Cross-session Classification of Mental Workload Levels using Recurrent Neural Networks

Vishnu KN
Neural Engineering Lab
Department of Bioscience and Bioengineering
IIT Guwahati Guwahati, India
kvishnu@iitg.ac.in

Srihari Madhavan
Neural Engineering Lab
Department of Bioscience and Bioengineering
IIT Guwahati Guwahati, India

Cota Navin Gupta
Neural Engineering Lab
Department of Bioscience and Bioengineering
IIT Guwahati Guwahati, India
cngupta@iitg.ac.in

***Abstract*—We propose a Long Short-Term Memory (LSTM) based RNN to tackle the three-level mental workload (MW) cross-session classification problem using an open-source electroencephalogram (EEG) dataset. We used average spectral power from all EEG frequency bands (delta, theta, alpha, beta, and gamma) and the approximate entropy of each trial as input features to the LSTM classification network. The proposed framework achieved a mean validation accuracy of 87.38% and a test accuracy of 43.79% on an unseen session of the same subject. Given the performance of our method, we speculate that our network may have tuned to the subject-specific features despite exhibiting generalized predictive capabilities. We also observed a decrease in classification accuracies (validation=86.63%) when the nonlinear feature was not considered. Our results suggest that LSTMs can model discriminative features of the different MW levels experienced by a passive brain-computer interface (PBCI) user.**

Keywords— LSTM, RNN, Deep Learning, Passive BCI, EEG, Mental Workload, Spectral features, Nonlinear Features, MATB-II Task.

I. Introduction

Accurate estimation of user mental workload (MW) levels is a useful tool for electroencephalogram (EEG) based Brain-Computer Interfaces (BCI)[1]. Passive EEG contains discernable correlates of mental workload levels and other task-related neural activities [2]. Automated online classification of the cognitive demand experienced in real-life scenarios like piloting aircraft and driving automobiles is a contemporary research problem in passive BCI technology [3], [4]. EEG is often the preferred modality for signal acquisition in BCI systems due to its high temporal resolution, portability, and cost-effectiveness [5]. However, it comes with a unique set of challenges arising from the nonstationarity of EEG signals across different recording sessions of the same task. The variability in recorded brain data across sessions is increased due to unrelated activities of subjects like eye-blinking, muscle movements, 50 Hz electrical interferences, etc., and variability due to the user's mental state such as mood or attention to the task [6]. Evidently, endowing cross-session classification abilities to an EEG classifier is a challenging problem [7].

Machine learning (ML) algorithms can easily approximate discriminative distributions of the high dimensional EEG data, conditional upon extracting task-relevant features from the signal [8]. Deep learning (DL) algorithms, such as recurrent neural networks (RNN) or convolutional neural networks (CNN) can entirely bypass this feature selection required by ML classifiers and completely automate signal classification [9], [10]. However, EEG-based BCI applications using deep neural networks (DNN) are currently limited due to the difficulties in obtaining a large sample size for analysis [11], [12]. DNNs tend to overfit the data if the depth of the model is not sufficiently supported by the sample size of training and validation datasets, limiting the broad usability of the learned model [13]. Long short-term memory (LSTMs) are tailor-made RNNs that can learn long temporal dependencies in sequential data [14]. The ability to approximate time-variant dynamical systems to learn temporal patterns that span large intervals makes LSTM-based RNNs a logical choice for classifying EEG signals [15], [16]. The memory state of the LSTM cell approximates the temporal dependencies in the data by recursively parsing both the input and memory state using the recurrent connection. LSTM architecture has many variants [17] in which Bi-directional LSTM cells (BiLSTMs) are particularly suited for EEG data since they model the input series in opposite directions using two parallelly stacked LSTM layers. The outputs of the central computational unit in both LSTM units are pooled together to model each element in the temporal sequence [18]. Even though LSTM is particularly effective in extracting temporal features by modeling the elements of the input sequence. CNNs, on the other hand, are adept at extracting high-level spatial features directly from the EEG time series. Despite the conceptual advantage of the recurrent operation in LSTM over the convolution operation in CNNs to parse temporal samples, LSTMs are still not as widely used for EEG classification as CNNs [12]. Overall, DNNs currently underperform ML classifiers in MW classification[11], [19].

Spectral features contain predictive information about the cognitive states of the brain [20]. Intra-session MW level classification with high accuracy can be achieved using spectral features [8], [21]. While average signal power in respective frequency bands signifies neural entrainment onto the ongoing dominant EEG oscillations [22], the approximate entropy quantifies the regularity of a time series signal and system complexity [23], [24]. We presumed that incorporating a nonlinear measure that quantifies trial variability could yield improved performance [25], as signal variability is a limiting factor in cross-session MW level classification [26].

MW activity may be reflected more in frontal, parietal, and occipital regions [27] than in other cortical regions. It can be detected reliably from even single-channel electrodes in the frontal regions [28]. Such differential localization of activity at certain scalp positions deems some electrodes less relevant than others for a given classification problem. This difference in the contribution of each electrode to the model's predictive power necessitates a manual intervention for selecting task-relevant channels (spatial feature selection). However, it may not hold true for DNNs as they can extract relevant features directly from the input data [9]. LSTMs have been used to model nonlinear data [29], [30], but more work is needed for neurophysiological signals. This article evaluates an LSTM network for its cross-session classification abilities by simple pooling of available EEG data sessions. To this effect, we adapted the LSTM network proposed by [31] and heuristically modified some of its parameters to suit the current

2022 IEEE Symposium Series on Computational Intelligence (SSCI) | 978-1-6654-8768-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/SSCI51031.2022.10022261

problem. The primary objective of this explorative analysis is to identify a robust RNN that can estimate universal discriminative patterns of MW levels.

The rigorous approaches for combining multiple sessions of EEG data are special cases of geometrical transformations performed on the pooled dataset, transforming them into a shared subspace defined in terms of the shared components across the multiple sessions [32]. However, we chose to address the cross-session classification problem by indiscriminately pooling multiple sessions of a given subject to build an inter-session signal classifier. We combined data from multiple sessions for training the network under the assumption that similarity in neural activity and inter-session signal variability will enable the algorithm to learn universal discriminative features of MW levels. Training the LSTM network on pooled data forces the network to assume that EEG signals from multiple sessions of a given subject are samples from the same statistical population. However, this assumption is false as EEG signals demonstrably exhibit non-stationarity and thus belong in different population distributions [33]. Nevertheless, we expected the proposed RNN to learn subject dependent features from pooled sessions.

II. Methods

A. Dataset and Cognitive Task

Reference [34] used a 64 Ag-AgCl electrode EEG acquisition device (ActiCap, Brain Products Gmbh) and an ActiCHamp amplifier (Brain Products, Gmbh). Only 62 out of the available 64 electrodes were used for EEG acquisition, placed in the 10-20 international system. The signal was sampled at 500 Hz. [34] recorded the original dataset and released it partially for the PBCI hackathon [19]. The open source EEG data from 15 subjects (19 males; 16 females, average age of 25 years) was recorded in three sessions (S1, S2, S3) conducted on three different days. The released dataset contains two recording sessions (S1 and S2) with their corresponding MW labels, and another session (S3) with unknown labels, reserved for the competition. The labels for the third session remained unreleased when this article was compiled. Therefore, all analyses except the competition results are based on only the first and second EEG recording sessions.

The MW levels were elicited using the Multi-Attribute Task Battery (MATB-II) protocol [35], a widely used BCI protocol designed by the National Aeronautics and Space Administration (NASA). It can elicit multiple MW levels by demanding more cognitive resources in the participating subject. The protocol involves varying simultaneous sub-tasks that generate graded mental workload levels depending upon the perceived task difficulty. The three MW levels generated by MATB-II task were labelled as "low," "medium," and "high". The sub-tasks included a "Tracking and System Monitoring" for the easy condition, a "Resource Management" task in addition to the previous one generates the medium levels of MW and a "Communication" task in addition to the previous two for the difficult task condition.

B. Signal Processing

The preprocessing pipeline implemented by [34] are, sequentially, extraction of task and resting state signals, removal of cardiac electrode, segmentation into 2s nonoverlapping epochs, referencing using the right mastoid electrode. Next high pass filtered (FIR; 1Hz), electrode rejection (average amplitude above 2 times the standard deviation across channels), SOBI for removing muscle, heart and eye components, low pass filter (FIR; 40 Hz), average re-referencing (CAR), and downsampling to 250Hz [19], [34].

Further, in our pipeline, we subtracted the mean resting-state amplitude of a given session from all the trials in that session to eliminate any offset in the signal [36]. We applied an additional windsorization step [37] for the classification using EEG time series.

C. Feature Extraction

We extracted five spectral features from each EEG channel for each trial namely the average power contained in the different EEG frequency bands such as delta (δ) (1-4 Hz), theta (θ) (4-8 Hz), alpha (α) (8-13 Hz), beta (β) (13-30 Hz), and, gamma(γ) (30-40 Hz) using Welch's power spectral density (PSD) estimate [38]. Then, approximate entropy which measures the signal variability was calculated for each trial across all channels. The concatenated feature vector to the LSTM classifier was thus of dimension 61*6. Further, we examined the contribution of non linearity to classification accuracy by eliminating the approximate entropy measure from the input feature vector and using only the PSD features for classification.

D. Deep Learning

We used the Deep Learning Toolbox from MathWorks to design RNN modules resembling convolutional modules in a typical CNN. The network architectures used for the EEG time series input is the LSTM – I (Fig. 1.B) network and for concatenated feature input is the LSTM – II network (Fig. 1.A). The LSTM-II was inspired by [31]. We adapted the original architecture and training variables to suit the present problem. We also changed the penultimate dense layer from 32 to 16 neurons (by trial and error), since the number of categories in our classification problem was halved compared to the original work.

We used two training strategies. The first being pooling of each subject's labeled EEG sessions S1 & S2 (which totaled to 894 trials) and a heldout 10 % of the above pooled as validation data. The second strategy uses either S1 or S2 for training with a 10 % held out as validation, while using the other (i.e S1 or S2) for testing. Each of the sessions (S1,S2 and S3) consisted of a total of 149 trials for each MW level per participant, bringing the total trials in individual sessions to 447 and in the combined two session to 894. Both training strategies were compared, and the strategy with pooled sessions was used for predicting the labels of the competition dataset (S3). A 10-fold cross-validation procedure was followed to verify the training accuracy of each subject.

III. Results

A. PBCI Hackathon

The concatenated feature vector used as input to LSTM network achieved the highest accuracy (Fig. 1.C, blue bars). Validation performance on EEG time series signals had considerably high variability across subjects compared to the concatenated feature vector. The proposed network (Fig. 1.A) worked best for the concatenated feature vector amongst other RNN architectures, and therefore was used to predict the labels of S3. The test accuracy (43.79%) on S3 was well over

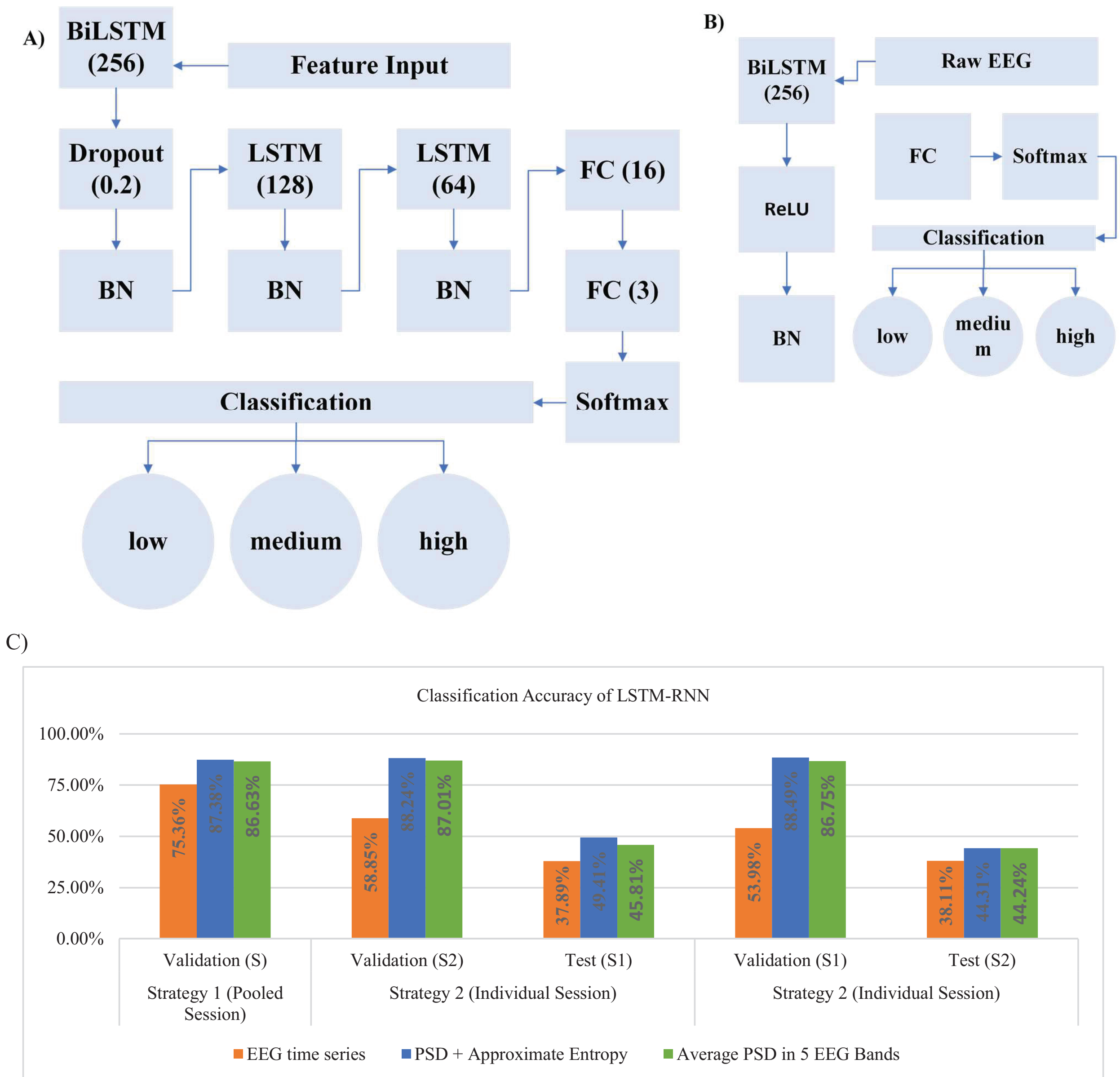

Fig. 1. **A)** RNN Architecture: The proposed RNN architecture (LSTM-II) contains one BiLSTM layer and two standard LSTM layers with 256, 128 64 recurrent neurons, respectively. Each of these LSTM modules subsequently has a batch normalization layer, except in the BiLSTM module, where a dropout layer (0.2) succeeds the BiLSTM layer, followed by the BN Layer. **B)** The LSTM-I network excludes the second and third LSTM modules and directly connects the ReLU then to BN to FC layer (BN: Batch Normalization Layer, FC: Fully Connected Layer). **C)** The classification accuracy using various strategies for training the classifier. (Orange: EEG time series, Blue: Spectral features and approximate entropy, Green: Spectral features only). Feature vector outperform EEG time series. Spectral features without entropy measure slightly decreased accuracy, the reduction was prominent in Test (S1) session.

both the chance level (33%) and adjusted chance level (38%) [19], but was lower than the validation accuracy (87.38%) on the pooled session (S1+S2) [39]. Interestingly, our network overfitted the pooled training data from both sessions as if all the training samples were drawn from the same distribution.

Most winning methods were ML algorithms (5 out of 7) and all except the proposed RNN [39] consisted of a spatial feature selection step to reduce the dimensionality of the data. The two DNNs that performed above the chance level were the proposed RNN [39] and a shallow CNN [40]. The number of learnable parameters in the network [40] was only a couple of hundred (hence the name 'shallow') compared to our network, which had about a million learnable parameters. Two other RNNs and one CNN did not surpass the chance level performance, and it may be considered as suggestive of the critical role of data pooling in imparting universal classification abilities to a DNN.

B. Pooling multi-sessions for training

Fig. 1.C describes the network's performance when different data pooling strategies were used for training the RNN. Training the model on stratrgy-2 produced high validation accuracy on the held-out sessions (88.24 ± 6.44%

(S1) and 88.49 ± 6.54% (S2)). The validation accuracy of strategy-1 was slightly lower (87.49 ± 4.5%). The standard deviation on the strategy-1 was considerably lower than training using strategy 2, suggestive of having estimated parameters that are tuned to both the sessions (S1 & S2). Strategy-2 produced highly variable performance upon testing on S1 or S2 suggesting poor generalizability. This is most likely a result of having trained only on intra-session data, thereby limiting the estimation of the essential nonstationary characteristics of EEG signals.

C. Nonlinear Feature

When the approximate entropy feature was eliminated from the concatenated feature vector, the accuracy reduced slightly (Fig. 1.C green bars). This result seemingly suggests that directly providing the network with a measure of trial variability enables better classification.

IV. Discussion

A total of five out of the ten models accepted for the competition were DNNs, and the rest were ML classifiers. However, out of the seven winning methods, only two methods relied on a DL algorithm for classification, the proposed LSTM-based RNN [39], and a shallow CNN [40]. The success of the shallow CNN is facilitated by the inclusion of task-specific priors on the spatial channels. Therefore, unlike the proposed RNN in this work, [40] has not utilized the depth of the network to extract relevant and optimal spatial features directly from the data. Since the CNN network was a 'shallow' network due to the low number of learnable parameters, our network was the only 'deep' neural network with desirable performance. The winning team used support vector machines for the classification along with a Riemannian-distance (RD) based electrode selection for spatial feature extraction. The robustness of using RD for feature extraction is attested by the fact that four out of the seven winning teams in the competition used it for feature extraction [19].

These results assert that ML algorithms still dominate EEG classification problems aided by spatial feature selection. In contrast, the proposed RNN achieved promising results without explicit spatial feature selection. The result agrees with the notion that DNNs do not need spatial feature extraction to model the high-dimensional EEG data. These results motivate further exploration of RNN architectures that may outperform current state-of-the-art methods.

The tuning of our network to the training data may also be attributed the dimensionality of the input. The EEG time series is continuous in the temporal dimension, while the extracted features are signal measurements averaged over a single trial. Therefore, based on the results, it can be hypothesized that the lower validation accuracy using EEG time series signals may have been due to decreased overfitting. Also, it is supported by the high-dimensional representation of the time domain EEG signals which are ideal for LSTM networks. However, we chose the concatenated feature vector over the EEG time series to predict the labels of S3, since the former had higher cross-validation accuracy. In hindsight, EEG signals may have resulted in a similar performance without overfitting. This speculation will be verified in a future study using EEG time series as input and training the classifier in an end-to-end fashion with no feature extraction. Training the network using the strategy-2 yielded higher classification accuracies than the strategy 1, though the performance variability is considerably lower in the latter. The similar performance using strategy-1 in comparison with stratrgy-2 is surprising. It may also suggest that the network has learned hypothetical signal features shared across the different intra-subject sessions.

The presence of both intra-session and inter-session signal variability in the pooled data increases the heterogeneity in the training dataset and may facilitate the estimation of universal discriminative features [41]. This phenomenon may have enabled our network to perform comparably to the other ML-based methods even without the spatial feature extraction that is critical to the performance of winning ML pipelines, evidenced in the competition [19]. The contribution of single trial variability measured using approximate entropy resulted in higher classification accuracies suggesting that trial-variability information provides discriminative features for the network (Fig. 1.C; blue bars).

We identified two directions to follow based on the current results. The proposal is to continue working with indiscriminate multi-session pooling to mitigate cross-session variability in other large multi-session datasets and the second being employing mathematical tools for data fusion of multiple EEG sessions on the small dataset described in this article.

V. Conclusion

Two conclusions can be outlined based on our network's performance evaluated by the PBCI Hackathon organizers. Unlike state-of-the-art ML methods, LSTM networks perform on par with them without explicit spatial feature extraction. The second is that LSTMs can learn a generalizable representation of MW levels even on small datasets by pooling the multiple available sessions and tackle the cross-session variability observed in nonstationary systems like EEG. The cross-session predictive capabilities exhibited by the proposed RNN indicate that our hypothesis of inter-session variability and inter-subject similarity positively contributing to the classification performance cannot be rejected. Further, we demonstrated that EEG time series could be used for MW classification using LSTM networks. Even though concatenated features outperformed the EEG time series, the results suggest that LSTM networks could be used to model EEG signal directly in time domain.

Acknowledgment

The MHRD Doctoral Scholarship from the Government of India funded Vishnu KN. Srihari Madhavan was employed as an Assistant Project Engineer sponsored by North East Centre for Biological Science and Healthcare Engineering (NECBH) grant, India, (Project No: 177). Cota Navin Gupta's time was funded by the Department of Science and Technology (DST), India, and the Swedish Research Council Grant (Project No: 14). The MATLAB code can be found at the below address. https://github.com/NeuralLabIITGuwahati/RNNClassifier

References

[1] C. Berka *et al.*, “EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks.,” *Aviation, space, and environmental medicine*, vol. 78 5 Suppl, pp. B231-44, 2007.

[2] A. Appriou, A. Cichocki, and F. Lotte, “Modern machine-learning algorithms: For classifying cognitive and affective states from electroencephalography signals,” *IEEE Systems, Man, and Cybernetics Magazine*, vol. 6, no. 3, pp. 29–38, Jul. 2020.

[3] A. Hernández-Sabaté, J. Yauri, P. Folch, M. À. Piera, and D. Gil, “Recognition of the Mental Workloads of Pilots in the Cockpit Using

EEG Signals," *Applied Sciences*, vol. 12, no. 5, 2022, doi: 10.3390/app12052298.

[4] S.-Y. Han, N.-S. Kwak, T. Oh, and S.-W. Lee, "Classification of pilots' mental states using a multimodal deep learning network," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 1, pp. 324–336, 2020, doi: https://doi.org/10.1016/j.bbe.2019.12.002.

[5] M. Rashid *et al.*, "Current status, challenges, and possible solutions of eeg-based brain-computer interface: A comprehensive review," *Frontiers in Neurorobotics*, vol. 14, Jun. 2020.

[6] J. van Erp, F. Lotte, and M. Tangermann, "Brain-Computer Interfaces: Beyond Medical Applications," *Computer*, vol. 45, no. 4, pp. 26–34, 2012, doi: 10.1109/MC.2012.107.

[7] H. Qu, M. Zhang, and L. Pang, "Mental Workload Classification Method Based on EEG Cross-Session Subspace Alignment," *Mathematics*, vol. 10, no. 11, 2022, doi: 10.3390/math10111875.

[8] Y. Zhou, S. Huang, Z. Xu, P. Wang, X. Wu, and D. Zhang, "Cognitive workload recognition using EEG signals and machine learning: A review," *IEEE Transactions on Cognitive and Developmental Systems*, pp. 1–1, 2021.

[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[10] 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. 051001, Aug. 2019, doi: 10.1088/1741-2552/ab260c.

[11] F. Lotte *et al.*, "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, Apr. 2018.

[12] 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, Apr. 2019, doi: 10.1088/1741-2552/ab0ab5.

[13] E. Lashgari, D. Liang, and U. Maoz, "Data augmentation for deep-learning-based electroencephalography," *Journal of Neuroscience Methods*, vol. 346, p. 108885, 2020, doi: https://doi.org/10.1016/j.jneumeth.2020.108885.

[14] S. Hochreiter and J. urgen Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[15] Xiao-Dong Li, J. K. L. Ho, and T. W. S. Chow, "Approximation of dynamical time-variant systems by continuous-time recurrent neural networks," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 52, no. 10, pp. 656–660, Oct. 2005.

[16] K. M. Tsiouris, V. C. Pezoulas, M. Zervakis, S. Konitsiotis, D. D. Koutsouris, and D. I. Fotiadis, "A Long Short-Term Memory deep learning network for the prediction of epileptic seizures using EEG signals," *Computers in Biology and Medicine*, vol. 99, pp. 24–37, 2018, doi: https://doi.org/10.1016/j.compbiomed.2018.05.019.

[17] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A Search Space Odyssey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.

[18] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.

[19] R. elle N. Roy *et al.*, "Retrospective on the first passive brain-computer interface competition on cross-session workload estimation," *Frontiers in Neuroergonomics*, vol. 3, Apr. 2022.

[20] S. Puma, N. Matton, P.-V. Paubel, É. Raufaste, and R. El-Yagoubi, "Using theta and alpha band power to assess cognitive workload in multitasking environments," *International Journal of Psychophysiology*, vol. 123, pp. 111–120, 2018, doi: https://doi.org/10.1016/j.ijpsycho.2017.10.004.

[21] P. Zarjam, J. Epps, and F. Chen, "Spectral EEG featuresfor evaluating cognitive load," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011, pp. 3841–3844. doi: 10.1109/IEMBS.2011.6090954.

[22] H. Zhou, L. Melloni, D. Poeppel, and N. Ding, "Interpretations of frequency domain analyses of neural entrainment: Periodicity, fundamental frequency, and harmonics," *Frontiers in Human Neuroscience*, vol. 10, Jun. 2016.

[23] S. M. Pincus, "Approximate entropy: A complexity measure for biological time series data."

[24] K. Natarajan, R. Acharya U, F. Alias, T. Tiboleng, and S. K. Puthusserypady, "Nonlinear analysis of EEG signals at different mental states," *BioMedical Engineering OnLine*, vol. 3, no. 1, Mar. 2004.

[25] U. R. Acharya, V. K. Sudarshan, H. Adeli, J. Santhosh, J. E. W. Koh, and A. Adeli, "Computer-Aided Diagnosis of Depression Using EEG Signals," *European Neurology*, vol. 73, no. 5–6, pp. 329–336, 2015, doi: 10.1159/000381950.

[26] Z. Yin and J. Zhang, "Cross-session classification of mental workload levels using EEG and an adaptive deep learning model," *Biomedical Signal Processing and Control*, vol. 33, pp. 30–47, Mar. 2017.

[27] E. Kutafina *et al.*, "Tracking of Mental Workload with a Mobile EEG Sensor," *Sensors*, vol. 21, no. 15, 2021, doi: 10.3390/s21155205.

[28] W. K. Y. So, S. W. H. Wong, J. N. Mak, and R. H. M. Chan, "An evaluation of mental workload with frontal EEG," *PLOS ONE*, vol. 12, no. 4, pp. 1–17, Apr. 2017, doi: 10.1371/journal.pone.0174949.

[29] J. Gonzalez and W. Yu, "Non-linear system modeling using LSTM neural networks," *IFAC-PapersOnLine*, vol. 51, no. 13, pp. 485–489, 2018, doi: https://doi.org/10.1016/j.ifacol.2018.07.326.

[30] Y. Wang, "A new concept using LSTM Neural Networks for dynamic system identification," in *2017 American Control Conference (ACC)*, 2017, pp. 5324–5329. doi: 10.23919/ACC.2017.7963782.

[31] P. Kaushik, A. Gupta, P. P. Roy, and D. P. Dogra, "EEG-Based age and gender prediction using deep BLSTM-LSTM network model," *IEEE Sensors Journal*, vol. 19, no. 7, pp. 2634–2641, Apr. 2019.

[32] M. Jiménez-Guarneros and P. Gómez-Gil, "Custom Domain Adaptation: A New Method for Cross-Subject, EEG-Based Cognitive Load Recognition," *IEEE Signal Processing Letters*, vol. 27, pp. 750–754, 2020, doi: 10.1109/LSP.2020.2989663.

[33] I. Albuquerque, J. Monteiro, O. Rosanne, A. Tiwari, J.-F. Gagnon, and T. H. Falk, "Cross-Subject Statistical Shift Estimation for Generalized Electroencephalography-based Mental Workload Assessment," in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 2019, pp. 3647–3653. doi: 10.1109/SMC.2019.8914469.

[34] Hinss *et al.*, "An EEG dataset for cross-session mental workload estimation: Passive BCI competition of the Neuroergonomics Conference 2021." Jun. 14, 2021. [Online]. Available: https://zenodo.org/record/4917218

[35] J. R. Comstock and R. J. Arnegard, "The multi-attribute task battery for human operator workload and strategic behavior research." [Online]. https://ntrs.nasa.gov/search.jsp?R=19920007912

[36] B. Chatterjee, R. Palaniappan, and C. N. Gupta, "Performance evaluation of manifold algorithms on a P300 paradigm based online BCI dataset," in *IFMBE Proceedings*, Cham: Springer International Publishing, 2019, pp. 1894–1898.

[37] U. Hoffmann, J.-M. Vesin, T. Ebrahimi, and K. Diserens, "An efficient P300-based brain–computer interface for disabled subjects," *Journal of Neuroscience Methods*, vol. 167, no. 1, pp. 115–125, Jan. 2008.

[38] P. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, Jun. 1967.

[39] S. Madhavan, V. K N, and C N, Gupta, "RNN Classification of Mental Workload EEG," presented at the Neuroergonomic Conference 2021, Germany, 2021. [Online]. Available: https://neuroergonomicsconference.um.ifi.lmu.de/wp-content/uploads/submissions/Madhavan-RNN%20Classification%20of%20Mental%20Workload%20EEG-226.pdf

[40] S. Sedlar, J. Benerradi, C. Le Breton, R. Deriche, T. Papadopoulo, and M. Wilson, "Rank-1 CNN for mental workload classification from EEG," presented at the Neuroergonomics Conference 2021, 2021.

[41] L. Raviv, G. Lupyan, and S. C. Green, "How variability shapes learning and generalization," *Trends in Cognitive Sciences*, vol. 26, no. 6, pp. 462–483, 2022, doi: https://doi.org/10.1016/j.tics.2022.03.007.