TIPTEKNO'24 MEDICAL TECHNOLOGIES CONGRESS
10-12 October 2024
Yasmin Bodrum Resort, Muğla, Türkiye

Comparison of Machine Learning Algorithms for Yes/No Decoding Using Functional Near-Infrared Spectroscopy (fNIRS)

Seray Erdoğan, Selin Ergün, Hande Giregiz, Bora Mert Şahin, Aykut Eken
Department of Biomedical Engineering
TOBB University of Economics and Technology
Ankara, Turkey
{seray.erdogan, selinergun, hgiregiz, bsahin, aykuteken}@etu.edu.tr

***Abstract*—Brain-computer interface (BCI) can be an alternative to speech production for people with disabilities. More recently, a non-invasive optical technique called functional near-infrared spectroscopy (fNIRS) has gained popularity in BCI studies due to several advantages such as high mobility, being inexpensive, and being tolerant to motion artifacts. In this study, we compared the performance of machine learning algorithms to decode fNIRS signals acquired during a binary decision paradigm for motor-independent communication. Twenty healthy participants were asked to perform mental arithmetic tasks for the "yes" decision and the rest for the "no" decision. Three trials for each decision were conducted and oxyhemoglobin concentration changes were used to classify the decision using machine learning algorithms: linear support vector machine (SVM), logistic regression (LR), naive Bayes, and k-nearest neighbors (KNN). We observed subject-wise average accuracies across twenty participants, with the logistic regression classifier achieving an average accuracy of 80.65% for training, 83.81% for validation, and 82.75% for testing. Similarly, the linear SVM classifier achieved an average accuracy of 83.89% for training, 82.34% for validation, and 81.67% for testing. Our findings suggest that both logistic regression and linear SVM classifiers, in combination with fNIRS, have the potential to be used in the binary classification with individuals having motor disabilities.**

Keywords—functional near-infrared spectroscopy; machine learning; brain-computer interface; yes/no decoding; mental arithmetics; binary communication

I. INTRODUCTION

Communication is vital for effective social interaction, and motor activity plays an essential role in maintaining the integrity of communication. Dysfunctions in the motor system can cause significant challenges in a person's ability to communicate with the outside world [1]. An alternative communication method using a brain-computer interface (BCI) can benefit individuals with motor control disabilities, such as amyotrophic lateral sclerosis (ALS), and those in a locked-in state, where a person is fully aware and awake but paralyzed [1]. A BCI is a communication system that uses brain waves to control or interact with computers or other desired devices [2]. Electroencephalogram (EEG) devices have long been used to develop non-invasive BCI systems. However, the low spatial resolution of EEG hinders source localization [3], and not everyone becomes adept at using EEG-based BCI, a phenomenon called BCI illiteracy [4], which led to the need for alternative instruments. BCI systems can also be developed using other neuroimaging techniques, such as functional magnetic resonance imaging (fMRI) and functional near-infrared spectroscopy (fNIRS). While fMRI offers higher spatial resolution than fNIRS and EEG, it is costly and immobility limits daily use [4]. In contrast, fNIRS is less expensive, has greater tolerance to motion artifacts, and allows for better mobility [5].

The majority of fNIRS-based BCI studies focus on binary communication, which includes methods such as mental arithmetic and motor imagery [3]. In this study, we conducted a mental arithmetic paradigm due to the challenges in extracting functional activity from individuals with long-standing impairments, as the signals from motor imagery are weaker than those from actual execution. Furthermore, mental arithmetic activates the prefrontal cortex, which makes data acquisition more accessible and reliable because the attenuation and artifacts caused by hair are less severe in this region [3,5].

fNIRS is a non-invasive optical technique that measures the concentration change in oxygenated (HbO2) and deoxygenated (HbR) hemoglobin caused by neural activation [6]. Neuroimaging allows us to analyze brain signals caused by functional activity. As a result of this activation, the metabolic requirements of neurons will change, and oxygen demands will be fulfilled by increasing cerebral blood flow. Therefore, fNIRS can calculate the concentration change of these two hemoglobins due to their distinctive absorption spectrum in near-infrared light [7].

Promising results were found in the previous studies of yes/no decoding using fNIRS. Nagels-Coune and colleagues found 75% accuracy in half of the participants with the motor

979-8-3315-2981-9/24/$31.00 ©2024 IEEE

imagery task [4]. Abdalmalak and colleagues achieved %79 sensitivity and 71% specificity and leave-one-out cross-validation was performed to achieve higher accuracy with a combination of 15 features [8]. Sereshkeh and colleagues also used leave-one-out cross-validation to optimize the hyperparameter and improve the accuracy of online classification [9]. Naseer and colleagues conducted a mental arithmetic task and found %74.28 average classification accuracy using linear discriminant analysis (LDA) and %82. 14 average classification accuracy by using a support vector machine (SVM) [5].

In this study, our objective was to develop multiple machine learning models to classify yes and no decisions using fNIRS and to identify which algorithm performs best. We aimed for approximately 80% accuracy in decoding yes/no questions to demonstrate the potential of fNIRS and machine learning methods addressing the primary needs of individuals with motor disabilities and communication challenges.

II. Materials and Methods

A. Signal Acquisition

fNIRS measurements were carried out using the "NIRSport2" fNIRS device (NIRx Medizintechnik GmbH, Berlin, Germany). This device utilizes a high power dual 32 mW light emitting diode (LED) technology with wavelengths of 760 nm and 850 nm as light sources and includes LED detectors. It consists of eight source and eight detector systems with a sampling frequency of 10 Hz [10] and fNIRS signals were acquired via Aurora data acquisition software.

B. Optode Configurations and Placements

Eight sources and seven detectors were placed on the forehead to measure prefrontal brain activation as seen in Fig. 1.

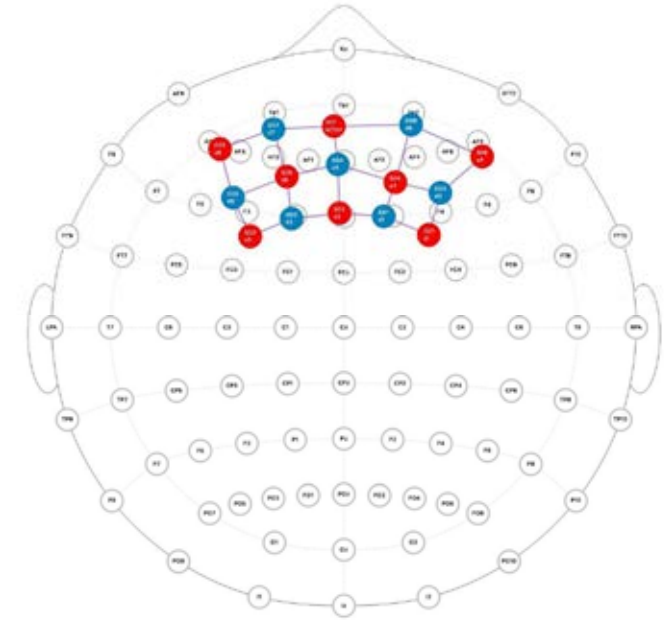


Fig. 1. Optode configuration and placement used in the experiment [11].

C. Subjects

20 healthy participants between the ages of 19 and 25 participated in the experiment with 11 female (mean of age 20.72 ± 1.19) and 9 male (mean of age 21.78 ± 1.99) participants.

Subjects were instructed not to smoke or consume coffee at least 3 hours before the start of the experiment. All participants signed a written state of consent before the experiment. It was emphasized that the researcher should be informed if there is any neurological disorder. The experimental procedure was approved by the TOBB University of Economics and Technology (TOBB ETU) Faculty of Medicine Clinical Research Ethics Committee (2024-08 number of Applications).

D. Experimental Procedure

In this study, we followed a similar experimental paradigm to previous studies, adjusting the response times based on several attempts to determine the participants' optimal response times [5]. Firstly, participants were asked to answer a survey containing 19 simple daily live questions that can be answered with yes or no such as "Are you above 18 years old?" or "Do you have a driver's license?" and random 3 yes and 3 no questions were selected to use in the experiment. The PsychoPy program was used to conduct the experiments [12]. Participants were positioned in a comfortable chair, with their field of view 70 cm away from the computer screen, and they were instructed to sit comfortably without making any movements such as coughing and shaking their heads during this period. During sessions, participants will be asked to perform tasks by the commands given on a computer screen. The experimental paradigm can be seen in Fig. 2. The participant first remained in a resting period for 20 seconds to set up a baseline condition and then saw a question period for 5 seconds. Before the response, the '+' sign was displayed on the screen for 1 second to increase concentration, and after that, the participant answered the questions for 20 seconds each with the given trajectories. The experimental procedure was repeated six times to answer randomly selected three yes and three no questions. The whole experimental procedure was conducted in approximately 4 minutes.



Fig. 2. The experimental paradigm consists of six trials.

When participants answer yes, they will need to perform simple mental arithmetic operations, such as addition, multiplication, or division of up to 4 digits (e.g. 1654-352). If they answer no, they will be expected to comply with the

instructions on the screen stating that they should wait in the resting state.

E. Data Preprocessing.

Signals from data acquisition were preprocessed using the HOMER-3 toolbox for MATLAB [13]. Initially, raw data were transformed into optical density changes (ΔOD). Heart pulses were checked in the power spectrum to assess the quality of the signal. Motion artifact removal was carried out using Wavelet and Spline interpolation approaches and a band-pass filter between 0.01 and 0.15 Hz was applied to filter the psychological noises that affect the signal, such as heart rate (0.6-2 Hz). The filtered ΔOD signals were converted into both oxy-hemoglobin (ΔHbO) and deoxy-hemoglobin (ΔHb) concentration changes using Modified Beer-Lambert Law [14], which was proposed in Formula 1, including the attenuation (G) of the light and the non-linear trajectory of the light in tissue, which is referred to as the differential pathlength factor (DPF) [7,14].

$$\begin{pmatrix}\Delta OD_{\lambda_1}\\ \Delta OD_{\lambda_2}\end{pmatrix} = \begin{bmatrix}\varepsilon_{HbO,\lambda_1}.d.DPF & \varepsilon_{Hb,\lambda_1}.d.DPF\\ \varepsilon_{HbO,\lambda_2}.d.DPF & \varepsilon_{Hb,\lambda_2}.d.DPF\end{bmatrix}\begin{pmatrix}\Delta \mathrm{HbO}\\ \Delta Hb\end{pmatrix} (1)$$

Where $\varepsilon\lambda$ is the molar extinction coefficient of hemoglobin in μM−1 mm−1, d·DPF is the total mean path length of detected photons and λ is the wavelength in nm.

Epochs were extracted from the 2 seconds before the answer period until the end of the period. Baseline correction was performed by subtracting the mean of the two seconds before the start of the period (resting period) from the entire epoch, followed by normalization. Due to noise, some channels were excluded. Additionally, some noisy trials were removed, with a maximum of two trials being excluded per participant, particularly the first trial, where participants had difficulty adapting to the experiment.

F. Classification

Feature extraction was performed using the mean, max, min, and slope values of all 22 channels of the ΔHbO epochs using MATLAB. Therefore achieved 88 features after the extraction. Feature selection was conducted using a linear support vector machine classifier to find the optimal features.

Table I HYPER-PARAMETERS USED IN MACHINE LEARNING

Model	Hyperparameters
Logistic Regression (LR)	penalty: l2 C: 0.0001,0.001, 0.01, 0.1, 1, 10
Nearest Neighbors (NN)	neighbors: 1,3,5…25 weights: uniform, distance
Naive Bayes	Learning rate: 1e-9, 1e-8, 1e-7, 1e-6, 1e-5
Linear SVM	C: 0.0001,0.001, 0.01, 0.1, 1, 10, 100, 1000 weights: None, 'balanced', {0: 1, 1: 10}

Logistic regression (LR), k-nearest neighbors (KNN), naive Bayes, and linear SVM models were analyzed to determine which model achieved the highest accuracy in binary decision, and hyperparameter optimization was carried out using nested cross-validation (CV). In nested CV, the outer loop performs the testing of the hyperparameter-optimized model, and the inner loop uses 10-fold cross-validation to arrange the optimal hyperparameters for the model. In the outer loop, leave-one-subject-out cross-validation was performed, which produces individual-based results. The trials from one subject are used as test data, and the other remaining data are used as train data. The hyperparameters used in machine learning are shown in Table 1.

III. RESULT AND DISCUSSION

The mean classification accuracies from all 20 participants across four machine learning models are summarized in Table 2. We optimized hyperparameters using nested cross-validation and grid search to identify the best settings for optimal accuracies [15].

Among the tested models, Naive Bayes was the least effective, with a mean training accuracy of 74.13%, validation accuracy of 67.42%, and test accuracy of 65.67%. Its AUC of 68% indicates low discriminative power. Although Naive Bayes accurately predicted trials for three participants, it struggled with generalization, achieving 50% or below for seven participants.

KNN demonstrated moderate performance, achieving mean accuracies of 90.01% for training, 75.34% for validation, and 73.67% for testing, along with an AUC of 75%. While it performed well for three participants, it also faced generalization challenges, with three participants scoring 50% or lower, and because of the gap between training and test accuracies.

In the linear SVM model, mean training, validation, and test accuracies were 83.89%, 82.34%, and 81.67%, respectively, indicating good generalization. The model correctly predicted all trials for eight participants, but one participant had an accuracy below 50%, which was also low in other models. The AUC of 91% highlights strong discriminative power.

In logistic regression, mean training accuracy was 80.65%, validation accuracy 83.81%, and test accuracy 82.75%, slightly higher than linear SVM. Participant-based mean accuracies of logistic regression can be seen in Fig. 3. Both models correctly predicted all trials for eight participants, with the same participant scoring below 50%. High variability in hemodynamic responses across subjects may contribute to low accuracies for some individuals. Logistic regression had fewer mismatches and a higher AUC of 94%, indicating stronger discriminative power.

Table II MEAN ACCURACIES ACCORDING TO THE MODELS

	Train Accuracy	Validation Accuracy	Test Accuracy
LR	80.65%	83.81%	82.75%
KNN	90.01%	75.34%	73.67%
Naive Bayes	74.13%	67.42%	65.67%
Linear SVM	83.89%	82.34%	81.67%

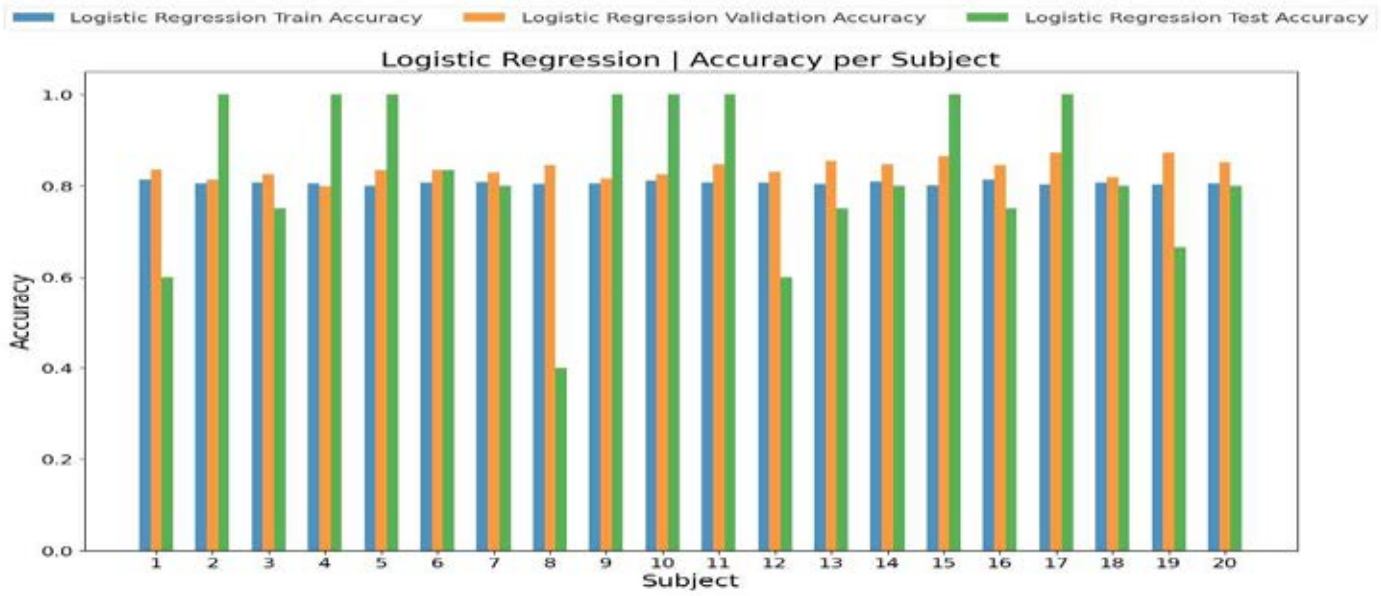


Fig. 3. Mean accuracy of trials across all participants in the logistic regression classifier.

Each machine learning method is tailored for specific applications; for example, logistic regression uses a sigmoid function to predict binary decisions [16], effectively mapping increasing neural signals for "yes" responses. Linear SVM employs support vectors to separate data linearly and has demonstrated high accuracy in previous studies [5]. Both logistic regression and linear SVM work effectively in this context. Conversely, KNN and naive Bayes may not be as suitable for this experiment, as indicated by their lower results.

IV. Conclusion

This study presents a binary decision-making paradigm based on prefrontal cortex activity measured via fNIRS, comparing popular machine learning models to identify the most suitable ones for yes/no decision tasks in fNIRS-based studies. Both logistic regression and linear SVM achieved the desired accuracy, demonstrating that fNIRS has the potential to deliver highly accurate results, particularly due to its advantages like mobility. However, it is important to note that this study was conducted solely on a homogeneous group of 20 healthy participants with similar ages. Although we employed a "leave one subject out" approach to mitigate small sample size issues, future studies could enhance the generalizability of these findings by including larger and more diverse participant groups, including individuals with relevant neurological or motor impairments. Moreover, future work will aim to incorporate more real-world scenarios to further evaluate the practical applications of this approach.

References

[1] Nagels-Coune, L., Benitez-Andonegui, A., Reuter, N., Lührs, M., Goebel, R., De Weerd, P., ... & Sorger, B. (2020). Brain-based binary communication using spatiotemporal features of fNIRS responses. Frontiers in human neuroscience, 14, 113.

[2] Saha, S., Mamun, K. A., Ahmed, K., Mostafa, R., Naik, G. R., Darvishi, S., ... & Baumert, M. (2021). Progress in brain computer interface: Challenges and opportunities. Frontiers in systems neuroscience, 15, 578875.

[3] Abdalmalak, A., Milej, D., Norton, L., Debicki, D. B., Owen, A. M., & Lawrence, K. S. (2021). The potential role of fNIRS in evaluating levels of consciousness. Frontiers in human neuroscience, 15, 703405.

[4] Nagels-Coune, L., Kurban, D., Reuter, N., Benitez, A., Gossé, L., Riecke, L., ... & Sorger, B. (2017). Yes or no?-Binary brain-based Communication utilizing motor imagery and fNIRS. GBCIC, 10, 978-3.

[5] Naseer, N., Hong, M. J., & Hong, K. S. (2014). Online binary decision decoding using functional near-infrared spectroscopy for the development of brain–computer interface. Experimental brain research, 232, 555-564.

[6] Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., & 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.

[7] Saliba, J., Bortfeld, H., Levitin, D. J., & Oghalai, J. S. (2016). Functional near-infrared spectroscopy for neuroimaging in cochlear implant recipients. Hearing research, 338, 64-75.

[8] Abdalmalak, A., Milej, D., Yip, L., Khan, A. R., Diop, M., Owen, A. M., & St Lawrence, K. (2020). Assessing time-resolved fNIRS for brain-computer interface applications of mental communication. Frontiers in Neuroscience, 14, 505426.

[9] Sereshkeh, A. R., Yousefi, R., Wong, A. T., & Chau, T. (2018). Online classification of imagined speech using functional near-infrared spectroscopy signals. Journal of neural engineering, 16(1), 016005.

[10] NIRSport2 | fNIRS Systems | NIRS Devices | NIRx. (n.d.). NIRx Medical Technologies. https://nirx.net/nirsport Accessed March 31 ,2024

[11] NIRSite 2.0 | fNIRS Systems | NIRS Devices | NIRx. (n.d.). NIRx Medical Technologies. https://nirx.net/nirsite

[12] Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. Frontiers in neuroinformatics, 2, 343.

[13] Huppert, T. J., Diamond, S. G., Franceschini, M. A., & Boas, D. A. (2009). HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. Appl Opt, 48(10), D280-298. doi:10.1364/ao.48.00d280

[14] Cope, M., Delpy, D. T., Reynolds, E. O. R., Wray, S., Wyatt, J., & Van der Zee, P. (1988). Methods of quantitating cerebral near infrared spectroscopy data. Oxygen transport to tissue X, 183-189.

[15] Ogunsanya, M., Isichei, J., & Desai, S. (2023). Grid search hyperparameter tuning in additive manufacturing processes. Manufacturing Letters, 35, 1031-1042.

[16] Liu, L. (2018, May). Research on logistic regression algorithm of breast cancer diagnose data by machine learning. In *2018 International Conference on Robots & Intelligent System (ICRIS)* (pp. 157-160). IEEE.