Predicting Discrete Levels of Cognitive and Perceptual Load Using Functional Near Infrared Spectroscopy Data

Kaunil Dhruv
University of Colorado, Boulder
Institute of Cognitive Science
Boulder, CO
kaunil.dhruv@colorado.edu

Trevor Grant
University of Colorado, Boulder
Institute of Cognitive Science
Boulder, CO
trevor.grant@colorado.edu

Leanne Hirshfield
University of Colorado, Boulder
Institute of Cognitive Science
Boulder, CO
leanne.hirshfield@colorado.edu

ABSTRACT

Predicting cognitive workload using physiological sensors has taken on a diffuse set of methods in recent years. Many of these methods, however, train models on small datasets with hand selected features, limiting a model's ability to transfer across participants, tasks, or experimental sessions. Here, we explore new potential methods of integrating data from and modeling on a large, cross-, participant, task, and session, set of high density functional near infrared spectroscopy (fNIRS) data by using an approach grounded in cognitive load theory and data warehousing techniques in combination with Long Short Term Memory Networks.

KEYWORDS

datasets, neural networks, cognitive load, perceptual load, adaptive systems

ACM Reference Format:
Kaunil Dhruv, Trevor Grant, and Leanne Hirshfield. 2019. Predicting Discrete Levels of Cognitive and Perceptual Load Using Functional Near Infrared Spectroscopy Data. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

If a system seeks optimal performance between a human agent and a computerized system then the amount of cognitive workload (CWL) on the part of the user must be reduced by the greatest degree possible. As computerized systems become more omnipresent in everyday life, and reliance on them increases, the impetus to achieve optimal performance between a system and a user becomes increasingly important. Past research indicates that increased CWL on the part of the user has had deleterious effects on both performance [41] as well as reaction times [22] when working in a simulated real world task environment. Other evidence indicates that these same demands elicit similar effects when interacting with computerized communication systems [33], as well as when performing basic cognitive tasks within a laboratory setting [2, 40]. If further increases in performance are to be had from these types of systems then accurate and effective measurement and prediction of CWL is crucial to driving these performance gains.

In recent years the use of physiological sensors has been gaining in popularity as a way to measure an individual's CWL during real work task performance. Previously, much work in the CWL domain was reliant either on behavioral scores, such as task performance and reaction time data, or subjective survey measures administered after the completion of the task. Though in some cases these measures may suffice, if the goal is to make adaptive systems more accurate in their predictions as to the user's internal state, then the use of real time physiological data to make these classifications may be a more robust measure by which to accomplish this. As the community has turned towards using physiological measures a device, functional near infrared spectroscopy (fNIRS), began to serve key purpose in the CWL problem space as a tool that allowed researchers the ability to record information about oxygenated and deoxygenated hemoglobin levels in the brain portably and non-invasively, which has allowed researchers access to real time information about individual's brains while they are subjected to ecologically valid (i.e. "real world") task environments.

Though the use of fNIRS has continued to gain traction in the research space through the years there still remain overarching concerns and shortcomings within the workload classification domain about fNIRS' efficacy at classifying cognitive states. Accuracy measures for predicting CWL are

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/1122445.1122456

often times reported to be fairly substantial [48], but these high accuracy measures come on the back of the classification models being trained on relatively small subject pools, with hand selected features, and fail to maintain these accuracy measures when tested on data taken from other sames of individuals, making it likely that the models themselves could have overfit to the either the individual or the entire subject pool. If further progress is to be made in this field then models ought to be trained on larger, and more varied data sets, and ground truth values that are more reflective of the underlying theories in CWL ought to be used. This can be accomplished by combining insights from lower level behavioral and self report measures used in previous literature to gain a better understanding of what types of load are known to elicit distinct responses in the human brain. The use of domain knowledge within the CWL as well as fNIRS fields should allow not only more accurate and transferable models, but also models that are able to make these predictions with a smaller set of data.

It is for this reason that in this submission we propose a data warehousing technique by which researchers can query over a set of data, taken from multiple participants in multiple experiments performing multiple tasks, and use a system developed by domain knowledge to label these sections of brain data based on the level of cognitive load undergone by the user during that task. By using this storage and analysis technique for labeling our ground truth measures for classification, we allow a more iterative approach to modeling and classification techniques, which may lead to broader insights about large data pools. It should also allow researchers to better leverage these insights and to allow for them to work in tandem with their modeling hypotheses. This application of labeling and storing fNIRS data in a researcher queryable format, known as PyoNIRS, should allow researchers working within the CWL and fNIRS research areas to reduce the problem space within datasets to a more manageable set of models, models which will be tied to a clearer set of hypotheses. Further, these insights from preliminary analysis should also allow users to down sample the size of the data being fed into the model by relying on what features lower level analysis deem most important for making predictions about individual cognitive load states. This, in turn, should allow for faster iterations in modeling procedures, resulting in more accurate models that require less input in order for their predictions to be made.

2 RELATED WORK

The Brain and Brain Measurement

The human brain is a complex structure, comprised of, on average, 86 billion neuronal cells [30]. Between these cells exist hundreds of trillions of synapses, or areas between cells where information is transferred from one neuron to another in the form of a chemical signal [58]. The work of multiple disciplines, ranging from cognitive neuroscience to computational biology, has yielded great advancements in our ability understand the human brain [26]. Though past research in the neuroscientific community has focused on differentiating these discrete cortical (outer surface) regions based on distinct functional specialization, further evidence is beginning to suggest that the interactions across these functionally specialized regions plays a role in higher level cognitive processing [62]. These functional systems, or networks, involve areas of cortex that are anatomically distant from one another in the brain, but whose patterns of connectivity are temporally correlated [17]. The neuroscience literature has linked many cognitive processes to specific brain areas, and these links are termed neural correlates [14, 18, 20]. While it is often assumed that there is a simple one-to-one mapping between processes and brain areas, in reality it is more complex with a many-to-many mapping between activations in certain regions and human processes [47]. These findings suggest that fNIRS may have a further roll to play in neruoscientific research as the strengths of fNIRS allow it to measure distal sections of brain activation with a greater temporal resolution than that of other brain measurement modalities, such as functional magnetic resonance imaging (fMRI). fNIRS systems work by the use of near-infrared light, which can penetrate through scalp and skull to reach the brain cortex. The optical fibers are placed on the surface of the head for illumination while detection fibers measure light which reflects back. Particularly, concentration changes in oxygenated and deoxygenated hemoglobin can be measured by measuring the amount of light that is reflected back from the brain into the light detector [15]. A review of the history of fNIRS, can be viewed in the work of Ferrari et al. [23] as well as Boas et al. [10]. fNIRS the beefit of being a go between in terms of spatial as well as temporal resolution when compared to other popular brain measurement equipment. fNIRS has a higher spatial resolution than EEG, making it possible to localize specific functional brain regions of activation, as could be done with the constrictive fMRI device [45]. The temporal resolution of fNIRS is better than that of fMRI, but is significantly less than that of EEG. The ability to spatially locate specific functional brain regions of interest enables high-density fNIRS sensors to identify specific neural correlates of CWL and other mental states of interest. The nature of brain correlations with certain cogntive tasks involving both a spatial as well a temporal relationship also makes brain data well suited for classification using deep neural network tecniques such as LSTMs and Convolutions Neural Nets [52].

Cognitive Workload Modeling

Recently, increasingly effective strides have been made in being able to predict and modulate the CWL level in tasks by using both behavioral [11, 29, 34] and, more recently, physiological and psychophysiological [2, 19, 40], measures to classify different levels of CWL. Though these steps have elucidated some of the underlying problems with classifying and predicting CWL we still lack a clear picture of what an ideal approach to creating more accurate and grounded predictions might be [16]. Differences in theoretical grounding of CWL lead to differences not only in CWL manipulations in experimental paradigms, but also competing evidence as to which measures (behavioral and physiological) are most effective at measuring CWL [16]. A subset of the field that shows increasing promise in terms of CWL level prediction is the use of non-invasive brain measurement modalities in order to predict a user's internal state [16]. Among these non-invasive measurements functional near infrared spectroscopy (fNIRS) has been increasingly adopted and utilized for it's, robustness to noise and ease of use [57].

Further complicating matters is that as more advanced methods of interaction between computerized systems and their users are developed, issues arise not only from the implementation considerations that are pertinent to the system [66], but also as a result of the lack of epistemic gains from the technologies used to achieve the implementation. This *information bottleneck* which is currently inherent the use of deep learning algorithms on data [60], is especially troubling when trying to optimize systems that are able to predict different types and different amounts of cognitive workload in the user. Even if accurate classification of CWL load level is accomplished it is often unclear as to how this accuracy was achieved. If researchers and developers of autonomous systems ever hope to achieve reliable results then new methods must be developed that mitigate the downsides of "black box" modeling while leveraging the components of the technology that make it so useful at generating accurate predictions.

Reasons why there may so many problems within this space is that CWL has always been viewed as a single construct, or CWL may be a component of an individual's state rather than a component of the task that the individual is undergoing. Recently, some have proposed the used of multiclass, multi-label classification as a way to break down the issue of cognitive load into more discreete pieces of load [48]. This approach, however, is not immune to it's own set of issues. One current problem within the multilabel classification domain that is exacerbated by the problem space of classifying mental states is that of label selection. Not only is the selection of a label schema that is representative of the underlying physiological data important, but one must also consider *how many* discrete labels should be used to accurately predict a given cognitive state. Though there has been success in using neuroimaging modalties coupled with machine learning techniques in the past [56], these successes are typically working within classifying discrete, tightly controlled, stimuli adjustments [63]. Past work also shown that a cognitive resources' (working memory) load level is able to be predicted using neurophysiological data [25, 27]. These results are promising, but the labeling schema in these studies may be too coarse-grained so as to miss out on crucial distinctions in the sub-components of the working memory system that cognitive models indicate can be distinguished from one another [4]. Though higher numbers of labels may be needed in order to properly carve a cognitive process at it's natural or theoretical joints, models trained on larger labeling schema could also become overly atomistic and therefore limit themselves in terms of their actual explanatory power. Others contend that theories may need to be reworked in order to accommodate new experimental as well as theoretical considerations [54]. A higher number of labels also carries of excess baggage of becoming more computationally expensive, especially in the case of binary relevance classification. Though, this concern may be abated by evidence which indicates that dimensionality reduction with large labeling schema is possible [8]. Proposed solutions to these issues are in no way exhaustive, and much work should continue be done within each of the mentioned problem points in order to work out potential solutions that may prove fruitful in this problem space.

3 METHODS

Current Dataset

The PyoNIRS system was developed from and for the purpose of mining knowledge from a large, cross experiment, cross participant dataset collected using a Hitachi ETG-4000 fNIRS device. The data set contains data from 11 different experiments with an average n of 21 data files per experiment (min = 10, max = 63). The total count of fNIRS data files within the current data repository is 245. To begin, only the two most recent data sets within the total dataset were selected for use within the PyoNIRS system. The reason for this approach is the most recent experiments have the vast majority of the data within the data set and are also the most complicated experiments run within the entire dataset, the rationale being that if the PyoNIRS system were able to accommodate the most complicated data files within the dataset then the ability to import data files from other experiments into the PyoNIRS systems should be relatively straightforward in comparison. After selecting for the two most recent experiments, the total number of data files involved in both the data cleaning pipeline as well as data analysis and modeling

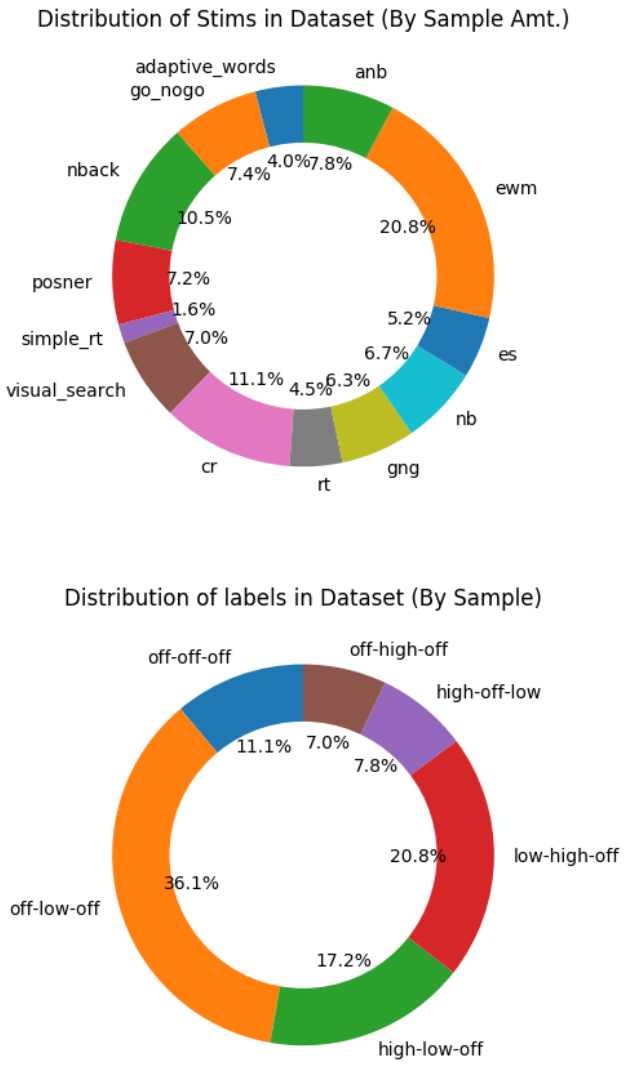

Figure 1: Above: The distribution of task (by number of samples [rows]) corresponding to each task). Below: The distribution of labels (by number of samples) corresponding to each label assigned to the tasks listed above.

portions of the project was reduced to 80 valid fNIRS data files. The total participant count within the selected dataset was therefore 61 (as one experiment had the participants use multiple sessions).

The ETG-4000 device uses near infrared light to measure levels of oxygenated and deoxygenated hemoglobin across the cerebral cortex of human subjects by creating multiple channels of light source and light detector pairs. The device samples at a rate of ten samples per second, and each experiment had an average data collection time of 36.4 minutes. Most common within the dataset is a probe configuration which uses a 3 X 11 array of optodes which were placed over participant's forehead area, measuring the prefrontal cortex of the brain. These datasets contain 52 separate points of measure on the prefrontal cortex. Within our dataset, however, there are also experiments which used a different array of optodes and therefore have 40 channels of data. Of these channels, 20 where placed on the prefrontal cortex and 10 were placed on both sides of the head over an area of the brain known and temporal-parietal junction (TPJ), or the area in which the brain's temporal lobe and its parietal lobe meet. The data set we currently have access to has human participants performing various tasks, from responding to tightly controlled stimuli adjustments, to performing more ecologically valid tasks. Though these data files with the 40 channel probe configuration exist within the total dataset, they were not present in the down selected sample of data we used for development and analysis. In order to accommodate those other files a separate system would need to be developed such that the channels were mapped to either the 10-20 EEG system for head space localization [31], or to use a more complex tool such as NIRS-SPM [68] to map the channels within the dataset to their underlying brain regions so that like channels across different probe configurations could be easily compared. This "channel mapping" could also be accomplished correlating the channels in different headsets to Broadmann areas [3] on the brain using techniques used by past researchers [42, 50]. Though this has been useful for combining data files from different experiments there are perhaps still better ways in which the data can be merged. The first of which being that rather than using Broadmann areas, which divide the brain's cerebral cortex based on different cell structures, more recent anatomical terminology and functionally differentiated regions of the brain could be use with which to map non-identical channels in different probe configurations. More work should be done to investigate whether or not this technique would bear any fruit and allow researchers to hold on to, or at least leverage more of the data sets that they have in their possession. Another option, which may be beyond the scope of this project but which is worth mentioning is that areas of the brain, though physically distal from one another may all be part of a functional network of brain regions. We were to combine channels of fNIRS data based instead on the functional connectivity we may be able to leverage a greater amount of the current dataset, and therefore derive more interesting patterns in the temporal nature of these networked areas [7, 32, 49]. As we planned on using LSTM modeling for the next step in the procedure, the use of probe configurations that we were confident represented the same spatial mappings to the human brain was important [38], and we were not satisfied by our efforts to algorithmic convert different probe mappings to a universal space (the sheer amount of

data within the dataset meant that doing so by hand would have been untenable given the time constraints). they we The dataset we currently have access to have not changed since the inception of this project, though the way in which we interact with the dataset has. Developing and using PyoNIRS has enabled the The bulk of our work has focused on creating a streamlined way to warehouse to current data into a more manageable and easy to access format.

As seen in Figure 1, there was a significant amount of difference between the distribution of end labels assigned within the dataset we were working on. This led to some difficulties in the modeling process that were attempted to be overcome first by down sampling the majority class, and then by novel modeling techniques, but even with these efforts the skew within the task distribution remained a problem point throughout the work of this project. The reason for this is that as we were trying to train many models on the different label types there were certain conditions for which we had very little accuracy for. Future work within the field should attempt to find ways to handle these sorts of data imbalances that may often come up as a result of experimental designs that were not developed with the idea of advanced statistical modeling in mind.

PyoNIRS Overview

This pipeline has been the bulk of our focus throughout this first period of development of the project as it has involved tying together multiple data streams and leaving open the possibility of allowing for further data entries in the future. A architecture diagram of the PyoNIRS system can be seen in Figure 2. The figure shows a birds eye view of the bulk of the systems within the PyoNIRS program that are used for cleaning and organizing the current dataset into a new format that is more conducive to the types of analysis that can be then carried out in the modeling steps. First, the data, and a special pointer file about the conditions a participant underwent during the data session, called a 'conditions file', as well as metadata information about the participant are all fed into the PyoNIRS application. Once all this information is present within PyoNIRS the application goes through the preprocessing procedures such as noise removal. This noise removal is achieved first by using a bandpass filter [21] to filter out unwanted signals within the overall optical signal recorded by each fNIRS device. As the signal is measuring light intensity, and this light intensity is used to measure Oxy- and DeOxy hemoglobin levels in particular areas of the cortical surface underlying physiological noise such as meyer waves [55] and heart rate [46] must be removed from the data to ensure that was is being measured is only the the signal of interest with respect to levels of Oxy and DeOxy hemogolin that are part of the BOLD (Blood Oxygen Level Dependent) signal [39]. After the data is filtered, noise removal techniques are used to deal with potential motion artefact within the data. Currently, our noise removal step for motion artefacts is wanting in that we discard the channel in which the artefact was found to be present (an event in which a light intensity jumps 8 standard deviations above the mean signal is determined to be a motion artefact). We plan to use better remove and reduction features in the future before the final modeling process begins. After this step the raw light intensity is converted into Oxy and DeOxy- hemoglobin levels using the modified beer-lambert law [5, 37]. As a pre-processing final step, all of the data is normalized by data channel using z-score normalization which allows not only the channels to be consistent with one another, but also allows for each data file within the dataset to be compared in even statistical space [55]. After the pre-processing step, PyoNIRS then, using the 'conditions' file, which gives a string label to each task the participant completed as well as the indexes for the beginning and end points of that task condition in the original data file, the data is cut into task chunks and stored in the database with both the task label, and metadata about the participant stored as a label for that particular data chunk. The data can then be indexed and queried by a user based on both task condition as well as by meta data parameters.

We will then leverage the information obtained from the previous step to begin to modeling process. We will to compare multiple models using multiple labels to determine which modeling yields the best accuracy for this type of data. We will be guided by, but not limited by, the previous work done in the domain, eluded to in section 2. For more information about proposed model inputs, training types and model architechtures, refer to fig. 4.

Data Pre-processing

Hemoglobin Conversion, Filtering, and Normalization. A particular challenge when working with fNIRS data as opposed to data obtained using other neuroimaging modalities is that the optical data captured by fNIRS systems is much more sensitive to forms of physiological noise that magnetic (fMRI) and electrical (EEG) signals are not sensitive to [35]. For this reason the noise removal steps currently implemented within the data pre-processing pipeline operate on the raw optical signal files obtained from the fNIRS device, and do not rely on the software within differing fNIRS systems to handle to pre-processing as the manufacturer's algorithms and parameters of these pre-processing steps may differ from one another and would not allow the comparison between fNIRS files obtained on different systems. For each file within the dataset the raw optical signals first converted into rate of change of oxy δHbO and de-oxy hemoglobin δHbR values using modified Beer-Lambert Law [5]. After the HbO and

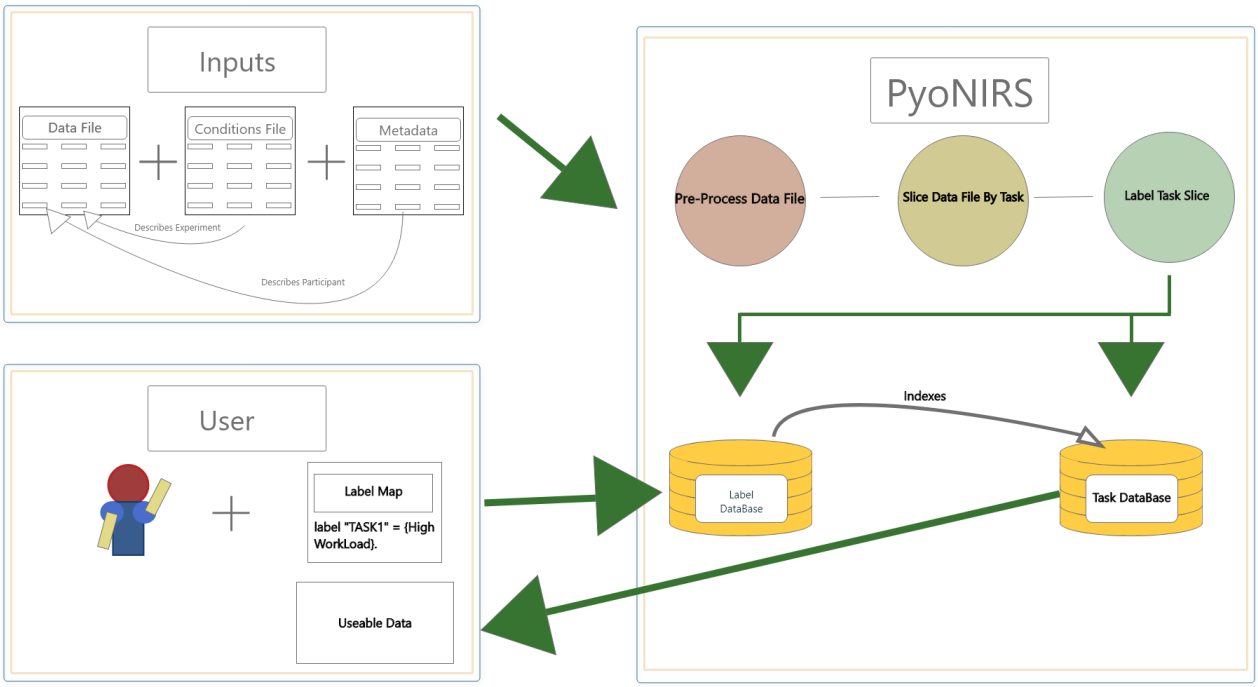

Figure 2: The architechture diagram of the PyoNIRS preprocessing and data warehousing application which will be use as a springboard for model development and further data mining techniques.

HbR values are obtained these values are then band-pass filtered using a 6^{th} order Butterworth filter with low and high frequency of $0.01Hz$ and $0.5Hz$. These filters serve the purpose of filtering out physiological noise from low frequency drifts in the optical signal as well as high frequency cardiac noise, respectively [1]. However, these efforts do not filter out the remaining physiological noise such as Meyer Waves (from respiration rate) and blood pressure artefacts as these signals occur at similar frequencies to that of the brain activation signal that is attempting to be measured using fNIRS [9]. To account for these other forms of physiological noise the general linear model (GLM) was used as an offline adaptive filter over the data [24]. Had the dataset not included pure fNIRS data and also included the physiology, there are other substantive measures that could have been used in stead of the GLM. Another method of this subtractive method for physiology removal involves the use of "short channel" fNIRS detectors, that would detects the physiological signal at the scalp level, and uses this physiological information as a method by which the remove the physiological noise [1]. As a result of the the measure obtained after these conversion and filtration steps being a delta all of the values within the data files, and does not reflect a total level of HbO and HbR, the resulting values are then z-score normalized so that consistent analysis can occur across participant files.

Noise Removal and Motion Artefacts. Working with a dataset this size also brings other challenges, especially in fNIRS data where motion artifacts and other types of experimental noise can reduce signal quality [12]. The data must first be passed through in its entirety to ensure that proper noise removal techniques and motion artifact algorithms have been run consistently across the entire dataset, so that the samples fed into the models are consistent. Many different methods to help combat this noise within the data from simple methods such as discarding or smoothing out of the signal, [12], to more complex methods such a wavelet-transform [43], to pre-whitening [6] and channel weighting [59] techniques such as those used in the NIRS-toolbox [51] application. Looking through the motion artefact removal algorithms in many of the modern fNIRS analysis packages show diverse methods for the removal of these noise artefacts. The simplest, a smoothing, algorithm was chosen for the current purposes of PyoNIRS as it did not disrupt the temporal nature of the data which would be use in the modeling process. Fortunately, many of these algorithms have been open sourced and can

easily be added to the preprocessing pipeline in future development iterations for this project if a more robust motion artefact remove system is needed. As a result of the temporal nature of fNIRS data, and the importance of maintaining the temporal information about the underlying brain structures, this dimension of the data ought be preserved and leveraged during our modeling process. This means that noise removal options such as removing that chunk of data from analysis may be deleterious to the features learning in more complex modeling operations. For this reason, it will be wise for researchers maintain the use of algorithms that do not disrupt the temporal aspect of the data.

Data Cleaning and Labeling Structure. Another, more theoretical challenge, with the current dataset is the establishment of ground truth for the labeling configuration. Though there is no direct way around this challenge, we have instead opted for a more user focused approach in our pre-processing pipeline. Our pipeline does not bake in the the end labels of the data and instead allows the user (in our case the modeler) to define their own labels based on the simple string labels that are included within the current dataset. The user, rather than labeling the data directly by the task, can instead pass a dictionary of labels to the dataset when retrieving the data from the data warehouse. This allows for a task, such as a working memory task, to be relabeled to the user's needs in each particular modeling situation and allows for greater flexibility and re use of the currently indexed data within the dataset. Along with not baking the labels in, we have provided the ability to run unsupervised learning methods on data currently pulled from the database, which could allow for some interesting explorations on similarities in brain patterns when doing novel tasks, which may give information about common patterns in brain activation in varying task environments.

Though our current infrastructure allows for multiple multi-labeling schema to be easily implemented prior to model training, our current labeling schema pulls heavily from CLT. Our current labeling schema consider the load that is placed on three of the four mental components theorized to be involved in CLT: Working Memory Load, Visual Perceptual Load, and Auditory Perceptual Load. Within these three variables we assign either a 'Off' (0) in the case of no load on that cognitive resource, a 'Low' (1) in the case of low load on that cognitive resource, and a 'High' (2) in the case of high load on the cognitive resource. As a result of this labeling infrastructure, we can also expand the number of labels to better fit differing cognitive theories. As mentioned previously, particular models of working memory show that different types of working memory can be experimentally isolated. Figure 3 shows examples of sample tasks, included within our current dataset and potential label schema for those tasks using a task based approach. If a researcher wished to instead devise the labeling schema as a variable of an individual participant, rather than as a variable of the task. These features are thus far hand selected features, but there is the possibility to add more advanced statistical and mining techniques to derive these features from the underlying data. To validate this reduction it might be important to model both the raw signal as well as the hand reduced and algorithm generated features to ensure that the way in which the dataset is being reduced does not have a deleterious effect on the accuracy of the models our efforts are trying to generate. Unsupervised learning methods might also be added to the beginning of the pipeline, this would allow a researcher to check against their current label schema before performing more advanced modeling to help in gauging as to whether or not their current subdivision of labeling shows any preliminary effect before more advanced modeling techniques are pursued.

Our research could help not only provide guidance as to which model configurations and labeling techniques will yield the best accuracy measures on fNIRS data, but preliminary pattern mining and feature extraction may also contribute to a better understanding of why these model architectures and label configurations achieve this level of performance. Our plan to use both a top down approach with modeling using deep learning architectures, where our trade off will be explainabilty for increased accuracy, as well as a bottom up approach with direct pattern mining and unsupervised learning to aid choosing the proper inputs to feed into the models. Thus far, we have managed to keep on track with this possible set of contributions as well as expand on it in a few key ways. For one, our up front work in developing a way to pre-process all of the data within the database in a similar way can allows for greater replication of whatever results our finally modeling process yields. Another upshot of this is that the low level pattern mining over feature extracted data allows for a further level of hypothesis generation, which can better inform what models should seek to take as inputs, not only to drive greater classification performance from the models, but to also allow researchers some insight into why a certain model may out perform another. This allows the modeling to not only have the pragmatic benefit of getting closer to making more accurate and perhaps more timely predictions over this type of data, but to do so with greater interpritability, which is an Achilles's heel of more advanced deep learning classification methods [60].

Our model generating infrastructure, in an attempt to remain as agnostic to theory as possible, is designed to accommodate multiple multi-labeling schema. There is debate within the cognitive load theory (CLT) community as to whether or not cognitive load is a component of the task or

is a component of the individual performing the task [44]. Our labeling schema attempts to curtail this concern by instead allowing for model generation that is able to be labeled by predefined task conditions (the task the participant was engaged in invoked high visual perceptual load), by self report measure (the participant indicated that their visual load was high for this particular task), as well as by task performance (the decrements of performance during this task indicate the participants level of load was high in this task).

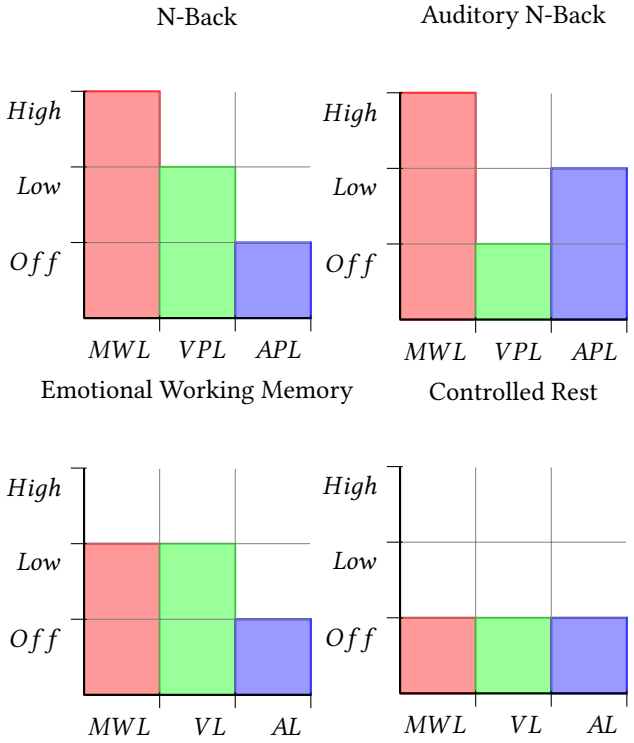

Figure 3: The Load Labels for Multiple Tasks used in Task Battery with labels in Working Memory Load (WML) in red, Vispual Perceptual Load (VPL) in green, and Auditory Perceptual Load (APL) in blue.

4 EVALUATION

In order to create a robust model with sound generalization capabilities, we accounted for the following main considerations pertaining to deep learning:

(1) **Modality of the dataset:** As described above, data gathered encompasses mental workload experienced by participants across 3 different classes - *MWL*, *VPL* and *APL*. At 2 different levels - *off* and *high* for each class. We trained 2 different models to predict mental workload at 2 levels *off* and *high* for MWL and VPL respectively. We also trained another model, to demonstrate the capability of the model to distinguish between perceptual modalities.

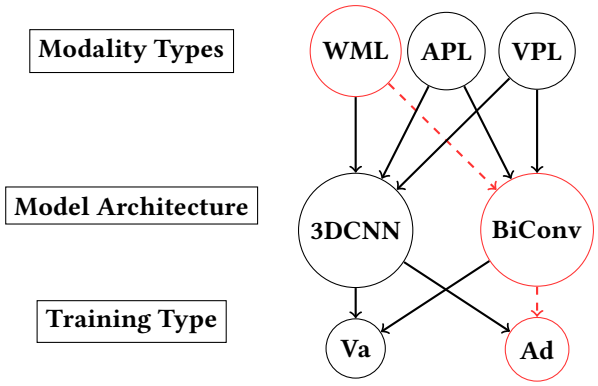

Figure 4: Models generated, with Modality Types being the different inputs (Working Memory Load, Auditory Perceptual Load, and Visual Perceptual Load), Model Architechtures being 3D Convolutional Neural Networks and Bi-Convoluational Long-Short Term Memory neural networks, and Training Types being both regular as well as Adversarial Training Methods. The path in red illustrates one of 12 possible models that will be trained for comparison.

(2) **Deep Learning Model Architecture:** Each sample of data obtained through high density fNIRS has spatio-temporal characteristics, as such it exhibits information across 4 dimensions viz. $TxCxWxH$ where T, C, W and H denote the time axis, the oxy and deoxy channels, the width of the channel matrix and the height of the channel matrix, respectively. Since these spatio-temporal characteristics are important to be maintained within the modeling architecture, we experimented with two commonly used and novel techniques. First, a variation of recurrent convolutional neural network-based architecture called Convolutional LSTM [67] and second, volumetric convolutions using 3D-CNN [61].

(3) **Training Methodology:** The data acquisition protocol followed resulted in a significant sampling bias towards low working memory workload (WM) population. As a result of this bias, conventional instance based supervised learning approaches, which map input space directly to WM labels, overfit low WM population. In order to combat the detrimental effects of sampling bias in the dataset, we utilized distance metric learning approach using Siamese Neural Networks [36]. Additionally, we also modified our training protocol to mimic adversarial training typically used in training GANs.

Siamese Neural Networks: Employing SNNs enabled us to combat the bias in our dataset, resulting in our best

performing model(s), originally introduced for verifying digitized signatures [13], SNNs have since been extended for one shot classification of images [36] and videos[65]. SNNs learn to generate lower dimensional representations of their high dimensional counterparts by learning to discriminate among pairs of similar and dissimilar inputs. This is achieved by using Contrastive loss [28] as the cost function. During training, an input pair is fed to the model with their true relationship y. In case of similar images, $y = 1$, otherwise $y = 0$. The model computes an embedding for each of the inputs in that pair. Contrastive loss maximizes the Euclidean distance between embeddings of dissimilar inputs, and minimizes the Euclidean distance between embeddings of similar inputs. The loss function is defined as follows:

$$L(\vec{X_1}, \vec{X_2}, y, W) = (1-y)^2 \frac{1}{2} {D_w}^2 + y \frac{1}{2} max(0, m - D_w)^2 \quad (1)$$

$\vec{X_1}, \vec{X_2}$ are a pair of input vectors from the high dimensional training set S_h, W is a parametric function that computes the corresponding lower dimensional mappings of X_i and D_w is the parameterized euclidean distance between the input pairs.

Adversarial Training: While employing a SNN did drastically improve the overall accuracy, we were also able to obtain a bump in the score by gradually changing the similarity factor between input pairs as the training progressed. In order to quantify the similarity between 2 inputs, we used mean SSIM [64] as the similarity metric. Originally used to quantify similarity between 2 images for image quality assessment, we extend SSIM over 3D image volume by using a 3-dimensional convolutional kernel. Using this training strategy we allow a randomly initialized model to become increasingly adept at generating separable and robust lower dimensional mappings for marginally similar and dissimilar input pairs.

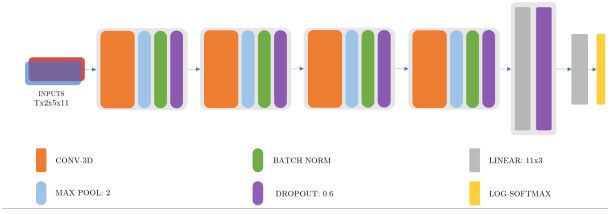

Figure 5: A diagram representing the 3D CNN model architecture.

Model Architectures

3D Convolutional Neural Network: We first employed a 3D Convolutional Neural Network (3D-CNN) based architecture [61]. 3D CNNs provide the ability to model spatio-temporal features unlike their 2D predecessors with the help of 3D convolutional and 3D pooling operations. Figure 5 shows the architecture of our 3D CNN-based model. One of the key hyper-parameters in this model is the 3D convolutional kernel depth and the number of 3D-CNN convolutional blocks. Through cross validation, we observed that a kernel depth of 2-1-1-1 for each of the respective CONV-3D blocks resulted in the highest validation performance.

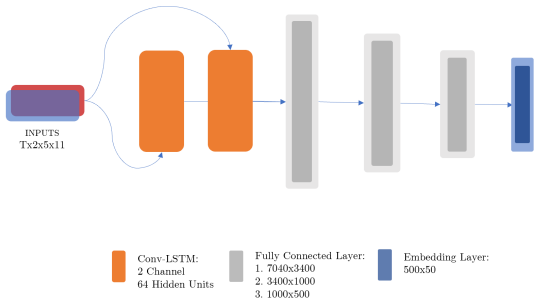

Figure 6: A diagram representing the Conv-LSTM based Bi-Directional LSTM model architecture.

Convolutional Long Short Term Memory Network: The second network we employed is based on a Convolutional LSTM (ConvLSTM) [67]. Conv-LSTM is a special case of 1 dimensional LSTMs in that each of the hidden, cell and input states as well as the input, forget and remember gates are 3D tensors, which are then operated on by a convolutional operation as opposed to a multiplication operation in dense LSTMs. In order to extract the spatio-temporal features from fNIRS data, we experimented with a 2-layer vanilla ConvLSTM and a single layer Bi-Directional ConvLSTM each of the blocks followed by a series of fully connected layers (FCNs). Our experiments concluded with the architecture delineated in Figure 6.

5 DISCUSSION

Performance metrics such as f1-score, precision, recall and accuracy for of each of the models formed by 12 (excluding auditory modality due to absence of labels for high CWL) total combinations of [modality-types]-[model-architecture]-[training-method] represented in Figure 7 are obtained by averaging confusion matrices from 10-fold cross validation (CV). For each CV run data is obtained by splitting the participants in 80:20 ratio for train and test set respectively. Figure 10 shows the performance of models trained using adversarial SNN training methodology only. Scores for models trained using vanilla supervised training have not been

shown for brevity as the former training method results in better performing models. However, in order to elucidate the superiority of adversarial training in our case, we have presented the confusion matrices of models obtained using each of the 2 methods in Figure 8.

Model	Precision	Recall	F1	Accuracy
	Working Memory			
3D-CNN	0.52	0.53	0.52	0.52
Bi-ConvLSTM	0.695	0.57	**0.62**	0.65
	Visual Perception			
3D-CNN	0.70	0.61	0.65	0.684
Bi-ConvLSTM	0.69	0.69	**0.69**	0.695
	Auditory vs. Visual			
3D-CNN	0.92	0.26	0.40	0.612
Bi-ConvLSTM	0.74	0.56	**0.63**	0.689

Figure 7: Classification performance of model architectures on different modalities.

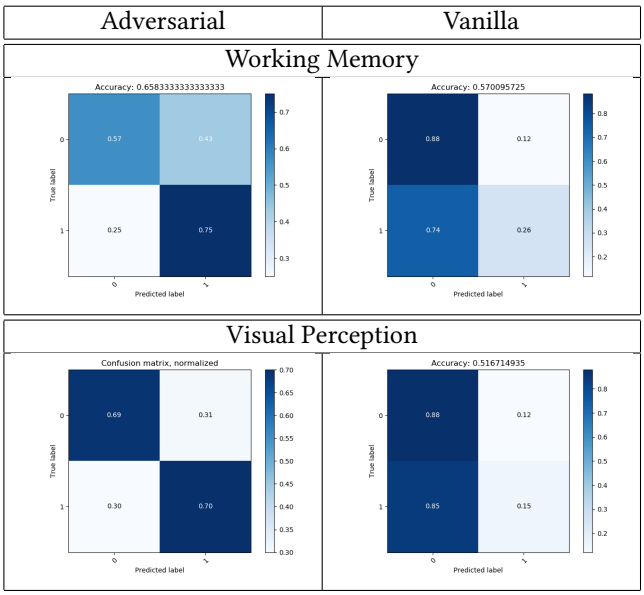

Figure 8: Robustness of Adversarial SNN Bi-ConvLSTM compared to vanilla Bi-ConvLSTM model to skew in the Working Memory dataset.

Although modeling techniques used by [53] result in a higher f1 score, it is to be noted that, our models are trained to generalize across participants as opposed [53] where the models are trained in between participants resulting in a participant agnostic model.

Significance of Adversarial SNN Architecture

The averaged confusion matrix for SNN based models are shown in Figure 9. A comparative confusion matrix for model trained using vanilla supervised learning approach with the same Bi-ConvLSTM architecture is presented to portray the significance of implementing Adversarial SNN. SNN based adversarial training methodology results in a model robust to inherent bias in the dataset producing a well balanced confusion matrix along the diagonal with higher f1-score. On the other hand, vanilla supervised training method is susceptible to bias in the dataset resulting in a skewed confusion matrix with a comparatively low f1-score.

6 CONCLUSION

Though the results generated by this initial exploration working with the current dataset were less than we had hoped, this does not mean that the use of this project was necessarily null and void. As the data had been in a less usable format than we had originally hoped, taking the time up front on the development of a system that allowed us to better handle that type of data may pay dividends later on through future interactions and explorations of the dataset.

Overall, the proposed architecture for developing a system to classify different states of CWL within human participants using fNIRS data may be a step in the right direction. Our methods, including a system for taking domain expertise to inform the models as well as using data mining and advanced classification techniques in order to feed back into the domain expertise could drastically improve the accuracy measures, transferablily, as well as the time window in which CWL classification with fNIRS data can take place. Further work is still needed within the PyoNIRS system, however. Including keeping things up to date in terms of the latest and greatest noise reduction, artefact removal and data preprocessing techniques. Other improvements to the system could include generating the results hypothesis tests, as well as simpler statistical models, off of features extracted from the data set that is selected, allowing fNIRS researchers who are not interested in the CWL or deep learning classification domains to use the software as a better way to organize their data and generate more interesting results. Several steps can be taken in order to make PyoNIRS more robust and usable to a wider audience within the fNIRS community, and these steps have currently been avoided in order to make room for the labeling and modeling features of the system.

The novel approach to labeling systems detailed within this report are not specific only to fNIRS data, these rules could apply for other types of brain data such as fMRI and EEG as why as other physiological measure provided the pre-processing steps needed within those types of data we replaced with the PyoNIRS pre-processing steps that are specific to the optical signals obtained from the fNIRS device. Though it may be difficult to translate the work presented in

this paper to other fields using advanced classification techniques it might be possible that this type of data warehousing and label selection based on domain expertise could help in other fields as well. Academic research should, perhaps, more broadly adopt some of the tools commonly used in industry in order to simply data analysis processes, which are over time cooked up independently from project to project.

Another possibility that could be opened up as a result of the adaptation of PyoNIRS, or of a PyoNIRS like system, is the possibility of creating a larger data repository amongst fNIRS research groups, which would allow the sharing of data sets coming from different labs, experimental procedures as well as devices. In it's current state the current code base of PyoNIRS would not be able to handle such an influx of data, but with some work and foresight the system should be able to reliably scale to be able to accept and index data in many different forms and translate it into a format that can be comparable and usable my other researchers. This idea is not without precedent within the neuroscientific community, who are beginning to see the benefits of openly sharing their MRI data and contributing to various open source projects such as as the Brain-Map or Human Connectome projects.

Many of the problems facing the CWL domain may be solved as a result of this sharing, and this ability to be better organize and draw insights from data. As the field continues to expand and continues to make increasing strides towards real time prediction of cognitive states then the ability to create other technologies that leverage these advances becomes more likely. It is crucial that these steps are therefore taken with care as the development of systems that rely on such predictions ought have a reasonable enough amount of accuracy measures before these adaptions are deployed. With a better handle and with better tools with with the classify and make sense of brain data, these issues will be minimized in the future.

REFERENCES

[1] A Farras Abdelnour and Theodore Huppert. 2009. Real-time imaging of human brain function by near-infrared spectroscopy using an adaptive general linear model. *Neuroimage* 46, 1 (2009), 133–143.

[2] Haleh Aghajani, Marc Garbey, and Ahmet Omurtag. 2017. Measuring mental workload with EEG+ fNIRS. *Frontiers in human neuroscience* 11 (2017), 359.

[3] Alfredo Ardila, Byron Bernal, and Monica Rosselli. 2016. How localized are language brain areas? A review of Brodmann areas involvement in oral language. *Archives of Clinical Neuropsychology* 31, 1 (2016), 112–122.

[4] Alan D Baddeley, Graham J Hitch, and Richard J Allen. 2019. From short-term store to multicomponent working memory: the role of the modal model. *Memory & cognition* 47, 4 (2019), 575–588.

[5] Wesley B Baker, Ashwin B Parthasarathy, David R Busch, Rickson C Mesquita, Joel H Greenberg, and AG Yodh. 2014. Modified Beer-Lambert law for blood flow. *Biomedical optics express* 5, 11 (2014), 4053–4075.

[6] Jeffrey W Barker, Ardalan Aarabi, and Theodore J Huppert. 2013. Autoregressive model based algorithm for correcting motion and serially correlated errors in fNIRS. *Biomedical optics express* 4, 8 (2013), 1366–1379.

[7] Erol Başar and Aysel Düzgün. 2016. The CLAIR model: Extension of Brodmann areas based on brain oscillations and connectivity. *International Journal of Psychophysiology* 103 (2016), 185–198.

[8] Wei Bi and James Kwok. 2013. Efficient multi-label classification with many labels. In *International Conference on Machine Learning*. 405–413.

[9] DA Boas, K Chen, D Grebert, and MA Franceschini. 2004. Improving the diffuse optical imaging spatial resolution of the cerebral hemodynamic response to brain activation in humans. *Optics letters* 29, 13 (2004), 1506–1508.

[10] David A Boas, Clare E Elwell, Marco Ferrari, and Gentaro Taga. 2014. Twenty years of functional near-infrared spectroscopy: introduction for the special issue.

[11] Erwin R Boer. 2000. Behavioral entropy as an index of workload. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 44. SAGE Publications Sage CA: Los Angeles, CA, 125–128.

[12] Sabrina Brigadoi, Lisa Ceccherini, Simone Cutini, Fabio Scarpa, Pietro Scatturin, Juliette Selb, Louis Gagnon, David A Boas, and Robert J Cooper. 2014. Motion artifacts in functional near-infrared spectroscopy: a comparison of motion correction techniques applied to real cognitive data. *NeuroImage* 85 (2014), 181–191.

[13] Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Säckinger, and Roopak Shah. 1994. Signature verification using a" siamese" time delay neural network. In *Advances in neural information processing systems*. 737–744.

[14] Colin F Camerer, George Loewenstein, and Drazen Prelec. 2004. Neuroeconomics: Why economics needs brains. *scandinavian Journal of Economics* 106, 3 (2004), 555–579.

[15] B Chance, Z Zhuang, Chu UnAh, C Alter, and L Lipton. 1993. Cognition-activated low-frequency modulation of light absorption in human brain. *Proceedings of the National Academy of Sciences* 90, 8 (1993), 3770–3774.

[16] Rebecca L Charles and Jim Nixon. 2019. Measuring mental workload using physiological measures: a systematic review. *Applied ergonomics* 74 (2019), 221–232.

[17] Michael W Cole, Jeremy R Reynolds, Jonathan D Power, Grega Repovs, Alan Anticevic, and Todd S Braver. 2013. Multi-task connectivity reveals flexible hubs for adaptive task control. *Nature neuroscience* 16, 9 (2013), 1348.

[18] Molly J Crockett, Luke Clark, Golnaz Tabibnia, Matthew D Lieberman, and Trevor W Robbins. 2008. Serotonin modulates behavioral reactions to unfairness. *Science* 320, 5884 (2008), 1739–1739.

[19] Deepika Dasari, Guofa Shou, and Lei Ding. 2017. ICA-derived EEG correlates to mental fatigue, effort, and workload in a realistically simulated air traffic control task. *Frontiers in neuroscience* 11 (2017), 297.

[20] Angelika Dimoka. 2012. How to conduct a functional magnetic resonance (fMRI) study in social science research. *MIS quarterly* (2012), 811–840.

[21] Gautier Durantin, Sebastien Scannella, Thibault Gateau, Arnaud Delorme, and Frederic Dehais. 2014. Moving Average Convergence Divergence filter preprocessing for real-time event-related peak activity onset detection: Application to fNIRS signals. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2107–2110.

[22] Johan Engström, Emma Johansson, and Joakim Östlund. 2005. Effects of visual and cognitive load in real and simulated motorway driving. *Transportation research part F: traffic psychology and behaviour* 8, 2

(2005), 97–120.

[23] Marco Ferrari and Valentina Quaresima. 2012. A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *Neuroimage* 63, 2 (2012), 921–935.

[24] Karl J Friston, Peter Jezzard, and Robert Turner. 1994. Analysis of functional MRI time-series. *Human brain mapping* 1, 2 (1994), 153–171.

[25] Alan Gevins, Michael E Smith, Harrison Leong, Linda McEvoy, Susan Whitfield, Robert Du, and Georgia Rush. 1998. Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. *Human factors* 40, 1 (1998), 79–91.

[26] Matthew F Glasser, Timothy S Coalson, Emma C Robinson, Carl D Hacker, John Harwell, Essa Yacoub, Kamil Ugurbil, Jesper Andersson, Christian F Beckmann, Mark Jenkinson, et al. 2016. A multi-modal parcellation of human cerebral cortex. *Nature* 536, 7615 (2016), 171.

[27] David Grimes, Desney S Tan, Scott E Hudson, Pradeep Shenoy, and Rajesh PN Rao. 2008. Feasibility and pragmatics of classifying working memory load with an electroencephalograph. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 835–844.

[28] Raia Hadsell, Sumit Chopra, and Yann LeCun. 2006. Dimensionality reduction by learning an invariant mapping. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, Vol. 2. IEEE, 1735–1742.

[29] Sandra G Hart and Christopher D Wickens. 1990. Workload assessment and prediction. In *Manprint*. Springer, 257–296.

[30] Suzana Herculano-Houzel. 2009. The human brain in numbers: a linearly scaled-up primate brain. *Frontiers in human neuroscience* 3 (2009), 31.

[31] Richard W Homan, John Herman, and Phillip Purdy. 1987. Cerebral location of international 10–20 system electrode placement. *Electroencephalography and clinical neurophysiology* 66, 4 (1987), 376–382.

[32] CJ Honey, O Sporns, Leila Cammoun, Xavier Gigandet, Jean-Philippe Thiran, Reto Meuli, and Patric Hagmann. 2009. Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences* 106, 6 (2009), 2035–2040.

[33] Shamsi T Iqbal, Xianjun Sam Zheng, and Brian P Bailey. 2004. Task-evoked pupillary response to mental workload in human-computer interaction. In *CHI'04 extended abstracts on Human factors in computing systems*. ACM, 1477–1480.

[34] Seongsik Jo, Rohae Myung, and Daesub Yoon. 2012. Quantitative prediction of mental workload with the ACT-R cognitive architecture. *International Journal of Industrial Ergonomics* 42, 4 (2012), 359–370.

[35] M Ahmad Kamran and Keum-Shik Hong. 2013. Linear parameter-varying model and adaptive filtering technique for detecting neuronal activities: an fNIRS study. *Journal of Neural Engineering* 10, 5 (2013), 056002.

[36] Gregory Koch. 2015. Siamese neural networks for one-shot image recognition.

[37] László Kocsis, Peter Herman, and Andras Eke. 2006. The modified Beer–Lambert law revisited. *Physics in Medicine & Biology* 51, 5 (2006), N91.

[38] Jun Liu, Amir Shahroudy, Dong Xu, and Gang Wang. 2016. Spatio-temporal lstm with trust gates for 3d human action recognition. In *European Conference on Computer Vision*. Springer, 816–833.

[39] Nikos K Logothetis and Brian A Wandell. 2004. Interpreting the BOLD signal. *Annu. Rev. Physiol.* 66 (2004), 735–769.

[40] Kevin Mandrick, Vsevolod Peysakhovich, Florence Rémy, Evelyne Lepron, and Mickaël Causse. 2016. Neural and psychophysiological correlates of human performance under stress and high mental workload. *Biological psychology* 121 (2016), 62–73.

[41] Bruce Mehler, Bryan Reimer, Joseph F Coughlin, and Jeffery A Dusek. 2009. Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation Research Record* 2138, 1 (2009), 6–12.

[42] Anat Mirelman, Inbal Maidan, Hagar Bernad-Elazari, Freek Nieuwhof, Miriam Reelick, Nir Giladi, and Jeffrey M Hausdorff. 2014. Increased frontal brain activation during walking while dual tasking: an fNIRS study in healthy young adults. *Journal of neuroengineering and rehabilitation* 11, 1 (2014), 85.

[43] Behnam Molavi and Guy A Dumont. 2012. Wavelet-based motion artifact removal for functional near-infrared spectroscopy. *Physiological measurement* 33, 2 (2012), 259.

[44] Fred Paas, Juhani E Tuovinen, Huib Tabbers, and Pascal WM Van Gerven. 2003. Cognitive load measurement as a means to advance cognitive load theory. *Educational psychologist* 38, 1 (2003), 63–71.

[45] Raja Parasuraman and Matthew Rizzo. 2008. *Neuroergonomics: The brain at work*. Vol. 3. Oxford University Press.

[46] Katherine L Perdue, Alissa Westerlund, Sarah A McCormick, and Charles A Nelson. 2014. Extraction of heart rate from functional near-infrared spectroscopy in infants. *Journal of biomedical optics* 19, 6 (2014), 067010.

[47] Russell A Poldrack. 2006. Can cognitive processes be inferred from neuroimaging data? *Trends in cognitive sciences* 10, 2 (2006), 59–63.

[48] Felix Putze, Sebastian Hesslinger, Chun-Yu Tse, YunYing Huang, Christian Herff, Cuntai Guan, and Tanja Schultz. 2014. Hybrid fNIRS-EEG based classification of auditory and visual perception processes. *Frontiers in neuroscience* 8 (2014), 373.

[49] Mikail Rubinov and Olaf Sporns. 2010. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* 52, 3 (2010), 1059–1069.

[50] Ryuji Sakakibara, Kuniko Tsunoyama, Osamu Takahashi, Megumi Sugiyama, Masahiko Kishi, Emina Ogawa, Tomoyuki Uchiyama, Tatsuya Yamamoto, Tomonori Yamanishi, Yusuke Awa, et al. 2010. Real-time measurement of oxyhemoglobin concentration changes in the frontal micturition area: An fNIRS study. *Neurourology and urodynamics* 29, 5 (2010), 757–764.

[51] Hendrik Santosa, Xuetong Zhai, Frank Fishburn, and Theodore Huppert. 2018. The NIRS brain AnalyzIR toolbox. *Algorithms* 11, 5 (2018), 73.

[52] Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. 2017. Deep learning with convolutional neural networks for EEG decoding and visualization. *Human brain mapping* 38, 11 (2017), 5391–5420.

[53] Susanne Schweizer and Tim Dalgleish. 2011. Emotional working memory capacity in posttraumatic stress disorder (PTSD). *Behaviour research and therapy* 49, 8 (2011), 498–504.

[54] Stoo Sepp, Steven J Howard, Sharon Tindall-Ford, Shirley Agostinho, and Fred Paas. 2019. Cognitive load theory and human movement: towards an integrated model of working memory. *Educational Psychology Review* (2019), 1–25.

[55] Adnan Shah and Abd-Krim Seghouane. 2013. Consistent estimation of the hemodynamic response function in fNIRS. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 1281–1285.

[56] Svetlana V Shinkareva, Robert A Mason, Vicente L Malave, Wei Wang, Tom M Mitchell, and Marcel Adam Just. 2008. Using fMRI brain activation to identify cognitive states associated with perception of tools and dwellings. *PLoS One* 3, 1 (2008), e1394.

[57] Erin Treacy Solovey, Audrey Girouard, Krysta Chauncey, Leanne M Hirshfield, Angelo Sassaroli, Feng Zheng, Sergio Fantini, and Robert JK

Jacob. 2009. Using fNIRS brain sensing in realistic HCI settings: experiments and guidelines. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*. ACM, 157–166.

[58] Olaf Sporns, Giulio Tononi, and Rolf Kötter. 2005. The human connectome: a structural description of the human brain. *PLoS computational biology* 1, 4 (2005), e42.

[59] Sungho Tak and Jong Chul Ye. 2014. Statistical analysis of fNIRS data: a comprehensive review. *Neuroimage* 85 (2014), 72–91.

[60] Naftali Tishby and Noga Zaslavsky. 2015. Deep learning and the information bottleneck principle. In *2015 IEEE Information Theory Workshop (ITW)*. IEEE, 1–5.

[61] Du Tran, Lubomir D. Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri. 2014. C3D: Generic Features for Video Analysis. *CoRR* abs/1412.0767 (2014). arXiv:1412.0767 http://arxiv.org/abs/1412.0767

[62] Martijn P van den Heuvel and Olaf Sporns. 2013. Network hubs in the human brain. *Trends in cognitive sciences* 17, 12 (2013), 683–696.

[63] Gael Varoquaux and Bertrand Thirion. 2014. How machine learning is shaping cognitive neuroimaging. *GigaScience* 3, 1 (2014), 28.

[64] Zhou Wang, Alan C Bovik, Hamid R Sheikh, Eero P Simoncelli, et al. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* 13, 4 (2004), 600–612.

[65] Lin Wu, Yang Wang, Junbin Gao, and Xue Li. 2018. Where-and-when to look: Deep siamese attention networks for video-based person re-identification. *IEEE Transactions on Multimedia* 21, 6 (2018), 1412–1424.

[66] Bin Xie and Gavriel Salvendy. 2000. Review and reappraisal of modelling and predicting mental workload in single-and multi-task environments. *Work & stress* 14, 1 (2000), 74–99.

[67] SHI Xingjian, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems*. 802–810.

[68] Jong Chul Ye, Sungho Tak, Kwang Eun Jang, Jinwook Jung, and Jaeduck Jang. 2009. NIRS-SPM: statistical parametric mapping for near-infrared spectroscopy. *Neuroimage* 44, 2 (2009), 428–447.