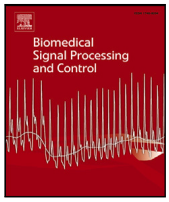
Biomedical Signal Processing and Control 89 (2024) 105703

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc



Integrated Spatio-Temporal Deep Clustering (ISTDC) for cognitive workload assessment

Debashis Das Chakladar [a,*], Partha Pratim Roy [b,1], Victor Chang [c,1]

[a] Department of Electronics and Communication Sciences Unit, Indian Statistical Institute Kolkata, West Bengal 700108, India
[b] Department of Computer Science and Engineering, Indian Institute of Technology Roorkee, Uttarakhand 247667, India
[c] Department of Operations and Information Management, Aston Business School, Aston University, UK

ARTICLE INFO

Keywords:
Cognitive workload
Electroencephalography
Deep clustering
Variational bayesian gaussian mixture model

ABSTRACT

Traditional high-dimensional electroencephalography (EEG) features (spectral or temporal) may not always attain satisfactory results in cognitive workload estimation. In contrast, deep representation learning (DRL) transforms high-dimensional data into cluster-friendly low-dimensional feature space. Therefore, this paper proposes an Integrated Spatio-Temporal Deep Clustering (ISTDC) model that uses DRL followed by a clustering method to achieve better clustering performance. The proposed model is illustrated using four Algorithms and Variational Bayesian Gaussian Mixture Model (VBGMM) clustering method. Temporal and spatial Variational Auto Encoder (VAE) models (mentioned in Algorithm 2 and Algorithm 3) learn temporal and spatial latent features from sequence-wise EEG signals and scalp topographical maps using the Long short-term memory and Convolutional Neural Network models. The concatenated spatio-temporal latent feature (mentioned in Algorithm 4) is passed to the VBGMM clustering method to efficiently estimate workload levels of *n*-back task. For the 0-back vs. 2-back task, the proposed model achieves the maximum mean clustering accuracy of 98.0%, and it improves by 11.0% over the state-of-the-art method. The results also indicate that the proposed multimodal approach outperforms temporal and spatial latent feature-based unimodal models in workload assessment.

1. Introduction

Cognitive workload can be defined as a multidimensional construct representing the load that performing a particular task imposes on the learner's cognitive system [1]. The workload level of an operator can be evaluated using subjective or physiological measures. To assess workload levels using traditional subjective measures, self-rating-based methods are used where results mostly depend on an individual subject's honesty. Therefore, an objective measurement based on physiological signals is indispensable. There exist several types of physiological measures that include cardiac activity: Electrocardiography (ECG), respiratory activity, eye activity, and brain activity: Electroencephalography (EEG), Functional magnetic resonance imaging (fMRI), Near-infrared spectroscopy (NIRS), Functional near-infrared spectroscopy (fNIRS). Due to the ability to reflect the electrical activities of the cortex, EEG is used as the most effective physiological measure for estimating cognitive workload levels [2]. EEG is used in several cognitive applications such as emotion classification [3,4], mental arithmetic tasks [5], *n*-back tasks [6], and simultaneous multitasking activities [7]. Mostly, high-dimensional power spectral density (PSD) features [8], or event-related potential (ERP) features [9] of EEG is used for workload estimation. EEG features have been used in several medical applications. A hybrid framework consisting of a complex brain network and Takagi–Sugeno–Kang fuzzy system has been used to identify Alzheimer's disease [10]. The topological features of functional brain networks have been used for efficient classification. The spectral power of EEG has been used in acupuncture stimulation [11]. Apart from PSD, topological graph features such as clustering coefficient, global efficiency, etc have been used in acupuncture manipulation [12]. However, the classification model using the unimodal hand-crafted EEG features (temporal: ERP or spectral: PSD) does not give a satisfactory result [13]. Multimodal fusion technique can be implemented by combining data of different sensors [14], by fusing different types of features of the same sensor [15] or based on the decision of different unimodal feature extractors [16]. Zhang et al. [17] have developed a deep multimodal framework consisting of temporal and spectral EEG features and a deep Convolutional Neural Network (CNN) model to

* Corresponding author.
E-mail addresses: ddaschakladar@gmail.com (D.D. Chakladar), partha@cs.iitr.ac.in (P.P. Roy), v.chang1@aston.ac.uk, victorchang.research@gmail.com (V. Chang).
[1] Senior Member, IEEE.

https://doi.org/10.1016/j.bspc.2023.105703
Received 10 June 2023; Received in revised form 9 October 2023; Accepted 5 November 2023
Available online 16 November 2023

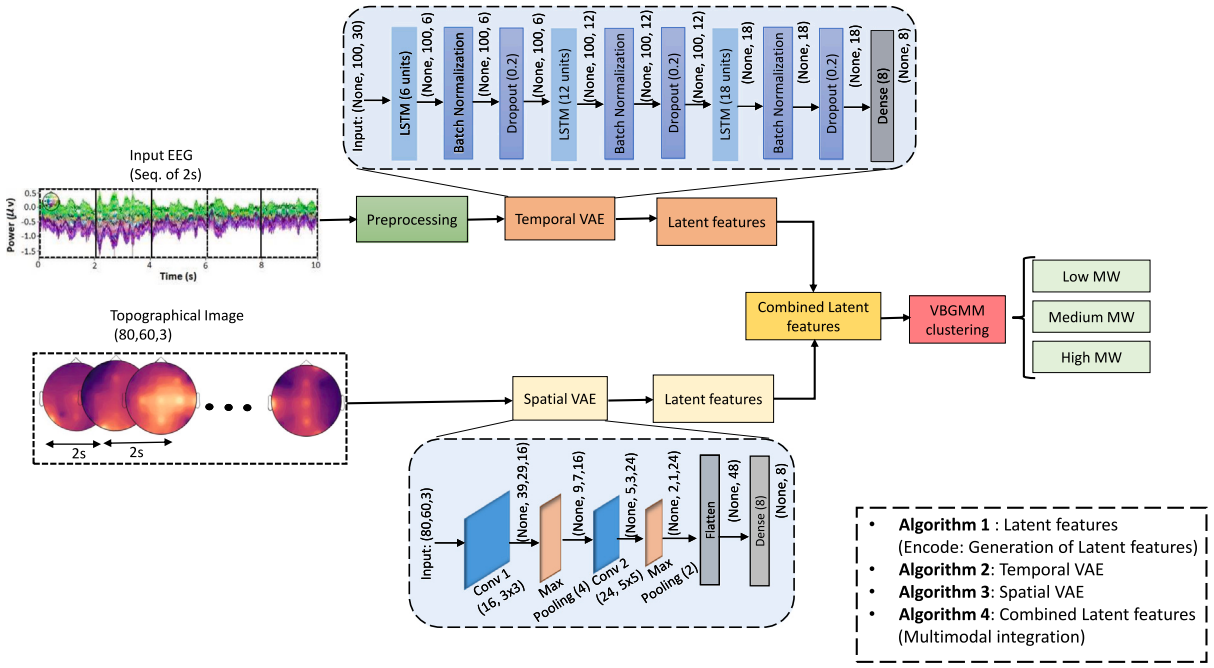


Fig. 1. Workload estimation using the proposed Integrated Spatio-Temporal Deep Clustering (ISTDC) framework. The framework is constructed using four proposed algorithms (Algorithm 1–Algorithm 4). Inputs are segregated based on stimulus representation of 2 s sequence window. The latent features (extracted from LSTM and CNN-based VAEs) are fused, and the merged latent feature is transmitted to the VBGMM clustering method to estimate workload levels.

estimate workload levels from *n*-back task. The multimodal-CNN model achieved 91.9% classification accuracy. Spatial–temporal autoencoder (STAE) constructed from CNN and Long short-term memory (LSTM) has been used to find the brain dynamics of Alzheimer patients [18]. The STAE model has achieved 96.30% classification accuracy. In [19], authors have merged multimodal features from EEG (PSD and ERP features) and ECG (Heart rate: HR feature) for the *n*-back experiment. In addition, HR and heart rate variability (HRV) features have been extracted from ECG signals. The fused features from both EEG and ECG sensors were passed to the different classifiers (k-nearest neighbors: kNN, SVM). Among different classifiers, SVM achieves the maximum classification accuracy of 90.6%. Lin et al. [20] have combined EEG (spectral power of EEG bands), and fNIRS (oxygenated hemoglobin: HbO and deoxygenated hemoglobin: HbR) signals to identify workload levels in the lane-deviation driving task. The results showed that an increased concentration of HbO and variation of EEG spectral power for theta, alpha, and beta bands were associated with poor driving performance. However, the feature space becomes large when combining multiple sensor/ modality data and high dimensional feature vectors, often giving poor classification results [21]. Moreover, traditional machine learning or deep learning-based classification methods generally suffer from high computational complexity on large-scale data [22]. To mitigate the computational complexity, the deep representation learning (DRL) method is extensively used alongside clustering to map the input data into feature space in such a way that separation in class becomes more relevant to the problem's context [23].

Deep neural network architecture such as AutoEncoder (AE) is mainly used in the DRL-based clustering techniques that embed the higher-level features from input data to capture task-specific contextual information. However, due to the complex behavior of EEG signals, AE and Sparse AutoEncoder (SAE)-based deep models [17,24] are unable to identify the underlying structure of the input EEG data. The Variational AutoEncoder (VAE) overcomes this issue by implementing the variational Bayesian method, which improves the scalability of the base network and, consequently, enhances the clustering performance [22]. The VAE model extracts the probabilistic properties from underlying EEG data [25]. Probabilistic properties allow better understanding and inference of underlying cognitive states by identifying the probability distributions of EEG features associated with different cognitive workload levels. In contrast, AE and sparse AE-based deep models are unable to find the underlying probability distributions of EEG features. Moreover, VAE also extracts noise-free localized latent features from input EEG, which improves classification results of EEG tasks [26]. EEG-based applications such as speech recognition [27] or emotion classification [28] have been implemented using the LSTM-based VAE model. VAE model constructed using CNN has been used for motor imagery classification [29]. However, for the large-sized EEG dataset, the computational complexity of deep models is expensive [22]. In contrast, in data clustering, samples (without class labels) are grouped into different clusters by minimizing inter-cluster similarity and maximizing intra-cluster similarity [21]. In the Variational Bayesian Gaussian Mixture Model (VBGMM), an appropriate model selection algorithm has been performed using the posterior distribution of data [30]. VBGMM has been used in EEG-based epileptic seizure detection [31], speaker identification [32] and speech recognition [33]. The VBGMM clustering method leads to better cluster covariance than other clustering methods [21].

Recent DRL-based clustering studies mainly highlighted the image recognition task [34,35], so designing a robust DRL-based clustering method that combines multiple EEG features (temporal and spatial) for workload estimation is of utmost demand. As EEG is suffering from the curse-of-dimensionality issue [21], and traditional classification techniques suffer from large computational complexity for executing large amounts of EEG data, it is essential to propose a DRL-based clustering model that can effectively estimate the workload levels from low-dimensional latent features. The proposed Integrated Spatio-Temporal Deep Clustering (ISTDC) framework is constructed using four algorithms (Algorithm 1–Algorithm 4). The temporal and spatial information from EEG data is captured using LSTM and CNN-based VAE in Algorithm 2 and Algorithm 3, respectively. Outputs of these algorithms (i.e., low-dimensional latent features of temporal and spatial VAEs) are combined into the final latent feature (output of Algorithm 4). The final merged latent feature is passed to the VBGMM clustering method for identifying workload levels. The proposed framework and its relationship with the Algorithms (1–4) are plotted in Fig. 1.

The contributions of this study are mentioned below:

(1) The proposed ISTDC framework is illustrated by four Algorithms (Encode, Temporal VAE, Spatial VAE, and Multimodal integration) followed by a deep clustering method. As the clustering efficiency mostly depends on learned feature representation [21], the VBGMM clustering method can effectively classify workload levels using the combined (temporal and spatial) deep latent feature.
(2) For the 0 vs. 2-back task, the proposed model achieves the maximum classification accuracy, and the proposed multimodal model demonstrates superior performance than the unimodal (spatial and temporal) VAE-based clustering approaches by 15.8% and 13.7%, respectively. Furthermore, the multimodal framework exceeds the current deep clustering approaches, improving clustering accuracy by 13.5%. The proposed model is also tested over two other open-access *n*-back datasets to demonstrate its effectiveness.
(3) Different kinds of comparison studies are performed to evaluate the efficiency of the proposed model. A significant performance improvement of the proposed model is observed for all types of comparisons.

The remainder of the paper is designed as follows. The proposed model is discussed in Section 2. The experimental results of the proposed model are represented in Section 3. A brief discussion is illustrated in Section 4. Finally, in Section 5, the paper is concluded with future work.

2. Methodology

The proposed ISTDC model is divided into three subsections: (A) Dataset and experiment analysis, (B) Integrated spatio-temporal VAE (IST-VAE) model, and (C) Cognitive workload estimation using VBGMM. A detailed description of each subsection is mentioned below.

2.1. Dataset and experimental analysis

An open-access public dataset [36] is used for evaluating the proposed model. The dataset contains EEG recordings of 26 subjects (9 males and 17 females, average age of 26.1 ± 3.5 years). EEG data were recorded using 30 EEG electrodes (Fp1, Fp2, AFF5 h, AFF6 h, AFz, F1, F2, FC1, FC2, FC5, FC6, Cz, C3, C4, T7, T8, CP1, CP2, CP5, CP6, Pz, P3, P4, P7, P8, POz, O1, O2, TP9 (reference) and TP10 (ground)) at a sampling rate of 1,000 Hz. Then, the raw EEG was filtered with a passband of 1–40 Hz to remove high-frequency noise from EEG. Next, the Independent component analysis (ICA) was applied to remove artifacts (ocular, cardiac) from EEG.

The experimental *n*-back dataset includes three sessions, where each session is divided into three series. In the experiment, nine series of *n*-back tasks are performed for each participant. Each series consists of 20 trials; thus the experiment includes a total of 180 (20 trials × 3 series × 3 sessions) trials for each *n*-back task. A single series was composed of a 2 s instruction showing the type of the task (0-, 2- or 3-back), a 40 s task period, and a 20 s rest period. In the *n*-back task, participants need to identify the letter/digit presented *n* trials earlier in the sequence. In the task state, a digit (i.e., stimulus) randomly appeared on a screen for 2 s. In the rest state, subjects were asked to keep their eyes closed without performing any task.

2.2. Integrated spatio-temporal VAE model

The DRL-based method, A non-linear mapping function $f_\theta : X \rightarrow Z$ converts the high-dimensional input $X \in \mathbb{R}^d$ into low-dimensional embedded feature space ($Z \in \mathbb{R}^k$, where $k \ll d$). The input and reduced dimensions are represented as d and k, respectively. Due to better feature learning capabilities and function approximation properties than other deep neural networks (DNN) methods, AEs are widely used for parameterization of f_θ. In VAE, the regularization effect on the latent variables overcomes the overfitting issue of AE. The entire process of the IST-VAE model is illustrated using four algorithms. Algorithm 1 depicts the encoding process of input data into a latent variable. Algorithm 2 and Algorithm 3 illustrate the construction process of the encoder in both VAEs. The two encoders ($TEncoder$ in Algorithm 2 and $SEncoder$ in Algorithm 3) of VAEs have extracted the corresponding latent features from inputs. Two encoders are built using LSTM and CNN models. Algorithm 4 concatenates the latent features ($SpLat$ and $TemLat$) obtained from $TEncoder$ and $SEncoder$ and passed to the VBGMM clustering method for workload classification. In the experiment, the sequence-specific (two seconds) EEG signals and topographical images were passed as input to the temporal and spatial VAEs. VAE extracts the localized features from topographical images of EEG [37].

Morlet wavelet features are useful to identify the localized changes of signal for different frequency components over time [38]. It can be noted that Common Spatial Pattern (CSP), spectral power and mean absolute band power features of the alpha band of EEG can effectively classify workload levels from different cognitive tasks such as "simultaneous capacity-based multitasking activity [7]", "mental arithmetic [39]" and "modified Sternberg task [40]" respectively. So, for different cognitive tasks, it can be observed that the alpha band of EEG can effectively classify the workload states than other EEG bands. Hence, the Morlet wavelet features are estimated based on the best EEG band (i.e., alpha band) for workload estimation. In the 40 s of the specific task (0/2/3-back) period in the experiment, each digit displays for 2 s; thus, 20 scalp topographical images ($\frac{40s}{2s} = 20$) are extracted for each type of task. So, 20 task-specific topographical images are constructed from one series. Thus, for nine such series, a total of 180 (20 × 9) images are constructed for each subject. Scalp topographical maps are generated from the sequence (time interval of 2 s.)-specific Morlet wavelet features. The topographical map refers to the spatial distribution of brain electrical activity for a specific frequency band (here, alpha band). The proposed spatial VAE model takes the topographical images (size: 80×60 × 3, where 80, 60, and 3 are represented as height, width, and number of image channels) as input (line 2 of Algorithm 3). The construction of the CNN-based VAE model (Spatial VAE) is illustrated in line 2 – 11 of Algorithm 3. In the spatial VAE model, 16 filters of size: 3 × 3 are applied on inputs, and outputs of the first convolutional layer are passed to the first Max pooling layer (width: 4 and stride: 2 × 2). The output of the first Max-pooling layer is passed to the second convolutional layer of 24 filters (size: 5 × 5). The output of the second convolutional layer is passed to the second Max pooling layer (width: 2 and stride: 1 × 1). The second Max pooling layer is followed by a Flatten layer. Finally, a dense layer of eight neurons (best latent feature dimension, refer to Fig. 4(a)) completes the model configuration. The construction process of temporal VAE using stacked LSTM layers is depicted in line 4 to 17 of Algorithm 2. Three LSTM layers (6, 12, and 18 units, respectively) are used to build the proposed temporal VAE, where the output of the first LSTM layer is passed as an input to the second LSTM layer. Next, the output of the second LSTM layer is fed into the third LSTM layer. A batch normalization and dropout layer (dropout rate: 0.2) are appended after each LSTM layer. A dense layer of eight neurons (best latent feature dimension, refer to Fig. 4(a)) followed by the third dropout layer (i.e., $d3$) completes the model configuration. For both VAEs (Algorithm 2 and Algorithm 3), the input is encoded into latent variables ($TemLat$ and $SpLat$) through the Encoding algorithm (Algorithm 1). The output dimension of both VAEs is maintained at the same value (i.e., eight) to combine the resultant latent features of both temporal and spatial VAEs into a shared embedded space. Fig. 1 shows the configuration of temporal and spatial VAE. The merging/concatenation process of two latent features is performed in line 1 in Algorithm 4.

Algorithm 1: Encode

Input: Model, Inputs (X)
Output: μ_z, σ_z, z

$\mu_z, \sigma_z \leftarrow Split(Model, X, 2)$
$batch \leftarrow \mu_z[0]$
$dim \leftarrow \mu_z[1]$
$eps \leftarrow$ Random vector with size (batch, dim)
$z \leftarrow \mu_z + \exp(0.5 * \sigma_z) * eps$
Return μ_z,σ_z, z

Algorithm 2: Temporal VAE

Input: $EEGdata(D) : \{x^1, x^2 .....x^N\}$
Output: $TemLat$

$ep \leftarrow 1e-04$, $M_2 \leftarrow Latent\ dimension$
$timesteps, features \leftarrow extracted\ fromD$
$Inputs \leftarrow \langle timesteps, features \rangle$
$Model \leftarrow Inputs$
$L1 \leftarrow Model.Add(LSTM(6, Inputs))$
$b1 \leftarrow Model.Add(BatchNormalization(ep, L1))$
$d1 \leftarrow Model.Add(Dropout(0.2, b1))$
$L2 \leftarrow Model.Add(LSTM(12, d1))$
$b2 \leftarrow Model.Add(BatchNormalization(ep, L2))$
$d2 \leftarrow Model.Add(Dropout(0.2, b2))$
$L3 \leftarrow Model.Add(LSTM(18, d2))$
$b3 \leftarrow Model.Add(BatchNormalization(ep, L3))$
$d3 \leftarrow Model.Add(Dropout(0.2, b3))$
$\mu, \sigma, TemLat \leftarrow Encode(Inputs)$
$M_2 \leftarrow dim(TemLat)$
$den \leftarrow Model.Add(Dense(M_2, d3))$
$TVAE \leftarrow Model(Inputs, [\mu, \sigma, TemLat], name = TEncoder)$
Return $TemLat$

Algorithm 3: Spatial VAE

Input: 3D Topographical Images: S
Output: $SpLat$

$M_1 \leftarrow Latent\ dimension$
$Model \leftarrow Inputs(80, 60, 3)$
$C1 \leftarrow Model.Add(Conv2D(16, Inputs)$
$M1 \leftarrow Model.Add(Maxpooling(4, C1))$
$C2 \leftarrow Model.Add(Conv2D(24, M1)$
$M2 \leftarrow Model.Add(Maxpooling(2, C2))$
$F \leftarrow Model.Add(Flatten(M2))$
$\mu, \sigma, SpLat \leftarrow Encode(Inputs)$
$M_1 \leftarrow dim(SpLat)$
$den \leftarrow Model.Add(Dense(M_1, F))$
$SVAE \leftarrow Model(Inputs, [\mu, \sigma, SpLat], name = SEncoder)$
Return $SpLat$

Algorithm 4: Multimodal integration

Input: TemLat, SpLat
Output: Merged Latent

$MergedLatent \leftarrow Concatenate(TemLat, SpLat)$
Return $MergedLatent$

2.3. Cognitive workload estimation using VBGMM

The concatenated latent feature vector (output of Algorithm 4) is passed to the VBGMM clustering for workload classification. In the Variational BGMM approach, approximate posterior distributions can be effectively determined using the variational inference algorithm while retaining the advantages of the Bayesian approach [41]. Since the probabilistic variables of each category (c) are statistically independent, the approximate posterior distribution of the VBGMM model can be expressed as:

$$p(\theta, Z \mid X, m) = \prod_c p(\theta_c \mid X_c, m) p(Z_c \mid X_c, m) \tag{1}$$

where, Z is the hidden variable of model m. $p(Z \mid X, m)$ is the approximate posterior distribution of model parameters, and $p(\theta \mid X, m)$ refers to the solutions of VB posterior distribution. The VBGMM clustering method is related to two parameters: (a) "prior type" (Dirichlet process or Dirichlet distribution) and (b) "weight_concentration_prior", which refers to the distribution of weights to each of the components based on prior type [30]. The output components or clusters can be inferred using this "weight" parameter. Here, the weight parameter value is selected as $1e+2$ and Dirichlet distribution is chosen as the prior type of the model. This extra parameterization is necessary for variational inference, but for the prior type "Dirichlet process", the inference process may be slower.

Here, the performance evaluation of VBGMM clustering is performed using three metrics, namely unsupervised clustering accuracy (Acc), Normalized Mutual Information (NMI), and Rand Index (RI). Acc (2) refers to the best matching between cluster assignments from the clustering method (t_i) and ground truth labels (y_i).

$$Acc = \max_m \frac{\sum_{i=1}^{n} 1\left\{ y_i = m(t_i) \right\}}{n} \tag{2}$$

where n is the number of samples and m ranges overall possible one-to-one mappings between clusters and labels. NMI (3) refers to the reduction of entropy information for cluster assignments with respect to the known ground truth labels.

$$NMI(y, m) = \frac{I(y, m)}{\frac{1}{2}[H(y) + H(m)]} \tag{3}$$

where the ground truth labels and cluster assignment are denoted by y and m, respectively. Mutual information between y and m is represented by I, and entropy is denoted by $H(.)$. The Rand Index (RI) checks the similarity between original and predicted class labels, represented as follows:

$$RI = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

where, TP, TN, FP, and FN signify true positive, true negative, false positive and false-negative rates, respectively.

3. Results

This section is divided into five subsections: (A) behavioral results, (B) integrated spatio-temporal VAE model analysis, (C) clustering results & performance analysis, (D) computational complexity analysis, and (E) comparative analysis. The detailed analysis of each section is mentioned below.

3.1. Behavioral results

As the task complexity level is related to the value of n (n = 2, 3) in the n-back task, the subject needs to remember more numbers of previous letters with the higher value of n. The Friedman test has been performed to check whether the means of each workload level are equal or not. The result ($p < 0.05$, $\chi^2 = 32.077$, $df = 30$) indicates that a significant difference exists in mean values of different workload levels. After the Friedman test, the Dunn–Bonferroni Posthoc test was conducted to determine the difference between each pair of workload levels. The Posthoc test (Table 1) validates that the mean differences between each pair of workload levels are statistically significant (p value < 0.05).

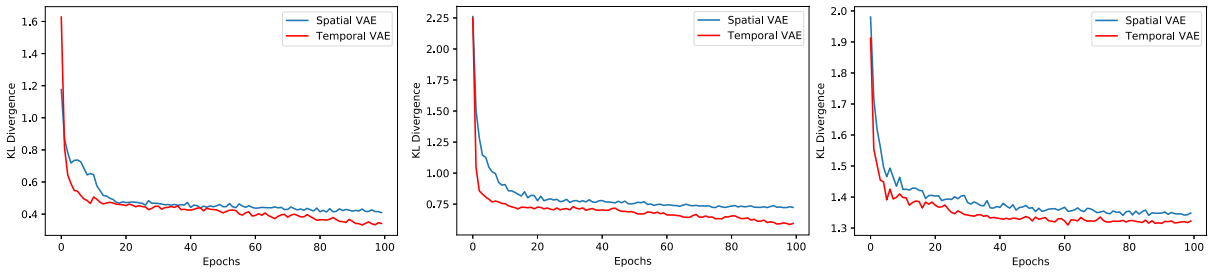


Fig. 2. Results of VAE training procedure with different combinations of spatial (Algorithm 3) and temporal VAE (Algorithm 2) models for best latent dimension=8 (refer to Fig. 4(a)). (a) Model 1 (proposed) configuration: temporal VAE → L6-L12-L18, spatial VAE → C16(3,3)-C24(5,5), (b) Model 2 configuration: temporal VAE → L24-L48, spatial VAE → C16(3,3)-C16(5,5)-C32(7,7) and (c) Model 3 configuration: temporal VAE → L12-L24-L36-L48, spatial VAE → C8(3,3)-C16(5,5)-C24(7,7)-C36(7,7). Here, the convolutional layer and the LSTM cells are denoted by C$x(k,k)$ and L, respectively. x and k are represented as the filter number and kernel size.

Table 1
Result of Dunn–Bonferroni Posthoc test.

Sample 1–Sample 2	Mean difference	Std. error	p value
2-back-0-back	1.056	0.297	0.007
3-back-2-back	1.114	0.321	0.011
3-back-0-back	1.127	0.336	0.012

3.2. Integrated spatio-temporal VAE model analysis

This section is further divided into three subsections: (1) selection of the VAE model & Spatial feature analysis, (2) effects of the latent dimension of VAEs, and (3) effects of optimizer and learning rate. The detailed discussions of each subsection are mentioned below.

3.2.1. Selection of the VAE model & spatial feature analysis

The construction of deep VAE models is illustrated in Algorithm 2 and Algorithm 3. Both algorithms are used only encoder parts of VAEs. Two types of losses are associated during the training process of VAE: reconstruction loss and Kullback–Leibler (KL) divergence. However, as the proposed model only uses the encoder's output, only KL divergence loss is considered. The KL divergence tends to regularize the organization of the latent space by making the distributions returned by the encoder close to a standard normal distribution. Hence, the objective is to select the optimal VAE model such that the KL divergence between the distribution of the encoder and input will be minimized. KL divergence values of different temporal and spatial VAE models are shown in Fig. 2. Different configurations of temporal and spatial VAE models are constructed using Algorithm 2 and Algorithm 3 by changing the layer-wise LSTM units or layer-wise filters of the CNN model. The training losses (KL divergence) of different model combinations (temporal-spatial VAE) are shown in Fig. 2. Since the first temporal and spatial VAE model combination (Model 1: Fig. 2(a)) incurs the lowest KL divergence loss, this configuration is selected for the proposed IST-VAE model. The temporal VAE of the proposed model is already analyzed with different combinations of LSTM layers (Model 1: 3 layers, Model 2: 2 layers, and Model 3: 4 layers) in Fig. 2. However, for generalization purposes, the temporal VAE is also experimented with one-layer LSTM model (L128) and another two-layered (L12-L18) LSTM model. Here, L represents the number of LSTM units. For the two differently configured temporal VAEs, the configuration of the best spatial VAE (in Model 1 of Fig. 2) is not changed while computing the combined latent features. For high workload condition (i.e., cond 4), the IST-VAE model achieves the mean clustering accuracy of 0.526 and 0.583, respectively, using these above-mentioned single and two-layered LSTM models, which are much lesser than the performance of three-layered LSTM model. Thus, the temporal VAE with three-layered LSTM (Model 1 in Fig. 2) is selected for the proposed IST-VAE. Each deep model is trained on 100 epochs. During training, the Adam optimizer (learning rate of 0.02) is used. A batch size of 32 was chosen during the training of the IST-VAE model. The hyperparameters (learning rate, batch size, number of epochs, number of hidden layers, number of neurons/layers etc.) of the IST-VAE model are tuned using the Random search method [42]. Like Grid search, the Random search does not exhaustively examine all conceivable parameter combinations; instead, it selectively draws hyperparameter values from predefined distributions in a random manner. Thus, this approach discovers effective hyperparameters quicker than the exhaustive grid search method.

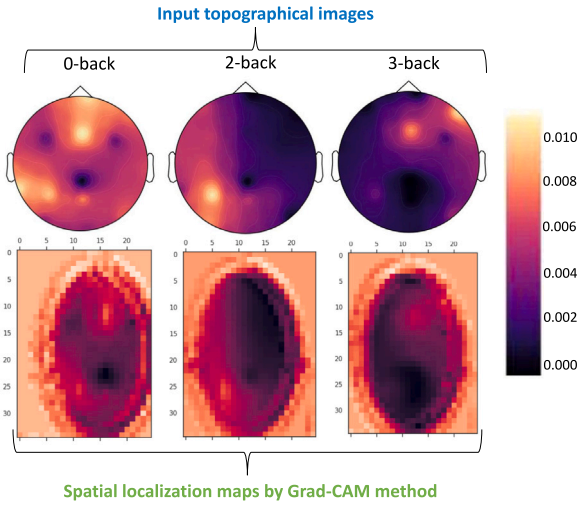


Fig. 3. Spatial localization maps of input topographical images using the Grad-CAM method. The localization feature maps are obtained by the Grad-CAM method from the last convolution layer of the proposed spatial VAE model. In the color bar, the brighter color represents the workload class-specific informative brain region.

Gradient-Class Activation Mapping (Grad-CAM) visualization technique utilizes gradients from the final convolutional layer to generate a coarse localization map, indicating the spatial areas where electrodes are positioned [43]. The feature map obtained by Grad-CAM from the last convolution layer contains high-level features and spatial information [43]. Thus, spatial localization maps extracted from the last convolution layer of the proposed spatial VAE model highlight the spatial information of scalp topographical images. Three randomly selected topographical images of each workload level and their spatial localization feature maps (obtained from Grad-CAM) are shown in Fig. 3. The brighter color in the localized map indicates the activated/informative region in the brain for the workload class. For the 0-back class, high activation is observed in the right frontal and left parietal brain lobes, whereas a high activation is observed only in the left parietal brain lobe for the 2-back task. Only the right frontal brain area is activated for the complex task (3-back).

Table 2
Results of the LSO experiment of the proposed model based on best latent feature vector (refer to Fig. 4(a)). Note: Accuracy (Acc), Normalized mutual information (NMI), and Rand index (RI). Workload conditions are: Cond 1: 0-back vs. 2-back, Cond 2: 0-back vs. 3-back, Cond 3: 2-back vs. 3-back, Cond 4: 0-back vs. 2-back vs 3-back. Best clustering accuracy (Acc) is marked with bold.

Subjects	Cond 1			Cond 2			Cond 3			Cond 4		
	ACC	NMI	RI	ACC	NMI	RI	ACC	NMI	RI	ACC	NMI	RI
Sub 1	0.988	0.869	0.882	0.771	0.803	0.643	0.886	0.901	0.832	0.792	0.722	0.801
Sub 2	0.965	0.802	0.723	0.602	0.556	0.627	0.785	0.568	0.622	0.707	0.779	0.639
Sub 3	0.989	0.823	0.798	0.781	0.765	0.711	0.987	0.876	0.933	0.658	0.593	0.689
Sub 4	0.981	0.844	0.811	0.689	0.623	0.598	0.773	0.782	0.675	0.734	0.767	0.597
Sub 5	**0.992**	0.881	0.827	0.708	0.646	0.602	0.926	0.891	0.912	0.845	0.723	0.698
Sub 6	0.979	0.913	0.933	0.856	0.798	0.766	0.978	0.821	0.856	0.798	0.767	0.803
Sub 7	0.987	0.893	0.922	0.932	0.898	0.853	0.836	0.804	0.866	0.933	0.887	0.813
Sub 8	0.952	0.833	0.721	0.779	0.736	0.799	0.988	0.786	0.877	0.784	0.763	0.722
Sub 9	0.985	0.923	0.907	0.658	0.558	0.561	0.982	0.996	0.893	0.813	0.724	0.766
Sub 10	0.978	0.883	0.743	0.723	0.593	0.634	0.923	0.866	0.794	0.834	0.798	0.771
Sub 11	0.981	0.923	0.899	0.738	0.792	0.635	0.993	0.962	0.902	0.592	0.632	0.602
Sub 12	0.982	0.995	0.949	0.689	0.568	0.637	0.985	0.936	0.903	0.906	0.853	0.778
Sub 13	0.977	0.889	0.836	0.934	0.884	0.865	0.813	0.723	0.668	0.756	0.623	0.644
Sub 14	0.974	0.936	0.811	0.941	0.742	0.833	0.998	0.922	0.798	0.823	0.846	0.755
Sub 15	0.978	0.973	0.944	0.863	0.722	0.744	0.906	0.789	0.802	0.956	0.845	0.811
Sub 16	0.985	0.962	0.884	0.778	0.652	0.689	0.882	0.797	0.791	0.689	0.65	0.627
Sub 17	0.984	0.923	0.889	0.924	0.881	0.857	0.592	0.623	0.556	0.592	0.56	0.522
Sub 18	0.972	0.856	0.923	0.899	0.865	0.798	0.658	0.612	0.605	0.782	0.705	0.732
Sub 19	0.986	0.902	0.823	0.952	0.872	0.822	0.856	0.798	0.831	0.785	0.685	0.653
Sub 20	0.985	0.853	0.771	0.851	0.782	0.749	0.962	0.889	0.901	0.807	0.748	0.656
Sub 21	0.968	0.828	0.745	0.798	0.712	0.698	0.946	0.902	0.879	0.822	0.765	0.689
Sub 22	0.991	0.923	0.887	0.923	0.836	0.789	0.956	0.872	0.778	0.736	0.638	0.611
Sub 23	0.986	0.905	0.899	0.911	0.844	0.792	0.962	0.892	0.762	0.802	0.736	0.633
Sub 24	0.977	0.825	0.786	0.836	0.768	0.812	0.933	0.862	0.736	0.845	0.766	0.744
Sub 25	0.983	0.893	0.843	0.893	0.824	0.872	0.955	0.923	0.879	0.889	0.798	0.765
Sub 26	0.979	0.901	0.916	0.928	0.798	0.812	0.936	0.868	0.902	0.836	0.813	0.796
Mean	**0.980**	**0.890**	**0.848**	**0.821**	**0.750**	**0.738**	**0.899**	**0.833**	**0.805**	**0.789**	**0.737**	**0.704**

3.2.2. Effects of latent dimension of VAEs

The image and EEG signal dimensions of the proposed IST-VAE model are M_1: $80 \times 60 \times 3 = 14\,400$ and $M_2 = 30$ (no. of EEG channels), respectively. As the outputs of two latent features (from temporal and spatial VAEs) are merged into an embedded feature space, thus, the resultant latent dimension (M) in the embedded feature space should be lower than M_2 (as $M_2 < M_1$). For simplicity, the experiment is analyzed between 0-back vs. 2-back workload levels. The proposed model is trained for all subjects using the leave-subject-out (LSO) test across different latent dimensions. In Fig. 4(a), the average clustering performance is evaluated based on the multimodal latent feature (output of Algorithm 4) of the IST-VAE model with different latent dimensions (by changing the different *eps* value of Algorithm 1). It can be noted that the maximum clustering performance is achieved with $M = 8$. Clustering performance varies with different latent dimensions of Temporal and Spatial VAEs (variable M_2 in Algorithm 2 and M_1 in Algorithm 3). The larger value of M may preserve more information but simultaneously minimize the discrimination power of learned representation, which decreases the clustering performance [44]. The same latent dimension (i.e., $M = 8$) is maintained for other workload conditions.

3.2.3. Effects of optimizer and learning rate

In this section, hyperparameters of the proposed IST-VAE model are tuned to obtain the best clustering performance. The convergence of deep neural networks largely depends on selecting the proper optimizer and learning rate (lr). The VAE model is trained using the leave-subject-out (LSO) method to avoid the overfitting issue. The proposed VAE model is trained with different optimizers and its learning rate (lr). Next, the clustering accuracy is computed for each subject of the test set. The process is repeated for all the subjects. The mean clustering accuracy of different optimizers (adam, stochastic gradient descent: SGD, adadelta, and RMSprop) and their learning rates lr (0.01 to 0.1) is reported in Fig. 4(b). It can be noted that the Adam optimizer with the lr of 0.02 achieves the maximum clustering accuracy. So, the Adam optimizer with the same lr is used for other workload conditions.

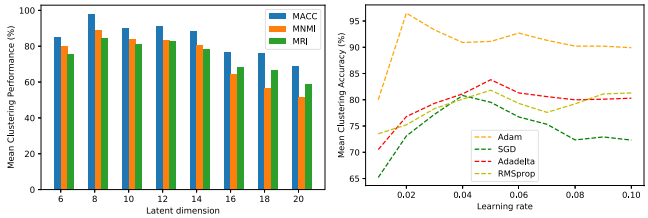


Fig. 4. Evaluation of clustering performance: (a) based on different latent dimensions, (b) based on different optimizers and their learning rates for workload condition 0-back vs. 2-back. Note: Mean clustering accuracy (MACC), Mean Normalized Mutual Information (MNMI), and Mean Rand Index (MRI).

However, the experiment result of the 0-back vs. 2-back workload condition is presented for simplicity.

3.3. Clustering results & performance analysis

This section is divided into two subsections: (1) Output analysis of the VBGMM clustering method and (2) performance analysis of VBGMM based on other datasets. A detailed description of each section is mentioned below.

3.3.1. Output analysis of the VBGMM clustering method

This section highlights the output analysis of the VBGMM clustering method and subject-wise clustering results of different workload levels. The merged latent feature (i.e., the combination of temporal and spatial latent features) created in Algorithm 4 is used as input for clustering. The LSO experiment (the training and testing set consists of entirely different subjects) is performed to evaluate the subject-wise variation across various workload levels. In the LSO experiment, the EEG signals and images of scalp topographical maps of 25 subjects are used as a training set. The testing set comprised the remaining subject's topographical maps and EEG data. After the LSO experiment,

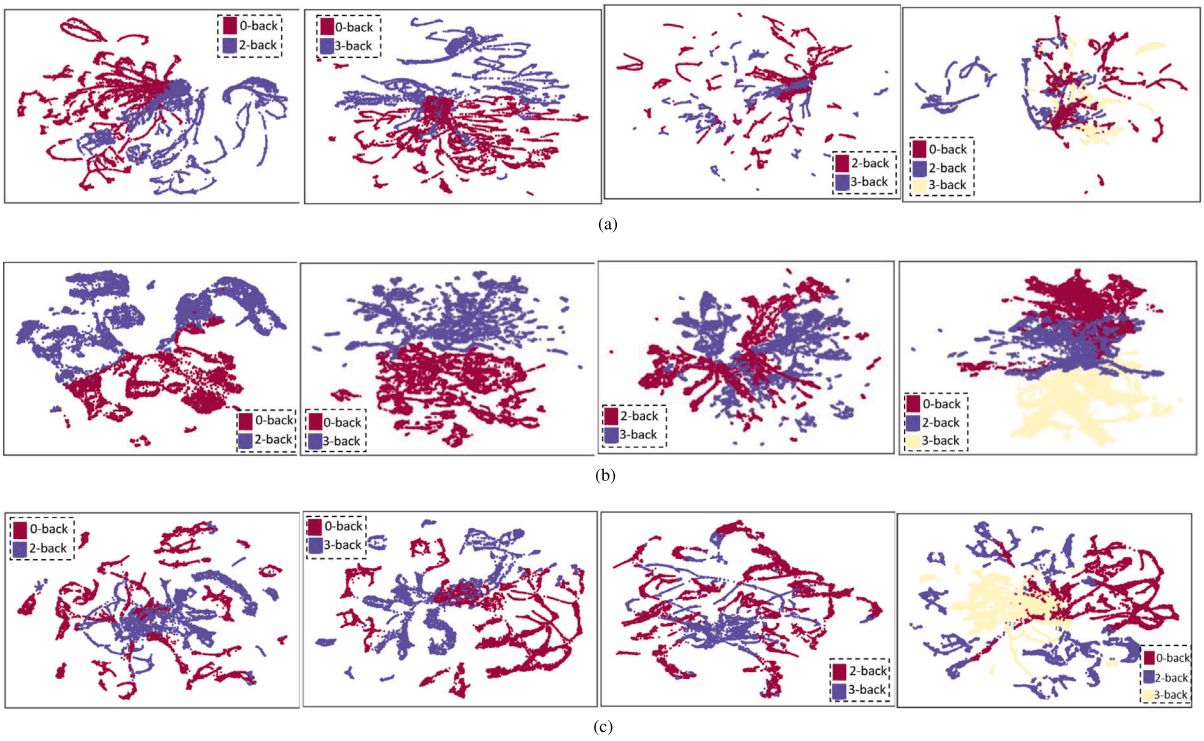


Fig. 5. VBGMM cluster representation (using UMAP plot) of S1 for different workload levels and latent dimensions: (a) latent dimension = 6, (b) latent dimension = 8, and (c) latent dimension = 10. The ordering of workload conditions (for all figures) are as follows: 0-back vs. 2-back, 0-back vs. 3-back, 2-back vs. 3-back, and all the workload levels (0-back vs. 2-back vs. 3-back). The parameters of the VBGMM method are as follows: weight concentration prior = 1e−1, weight concentration prior type = Dirichlet process, and maximum iteration to converge = 100.

the trained latent feature vector of VAE is used for the training of VBGMM, and the predicted latent feature vector is used to evaluate clustering performance. The training parameters of VBGMM are as follows: maximum iterations = 150, weight_concentration_prior = 0.01, weight_concentration_priortype = Dirichlet process, covariance_type =full, and mean_precision_prior = 0.1. The LSO experiment is performed for all subjects with different train-test sets, and the result is shown in Table 2. The average clustering performance result (MACC and MNMI) based on the latent features generated from the proposed IST-VAE (output of Algorithm 4) and two unimodal VAE models' (outputs of Algorithm 2 and Algorithm 3) is displayed in Table 3. The result shows that the proposed model achieves the best clustering performance regardless of workload conditions. The best mean clustering accuracy for the 0 vs. 2-back condition for the VBGMM clustering method is 98.0%, whereas the subject-wise best clustering accuracy for the subject S5 is 99.2% (refer to Table 2).

In manifold learning, the high-dimensional data can be represented in low-dimensional space for better visualization. Through the non-linear mapping operation, uniform manifold approximation and projection (UMAP) transforms higher-dimensional data into a lower-dimensional manifold with the fastest execution time [45]. The UMAP cluster plots (Fig. 5) of the VBGMM clustering method are shown for three different latent dimensions (6, 8, and 10) of the IST-VAE model. The different-sized latent dimensions are obtained by changing the *eps* variable in Algorithm 1. The clustering quality is highly dependent on the learned representation of data [21]. It can be noted that the cluster representation for latent dimension 8 is better than other latent dimensions (6 and 10). For latent dimension 8, clusters are clearly distinguished for the lowest workload level (0-back vs. 2-back in Fig. 5(b)), but with the more number of workload levels, some clusters get overlapped (0-back vs. 2-back vs. 3-back in Fig. 5(b)) for the complicated pattern of brain signals. For latent dimensions 6 (Fig. 5(a)), and 10 (Fig. 5(c)), the learned representation of the IST-VAE model is poor, resulting a scattered or overlapped clusters for all workload conditions.

Table 3
Performance analysis of the VBGMM clustering method using the latent features (output of Algorithm 4) of proposed IST-VAE and unimodal temporal/ spatial VAEs (outputs of Algorithm 2, Algorithm 3). For workload conditions, refer to Table 2. Note: Mean Accuracy (MAcc), Mean Normalized mutual information (MNMI).

Conditions	IST-VAE		Spatial VAE		Temporal VAE	
	MACC	MNMI	MACC	MNMI	MACC	MNMI
Cond 1	0.980	0.890	0.822	0.782	0.843	0.801
Cond 2	0.821	0.750	0.689	0.665	0.742	0.703
Cond 3	0.899	0.833	0.726	0.623	0.771	0.724
Cond 4	0.789	0.737	0.623	0.556	0.649	0.623

Table 4
VBGMM clustering result based on other *n*-back datasets. Idle state of [47] denotes 0-back state.

Dataset 1 [46]			Dataset 2 [47]		
Conditions	MACC	MNMI	Conditions	MACC	MNMI
0-1 back	0.923	0.878	0-1 back	0.911	0.802
0-2 back	0.917	0.852	0-2 back	0.802	0.723
1-2 back	0.825	0.773	1-2 back	0.769	0.703
0-1-2 back	0.703	0.683	0-1-2 back	0.733	0.705

3.3.2. Performance analysis of VBGMM based on other datasets

For generalization, the proposed model has been implemented to two publicly open-access *n*-back datasets [46,47]. Each open-access dataset has a different workload condition. The LSO experiment is performed for each dataset, and the VBGMM clustering result (Table 4) is produced based on the MACC and MNMI matrices. It can be observed that the proposed model achieves the best clustering performance (MACC: 0.923, MNMI: 0.878) for the workload level (0-back vs. 1-back) on the dataset [46]. Therefore, it can be stated that the proposed model is capable of accurately estimating different workload conditions across various datasets.

Table 5
Performance analysis of VBGMM using proposed multimodal spatio-temporal features and other latent feature estimation methods. Workload conditions are mentioned in Table 2. Note: AutoEncoder (AE), sparse autoencoder (SAE), Denoising AutoEncoder (DAE), Convolutional autoencoder (CAE), LSTM autoencoder (LAE), Proposed: IST-VAE.

Conditions	Mean clustering accuracy						
	PCA	AE	SAE	CAE	LAE	DAE	Proposed
Cond 1	0.475	0.536	0.663	0.709	0.732	0.829	0.980
Cond 2	0.458	0.507	0.612	0.673	0.741	0.765	0.821
Cond 3	0.512	0.529	0.589	0.625	0.656	0.633	0.899
Cond 4	0.388	0.428	0.546	0.612	0.536	0.707	0.789

Table 6
Comparison of the proposed VBGMM with different clustering techniques for all workload conditions (refer to Table 2).

Methods	MACC			
	Cond 1	Cond 2	Cond 3	Cond 4
IST-VAE + K-means	0.563	0.513	0.602	0.523
IST-VAE + DB-Scan	0.803	0.823	0.746	0.692
IST-VAE + Hierarchical	0.742	0.646	0.563	0.502
IST-VAE + Optics	0.708	0.801	0.731	0.712
IST-VAE + Birch	0.803	0.692	0.727	0.704
IST-VAE + GMM	0.789	0.724	0.743	0.733
IST-VAE + Spectral	0.713	0.687	0.609	0.538
IST-VAE + VBGMM	0.980	0.821	0.899	0.789

3.4. Computational complexity analysis

This section discusses the computational complexity of the proposed model. In the proposed model, the LSTM and CNN-based VAE models are used to extract temporal and spatial latent features from inputs. The time complexity of the single LSTM network is $O(tW)$, where t and W are the time step and weights, respectively [48]. Here, each weight is associated with a single node. The computational complexity of the CNN model is $O(N)$, where $N \times N$ is the input image size [49]. The concatenation of latent features (Algorithm 4) and the encoding operation (Algorithm 1) takes $O(1)$ time. The complexity of the VBGMM clustering algorithm is $O(sD^3)$, where s and D are the numbers of iterations and feature dimensions [50]. Therefore, the overall complexity of the proposed model is $O(tW + N + sD^3)$.

3.5. Comparison study

Several comparison studies are performed in this section, namely, (1) a comparison among deep models for cluster analysis, (2) a comparison among clustering methods, (3) a comparison among deep clustering methods, and (4) a comparative analysis among existing studies. The detailed discussions of each subsection are mentioned below.

3.5.1. Comparison among deep models for cluster analysis

Table 5 compares the VBGMM clustering results using the proposed multimodal latent feature and other latent feature estimation methods, such as PCA and deep latent features (extracted from different AE-based deep models). In Convolutional autoencoder (CAE) and LSTM autoencoder (LAE), clustering performance is evaluated using deep spatial and temporal latent features (i.e., outputs of the CNN and LSTM-based encoders), respectively. The best LSTM and CNN model configuration (refer to Fig. 2(a)) is used in LAE and CAE models. In the denoising autoencoder (DAE), noise is added to the input topographical image with the optimal noise factor of 0.8 [44]. Then, the reconstructed robust (noise-free) spatial feature (from the decoder of DAE) is used as a latent feature for clustering. For Sparse Autoencoder (SAE), an L2 regularizer was utilized with a weight decay factor of 0.003 [51]. The autoencoder model only consists of dense layers. The encoder part is only utilized for extracting latent features in all the autoencoder models except for the DAE. The configuration of all autoencoders are as follows: SAE/AE→ D(128)-D(32)-D(8), DAE → C16(3,3)-M(2)-C32(3,3)-M(2)-DC32(3,3)-U(2)-DC16(3,3)-U(2)-DC8(3,3), CAE→ C16(3,3)-M(4)-C24(5,5)-M(2)-F-D(8), LAE→ L(6)-L(12)-L(18)-D(8), where C$x(k,k)$: convolutional layer with x filters (k = filter size), DC: deconvolution, $M(n)$: max-pooling (width), F: flatten, D: dense, U: upsampling, L: LSTM layer. The CAE, DAE, SAE, and AE used the topographical images, whereas LAE used the preprocessed EEG signal for training. It can be identified that the proposed model achieves the best clustering results among all the AE-based models across all the workload conditions.

Table 7
Comparison of the proposed model with existing deep clustering techniques for all workload conditions (refer to Table 2). Note: Semantic Clustering by Adopting Nearest neighbors (SCAN), Not too Deep (N2d) clustering, Deep Embedded Clustering (DEC), and Discriminatively Boosted Clustering (DBC)

Deep clustering methods	MACC			
	Cond 1	Cond 2	Cond 3	Cond 4
N2D [52]	0.712	0.663	0.623	0.573
SCAN [53]	0.845	0.768	0.757	0.702
DEC [34]	0.782	0.745	0.702	0.663
DBC [54]	0.801	0.773	0.745	0.712
Proposed method	0.980	0.821	0.899	0.789

3.5.2. Comparison among clustering methods

This section highlights the effectiveness of the proposed VBGMM clustering methods over other clustering methods for all the workload conditions. Here, all the clustering methods are evaluated based on latent features (i.e., the output of Algorithm 4) extracted from the proposed IST-VAE model, and the comparison (Table 6) is performed based on the average clustering accuracy. It can be shown that VBGMM outperforms all other clustering algorithms. VBGMM outperforms density or partition-based clustering techniques because of the overlapping distribution of the input data.

3.5.3. Comparison among deep clustering methods

This section highlights the comparison between the proposed multimodal deep clustering method and some popular deep clustering methods for the experimental dataset. The spatial topographical images are only used in the image-based deep clustering methods. The comparison is shown in Table 7. As the proposed model captures both the temporal and spatial features for clustering, it enhances the cluster quality. For all of the experimental conditions, the proposed deep clustering performs better than other deep clustering methods.

3.5.4. Comparative analysis among existing studies

In this section, the comparison is performed between the proposed model and the studies that used the same experimental dataset [36]. The comparison (refer to Table 8) is performed based on the methods and results of those studies. It can be concluded that the low-dimensional latent feature performs better than the traditional EEG features (PSD, ERD) for the same experimental dataset. Thus, the proposed model achieves better accuracy than other models.

4. Discussion

This study combines a multimodal deep VAE model and the VBGMM clustering method for workload classification. In this paper, the VBGMM-based probabilistic clustering method is employed since it emerges as a suitable choice for addressing the stochastic nature of human responses [59]. The performance of a proposed IST-VAE model largely depends on the VAE model structure and hyperparameters. The proposed multimodal deep clustering is evaluated with the three different workload conditions with varying degrees of task complexity. The proposed model achieves 98.0% and 99.2% as the best mean

Table 8
Comparative study among the proposed method and other studies based on the same experimental dataset. Note: ERD: event-related desynchronization, ERS: event-related synchronization, DWT: discrete wavelet transform, DNN: Deep Neural Network, FBC: Functional Brain Connectivity, The highest accuracy of the proposed models is shown in the "Acc" column.

Study	Methods	Acc(%)
Saadati et al. [55]	ERD/ERS + Hybrid DNN	0.74
Saadati et al. [56]	ERD/ERS + Hybrid CNN	0.69
Khanam et al. [57]	DWT features and SVM	0.87
Cao et al. [58]	PSD, FBC + SVM	0.77
Proposed method	Temporal-spatial VAE models +VBGMM	0.98

clustering accuracy and subject-wise clustering accuracy, respectively (refer to Table 2). The performance of the proposed deep IST-VAE framework also outperforms unimodal-based VAE models (Table 3) and other AE-based deep models (Table 5). The result indicates that the fusion of low dimensional temporal and spatial EEG features can effectively identify the workload levels of the *n*-back task. For generalization purposes, the proposed model is examined on two open-access *n*-back datasets (refer to Table 4), and promising results are found for both datasets. According to comparison studies, the proposed VBGMM clustering approach performs better than traditional well-known clustering methods (refer to Table 6). The proposed multimodal-based clustering method can efficiently classify the different workload levels. Thus, the proposed model can be effectively used to identify human behavior while accomplishing EEG-based cognitive tasks.

5. Conclusion & future work

In this paper, two modalities (temporal and spatial) of EEG are fused to estimate the workload levels of a participant efficiently. Four distinctive algorithms (Encode, Temporal VAE, Spatial VAE, and Multimodal integration) are developed and used to build the multimodal workload estimation framework. The latent feature creation process for the temporal and spatial encoders is performed by Algorithm 1. The selection of a proper latent feature dimension is of utmost importance in DRL-based clustering applications, as changing the latent feature's dimension can affect the clustering performance. The clustering performance with different latent dimensions is plotted in Fig. 4(a). In Algorithm 2 and Algorithm 3, both temporal and spatial encoders (using deep VAE) extract the layer-wise temporal/spatial meaningful information for classification. Moreover, the LSTM and CNN-based VAEs used in Algorithm 2 and Algorithm 3 are applicable for sequence learning contexts with long-term dependency problems [17]. As the extracted low-dimensional latent feature contains significant information related to the problem context, merging both of these feature vectors into a common embedded feature vector highlights more relevant information than the uni-modal approach. As the clustering efficiency especially depends on learned feature representation [21], thus, the VBGMM clustering method can efficiently estimate workload-specific clusters using the merged latent feature (output of Algorithm 4) having the temporal and spatial knowledge of input data.

The experimental dataset consists of the EEG and NIRS data of 26 subjects; however, the experiment is conducted only on EEG signals. In the near future, sensor-level multimodal capabilities (EEG and NIRS) can be implemented in the proposed model to enhance its performance.

CRediT authorship contribution statement

Debashis Das Chakladar: Conceptualization, Methodology, Data curation, Software, Writing – original draft. **Partha Pratim Roy:** Writing – review & editing, Supervision. **Victor Chang:** Resources, Writing – review & editing, Supervision.

Declaration of competing interest

This is hereby certify that the paper is original, neither the paper nor a part of it is under consideration for publication anywhere else. Also, we have no conflicts of interest to disclose.

Data availability

The authors do not have permission to share data.

Acknowledgment

Part of Prof Chang's work is partly supported by VC Research, UK (VCR0000221). We confirm that the manuscript has been read and approved by all named authors.

References

[1] F.G. Paas, J.J. Van Merriënboer, Instructional control of cognitive load in the training of complex cognitive tasks, Educ. Psychol. Rev. 6 (1994) 351–371.
[2] P. Antonenko, F. Paas, R. Grabner, T. Van Gog, Using electroencephalography to measure cognitive load, Educ. Psychol. Rev. 22 (4) (2010) 425–438.
[3] D. Panda, D.D. Chakladar, 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.
[4] D.D. Chakladar, S. Chakraborty, EEG based emotion classification using "correlation based subset selection", Biol. Inspired Cogn. Archit. 24 (2018) 98–106.
[5] D.D. Chakladar, P.P. Roy, M. Iwamura, EEG-based cognitive state classification and analysis of brain dynamics using deep ensemble model and graphical brain network, IEEE Trans. Cogn. Dev. Syst. (2021).
[6] D.D. Chakladar, D. Samanta, P.P. Roy, Multimodal deep sparse subspace clustering for multiple stimuli-based cognitive task, in: 2022 26th International Conference on Pattern Recognition, ICPR, IEEE, 2022, pp. 1098–1104.
[7] D.D. Chakladar, S. Dey, P.P. Roy, D.P. Dogra, EEG-based mental workload estimation using deep BLSTM-LSTM network and evolutionary algorithm, Biomed. Signal Process. Control 60 (2020) 101989.
[8] R. Chai, G.R. Naik, T.N. Nguyen, S.H. Ling, Y. Tran, A. Craig, H.T. Nguyen, Driver fatigue classification with independent component by entropy rate bound minimization analysis in an EEG-based system, IEEE J. Biomed. Health Inf. 21 (3) (2016) 715–724.
[9] A.-M. Brouwer, M.A. Hogervorst, J.B. Van Erp, T. Heffelaar, P.H. Zimmerman, R. Oostenveld, Estimating workload using EEG spectral power and ERPs in the n-back task, J. Neural Eng. 9 (4) (2012) 045008.
[10] H. Yu, X. Lei, Z. Song, C. Liu, J. Wang, Supervised network-based fuzzy learning of EEG signals for Alzheimer's disease identification, IEEE Trans. Fuzzy Syst. 28 (1) (2019) 60–71.
[11] H. Yu, X. Wu, L. Cai, B. Deng, J. Wang, Modulation of spectral power and functional connectivity in human brain by acupuncture stimulation, IEEE Trans. Neural Syst. Rehabil. Eng. 26 (5) (2018) 977–986.
[12] H. Yu, X. Li, X. Lei, J. Wang, Modulation effect of acupuncture on functional brain networks and classification of its manipulation with EEG signals, IEEE Trans. Neural Syst. Rehabil. Eng. 27 (10) (2019) 1973–1984.
[13] A. Supratak, H. Dong, C. Wu, Y. Guo, DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG, IEEE Trans. Neural Syst. Rehabil. Eng. 25 (11) (2017) 1998–2008.
[14] M.A. Hogervorst, A.-M. Brouwer, J.B. Van Erp, Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload, Front. Neurosci. 8 (2014) 322.
[15] D.D. Chakladar, P. Kumar, P.P. Roy, D.P. Dogra, E. Scheme, V. Chang, A multimodal-Siamese Neural Network (mSNN) for person verification using signatures and EEG, Inf. Fusion 71 (2021) 17–27.
[16] Y. Tomita, F.-B. Vialatte, G. Dreyfus, Y. Mitsukura, H. Bakardjian, A. Cichocki, Bimodal BCI using simultaneously NIRS and EEG, IEEE Trans. Biomed. Eng. 61 (4) (2014) 1274–1284.
[17] P. Zhang, X. Wang, J. Chen, W. You, W. Zhang, Spectral and temporal feature learning with two-stream neural networks for mental workload assessment, IEEE Trans. Neural Syst. Rehabil. Eng. 27 (6) (2019) 1149–1159.
[18] L. Wu, Q. Zhao, J. Liu, H. Yu, Efficient identification of Alzheimer's brain dynamics with spatial-temporal autoencoder: A deep learning approach for diagnosing brain disorders, Biomed. Signal Process. Control 86 (2023) 104917.
[19] P. Zhang, X. Wang, J. Chen, W. You, Feature weight driven interactive mutual information modeling for heterogeneous bio-signal fusion to estimate mental workload, Sensors 17 (10) (2017) 2315.
[20] C.-T. Lin, J.-T. King, C.-H. Chuang, W. Ding, W.-Y. Chuang, L.-D. Liao, Y.-K. Wang, Exploring the brain responses to driving fatigue through simultaneous EEG and fNIRS measurements, Int. J. Neural Syst. 30 (01) (2020) 1950018.

[21] M.R. Karim, O. Beyan, A. Zappa, I.G. Costa, D. Rebholz-Schuhmann, M. Cochez, S. Decker, Deep learning-based clustering approaches for bioinformatics, Brief. Bioinform. 22 (1) (2021) 393–415.
[22] E. Min, X. Guo, Q. Liu, G. Zhang, J. Cui, J. Long, A survey of clustering with deep learning: From the perspective of network architecture, IEEE Access 6 (2018) 39501–39514.
[23] R.G. Hefron, B.J. Borghetti, J.C. Christensen, C.M.S. Kabban, Deep long short-term memory structures model temporal dependencies improving cognitive workload estimation, Pattern Recognit. Lett. 94 (2017) 96–104.
[24] Y. Yuan, G. Xun, Q. Suo, K. Jia, A. Zhang, Wave2vec: Deep representation learning for clinical temporal data, Neurocomputing 324 (2019) 31–42.
[25] T. Behrouzi, D. Hatzinakos, Graph variational auto-encoder for deriving EEG-based graph embedding, Pattern Recognit. 121 (2022) 108202.
[26] J.F. Hwaidi, T.M. Chen, A novel KOSFS feature selection algorithm for EEG signals, in: IEEE EUROCON 2021-19th International Conference on Smart Technologies, IEEE, 2021, pp. 265–268.
[27] W.-N. Hsu, Y. Zhang, J. Glass, Unsupervised domain adaptation for robust speech recognition via variational autoencoder-based data augmentation, in: 2017 IEEE Automatic Speech Recognition and Understanding Workshop, ASRU, IEEE, 2017, pp. 16–23.
[28] X. Li, Z. Zhao, D. Song, Y. Zhang, J. Pan, L. Wu, J. Huo, C. Niu, D. Wang, Latent factor decoding of multi-channel EEG for emotion recognition through autoencoder-like neural networks, Front. Neurosci. 14 (2020) 87.
[29] M. Dai, D. Zheng, R. Na, S. Wang, S. Zhang, EEG classification of motor imagery using a novel deep learning framework, Sensors 19 (3) (2019) 551.
[30] D.M. Blei, M.I. Jordan, et al., Variational inference for Dirichlet process mixtures, Bayesian Anal. 1 (1) (2006) 121–143.
[31] A.S. Zandi, R. Tafreshi, M. Javidan, G.A. Dumont, Predicting epileptic seizures in scalp EEG based on a variational Bayesian Gaussian mixture model of zero-crossing intervals, IEEE Trans. Biomed. Eng. 60 (5) (2013) 1401–1413.
[32] X. Zhao, Y. Dong, Variational bayesian joint factor analysis models for speaker verification, IEEE Trans. Audio Speech Lang. Process. 20 (3) (2011) 1032–1042.
[33] S. Watanabe, Bayesian approaches in speech recognition, 2011, APSIPA NTT Communication Science Laboratories, NTT Corporation.
[34] J. Xie, R. Girshick, A. Farhadi, Unsupervised deep embedding for clustering analysis, in: International Conference on Machine Learning, PMLR, 2016, pp. 478–487.
[35] K. Ghasedi Dizaji, A. Herandi, C. Deng, W. Cai, H. Huang, Deep clustering via joint convolutional autoencoder embedding and relative entropy minimization, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 5736–5745.
[36] J. Shin, A. Von Lühmann, D.-W. Kim, J. Mehnert, H.-J. Hwang, K.-R. Müller, Simultaneous acquisition of EEG and NIRS during cognitive tasks for an open access dataset, Sci. Data 5 (2018) 180003.
[37] D.D. Chakladar, S. Datta, P.P. Roy, A. Vinod, Cognitive workload estimation using variational auto encoder & attention-based deep model, IEEE Trans. Cogn. Dev. Syst. (2022).
[38] M.X. Cohen, Analyzing Neural Time Series Data: Theory and Practice, MIT Press, 2014.
[39] D.D. Chakladar, S. Dey, P.P. Roy, M. Iwamura, EEG-based cognitive state assessment using deep ensemble model and filter bank common spatial pattern, in: 2020 25th International Conference on Pattern Recognition, ICPR, IEEE, 2021, pp. 4107–4114.
[40] R.N. Roy, S. Charbonnier, A. Campagne, S. Bonnet, Efficient mental workload estimation using task-independent EEG features, J. Neural Eng. 13 (2) (2016) 026019.
[41] S. Watanabe, Y. Minami, A. Nakamura, N. Ueda, Variational Bayesian estimation and clustering for speech recognition, IEEE Trans. Speech Audio Process. 12 (4) (2004) 365–381.
[42] J. Bergstra, Y. Bengio, Random search for hyper-parameter optimization, J. Mach. Learn. Res. 13 (2) (2012).
[43] Y. Li, H. Yang, J. Li, D. Chen, M. Du, EEG-based intention recognition with deep recurrent-convolution neural network: Performance and channel selection by grad-CAM, Neurocomputing 415 (2020) 225–233.
[44] X. Peng, H. Zhu, J. Feng, C. Shen, H. Zhang, J.T. Zhou, Deep clustering with sample-assignment invariance prior, IEEE Trans. Neural Netw. Learn. Syst. 31 (11) (2019) 4857–4868.
[45] E. Becht, L. McInnes, J. Healy, C.-A. Dutertre, I.W. Kwok, L.G. Ng, F. Ginhoux, E.W. Newell, Dimensionality reduction for visualizing single-cell data using UMAP, Nature Biotechnol. 37 (1) (2019) 38–44.
[46] A.J. Ries, J. Touryan, B. Ahrens, P. Connolly, The impact of task demands on fixation-related brain potentials during guided search, PLoS One 11 (6) (2016) e0157260.
[47] B.S. Cheema, S. Samima, M. Sarma, D. Samanta, Mental workload estimation from EEG signals using machine learning algorithms, in: International Conference on Engineering Psychology and Cognitive Ergonomics, 2018, pp. 265–284.
[48] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[49] O. Alaca, S. Althunibat, S. Yarkan, S.L. Miller, K.A. Qaraqe, CNN-based signal detector for IM-OFDMA, in: 2021 IEEE Global Communications Conference, GLOBECOM, IEEE, 2021, pp. 01–06.
[50] M. Cai, Y. Shi, J. Liu, J.P. Niyoyita, H. Jahanshahi, A.A. Aly, DRKPCA-VBGMM: Fault monitoring via dynamically-recursive kernel principal component analysis with variational Bayesian Gaussian mixture model, J. Intell. Manuf. 34 (6) (2023) 2625–2653.
[51] E. Hosseini-Asl, J.M. Zurada, O. Nasraoui, Deep learning of part-based representation of data using sparse autoencoders with nonnegativity constraints, IEEE Trans. Neural Netw. Learn. Syst. 27 (12) (2015) 2486–2498.
[52] R. McConville, R. Santos-Rodriguez, R.J. Piechocki, I. Craddock, N2d:(not too) deep clustering via clustering the local manifold of an autoencoded embedding, in: 2020 25th International Conference on Pattern Recognition, ICPR, IEEE, 2021, pp. 5145–5152.
[53] W. Van Gansbeke, S. Vandenhende, S. Georgoulis, M. Proesmans, L. Van Gool, Scan: Learning to classify images without labels, in: European Conference on Computer Vision, Springer, 2020, pp. 268–285.
[54] F. Li, H. Qiao, B. Zhang, Discriminatively boosted image clustering with fully convolutional auto-encoders, Pattern Recognit. 83 (2018) 161–173.
[55] M. Saadati, J. Nelson, H. Ayaz, Multimodal fNIRS-EEG classification using deep learning algorithms for brain-computer interfaces purposes, in: Advances in Neuroergonomics and Cognitive Engineering: Proceedings of the AHFE 2019 International Conference on Neuroergonomics and Cognitive Engineering, and the AHFE International Conference on Industrial Cognitive Ergonomics and Engineering Psychology, July 24-28, 2019, Washington DC, USA 10, Springer, 2020, pp. 209–220.
[56] M. Saadati, J. Nelson, H. Ayaz, Convolutional neural network for hybrid fNIRS-EEG mental workload classification, in: Advances in Neuroergonomics and Cognitive Engineering: Proceedings of the AHFE 2019 International Conference on Neuroergonomics and Cognitive Engineering, and the AHFE International Conference on Industrial Cognitive Ergonomics and Engineering Psychology, July 24-28, 2019, Washington DC, USA 10, Springer, 2020, pp. 221–232.
[57] F. Khanam, A.A. Hossain, M. Ahmad, Electroencephalogram-based cognitive load level classification using wavelet decomposition and support vector machine, Brain-Comput. Interfaces 10 (1) (2023) 1–15.
[58] J. Cao, E.M. Garro, Y. Zhao, EEG/fNIRS based workload classification using functional brain connectivity and machine learning, Sensors 22 (19) (2022) 7623.
[59] H. Rajaguru, S.K. Prabhakar, Variational Bayesian matrix factorization and certain post classifiers for classification of epilepsy from EEG signals, Res. J. Pharmacy Technol. 9 (6) (2016) 1–5.