) layer cannot be used at the beginning of the network. As illustrated in Figure 4, the key point is to replace the first GC layer with a depth-wise convolution (DC) layer before the self-attention mechanism finds the important channels. Nevertheless, the other convolution layers are still GC layers. The DC layer ignores the interchannel information, which must be remedied with a point-wise convolution (PC) layer. This will complicate the overall structure of our neural network.

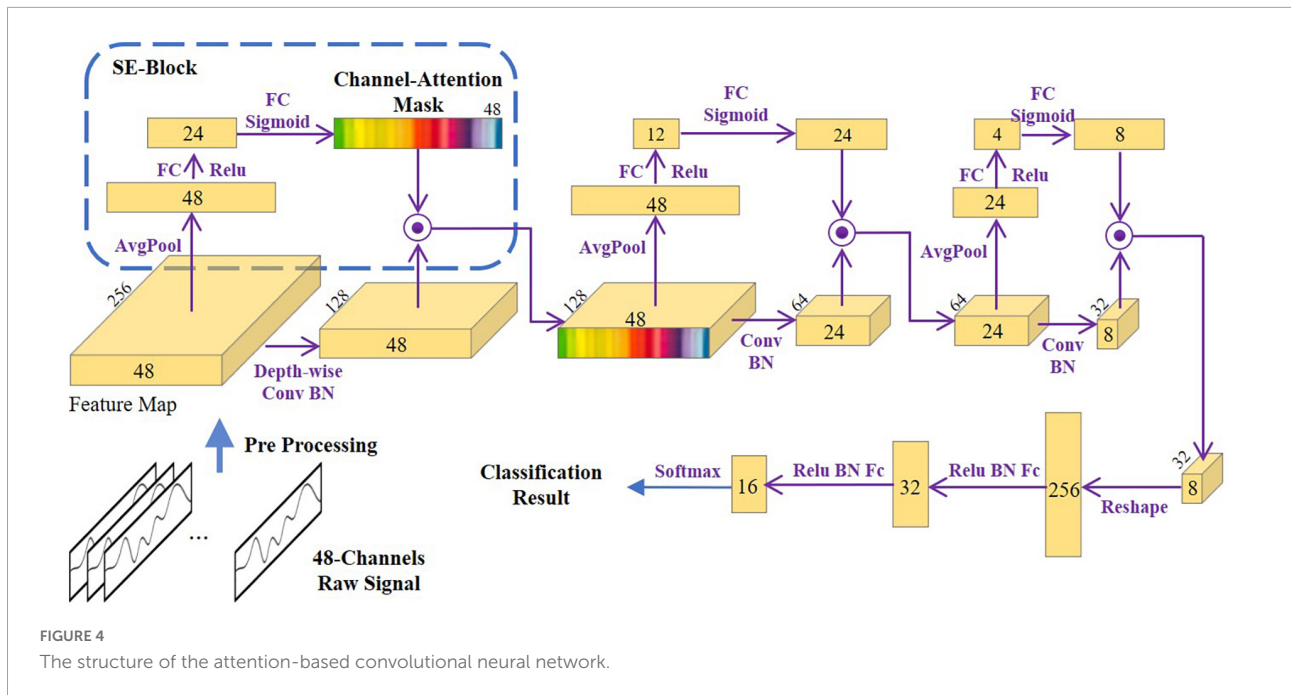
Due to the limited amount of training data in this study, the neural network was easily overfitted. However, in our training procedure, we were not overtly concerned with the generalization properties of the neural network. Instead, we aimed to reveal the importance of channels. The main purpose of this model is to train the channel-attention block, and some degree of overfitting can help make important channels more obvious (Hu et al., 2020). Thus, model overfitting can be acceptable for general inference.

Results

The criterion for statistical significance was set at $p < 0.05$. The Greenhouse–Geisser correction was used to compensate for sphericity violations. Effect sizes were measured by η_p^2 , with $\eta_p^2 = 0.01$, 0.06, and 0.14 indicating small, medium, and large effects, respectively (Fritz et al., 2012).

n-back performance

The descriptive statistics for accuracy and reaction time in each group are presented in Table 1. Accuracy was calculated by determining the average percentage of correct trials under



each back condition, while reaction time was computed by determining the mean across correct trials for each back condition. As shown in **Table 1**, novices performed the trials with lower accuracy and slower reaction times than advanced students over each back level.

Paired *t*-tests indicated that the accuracies for the 1-back and 2-back conditions were near the ceiling levels and did not differ in both advanced students and novices. Accuracy for the 3-back condition was marginally significantly lower than those for the 1-back [$t(16) = 1.984, p = 0.073$] and the 2-back [$t(16) = 2.072, p = 0.063$] conditions among the advanced students. However, accuracy for the 3-back condition was significantly lower than those for the 1-back [$t(16) = 4.690, p = 0.001$] and the 2-back [$t(16) = 3.801, p = 0.003$] conditions among the novices.

Paired *t*-tests indicated that the 1-back task was performed faster than the 2-back [$t(16) = -4.641, p = 0.001$; $t(16) = -4.935, p < 0.001$] and 3-back [$t(16) = -8.567, p < 0.001$; $t(16) = -11.950, p < 0.001$] tasks by the advanced students and the novices, respectively, while the 2-back task was performed faster than the 3-back [$t(16) = -7.637, p < 0.001$; $t(16) = -8.368, p < 0.001$] task by both advanced students and the novices.

To examine group differences, we conducted one-way repeated-measures analysis of variance (ANOVA) with programming ability (novices vs. advanced students) as the between-subjects factor and n-back levels (1-back, 2-back, and 3-back) as the within-subject factor.

One-way repeated-measures ANOVA revealed a marginally significant main effect of programming ability on accuracy [$F(1, 32) = 3.867, p = 0.075, \eta_p^2 = 0.260$]. The interaction between

programming ability and the n-back task was not significant [$F(2, 64) = 1.692, p = 0.207, \eta_p^2 = 0.133$].

One-way repeated-measures ANOVA with Greenhouse-Geisser correction revealed a main effect of programming ability on reaction time [$F(1, 32) = 5.650, p = 0.029, \eta_p^2 = 0.239$]. The interaction between programming ability and the n-back task was also not significant [$F(2, 64) = 1.177, p = 0.304, \eta_p^2 = 0.061$].

Figure 5 illustrated that the correlations between reaction time and programming score were negative for 1-, 2-, and 3-back levels ($r = -0.75, r = -0.71, \text{ and } r = -0.81$), i.e., the reaction time was faster for a higher programming score.

Functional near-infrared spectroscopy hemodynamic responses

To determine the differences in hemodynamic responses between novices and advanced students under the three n-back conditions, we first analyzed the grand averages of all participants.

Figure 6 exhibited the grand averages of all participants for the three n-back levels. The blue lines showed the grand averages for novices, while the magenta line showed the overall mean for advanced students. For HbO, a clear increase was observed at the 1-, 2-, and 3-back levels, and the slope was positive for all three n-back conditions in the left, center, and right prefrontal cortices. The grand average increase was steeper in the 2-back task than in the 1-back task and was the steepest in the 3-back task.

For HbR, a slight decrease in concentration changes can be seen for all three n-back conditions, and the slope was negative

TABLE 1 Means and standard deviations of accuracy and reaction time during the n-back task.

Dependent variable	Accuracy				Reaction time (ms)			
	Novices		Advanced students		Novices		Advanced students	
	M	SD	M	SD	M	SD	M	SD
1-back	0.986	0.023	0.992	0.019	550	113	474	53.2
2-back	0.983	0.033	0.986	0.032	751	152	654	75.5
3-back	0.845	0.080	0.931	0.065	1276	222	1087	196

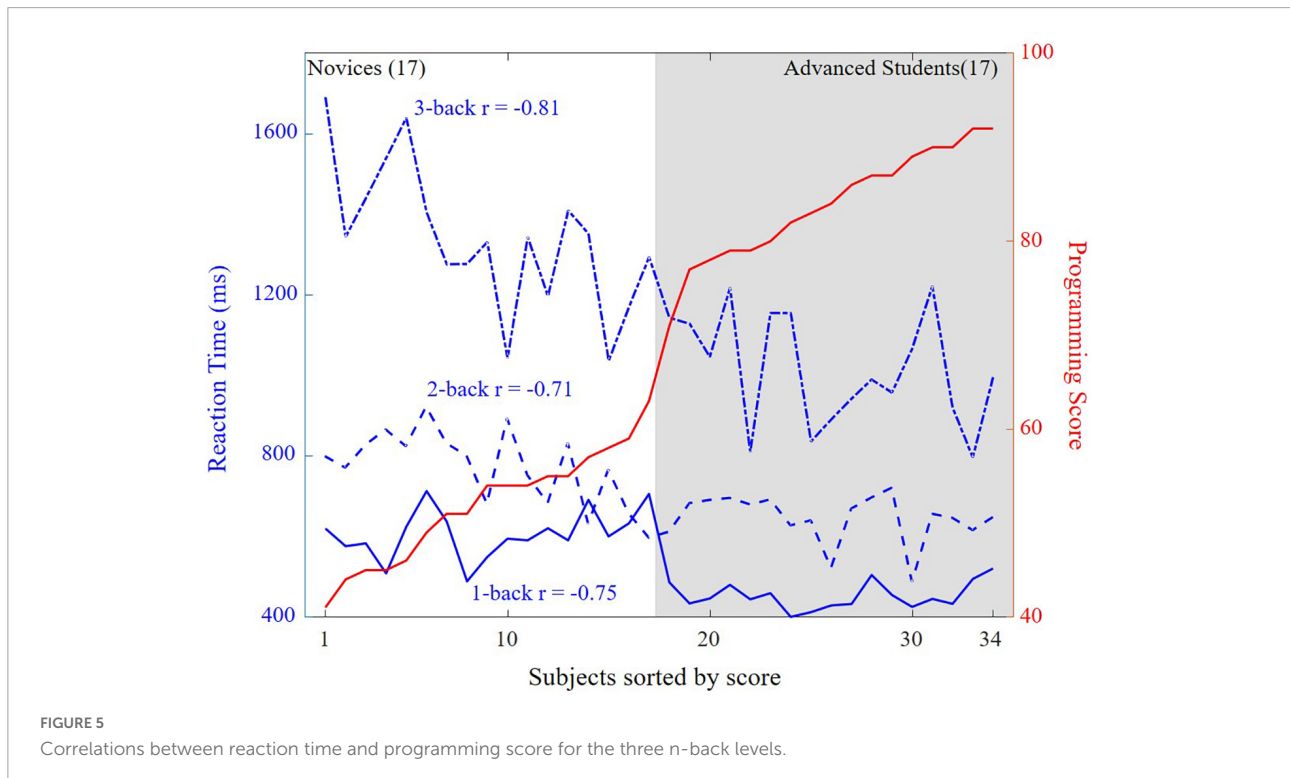


FIGURE 5 Correlations between reaction time and programming score for the three n-back levels.

for the three n-back levels; conversely, there was no obvious difference between the 1- and 2-back grand averages and the 3-back grand average.

One-way repeated-measures ANOVA revealed a main effect of programming ability on HbO [$F(1, 32) = 8.838, p = 0.007, \eta_p^2 = 0.287$; $F(1, 32) = 12.713, p = 0.002, \eta_p^2 = 0.366$; $F(1, 32) = 25.805, p < 0.001, \eta_p^2 = 0.540$] in the right, center, and left prefrontal cortices. The interaction between programming ability and the n-back task was not significant.

Figure 7 illustrated that the correlations between HbO and programming score were negative for the three n-back levels in the left ($r = -0.50, -0.51, -0.54$), center ($r = -0.36, -0.17, -0.38$), and right ($r = -0.09, r = -0.23, r = -0.38$) prefrontal cortices, i.e., HbO is lower for higher programming scores.

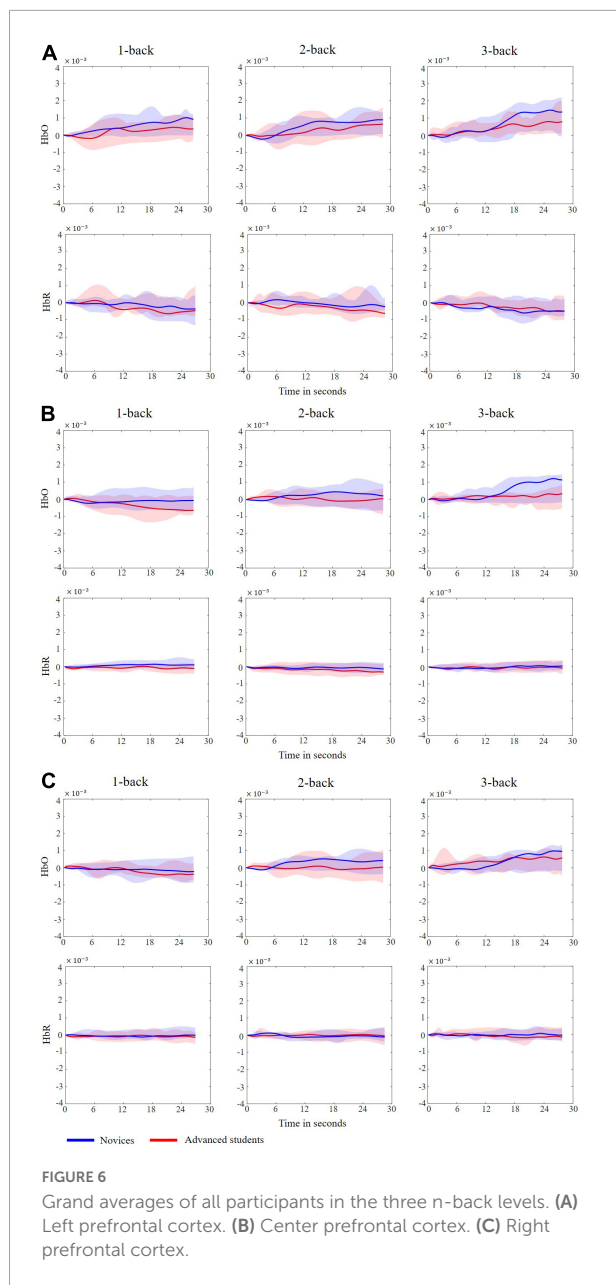
One-way repeated-measures ANOVA revealed that the main effect of programming ability on HbR [$F(1, 32) = 0.001, p = 0.980, \eta_p^2 < 0.001$; $F(1, 32) = 2.716, p = 0.114, \eta_p^2 = 0.110$;

$F(1, 32) = 0.104, p = 0.750, \eta_p^2 = 0.005$] was not significant in the right, center, or left prefrontal cortices, and there was no interaction between programming ability and the n-back task.

The results indicated that HbO can indicate working-memory load and show significant associations between brain activity and programming ability, but HbR cannot.

Functional near-infrared spectroscopy feature analysis using attention-based convolutional neural network

To obtain the most important channels in the fNIRS signals, we constructed a virtual classification task, and tried letting the neural network model illustrate the importance of the fNIRS channels through the virtual training task. Under this virtual task, we directly combined all subject data into a signal batch



and used MSE loss to train the network. After approximately 50 epochs, the loss stopped falling, and we obtained almost 100% accuracy. Because the amount of data was not very large, the generalization ability of the network was weak. However, we were not going to use this trained network for general classification in a new set of data. We only needed to observe the network's understanding of channel importance under this task. In the present study, we used the mean value of 10 training sessions for subsequent analysis.

The left panels in **Figure 8** showed the channel weights fitting with the training process. The left panel showed that the accuracy (red curve) was close to 100% with the loss (blue curve) down to zero. Furthermore, we distinguished those weights into

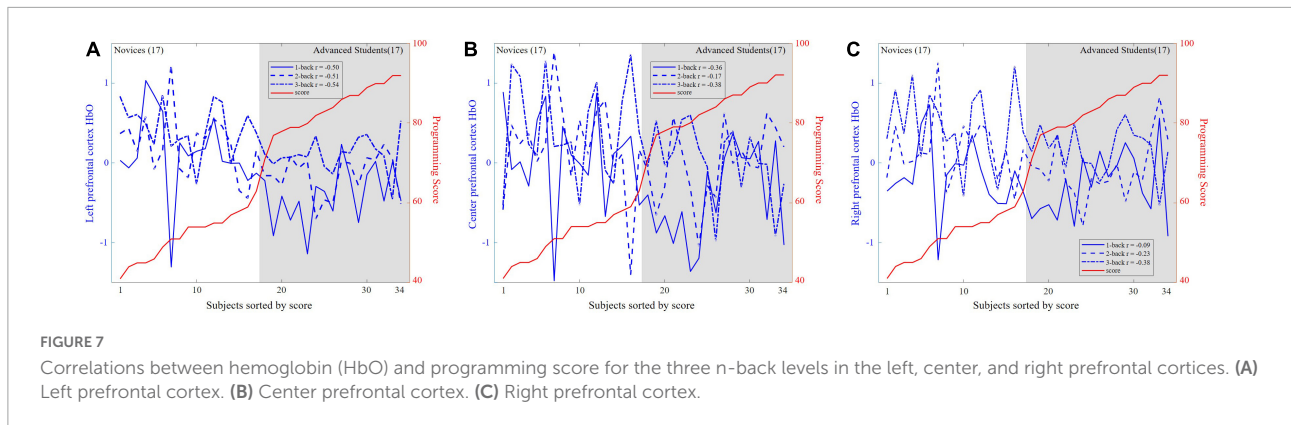
three brain regions (see **Figure 2**), as shown in the right panels. After fitting, the left prefrontal cortex showed an obviously high weight in the three n-back levels. These findings were consistent with the results of one-way repeated-measures ANOVA, in which the left prefrontal cortex had a larger effect size than the right and center prefrontal cortices.

Discussion

The main purpose of the present study was to investigate the brain mechanisms underlying the relationship between working memory and programming ability by using fNIRS signals.

The analysis of participants' n-back performance showed differences in the accuracy and reaction time depending on the n-back level between novices and advanced students. Advanced students performed better than novices in terms of both accuracy and reaction time. Since the n-back task is recognized as an effective method to measure working memory, a better n-back performance indicated higher working-memory capacity (Kirchner, 1958). The study results validated the relationship between programming ability and working memory, and students with higher working-memory capacity showed better programming ability. The results also provided evidence that limited working-memory capacity has negative effects on learning (Alloway, 2009).

Since the n-back task may be easier for advanced students, the advanced participants were expected to show less prefrontal cortex activation during each n-back experiment (Asgher et al., 2019; Khoe et al., 2020). **Figures 6, 7** illustrated the neural evidence of this finding. The hemodynamic responses of HbO associated with n-back stimulus presentation increased more in novices than in advanced students. The results of the statistical analyses revealed that HbO in the prefrontal cortex showed significant differences between novices and advanced students during the n-back task. Thus, HbO signals measured during the n-back test can be used to robustly predict the programming ability of students. The changes in cerebral blood oxygen signals represent the changes in local oxygen consumption caused by local brain activity and reflect the activity state of the human brain (Strangman et al., 2002). According to the neural efficiency hypothesis (Haier et al., 1988), the higher the performance in related fields (the higher the cognitive ability), the lower the activation degree of the cerebral cortex, showing a negative correlation. Thus, in comparison with a lower cognitive ability group, a higher cognitive ability group shows lower activation of brain regions when performing tasks with the same difficulty (Dunst et al., 2014; Genc et al., 2018). A higher working-memory load tends to produce greater prefrontal cortex activation (Herff et al., 2013). The novices exhibited significantly higher HbO concentration increments than their advanced counterparts during the n-back tests. Thus, the working-memory load in novices was higher and



consumed more mental resources. In comparison with the novices, the advanced students illustrated lower prefrontal cortex activation for the n-back task, which was considered to place less of a demand on working-memory load and was easier to complete for the advanced students. This is the underlying basis for the measurement of prefrontal cortex activation of fNIRS signals during n-back tasks to predict an individual's programming ability.

The results also indicated the possibility of predicting students' programming potential through general psychological experiments such as n-back test (as shown in Figure 5). This method may be especially useful for evaluating individuals with no programming foundation.

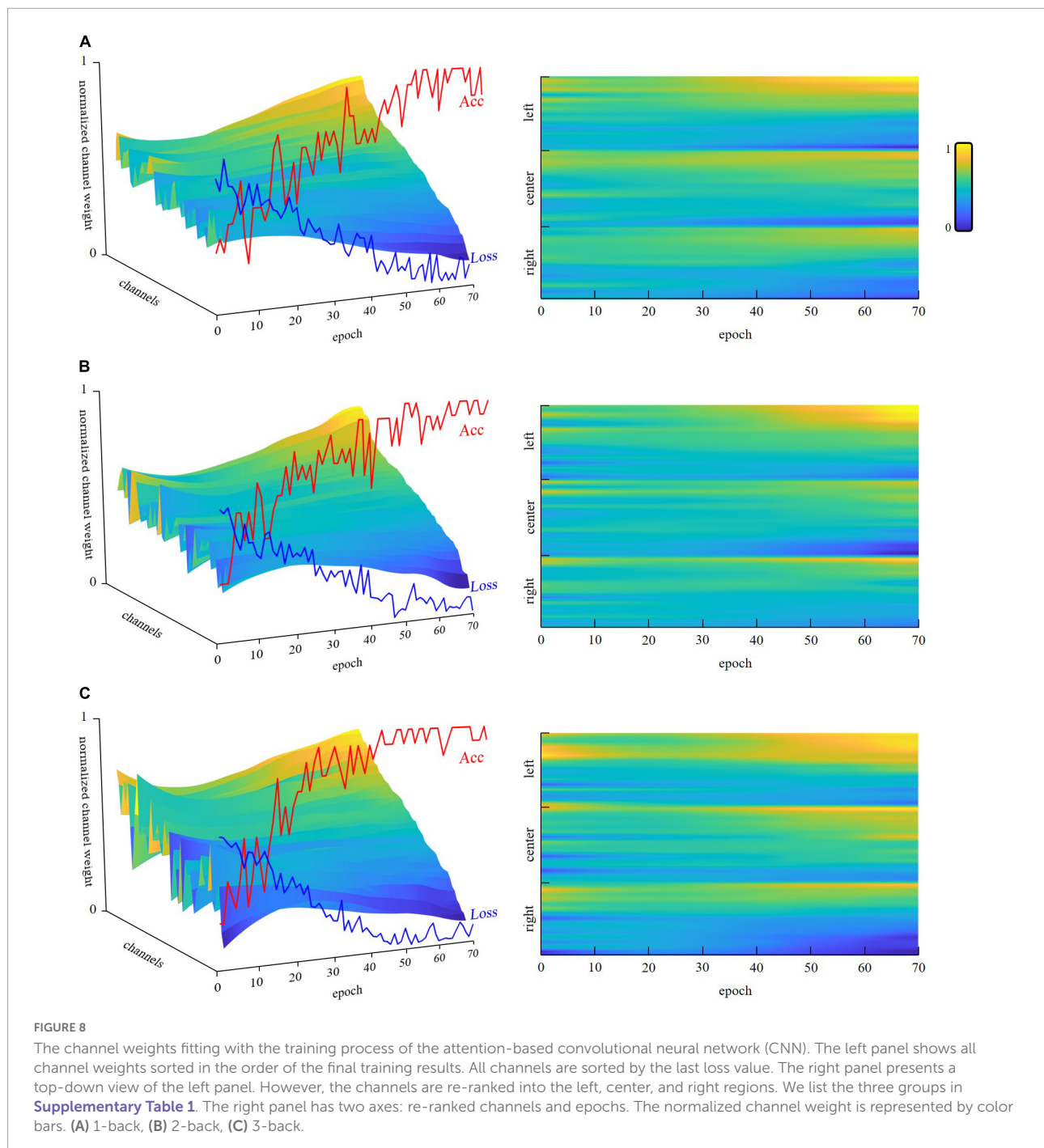
Additionally, HbR activation reduced slightly during the n-back task, as illustrated in Figure 6, and programming ability showed no main effect on HbR. This may be because relative to HbO, the HbR concentration is weak and difficult to detect in real time (Abibullaev and An, 2012), making it harder to detect significant effects on HbR activation in task-based fNIRS (Huppert et al., 2006).

Although the main effects of programming ability on HbO were significantly different in the left, middle, and right prefrontal cortices, by applying the deep learning method of attention-based CNN to the fNIRS signals, we found that the channels that can better distinguish programming ability were in the left region of the prefrontal cortex (see Figure 8). CNNs can extract microfeatures from temporal domain signals that may be corrupted by statistical analysis. CNNs can also be used to discover and extract the appropriate internal structure through convolution and pooling operations and automatically generate the deep features of the raw data. Moreover, the deep features are robust against translation and scaling (Zhao et al., 2017); they work well in discarding noisy series and can extract meaningful patterns while ignoring patterns without value (Aussem and Murtagh, 1997). By introducing attention modules (i.e., the SE block), we can open the black box to see which feature the CNN network relies on to identify those signals. As the CNN network is gradually fitted, the attention modules indicate the important channels, as shown in

Figure 8. These high-weight channels are those that the neural network uses to understand and classify. In other words, these channels and their corresponding brain regions have higher resolution in this task. The SE block can also improve the representational power of the regular CNN by offering it a kind of dynamic channel-wise fixing feature (Hu et al., 2020). Furthermore, the feature importance values produced by the self-attention operation can be used for model pruning, which can lead to the construction of more efficient physiological signal analysis networks.

The results of the current study also provided further evidence to support the lateralization of brain functions. The left prefrontal cortex was more important in programming ability prediction, as demonstrated in Figures 7, 8. Many studies have reported functional hemispheric asymmetry in cognitive processes (Gazzaniga, 1989, 1995; Goel, 2019). Smith et al. (1996) and Smith and Jonides (1997) used PET technology to study the neural basis of working memory with the n-back paradigm. Their results showed that the activation areas in the verbal and spatial n-back tasks are different: the former activates the left hemisphere, and the latter activates the right hemisphere. Baddeley and Logie (1999) also reviewed the evidence showing that the left hemisphere is associated with verbal working-memory tasks. As shown in Figure 6, our results also demonstrated that left prefrontal cortex regional activation was more dynamic during the verbal n-back test.

This study had some limitations that should be addressed in future studies. First, we did not consider the mediating factors between working memory and programming ability. Previous studies have shown that the relationship between working memory and academic performance is mediated by visuospatial abilities (Logie et al., 2000) and the ability to control attention (Kane and Engle, 2003). Future studies should aim to control these mediating factors to acquire more rigorous results. Second, the number of participants who met the criteria for advanced students was relatively small. Further studies with additional data are required to improve the generalizability of the findings.



Conclusion

To the best of our knowledge, few empirical studies have directly examined the relationship between working-memory capacity and programming ability, and no studies have provided direct neural evidence to support this relationship. The present study attempts to fill this gap and demonstrates that students' programming ability can be predicted by evaluation

of their working-memory capacity while providing direct neural evidence supporting this relationship. The results of our analyses indicate that fNIRS detected functional neural changes associated with the workload in the prefrontal cortex, demonstrating that the hemodynamic responses measured in the prefrontal cortex can be used to discriminate between novices and advanced students. Additionally, we utilized an attention-based CNN to analyze the spatial domains of the fNIRS signals and demonstrated that the left prefrontal cortex

was more important than other brain regions for programming ability prediction.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://github.com/GiantMushroom/NIRdataset.git>.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of the School of Information and Electronic Engineering, Zhejiang University of Science and Technology. The participants provided their written informed consent to participate in this study.

Author contributions

XG: methodology and writing – original draft preparation. YL: conceptualization, validation, writing – review and editing, and funding acquisition. YZ: formal analysis and writing – original draft preparation. CW: investigation and data curation. All authors contributed to the article and approved the submitted version.

References

- Abibullaev, B., and An, J. (2012). Classification of frontal cortex haemodynamic responses during cognitive tasks using wavelet transforms and machine learning algorithms. *Med. Eng. Phys.* 34, 1394–1410. doi: 10.1016/j.medengphy.2012.01.002
- Alloway, T. P. (2009). Working memory, but not IQ, predicts subsequent learning in children with learning difficulties. *Eur. J. Psychol. Assess.* 25, 92–98. doi: 10.1027/1015-5759.25.2.92
- Anmarkrud, Ø, Andresen, A., and Bråten, I. (2019). Cognitive load and working memory in multimedia learning: conceptual and measurement issues. *Educ. Psychol.* 54, 61–83. doi: 10.1080/00461520.2018.1554484
- Ashger, U., Ahmad, R., Naseer, N., Ayaz, Y., Khan, M. J., and Amjad, M. K. (2019). Assessment and classification of mental workload in the prefrontal cortex (PFC) using fixed-value modified beer-lambert law. *IEEE Access* 7, 143250–143262. doi: 10.1109/access.2019.2944965
- Aussem, A., and Murtagh, F. (1997). Combining neural network forecasts on wavelet-transformed time series. *Connect. Sci.* 9, 113–122. doi: 10.1080/095400997116766
- Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., and Onaral, B. (2012). Optical brain monitoring for operator training and mental workload assessment. *Neuroimage* 59, 36–47. doi: 10.1016/j.neuroimage.2011.06.023
- Baddeley, A. (2003). Working memory: looking back and looking forward. *Nat. Rev. Neurosci.* 4, 829–839. doi: 10.1038/nrn1201
- Baddeley, A. (2010). Working memory. *Curr. Biol.* 20, R136–R140. doi: 10.1016/j.cub.2009.12.014
- Baddeley, A. D., and Hitch, G. (1974). Working memory. *Psychol. Learn. Motiv.* 8, 47–89. doi: 10.1016/s0079-7421(08)60452-1
- Baddeley, A. D., and Logie, R. H. (1999). “Working memory: the multiple-component model,” in *Models of Working Memory*, eds A. Miyake and P. Shah (Cambridge, MA: Cambridge University Press), 28–61.
- Bai, S., Kolter, J. Z., and Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv [Preprint]*. Available online at: <https://arxiv.org/abs/1803.01271> (accessed April 19, 2018).
- Barrouillet, P., and Lépine, R. (2005). Working memory and children’s use of retrieval to solve addition problems. *J. Exp. Child Psychol.* 91, 183–204. doi: 10.1016/j.jecp.2005.03.002
- Borovykh, A., Bohte, S., and Oosterlee, C. W. (2018). Dilated convolutional neural networks for time series forecasting. *J. Comput. Finan.* 22, 73–101. doi: 10.21314/jcf.2019.358
- Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., and Noll, D. C. (1997). A parametric study of prefrontal cortex involvement. *Neuroimage* 5, 49–62.
- Buitrago Flórez, F., Casallas, R., Hernández, M., Reyes, A., Restrepo, S., and Danies, G. (2017). Changing a generation’s way of thinking: teaching computational thinking through programming. *Rev. Educ. Res.* 87, 834–860. doi: 10.3102/0034654317710096
- Cantin, R. H., Gnaedinger, E. K., Galloway, K. C., Hesson-McInnis, M. S., and Hund, A. M. (2016). Executive functioning predicts reading, mathematics, and

Funding

This research was supported by the Research Projects of the Humanities and Social Sciences Foundation of the Ministry of Education of China (grant no. 20YJA880034).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2022.1058609/full#supplementary-material>

theory of mind during the elementary years. *J. Exp. Child Psychol.* 146, 66–78. doi: 10.1016/j.jecp.2016.01.014

Chance, B., Zhuang, Z., UnAh, C., Alter, C., and Lipton, L. (1993). Cognition-activated low-frequency modulation of light absorption in human brain.pdf. *Proc. Natl. Acad. Sci. U.S.A.* 90, 3770–3774. doi: 10.1073/pnas.90.8.3770

Chen, Y., Kang, Y., Chen, Y., and Wang, Z. (2019). Probabilistic forecasting with temporal convolutional neural network. *arxiv* [Preprint]. doi: 10.48550/arXiv.1906.04397

Choi, J., Kim, J., Hwang, G., Yang, J., Choi, M., and Bae, H. (2016). Time-divided spread-spectrum code-based 400 fW-detectable multichannel fNIRS IC for portable functional brain imaging. *IEEE J. Solid State Circ.* 51, 484–495. doi: 10.1109/jssc.2015.2504412

Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., et al. (1997). Temporal dynamics of brain activation during a working memory task.pdf. *Nature* 386, 604–608. doi: 10.1038/386604a0

Dunst, B., Benedek, M., Jauk, E., Bergner, S., Koschutnig, K., Sommer, M., et al. (2014). Neural efficiency as a function of task demands. *Intelligence* 42, 22–30. doi: 10.1016/j.intell.2013.09.005

Ferrari, M., and Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *Neuroimage* 63, 921–935. doi: 10.1016/j.neuroimage.2012.03.049

Franceschini, M. A., Joseph, D. K., Huppert, T. J., Diamond, S. G., and Boas, D. A. (2006). Diffuse optical imaging of the whole head. *J. Biomed. Opt.* 11:054007.

Fritz, C. O., Morris, P. E., and Richler, J. J. (2012). Effect size estimates: current use, calculations, and interpretation. *J. Exp. Psychol. Gen.* 141, 2–18. doi: 10.1037/a0024338

Gazzaniga, M. (1989). Organization of human brain. *Science* 245, 947–952.

Gazzaniga, M. (1995). Principles of human brain organization derived from split-brain studies. *Neuron* 14, 217–228. doi: 10.1016/0896-6273(95)90280-5

Genc, E., Fraenz, C., Schluter, C., Friedrich, P., Hossiep, R., Voelkle, M. C., et al. (2018). Diffusion markers of dendritic density and arborization in gray matter predict differences in intelligence. *Nat. Commun.* 9:1905. doi: 10.1038/s41467-018-04268-8

Goel, V. (2019). Hemispheric asymmetry in the prefrontal cortex for complex cognition. *Handb. Clin. Neurol.* 163, 179–196. doi: 10.1016/B978-0-12-804281-6.00010-0

Guo, M.-H., Xu, T.-X., Liu, J.-J., Liu, Z.-N., Jiang, P.-T., Mu, T.-J., et al. (2022). Attention mechanisms in computer vision: a survey. *Comput. Vis. Media* 8, 331–368. doi: 10.1007/s41095-022-0271-y

Haier, R. J., Siegel, B. V., Nuechterlein, K. H., Hazlett, E., Wu, J. C., and Paek, J. (1988). Cortical glucose metabolic rate correlates of abstract reasoning and attention studied with positron emission tomography. *Intelligence* 12, 199–217. doi: 10.1016/j.neubiorev.2009.04.001

Herff, C., Heger, D., Fortmann, O., Hennrich, J., Putze, F., and Schultz, T. (2013). Mental workload during n-back task-quantified in the prefrontal cortex using fNIRS. *Front. Hum. Neurosci.* 7:935. doi: 10.3389/fnhum.2013.00935

Ho, T. K. K., Gwak, J., Park, C. M., and Song, J.-I. (2019). Discrimination of mental workload levels from multi-channel fNIRS using deep learning-based approaches. *IEEE Access* 7, 24392–24403. doi: 10.1109/access.2019.2900127

Hocke, L. M., Oni, I. K., Duszynski, C. C., Corrigan, A. V., Frederick, B. D., and Dunn, J. F. (2018). Automated processing of fNIRS data—a visual guide to the pitfalls and consequences. *Algorithms* 11:67. doi: 10.3390/a11050067

Hong, K.-S., and Yaqub, M. A. (2019). Application of functional near-infrared spectroscopy in the healthcare industry: a review. *J. Innov. Opt. Health Sci.* 12:1930012. doi: 10.1142/s179354581930012x

Hoshi, Y., and Tamura, M. (1993). Detection of dynamic changes in cerebral oxygenation coupled to neuronal function mental work in man. *Neurosci. Lett.* 150, 5–8. doi: 10.1016/0304-3940(93)90094-2

Hu, J., Shen, L., Albanie, S., Sun, G., and Wu, E. (2020). Squeeze-and-Excitation Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 42, 2011–2023.

Huppert, T. J., Diamond, S. G., Franceschini, M. A., and Boas, D. A. (2009). HomER: A review of time-series analysis methods for near-infrared spectroscopy of the brain. *Appl. Opt.* 48, 280–298.

Huppert, T. J., Hoge, R. D., Diamond, S. G., Franceschini, M. A., and Boas, D. A. (2006). A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. *Neuroimage* 29, 368–382. doi: 10.1016/j.neuroimage.2005.08.065

Isbilir, E., Cakir, M. P., Acarturk, C., and Tekerek, A. S. (2019). Towards a multimodal model of cognitive workload through synchronous optical brain imaging and eye tracking measures. *Front. Hum. Neurosci.* 13:375. doi: 10.3389/fnhum.2019.00375

Ivanova, A. A., Srikant, S., Sueoka, Y., Kean, H. H., Dhamala, R., O'Reilly, U. M., et al. (2020). Comprehension of computer code relies primarily on domain-general executive brain regions. *eLife* 9:e58906. doi: 10.7554/eLife.58906

Janani, A., Sasikala, M., Chhabra, H., Shajil, N., and Venkatasubramanian, G. (2020). Investigation of deep convolutional neural network for classification of motor imagery fNIRS signals for BCI applications. *Biomed. Signal Process. Control* 62:102133. doi: 10.1016/j.bspc.2020.102133

Kane, M. J., and Engle, R. W. (2003). Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. *J. Exp. Psychol. Gen.* 132, 47–70. doi: 10.1037/0096-3445.132.1.47

Khoe, H. C. H., Low, J. W., Wijerathne, S., Ann, L. S., Salgaonkar, H., Lomanto, D., et al. (2020). Use of prefrontal cortex activity as a measure of learning curve in surgical novices: results of a single blind randomised controlled trial. *Surg. Endosc.* 34, 5604–5615. doi: 10.1007/s00464-019-07331-7

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information.pdf. *J. Exp. Psychol.* 55, 352–358. doi: 10.1037/h0043688

Koh, P. H., Glaser, D. E., Flandin, G., Kiebel, S., Butterworth, B., Maki, A., et al. (2007). Functional optical signal analysis: a software tool for near-infrared spectroscopy data processing incorporating statistical parametric mapping. *J. Biomed. Opt.* 12:064010. doi: 10.1117/1.2804092

Logie, R. H., Della Sala, S., Wynn, V., and Baddeley, A. D. (2000). Visual similarity effects in immediate verbal serial recall. *Q. J. Exp. Psychol. A* 53, 626–646. doi: 10.1080/1713755916

Lv, J., Jiang, X., Li, X., Zhu, D., Chen, H., Zhang, T., et al. (2015). Sparse representation of whole-brain fMRI signals for identification of functional networks. *Med. Image Anal.* 20, 112–134. doi: 10.1016/j.media.2014.10.011

Lv, J., Jiang, X., Li, X., Zhu, D., Zhang, S., Zhao, S., et al. (2014). Holistic atlases of functional networks and interactions reveal reciprocal organizational architecture of cortical function. *IEEE Trans. Biomed. Eng.* 62, 1120–1131. doi: 10.1109/TBME.2014.2369495

Maki, A., Yamashita, Y., Ito, Y., Watanabe, E., Mayanagi, Y., and Koizumi, H. (1995). Spatial and temporal analysis of human motor activity using noninvasive NIR topography. *Med. Phys.* 22, 1997–2005. doi: 10.1118/1.597496

Matthes, K., and Gross, F. (1938). Fortlaufende registrierung der lichtabsorption des blutes in zwei verschiedenen spektralbezirken. *Nauyny Schmiedebergs Arch. Pharmacol.* 191, 381–390.

Meidenbauer, K. L., Choe, K. W., Cardenas-Iniguez, C., Huppert, T. J., and Berman, M. G. (2021). Load-dependent relationships between frontal fNIRS activity and performance: a data-driven PLS approach. *Neuroimage* 230:117795. doi: 10.1016/j.neuroimage.2021.117795

Nystrom, L. E., Braver, T. S., Sabb, F. W., Delgado, M. R., Noll, D. C., and Cohen, J. D. (2000). Working memory for letters, shapes, and locations: fMRI evidence against stimulus-based regional organization in human prefrontal cortex. *Neuroimage* 11, 424–446. doi: 10.1006/nimg.2000.0572

Owen, A. M., McMillan, K. M., Laird, A. R., and Bullmore, E. (2005). N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Hum. Brain Mapp.* 25, 46–59. doi: 10.1002/hbm.20131

Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., et al. (2020). The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. *Ann. N.Y. Acad. Sci.* 1464, 5–29. doi: 10.1111/nyas.13948

Purpura, D. J., and Ganley, C. M. (2014). Working memory and language: skill-specific or domain-general relations to mathematics? *J. Exp. Child Psychol.* 122, 104–121. doi: 10.1016/j.jecp.2013.12.009

Quaresima, V., and Ferrari, M. (2019). A mini-review on functional near-infrared spectroscopy (fnirs): where do we stand, and where should we go? *Photonics* 6:87. doi: 10.3390/photonics6030087

Ragland, J. D., Turetsky, B. I., Gur, R. C., Gunning-Dixon, F., Turner, T., Schroeder, L., et al. (2002). Working memory for complex figures: an fMRI comparison of letter and fractal n-back tasks. *Neuropsychology* 16, 370–379. doi: 10.1037/0894-4105.16.3.370

Relkin, E., de Ruiter, L. E., and Bers, M. U. (2021). Learning to code and the acquisition of computational thinking by young children. *Comput. Educ.* 169:104222. doi: 10.1016/j.compedu.2021.104222

Saadati, M., Nelson, J., and Ayaz, H. (2020). “Convolutional neural network for hybrid fNIRS-EEG mental workload classification,” in *Advances in Neuroergonomics and Cognitive Engineering*, ed. H. Ayaz (Cham: Springer), 221–232. doi: 10.3389/fnbot.2022.873239

Sassaroli, A., and Fantini, S. (2004). Comment on the modified Beer-Lambert law for scattering media. *Phys. Med. Biol.* 49, N255–N257. doi: 10.1088/0031-9155/49/14/n07

- Shute, V. J. (1995). Who is likely to acquire programming skills? *J. Educ. Comput. Res.* 7, 1–24. doi: 10.2190/vqjd-tlyd-5wvb-rypj
- Singh, A. K., and Dan, I. (2006). Exploring the false discovery rate in multichannel NIRS. *Neuroimage* 33, 542–549. doi: 10.1016/j.neuroimage.2006.06.047
- Smith, E. E., and Jonides, J. (1997). Working memory a view from neuroimaging. *Cogn. Psychol.* 33, 5–42.
- Smith, E. E., Jonides, J., and Koeppel, R. A. (1996). Dissociating verbal and spatial working. *Cerebr. Cortex* 6, 11–20. doi: 10.1093/cercor/6.1.11
- Strangman, G., Culver, J. P., Thompson, J. H., and Boas, D. A. (2002). A quantitative comparison of simultaneous BOLD fMRI and NIRS recordings during functional brain activation. *Neuroimage* 17, 719–731. doi: 10.1006/nimg.2002.1227
- Swanson, H. L., and Alloway, T. P. (2012). “Working memory, learning, and academic achievement,” in *APA Educational Psychology Handbook, Vol 1: Theories, Constructs, and Critical Issues*, eds K. R. Harris, S. Graham, T. Urdan, C. B. McCormick, G. M. Sinatra, and J. Sweller (Cham: Springer), 327–366.
- Tong, Y., Lindsey, K. P., and deB Frederick, B. (2011). Partitioning of physiological noise signals in the brain with concurrent near-infrared spectroscopy and fMRI. *J. Cerebr. Blood Flow Metab.* 31, 2352–2362.
- Trakoolwilaian, T., Behboodi, B., Lee, J., Kim, K., and Choi, J. W. (2017). Convolutional neural network for high-accuracy functional near-infrared spectroscopy in a brain–computer interface: three-class classification of rest, right-, and left-hand motor execution. *Neurophotonics* 5:011008. doi: 10.1117/1.NPh.5.1.011008
- Tu, J.-J., and Johnson, J. R. (1990). Can computer programming improve problem-solving ability? *SIGCSE Bull.* 22, 30–33.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). “Attention is all you need,” in *Proceedings of the 31st Conference on Neural Information Processing Systems*, Long Beach, CA.
- Villringer, A., Planck, J., Hock, C., Schleinkofer, L., and Dimagl, U. (1993). Near infrared spectroscopy (NIRS): a new tool to study hemodynamic changes during activation of brain function in human adults. *Neurosci. Lett.* 154, 101–104.
- Werner, L., Denner, J., Campe, S., and Kawamoto, D. C. (2012). “The fairly performance assessment: measuring computational thinking in middle school,” in *SIGCSE '12 Proceedings of the 43rd ACM Technical Symposium on Computer Science Education*, Raleigh.
- Wing, J. M. (2008). Computational thinking and thinking about computing. *Philos. Trans. A Math. Phys. Eng. Sci.* 366, 3717–3725. doi: 10.1098/rsta.2008.0118
- Wolf, M., Wolf, U., Toronov, V., Michalos, A., Paunescu, L. A., Choi, J. H., et al. (2002). Different time evolution of oxyhemoglobin and deoxyhemoglobin concentration changes in the visual and motor cortices during functional stimulation: a near-infrared spectroscopy study. *Neuroimage* 16(3 Pt 1), 704–712. doi: 10.1006/nimg.2002.1128
- Yang, D., Hong, K. S., Yoo, S. H., and Kim, C. S. (2019). Evaluation of neural degeneration biomarkers in the prefrontal cortex for early identification of patients with mild cognitive impairment: an fNIRS study. *Front. Hum. Neurosci.* 13:317. doi: 10.3389/fnhum.2019.00317
- Yang, D., Huang, R., Yoo, S. H., Shin, M. J., Yoon, J. A., Shin, Y. I., et al. (2020). Detection of mild cognitive impairment using convolutional neural network: temporal-feature maps of functional near-infrared spectroscopy. *Front. Aging Neurosci.* 12:141. doi: 10.3389/fnagi.2020.00141
- Yeung, M. K., Lee, T. L., Han, Y. M. Y., and Chan, A. S. (2021). Prefrontal activation and pupil dilation during n-back task performance: a combined fNIRS and pupillometry study. *Neuropsychologia* 159:107954. doi: 10.1016/j.neuropsychologia.2021.107954
- Yu, F., and Koltun, V. (2016). Multi-scale context aggregation by dilated convolutions. *arxiv* [Preprint]. Available online at: <https://arxiv.org/abs/1511.07122> (accessed April 30, 2016).
- Zhao, B., Lu, H., Chen, S., Liu, J., and Wu, D. (2017). Convolutional neural networks for time series classification. *J. Syst. Eng. Electron.* 28, 162–169. doi: 10.21629/jsee.2017.01.18