Latent Space Coding Capsule Network for Mental Workload Classification

Yinhu Yu, Anastasios Bezerianos, *Senior Member, IEEE*, Andrzej Cichocki, *Life Fellow, IEEE*, and Junhua Li, *Senior Member, IEEE*

Abstract—**Mental workload can be monitored in real time, which helps us improve work efficiency by maintaining an appropriate workload level. Based on previous studies, we have known that features, such as band power and brain connectivity, can be utilized to classify the levels of mental workload. As band power and brain connectivity represent different but complementary information related to mental workload, it is helpful to integrate them together for workload classification. Although deep learning models have been utilized for workload classification based on EEG, the classification performance is not satisfactory. This is because the current models cannot well tackle variances in the features extracted from non-stationary EEG. In order to address this problem, we, in this study, proposed a novel deep learning model, called latent space coding capsule network (LSCCN). The features of band power and brain connectivity were fused and then modelled in a latent space. The subsequent convolutional and capsule modules were used for workload classification. The proposed LSCCN was compared to the state-of-the-art methods. The results demonstrated that the proposed LSCCN was superior to the compared methods. LSCCN achieved a higher testing accuracy with a relatively smaller standard deviation, indicating a more reliable classification across participants. In addition, we explored the distribution of the features and found that top discriminative features were localized in the frontal, parietal, and occipital regions. This study not only provides a novel deep learning model but also informs further studies in workload classification and promotes practical usage of workload monitoring.**

Index Terms—**Mental workload classification, latent space coding, capsule networks, EEG, brain connectivity, band power.**

I. Introduction

WITH the advancement of information technology and the wide application of automatic machines, workers have ever changed their working manner with a reduction of physical load but an increased mental workload. Mental workload (MW) refers to the psychological and physiological burden generated by the operator when performing specific cognitive tasks, which is generally related to situational awareness, emotion, and alertness [1]. Task execution is closely related to the psychological state of the operator. However, our brain capacity seems not to be enhanced in line with the increasing requirement of mental resources in jobs. In such a situation, the required mental resources may exceed people's capacity, resulting in mental fatigue, low performance, and dull cognition [2]. In order to reduce the incidence of the above problems and maintain the health of workers, an appropriate workload should be assigned to avoid overloading [3].

Mental workload was assessed mainly based on behaviors and subjective feedback in early research studies [4]. Usually, a questionnaire, such as the National Aeronautics and Space Administration's Task Load Index (NASA-TLX), or task accuracy and reaction time were utilized to assess workload level. This assessment can only be carried out after the completion of a task and cannot provide a real-time indicator of workload. With the advancement of physiological signal measurement and processing technology, more researchers have attempted to utilize physiological signals to obtain an objective and real-time assessment of mental workload [5]. These physiological signals include electroencephalogram (EEG), heart rate variability (HRV), electrooculogram (EOG), electrocardiogram (ECG), and functional near-infrared spectroscopy (fNIRS) [6], [7], [8], [9]. In this study, we used EEG for the assessment because EEG exhibits suitability for reflecting brain activities associated with the mental workload. Moreover, EEG has highly temporal resolution and can be recorded in a non-invasive way and at a low cost [10].

A few categories of features extracted from EEG can be used for mental workload classification. One of the frequently used categories is band power. For instance, the theta power in the frontal midline area increased along with the elevation of mental workload [11]. Brouwer et al. found that the alpha power in the parietal area was decreased with the increase

of mental workload [12]. Besides, delta, beta, and gamma bands were also associated with mental workload [13], [14], [15]. Considering the interactions between brain regions, functional connectivity can be combined with band powers to enhance the performance of mental workload classification. This has been evidenced in our previous study [16]. The study demonstrated that the features of band power and functional connectivity were complementary and the workload classification was improved by combining them [16]. According to the comparison results, the connection features in the gamma band were relatively better than that in other bands in terms of workload classification. The band powers reflect the information existing in individual channels (corresponding to each brain region) while brain connectivity reflects interaction information between brain regions. As shown in the study [17], [18], the complex interactions among brain regions associated with mental workload states can be robustly displayed through functional connectivity. Based on these previous findings, we, therefore, determined to fuse the features of functional connectivity from the gamma band and the features of band power from all typical bands in this study.

Traditional machine learning classifiers, such as k-Nearest Neighbors (k-NN) [18], Support Vector Machine (SVM) [12], [17], random forest (RF) [16], and Linear Discriminant Analysis (LDA) [3], [19], have been used for mental workload classification. Blanco et al. trained different classifiers (i.e., k-Nearest Neighbors, Discriminant Analysis, Naïve Bayes, Decision Trees, and SVM) with different parameter settings to predict different levels of workload in simulated flight missions, among which LDA achieved the best results [19]. These traditional machine learning classifiers can be combined to improve classification performance. For example, Gu et al. combined Extreme Learning Machine (ELM) and Support Vector Machine (SVM) to develop the ELM-SVM model, which achieved higher accuracy than separate classifiers (i.e., the single SVM and the single ELM) in the mental workload classification [20]. Besides, deep learning models were also used for workload classification. Deep learning is more powerful in feature extraction than traditional methods and exhibits advantages in mental workload classification. Convolutional neural network (CNN) is one of the representative algorithms of deep learning for classification based on EEG [21], [22], [23]. Lee et al. proposed a convolutional neural network (CNN) based on multiple feature blocks that enhanced classification accuracy to 0.75 for the pilots' mental states' analysis [24]. Zhang et al. proposed two-stream neural networks (TSNN) that combined CNN and a temporal convolutional network (TCN) for three-class mental workload classification and also proved the effectiveness of CNN in extracting information [25]. Long Short-Term Memory (LSTM) can also be used to retain long-term context for workload classification [26], [27]. Moreover, Bashivan et al. introduced CNN into LSTM to develop deep recurrent convolutional neural networks (RCNNs) for preserving EEG's spatial, spectral, and temporal structure in classifying mental workload [28]. The above methods are to add classifiers after feature extraction for classification, while Li et al. proposed

supervised Canonical Polyadic Decomposition, which directly adds auxiliary label information in the decomposition process so that the classification can be implemented without an additional classifier [29].

Features extracted from non-stationary EEG are not robust enough in terms of classification, especially for mental workload classification. In this study, we attempted to mitigate this problem by embedding a variational auto-encoder (VAE) into the model to create latent variables [30], [31], [32]. The VAE establishes the probability distribution of features in the latent space to reduce the impact of the handcrafted EEG features on the classification robustness. Considering that deep learning approaches such as CNN are not good at capturing structural information, the performance of workload classification could be improved by the capsule network (CapsNet) [33], and it could achieve good performance without massive training data [34]. The capsule network was also applied in schizophrenia identification and P300 detection successfully and proved to be an effective method for neurophysiological signal classification [35], [36].

Taking all the above into consideration, we proposed a latent space coding capsule network (LSCCN) for workload classification. In this model, the features of band power and brain connectivity were modelled in the VAE module to form latent variables. The relationships between the latent variables were captured via subsequent convolutional and capsule layers. The proposed model was compared to the state-of-the-art methods, and detailed comparison results were given in this paper. In addition, we explored the features to show their distributions for deepening the understanding of mental workload.

II. METHOD

A. Participants and Experiment

A prescreen was administered to exclude those who had attended any EEG-related experiments or had the experience of the use of the aircraft simulation. Seven healthy participants were included in this study. The institutional review board of the National University of Singapore approved the experimental protocol. All participants gave their written consent forms for participation in the experiment. In the experiment, participants needed to wear the Oculus Rift virtual reality headset, which was used to display virtual aircraft, and to manipulate a simulated aircraft via a joystick. The participants experienced three levels of manipulation difficulty, which correspond to three workload levels: low, medium, and high workload. Under the low workload, participants did not need to do any tasks and just looked at an autonomous simulated aircraft. Under the medium workload, participants needed to manipulate the simulated aircraft manually. Under the high workload, malfunctions, such as engine failure, occurred, which led to unstable aircraft. Participants had to pay more effort to manipulate the unstable aircraft. Accordingly, the required workload was more than that under the medium workload condition. Each workload level lasted 2 minutes, resulting in a total of 6 minutes for a session. Each participant performed three sessions.

B. Feature Extraction and Fusions

EEG signals were recorded by 62 electrodes at a sampling rate of 256 Hz. A standard EEG preprocessing was adopted to remove artifacts, including bandpass filter (0.5~48 Hz) and independent component analysis (ICA). Then the preprocessed EEG signals were divided into two-second segments, resulting in 540 segments (62 channels × 512 data points) for each participant. The data of each channel were transformed into power spectral density (PSD) using short-time Fourier transform (STFT) with a one-second sliding time window (i.e., 256 data points). Different overlapping settings were used and compared in terms of classification performance. After STFT, we obtained the band power features by summing up the powers of each frequency within respective frequency bands (i.e., Delta: 1~4 Hz, Theta: 4~8 Hz, Alpha: 8~12 Hz, Beta: 12~30 Hz, and Gamma: 30~45 Hz), resulting in a feature matrix (number of time windows × 5 bands) for each channel.

The connectivity strengths between channels were quantified by either phase locking value (PLV) [37], [38], [39] or phase lag index (PLI) [40]. Specifically, the preprocessed EEG signals were firstly filtered into the gamma band (rather than all five bands) for extracting brain connectivity features, because the gamma band has an advantage in the connectivity feature extraction for the workload classification according to our previous study [16]. The PLV of EEG signals of channel k and channel l over time span $t = t_1, t_2, \ldots, t_n$ was calculated by:

$$PLV_{k,l} = \langle e^{j(\phi_k(t) - \phi_l(t))} \rangle, \quad (1)$$

where $\langle \cdot \rangle$ represents the average over the time span, ϕ_k and ϕ_l are the signal phases of channel k and l. The range of PLV is [1, 0], of which 0 indicates no phase synchronization and 1 indicates perfect phase synchronization [37], [38], [39].

PLI is also a phase-based method and was used to measure consistent and nonzero phase lag between channels. PLI is calculated as:

$$PLI_{k,l} = |\langle sign[\Delta\varphi_{k,l}(t)] \rangle|, \quad (2)$$

where $\Delta\varphi_{k,l}(t)$ represents the phase difference between channel k and channel l at time t, and $sign$ is a signum function, and $|\cdot|$ denotes an absolute value function. The PLI value range is also from 0 to 1. A value of 0 indicates either no synchronization or phase synchronization difference centered around 0 and π. In contrast, a value of 1 indicates perfect phase synchronization with consistent phase differences other than 0 and π [40]. After calculating PLV or PLI values for any two channels, all these values were formed a functional connectivity matrix. The size of this matrix is 62 × 62, which is for a segment. Subsequently, the band power features were merged into functional connection features. Moreover, the merged features were normalized into [1, 0] across channels. For functional connection features, the entries on the main diagonal (i.e., self-connections) were excluded in the feature normalization.

C. Model Architecture

The model architecture of the proposed LSCCN is illustrated in Fig. 1. The proposed model includes the VAE
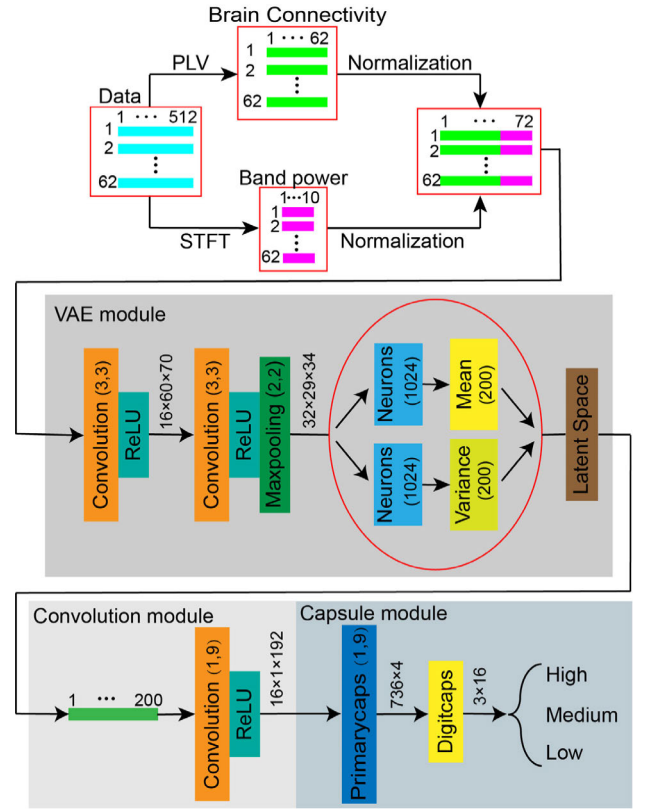

Fig. 1. The architecture of the proposed model. The brain connectivity features and band power features were extracted through PLV and STFT, respectively. Then, these two kinds of features were fused and fed into the following module. There are VAE module, convolution module, and capsule module in the proposed model. Notes: this illustration is based on the case of the no overlapping and the use of PLV.

module, the convolution module, and the capsule module. We took the case of PLV and STFT without overlapping as an example for the illustration. In this case, the size of the merged feature matrix X was 62 × 72 for each segment. This matrix X was fed into the first module (i.e., the VAE module).

The VAE module was designed to learn the latent variable from sample X. It is assumed that the sample X can be generated based on the unobserved latent variable z. Specifically, we extracted information by 2 convolutional layers containing 16 and 32 convolution kernels (3 × 3) with a stride of 1 and a max-pooling layer with the kernel of 2 × 2. After two convolutional layers and a max-pooling layer, the output (32 × 29×34) was flattened into a vector x. The latent variable is generated by the prior distribution $p_\theta(z)$. And x can be determined by the conditional distribution $p_\theta(x|z)$. However, the parameter θ and the values of the latent variable z are both unknown, and the true posterior $p_\theta(z|x) = p_\theta(x|z)p_\theta(z)/p_\theta(x)$ is intractable. Hence, Kingma et al. introduced the encoder $q_\phi(z|x)$ to generate a distribution of the possible values of the z from the sample x that aims to approximate the true posterior $p_\theta(z|x)$ [31]. Similarly, the decoder $p_\theta(x|z)$ generates a distribution over the corresponding values of x based on the latent variables z. The below formula represents the process of approximate inference of

the posterior distribution of hidden variables:

$$z \sim q_\phi(z|x). \tag{3}$$

The encoder $q_\phi(z|x)$ is a multivariate Gaussian with a diagonal covariance structure, as shown in

$$\log q_\phi(z|x) = \log N(z; \mu, \sigma^2 I), \tag{4}$$

where I is an identity matrix, μ and σlog^2 are derived by the neural networks (i.e., a fully-connected neural network). In addition, the reparameterization trick is applied. The latent variable z is sampled by using

$$z = \mu + \sigma \epsilon, \tag{5}$$

where ϵ is an auxiliary noise variable $\epsilon \sim N(0, 1)$. Specifically, the output was converted into 200 means and 200 variances by each individual neural network layer, forming a latent variable z (1×200) in the latent space.

Then, the output z from the VAE module was fed into the convolutional module. The convolutional module has 16 kernels with the size of 1×9 and a stride of 1, and the rectified linear unit (ReLU) is set as the activation function.

The output of the convolutional module was then fed into the capsule module. The capsule module has two layers (i.e., a Primarycaps layer, and a Digitcaps layer). The primarycaps layer is designed to generate primary capsules and contains 32 convolutional filters with a 1×9 kernel and a stride of 2. The feature matrixes of 1×92 were achieved by 32 filters. Each primary capsule's depth was set as 4. Specifically, the features were grouped with 4 as a unit to form $(32/4) \times 1 \times 92$ primary capsules (i). The final layer (Digitcaps layer) has three capsules (j) with a depth of 16 to represent different levels of mental workload. The length of each capsule indicates the probability of each mental workload level, and the capsules were trained by dynamic routing.

D. Dynamic Routing

The dynamic routing algorithm [34], [35], [41], [42] is used to predict higher-level capsules. Specifically, a "predicted vector" $\hat{u}_{j|i}$ is obtained through multiplying the output u_i of the i-th primary capsule by the weight matrix. The formula is as follows:

$$\hat{u}_{j|i} = W_{ij} u_i, \tag{6}$$

where $W_{ij}(j = 1, 2, 3)$ represents the weight matrix.

Secondly, the total input s_j of the j-th mental workload capsule is obtained by the weighted sum of the "predicted vector" $\hat{u}_{j|i}$ of the lower capsules. The formula is as follows:

$$s_j = \Sigma_i c_{ij} \hat{u}_{j|i}, \tag{7}$$

where c_{ij} is the coupling coefficient. The sum of all coupling coefficients between the lower capsule and all the higher capsules is 1. The coupling coefficients are determined by the SoftMax function. The SoftMax function is as follows:

$$c_{ij} = \frac{\exp(b_{ij})}{\Sigma_k \exp(b_{ik})}, \tag{8}$$

where b_{ij} is the log prior probability that the i-th primary capsule is connected to the j-th mental workload capsule, and b_{ij} is initialized with zero.

Then, the vector output v_j of j-th mental workload capsule is determined by a no-linear activation function. The formula is as follows:

$$v_j = \frac{||s_j||^2}{1 + ||s_j||^2} \frac{s_j}{||s_j||}, \tag{9}$$

This operation normalized the length of the vector output of j-th mental workload capsule to be between 0 and 1. And we updated log prior probabilities b_{ij} with the routing iterations by

$$b_{ij} \leftarrow b_{ij} + \hat{u}_{j|i} \cdot v_j, \tag{10}$$

In addition, the dynamic routing has a limit of iteration number. That is, it is stopped when a predefined number of iterations is reached. After the dynamic routing, the outputs of the capsule layer are used for the classification.

E. Loss Function

We proposed to use a combination of margin loss, reconstruction loss and the KL (Kullback-Leible) divergence to minimize the composite loss during the training. The margin loss for each mental workload capsule is calculated according to the following formula:

$$L_k = T_k \max\left(0, m^+ - ||v_k||\right)^2 + \lambda(1 - T_k) \max\left(0, ||v_k|| - m^-\right)^2, \tag{11}$$

where T_k is an indicator of the class. T_k is equal to 1 if the mental workload of class k is present, otherwise $T_k = 0$. m^+ and m^- were set as 0.9 and 0.1, which were used to punish false positives and false negatives, respectively. λ was used to adjust the proportion of the loss for absent mental workload classes, and it was set as 0.5 in our case. The reconstruction loss acted as regularization for the model [41]. The process was performed by three layers that are fully connected to the output vector v_j. The reconstruction loss of v_j was calculated by mean square error (MSE). We scaled down the reconstruction loss by 0.005 (η_1) so that it did not affect the margin loss, which was the dominant part, during training.

Moreover, while training the VAE module, the optimization objective of the VAE module consisted of two parts. The first part was the reconstruction loss calculated by binary cross entropy (BCE) through three layers fully connected to the latent variables, while the second term is the KL divergence of the approximate posterior from the prior [31]. In order that the capsule module dominated the total loss, identical to the reconstruction process in the capsule module, the optimization objective of the VAE module was also scaled down by 0.00001 (η_2). And the KL divergence was scaled down by 0.1 (η_3). Finally, the total loss included the margin losses of all mental workload capsules, the two parts of reconstruction losses, and
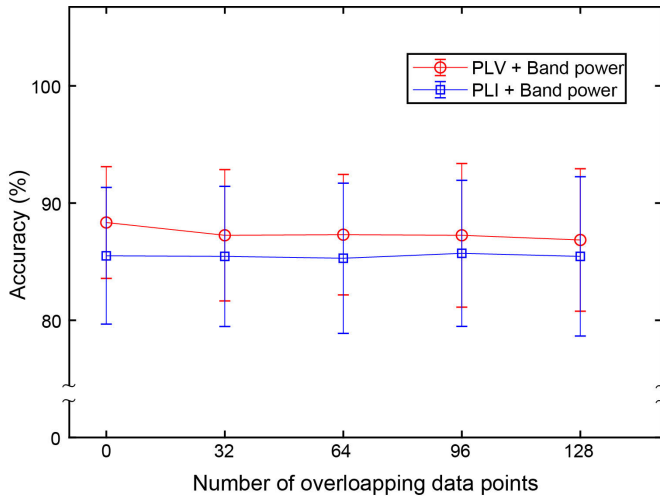
Fig. 2. Classification performance comparison of different connectivity feature extraction methods and different numbers of overlapping data points.
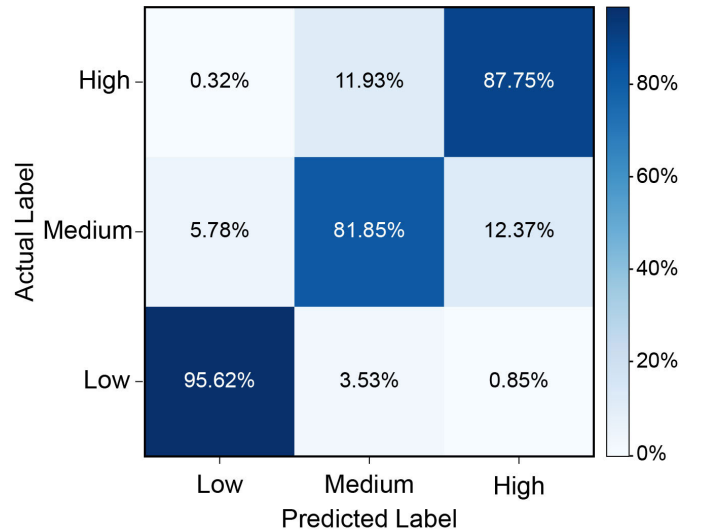

Fig. 3. The confusion matrix was averaged across all participants for the LSCCN using the combination of power features and connectivity features extracted by STFT under the condition of no overlapping and PLV.

the KL divergence:

$$\text{Total loss} = \text{margin loss}\eta_1 \times \text{reconstruction loss of } v_j$$
$$+ \eta_2 \times (\text{reconstruction loss of } z + \eta_3$$
$$\times KL(q_\phi(z|x)||p_\theta(z))). \quad (12)$$

This total loss was used during the model's training. The training process was terminated when the total loss was less than 10^{-5} or the iteration reached the predefined maximum number (i.e., 200). Also, we used an exponentially decaying learning rate to optimize the learning. The learning rate was changed with the iteration and was calculated by

$$lr_{\text{current}} = lr_{last} \times a^{epoch}, \quad (13)$$

where $lr_{current}$ represents the learning rate of the current iteration, and lr_{last} represents the learning rate of the last iteration. a is the base of the decaying learning rate during training, and $epoch$ represents the number of the current epoch.

III. RESULT AND DISCUSSIONS

A. Effects of Connectivity Extraction Methods and Overlapping Percentages

As the combination of power features and connectivity features benefits workload classification, we used feature combinations in this study. We first investigated the effects of different connectivity feature extraction methods (i.e., PLV and PLI) and different numbers of overlapping data points (0, 32, 64, 96, 128) in STFT when extracting power and connectivity features. Five-fold cross-validation was employed to evaluate the performance of mental workload classification. The classification results for each participant were listed in Table I. The classification performance was the best in participant 3. More than 90% was obtained no matter which connectivity extraction method was used and how many data points were overlapped. The lowest performance was observed in participant 5 (more than 70%). For each overlapping case, we averaged accuracies across all participants. These

average accuracies are shown in Fig. 2. The results showed that, regardless of the number of overlapping data points, the performance with PLV was higher than that with PLI. PLI could remove some information useful for classification because it employed the sign function for robustness [43]. Based on the comparison between these overlapping cases, the best performance (88.34% ± 4.77%) was obtained using PLV and STFT without any overlapping data points (i.e., the number of overlapping data points is zero). In the subsequent method comparison, the combination of power features and connectivity features was extracted by STFT under the condition of no overlapping and PLV. The confusion matrix, in this case, is shown in Fig. 3. It showed that the identification of the low workload level outperformed that of the other two levels.

B. Methods Comparison

In this study, we compared the proposed LSCCN to the state-of-the-art methods and investigated the role of each module in the LSCNN in terms of classification accuracy. These are LSTM [27], CNN [44], RF [16], a single VAE classifier, and a single capsule network. The model parameters were tuned to have optimal classification performance. The parameters of LSCCN are listed in Table II.

a) LSTM [27]: According to the existing study [27], the model consists of two LSTM layers and 100 units. Dropout rate was set to 20% and located between the first LSTM and the input layer. A fully-connected layer served at the last layer.

b) CNN [44]: The model consists of two convolution layers based on 3×3 convolutional kernels with a stride of 1. All convolutional layers are activated by the rectified linear unit (ReLU). A max-pooling layer with 2×2 kernels and a *softmax* were adopted.

c) RF [16]: In our previous study, compared with other combination features, the combination of connection and power features can achieve a higher classification accuracy (i.e., 83.12%) by the random forest (RF, 500 trees). After

TABLE I
RESULTS OF PLV AND PLI UNDER DIFFERENT OVERLAPPING DATA POINTS

Brain Connectivity	Overlapping Data Points in STFT	P1 (%)	P2 (%)	P3 (%)	P4 (%)	P5 (%)	P6 (%)	P7 (%)	Mean ± Standard Deviation (%)
PLV	0	92.04	87.22	95.93	87.78	81.48	89.45	84.45	88.34 ± 4.77
	32	91.85	86.11	95.37	86.48	79.26	89.81	81.85	87.25 ± 5.61
	64	91.67	86.30	94.81	87.41	79.82	88.70	82.41	87.30 ± 5.14
	96	91.85	86.11	95.37	88.15	76.85	89.82	82.59	87.25 ± 6.13
	128	91.85	85.93	95.54	85.93	77.59	89.44	81.67	86.85 ± 6.08
PLI	0	89.81	82.97	95.18	83.15	77.59	87.78	82.04	85.50 ± 5.83
	32	90.56	82.59	95.37	82.41	77.41	86.85	82.96	85.45 ± 5.98
	64	91.11	82.59	95.37	82.59	76.48	87.41	81.48	85.29 ± 6.41
	96	91.11	83.52	95.37	82.59	76.48	87.78	83.15	85.71 ± 6.23
	128	92.22	82.78	95.74	82.22	76.30	87.78	81.11	85.45 ± 6.80

TABLE II
MODEL PARAMETERS

	Name of Parameters	Value
Training Settings	Learning Rate (lr)	0.001
	Coefficient of Reconstruction Loss of Output Vector (η_1)	0.005
	Coefficient of Optimization Objective in VAE Module (η_2)	0.00001
	Coefficient of KL Divergence (η_3)	0.1
	Epochs	200
	Batch Size	32
	Base of Decay of Learning Rate (a)	0.9
VAE Module	Kernel Size	3×3
	Padding	None
	Stride	1
	Kernel Size	3×3
	Padding	None
	Stride	1
	Number of Neurons	1024
	Number of Neurons	200
Convolution Module	Kernel Size	1×9
	Padding	None
	Stride	1
Capsule Module	1 Kernel Size	1×9
	1 Padding	None
	1 Stride	2
	2 Vector Length	4
	2 Routings	3
	2 Vector Length	16

Fig. 4. Classification performance comparison using different methods.
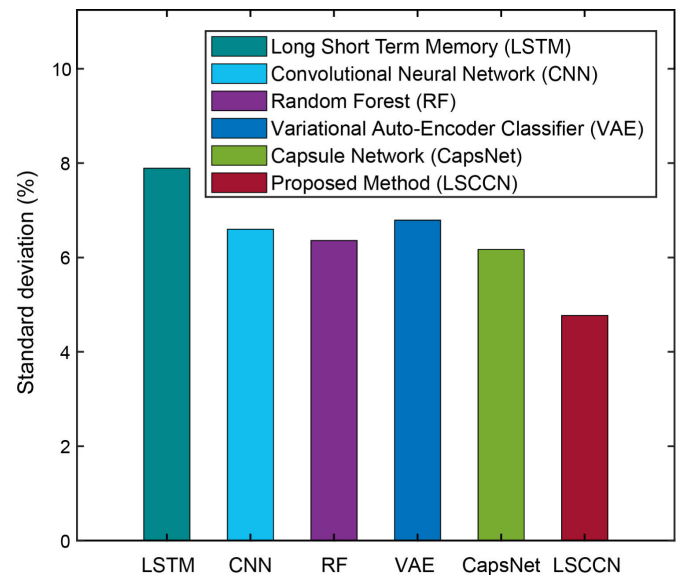
Fig. 5. Standard deviation comparison using different methods.

using feature selection, the classification accuracy of RF was improved to 84.34% with the combination of features of power features, graph metric (PLV), connection features (PLV), and F-score.

d) A single VAE classifier: The model consists of the VAE module and a *softmax*. The VAE module was kept the same as that in LSCCN.

e) A single CapsNet: The model consists of the convolution layers which were the same as those in LSCCN. In the capsule layers, each primary capsule's depth was set as 8, and each mental workload capsule was set as 32 in the Digitcaps layer.

Five-fold cross-validation was employed to evaluate the classification performance of the models and compared them in terms of accuracy mean ± standard deviation (see Fig. 4 and Fig. 5). The results showed that LSCCN had the highest average accuracy of 88.34% ± 4.77%. Th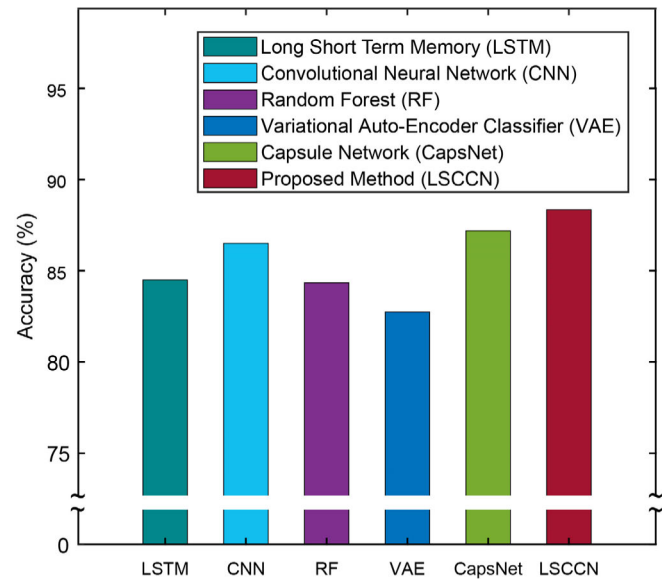e other models, LSTM [27], CNN [44], RF [16], the single VAE classifier, and the single capsule network achieved accuracies of 84.50% ± 7.89%, 86.51% ± 6.60%, 83.34% ± 6.36%, 82.75% ± 6.79%, and 87.20% ± 6.17%. LSCCN also had a
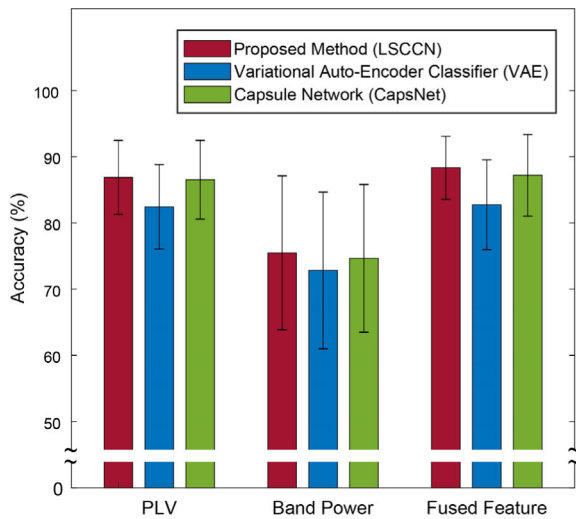
Fig. 6. Comparison of classification results for different features of each module and LSCCN.

relatively smaller standard deviation (i.e., 4.77%), indicating a more reliable classification performance across participants. The detail of individual accuracies for each participant is shown in Table III.

C. Ablation Exploration

We further conducted ablation exploration to understand the contribution to the classification performances for each module in our proposed LSCCN and individual feature categories. We disassembled the LSCCN into two main modules (i.e., VAE module and CapsNet) and checked whether or not the performance was decreased. We also checked the performance when either brain functional connectivity features or band power features were used individually. The results can be found in Tables IV. The results showed that, regardless of the feature category, the classification results of our proposed LSCCN were higher than those of the VAE classifier and the CapsNet. It indicated that our application of the VAE module was effective, and our proposed LSCCN also had a good robustness in dealing with different kinds of input features. We compared the classification results based on the individual feature category with that of the fused features (see Fig. 6). We found that the classification performance of PLV features was better than that of band power features, indicating that the information transmission between brain regions was more conducive to mental workload classification. Besides, the classification results of fused features were higher than those of individual feature categories (i.e., brain functional connectivity features and band power features), indicating that the fusion of different features could supplement each other's information to improve classification performance.

D. Features Associated With Mental Workload

We also explored band power features to gain insights into the features in relation to the differences among the levels of mental workload. Band powers were averaged over all segments and all participants for each frequency band
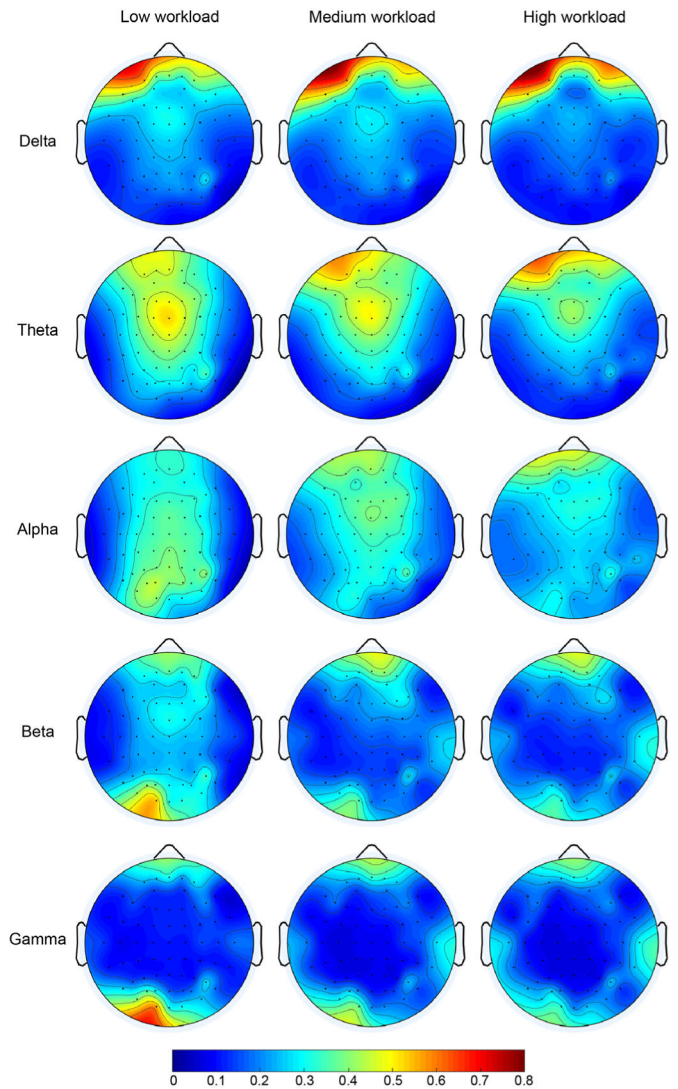

Fig. 7. The topographies of power features under low, medium, and high workload were averaged across all segments and all participants.

respectively. The average band powers are visualized in Fig. 7. The elevation of the theta power in the prefrontal area was associated with the increase of the workload. In addition, an obvious decrease in the parieto-occipital region was observed in the alpha band when the workload increased, which was in line with the finding in Kakkos etal.'s study [17]. Also, Roy et al. found a reduction in the alpha power and the beta power at the midline region when the workload increased [45]. This phenomenon was also been found in our study. Besides the power features from the low-frequency bands, the higher frequency band (i.e., gamma) also exhibited discriminative information, which can provide complementary information [46].

Likewise, we explored functional connectivity features and compared connectivity strengths between mental workload levels. The connectivity strength differences are shown in Fig. 8. It shows that connectivity strengths are dominantly different between the low workload level and the other two workload levels. The strength differences between the medium workload level and the high workload level are relatively

TABLE III
MODELS COMPARISON

Model	P1 (%)	P2 (%)	P3 (%)	P4 (%)	P5 (%)	P6 (%)	P7 (%)	Mean ± Standard Deviation (%)
LSTM [27]	91.85	83.52	95.37	83.15	72.22	87.22	78.15	84.50 ± 7.89
CNN [44]	92.22	85.56	95.37	84.63	75.18	89.26	83.34	86.51 ± 6.60
RF [16]	88.89	84.44	93.15	79.81	76.11	89.26	78.70	84.34 ± 6.36
VAE Classifier	89.44	81.30	92.78	81.48	72.22	83.33	78.70	82.75 ± 6.79
CapsNet	91.67	85.93	96.85	85.37	77.22	88.89	84.45	87.20 ± 6.17
LSCCN	92.04	87.22	95.93	87.78	81.48	89.45	84.45	88.34 ± 4.77

TABLE IV
THE CLASSIFICATION PERFORMANCES FOR INDIVIDUAL FEATURE CATEGORIES

| Feature	Model	P1 (%)	P2 (%)	P3 (%)	P4 (%)	P5 (%)	P6 (%)	P7 (%)	Mean ± Standard Deviation (%)
PLV	LSCCN	91.30	84.07	95.00	88.15	77.78	87.78	84.07	86.88±5.58
	VAE Classifier	88.33	80.74	92.22	82.04	72.59	82.41	78.70	82.43±6.38
	CapsNet	91.29	83.33	95.74	84.26	77.41	88.33	85.37	86.53±5.93
Band Power	LSCCN	86.67	74.07	91.11	68.52	61.11	82.78	64.07	75.48±11.62
	VAE Classifier	84.07	73.33	87.22	66.67	57.78	81.11	59.63	72.83±11.84
	CapsNet	84.26	75.19	89.81	67.59	61.29	81.85	62.59	74.65±11.17
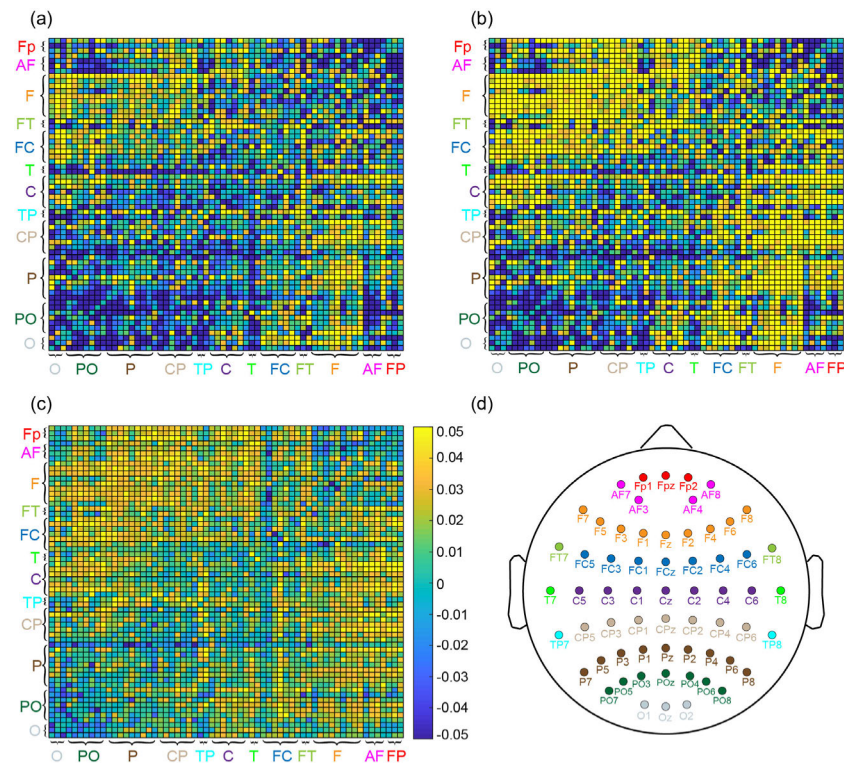
Fig. 8. Differences of connectivity strengths (obtained by PLV) averaged over all segments and all participants between mental workload levels. (a), (b), and (c) illustrates the differences between the low workload and the medium workload, between low workload and the high workload, and between medium workload and high workload, respectively. (d) illustrates the layout of electrodes used in this study.

small. This supports that the classification between medium and high levels is more difficult, and the misclassification is higher (see confusion matrix in Fig. 3). It suggests that the difference in PLV may reflect the significantly changed information transmission activity of the brain during processing tasks with the different mental workload. This may reflect the increased information transmission between brain regions when performing high mental workload tasks.

In addition, we found that most of the PLVs were significant differences in the frontal and parietal regions. In addition, the connection of some nodes in the occipital region also showed differences under the different kinds of mental workload. It indicates that the functional connection features between the different brain regions considerably impact the level of mental workload. This corroborates with previous studies indicating frontal-parietal modulations in regard to mental workload [18],
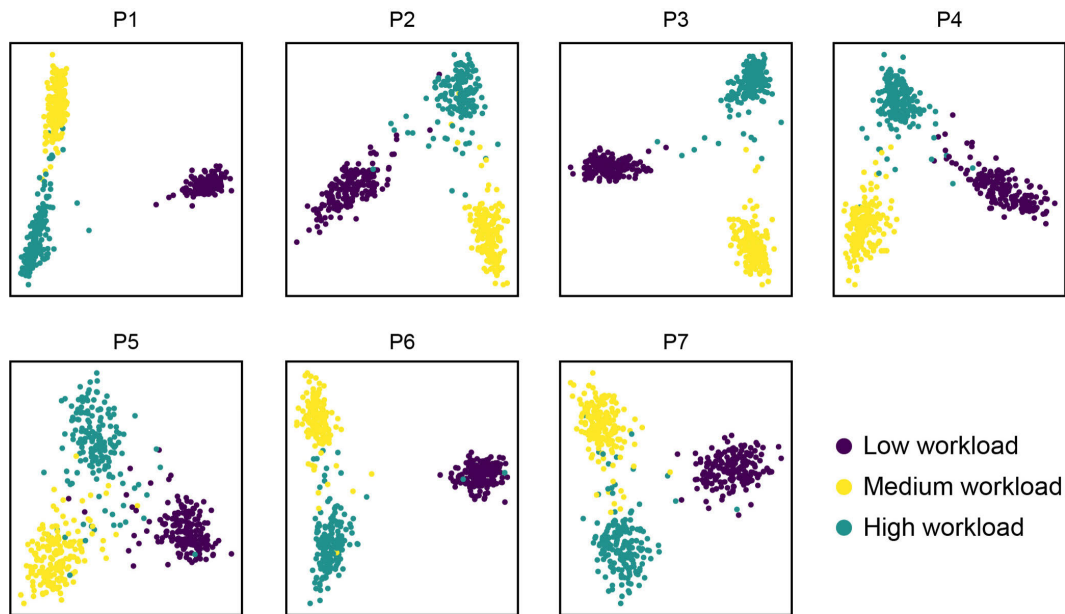
Fig. 9. Visualization of the latent variables after dimension reduction using principal component analysis.

[46], [47], [48], [49]. For example, Guan et al. found that some nodes and some connectivities changed significantly in the frontal-parietal region with the mental workload [49]. Similarly, Kakkos's study found that workload was associated with elevated frontal and parietal regions in the n-Back task and the Mental Arithmetic task [17]. The superior frontal region supports working memory processes. It has been proved that the frontal region is involved in mental arithmetic-induced workload [50] and the fronto-executive dysfunction may lead to impairment of arithmetic function [51]. In addition to the prefrontal cortex, working memory was also supported by the parietal region [52]. Our study added favorable evidence for the effect of the frontal and parietal regions on the mental workload.

E. Insight Into LSCCN

In the LSCCN model, the VAE module can encode input data into latent variables in the latent space, which can be decoded to recover the input data as much as possible. It ensures that the input data can be embedded into a lower dimension space while retaining the discriminative information. In order to gain insights into the VAE module, we visualized the latent variables. Because the dimension of features was still much more than two, even if it had been reduced compared to the dimension of the input data, we used principal component analysis to reduce the dimension of features to two. The samples were visualized by these two-dimensional features in Fig. 9. From the visualization, it is obvious that the samples from the same class were generally gathered together to form a cluster. For example, samples were clearly separated among low, medium, and high workload levels in participant 3. Separation is also quite good in participant 1. It means that the features extracted by the VAE module well represent discriminative information between workload levels. This can be proven by the results

of workload classification. The classification accuracies are high (more than 90%) for these two participants. In contrast, samples are relatively overlapped in participant 5 (see P5 in Fig. 9). The samples from different workload levels are superimposed between each other. Accordingly, the workload classification performance is relatively low in participant 5. It is also found that the overlap is large between the medium workload level and the high workload level. The low workload level is generally away from the other two workload levels. This is consistent with the result of the confusion matrix. It is also in agreement with the participant's involvement in the tasks of the experiment. Participants were required to pay much more effort in the medium and high workload tasks compared to that in the low workload task. The brain tended to enhance the efficiency of information transmission between brain regions to deal with the pressure caused by higher workload tasks [49].

F. Limitations

In this study, the number of participants is not large, which limits the confidence of the findings. The findings derived from this study need a further study with a much larger number of participants to be confirmed. However, the limitation of a small number of participants does not degrade the contribution of the study because the main contribution of the study is to provide an effective classification model for the mental workload. The reported findings are added contributions.

IV. CONCLUSION

In this study, we proposed a novel deep learning model, latent space coding capsule network (LSCCN), to classify multiple levels of mental workload. The LSCCN mainly includes the VAE module, the convolution module, and the capsule module. The VAE module creates latent variables learned from the fused features of EEG to enhance the

robustness of feature representation. The capsule module can capture the relationships between the latent variables for better mental workload classification. The proposed model was compared to LSTM, CNN, RF, the single VAE classifier, and the single capsule network in terms of the classification performance. The comparison results demonstrated that the proposed model outperformed all these compared methods. We further explored band powers, connectivity strengths, and latent variables to gain insights into the mental workload. We found that modulations in the frontal, parietal and occipital regions were closely related to mental workload. Through the visualization of the latent variables, we can see that the latent variables are vital for the workload classification. In this study, we not only proposed a promising model for mental workload classification, but also provided the insights into mental workload from the perspective of feature representation. These could give a positive effect on the investigation of mental workload and the development of workload classification models in the future.

REFERENCES

[1] Y. Kato, H. Endo, and T. Kizuka, "Mental fatigue and impaired response processes: Event-related brain potentials in a Go/NoGo task," *Int. J. Psychophysiol.*, vol. 72, no. 2, pp. 204–211, May 2009.

[2] D. Kamioka, K. Murai, K. Kitamura, and S.-I. Wakida, "Construction of port-coordinator duties list for mental workload assessment," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2017, pp. 3409–3414.

[3] J. Kohlmorgen et al., "Improving human performance in a real operating environment through real-time mental workload detection," *Toward Brain-Comput. Interfacing*, vol. 409422, pp. 409–422, Jul. 2007.

[4] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (task load index): Results of empirical and theoretical research," in *Advances in Psychology*. Amsterdam, The Netherlands: Elsevier, 1988, pp. 139–183.

[5] R. Parasuraman, "Neuroergonomics: Brain, cognition, and performance at work," *Current Directions Psychol. Sci.*, vol. 20, no. 3, pp. 181–186, Jun. 2011.

[6] Y. Wu, T. Miwa, and M. Uchida, "Using physiological signals to measure operator's mental workload in shipping-an engine room simulator study," *J. Mar. Eng. Technol.*, vol. 16, no. 2, pp. 61–69, 2017.

[7] K. Ryu and R. Myung, "Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic," *Int. J. Ind. Ergonom.*, vol. 35, no. 11, pp. 991–1009, Nov. 2005.

[8] J. Kopf, T. Dresler, P. Reicherts, M. J. Herrmann, and A. Reif, "The effect of emotional content on brain activation and the late positive potential in a word n-back task," *PLoS ONE*, vol. 8, no. 9, Sep. 2013, Art. no. e75598.

[9] K. Yokoyama, J. Ushida, Y. Sugiura, M. Mizuno, Y. Mizuno, and K. Takata, "Heart rate indication using musical data," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 7, pp. 729–733, Jul. 2002.

[10] C. Berka et al., "EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks," *Aviation, Space, Environ. Med.*, vol. 78, no. 5, pp. B231–B244, 2007.

[11] A. Gevins, "High-resolution EEG mapping of cortical activation related to working memory: Effects of task difficulty, type of processing, and practice," *Cerebral Cortex*, vol. 7, no. 4, pp. 374–385, Jun. 1997.

[12] A.-M. Brouwer, M. A. Hogervorst, J. B. F. van Erp, T. Heffelaar, P. H. Zimmerman, and R. Oostenveld, "Estimating workload using EEG spectral power and ERPs in the n-back task," *J. Neural Eng.*, vol. 9, no. 4, Aug. 2012, Art. no. 045008.

[13] J. C. Christensen, J. R. Estepp, G. F. Wilson, and C. A. Russell, "The effects of day-to-day variability of physiological data on operator functional state classification," *NeuroImage*, vol. 59, no. 1, pp. 57–63, Jan. 2012.

[14] M. Pesonen, H. Hämäläinen, and C. M. Krause, "Brain oscillatory 4–30 Hz responses during a visual n-back memory task with varying memory load," *Brain Res.*, vol. 1138, pp. 171–177, Mar. 2007.

[15] T. I. Laine, K. W. Bauer, J. W. Lanning, C. A. Russell, and G. F. Wilson, "Selection of input features across subjects for classifying crewmember workload using artificial neural networks," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 32, no. 6, pp. 691–704, Nov. 2002.

[16] Z. Pei, H. Wang, A. Bezerianos, and J. Li, "EEG-based multiclass workload identification using feature fusion and selection," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–8, 2021.

[17] I. Kakkos et al., "EEG fingerprints of task-independent mental workload discrimination," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 10, pp. 3824–3833, Oct. 2021.

[18] J. Fan, J. W. Wade, A. P. Key, Z. E. Warren, and N. Sarkar, "EEG-based affect and workload recognition in a virtual driving environment for ASD intervention," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 1, pp. 43–51, Jan. 2018.

[19] J. A. Blanco et al., "Quantifying cognitive workload in simulated flight using passive, dry EEG measurements," *IEEE Trans. Cognit. Develop. Syst.*, vol. 10, no. 2, pp. 373–383, Jun. 2018.

[20] H. Gu, Z. Yin, and J. Zhang, "EEG based mental workload assessment via a hybrid classifier of extreme learning machine and support vector machine," in *Proc. Chin. Control Conf. (CCC)*, Jul. 2019, pp. 8398–8403.

[21] J. Cui, Z. Lan, O. Sourina, and W. Müller-Wittig, "EEG-based cross-subject driver drowsiness recognition with an interpretable convolutional neural network," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Feb. 16, 2022, doi: 10.1109/TNNLS.2022.3147208.

[22] O.-Y. Kwon, M.-H. Lee, C. Guan, and S.-W. Lee, "Subject-independent brain–computer interfaces based on deep convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 10, pp. 3839–3852, Oct. 2020.

[23] Z. Sun et al., "WLNet: Towards an approach for robust workload estimation based on shallow neural networks," *IEEE Access*, vol. 9, pp. 3165–3173, 2021.

[24] D.-H. Lee, J.-H. Jeong, K. Kim, B.-W. Yu, and S.-W. Lee, "Continuous EEG decoding of pilots' mental states using multiple feature block-based convolutional neural network," *IEEE Access*, vol. 8, pp. 121929–121941, 2020.

[25] P. Zhang, X. Wang, J. Chen, W. You, and W. Zhang, "Spectral and temporal feature learning with two-stream neural networks for mental workload assessment," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 6, pp. 1149–1159, Jun. 2019.

[26] R. G. Hefron, B. J. Borghetti, J. C. Christensen, and C. M. S. Kabban, "Deep long short-term memory structures model temporal dependencies improving cognitive workload estimation," *Pattern Recognit. Lett.*, vol. 94, pp. 96–104, Jul. 2017.

[27] U. E. Manawadu, T. Kawano, S. Murata, M. Kamezaki, J. Muramatsu, and S. Sugano, "Multiclass classification of driver perceived workload using long short-term memory based recurrent neural network," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 1–6.

[28] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, "Learning representations from EEG with deep recurrent-convolutional neural networks," 2015, *arXiv:1511.06448*.

[29] J. Li, C. Li, and A. Cichocki, "Canonical polyadic decomposition with auxiliary information for brain–computer interface," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 1, pp. 263–271, Jan. 2017.

[30] X. Li et al., "Variational autoencoder based latent factor decoding of multichannel EEG for emotion recognition," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Nov. 2019, pp. 684–687.

[31] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*.

[32] S. You, B. Hwan Cho, Y.-M. Shon, D.-W. Seo, and I. Y. Kim, "Semi-supervised automatic seizure detection using personalized anomaly detecting variational autoencoder with behind-the-ear EEG," *Comput. Methods Programs Biomed.*, vol. 213, Jan. 2022, Art. no. 106542.

[33] Y. Yu and J. Li, "Feature fusion-based capsule network for cross-subject mental workload classification," in *Brain Informatics*, M. Mahmud, J. He, S. Vassanelli, A. van Zundert, and N. Zhong, Eds. Cham, Switzerland: Springer, 2022, pp. 164–174.

[34] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proc. Adv. Neural Inf. Process Syst.*, vol. 30, 2017.

[35] T. Wang, A. Bezerianos, A. Cichocki, and J. Li, "Multikernel capsule network for schizophrenia identification," *IEEE Trans. Cybern.*, vol. 52, no. 6, pp. 4741–4750, Jun. 2022.

[36] Z. Wang, C. Chen, J. Li, F. Wan, Y. Sun, and H. Wang, "ST-CapsNet: Linking spatial and temporal attention with capsule network for P300 detection improvement," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 991–1000, 2023.

[37] P. Tass et al., "Detection of *n:m* phase locking from noisy data: Application to magnetoencephalography," *Phys. Rev. Lett.*, vol. 81, no. 15, pp. 3291–3294, 1998.

[38] P. Celka, "Statistical analysis of the phase-locking value," *IEEE Signal Process. Lett.*, vol. 14, no. 9, pp. 577–580, Sep. 2007.

[39] S. Aydore, D. Pantazis, and R. M. Leahy, "A note on the phase locking value and its properties," *NeuroImage*, vol. 74, pp. 231–244, Jul. 2013.

[40] C. J. Stam, G. Nolte, and A. Daffertshofer, "Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources," *Hum. Brain Mapping*, vol. 28, no. 11, pp. 1178–1193, 2007.

[41] Y. Liu et al., "Multi-channel EEG-based emotion recognition via a multi-level features guided capsule network," *Comput. Biol. Med.*, vol. 123, Aug. 2020, Art. no. 103927.

[42] K.-W. Ha and J.-W. Jeong, "Motor imagery EEG classification using capsule networks," *Sensors*, vol. 19, no. 13, p. 2854, Jun. 2019.

[43] S.-E. Moon, S. Jang, and J.-S. Lee, "Convolutional neural network approach for EEG-based emotion recognition using brain connectivity and its spatial information," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 2556–2560.

[44] U. Asgher, K. Khalil, Y. Ayaz, R. Ahmad, and M. J. Khan, "Classification of mental workload (MWL) using support vector machines (SVM) and convolutional neural networks (CNN)," in *Proc. 3rd Int. Conf. Comput., Math. Eng. Technol. (iCoMET)*, Jan. 2020, pp. 1–6.

[45] R. N. Roy, S. Charbonnier, A. Campagne, and S. Bonnet, "Efficient mental workload estimation using task-independent EEG features," *J. Neural Eng.*, vol. 13, no. 2, Apr. 2016, Art. no. 026019.

[46] C. Mühl, C. Jeunet, and F. Lotte, "EEG-based workload estimation across affective contexts," *Frontiers Neurosci.*, vol. 8, p. 114, Jun. 2014.

[47] G. N. Dimitrakopoulos et al., "Task-independent mental workload classification based upon common multiband EEG cortical connectivity," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 11, pp. 1940–1949, Nov. 2017.

[48] J. Kaiser and W. Lutzenberger, "Cortical oscillatory activity and the dynamics of auditory memory processing," *Rev. Neurosci.*, vol. 16, no. 3, pp. 239–254, Jan. 2005.

[49] K. Guan, Z. Zhang, X. Chai, Z. Tian, T. Liu, and H. Niu, "EEG based dynamic functional connectivity analysis in mental workload tasks with different types of information," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 632–642, 2022.

[50] G. Sammer et al., "Relationship between regional hemodynamic activity and simultaneously recorded EEG-theta associated with mental arithmetic-induced workload," *Hum. Brain Mapping*, vol. 28, no. 8, pp. 793–803, 2007.

[51] M. Delazer, F. Domahs, A. Lochy, E. Karner, T. Benke, and W. Poewe, "Number processing and basal ganglia dysfunction: A single case study," *Neuropsychologia*, vol. 42, no. 8, pp. 1050–1062, Jan. 2004.

[52] P. Sauseng, W. Klimesch, M. Schabus, and M. Doppelmayr, "Fronto-parietal EEG coherence in theta and upper alpha reflect central executive functions of working memory," *Int. J. Psychophysiol.*, vol. 57, no. 2, pp. 97–103, Aug. 2005.