Performance of 1D-CNNs for EEG-Based Mental State Classification: Effects of Domain, Window Size and Electrode Montage

Veda Narayana Koraganji[1,2], Aidan J. Whelan[1], Akhil Mokkapati[1],
Juliette N. Zerick[1], R. Michael Winters[2] & Gregory F. Lewis[1,2]

***Abstract*— Deep learning paradigms have revolutionized the field of brain-computer interfacing and enabled the use of complex, nuanced methods for recognizing mental states. Prior work has demonstrated that these models can recognize the mental state in a variety of tasks, but few have specifically explored their performance concerning human factors that come into play with the use of brain-computer interfaces in day-to-day applications. For this research, we explored the use of 1D-convolutional neural networks to recognize two mental states–mental arithmetic and rest–from electroencephalograph signals. We focused our analysis on three parameters that affect the design and usability of brain-computer interfaces: input data representation (i.e. domain), window size (i.e. latency), and electrode montage (i.e. form-factor). In line with prior work, we found a clear bias in performance towards the frequency domain representation. We also found that training our model with short windows of time (i.e. 0.25s) provided close to peak accuracy. Furthermore, high accuracy was maintained with sparse electrode subsets of the full 10-20 system. We discuss these findings and how they can contribute to ongoing work to bring deep learning enabled brain-computer interfaces into day-to-day applications.**

I. INTRODUCTION

Mental state classification is a goal of many emerging fields of research, including neuroergonomics [1], neuroprosthetics [2], mental workload [3] and mental health studies [4], [5], [6]. It is increasingly adopted into new research areas and commercial applications to promote well-being, improve quality of life, and enhance the cognitive efficiency of subjects [3], [6], [7]. One promising method of mental state classification is through the analysis of real-time neurophysiological signals using Brain Computer Interfaces (BCIs). BCIs, introduced a few decades ago, originally suffered from limited signal processing capabilities, long training times, bulky setups, and were limited to neuroprosthetic applications [2], [8], [9]. However, with advances in imaging hardware, signal processing, and machine learning methods, the capabilities of BCIs significantly improved and enabled their use for mental state classification in more data-intensive applications such as military, entertainment, and research-related areas [6], [10], [11].

A typical modern BCI consists of a hardware stage and a software stage. The hardware stage acquires and filters the physiological signal and the software stage extracts significant features from the signal, classifies the mental state from those features, and responds by performing relevant application-dependent actions. Signal acquisition can be performed with many methods including electroencephalography(EEG) [9], electrocorticography [12], magnetoencephalography [13], or direct electrical signal acquisition in neurons [14]. Among these, EEG signals are preferred for day-to-day applications–and the focus of our work–owing to its excellent temporal resolution, high portability, and relatively low cost [15]. However, EEG suffers from low signal-to-noise ratio (SNR) and high inter-subject variability [16]. Software systems utilizing advanced machine learning (ML) methods increased the classification accuracy of BCIs above 70%, a commonly accepted threshold for BCI performance [17], but the accuracy of these ML classifiers largely depends on human expertise for feature extraction and suffers from low generalizability across subjects [18], [19].

In recent years deep learning has been adopted into BCIs to address some of these challenges. Deep learning networks can accept raw signal data and extract innate features, eliminating time-consuming preprocessing and reducing reliance on human expertise [19], [20]. However, these models do not offer the level of transparency as other machine learning models, making them less intuitive to comprehend the features extracted from the EEG signals. Nonetheless, several deep learning models were implemented in BCIs based on the application and the unique advantages those models provide [16], [19].

A. Current Challenges with BCIs

Although modern BCIs–using EEG acquisition and equipped with deep learning–have shown promise in research settings, they have yet to cross over into mainstream, day-to-day applications. Key to the success of widespread adoption of BCIs are several factors: high accuracy, small form factor, few calibrations, low latency, and subject independent performance [15], [16], [21]. In addition to those factors, deep learning presents its own questions such as the effects of different hyper-parameters, input representations, learning efficiency, and transfer learning. BCIs with many electrodes require more training and configuration time compared with fewer electrodes. Previous deep learning methods have also used window sizes on the order of seconds, introducing substantial delays in interaction and limiting their utility in day-to-day applications. These lingering challenges that prevent the full adoption of BCIs can be split into two categories: *system limitations* and *human factor limitations*. System limitations refer to inherent hardware and software constraints [21], [22]. Human factors on the other hand are

[1]Luddy School of Informatics, Computing, and Engineering, Indiana University Bloomington
[2]Kinsey Institute, Indiana University Bloomington
Correspondence: vkoragan@iu.edu

related to limitations arising from the usage of BCIs in day-to-day applications [15], [21].

1) System Limitations: Modern BCI hardware is capable of higher transfer rates and temporal resolution, but is vulnerable to ocular movement and requires high pressure and large contact area at electrode locations [4], [23]. On the software front, deep learning models yield higher accuracy and reduce dependence on human expertise, but require long training periods, frequent re-calibration, and are susceptible to inter-subject variability of EEG signals [16], [24]. They are also sensitive to input data representation and require proper domain selection (time- or frequency-domain representation of input).

2) Human Factors Limitations: In a typical laboratory setting, BCIs are equipped with bulky imaging systems: containing many electrodes and wires, affecting aesthetics, and adding significant weight. To offer high classification accuracy BCIs use window sizes on the order of a few seconds: increasing the overall latency for real-time mental state classification. For day-to-day applications, small aesthetic form-factor and quick real-time estimates are preferred [5], [25], [26]. BCIs need to be designed with fewer electrodes at aesthetically pleasing locations and small window size for rapid classification.

II. Background & Research Scope

A. Background

Several deep learning models have been applied to BCIs with varying degrees of success depending on the advantages that particular models provide [19]. *Representative* models such as Restricted Boltzmann Machines (RBMs), deep belief networks (DBN), and autoencoders were used primarily for feature extraction in BCIs as they are not ideal for classifications [27], [28]. *Discriminative* models such as the Multilayer Perceptron (MLP), Recurrent Neural Network (RNN), and Convolutional Neural Network (CNN) can be used for feature extraction and classification. [29] adopted a DBN-RBM algorithm to detect sleep spindle based on Power Spectral Density (PSD) features extracted from sleep EEG signals and achieved an F-1 of 92.78% on a local dataset. In [30], a combination of manually extracted features and deep autoencoder neural networks for classification on sleep state data yielded 80.4% accuracy. In recent years CNNs have gained popularity for EEG-based classification by consistently offering very high accuracies with end-to-end classification capabilities: the networks take EEG data and produce classification bypassing feature extractions and other steps. [31] reported an accuracy of 86% using a CNN for feature extraction and classification. [32] used a 1D convolution layer on multiclass and binary class datasets and reported 99.7% and 98% accuracies. [33] reported accuracy above 99% using a hybrid time-frequency input representation with a 2D CNN. Several other studies such as [34], [35], [36] have also utilized CNNs with different architectures, window sizes, and electrode montages. But these studies did not focus on the performance of those classifiers concerning day-to-day applications.

B. Research Scope

For successful crossover of deep learning enabled BCIs into day-to-day applications, we argue that their performance be evaluated on the following qualities: proper selection of domain to maintain high accuracy, small form-factor suitable for daily use, low latency to minimize delay, resistance against inter-subject variability, and transfer learning to reduce the need for re-training.

Prior work on BCIs utilizing deep learning has been focused either on the domain–using time domain or frequency domain or a hybrid time-frequency representation of the data–or minimizing the number of electrode channels used [36], [33], [32]. Numerous works have explored the usage of BCIs in day-to-day applications such as driving, aviation, gaming, attention monitoring, and psychotherapy [37], [11], [38], [6], [4] but fall short in comprehensively addressing the human factors. Most of these works used data collected in a laboratory setting and restricted their focus to addressing only some of the above qualities; a comprehensive analysis studying the combined effect of these qualities remains undone.

To address this deficit, we have constructed a 1D-CNN, representing a BCI, and studied its performance against the combined effect of the domain, electrode montage, and window size. We selected electrode montages that either have significance based on prior literature or have small and aesthetically pleasing form-factor. We picked window sizes based on existing literature and small window sizes that might be suitable for day-to-day applications. We then tested the network performance for different electrode montages and window sizes for both time and frequency domains. Other system limitations of BCIs, including hyper-parameter optimization for CNNs are out of the scope of this work.

III. Materials and Methods

A. Dataset Description

For our experiment, we used the dataset ”EEG During Mental Arithmetic Tasks Dataset”, publicly available on PhysioNet and Kaggle [39]. It contains data from 36 healthy participants (Median age = 18.25, SD = 2.14) selected with inclusion criteria: normal vision and no mental, cognitive, or communication disabilities and exclusion criteria: usage of psychoactive medication, drug or alcohol addiction. Each participant data is composed of 23 EEG electrode data collected with Neurocom Monopolar headset (Ukraine, XAI-MEDICA) using the 10/20 international electrode scheme. In the experiment, subjects switched between performing the rest phase and the mental arithmetic phase. Serial subtraction–a standard stress-inducing protocol in many studies [40], [41], [42], [43]–was used during the mental arithmetic phase as an intense cognitive task. EEG data were recorded for 180s during the rest phase and 60s during the mental arithmetic phase, sampled at 500Hz and filtered with 0.5Hz high pass filter and 45Hz low pass filter. The database consists of 72 EDF format files: 1 rest EEG recording and 1 mental arithmetic EEG recording from each participant, and one CSV file with basic information about the participants.

B. CNN Architecture

A CNN is a neural network that performs convolution on input data to extract useful information (features). A typical CNN has the following layers: input layer, convolution layer, pooling layer, dropout layer, flattening layer, and dense layer. The first layer is the input layer which receives the preprocessed data and feeds it to the convolution layer. The convolution layer performs systematic convolution operations on the data using kernels (a filter that helps extract specific features from the data) and creates an abstracted representation of features called a feature map. These feature maps are large and highly sensitive to the location of features in data . To avoid these problems, pooling and dropout layers are used after convolution layers. Pooling layers use max-pooling operation on feature maps and downsample them, reducing computational load and making them less translationally variant. The dropout layer helps avoid overfitting the data by randomly setting the activation of some of the nodes to zero, a number determined by dropout rate. The flattening layer transforms the data into a 1D vector. Dense layers are fully connected layers that perform further computation and generate the classification. Readers interested in knowing more about CNNs are referred to [44], [45].

For our research, we constructed a 1D-CNN with a varying number of convolutional layers. The generic CNN has an input layer, convolution layers (immediately followed by pooling and dropout layers), a flattening layer after the last convolution layer, a final dropout layer, and three dense layers–the last of which makes the classification. The dropout rate for all the drop-out layers in the model was set to 0.5. A high-level illustration of a CNN with two convolutional layers is presented in figure 1.

C. Research Context

1) Parameter Selection: To test the suitability of the classifier for day-to-day applications, we evaluated its performance for various combinations of window sizes and electrode montages for both domains. We picked window sizes 1, 2, 4, and 8 seconds based on existing literature that used deep learning and EEG signals for classification [16], [35], [34] and compared them with smaller window sizes 0.5 and 0.25 seconds that are more suitable in day-to-day application scenarios. For electrode montages, we selected the prefrontal cortex and headband system that closely resemble commercially available EEG measurement products. These products were designed with day-to-day applications in mind; they have fewer electrodes at aesthetically pleasing locations. We also selected montages commonly used for Motor Imagery and P300 type datasets to test if the classifier can deduce information about the mental state from areas of the brain other than the prefrontal cortex. The list of electrode montages and the corresponding electrodes in the 10/20 system are given in table I.

TABLE I
ELECTRODE MONTAGES

Electrode Montage	EEG Electrodes
All	Fp1, Fp2, F3, F4, Fz, F7, F8, C3, C4, Cz, P3, P4, Pz, T3, T4, T5, T6, O1, O2, A1-A2
Prefrontal Cortex	Fp1, Fp2, F7, F8, A1-A2
Headband	Fp1, Fp2, F8, F7, T3, T4, T5, T6, O1, O2, A1-A2
Without PFC	F3, F4, Fz, C3, C4, Cz, P3, P4, Pz, T3, T4, T5, T6, O1, O2, A1-A2
Motor-Imagery System	Fz, C3, C4, A1-A2
P300	Pz, A1-A2

2) Preprocessing: The training and test sets were generated by randomly splitting the dataset into 27 and 9 subjects and downsampling the EEG data to 128Hz. Based on the window size and the electrode montage, the corresponding electrode data was extracted and segmented into non-overlapping window segments. The resultant window segments were processed into time and frequency domains. For the time domain, the input was constructed by combining the segmented windows for each of the extracted EEG channels

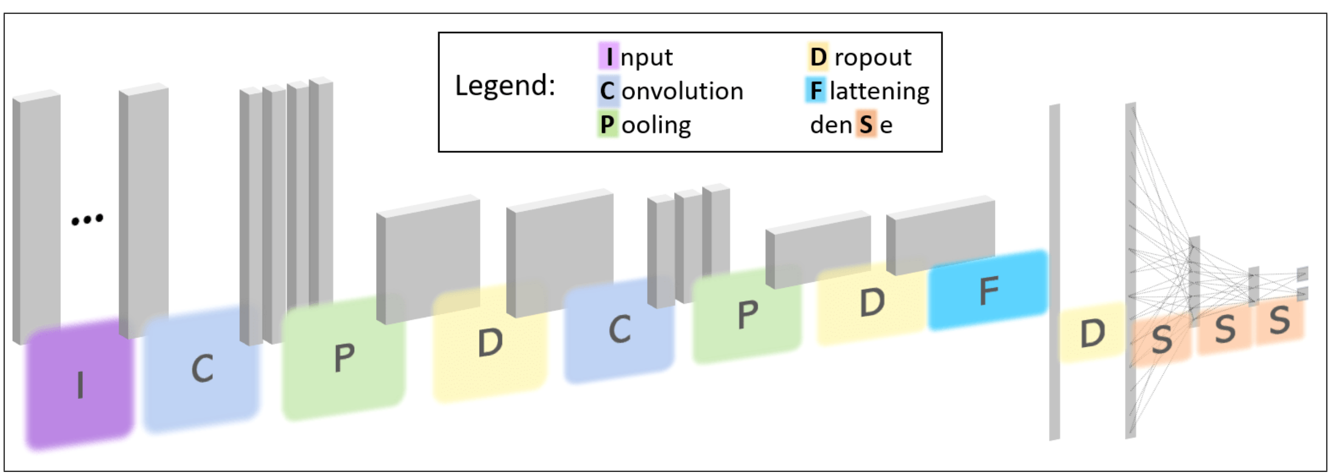

Fig. 1. A high-level depiction of our 2-layer CNN architecture.

into a 2D matrix, along with a one-hot vector containing the corresponding labels for rest or mental arithmetic states. For the frequency domain, a 2D matrix was constructed by calculating the power spectral density (PSD) from the segmented data using Welch's method [46] for each of the extracted EEG channels and a one-hot vector containing the corresponding labels.

3) Experiments: We began by selecting suitable hyperparameters for comparing the performance of four 1D-CNN architectures with a varying number of convolutional layers. Based on the suggestions by [47] we set the learning rate to 0.001 and the batch size to 64. The choice of window size for CNN architecture comparison was a difficult one, but we settled on 1s based on its overall performance in various studies using CNNs for EEG data [48], [35], [34]. Using all the electrode data provided in the dataset, we ran the four 1D-CNN architectures with both the domains for 50 trials each. Finally, while keeping the learning rate and batch size the same, we picked the best performing architecture and ran 5 trials (5-fold cross validation) for every window size and electrode montage combination for both the domains.

IV. Results

The performance metric chosen for this study was area under the curve (AUC) score [49].

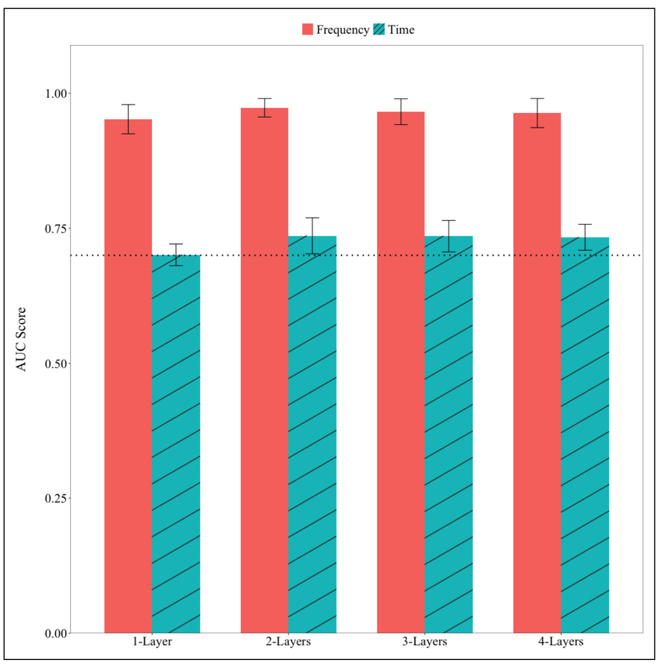

Fig. 2. Comparison of different CNN architectures. The dotted black line represents the acceptable threshold for BCIs.

A. CNN architecture

Figure 1 depicts the AUC scores of the classifiers for time and frequency domains. Improvements were marginal in magnitude, but the ANOVA indicated a significant improvement in AUC for the 2-layer architecture over the 1- and 4-layer architectures in both the time and frequency domain ($p < .05$) and greater performance than 3-layer architecture ($p > 0.1$), similar to [32].

B. Window Size

Figure 3 depicts the AUC scores of the classifier for the time and the frequency domain representation for window sizes ranging from 0.25s to 8s. We applied a multivariate analysis of variance (MANOVA) to test the impact of window size on both frequency and time domain AUC, using the all electrodes montage. Levene's test indicated unequal variance in the metrics in the frequency domain, so we selected the Games-Howell post hoc test for investigation of significant differences in all multiple comparison tests for consistency. The multivariate result was significant for window size, Pillai's Trace = 1.07, $F = 5.50, df = (10, 48), p < .001$. Post hoc tests on the time and frequency domain differences in AUC with respect to window size were used to follow-up on the MANOVA result. For the time domain model, the window size of 1s performed best, but the difference between 0.25s, 0.5s, and 2s was not significant. The 1s window was significantly better in AUC than 4s and 8s window lengths (all $p's < 0.01$) in the time domain model. For the frequency domain 1s window sizes were significantly better than larger sizes, (all $p's < 0.05$), and not significantly different than shorter sizes (all $p's > 0.05$).

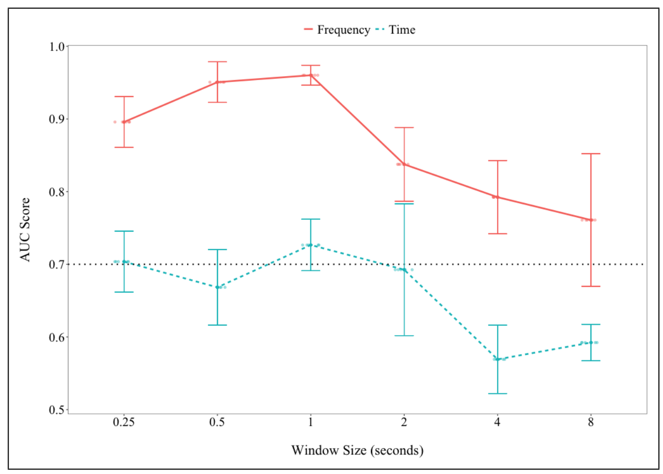

Fig. 3. Performance across window sizes(note that the x-axis is not linear). The dotted black line represents the acceptable threshold for BCIs.

C. Electrode Montage

Figure 4 depicts the AUC score of the classifier for the time and frequency domain representation for the previously mentioned electrode montages quantified with 1s windows. Our results indicate that for the time domain based model the classifier performance decreased with decrease in the number of electrodes. A Multivariate ANOVA was again used to test for the impact of electrode montage on model performance with AUC for the time and frequency domain models as performance metric. The multivariate result was significant for electrode montage, Pillai's Trace = 0.94, $F = 4.26, df = (10, 48), p < .001$. Due to unequal variances we again used the Games-Howell post hoc test to follow-up on montage differences for both domains. This test indicated that for the time domain models, the all electrode montage was only significantly better than the Muse and Motor-imagery system montages, $p < .05$. In the frequency-domain models, again the all electrode montage performed best, but only the contrast with the P300 system was significant, $p < .001$.

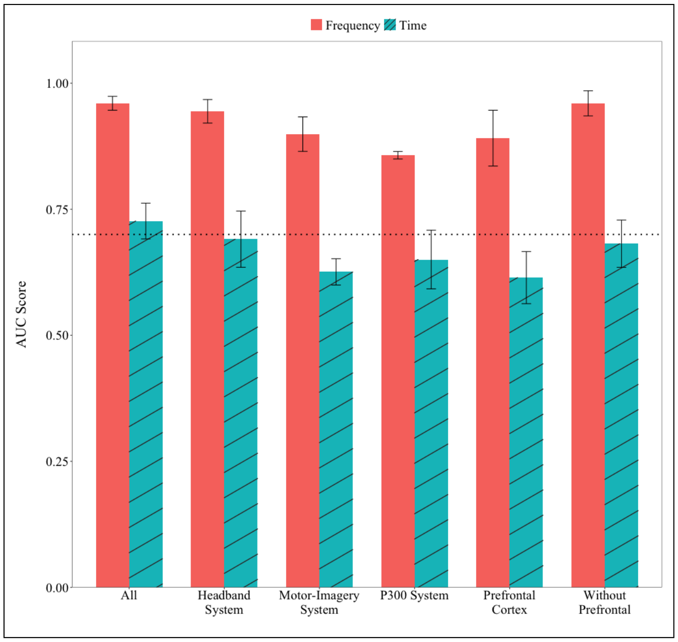

Fig. 4. Performance across electrode montages (window size = 1s). The dotted black line represents the acceptable threshold for BCIs.

D. Domain

A simple paired t-test was used to contrast AUC from the time and frequency domain representation of all models, yielding a test set of 180 comparisons. As expected, the frequency domain AUC was consistently greater [0.84 vs 0.62] across these tests, $t(179) = 34.12$, $p < .001$.

V. Discussion

Our research explored the effects of three independent variables; domain, window size, and electrode montage; on the performance of 1D-CNNs. We found that these variables affect the performance of BCI in day-to-day applications to recognize mental states from real-time EEG. Previous works have evaluated various CNN architectures for EEG-based mental state detection but did not evaluate their performance for day-to-day applications where the combined effect of the three independent variables plays an important role. A comparison of selected works that used either the same dataset or used CNN for EEG-based classification with our work has been presented in table II.

The particular aspects of BCI operation in day-to-day to applications that were studied in our work are categorized as: accuracy, latency, form factor, and inter-subject variability.

A. Accuracy

Our work explored how a difference in EEG basis representation (time- & frequency-domain) affected CNN performance on the same dataset. We found that domain had a significant effect on performance. In particular, the frequency domain outperformed the same dataset when represented as a time-domain signal. Although there is a considerable discussion within the field over the optimal metric to compare model performance when trained on different representations of the same data, we would like to point out that in all iterations the Frequency domain AUC was greater than the Time domain AUC; suggesting that the result is quite robust to analysis metric. This result is unexpected because the data contains the same underlying information in both domains as related through the Discrete Fourier Transform. Our results support prior work that has suggested that CNNs can more effectively learn one representation over the other [50], [51], [20]. This result has direct application to future BCI recognition algorithms targeting real-time mental-state recognition. Researchers exploring the performance of learning algorithms on similar cognitive tasks may expect that a frequency-domain input representation will significantly improve CNN performance relative to the same data in the time domain.

TABLE II

Comparison with prior work

Publication	Form-factor	Latency	Time-domain	Frequency-domain
[32]	No	Yes	Yes	No
[33]	No	No	No	Yes
[34]	No	Yes	No	Yes
[35]	Yes	Yes	Yes	No
[36]	No	No	Yes	Yes
Our work	Yes	Yes	Yes	Yes

B. Latency

Section IV-B presented a systematic analysis of the performance of both time- and frequency-domains for different window sizes. In a real-time BCI system, window size defines the minimum latency for mental-state identification from real-time EEG. Consistent with the literature, we found that window size significantly affected performance. In particular, longer windows had worse performance than shorter windows. One possible explanation for the higher performance for shorter windows is that shorter windows effectively created more training samples in the same amount of time.

Upon evaluating the performance of the classifier for varying time windows, we found an optimal learning performance at 1s time windows. Further, there was only a mild decrease in performance for 0.25s window size, for all the electrode montages, as seen in figure 5. Future researchers may be able to use this intuition to increase the performance of their learning algorithms. The optimal time window might be unknown a priori, and we argue that future researchers should tune input window size as an important parameter determining CNN performance. This has immediate implications for BCI user experience in the form of latency.

C. Form factor

In Section IV-C we used the same dataset to explore how the performance of our classifier varied with subsets of the all EEG channel montage. Instead of randomly sampling subsets of electrodes, we guided our selection to focus on electrode montages that are: widely used in BCI protocols (e.g. Motor Imagery, P300), commercially available EEG form-factors (e.g. Headband, Muse, Neurosky), and multimodal EEG-based BCIs (combined prefrontal fNIRS and EEG). We found that performance decreased only marginally when fewer electrodes were used. In practice, this means that BCI users can make similar levels of performance with fewer EEG channels. This result has immediate relevance to the design of BCI systems; if comparable performance can be achieved with fewer electrodes, then it would be advantageous to use fewer electrodes for the sake of simplicity and speed. Determining which electrode subset to use will depend upon context: what is the ”main” BCI protocol used, and how the position of the sensors may affect system usability.

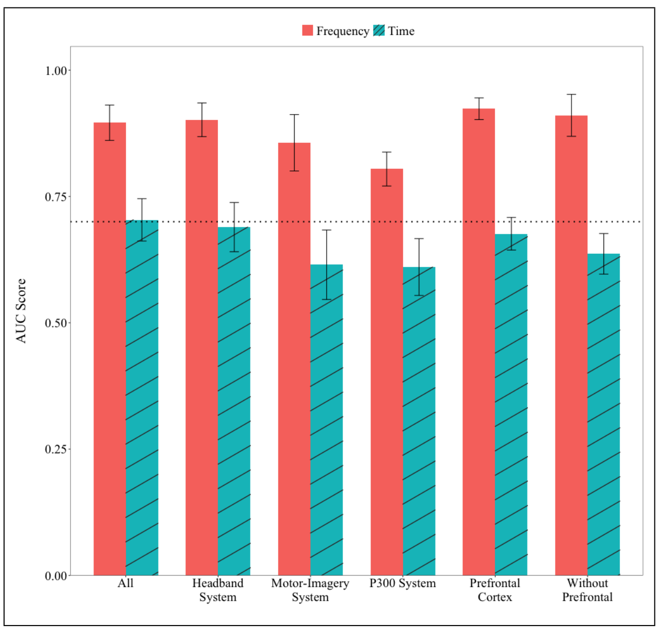

Fig. 5. Performance across electrode montages(window size = 0.25s). The dotted black line represents the acceptable threshold for BCIs.

D. Inter-subject variability and re-training

In our work, we followed a user-independent training approach, where the classifier is trained on a group of subjects and then tested on novel subjects. The classifier was able to generalize well for a large population displaying its resistance to inter-subject variability and its ability to generalize well for new subjects. This has various benefits, especially it does not need re-training for new subjects and can be used directly, which is desirable for day-to-day applications where subjects might change on a regular basis.

VI. Limitations & Future Work

The significant advantage we observed when using the frequency domain representation may be an artifact of the mental arithmetic task. It is also possible to measure mental state with an Event-Related Potential (ERP) paradigm. In this paradigm, time-windows are synchronized, or ”time-locked,” to precisely timed external events. It is possible that another dataset that uses precisely-defined external events (e.g. arrow) to classify mental state might have higher performance with a time domain representation. Future work can test this by finding an existing ERP dataset and applying a similar time and frequency comparison approach.

Our work demonstrated an optimal performance for our algorithms using time windows of 1s. However, performance with 0.25s windows only decreased slightly. There are some interesting usability questions that arises for shorter window sizes. In particular, are shorter windows with slightly lower accuracy better than longer windows with higher accuracy for overall accuracy of the BCI? Are BCI users willing to sacrifice some degree of classification performance if it meant that recognition could occur faster? We also found a similar effect by varying the electrode montage. Although the best performance occurred with all electrodes, montages with even one electrode (e.g. P300) performed comparably well. Future work would benefit from a better understanding of human factors in the trade-off between classification accuracy, latency and electrode montage. Because human factors can be abstracted from the particulars of many learning algorithms, future studies could simulate BCIs with different performance, window sizes and electrode montages to explore these effects systematically and without being tethered to any particular learning algorithm.

VII. Conclusion

An objective analysis of the the performance of 1D-CNNs with time and frequency domain representations of a publicly available and vetted dataset has been presented in this paper. Our analysis demonstrated that the frequency domain-based approach outperforms that of the time domain for the same dataset. We also studied the impact of possible limitations to BCI systems such as latency and form factor on the accuracy of the software classifier by experimenting with varied window sizes and electrode montages. Results showed that mental state estimation can be reliably performed on window sizes as short as or even shorter than 1 second. Furthermore, while using frequency domain methods, decreasing the number of electrodes did not necessarily decrease the performance of the CNN significantly. These findings, along with suggested improvements in other BCI limitations discussed in the paper, can be used to translate BCI research into day-to-day applications.

VIII. Acknowledgements

We are very grateful to Evan Nix for his help generating our results figures. We are also grateful to other members of the lab including Larry Zhang and Shyam Shah, who provided input and perspective at various points in the

process. We are grateful for the participants who contributed their EEG data and the people who created this dataset [39] and made it publicly available for research.

REFERENCES

[1] K. Mandrick, Z. Chua, M. Causse, S. Perrey, and F. Dehais, "Why a Comprehensive Understanding of Mental Workload through the Measurement of Neurovascular Coupling Is a Key Issue for Neuroergonomics?" *Frontiers in Human Neuroscience*, vol. 10, p. 250, 2016. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnhum.2016.00250

[2] R. Rothschild, "Neuroengineering Tools/Applications for Bidirectional Interfaces, Brain–Computer Interfaces, and Neuroprosthetic Implants – A Review of Recent Progress," *Frontiers in Neuroengineering*, vol. 3, p. 112, 2010. [Online]. Available: https://www.frontiersin.org/article/10.3389/fneng.2010.00112

[3] F. Dehais, A. Lafont, R. Roy, and S. Fairclough, "A Neuroergonomics Approach to Mental Workload, Engagement and Human Performance," *Frontiers in Neuroscience*, vol. 14, p. 268, 2020. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnins.2020.00268

[4] D. J. McFarland, J. Daly, C. Boulay, and M. A. Parvaz, "Therapeutic applications of BCI technologies," *Brain-Computer Interfaces*, vol. 4, no. 1-2, pp. 37–52, 2017. [Online]. Available: https://doi.org/10.1080/2326263X.2017.1307625

[5] T. Gateau, H. Ayaz, and F. Dehais, "In silico vs. Over the Clouds: On-the-Fly Mental State Estimation of Aircraft Pilots, Using a Functional Near Infrared Spectroscopy Based Passive-BCI," *Frontiers in Human Neuroscience*, vol. 12, p. 187, 2018. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnhum.2018.00187

[6] D. Erdogmus, A. Adami, M. Pavel, Tian Lan, S. Mathan, S. Whitlow, and M. Dorneich, "Cognitive State Estimation Based on EEG for Augmented Cognition," in *Conference Proceedings. 2nd International IEEE EMBS Conference on Neural Engineering, 2005.* IEEE, 2005, pp. 566–569. [Online]. Available: http://ieeexplore.ieee.org/document/1419686/

[7] R. K. Mehta and R. Parasuraman, "Neuroergonomics: a review of applications to physical and cognitive work," *Frontiers in Human Neuroscience*, vol. 7, p. 889, 2013. [Online]. Available: http://journal.frontiersin.org/article/10.3389/fnhum.2013.00889/abstract

[8] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1388245702000573

[9] E. A. Curran and M. J. Stokes, "Learning to control brain activity: A review of the production and control of EEG components for driving brain-computer interface (BCI) systems," *Brain and Cognition*, vol. 51, no. 3, pp. 326–336, 4 2003. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0278262603000368

[10] S. Bulárka and A. Gontean, "Brain-Computer Interface review," in *2016 12th IEEE International Symposium on Electronics and Telecommunications (ISETC)*, 10 2016, pp. 219–222.

[11] F. Dehais, A. Duprès, S. Blum, N. Drougard, S. Scannella, R. N. Roy, and F. Lotte, "Monitoring Pilot's Mental Workload Using ERPs and Spectral Power with a Six-Dry-Electrode EEG System in Real Flight Conditions," *Sensors*, vol. 19, no. 6, 2019. [Online]. Available: https://www.mdpi.com/1424-8220/19/6/1324

[12] C. Kapeller, K. Kamada, H. Ogawa, R. Prueckl, J. Scharinger, and C. Guger, "An electrocorticographic BCI using code-based VEP for control in video applications: a single-subject study," *Frontiers in Systems Neuroscience*, vol. 8, p. 139, 2014. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnsys.2014.00139

[13] J. Mellinger, G. Schalk, C. Braun, H. Preissl, W. Rosenstiel, N. Birbaumer, and A. Kübler, "An MEG-based brain–computer interface (BCI)," *NeuroImage*, vol. 36, no. 3, pp. 581–593, 2007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811907002261

[14] D. M. Taylor, S. I. H. Tillery, and A. B. Schwartz, "Direct Cortical Control of 3D Neuroprosthetic Devices," *Science*, vol. 296, no. 5574, pp. 1829–1832, 2002. [Online]. Available: https://science.sciencemag.org/content/296/5574/1829

[15] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain Computer Interfaces, a Review," *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012. [Online]. Available: https://www.mdpi.com/1424-8220/12/2/1211

[16] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: a systematic review," *Journal of Neural Engineering*, vol. 16, no. 5, p. 51001, 8 2019. [Online]. Available: https://doi.org/10.1088/1741-2552/ab260c

[17] A. Kübler, N. Neumann, B. Wilhelm, T. Hinterberger, and N. Birbaumer, "Predictability of Brain-Computer Communication," *Journal of Psychophysiology*, vol. 18, no. 2/3, pp. 121–129, 1 2004. [Online]. Available: https://doi.org/10.1027/0269-8803.18.23.121

[18] A. Bashashati, M. Fatourechi, R. K. Ward, and G. E. Birch, "A survey of signal processing algorithms in brain{\textendash}computer interfaces based on electrical brain signals," *Journal of Neural Engineering*, vol. 4, no. 2, pp. R32–R57, 3 2007. [Online]. Available: https://doi.org/10.1088/1741-2560/4/2/r03

[19] X. Zhang, L. Yao, X. Wang, J. Monaghan, D. McAlpine, and Y. Zhang, "A survey on deep learning-based non-invasive brain signals: recent advances and new frontiers," *Journal of Neural Engineering*, vol. 18, no. 3, p. 31002, 3 2021. [Online]. Available: https://doi.org/10.1088/1741-2552/abc902

[20] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: a review," *Journal of Neural Engineering*, vol. 16, no. 3, p. 031001, 6 2019. [Online]. Available: https://doi.org/10.1088/1741-2552/ab0ab5

[21] P. Aricò, G. Borghini, G. Di Flumeri, N. Sciaraffa, and F. Babiloni, "Passive BCI beyond the lab: current trends and future directions," *Physiological Measurement*, vol. 39, no. 8, p. 08TR02, 8 2018. [Online]. Available: https://doi.org/10.1088/1361-6579/aad57e

[22] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update," *Journal of Neural Engineering*, vol. 15, no. 3, p. 031005, 6 2018. [Online]. Available: https://doi.org/10.1088/1741-2552/aab2f2

[23] O. G. Lins, T. W. Picton, P. Berg, and M. Scherg, "Ocular artifacts in recording EEGs and event-related potentials II: Source dipoles and source components," *Brain Topography*, vol. 6, no. 1, pp. 65–78, 9 1993. [Online]. Available: https://doi.org/10.1007/BF01234128

[24] F. Fahimi, Z. Zhang, W. B. Goh, T.-S. Lee, K. K. Ang, and C. Guan, "Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI," *Journal of Neural Engineering*, vol. 16, no. 2, p. 026007, 4 2019. [Online]. Available: https://dx.doi.org/10.1088/1741-2552/aaf3f6 https://iopscience.iop.org/article/10.1088/1741-2552/aaf3f6

[25] J. Wolpaw and E. W. Wolpaw, *Brain–Computer InterfacesPrinciples and Practice*. Oxford University Press, 1 2012, vol. 23, no. 2. [Online]. Available: 10.1093/acprof:oso/9780195388855.001.0001

[26] A. Rezeika, M. Benda, P. Stawicki, F. Gembler, A. Saboor, and I. Volosyak, "Brain–Computer Interface Spellers: A Review," *Brain Sciences*, vol. 8, no. 4, 2018. [Online]. Available: https://www.mdpi.com/2076-3425/8/4/57

[27] J. Li, Z. Struzik, L. Zhang, and A. Cichocki, "Feature learning from incomplete EEG with denoising autoencoder," *Neurocomputing*, vol. 165, pp. 23–31, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231215004282

[28] S. Kumar, A. Sharma, K. Mamun, and T. Tsunoda, "A Deep Learning Approach for Motor Imagery EEG Signal Classification," in *2016 3rd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE)*. IEEE, 12 2016, pp. 34–39. [Online]. Available: http://ieeexplore.ieee.org/document/7941937/

[29] D. Tan, R. Zhao, J. Sun, and W. Qin, "Sleep spindle detection using deep learning: A validation study based on crowdsourcing," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 8 2015, pp. 2828–2831. [Online]. Available: https://ieeexplore.ieee.org/document/7318980/

[30] L. Fraiwan and K. Lweesy, "Neonatal sleep state identification using deep learning autoencoders," in *2017 IEEE 13th International Colloquium on Signal Processing its Applications (CSPA)*, 2017, pp. 228–231.

[31] A. Vilamala, K. H. Madsen, and L. K. Hansen, "Deep convolutional neural networks for interpretable analysis of EEG sleep stage scoring," in *2017 IEEE 27th International Workshop on Machine*

Learning for Signal Processing (MLSP). IEEE, 9 2017, pp. 1–6. [Online]. Available: http://ieeexplore.ieee.org/document/8168133/

[32] M. Saini, U. Satija, and M. Deo Upadhayay, "Light-Weight 1-D Convolutional Neural Network Architecture for Mental Task Identification and Classification Based on Single-Channel EEG," *arXiv e-prints*, p. arXiv:2012.06782, 12 2020.

[33] S. Madhavan, R. K. Tripathy, and R. B. Pachori, "Time-Frequency Domain Deep Convolutional Neural Network for the Classification of Focal and Non-Focal EEG Signals," *IEEE Sensors Journal*, vol. 20, no. 6, pp. 3078–3086, 3 2020.

[34] S. Vandecappelle, L. Deckers, N. Das, A. H. Ansari, A. Bertrand, and T. Francart, "EEG-based detection of the locus of auditory attention with convolutional neural networks," *eLife*, vol. 10, p. e56481, 4 2021. [Online]. Available: https://doi.org/10.7554/eLife.56481

[35] P. Keelawat, N. Thammasan, M. Numao, and B. Kijsirikul, "A Comparative Study of Window Size and Channel Arrangement on EEG-Emotion Recognition Using Deep CNN," *Sensors*, vol. 21, no. 5, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/5/1678

[36] O. Attallah, "An Effective Mental Stress State Detection and Evaluation System Using Minimum Number of Frontal Brain Electrodes," *Diagnostics*, vol. 10, no. 5, 2020. [Online]. Available: https://www.mdpi.com/2075-4418/10/5/292

[37] M. Takahashi, O. Kubo, M. Kitamura, and H. Yoshikawa, "Neural network for human cognitive state estimation," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'94)*, vol. 3. IEEE, 1994, pp. 2176–2183. [Online]. Available: http://ieeexplore.ieee.org/document/407565/

[38] K. Das, D. S. Rizzuto, and Z. Nenadic, "Mental State Estimation for Brain–Computer Interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 8, pp. 2114–2122, 8 2009.

[39] I. Zyma, S. Tukaev, I. Seleznov, K. Kiyono, A. Popov, M. Chernykh, and O. Shpenkov, "Electroencephalograms during Mental Arithmetic Task Performance," *Data*, vol. 4, no. 1, 2019. [Online]. Available: https://www.mdpi.com/2306-5729/4/1/14

[40] X. Yu, J. Zhang, D. Xie, J. Wang, and C. Zhang, "Relationship between scalp potential and autonomic nervous activity during a mental arithmetic task," *Autonomic Neuroscience*, vol. 146, no. 1, pp. 81–86, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566070208002609

[41] M. Gärtner, S. Grimm, and M. Bajbouj, "Frontal midline theta oscillations during mental arithmetic: effects of stress." *Frontiers in behavioral neuroscience*, vol. 9, p. 96, 2015.

[42] Q. Wang and O. Sourina, "Real-Time Mental Arithmetic Task Recognition From EEG Signals," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 21, no. 2, pp. 225–232, 3 2013. [Online]. Available: http://ieeexplore.ieee.org/document/6408308/

[43] A. Masuo, S. Kaga, and S. Kato, "Discrimination of Brain Activity using Near-Infrared Spectroscopy Signal Change in Prefrontal Cortex During Mental Arithmetic Task," in *2019 IEEE 8th Global Conference on Consumer Electronics (GCCE)*. IEEE, 10 2019, pp. 734–735. [Online]. Available: https://ieeexplore.ieee.org/document/9015532/

[44] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015. [Online]. Available: https://dx.doi.org/10.1038/nature14539

[45] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. [Online]. Available: http://www.deeplearningbook.org

[46] A. S. Al-Fahoum and A. A. Al-Fraihat, "Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains," *ISRN neuroscience*, vol. 2014, p. 730218, 2 2014. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4045570/

[47] I. Kandel and M. Castelli, "The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset," *ICT Express*, vol. 6, no. 4, pp. 312–315, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405959519303455

[48] A. Ravi, N. H. Beni, J. Manuel, and N. Jiang, "Comparing user-dependent and user-independent training of {CNN} for {SSVEP} {BCI}," *Journal of Neural Engineering*, vol. 17, no. 2, p. 26028, 4 2020. [Online]. Available: https://doi.org/10.1088/1741-2552/ab6a67

[49] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognition*, vol. 30, no. 7, pp. 1145–1159, 7 1997. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0031320396001422

[50] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces," *Journal of Neural Engineering*, vol. 15, no. 5, p. 056013, 10 2018. [Online]. Available: https://doi.org/10.1088/1741-2552/aace8c

[51] R. T. Schirrmeister and others, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapp.*, vol. 38, pp. 5391–5420, 2017. [Online]. Available: https://dx.doi.org/10.1002/hbm.23730