Deep Learning Forecast of Cognitive Workload Using fNIRS Data

Nicolas Grimaldi
Industrial and Systems Engineering
University of Florida
Gainesville, FL, USA
n.grimaldi@ufl.edu

Yunmei Liu
Industrial Engineering
University of Louisville
Louisville, KY, USA
yunmei.liu@louisville.edu

Ryan McKendrick
Mission Systems
Northrop Grumman
Falls Church, VA, USA
ryan.mckendrick@ngc.com

Jaime Ruiz
Computer & Information Science & Engineering
University of Florida
Gainesville, FL, USA
jaime.ruiz@ufl.edu

David Kaber
Industrial and Systems Engineering
University of Florida
Gainesville, FL, USA
dkaber@ise.ufl.edu

Abstract—

Introduction: In the domain of helicopter piloting, the pilot's performance is driven by many cognitive processes, demanding substantial cognitive resources. The pilot must maintain situation awareness and perform rapid decision-making. An objective of integrated helicopter technologies is to predict and effectively manage pilot cognitive workload to ensure safety and efficiency throughout flight. Methods: In this study, we collected data on seven participants, including three experienced pilots, using a UH-60V cockpit simulator to perform 46 distinct trials under various flight conditions. fNIRS neuroimaging was used to collect high-resolution neurophysiological data for exploring and forecasting cognitive workload using a collection of deep learning models. Model implementation: Three deep learning architectures are detailed in this work: a stacked LSTM model, a CNN-LSTM hybrid, and a transformer model. Results: An evaluation of three Seq2Seq models, each with two distinct forecasting lengths (10s and 30s), revealed LSTM-based architectures as superior performers for 10s forecasting tasks. Discussion: The LSTM-based models' superior performance suggested potential limitations with the transformer's self-attention mechanisms for our specific application. Surprisingly, the CNN-LSTM architecture did not surpass the stacked LSTM model's performance during forecasting tasks. Conclusion: Future research directions include exploring diverse time-series Seq2Seq methods and forecasting cognitive workload as ordinal measures, offering insights into shifting cognitive demands.

Keywords— cognitive workload, forecasting, adaptive human systems, fNIRS, deep learning

1. INTRODUCTION

1.1 Cognitive Workload

The task of helicopter piloting engages a complex ensemble of cognitive processes that contribute cumulatively to the pilot's cognitive workload [1]. From the perspective of information processing theory, each cognitive task represents a demand on a pilot's finite cognitive resources [2, 3]. Tasks requiring rapid decision-making, for example, impose substantial loads on working memory and decision-making processes, with pilots often having to quickly interpret dynamic flight data, fluctuating weather conditions, and unexpected emergencies. Multi-tasking further stretches these cognitive resources as it necessitates simultaneous attention switching, control, and coordination of multiple flight parameters and systems [4]. Maintaining situation awareness, on the other hand, requires continuous encoding, storage, and retrieval of environmental information, thereby adding to overall cognitive load [5]. Finally, complex manual control tasks require the integration of sensory input, motor control, and the executive function, further compounding overall cognitive workload [6]. In scenarios characterized by high-stakes, evolving conditions and intense emotions, these processes can push cognitive workload beyond capacity, posing safety risks.

To date, cognitive workload management mainly consists of cognitive workload monitoring and assessment using peripheral physiological or neurophysiological signals. Cognitive workload prediction, which requires a (near) real-time assessment of an individual's cognitive state, based on immediate physiological and behavioral data, is a strategy that has been investigated for managing cognitive workload [7-9]. While valuable for providing a snapshot of the cognitive load at any given moment, this modeling approach is inherently reactive and doesn't offer the foresight necessary to mitigate impending periods of high cognitive workload for pilots that forecasting approaches can provide.

Inspired by advancements in deep learning and successful applications of forecasting in various domains, such as finance for gold price prediction [10], and meteorology for "nowcasting" convective precipitation [11], our study proposes a shift towards forecasting of cognitive workload. This approach aims to project the future state of cognitive workload based on current and past data, thus offering a lookahead into potential changes in cognitive demands. This approach can

979-8-3503-1579-0/24/$31.00 ©2024 IEEE

support pilots through foresight of workload allowing for preparation and strategic allocation of cognitive aiding in response to upcoming complex tasks or potentially stressful situations. The versatile and cognitively complex nature of piloting tasks renders them an ideal case study for investigating the feasibility of cognitive workload forecasting. By transitioning from prediction to forecasting, we can facilitate proactive task management and decision-making towards enhancing piloting safety.

There are several methods that can be utilized to predict cognitive workload, encompassing both physiological and performance-based measures. Among these, a prominent approach has been the application of neurophysiological tools such as Functional Near-Infrared Spectroscopy (fNIRS). This non-invasive neuroimaging technology offers valuable insights into neural activity patterns associated with diverse cognitive workload levels. Its distinctive strengths lie in its portability, reduced susceptibility to movement artifacts, and lower sensitivity to the surrounding electromagnetic environment, all of which make fNIRS exceptionally suitable for real-world research contexts [12]. Crucially, a wealth of research has established the sensitivity of fNIRS to cognitive workload, demonstrating that workload induces detectable changes in cerebral blood flow, predominantly in the prefrontal cortex. This pivotal region in the brain, involved in executive functions, mirrors variations in cognitive workload through fluctuations in oxygenated and deoxygenated hemoglobin levels effectively captured by fNIRS [12-16]. These hemodynamic changes serve as indices of cognitive workload, thus providing a real-time, objective, and quantifiable measure.

1.2 Forecasting Models

Deep learning models are becoming increasingly common in forecasting applications. One of the unique strengths of deep learning lies in its capability to extract relevant features/patterns from unstructured datasets autonomously, bypassing the need for manual feature engineering, as often required in traditional machine learning approaches [17]. This autonomy not only reduces the dependency on domain knowledge but also mitigates the risk of human bias being introduced during the feature engineering process [18]. Therefore, the adoption of deep learning models for cognitive workload forecasting from physiological and neurophysiological signals may provide an efficient and objective framework for this application.

Sequence-to-sequence (Seq2Seq) modeling tasks represent deep learning algorithms adept for handling and reproducing sequential data, such as natural language and protein sequencing. These sequential models include the likes of recurrent neural networks (RNNs), long short-term memory (LSTM) models, temporal convolution networks (TCNs), and transformers. Their inherent capability to learn and generate sequential data makes such models ideal for forecasting time-series data. Conventional time-series forecasting tasks include stock prices [19], seasonal influenza [20], and weather [21] among others. Similarly, cognitive workload forecasting can be viewed as a Seq2Seq modeling problem, where the physiological signals and *apriori* classifications of workload states can be viewed as the input sequence and future states of workload can be viewed as the output sequence.

1.3 Objectives and Hypotheses

The primary objective of this study was to advance the domain of cognitive workload management by developing a deep learning-based model capable of forecasting cognitive workload from fNIRS data. This "look-ahead" capability offers significant potential utility for complex and dynamic tasks like piloting, where a short window of foresight could be instrumental in proactively managing cognitive resources and enhancing overall safety. On this basis, we propose the following hypotheses:

Hypothesis 1 (H1): Deep learning models can successfully extract meaningful temporal features from fNIRS data related to cognitive workload, enabling not only an accurate prediction of current cognitive workload levels but also a forecast of the workload.

Hypothesis 2 (H2): Owing to their capability to capture both spatial and temporal dynamics in the data, the hybrid CNN-LSTM model will demonstrate superior performance in forecasting cognitive workload, as compared to a stacked LSTM model.

2. METHOD

2.1 Participants and Apparatus

A total of seven participants, including three with previous piloting experience, were recruited for this experiment. All participants were male, aged 24-60, with a mean age of 37. Participants were then seated in a UH-60V cockpit simulator, which is designed to emulate the UH-60V Blackhawk Helicopter. Participants had access to two multi-function displays (MFDs), a control display unit (CDU), a flight display (FDVCP), and the mission system control unit (MSCU). The device used to measure participants' brain activity was an fNIRS Devices fNIR 2000 S system [22].

2.2 fNIRS Data Processing

The raw light intensity from 16 optodes at a 730nm channel, an 850nm channel, and an ambient channel were captured using the fNIRS neuroimaging system. Firstly, raw light intensity was motion corrected before undergoing a filtering process. The filtering procedure involved using a Daubechies 5 wavelet filter used to calculate wavelet coefficients with an α=0.1 threshold, followed by a 0.12Hz low-pass filter. Then, the data is converted into optical density, and eventually deoxygenated hemoglobin (Hb) and oxygenated hemoglobin (HbO2) is obtained. Afterwards, oxygenation levels (Oxy) and percent change in blood volume (HbT) are captured by finding the difference between HbO2 and Hb and the summation of HbO2 and Hb, respectively. Finally, the quadratic and cubic polynomials of the aforementioned hemodynamic values are computed. A comprehensive data preprocessing description can be found in [23].

2.3 Experiment Design

A total of 46 trials were conducted with varying flight conditions. For example, some conditions included weather events, some included reduced pre-flight check time, and some had no aberrant conditions, among others. Variations on each trial condition were designed in order to induce varying levels of cognitive workload. For example, convective activity "route arounds", engine fuel pressure warnings, and reduced pre-flight check times were introduced among other off-nominal events.

2.4 Task Procedure

During each trial participants were instructed to perform a full flight, consisting of pre-flight and flight phases. The participants played the role of a co-pilot (flying; PF) during the experiment, while a confederate played the role of pilot (not flying; PNF) and provided minimal guidance to the participant. During the pre-flight, the PF was asked to ensure that a series of checks were completed before takeoff. Afterwards, during flight operations the PF was tasked with monitoring the available instruments to verify whether any issue arose. When an emergency condition was identified, the PF was required to acknowledge it and request PNF assistance in resolving the issue, if necessary, via their own actions or commands to the PNF.

2.5 Workload Labeling

Previous observations utilizing Rasch labeling, a modeling technique adapted from item response theory, for cognitive workload demonstrated better model fits than other modeling approaches [23]. Rasch modeling was used to obtain a representation of users' capacity or latent trait. In this manner, three load levels (underload, optimal, and overload) can be extracted by accounting for: 1) task difficulty, 2) individual capacity, 3) observed performance, and 4) projected performance. A multinomial regression model was then trained using such labels to predict participant (PF) workload classifications in real time. The model structure was determined via grammatical evolution [24] and demonstrates similar performance to the Random Forest model highlighted by McKendrick et al. [23]. The full dataset, including both the workload classification confidences for each level and 64 processed fNIRS features, was used to train the selected forecasting models.

2.6 Forecasting Model Selection

The decision to implement and compare Seq2Seq models to forecast cognitive workload was motivated by the intrinsic nature of the problem at hand, specifically interpreting past physiological signals and workload trends and generating future workload confidence states. Two types of models are prominent in the forecasting field, including LSTM and transformer models. LSTM models [10, 19] have demonstrated superior performance compared to other forecasting techniques, such as RNNs, support vector regression, multilayer perceptron networks, and convolutional networks. Transformers are one of the most contemporary models in the field, yet they have already demonstrated superior performance over traditional methods in some forecasting tasks [20]. Several variations on the original transformer model, such as Informer [21] and Temporal Fusion Transformer [25], have also emerged and exploited the transformer's self-attention mechanism. These approaches have been shown to improve performance in forecasting tasks.

Several LSTM and transformer hybrids have been enhanced with convolutional neural networks (CNNs), given their capability to capture temporally localized features in datasets [10, 19, 26]. Based on our review of modeling methods, we decided to implement three models, including a stacked LSTM model, a CNN-LSTM model hybrid, and a convolutional transformer model to evaluate their performance at forecasting cognitive workload.

3. Model Implementation

It is necessary to understand the nature of any dataset before visiting model architectures. The input sequence for all models has a dimensionality of *[batch size, feature size, sequence length]*. Feature size encompasses the 64 processed fNIRS features and the confidence of each workload prediction (on a scale from 0-1), which were considered as features 65-67. The sequence length was defined as 120s, as this duration was found to be the best performing length for workload classification (before diminishing returns). The output sequence for each model has a dimensionality of *[batch size, workload levels, sequence length]*, where the selected sequence length was either 10s or 30s. Here we do not have a defined feature size dimension and instead have a dimension of size 3, corresponding to the number of predicted workload levels. The motivation for selecting 10s and 30s as the target output lengths was twofold. First, pilots require an initial adjustment period to respond to recommendations from automated guidance systems, which could take up to several seconds. Additionally, the limits of forecasting models in the field are uncertain, which led to a relatively wide disparity between the two selected output lengths.

The stacked LSTM architecture, as presented in Fig. 1, consists of two LSTM layers, to capture temporal dependencies, followed by a linear mapping layer. Similarly, the CNN-LSTM hybrid model, consists of a convolutional layer, to extract local temporal features, followed by a single LSTM layer, and finally a linear mapping layer, as depicted in Fig. 2. Finally, the convolutional transformer model, shown in Fig. 3, has two inputs, one input for the encoder and one for the decoder. The input for the encoder encompasses the entire 120s sequence length, however, the decoder input is defined as 10s or 30s (equal to the desired output length) to remove the autoregressive task in transformers and speed up real-time predictions. This method did not see a decrease in performance, as compared to an autoregressive technique in which each data point is produced iteratively. Both inputs are then processed differently, where the encoder input uses a convolutional layer, and the decoder input uses a linear layer. The inputs are then passed onto the transformer model (stacked encoder and decoder layers) and finally the output of the decoder undergoes a linear mapping.

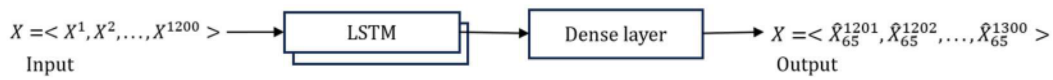

Fig. 1. Stacked LSTM architecture (10s model), displaying the output of a single confidence value sequence (65th feature).

Fig. 2. CNN-LSTM hybrid architecture (10s model), displaying the output of a single confidence value sequence (65th feature).

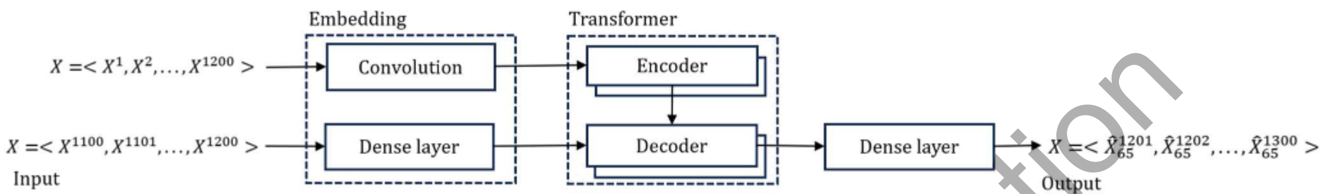

Fig. 3. Convolutional transformer architecture (10s model), displaying the output of a single confidence value sequence (65th feature).

The models were trained using a single random 80-20 split that consisted of data from all participants for a total of 45 out of 46 trials. A single hold-out evaluation trial was reserved for validation purposes. Sequences were generated from each trial. Each input sequence was individually normalized using its own mean and standard deviation (excluding confidence values), ensuring that the standardization was specific to that sequence rather than being based on the overall population or individual participants. This approach also accounted for the variations in trial conditions. The models use an Adam optimizer, and a mean absolute error (MAE) loss function. Models were trained with a batch size of 16 and executed over a total of 5 epochs for LSTM-based architectures and 20 epochs for the Transformer architecture. For the convolutional transformer model, we implemented a dropout rate of 0.1 across the embedding, positional encoding, transformer, and linear mapping layer.

4. Results

A total of six Seq2SSeq models were trained, including two differing forecasting lengths – 10s and 30s – for each of the three model architectures. We then evaluated the performance on an isolated hold-out evaluation trial which was not included in the training or testing sets. The total length of the trial was 1360s (22.67 minutes). It is worth noting that there were significant imbalances in the dataset in terms of workload level classifications. The distribution of workload classification from the multinomial regression model included: 'underload' 22.5%, 'optimal' 46.5%, 'overload' 31%.

A summary of the regression metrics, as well as computation time for each model forecast, is shown in Table I. Both MAE and mean squared error (MSE) were evaluated when comparing model regression performance (between forecasted and real confidence levels). The stacked LSTM and CNN-LSTM models achieved a comparable regression performance during both the 10s and 30s forecasting tasks. Meanwhile, the Transformer model underperformed, while achieving surprisingly similar metrics in both variations.

A depiction of the classification performance of each model for the hold-out evaluation trial is shown in Table II. During this trial, the continuous confidence levels were mapped to one of three classification levels (underload, optimal, and overload), based on the highest respective classification value. For the purposes of evaluation, the classification predictions were performed following the first 2 minutes (1200 datapoints) of the trial, which provided a baseline input for the first forecasted output. Subsequently, after each forecast the input data shifts either 10s or 30s to the most recently available 120s. At this point a new forecast is formulated and the process is repeated until the end of the trial.

When accounting for all metrics in the hold-out evaluation set, including macro and weighted precision, recall, and F1 score, the CNN-LSTM (94% accuracy) and stacked LSTM (93% accuracy) models achieved very similar performances during the 10s model variations, while the Transformer model (88% accuracy) underperformed once again. However, during the 30s variations, the stacked LSTM (87% accuracy) slightly outperforms all other architectures, followed by the Transformer model (85% accuracy) and finally the CNN-LSTM model (85% accuracy).

Models were trained and evaluated using an NVIDIA RTX 4090 GPU and an Intel i7-12700kf CPU. A comparison of the average computation time required for prediction of a single output was also conducted, as real-time forecasting can benefit from faster computation times. The CNN-LSTM model was found to achieve the fastest computation time during 10s and 30s forecasting tasks, achieving roughly half the computation time of the second fastest architecture (Transformer).

TABLE I. REGRESSION METRICS AND COMPUTATION TIME FOR ALL MODEL ARCHITECTURES AT DIFFERING FORECASTING LENGTHS (VARIATION). THEY ARE COMPUTED USING THE DIFFERENCE BETWEEN THE REAL AND FORECASTED CONFIDENCE VALUES FOR ALL 3 LOAD LEVELS IN THE TESTING SET.

Variation	LSTM		CNN-LSTM		Transformer	
	10s	30s	10s	30s	10s	30s
MAE	0.051	0.060	0.056	0.064	0.147	0.147
MSE	0.017	0.018	0.018	0.020	0.052	0.052
Time (ms)	23	25	7	8	15	18

TABLE II. CLASSIFICATION PERFORMANCE IN THE HOLD-OUT EVALUATION TRIAL, WHERE THE HIGHEST CONFIDENCE VALUE WAS CHOSEN AS THE LOAD LEVEL. THE MACRO AND WEIGHTED AVERAGE VALUES FOR PRECISION, RECALL, AND F1 SCORE OF EACH MODEL ARE SHOWN.

Variation	Metric	LSTM		CNN-LSTM		Transformer	
		Macro	Weighted	Macro	Weighted	Macro	Weighted
10 Second	Precision	0.92	0.93	0.92	0.93	0.85	0.88
	Recall	0.83	0.93	0.83	0.94	0.64	0.88
	F1 Score	0.87	0.93	0.87	0.93	0.70	0.87
30 Second	Precision	0.50	0.86	0.48	0.81	0.47	0.82
	Recall	0.52	0.87	0.40	0.85	0.44	0.85
	F1 Score	0.51	0.86	0.41	0.82	0.45	0.83

5. DISCUSSION

Contrary to our expectations (H2), the hybrid CNN-LSTM architecture did not demonstrate superior performance when compared to the stacked LSTM. In fact, diminished results were observed on the 30s variant of the models. However, it is noteworthy that the CNN-LSTM model did achieve much faster computation times than the stacked model in all variations. When examining the forecasting resolutions, 10s variants consistently yielded overall stronger classification results when compared to their 30s counterparts, suggesting that forecasting over a 30s interval may introduce uncertainties. This observation supports an assertion that a 10-second window may be preferred for accurate estimations of impeding cognitive workload.

A plausible explanation for the subpar performance of the Transformer model may lie in the intrinsic nature of self-attention mechanisms, which can compromise the retention of temporal information, as postulated by Zeng et al. [27]. Furthermore, uncertainties encountered during trials, particularly in the complex piloting tasks that we examined, can lead to unforeseen errors, such as component failures. These unexpected variables may induce unpredictable fluctuations in cognitive workload, which are challenging to account for solely based on physiological measures and historical workload data.

6. CONCLUSION

In this study, we investigated the potential of Seq2Seq models for predicting future cognitive workload levels, leveraging auxiliary physiological data with real-time workload classifications. In support of our primary hypothesis, we found that deep learning models can successfully extract meaningful temporal features from fNIRS data and classified workload confidence states to enable accurate forecasting of cognitive workload levels. LSTM-based model architectures emerged as notably efficient, in contrast to the Transformer model, which revealed limitations, possibly due to self-attention mechanisms [27]. Additionally, we showed that for a 10s forecasting length, we can precisely and reliably estimate changes in cognitive workload. However, extending the forecast to 30s introduces greater inaccuracies, potentially hindering practical application in the piloting context. These findings lay a foundational steppingstone for future research in cognitive workload forecasting.

Looking ahead, prospective research should include a more comprehensive analysis of a wider array of time-series forecasting methods, including deep learning models such as, the TCN model [28], TiDE model [29], and N-HiTS [30], as well other traditional statistical techniques and regression models. Further work should also consider the forecasting of cognitive workload levels strictly as ordinal measures – such as “low”, “medium”, and “high”. Leveraging loss functions and neural architectures intended to interpret and more directly predict such workload levels.

In the context of piloting tasks, forecasting cognitive workload has potential for significant safety advancements. Through real-time workload forecasting, adaptations to pilot guidance systems, such as providing information aiding during spikes in workload, could prove imperative to preventing pilot errors. However, some limitations in forecasting need to be identified. Type I predictions errors, or incorrectly predicting a change in workload, could result in unnecessary adjustments to a guidance system. In addition, forecasting protocols could also lead to pilot over-reliance on guidance systems to proactively support flight operations during pending impediments. Such limitations should be closely monitored during applications of forecasting frameworks for pilot guidance systems. In conclusion, the integration of cognitive workload forecasting in pilot guidance represents a significant stride towards enhancing aviation safety and efficiency. As this technology matures, it promises not only to reduce pilot errors but also to set a new standard in the evolution of smart, adaptive flight systems.

ACKNOWLEDGEMENT

This work was partially funded by the U.S. Defense Advanced Research Projects Agency under contract #HR00112220004. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect these agencies' views.

REFERENCES

[1] Gaetan, S., Dousset, E., Marqueste, T., Bringoux, L., Bourdin, C., Vercher, J. L., & Besson, P. (2015). Cognitive workload and psychophysiological parameters during multitask activity in helicopter pilots. Aerospace medicine and human performance, 86(12), 1052-1057.

[2] Nachreiner, F. (1995). Standards for ergonomics principles relating to the design of work systems and to mental workload. Applied Ergonomics, 26(4), 259-263.

[3] Miyake, A., & Shah, P. (1999). Models of working memory (pp. 442-481). Cambridge: Cambridge University Press.

[4] Wickens, C. D. (2002). Multiple resources and performance prediction. Theoretical issues in ergonomics science, 3(2), 159-177.

[5] Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. Human factors, 37(1), 32-64.

[6] Parasuraman, R., & Mouloua, M. (Eds.). (2018). Automation and human performance: Theory and applications. Routledge.

[7] Hincks, S. W., Afergan, D., & Jacob, R. J. (2016). Using fNIRS for real-time cognitive workload assessment. In Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience: 10th International Conference, AC 2016, Held as Part of HCI International 2016, Toronto, ON, Canada, July 17-22, 2016, Proceedings, Part I 10 (pp. 198-208). Springer International Publishing.

[8] Boehm, U., Matzke, D., Gretton, M., Castro, S., Cooper, J., Skinner, M., ... & Heathcote, A. (2021). Real-time prediction of short-timescale fluctuations in cognitive workload. Cognitive research: principles and implications, 6(1), 1-29.

[9] Udo, B., Dora, M., Gretton, M., Castro, S., Cooper, J., Skinner, M., ... & Heathcote, A. (2021). Real-time prediction of short-timescale fluctuations in cognitive workload. Cognitive Research, 6(1).

[10] Livieris, I. E., Pintelas, E., & Pintelas, P. (2020). A CNN–LSTM model for gold price time-series forecasting. Neural computing and applications, 32, 17351-17360.

[11] Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Advances in neural information processing systems, 28.

[12] Pinti, P., Aichelburg, C., Lind, F., Power, S., Swingler, E., Merla, A., ... & Tachtsidis, I. (2015). Using fiberless, wearable fNIRS to monitor brain activity in real-world cognitive tasks. JoVE (Journal of Visualized Experiments), (106), e53336.

[13] Kojima, T., Tsunashima, H., Shiozawa, T., Takada, H., & Sakai, T. (2005). Measurement of train driver's brain activity by functional near-infrared spectroscopy (fNIRS). Optical & Quantum Electronics, 37.

[14] Tomioka, H., Yamagata, B., Takahashi, T., Yano, M., Isomura, A. J., Kobayashi, H., & Mimura, M. (2009). Detection of hypofrontality in drivers with Alzheimer's disease by near-infrared spectroscopy. Neuroscience letters, 451(3), 252-256.

[15] Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., & Onaral, B. (2012). Optical brain monitoring for operator training and mental workload assessment. Neuroimage, 59(1), 36-47.

[16] Gateau, T., Durantin, G., Lancelot, F., Scannella, S., & Dehais, F. (2015). Real-time state estimation in a flight simulator using fNIRS. PloS one, 10(3), e0121279.

[17] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.

[18] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.

[19] Lu, W., Li, J., Li, Y., Sun, A., & Wang, J. (2020). A CNN-LSTM-based model to forecast stock prices. Complexity, 2020, 1-10.

[20] Wu, N., Green, B., Ben, X., & O'Banion, S. (2020). Deep transformer models for time series forecasting: The influenza prevalence case. arXiv preprint arXiv:2001.08317.

[21] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021, May). Informer: Beyond efficient transformer for long sequence time-series forecasting. In Proceedings of the AAAI conference on artificial intelligence (Vol. 35, No. 12, pp. 11106-11115).

[22] fNIR Devices LLC. (n.d.). *fNIR Devices*. Retrieved January 21, 2024, from https://fnirdevices.com/

[23] McKendrick, R., Feest, B., Harwood, A., & Falcone, B. (2019). Theories and methods for labeling cognitive workload: Classification and transfer learning. Frontiers in human neuroscience, 13, 295.

[24] O'Neill, M., & Ryan, C. (2001). Grammatical evolution. IEEE Transactions on Evolutionary Computation, 5(4), 349-358.

[25] Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting, 37(4), 1748-1764.

[26] Huang, S., Wang, D., Wu, X., & Tang, A. (2019, November). Dsanet: Dual self-attention network for multivariate time series forecasting. In Proceedings of the 28th ACM international conference on information and knowledge management (pp. 2129-2132).

[27] Zeng, A., Chen, M., Zhang, L., & Xu, Q. (2023, June). Are transformers effective for time series forecasting?. In Proceedings of the AAAI conference on artificial intelligence (Vol. 37, No. 9, pp. 11121-11128).

[28] Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:1803.01271.

[29] Das, A., Kong, W., Leach, A., Sen, R., & Yu, R. (2023). Long-term Forecasting with TiDE: Time-series Dense Encoder. arXiv preprint arXiv:2304.08424.

[30] Challu, C., Olivares, K. G., Oreshkin, B. N., Ramirez, F. G., Canseco, M. M., & Dubrawski, A. (2023, June). NHITS: Neural Hierarchical Interpolation for Time Series Forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 37, No. 6, pp. 6989-6997).