

One Page Summary

The Tufts fNIRS to Mental Workload (Tufts fNIRS2MW) dataset contains brain activity recordings and other data from adult humans performing controlled cognitive workload tasks. We developed this dataset as a step towards a vision for non-intrusive, implicit brain-computer interfaces that can accurately detect the current intensity of a person's mental workload (low or high) and respond by adjusting an interface accordingly.

Contents: For each participant (68 recommended; 87 total), the dataset contains the following records obtained during one 30-60 minute experimental session:

- fNIRS recordings - a multivariate time-series representing brain activity throughout the session, recorded by a sensor probe placed on the forehead and secured via headband
 - All measurements are recorded at a regular sampling rate of 5.2 Hz.
 - At each timestep, we record 8 real-valued measurements, one for each combination of 2 measured concentration changes (oxygenated hemoglobin and deoxygenated hemoglobin), 2 optical data types (intensity and phase), and 2 spatial locations on the forehead. The units of each measurement are (change in) micro-moles of (oxy-/deoxy-)hemoglobin per liter of tissue.
- Activity labels - Annotations of the experimental task activity the subject performed throughout the session, including instruction, rest, and active experiment segments.
 - We label each segment of the active experiment as one of four possible n-back working memory intensity levels (0-back, 1-back, 2-back, or 3-back). Increased intensity levels are intended to induce an increased level of cognitive workload.
- Activity performance - The participant's correctness at the working memory tasks.
- Demographics - The participant's age, gender, race, handedness, and other attributes.

Context: Data was collected from January - July 2021 using a convenience sample of individuals in the Medford, MA area who responded to advertisements and gave informed consent. Data has been fully de-identified and approved for release by the Tufts University IRB.

Recommended Preprocessing: We release files for all 87 total participants, but we recommend excluding the fNIRS data from 19 participants for data quality reasons (see our paper). Our released files indicate the 68 eligible participants.

Recommended Machine Learning Task: sliding-window time-series classification

- *In:* fixed length (e.g. 30 seconds) multivariate time series of fNIRS brain signals
- *Out:* predicted workload intensity level, with up to 4 possibilities (0-, 1-, 2-, or 3-back)
- *Possible Protocols to compare generalization potential fairly:*
 - subject-specific : Train and test data (no overlap) come from one target subject
 - generic : Train on data from S other subjects, evaluate on a target subject
 - generic + fine-tuning : Fine-tune generic model on data from a target subject

Datasheet Contents

- [Motivation](#)
- [Composition](#)
- [Collection Process](#)
- [Preprocessing](#)
- [Uses](#)
- [Distribution](#)
- [Maintenance](#)

Motivation

For what purpose was the dataset created?

Functional near-infrared spectroscopy (fNIRS) promises a non-intrusive way to measure real-time brain activity and build responsive brain-computer interfaces (BCIs). To make such interfaces possible, we imagine a classifier that can consume recent brain activity signals and producing a prediction of the user's current mental workload (e.g. low or high). Using knowledge of the user's current state, the interface could respond accordingly (e.g. reduce clutter to allow focus).

However, building such a classifier is difficult due to several key barriers:

- *the lack of available datasets* : Only a few open datasets exist for fNIRS data paired with mental workload labels; none of these contains more than 30 subjects.
- *the high variation in fNIRS data* : This makes generalizing to a new subject or session difficult, especially if the available data for training a model is limited.

We thus created this dataset to provide the largest available open-access dataset for training and evaluating classifiers of mental workload. We hope it helps benchmark progress toward the goal of real-time responsive BCI for everyday use cases.

Who created the dataset and on behalf of which entity?

The data was designed and collected by two groups of collaborating researchers from the Tufts University Department of Computer Science, with fNIRS sensor design and implementation help from experts in the Tufts Dept. of Biomedical Engineering. Data collection was led by researchers in the Tufts Human Computer Interaction laboratory. Efforts to build and evaluate classifiers using this data were led by researchers in the Machine Learning Research Group.

Who funded the creation of the dataset?

The dataset collection was funded by grants from the U.S. National Science Foundation and Google. This included the purchase and setup of the fNIRS sensor device and \$20 US compensation for each participant.

Composition

What do the instances that comprise the dataset represent?

The dataset contains records from many adult human participants. Each participant completed a 30-60 minute experimental session, in which they received brief instruction, then performed a series of experimental tasks designed to induce varying intensity levels of working memory.

How many instances are there in total (of each type, if appropriate)?

The entire dataset contains all 87 participants.

Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set?

All 87 participants who completed our study protocol are represented.

What data does each instance consist of?

For each participant, we have recorded and made available:

- fNIRS recordings representing metabolic brain activity throughout the session
- activity labels indicating which time intervals correspond to instruction, rest, and experimental tasks (especially labeling the 0-back, 1-back, 2-back, and 3-back segments)
- Correctness labels for each response to stimuli in the experimental n-back tasks
- Pre-survey results indicating demographics and other contextual information (sleep, history of brain injury, etc.)
- Post-experiment survey

Is there a label or target associated with each instance?

Yes, the labels of experimental activities performed in each time segment of the session can be used as the target outcome of a prediction task.

- Is any information missing?

Across the entire set of records, only one entry is missing (for one subject, we do not know handedness). Otherwise, all records are present.

Are relationships between individual instances made explicit?

Each record represents a separate subject. A subject id number is consistently associated with each record throughout.

Are there recommended data splits into train/validation/test?

Yes, for each supported evaluation paradigm (subject-specific, generic, and generic + fine-tuning), we provided recommended train/validation/test splits that allow each subject's data a turn as the out-of-sample test set.

Are there sources of error / noise / redundancy?

fNIRS sensor recordings are known to contain noise and artifacts. We provide a preprocessed version of all recordings that performed bandpass filtering from 0.001 - 0.2 Hz to remove some drift, respiration, and heartbeat issues. We also recommend excluding some participants' data entirely due to quality issues. Users are still cautioned that the provided data may benefit from further processing to address remaining noise issues.

Is the dataset self-contained?

Yes, all files are provided from the same stable source (our website).

Does the dataset contain data that might be considered confidential?

All data is de-identified.

Does the dataset relate to people?

Yes, this data was collected from adult humans.

Does the dataset identify subpopulations?

Yes, the provided demographics data allows identifying each participant's age, gender, race, handedness, and other attributes. These attributes can be used to form subpopulations.

The following table summarizes the composition of our recommended eligible cohort of 68 individuals:

	Asian	White	Latin	Black	Pac. Isl.	other
race	32	27	3	2	1	3
	M	F	other	decline		
gender	26	39	2	1		
	right	left	unreported			
handedness	64	3	1			
	min.	max.	mean	std		
age	18.0	44.0	21.71	4.01		
sleep last night (hr.)	3.5	10.0	7.29	1.21		

Is it possible to identify individuals?

No personally identifiable information has been released.

Does the dataset contain data that might be sensitive in any way?

The released data does identify self-reported racial and ethnic origins, as well as gender.

Otherwise, the released data does not indicate other sensitive information that we can foresee could be used to enable discrimination or mistreatment of any individual in our study in the unlikely event of a bad actor who is able to perform re-identification.

Collection Process

How was the data associated with each instance acquired?

Each voluntary participant was scheduled for an individual experimental session in our laboratory at Tufts University, where they sat in a standard chair in front of a desktop computer with screen, keyboard and mouse for 30-60 minutes. During this session, we collected a pre-experiment survey for demographics and information, provided brief instruction, obtained informed consent, placed the fNIRS sensor carefully on the subject's forehead, performed a series of experimental tasks, and then collected post-experiment survey data to verify quality.

What mechanisms or procedures were used to collect the data?

The experimental protocol followed a standard procedure for n-back working memory tasks: subjects performed 16 different blocks of n-back trials, using a counter-balanced Latin square design. Each block consisted of 80 seconds of experimental activity followed by a brief rest (10-30 sec.). See our paper for further details, where all protocols are fully documented.

For fNIRS sensing, we used Imagent frequency-domain (FD) NIRS instrument manufactured by ISS (Champaign, IL, USA). Two custom probes (right and left) were secured to the subject's forehead via a removable headband. Raw FD-NIRS measurements contain temporal traces of alternating-current intensity and changes in phase at two wavelengths (690 and 830 nm), using a 110 MHz modulation frequency. From these measurements, we recover time traces of dynamic changes in oxyhemoglobin (ΔHbO ; units: $\mu\text{mol/L}$) and deoxyhemoglobin (ΔHbR ; units: $\mu\text{mol/L}$) from either intensity or phase. See our paper for further details, where all fNIRS sensing and processing details are fully documented.

How were these mechanisms or procedures validated?

We verified fNIRS data quality by ensuring consistent sensor settings (esp DC intensity) and examining each participant's fNIRS time series for evidence of severe data quality issues.

We verified that participants performed the desired working memory tasks at appropriate intensity levels by recording a binary response to every n-back stimulus (either "left arrow" for the stimulus matches the target or "right arrow" for no match). For each participant, we checked the accuracy of their responses compared to the correct answers.

If the dataset is a sample from a larger set, what was the sampling strategy (e.g., deterministic, probabilistic with specific sampling probabilities)?

Our data was a convenience sample of people who responded to email and posted flyer advertisements. We intended to reach individuals across Tufts University, and our resulting cohort roughly matches the university community's overall demographics in terms of age, gender, and race.

Who was involved in the data collection process and how were they compensated?

Each participant was compensated \$20 US.

Two operators in the experiment were a paid student researcher and a volunteer student research assistant.

Over what timeframe was the data collected? Does this timeframe match the creation

timeframe of the data associated with the instances?

The data was collected from January to July, 2021. All data was created during experimental sessions that took place during this time window.

Were any ethical review processes conducted?

Yes, our study design (**STUDY00000959**) was approved by the [Tufts University Institutional Review Board](#).

Because collection occurred during the COVID-19 pandemic in early 2021, we also got approval from the Tufts Integrative Safety Committee (ISC). We followed required sanitation and social distancing practices: experimenters used personal protective equipment and disinfected the fNIRS probe for each subject. From Sep 2020 to May 2021, we were approved for one subject to perform the experiment every 72 hours, to ensure distancing and prevent disease spread. From May 2021 to Aug 2021, we were approved for a shorter time gap (30 minutes between subjects). Actual data collection occurred from January - July 2021.

The IRB approval certificate is shown below:

Approved

Entered IRB: 8/14/2020 3:53 PM
Initial approval: 9/11/2020
Initial effective: 9/11/2020
Effective: 12/16/2020
Approval end: 9/10/2021
Last updated: 6/17/2021 12:24 PM

Next Steps

View Study

Print Version

STUDY00000959: Automated Multimodal real-time BCI Framework

Principal Investigator: Liang Wang	IRB office: SBER IRB
Submission type: Initial Study	IRB coordinator: Christine Pelletier
Primary contact: Liang Wang	Letter: Expedited Approval STUDY00000959.pdf(0.01)
PI proxies: Robert Jacob	Regulatory authority: 2018 Requirements

```

graph LR
    A[Pre-Submission] --> B[Pre-Review]
    A --> C[IRB Review]
    B --> D[Clarification Requested]
    D --> B
    C --> E[Clarification Requested]
    E --> C
    C --> F[Post-Review]
    F --> G[Modifications Required]
    G --> F
    F --> H[Review Complete]
  
```

Does the dataset relate to people?

Yes this was collected from volunteer adult human participants.

Did you collect the data from the individuals in question directly, or obtain it via third parties?

Yes we collected data directly from the participants.

Were the individuals in question notified about the data collection?

Yes, all individuals were recruited volunteers and compensated for their time. Each individual was fully informed about all data collected (pre survey, fNIRS recordings, and post survey).

We recruited participants from the Tufts University and broader Boston community through fliers posted on campus. The fliers contained the following text:

*Human-Computer Interaction Lab Looking for Participants for a Research Study !
\$20 for one hour*

*Do basic tasks on the computer while wearing biological sensors to help advance the study of
Brain-Computer Interfaces and Human-Computer Interaction!*

*You must be at least 18 years old and an English speaker to participate. You should not have a
history of seizures/epilepsy. For more information, please contact leonwang@cs.tufts.edu*

While we were approved for recruitment via emails using the same text above to official listservs of undergraduate and graduate students in the Computer Science department at Tufts, we did not end up using this recruitment tool. We found sufficient participation from flyer advertisements.

Did the individuals in question consent to the collection and use of their data?

Yes, all participants consent to the collection and use of their data.

Consent form for collection of initial data:

<https://drive.google.com/file/d/1p82pY6SmVs9VW0Z0a4HhxBnt4gZOfHvd/view?usp=sharing>

Re-consent form for approval for public release:

<https://drive.google.com/file/d/1EK-nVSbp3CmNVBqOLW3sDnuUmBKaBYQy/view?usp=sharing>

If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses?

Yes, it is clearly described in the consent form (Section "Is my participation voluntary, and can I withdraw") and the re-consent form (Section "How can I be removed").

Has an analysis of the potential impact of the dataset and its use on data subjects (e.g., a data protection impact analysis) been conducted?

The IRB approval process described above required us to describe the potential positive impact as well as risks to subjects. Because the risk of harms was judged to be minimal, our study was approved for release.

Preprocessing

Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucketing, removal of instances, processing of missing values)?

Yes, we applied several preprocessing steps:

- Selecting a recommended cohort of 68 individuals
 - Out of all 87 participants, we excluded 19 for a variety of data quality reasons, as described in our paper.
- Bandpass filtering to remove some heartbeat and respiration artifacts
- Extracting sliding windows from the active experimental segments of the raw full-session recording

Was the “raw” data saved in addition to the preprocessed data (e.g., to support unanticipated future uses)?

Yes we make all data collected from all 87 participants available, so other inclusion criteria could be tried in the future.

Yes raw full-session fNIRS recordings (without the bandpass filter) are provided.

Is the software used to preprocess/clean/label the instances available?

Yes we make available the software to do the full experimental data collection, including the active n-back experimental tasks (how participants are shown visual stimuli on screen, collecting keyboard input as response, etc.).

We also make available the software to perform the dual-slope preprocessing, the bandpass filtering and the sliding window extraction. Some pieces of this software are written in Matlab, and other parts are in Python.

Uses

Has the dataset been used for any tasks already?

Yes, we have used the data to train and evaluate mental workload classifiers under subject-specific, generic, and fine-tuning paradigms. See our paper for details.

Is there a repository that links to any or all papers or systems that use the dataset?

Our dataset website contains a link to our current publication, describing the dataset and our analysis of the results. We plan to document future relevant papers that also use our dataset on the same website, to provide a comprehensive resource for future users.

What (other) tasks could the dataset be used for?

The data could perhaps also be used to predict if a user is correctly following experimental instructions or not (using the recorded responses to stimuli).

Is there anything about the composition of the dataset or the way it was collected and preprocessed that might impact future uses?

The data was collected with a custom-designed sensor probe. Further researchers would need to create a similar probe; the specifications and dimensions are provided in the paper.

Are there tasks for which the dataset should not be used?

Future users should not try to use brain recordings to predict any sensitive attributes (e.g. race) available in the demographics data.

Future users should never try to reidentify the subjects included in our dataset.

Distribution

How will the dataset be distributed?

The dataset will be released via a website controlled by the study creators.

All data will be released in standard plain-text formats (comma-separated value files), with data dictionaries and full documentation available on the website.

Will the dataset be distributed to third parties outside of the entity on behalf of which the dataset was created?

The dataset will be released via a website controlled by the study creators. The data files themselves will be hosted on Box.com, or in the future on a similar vendor approved by Tufts Technology Services.

Does the dataset have a digital object identifier (DOI)?

The dataset is available at a stable URL:

https://tufts-hci-lab.github.io/code_and_datasets/fNIRS2MW.html

We are exploring how to best obtain a DOI.

When will the dataset be distributed?

The data files will be available from early September 2021 onward.

Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)?

The data and documentation is released under a CC-BY 4.0 license.

Have any third parties imposed IP-based or other restrictions on the data associated with the instances?

No.

Do any export controls or other regulatory restrictions apply to the dataset or to individual instances?

No.

Maintenance

Who will be supporting/hosting/maintaining the dataset?

Members of the Tufts HCI laboratory and the Tufts Machine Learning research group.

How can the owner/curator/manager of the dataset be contacted?

Please email the following individuals:

- Liang Wang : leonwang(AT)cs.tufts.edu
- Robert Jacob : jacob(AT)cs.tufts.edu

Updated contact information is also maintained on our website.

Is there an erratum? Will the dataset be updated?

The dataset website will be updated to report any changes or errors in the dataset. Users can check the website for announcements of any changes.

If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances?

We have consent and approval for making these files available for at least 10 years.

Will older versions of the dataset continue to be supported/hosted/maintained?

Yes. We plan to maintain this version of the dataset continually. If updates occur or even more data is collected, we will release a new version (with appropriate major/minor version number). Older versions will always be available, and a clear release log will document any changes between versions.

If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so?

Our data is released under a CC BY 4.0 license, which allows others to reuse and remix the dataset as long as original credit is attributed to us (please cite our publication).

Any reuse or remixing is the responsibility of the downstream creators.