

2024

Analyzing human brain functional near-infrared spectroscopy data with transformer model

<https://hdl.handle.net/2144/48880>

"Downloaded from OpenBU. Boston University's institutional repository."

BOSTON UNIVERSITY
COLLEGE OF ENGINEERING

Thesis

**ANALYZING HUMAN BRAIN FUNCTIONAL
NEAR-INFRARED SPECTROSCOPY DATA WITH
TRANSFORMER MODEL**

by

SHANGZHOU YIN

B.S., Queen's University, 2022

Submitted in partial fulfillment of the
requirements for the degree of
Master of Science

2024

© 2024 by
SHANGZHOU YIN
All rights reserved

Approved by

First Reader

Xiaojun Cheng, PhD
Research Assistant Professor of Biomedical Engineering

Second Reader

Meryem Ayşe Yücel, PhD
Research Associate Professor of Biomedical Engineering

Third Reader

David A. Boas, PhD
Arthur G.B. Metcalf Chair
Professor of Biomedical Engineering
Professor of Electrical and Computer Engineering

Fourth Reader

Lei Tian, PhD
Assistant Professor of Electrical and Computer Engineering
Assistant Professor of Biomedical Engineering

Acknowledgments

I would like to say huge thank you to my advisor Prof. Xiaojun Cheng and Prof. Meryem Yücel, for providing the original idea of this research and helping me go through it. It is very grateful to have the opportunity to work in this Lab.

Second, I would also like to express my gratitude to my collaborator Teah Serani who played crucial role in this research work.

Last, I would like to thank everyone who has helped and supported me throughout this journey.

ANALYZING HUMAN BRAIN FUNCTIONAL NEAR-INFRARED SPECTROSCOPY DATA WITH TRANSFORMER MODEL

SHANGZHOU YIN

ABSTRACT

Functional Near-infrared Spectroscopy (fNIRS) is an optical neuroimaging technology measuring local cortical concentration changes of oxygenated and deoxygenated hemoglobin, which are associated with brain activities. For the robust estimation of hemodynamic brain responses, the current best practice is General Linear Model (GLM) with temporally embedded Canonical Correlation Analysis (tCCA), which has been tested to improve the performance by reducing nuisance signals from systemic physiology and motion. However, some challenging confounding signals from motions, including optode shifts or non-linear correlation between motion and physiological responses, are hard to be reduced or eliminated with traditional methods including tCCA-GLM. This study proposes a transformer-based model aiming to remove local and systemic physiological confounding signals and increase the detection accuracy of hemodynamic responses.

Contents

1	Introduction	1
1.1	Review of fNIRS' history and applications	1
1.2	The fundamental principle of fNIRS	3
1.3	Motion artifacts	5
2	Analyzing Motion-related Artifacts in fNIRS data	6
2.1	Previous Studies	6
2.2	Aim of this Study	7
3	Methods	9
3.1	Transformer	9
3.1.1	Overview of Transformer	9
3.1.2	Proposed Transformer Architecture	11
3.2	Experiments	13
3.2.1	Data Sets Preparation	13
3.2.2	Motion	15
3.2.3	Training & Fine-tuning	15
3.2.4	Metrics	16
4	Results	18
4.1	Impact of adding Motion data for model training	18
4.2	Predictions on one subject	19
4.3	Predictions on multiple subjects	20

5	Conclusions	21
A	Additional Model fine-tuning results	22
A.1	Positional Encoding	22
A.2	Segmentation Size	22
A.3	Normalization Methods	23
	References	25
	Curriculum Vitae	28

List of Figures

1·1	The current state-of-the art fNIRS systems (a) The cutting edge multi-channel fNIRS measurement system to measure adults; and (b) fNIRS systems that are commonly used for infants.	2
1·2	(a) Banana-shaped trajectory of photons from sources to detectors (Pinti et al., 2020); and (b) A typical hemodynamic response recorded by fNIRS instrument (Huppert et al., 2006).	4
3·1	Overview of neural network architectures evaluated in this study. (a) The state-of-the-art transformers model (Vaswani et al., 2017). (b) The proposed architecture used for this thesis.	11
3·2	Synthetic HRF with three different amplitudes (100/50/20%)	14
4·1	HRF predictions with different low-pass filtered motion data	18
4·2	HRF predictions using synthesized data for one subject.	19
4·3	HRF predictions using one subject	20
A·1	Impact of adding positional encoding for model(a) HRF predictions with positional encoding; and (b) HRF predictions without positional encoding.	23
A·2	Impact of using different segmentation size.	24
A·3	Impact of using different normalization methods(a) HRF predictions using min max normalization; and (b) HRF predictions using zero mean and unit variance.	24

List of Abbreviations

CCA	Canonical Correlation Analysis
EEG	Electroencephalogram
fMRI	Functional magnetic resonance imaging
fNIRS	Functional Near-Infrared Spectroscopy
GLM	General Linear Model
Hb	Hemoglobin
HbO ₂	Oxygenated Hemoglobin
HbR	Deoxygenated Hemoglobin
HbT	Total Hemoglobin
NIR	Near-Infrared
\mathbb{R}^2	the Real Plane
RNN	Recurrent Neural Network
SS	Short Separation Channels
tCCA	Temporally Canonical Correlation Analysis

Chapter 1

Introduction

1.1 Review of fNIRS' history and applications

The development of optical methods originated from a lightweight ear oxygen meter invented by Glen Mikikan in 1942 ([Chance, 1991](#)). The researchers began to explore the potential of using different wavelengths of light to analyze tissue properties. In 1977, Frans Jöbsis reported that transillumination spectroscopy can be used to monitor the hemoglobin (Hb) oxygenation in real-time due to the relatively high degree of brain tissue transparency in the near-infrared (NIR) range ([Jöbsis, 1977](#)). After conducting several experiments on laboratory animals, NIR studies were applied to newborns and adult cerebrovascular patients ([Brazy et al., 1985](#)). In the 1980s, Marco Ferrari started to measure changes of oxygen in experimental animals ([Ferrari et al., 1980](#); [Giannini et al., 1982](#)) and human adults' brains ([Ferrari et al.,](#) ; [Ferrari et al., 1985](#)). The discovery of fNIRS can be traced back to 1992 and early studies primarily focused on developing instrumentation and methodology for measuring the changes in brain tissues, starting from a single-channel system ([Ferrari and Quaresima, 2012](#)). To date, the culmination of decades of research and technological innovation has led to the realization of wearable, high-density fNIRS systems.

The past few decades have witnessed a rapid increase in the use of fNIRS in various fields, including neuroscience, cognitive science, psychology and clinical research, due to its portability, movement tolerability and safety of use as compared to Functional magnetic resonance imaging (fMRI) which requires subjects to remain stationary in

a scanner. Researchers use fNIRS to measure changes in the cortical activity during various motor tasks, ranging from finger tapping, hand movement to gait analysis (Kim et al., 2017). Many studies were conducted to explore the prefrontal cortex during cognitive analysis involving decision-making and language processing. For instance, the researchers examine the brain area associated with language comprehension and generation. It can also be used to analyze the brain activity related to the development of mathematics and language skills in schoolchildren (Soltanlou et al., 2018). This technique has been employed in the clinical perspective as well, particularly in investigating and helping diagnose neural relatedness of various mental health conditions such as Alzheimer’s disease, schizophrenia, Parkinson’s disease and children disorders etc. (Rahman et al., 2020). Recently with the portability of fNIRS, it can also be utilized to measure brain signals in the everyday world with applications such as non-invasive brain health monitoring, brain-computer interfaces, and rehabilitation. Although fNIRS has already been demonstrated to be more robust against motion artifacts than fMRI or electroencephalogram (EEG), measurements in everyday world with the subject freely moving will still create systemic physiological responses and motion artifacts, which will be the focus of this thesis.



Figure 1-1: The current state-of-the art fNIRS systems (a) The cutting edge multi-channel fNIRS measurement system to measure adults; and (b) fNIRS systems that are commonly used for infants.

1.2 The fundamental principle of fNIRS

The fNIRS as an optical imaging technique uses NIR light with wavelength of 650-950 nm to measure the concentration changes of oxygenated (HbO₂) and deoxygenated (HbR) hemoglobin in brain cortical area (Ferrari and Quaresima, 2012). The NIR light is continuously emitted by light sources (emitting optodes) through the brain tissues and collected by light detectors (detecting optodes). The typical trajectory of photons from sources to detectors can be depicted as banana-shaped. Human body is made of approximately 70% of water and its absorption is minimum within the NIR optical window, so NIR light has the ability to penetrate through the scalp, skull and into brain tissues, and is absorbed by two main chromophores (HbO₂ and HbR) existing in the blood with different absorption coefficients (Pinti et al., 2020). The scattered-back light can provide information about the amount of light absorption by chromophores using the modified Beer-Lambert Law (mBLL) (Soltanlou et al., 2018; Pinti et al., 2020). HbO₂ and HbR absorb the NIR light differently: HbO₂ absorption coefficient is higher for $\lambda > 800$ nm while HbR has a higher absorption coefficient for $\lambda < 800$ nm (Pinti et al., 2020). From the absorption coefficients of two or more wavelengths, the HbO₂ and HbR concentrations can be calculated. Mathematically, it is solving matrix equation of

$$\begin{bmatrix} \Delta OD_{\lambda_1} \\ \Delta OD_{\lambda_2} \end{bmatrix} = \frac{1}{BL} \begin{bmatrix} \epsilon_{HbO}(\lambda_1) & \epsilon_{HbR}(\lambda_1) \\ \epsilon_{HbO}(\lambda_2) & \epsilon_{HbR}(\lambda_2) \end{bmatrix}^{-1} \begin{bmatrix} \Delta[HbO] \\ \Delta[HbR] \end{bmatrix} \quad (1.1)$$

Where $\Delta OD = -\log(I/I_0)$ is the change of the optical density, I is the light intensity measured by the detector, I_0 is the baseline intensity, ϵ_{HbO} and ϵ_{HbR} are the absorption coefficients of HbO₂ and HbR, $\Delta[HbO]$ and $\Delta[HbR]$ are the changes of the concentrations of HbO₂ and HbR, λ_1 and λ_2 are the wavelengths, which are set to be 690 nm and 830 nm in this project. L is the distance between source and detector and B is a differential pathlength factor that takes into account the ratio between

the average photon pathlengths after scattering and L .

In depth, the activation of brain regions requires energy derived from glucose metabolism, which relies on the oxygen delivered by blood by combining with hemoglobin as HbO_2 (Delpy and Cope, 1997). This results in changes with an increases in HbO_2 and decreases in HbR in the locally activated region that can be utilized to measure brain activities. Therefore fNIRS measures hemodynamic responses of brain activation similar to fMRI, but is also a direct measurement of HbO_2 and HbR concentration variations.

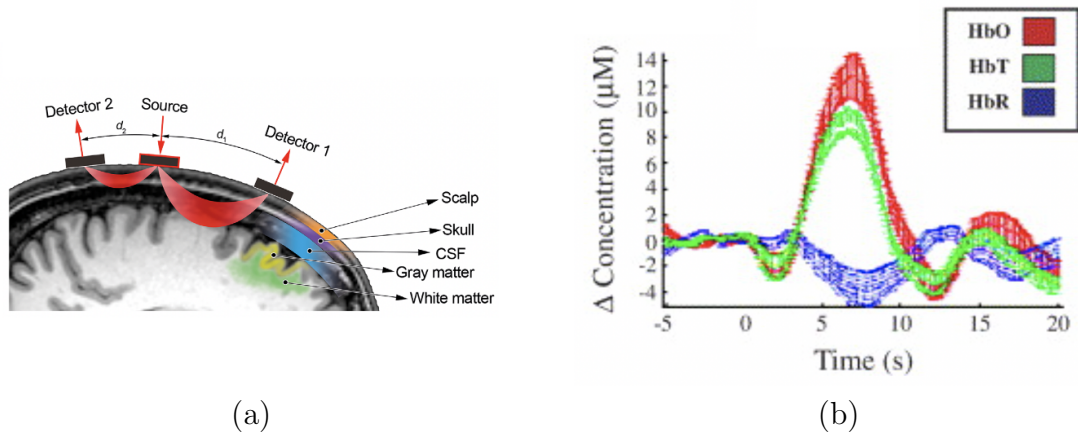


Figure 1.2: (a) Banana-shaped trajectory of photons from sources to detectors (Pinti et al., 2020); and (b) A typical hemodynamic response recorded by fNIRS instrument (Huppert et al., 2006).

A channel in fNIRS refers to a source-detector pair (Figure.1.2(a)). The portion of tissue analyzed by NIR light is positioned at the midpoint of source and detector at a depth of approximately half of the distance between source and detector (Patil et al., 2011). Several studies evaluated the spatial and depth sensitivity of fNIRS to the brain tissues as a function of different distance between sources and detectors with Monte Carlo simulations. Increasing the distance between sources and detectors will increase the depth sensitivity, but results in a lower signal-to-noise ratio (SNR) due to

fewer photons arriving at the detector at large source-detector separations ([Calderon-Arnulphi et al., 2009](#)). Standard separations are between 30-35 mm from sources and detectors are commonly used for adults. An example of a typically measured HbO₂, HbR, and total hemoglobin concentrations (HbT) which is the sum of HbO₂ and HbR and is proportional to the cerebral blood volume is shown in [Figure.1.2\(b\)](#).

1.3 Motion artifacts

Motion artifacts significantly impact the quality of fNIRS measured signals. One primary issue arises from the uncoupling of the source or detector from the skin, leading to the abrupt increase or decrease in measured light attenuation. Additionally, gravitational effects might alter blood flow or volume within different regions of the head ([Robertson et al., 2010](#)). Motion artifacts are often a significant component in fNIRS signal, which can be categorized into three types, spikes, baseline-shifts and low frequency variations. Motion artifacts varies in shapes, frequency content and timing; they can be easily detected with high amplitude, high frequency spikes while can be harder to be extracted from the HRF with low frequency content ([Brigadoi et al., 2014](#)).

Chapter 2

Analyzing Motion-related Artifacts in fNIRS data

2.1 Previous Studies

One of accelerometer-based methods introduced an active-noise cancellation method which assumes that the measured signals are the combination of motionless signals and motion related artifacts ([Kim et al., 2011](#)). The goal of this method is to reduce the power of the recovered signals. This method is applied to optical intensities in real time, but it is not clear to apply it to optical densities or other metrics. Another accelerometer-based method is accelerometer-based motion artifact removal (ABAMAR) which is an offline analysis method that the accelerometer outputs are used for motion related artifacts detection and removal process is dependent on the measured fNIRS signals ([Virtanen et al., 2011](#)). This solution is used for optical intensities, optical densities, and concentration changes and it can effectively reduce step-like artifacts but some signal details during motion events will be lost due to its correction method. Moreover, linearly polarized light sources and orthogonally polarized analyzers was proposed to eliminate the hair-reflected light caused by the optode fluctuation during tasks involving movement. ([Yamada et al., 2015](#)).

Later, more advanced solutions were introduced without additional hardware. Wiener filter assumes the fNIRS are a simple addition between the actual fNIRS signals and motion artifacts and they are stationary and uncorrelated, but it re-

quires the prior knowledge of the power spectral densities (PSDs) (Izzetoglu et al., 2005). The Kalman filter is derived from the Wiener filter, utilizing the auto-regressive model (AR) but cannot handle non zero-mean Gaussian measurement noise (Izzetoglu et al., 2010). Machine learning methods are also explored for HRF extraction and motion-related artifacts correction for fNIRS measurements (von Lühmann et al., 2020b). General linear model is an optimal method treating signals as a combination linear functions and noise, which recovers the estimates of evoked hemodynamic response function from observed Long-Separation (LS) fNIRS measurements. Short-Separation (SS) signals are incorporated in General Linear Model (GLM) as physiological noise regressors for increasing the accuracy of estimating HRF. (von Lühmann et al., 2020c). However, some signal characteristics such as non-instantaneous or constant non-coupling are not well resolved and auxiliary signals are not explored by this approach. Recently, temporally embedded Canonical Correlation Analysis (tCCA) is incorporated in the supervised GLM. SS signals and other auxiliary signals are better exploited with the help of tCCA (von Lühmann et al., 2020b).

2.2 Aim of this Study

fNIRs is a non-invasive, portable technique for functional monitoring and imaging of human brain hemodynamics. Two primary types of noise can contaminate the cerebral hemodynamics: physiological and non-physiological noise. Physiological noise includes the systemic interference arised from the changes in the blood pressure such as cardiac, respiration, Mayer waves, and low-frequency oscillations, while non-physiological noise involves motion artifacts from optode-decoupling and instrumental noise (von Lühmann et al., 2020b). The current best practice, GLM with tCCA, cannot address some non-linear issues caused by motion related artifacts such as optode shifts or non-linear correlations between motion and physiological responses. A great

variety of signal processing and time series forecasting problems employ the Deep Neural Networks, However, they are rarely incorporated in the fNIRS or related area. Transformer, as an innovative model, particularly its strong ability to capture the complex relationship in sequential data through the self-attention mechanism. This mechanism enables to identify the correlation between fNIRS data and HRF. Therefore, Transformer stands out as a perfect candidate to extract the HRF from fNIRS data.

Chapter 3

Methods

3.1 Transformer

3.1.1 Overview of Transformer

The Transformer model stands at the cutting-edge of Deep Neural Networks technology, developed by Google. It was originally designed to handle sequentially ordered input data (e.g. text, image). Transformer has advantage over traditional deep learning architectures like Recurrent Neural Networks(RNNs). Its key strength lies in the utilization of attention mechanism, as detailed by ([Vaswani et al., 2017](#)), which has the strong ability to capture complex relationship within data and improve the efficiency to understand the contexts and significance of each token relative to other tokens through parallel token processing.

Transformers usually consist of encoders and decoders structures. The encoder layer processes the input sequence and generates the representations that capture the meanings of tokens in input and the relevance of each tokens to the other tokens, while the decoder layers take all the representations generated by the encoders and use their contextual information to generate the ordered output sequence ([Katrompas et al., 2022](#)).

The attention mechanism is the key to achieve this for both encoders and decoders, which allows each token to attend to all other tokens in the same sequence, regardless of their positions, and compute the attention scores of all pairs of tokens.

The attention score is explained as

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3.1)$$

where Query (Q), Key(K) and Value(V) are three vectors created by multiplying the token embedding by three weighted matrices that are randomly initialized and trained during the training process. Attention scores are calculated by taking the dot product of Q and K for each embedding and dividing by the square root of the dimension of K to stabilize gradients during training. Softmax function is then applied to calculate the probability distribution. Multiplying the distribution by each V and summing up the weighted V to yield attention score.

The principle of multi-head attention is to split the Q , K and V into h matrices, corresponding to h heads, the attention function 3.1 is then applied to each matrix in parallel. The resulting matrices are concatenated and projected to yield the final attention score, as shown in 3.2.

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O \quad (3.2)$$

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (3.3)$$

Where matrices $W_i^Q \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^K \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^V \in \mathbb{R}^{d_{model} \times d_v}$, $W^O \in \mathbb{R}^{h d_v \times d_{model}}$ and dimension of input representation is denoted as d_{model} .

The transformer does not understand and process the sequential data in order at the encoder, so the positional encoding is added to provide the position information to maintain the order in output. The feedforward neural network is applied independently to each position in the sequence within the encoder and decoder, with two linear transformations and a non-linear activation function. Residual connections and layer normalization are employed to facilitate the training by allowing gradients to

flow easily and stabilizing the training.

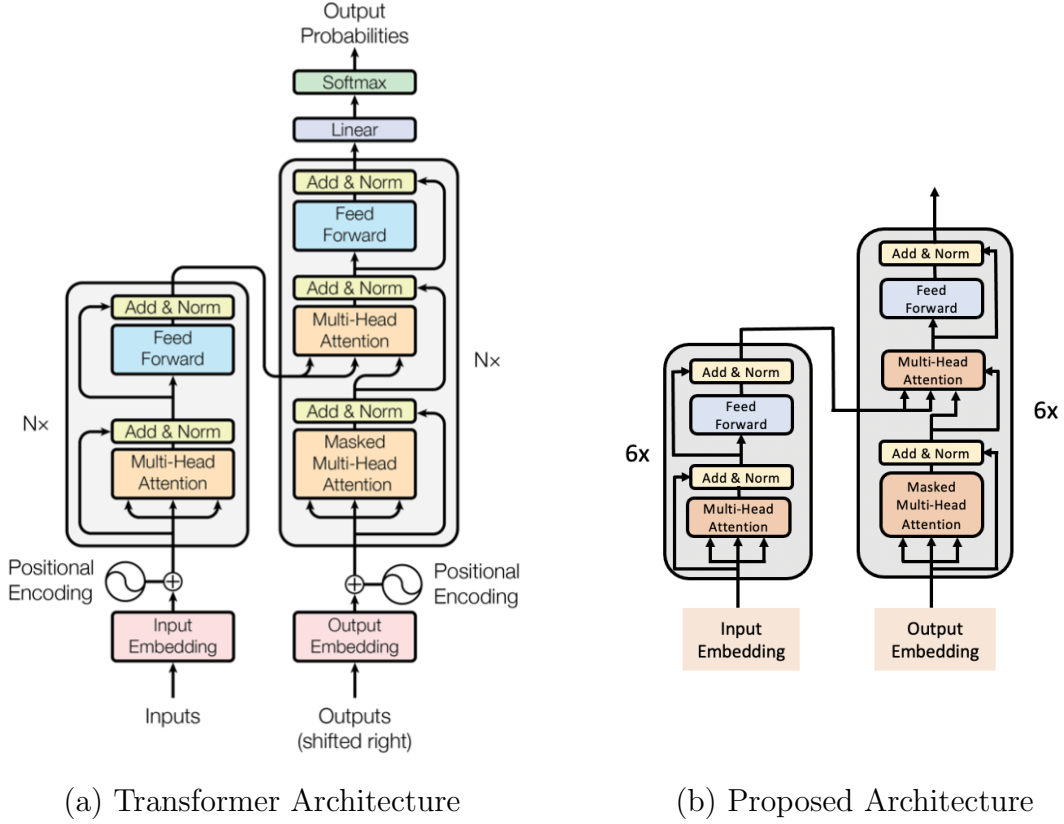


Figure 3.1: Overview of neural network architectures evaluated in this study. (a) The state-of-the-art transformers model (Vaswani et al., 2017). (b) The proposed architecture used for this thesis.

3.1.2 Proposed Transformer Architecture

The purpose of utilizing the transformer-based architecture for this study is its potential ability of learning tokens and valuing the importance of each token in the sequence. In model design we follow the original Transformer architecture. This study primarily focuses on discovering the relationship between fNIRS measurement data and synthetic HRF, predicting the HRF given fNIRS measurement data and comparing it with the tCCA-GLM method. A time-series sequence to sequence transformer

is built based on the transformer described above for the fNIRS measurement data. The architecture eliminates the input embedding, output embedding, positional encoding, linear and softmax layer. The encoder and decoder architectures are identical to the original architecture shown in Figure 3.1.

Input embedding and output embedding are used to represent the words for input and output sequence to facilitate language translation (Vaswani et al., 2017). However, the input and output sequence (respectively, fNIRS measurement data and synthetic HRFs) are time-series signal measurement appearing as a 1D sequence of real numbers, so the embedding does not convert the data into meaningful information but rather adds complexity to the model.

The relative and/or absolute positional encoding is designed to maintain the order within a sequence and its output, and to distinguish the same tokens based on their positions in different sequence samples. However, each token represents a single time-stamp for the signal measurement and may appear only once or multiple times within the sequence; the positional encoding does not provide any meaningful information but increases the model complexity by adding additional features for the training process. The detailed explanation can be found in Appendix A.

Linear layer is designed to project the high-dimensional representation into a space that matches the desired output dimensionality. The softmax layer follows the linear layer to compute the probability of each token being the next token in the output sequence. These two steps enable the transformer model to generate the next token in the output sequence with the highest probability. However, since this study handles the real numbers, it is unnecessary to compute probabilities of a signal time-stamp data point as the next possible token.

3.2 Experiments

3.2.1 Data Sets Preparation

This section is to describe the process of dataset creation for transformer model training, validation and testing. The process involves two major steps: adding synthetic HRF to fNIRS data and segmenting and normalizing data using python. The existing Dataset consists of 10-min resting state data from 14 healthy participants(von Lühmann et al., 2020a). The optode array for Dataset has 16 sources, 24 long-separation detectors and 8 short-separation detectors covering the head from frontal to parietal regions bilaterally.

Adding synthetic HRFs to the fNIRS data

The Synthetic HRF shown in Figure 3.2 is generated at three amplitudes (100/50/20%) using a gamma function with a time-to-peak of 6s and a total duration of 16.5s. In this study, the HRF with an amplitude of 100% is used for the Transformer. The shape of the HRF is represented by a variant gamma function (Huppert et al., 2006), expressed as follows:

$$hrf(t) = \sum_{h=1}^H b_i(t - \Delta t \cdot h)\beta_h \quad (3.4)$$

All recorded channels were converted to optical density and subsequently into concentration. A lowpass filter was applied to the concentration. The channels were segmented into intervals of 60 seconds each. Random onsets ranging from 0 to 15 seconds were chosen and the HRF was convolved at each onset.

Segmenting and normalizing processed fNIRS data

The generated datasets were saved in the comma-separated value format(CSV) files, which consists of 56 channels of filtered concentration of fNIRS measurements with corresponding synthetic HRF but random onsets. Due to limited resting state data

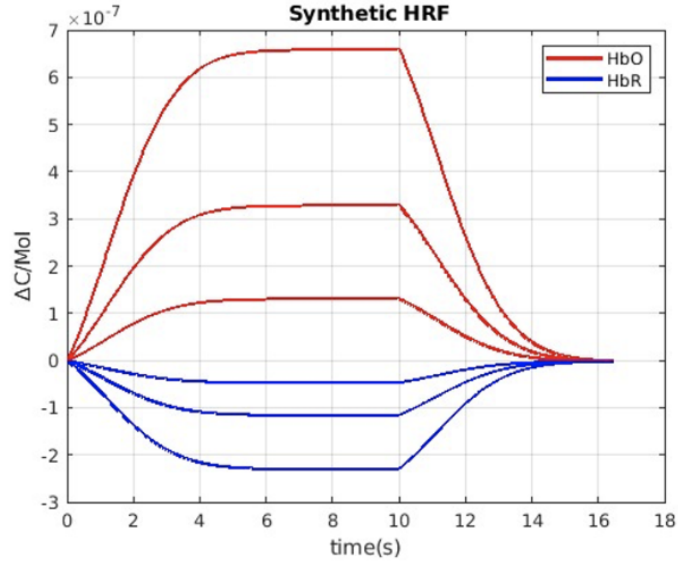


Figure 3-2: Synthetic HRF with three different amplitudes (100/50/20%)

collection for each subject, we augmented and created 4 variations for each subject with same channel data but random onsets of HRF. Since the data collection for each subject did not occur at precisely the same time point, all channels were trimmed to match the shortest one in order to maintain uniform length across all subjects' channels. The fNIRS data and synthetic HRF were then concatenated for each channel separately, serving as the input and output sequence respectively. The next step involves splitting the data into segments or tokens in preparation for training. The detailed explanation of choosing size for segments can be found in Appendix A. The values in the processed data of fNIRS data and synthetic HRFs are roughly 10^{-7} , which are relatively small for the model, leading to inefficiency. MinMax Scalar normalization was adopted for data preprocessing to rescale all the data to fall between 0 and 1, which is explained in equation 3.5.

$$x_{\text{rescaled}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (3.5)$$

3.2.2 Motion

Head motions were simultaneously collected in x , y , and z directions using a 3-axis accelerometer secured on the head with a headband. Motion data(M) was considered to assist in effectively discovering similar motion patterns between the fNIRS data and HRF for better extraction using Transformer. To implement the motion data, we calculated the magnitude of motion as a second input sequence for Transformer. It is expressed as

$$M = \sqrt{(x^2 + y^2 + z^2)} \quad (3.6)$$

3.2.3 Training & Fine-tuning

The input sequence and the ground truth are 1D signals. To handle the large number of timestamp data points in each signal while requiring fewer computation resources, we segmented the signals into $x \in \mathbb{R}^{D \times N}$, where D represents the segmentation size, and n represents the number of segments. Since we used the multihead attention mechanism in the training, the signal is expressed as $x \in \mathbb{R}^{h \times d \times N}$ where h represents the number of parallel attention layers, or head d is the segmentation size in each head. D also serves as the effective input sequence dimension for the transformer. The dimension D of 5000 is used consistently throughout the Transformer to ensure that the size of the input and output vectors remain constant through all of its layers.

In this study, we employed a configuration consisting of 6 encoders, 6 decoders and 8 heads for the multihead attention mechanism. We experimented with the number of encoders and decoders(i.e. 2, 3 and 4), and found that 6 for each yielded the best performance.

Adaptive Moment Estimation(Adam) optimizer was selected for this study, which is one of the most commonly used optimizers to train Transformers because it achieves

good convergence by storing rolling average of the previous gradients. Through experimentation, we found the optimal settings to be $\beta_1 = 0.9$, $\beta_2 = 0.98$, $l_{rate} = 0.0001$ and $\epsilon = 10^{-9}$.

Early stopping mechanism was adopted by using a patience parameter for monitoring the validation loss during training and stopping the training process when the metric starts to degrade, preventing overfitting and improving generalization.

The loss function employed the Mean Square Error (MSE) for sequence data prediction, which calculates the difference between predictions and ground truth. Mathematically, it is expressed as

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (3.7)$$

where N is the number of samples, y_i are observed values and \hat{y}_i are predicted values.

We maintained the rest of the configurations as described in (Vaswani et al., 2017). We followed the original regularization for the transformer by applying dropout to the output of each sub-layer before it is added to the sub-layer input and normalized, with a dropout rate of $P_{drop} = 0.1$.

3.2.4 Metrics

Root Mean Square Error (RMSE) and Pearson Correlation Coefficient(r) were used to evaluate the difference between ground truth and the predictions.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (3.8)$$

where \hat{y}_i are predicted values, y_i are observed values and n is the number of observations.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (3.9)$$

where x_i, y_i are observed values, \bar{x}, \bar{y} are mean of the observed values.

Chapter 4

Results

4.1 Impact of adding Motion data for model training

We selected a single subject from the Dataset to explore the efficiency of motion data on HRF predictions. We tested the efficiency of motion data in HRF predictions. In principle, motion data should be helpful in identifying the correlation between motion recorded by IMUs and motion-related fNIRS signals. A low-pass filter with different cutoff frequency is used, implemented using the HOMER3 built-in function, aimed to reduce noise. However, as shown in Figure 4.1, the predictions contain a significant amount of noise, and both the onsets and amplitude of HRF are off. This suggests that noise in the motion data has propagated to the result through the transformer model. Therefore, we decided to not use the motion data in this study.

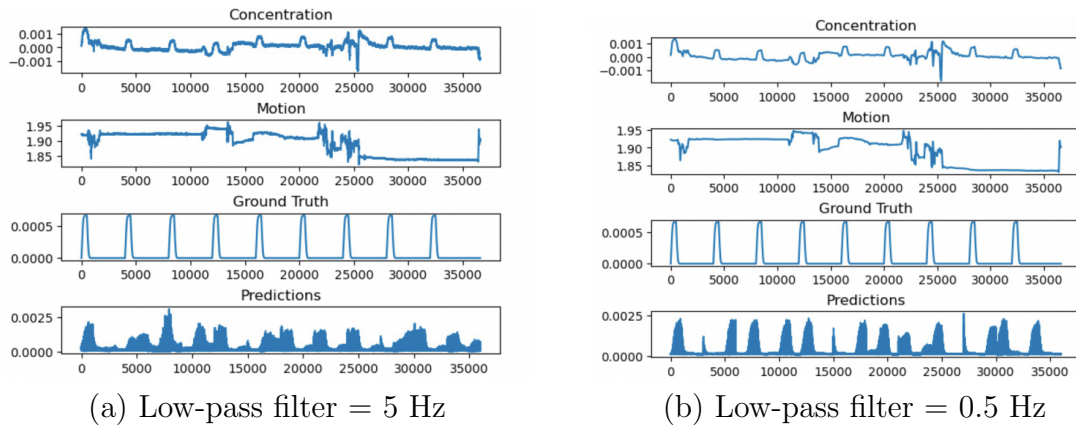


Figure 4.1: HRF predictions with different low-pass filtered motion data

4.2 Predictions on one subject

We selected a single subject from the Dataset to explore the capability of our proposed model. Given the limited data size of subjects, we generated 14 variations of this subject. These variations maintained the same channels but introduced different random onsets. Each variation was split into train, validation and test sets and subsequently all variations were concatenated. In figure 4.2, the predictions are almost perfectly aligned with the ground truth, especially the onsets and amplitudes. This demonstrates the potential of the transformer model. But this result could indicate overfitting. One possible reason is that even though we trained and tested on different channels, some sources and detectors used for training are physically close to those used for testing on the fNIRS device, and this can lead to strong correlations between signals for these channels. The other possible reasons for overfitting are small datasets and high complexity of Transformer model. Therefore, from this figure, we can see the potential of Transformer model for better extraction of HRF. For these results the RMSE for all channels is 0.27, and the correlation is 0.67.

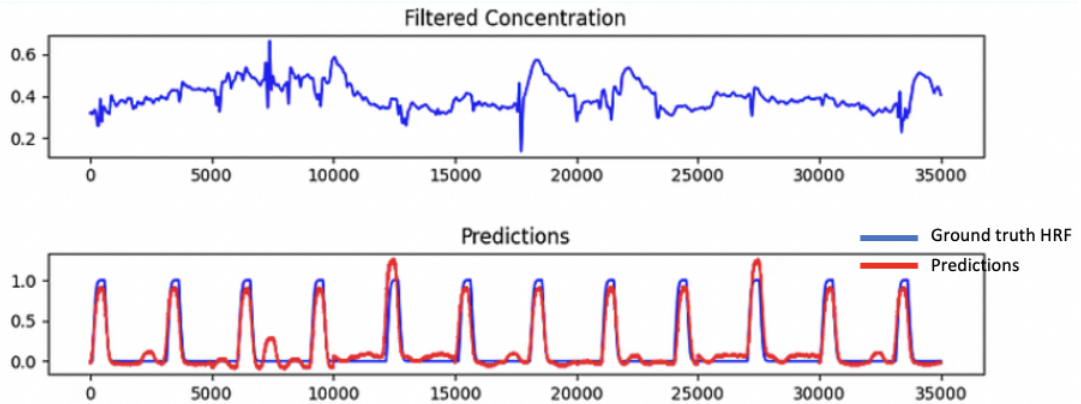


Figure 4.2: HRF predictions using synthesized data for one subject.

4.3 Predictions on multiple subjects

The dataset generation is mentioned in Section 3.2.1. We trained the model with 10 subjects. To improve the robustness and diversity of the dataset, each subject was expanded with 4 variations. We then test on a subject that was not included in the training. This approach was implemented with the aim of mitigating the risk of overfitting. In figure 4.3, we can see the HRF predictions does not closely align with the ground truth, particularly the amplitudes of HRF. However, the model demonstrated the capability of capturing the timing of onsets. The predictions of the onset positions are almost perfectly aligned with the ground truth. For these results the RMSE for all channels is 0.51, and the correlation is 0.34.

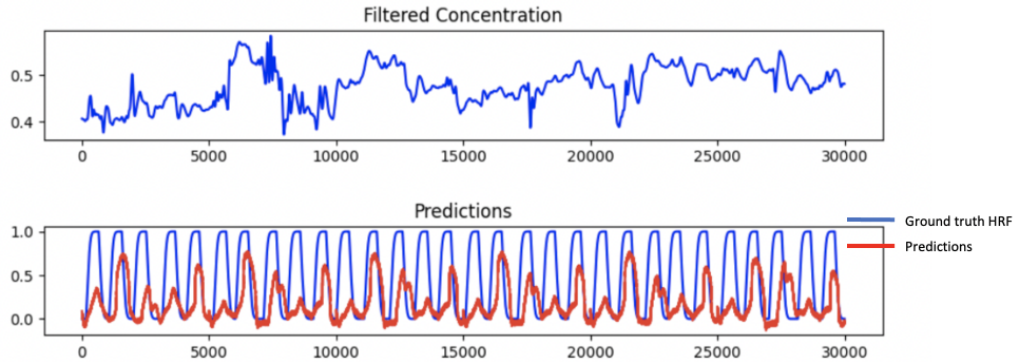


Figure 4.3: HRF predictions using one subject

Chapter 5

Conclusions

This study introduced a Transformer-based architecture for predicting HRF from fNIRS data. Results from using motion data to improve the prediction accuracy indicated the motion data does not contribute to the model. Furthermore, this study explored the model's performance across the different scales of data, from small to large datasets. These experiments demonstrated the potential of the proposed model in effectively capturing and predicting HRF, highlighting its applicability and effectiveness in the field. Future work should further collect various data type, conditions and experimental settings, including but not limited to physical activities such as walking and running. Moreover, fine-tuning the models with varied datasets can further enhance the predictive capabilities.

Appendix A

Additional Model fine-tuning results

A.1 Positional Encoding

A single subject from the Dataset was selected to explore the necessity of positional encoding. Positional Encoding was added to each segments after input embedding shown in [3.1](#) and before sending to the model. Mathematically, it is explained as

$$PE_{\text{pos},2i} = \sin(\text{pos}/10000^{2i/d_{\text{model}}}) \quad (\text{A.1})$$

$$PE_{\text{pos},2i+1} = \cos(\text{pos}/10000^{2i/d_{\text{model}}}) \quad (\text{A.2})$$

where pos is the position and i is the dimension. Every dimension of the positional encoding is represented by a sinusoid function.

From Figure [A.1](#), the HRF prediction has more noise, so we decided to not use positional encoding.

A.2 Segmentation Size

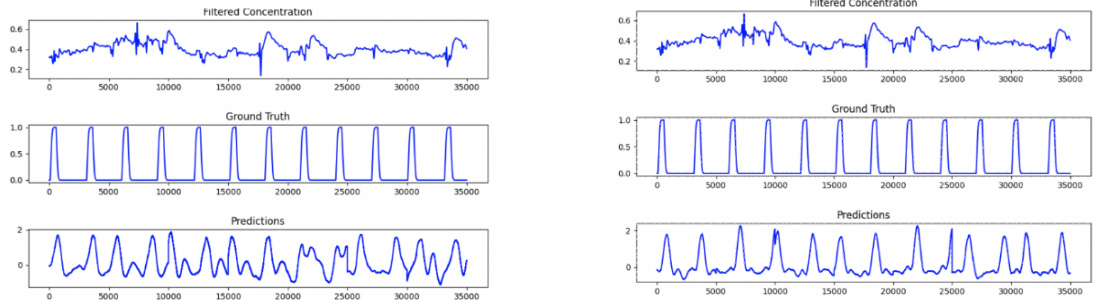
A single subject was chosen to explore the impact of changing segmentation size on prediction accuracy. We experimented with segmentation size of 512, 1000, 2000, 3000, 4000, 5000, 6000, 7000, and found the model performs best with a segmentation size of 5000. Some experiments examples results shown in [A.2](#)

A.3 Normalization Methods

In this study, we compared two common normalization methods for data preprocessing using a single subject: min-max normalization and zero mean and unit variance normalization. Min-max normalization has been explained in 3.5. Mathematically, zero mean and unit variance can be expressed as

$$X_{\text{rescaled}} = \frac{x - \mu}{\sigma} \quad (\text{A.3})$$

where μ is mean of observed data and σ is standard deviation of observed data. Through these two figures, min-max normalization performs better as it is closer to the ground truth.



(a) Prediction with positional encoding (b) Prediction without positional encoding

Figure A.1: Impact of adding positional encoding for model(a) HRF predictions with positional encoding; and (b) HRF predictions without positional encoding.

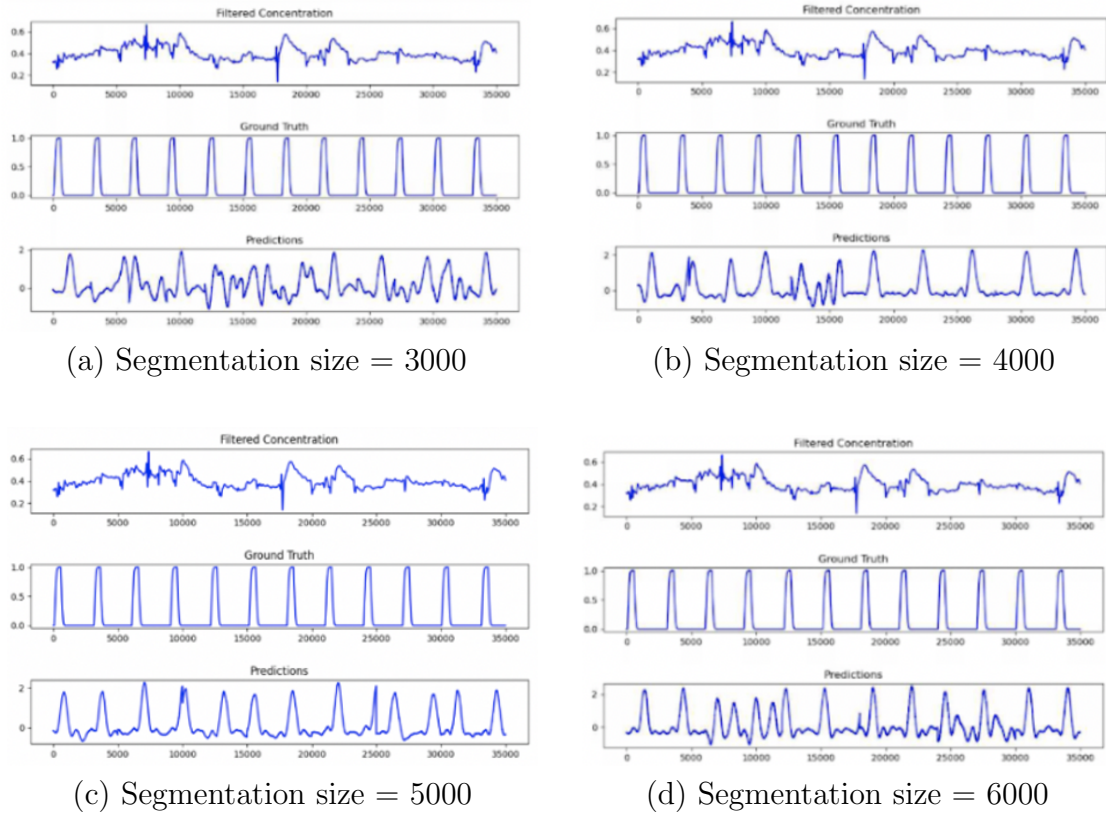


Figure A.2: Impact of using different segmentation size.

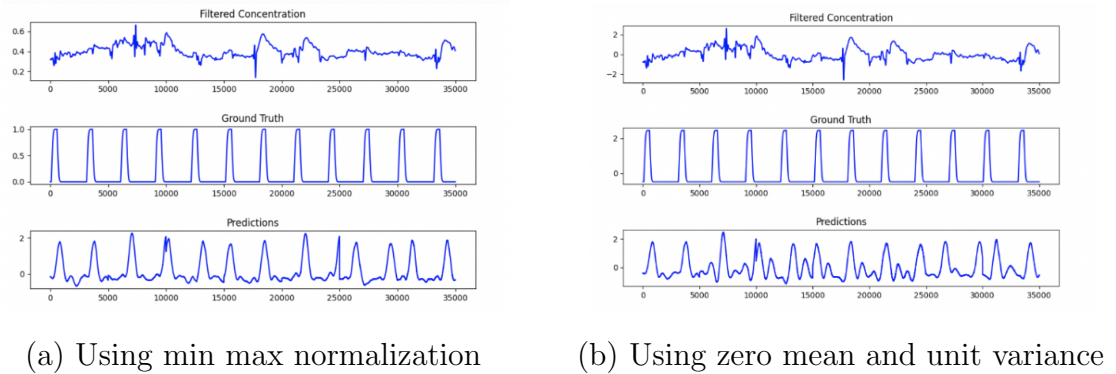


Figure A.3: Impact of using different normalization methods(a) HRF predictions using min max normalization; and (b) HRF predictions using zero mean and unit variance.

References

- Brazy, J. E., Lewis, D. V., Mitnick, M. H., and vander Vliet, F. F. J. (1985). Non-invasive monitoring of cerebral oxygenation in preterm infants: preliminary observations. *Pediatrics*, 75(2):217–225.
- Brigadoi, S., Ceccherini, L., Cutini, S., Scarpa, F., Scatturin, P., Selb, J., Gagnon, L., Boas, D. A., and Cooper, R. J. (2014). Motion artifacts in functional near-infrared spectroscopy: a comparison of motion correction techniques applied to real cognitive data. *Neuroimage*, 85:181–191.
- Calderon-Arnulphi, M., Alaraj, A., and Slavin, K. V. (2009). Near infrared technology in neuroscience: past, present and future. *Neurological research*, 31(6):605–614.
- Chance, B. (1991). Optical method. *Annual review of biophysics and biophysical chemistry*, 20(1):1–30.
- Delpy, D. and Cope, M. (1997). Quantification in tissue near-infrared spectroscopy. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 352(1354):649–659.
- Ferrari, M., Giannini, I., Carpi, A., and Fasella, P. Near i.r. spectroscopy in non invasive monitoring of cerebral function. In *Proc. World Congress on Medical Phys. and Biomed. Eng 1982, Hamburg, September 5–11, Ed. by Bleifeld W., Harder D., Leetz H.K. and Schaldach M., MPBE 1982 e. V., Hamburg, abs. 22.17.*
- Ferrari, M., Giannini, I., Carpi, A., Fasella, P., Fieschi, C., and Zanette, E. (1980). Non invasive infrared monitoring of tissue oxygenation and circulatory parameters. In *XII World Congress of Angiology, Athens, September 7–12, abs. 663.*
- Ferrari, M., Giannini, I., Sideri, G., and Zanette, E. (1985). Continuous non invasive monitoring of human brain by near infrared spectroscopy. In *Oxygen transport to tissue VII*, pages 873–882. Springer.
- 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(2):921–935.
- Giannini, I., Ferrari, M., Carpi, A., and Fasella, P. (1982). Rat brain monitoring by near-infrared spectroscopy: an assessment of possible clinical significance. *Physiological chemistry and physics*, 14(3):295–305.

- 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(2):368–382.
- Izzetoglu, M., Chitrapu, P., Bunce, S., and Onaral, B. (2010). Motion artifact cancellation in nir spectroscopy using discrete kalman filtering. *Biomedical engineering online*, 9:1–10.
- Izzetoglu, M., Devaraj, A., Bunce, S., and Onaral, B. (2005). Motion artifact cancellation in nir spectroscopy using wiener filtering. *IEEE Transactions on Biomedical Engineering*, 52(5):934–938.
- Jöbsis, F. F. (1977). Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. *Science*, 198(4323):1264–1267.
- Katrompas, A., Ntakouris, T., and Metsis, V. (2022). Recurrence and self-attention vs the transformer for time-series classification: a comparative study. In *International Conference on Artificial Intelligence in Medicine*, pages 99–109. Springer.
- Kim, C.-K., Lee, S., Koh, D., and Kim, B.-M. (2011). Development of wireless nirs system with dynamic removal of motion artifacts. *Biomedical Engineering Letters*, 1:254–259.
- Kim, H. Y., Seo, K., Jeon, H. J., Lee, U., and Lee, H. (2017). Application of functional near-infrared spectroscopy to the study of brain function in humans and animal models. *Molecules and cells*, 40(8):523–532.
- Patil, A. V., Safaie, J., Moghaddam, H. A., Wallois, F., and Grebe, R. (2011). Experimental investigation of nirs spatial sensitivity. *Biomedical optics express*, 2(6):1478–1493.
- Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., and Burgess, P. W. (2020). The present and future use of functional near-infrared spectroscopy (fnirs) for cognitive neuroscience. *Annals of the new York Academy of Sciences*, 1464(1):5–29.
- Rahman, M. A., Siddik, A. B., Ghosh, T. K., Khanam, F., and Ahmad, M. (2020). A narrative review on clinical applications of fnirs. *Journal of Digital Imaging*, 33(5):1167–1184.
- Robertson, F. C., Douglas, T. S., and Meintjes, E. M. (2010). Motion artifact removal for functional near infrared spectroscopy: a comparison of methods. *IEEE Transactions on Biomedical Engineering*, 57(6):1377–1387.
- Soltanlou, M., Sitnikova, M. A., Nuerk, H.-C., and Dresler, T. (2018). Applications of functional near-infrared spectroscopy (fnirs) in studying cognitive development: The case of mathematics and language. *Frontiers in psychology*, 9:311631.

- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. https://papers.nips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.
- Virtanen, J., Noponen, T., Kotilahti, K., Virtanen, J., and Ilmoniemi, R. J. (2011). Accelerometer-based method for correcting signal baseline changes caused by motion artifacts in medical near-infrared spectroscopy. *Journal of biomedical optics*, 16(8):087005–087005.
- von Lühmann, A., Li, X., Gilmore, N., Boas, D. A., and Yücel, M. A. (2020a). Open access multimodal fnirs resting state dataset with and without synthetic hemodynamic responses. *Frontiers in Neuroscience*, 14:579353.
- von Lühmann, A., Li, X., Müller, K.-R., Boas, D. A., and Yücel, M. A. (2020b). Improved physiological noise regression in fnirs: a multimodal extension of the general linear model using temporally embedded canonical correlation analysis. *NeuroImage*, 208:116472.
- von Lühmann, A., Ortega-Martinez, A., Boas, D. A., and Yücel, M. A. (2020c). Using the general linear model to improve performance in fnirs single trial analysis and classification: a perspective. *Frontiers in human neuroscience*, 14:30.
- Yamada, T., Umeyama, S., and Ohashi, M. (2015). Removal of motion artifacts originating from optode fluctuations during functional near-infrared spectroscopy measurements. *Biomedical optics express*, 6(12):4632–4649.

CURRICULUM VITAE

