EEG-Based Cognitive State Assessment Using Deep Ensemble Model and Filter Bank Common Spatial Pattern

Debashis Das Chakladar[1*]   Shubhashis Dey[2]   Partha Pratim Roy[1]   Masakazu Iwamura[3]

[1]Indian Institute of Technology Roorkee, India. [2]Institute of Engineering and Management Kolkata, India .
[3]Osaka Prefecture University, Japan .

ddaschakladar@gmail.com[1*]

Abstract—Electroencephalography (EEG) is the most used physiological measure to evaluate the cognitive state of a user efficiently. As EEG inherently suffers from a poor spatial resolution, features extracted from each EEG channel may not be efficiently used for the cognitive state assessment. In this paper, the EEG-based cognitive state assessment has been performed during the mental arithmetic experiment, which includes two cognitive states (task and rest) of a user. To obtain the temporal as well as the spatial resolution of the EEG signal, we combined the Filter Bank Common Spatial Pattern (FBCSP) method and Long Short-Term Memory (LSTM)-based deep ensemble model for classifying the cognitive state of a user. Subject-wise data distribution has been performed due to the execution of a large volume of data in a low computing environment. In the FBCSP method, the input EEG is decomposed into multiple equal-sized frequency bands, and spatial features of each frequency bands are extracted using the Common Spatial Pattern (CSP) algorithm. Next, a feature selection algorithm has been applied to identify the most informative features for classification. The proposed deep ensemble model consists of multiple similar structured LSTM networks that work in parallel. The output of the ensemble model (i.e., the cognitive state of a user) is computed using the average weighted combination of the individual model prediction. This proposed model achieves 87% classification accuracy, and it can also effectively estimate the cognitive state of a user in a low computing environment.

I. INTRODUCTION

In literature, the internal cognitive state of a person is often expressed by mental workload, task demand, or vigilance detection. The cognitive state of a user can be recognized using subjective measures or physiological measures. In subjective measure, the workload level of a person can be assessed by rating-based questionnaires. There exist several types of physiological measures that include cardiac activity (Electrocardiogram (ECG)), respiratory activity, eye activity, and brain activity (Electroencephalogram (EEG), Functional magnetic resonance imaging (fMRI), Near-infrared spectroscopy (NIRS), Functional near-infrared spectroscopy (fNIRS)). Among them, brain activity is mostly used to measure the cognitive state of a user. EEG-based cognitive state assessment can be performed through long time attention-demanding tasks, such as driving a car [1], vigilance detection [2] etc.. Moreover, EEG has also been used in several applications like epileptic seizure detection [3], user verification [4], emotion

recognition [5]–[7], speech recognition [8], word-familiarity detection [9], and sentiment analysis of consumer [10]. In binary classification problems, Common spatial patterns (CSP) produce spatial filters by maximizing the variance of one class while minimizing the variance of another. Spatial filters of different EEG bands along with event-related potentials (ERPs) had been used to evaluate mental workload in n-back tests [11] and Sternberg memory task [12].

Most of the existing studies had used CSP features of EEG to evaluate the cognitive states of a user. However, the selection of the frequency band for each subject is a major drawback of the CSP algorithm. To overcome this issue, in this paper, we use Filter Bank Common Spatial Patterns (FBCSP) method [13] that used all the frequency bands of EEG for each subject. A sequential model often suffers from overfitting, local optima issues that can overwhelm by the ensemble model. The ensemble classifier consists of multiple classifiers that efficiently predict the unseen data by combining the prediction of each classifier using some voting system. Moreover, the ensemble model significantly reduces the bias and variance of the machine learning model.

In this paper, we have implemented a framework that consists of the FBCSP method and the LSTM-based deep ensemble model. To the best of our knowledge, there exists no such framework to estimate the cognitive state of a subject. In the FBCSP method, the filter bank consists of multiple filtered signals with a specific frequency band of EEG. Spatial features from each bank have been extracted using the CSP algorithm. Next, the optimum features have been selected using the Mutual Information-based Best Individual Feature (MIBIF) algorithm. The optimum spatial features of each subject are fed into each LSTM network in the ensemble model. All the LSTM networks are executed parallelly for each subject. The framework of this proposed model is shown in Fig. 1. The proposed model is divided into four components, namely: filter bank, feature extraction, feature selection, and a deep ensemble model. Each of the components has been discussed in Section III.

The rest of this paper is organized as follows: Related works are summarized in Section II; the detailed discussion of the proposed work is presented in Section III; Section IV

represents the experimental results of the proposed model; finally, in Section V we conclude the paper.

II. RELATED WORK

In [14], authors have implemented the FBCSP-based framework to identify cognitive workload from the spatial and verbal n-back task. They observed that spatial filters of the theta band were highly excited in the frontal lobe whereas, filters of the alpha band show more activation in the temporal and parietal lobe of the brain. Yang *et al.* [15] has proposed an ensemble model using stacked denoising autoencoders (SDAE) for estimating cognitive workload. They have achieved 92% average classification accuracy. A stacked autoencoder-based deep ensemble model has been implemented for recognizing the emotion of a participant [16]. Their model has achieved 83.61% accuracy. In [17], authors have evaluated different levels (low, medium, and high) of cognitive workload during the mental arithmetic task. They have used CSP features of the EEG signal and Linear Discriminant Classifier to distinguish different workload levels. Multi-channel and single-channel EEG-based workload estimation during the mental arithmetic task had been proposed in [18]. For multi-channel and single-channel-based EEG systems, their model has achieved 97.87% and 84.15% classification accuracy. Unimodal, as well as multimodal physiological measures (EEG, fMRI), have been used to estimate the cognitive state of a participant during mental arithmetic test [19] - [20]. A deep neural network and evolutionary algorithm-based hybrid model has been proposed for estimating mental workload during "simultaneous capacity (SIMKAP)"-based multitasking activity [21]. They have achieved 86.33% classification accuracy using deep BLSTM-LSTM network and Grey Wolf Optimizer. In [22], the author discussed the association between the Heart Rate Variability (HRV) measure of the ECG signal and the complexity of the mental arithmetic (easy and calculative) task. They observed that the value of HRV is inversely proportional to the complexity of tasks. A deep ensemble model has been developed for estimating subject-specific mental workload during the automatic enhanced cabin air management system (ACAMS) experiment [15]. Subject-wise local features have been extracted from EEG. Then, classification has been performed using the stacked denoising autoencoder. They reported an average classification accuracy of 92%. EEG-based cognitive state (focused, unfocused, and drowsy) assessment of an operator has been evaluated in [23]. They have achieved an average accuracy of 91.72% with the SVM classifier. In [24], authors have implemented an EEG-based spatial-temporal convolutional neural network (ESTCNN) to detect driver fatigue. They fused temporal and spatial features of EEG, and the combined features have been used for classification. They have achieved 97.37% classification accuracy with their model.

III. PROPOSED METHOD

In this section, we explain the subject-wise cognitive state estimation using the FBCSP method and the deep ensemble model. Initially, the multichannel EEG signal has been
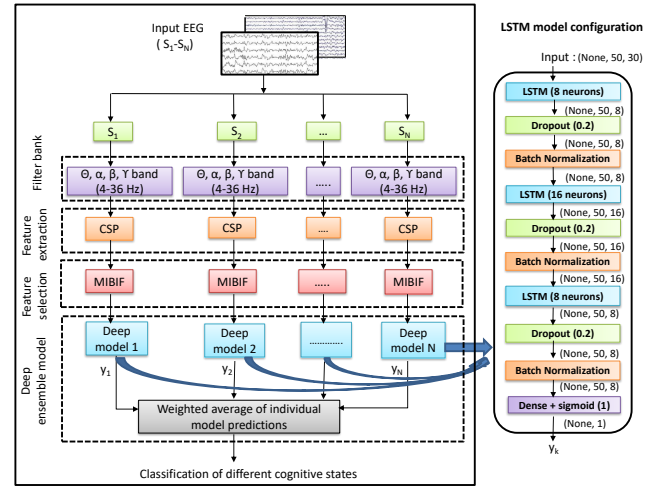

Fig. 1. Framework of the proposed LSTM-based deep ensemble model for estimating the cognitive state of a subject. The configuration of an individual deep LSTM model is shown on the right side. The three-dimensional LSTM input is represented as (samples, time steps, and features), where time steps and features are 50 and 30, respectively. The sample size is determined during the training process.

decomposed into different equal-sized frequency bands (i.e., filter bank). The band-specific spatial features have been extracted using the CSP method, and relevant features have been selected using the feature selection algorithm (MIBIF). Finally, the classification of the cognitive state for each subject has been performed using the LSTM-based deep ensemble model. The component-wise proposed framework is shown in Fig. 1.

A. Filter Bank Common Spatial Pattern (FBCSP)

Generally, the FBCSP method consists of four stages, namely, creation of filter bank, extraction of spatial features, feature selection, and classification. Here, we use the first three stages of the FBCSP method as the deep ensemble model performs the classification. In Fig. 1, component 1 to 3 of the proposed framework falls under FBCSP section. Here, the filter bank was created by decomposing the EEG signal into eight equal frequency bands, namely 4-8, 8-12, .., 32-36 Hz using bandpass filters. The filter bank consists of all four commonly used EEG bands, namely: theta (θ: 4-7 Hz), alpha (α: 8 -15 Hz), beta (β: 16-31 Hz), and gamma (γ: 32-36 Hz) respectively. CSP algorithm has been applied to extract the spatial features from each of those bands. Next, the most discriminate CSP features from each filter bank have been identified using the MIBIF feature selection method. The subject-specific optimum CSP features have been fed into the LSTM model. In the mental arithmetic test, the output classes represent the cognitive states (task and rest) of a subject during the experiment. Spatial filters (calculated from the CSP algorithm) maximize the ratio of data variance for two classes. For example, spatial filtered signal (F) of a single trial EEG (E) is represented as (1).

$$F = W_{csp} \times E \qquad (1)$$

Where E is an $C \times S$ matrix, where C and S represents the number of channels and number of samples per channel, respectively. In the CSP projection matrix (W_{csp}), rows and columns represent spatial filters and the common spatial patterns, respectively. The W_{csp} for each class is computed by maximizing the ratio between the mean covariance matrices of each class. Here, we consider a small number of spatially filtered signals of each class and variance of those signals represented as features for classification. Here, the feature vector (X_k) is computed from the first n rows of F (i.e. F_k where $k \in [1, ..n]$). The computation of X_k using log-variance method is mentioned in (2).

$$X_k = log\left(\frac{var(F_k)}{\sum_{i=1}^{n} var(F_i)}\right) \qquad (2)$$

B. Mutual Information-based Best Individual Feature (MIBIF)

In information theory, mutual information is computed between two variables to reduce the uncertainty of one variable based on the known value of another. MIBIF algorithm follows the filter approach of feature selection. In the MIBIF algorithm, initially, the mutual information of each feature with output class is being computed, the first k features ($k < d$) with maximum mutual information are being selected as optimum features. Note that k is a user-defined parameter.

C. Deep ensemble model

This subsection consists of two parts; framework of the LSTM-based deep model and ensemble technique to combine LSTM models.

1) Framework of the LSTM-based deep model: Each LSTM-based deep model within the ensemble framework consists of several stacked LSTMs, dropout, batch normalization, and dense layers. The pictorial representation of a single LSTM-based deep model is shown on the right side of Fig. 1. In the ensemble, all the LSTM model follows the same structure. The dense layers are useful for applying more hidden layers in the model. The dropout layers are mainly used to avoid overfitting, whereas the batch normalization layers are useful for normalizing the output of each layer. The first LSTM layer consists of 8 units, accompanied by a dropout and batch normalization layer. The output of the first LSTM layer passed as an input to the second LSTM layer with 16 units, accompanied by a dropout and batch normalization layer. The last LSTM layer consists of 8 units, further followed by a dropout and batch normalization layer. Finally, a dense layer with a single neuron finished the model configuration.

2) Ensemble technique to combine LSTM models: The ensemble process is divided into five steps. The dataset of each subject has been divided into train, validation, and test set in the ratio of 7:2:1. The optimum CSP features (computed by the MIBIF algorithm) are represented as the inputs to the proposed ensemble model. Each LSTM model has been trained based on the optimum feature set. Next, the AUC score of each model has been calculated on the validation set. Ensemble weight $\left(w_n^{(t)}\right)$ at t^{th} time step for each model has been computed using average AUC score as shown in (3). The prediction of N

number of LSTM models at t^{th} timestep are $y_1^{(t)}, y_2^{(t)}, ...y_N^{(t)}$ respectively. Finally, the output of the ensemble model (4) at $(t+1)^{th}$ time step is computed by the combined results of prediction (y_n) and weight (w_n) of each LSTM model at t^{th} time step. The entire procedure of the proposed ensemble model is shown in Fig. 2.

$$w_n^{(t)} = \frac{AUC(model^{(t)})}{\sum_{n=1}^{N} AUC(model_n^{(t)})} \qquad (3)$$

$$y_{ensem}^{(t+1)} = \sum_{n=1}^{N} y_n^{(t)}.w_n^{(t)} \qquad (4)$$

The weight parameter follows two properties: $0 \leq w_n \leq 1$, and $\sum_{n=1}^{N} w_n = 1$. AUC score of each model at t^{th} time step is represented by $AUC(model^{(t)})$. Here, N represents the number of parallel LSTMs.
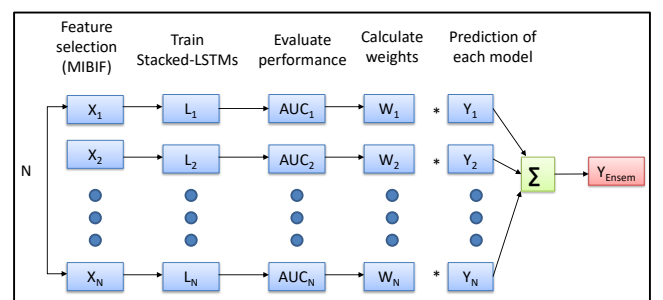

Fig. 2. Procedure of the proposed ensemble model.

IV. RESULTS

This section consists of four subsections, namely; (A) dataset analysis, (B) analysis of the FBCSP method, (C) analysis of the deep ensemble model, and (D) comparative analysis.

A. Dataset analysis

We have used the open-access dataset [25] for our experiment. The study contains three datasets: (A) motor imagery (left vs. right hand), (B) mental arithmetic vs. baseline test, and (C) motion artifacts of 30 subjects. In our experiment, we have used mental arithmetic data (dataset B) captured from 30 channel EEG. The sampling frequency of EEG data was set at 200Hz. Here, the task state refers to the cognitive state of a user while performing the mental arithmetic test, whereas the rest state represents the baseline test. The artifacts removal from raw EEG has been performed using Independent component analysis (ICA). Here, we have performed ICA-based electrooculogram (EOG) rejection using Automatic Artifact Removal (AAR) toolbox in EEGLAB [26]. ICA can be successfully utilized for finding consistent spatial components from brain signal [27]. The artifacts-free spatial components are useful for finding the proper CSP patterns in the FBCSP method, which leads to better classification performance in the later stage.

4109

B. Analysis of the FBCSP method

In this paper, we have applied three parts of the general FBCSP method. The first part consists of filter bank creation based on a specific frequency range (based on bandpass filters). The second and third part contains the spatial feature extraction from each filter bank using the CSP method and relevant feature selection using the MIBIF algorithm, respectively. In the case of binary classification, the CSP filters minimize the variance of one class. Besides this, it maximizes the variance of another class. Here, the variance computed from each bandpass filtered signal denotes the band-power of the signal. CSP filters are applied to bandpass filtered signals (frequency range: 4-36 Hz) to discriminate two cognitive states (task and rest). For a subject, discrimination of cognitive states using band power is shown in Fig. 3. It can be noted that the band power for the task is higher than the rest. In Fig. 3, the band power of CSP filters has been computed for frequency range (4-36 Hz), but due to negligible power after frequency 35Hz, we have omitted that part from the graph.

The mental arithmetic experiment is divided into two parts according to different time intervals: mental arithmetic (MA)/ task (0 to 10 seconds) followed by baseline task/rest (11 to 20 seconds). CSP filters are being applied over the mentioned time intervals. The cognitive state-wise topographical map of a subject for all the four frequency bands is shown in Fig. 4, and Fig. 5, respectively. Here, CSP filters are being represented by the topographical map for a specific time range. It can be noted that the spatial filter obtained from the alpha band (8-15 Hz) clearly distinguishes two cognitive states of a subject. In Fig. 4(c) all the CSP filter gives more weights to the frontal electrodes during the cognitive state task, while less weights are being assigned to frontal lobes in the rest state (Fig. 4(d)). For the beta band (Fig. 5(a), and Fig. 5(b)), excitation of frontal electrodes of 3rd and 5th CSP filters can distinguish two cognitive states. CSP filters of theta and gamma band have no major significance in distinguishing the cognitive state.

After extracting the spatial features by the CSP algorithm, we have proceeded to find the relevant features from the input feature set. The k best features having maximum mutual information have been selected using the MIBIF algorithm. Classification accuracy after varying the size of the feature subset (k) has been shown in Table I. The input feature set
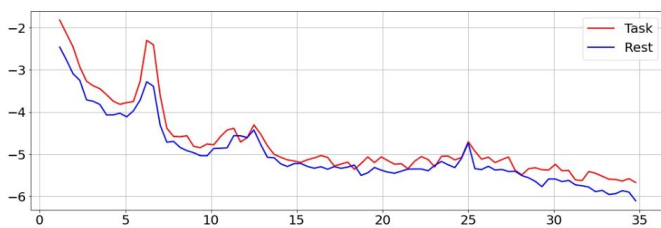

Fig. 3. Spatial pattern obtained from CSP filter of a subject. Band power clearly distinguishes the two cognitive states (task, rest) of the subject. Here, X-axis and Y-axis represents the frequency (Hz) and band power (dB), respectively.

contains 30 features, from which we select the top k features ($k < 30$) for classification. It can be noted that the maximum accuracy has been obtained with a feature subset having size 15.

TABLE I
CLASSIFICATION ACCURACY BASED ON SIZE OF FEATURE SUBSET

Size of subset	Classification accuracy (%)
5	70.25
10	85.56
15	87
25	83.78

C. Analysis of the ensemble model

This subsection consists of five parts: (1) analysis based on different structures of LSTM network, (2) output and performance analysis, (3) runtime performance & computational complexity, (4) scalability test of the proposed model, and (5) performance analysis based on other datasets.

1) Analysis based on different structures of LSTM network: Here, we discuss the structure of each LSTM network in terms of LSTM layers and the number of units in each LSTM layer. Number of units at i^{th} LSTM layer represents as h_i. At layer 1, h_1 started with 2 and increased up to 24. The maximum classification accuracy has been obtained with h_1 equal to 8, so we set the starting value of h_2 as 8. For two layers, h_2 varies from 8 to 20, and maximum accuracy has been obtained with h_2 equal to 16. In the second layer, model performance hardly improves after h_2 equal to 16. Finally, in the third layer, we set h_3 between 8 to 16. Similar to the second layer, the third layer's performance cannot be further enhanced after extending more units. The number of units in the first, second, and third LSTM layer ($h_1 = 8$, $h_2 = 16$, and $h_3 = 8$) with maximum accuracy have been employed in each LSTM network in the ensemble model. The classification accuracy of the proposed ensemble model with different structured LSTMs is plotted in Fig. 6.

2) Output and performance analysis: This section discusses the output and performance analysis of the proposed ensemble model. For binary classification, binary cross-entropy loss function has been used. We train the model with the batch size of 2000. Adam optimizer, with the initial learning rate of 1e-03, has been used. Due to low computational resources, we performed the experiment on fourteen subjects. For evaluating the subject-wise performance of the model, we perform the leave-subject-out experiment, where the training/testing set contains an entirely different subject. We trained the model with thirteen subjects and predicted the output on the remaining subject (i.e., "Subject in Test set" in Fig. 8). The result of the leave-subject-out experiment is shown in Fig. 8. It can be observed that the maximum model performance has been achieved for subject 10 (S10). The learning curve of a model indicates the training and validation score for varying numbers of training samples. The learning curve of an individual LSTM-based deep model is shown in Fig. 7(a). Here, both the validation and the training score of the model converge
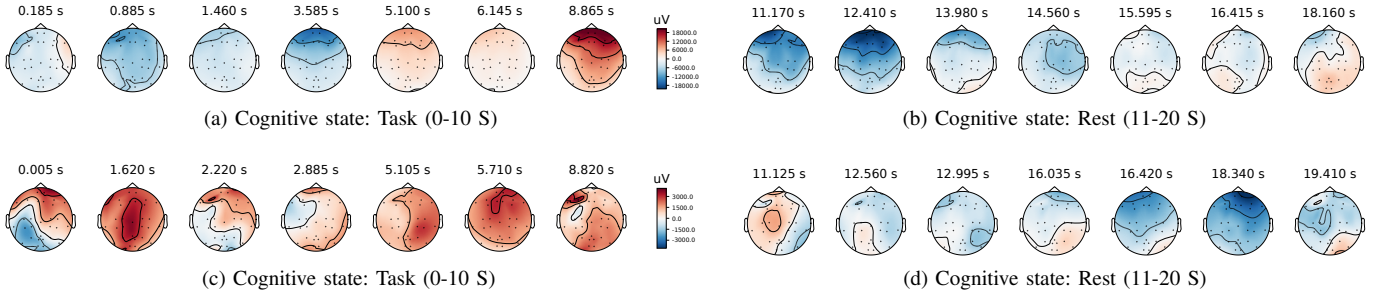
4110

Fig. 4. CSP filters of different frequency bands of a subject during the mental arithmetic (MA) experiment. (a-b): theta band (4-7 Hz), (c-d): alpha band (8-15 Hz). Cognitive state-wise CSP filters have been obtained during the specified time interval (task: 0-10 seconds and rest: 11-20 seconds) of the experiment.
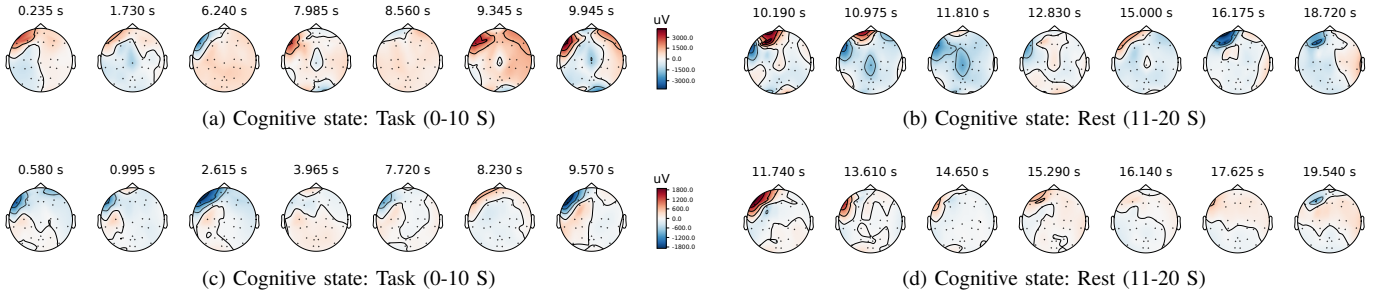

Fig. 5. CSP filters of different frequency bands of a subject during the mental arithmetic (MA) experiment. (a-b): beta band (16-31 Hz), and (c-d): gamma band (32-36 Hz). Cognitive state-wise CSP filters have been obtained during the specified time interval (task: 0-10 seconds and rest: 11-20 seconds) of the experiment.
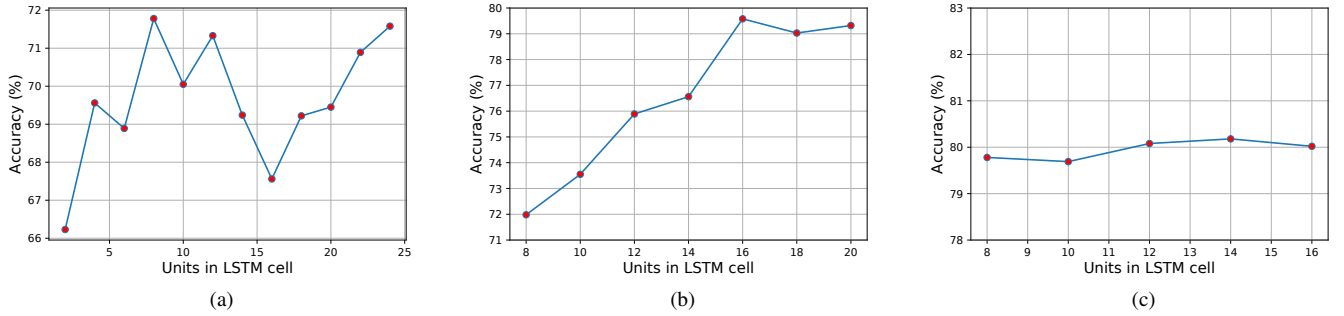

Fig. 6. Classification accuracy of the proposed ensemble model using various configurations of the LSTM model: (a) single layered LSTM, (b) two-layered LSTM, and (c) three-layered LSTM.

at high value with an increasing number of training samples, which will increase the generalization of the model. As all the LSTM model follows the same structure, and data distribution is identical for all the subjects, we conclude that the proposed ensemble model becomes highly generalized to predict the output classes. The Receiver Operating Characteristic (ROC) curve of the proposed model is plotted in Fig. 7(b). The AUC score of the model achieves 0.91, which leads to a significant result in classification analysis. The classification result of the proposed model is discussed in Table II. This table shows the class-wise (task and rest) prediction results of the proposed model. The result is presented based on different classification parameters, such as precision, recall, and F1 score.

TABLE II
CLASSIFICATION ANALYSIS OF THE PROPOSED MODEL

Class	Precision (%)	Recall (%)	F1 Score (%)
Task	86.38	82.13	84.20
Rest	85.73	81.32	83.47

3) Runtime performance & Computational complexity: Here, we first analyzed the runtime performance of the ensemble LSTM network; then, we estimate the overall model (consisting of the FBCSP method, MIBIF method, and ensemble LSTM network) complexity. We perform the analysis for a maximum of fourteen subjects due to the limitation of a low memory environment. Our hardware configuration comprises
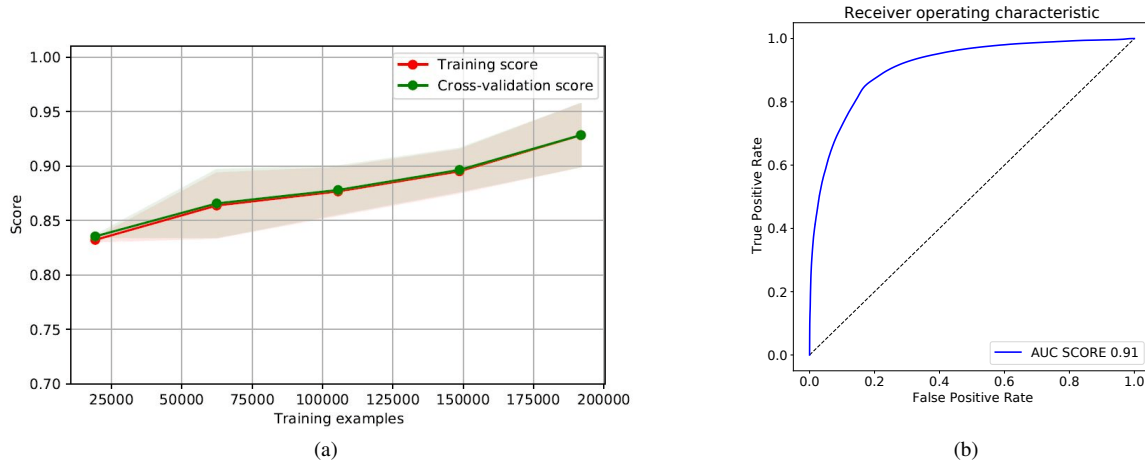
4111

Fig. 7. Performance analysis of the proposed model: (a) Learning curve of a single LSTM model, (b) ROC curve of the proposed ensemble model.
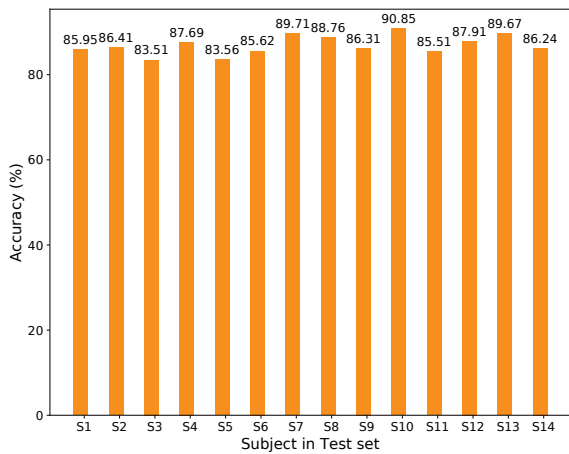

Fig. 8. Subject-wise performance analysis: Result of Leave-subject-out experiment.

a 2-GHz CPU and a 24G RAM. For comparison, we have built a sequential model with a similar LSTM configuration as per the ensemble model. The model performance has been evaluated in terms of execution time. For a subject, the total time required to train a sequential LSTM model is about 28.20 minutes; for fourteen subjects, it takes nearly 395 minutes. On the other side, the training time required for the proposed ensemble model (consists of fourteen parallel models) is approximately 349 minutes. In the ensemble model, the training time required for each subject is 24.93 minutes. It can be noted that the sequential model takes more time compared to the proposed model for the same runtime environment. Hence, we can conclude that the proposed ensemble model outperforms a sequential model in a low memory environment.

The computational complexity of CSP algorithm is $O(m^3 + Nm^2)$, where m and N represents number of features and number of samples respectively [28]. So, the computational complexity for extracting the spatial features from P fil-

ter banks is $O(P(m^3 + Nm^2))$ or simply $O(m^3 + Nm^2)$ (as $P << m$, and $P << N$). In the MIBIF method, first, the inter-feature mutual information is computed in $O(N\sqrt{KN})$ [29], where N is the total number of data points and K ($K < N$) is a constant. Next, the feature set has been sorted ($O(N \log N)$), and the top f features have been selected based on the individual feature's rank ($O(1)$). Thus, the total time taken for sorting and selecting f features is $O(N \log N)$. Hence, total complexity required for MIBIF algorithm is $O(N \log N + N\sqrt{KN})$. It can be simply written as $O(N\sqrt{KN})$, ($N \log N < N\sqrt{KN}$, when $N \to \infty$). The computational complexity of a single LSTM network for t time step and W weights is $O(tW)$ [30]. Here, each weight associated with a singe node. Thus, the overall complexity of the proposed ensemble model is $O(S(tW + (m^3 + Nm^2) + (N\sqrt{KN}))))$, where S is the number of subjects.

4) Scalability test of the proposed ensemble model: In this subsection, we discuss the scalability test of the proposed model by varying the number of subjects. For the low computing resources, we only perform the classification analysis up to fourteen subjects. The scalability test of the proposed model is shown in Fig. 9. It can be noted that maximum accuracy has been obtained for ten subjects; the performance of the proposed model cannot be further enhanced after extending more number of subjects.

5) Performance analysis based on other datasets: To evaluate the robustness, we have applied the proposed method on other mental arithmetic datasets. The recording protocol of each dataset is different (EEG, NIRS). The performance analysis of each dataset is shown in Table III. A notable performance of the proposed method has been observed irrespective of dataset and recording protocol.

D. Comparative analysis

This section comprises two subsections: (a) comparative analysis with other neural network models and (b) comparative analysis with existing studies.
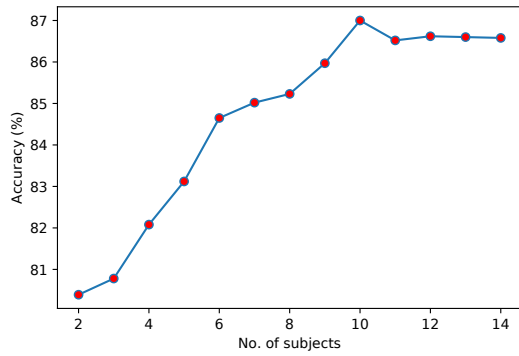
Fig. 9. Scalability test of the proposed ensemble model.

TABLE III
PERFORMANCE ANALYSIS BASED ON OTHER DATASETS. NOTE: MA:
MENTAL ARITHMETIC

Dataset - Recording protocol	Study	Accuracy (%)
MA vs. Baseline task [25] - NIRS	Shin *et al.* [25]	80.70
	Shin *et al.* [31]	80.00
	Ergun *et al.* [32]	84.94
	Proposed method	**89.52**
MA vs. Resting task [33] - NIRS	Bauernfeind *et al.* [33]	79.70
	Proposed method	**89.76**
MA vs. Resting task [34] - EEG	Ahammed *et al.* [35]	87.50
	Rani *et al.* [36]	77.58
	Proposed method	**91.11**

1) Comparative analysis with other neural network models:
Here, the comparative analysis between the proposed model with other neural network methods (GRU, BLSTM-GRU, stacked LSTM, BLSTM-LSTM) has been discussed. The comparative analysis is presented in Table IV. It can be noticed that the proposed model performs better than other neural network models with higher classification accuracy and lesser computational resources (in terms of the number of nodes used in the LSTM network).

TABLE IV
COMPARATIVE ANALYSIS WITH OTHER DEEP NEURAL NETWORK MODELS.
NOTE: L: LSTM, D: DENSE, B: BLSTM, G: GRU, TD: TIME
DISTRIBUTED LAYER

Model	Precision (%)	Recall (%)	Accuracy (%)	Layers/nodes
GRU	75.78	75.64	75.65	G30-G20-G10-D100-D1
BLSTM-GRU	81.60	81.56	81.54	B30-G10-G10-TD1-D1
Stacked LSTM	81.83	80.66	80.98	L30-L20-L10-TD1-D1
BLSTM-LSTM	82.61	82.52	82.53	B30-B10-L10-D100-D1
Proposed	**86.55**	**81.72**	**87.00**	**L8-L16-L8-D1**

2) Comparative analysis with existing studies: Here, we have performed a comparative analysis between the proposed ensemble model and other existing studies for the same experiment. We have implemented the existing studies and examined them on our dataset. The comparative analysis is shown in Table V. All the existing methods have been examined in the same hardware environment (2-GHz CPU and a 24G RAM). We keep the same configuration of parameters for classifiers (shrinkage LDA, SVM, kNN, etc.) as per existing studies, and compute the runtime of those methods. From the Table V, we can conclude that the proposed model effectively distinguishes the mental state of a subject with high classification accuracy and reasonable runtime.

TABLE V
COMPARATIVE ANALYSIS WITH EXISTING STUDIES

Study	Method	Accuracy (%)	Runtime (min.)
Shin *et al.* [25]	CSP + Shrinkage LDA	83.60	284.17
Sammer *et al.* [20]	PCA + ANOVA	71.40	142.78
Rebsamen *et al.* [19]	Sparse Multinomial Logistic Regression	65.02	120.05
Dimitriadis *et al.* [37]	Functional connectivity graph (parieto occipital lobes) + kNN	84.12	245.91
Al-shargie *et al.* [38]	Two sample t-test + SVM	85.00	356.89
Subhani *et al.* [39]	SVM + t-test	83.43	368.23
Proposed method	**Ensemble deep LSTM networks**	**87.00**	**349.00**

V. CONCLUSION

This paper combines two methods, namely, the FBCSP method and the deep ensemble model for identifying the cognitive state of a subject during a mental arithmetic test. The proposed deep ensemble model can efficiently identify the cognitive state of a subject with 87% classification accuracy. The model can be effectively utilized for the execution of a deep model over a large volume of data in the low memory environment. The proposed ensemble model takes less computational time compared to an equivalent sequential model and other state-of-the-art methods. The results of Table IV proves the superiority of the proposed ensemble model over other deep neural network models. The robustness of the proposed model has been evaluated in Table III. In comparative analysis (Table V), we also noticed that the proposed ensemble model outperforms other existing studies related to the mental arithmetic experiment.

Due to low-computational resources, performance evaluation of the proposed model has been performed for less number of subjects. Moreover, the source dataset [25] contains multimodal (EEG and NIRS) physiological signals of thirty subjects. In contrast, we have conducted our experiment only on EEG signals. Therefore, our experimental analysis is limited to the EEG-based unimodal approach rather than the multimodal (EEG and NIRS) one. For the same reason, we could not perform the unimodal vs. multimodal comparison of the same experiment.

The effectiveness of the proposed model can be further improved by increasing the computational resources. In the near future, we will revise the proposed model by adding

multimodal functionality (EEG and NIRS) for more number of subjects. We will further compare the performance of the proposed model based on a unimodal vs. multimodal scenario. The comparison will be helpful in identifying the best modality that can effectively estimate the cognitive workload of a person.

REFERENCES

[1] M. Hajinoroozi, Z. Mao, T.-P. Jung, C.-T. Lin, and Y. Huang, "EEG-based prediction of driver's cognitive performance by deep convolutional neural network," *Signal Processing: Image Communication*, vol. 47, pp. 549–555, 2016.

[2] S. Charbonnier, R. N. Roy, S. Bonnet, and A. Campagne, "EEG index for control operators' mental fatigue monitoring using interactions between brain regions," *Expert Systems with Applications*, vol. 52, pp. 91–98, 2016.

[3] K. Das, D. Daschaklader, P. P. Roy, A. Chatterjee, and S. P. Saha, "Epileptic seizure prediction by the detection of seizure waveform from the pre-ictal phase of EEG signal," *Biomedical Signal Processing and Control*, vol. 57, p. 101720, 2020.

[4] R. Saini, B. Kaur, P. Singh, P. Kumar, P. P. Roy, B. Raman, and D. Singh, "Don't just sign use brain too: A novel multimodal approach for user identification and verification," *Information Sciences*, vol. 430, pp. 163–178, 2018.

[5] D. Panda, D. D. Chaklader, and T. Dasgupta, "Multimodal system for emotion recognition using EEG and customer review," in *Proceedings of the Global AI Congress 2019*. Springer, 2020, pp. 399–410.

[6] B. Kaur, D. Singh, and P. P. Roy, "EEG based emotion classification mechanism in BCI," *Procedia computer science*, vol. 132, pp. 752–758, 2018.

[7] D. D. Chaklader and S. Chakraborty, "EEG based emotion classification using "correlation based subset selection"," *Biologically inspired cognitive architectures*, vol. 24, pp. 98–106, 2018.

[8] P. Kumar, R. Saini, P. P. Roy, P. K. Sahu, and D. P. Dogra, "Envisioned speech recognition using EEG sensors," *Personal and Ubiquitous Computing*, vol. 22, no. 1, pp. 185–199, 2018.

[9] V. Khurana, P. Kumar, R. Saini, and P. P. Roy, "EEG based word familiarity using features and frequency bands combination," *Cognitive Systems Research*, vol. 49, pp. 33–48, 2018.

[10] S. Kumar, M. Yadava, and P. P. Roy, "Fusion of EEG response and sentiment analysis of products review to predict customer satisfaction," *Information Fusion*, vol. 52, pp. 41–52, 2019.

[11] A.-M. Brouwer, M. A. Hogervorst, J. B. Van Erp, T. Heffelaar, P. H. Zimmerman, and R. Oostenveld, "Estimating workload using EEG spectral power and ERPs in the n-back task," *Journal of neural engineering*, vol. 9, no. 4, p. 045008, 2012.

[12] R. N. Roy, S. Charbonnier, A. Campagne, and S. Bonnet, "Efficient mental workload estimation using task-independent EEG features," *Journal of neural engineering*, vol. 13, no. 2, p. 026019, 2016.

[13] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 2008, pp. 2390–2397.

[14] M. Arvaneh, A. Umilta, and I. H. Robertson, "Filter bank common spatial patterns in mental workload estimation," in *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 4749–4752.

[15] S. Yang, Z. Yin, Y. Wang, W. Zhang, Y. Wang, and J. Zhang, "Assessing cognitive mental workload via EEG signals and an ensemble deep learning classifier based on denoising autoencoders," *Computers in biology and medicine*, vol. 109, pp. 159–170, 2019.

[16] Z. Yin, M. Zhao, Y. Wang, J. Yang, and J. Zhang, "Recognition of emotions using multimodal physiological signals and an ensemble deep learning model," *Computer methods and programs in biomedicine*, vol. 140, pp. 93–110, 2017.

[17] Z. Y. Chin, X. Zhang, C. Wang, and K. K. Ang, "EEG-based discrimination of different cognitive workload levels from mental arithmetic," in *40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 1984–1987.

[18] 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, 2013.

[19] B. Rebsamen, K. Kwok, and T. B. Penney, "EEG-based measure of cognitive workload during a mental arithmetic task," in *International Conference on Human-Computer Interaction*, 2011, pp. 304–307.

[20] G. Sammer, C. Blecker, H. Gebhardt, M. Bischoff, R. Stark, K. Morgen, and D. Vaitl, "Relationship between regional hemodynamic activity and simultaneously recorded EEG-theta associated with mental arithmetic-induced workload," *Human brain mapping*, vol. 28, pp. 793–803, 2007.

[21] D. D. Chaklader, S. Dey, P. P. Roy, and D. P. Dogra, "EEG-based mental workload estimation using deep BLSTM-LSTM network and evolutionary algorithm," *Biomedical Signal Processing and Control*, vol. 60, p. 101989, 2020.

[22] N. Singh, Y. Aggarwal, and R. K. Sinha, "Heart rate variability analysis under varied task difficulties in mental arithmetic performance," *Health and Technology*, vol. 9, no. 3, pp. 343–353, 2019.

[23] Ç. İ. Acı, M. Kaya, and Y. Mishchenko, "Distinguishing mental attention states of humans via an EEG-based passive BCI using machine learning methods," *Expert Systems with Appl.*, vol. 134, pp. 153–166, 2019.

[24] Z. Gao, X. Wang, Y. Yang, C. Mu, Q. Cai, W. Dang, and S. Zuo, "EEG-based spatio–temporal convolutional neural network for driver fatigue evaluation," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 9, pp. 2755–2763, 2019.

[25] J. Shin, A. von Lühmann, B. Blankertz, D.-W. Kim, J. Jeong, H.-J. Hwang, and K.-R. Müller, "Open access dataset for EEG + NIRS single-trial classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 10, pp. 1735–1745, 2016.

[26] G. Gómez-Herrero, W. De Clercq, H. Anwar, O. Kara, K. Egiazarian, S. Van Huffel, and W. Van Paesschen, "Automatic removal of ocular artifacts in the EEG without an EOG reference channel," in *Proceedings of the 7th Nordic Signal Processing Symposium-NORSIG*, 2006, pp. 130–133.

[27] L. Griffanti, G. Salimi-Khorshidi, C. F. Beckmann, E. J. Auerbach, G. Douaud, C. E. Sexton, E. Zsoldos, K. P. Ebmeier, N. Filippini, C. E. Mackay *et al.*, "ICA-based artifact removal and accelerated fMRI acquisition for improved resting state network imaging," *Neuroimage*, vol. 95, pp. 232–247, 2014.

[28] L.-C. Shi, Y. Li, R.-H. Sun, and B.-L. Lu, "A sparse common spatial pattern algorithm for brain-computer interface," in *International Conference on Neural Information Processing*, 2011, pp. 725–733.

[29] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Physical review E*, vol. 69, no. 6, p. 066138, 2004.

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

[31] J. Shin and C.-H. Im, "Performance prediction for a near-infrared spectroscopy-brain–computer interface using resting-state functional connectivity of the prefrontal cortex," *International journal of neural systems*, vol. 28, no. 10, p. 1850023, 2018.

[32] E. Ergün and Ö. Aydemir, "Decoding of binary mental arithmetic based near-infrared spectroscopy signals," in *3rd International Conference on Computer Science and Engineering (UBMK)*, 2018, pp. 201–204.

[33] G. Bauernfeind, R. Scherer, G. Pfurtscheller, and C. Neuper, "Single-trial classification of antagonistic oxyhemoglobin responses during mental arithmetic," *Medical & biological engineering & computing*, vol. 49, no. 9, p. 979, 2011.

[34] 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, p. 14, 2019.

[35] K. Ahammed and M. U. Ahmed, "Quantification of mental stress using complexity analysis of EEG signals," *Biomedical Engineering: Applications, Basis and Communications*, vol. 32, p. 2050011, 2020.

[36] M. Rani, S. B. Dhok, R. B. Deshmukh, and P. Kumar, "Overlap aware compressed signal classification," *IEEE Access*, vol. 8, pp. 52 950–52 967, 2020.

[37] S. I. Dimitriadis, Y. Sun, K. Kwok, N. A. Laskaris, and A. Bezerianos, "A tensorial approach to access cognitive workload related to mental arithmetic from EEG functional connectivity estimates," in *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013, pp. 2940–2943.

[38] F. Al-Shargie, T. B. Tang, N. Badruddin, and M. Kiguchi, "Mental stress quantification using EEG signals," in *International Conference for Innovation in Biomedical Engineering and Life Sciences*, 2015, pp. 15–19.

[39] A. R. Subhani, W. Mumtaz, M. N. B. M. Saad, N. Kamel, and A. S. Malik, "Machine learning framework for the detection of mental stress at multiple levels," *IEEE Access*, vol. 5, pp. 13 545–13 556, 2017.