An ELM-based Deep SDAE Ensemble for Inter-Subject Cognitive Workload Estimation with Physiological Signals

Zhanpeng Zheng[1], Zhong Yin[1,*], Jianhua Zhang[2]

1. Engineering Research Center of Optical Instrument and System, Ministry of Education, Shanghai Key Lab of Modern Optical System, University of Shanghai for Science and Technology, Shanghai, 200093, P. R. China.
E-mail: 182560473@st.usst.edu.cn; yinzhong@usst.edu.cn

2. OsloMet Artificial Intelligence Lab, Department of Computer Science, Oslo Metropolitan University, Oslo, N-0130, Norway.
E-mail: jianhuaz@oslomet.no

Abstract: Evaluating operator cognitive workload (CW) levels in human-machine systems based on neurophysiological signals is becoming the basis to prevent serious accidents due to abnormal state of human operators. This study proposes an inter-subject CW classifier, extreme learning machine (ELM)-based deep stacked denoising autoencoder ensemble (ED-SDAE), to adapt the variations of the electroencephalogram (EEG) feature distributions across different subjects. The ED-SDAE consists of two cascade-connected modules, which are termed as high level personalized feature abstractions and abstraction fusion. The combination of SDAE and locality preserving projection (LPP) technique is regarded as base learner to obtain ensemble members for training meta-classifier by stacking-based approach. The ELM model with Q-statistics diversity measurement is acted as meta-classifier to fuse above inputs to improve classification performance. The feasibility of the SD-SDAE is tested by two EEG databases. The multi-class classification rate achieves 0.6353 and 0.6747 for T1 and T2 respectively, and significantly outperforms several shallow and deep CW estimators. By computing the main time complexity, the computational workload of the ED-SDAE is also acceptable for high-dimensional EEG features.

Key Words: Cognitive workload, EEG, ensemble learning, extreme learning machine, stacked denoising autoencoder

1 Introduction

Investigating the intelligent computers and the cognitive workload (CW) experienced by human operators becomes a primordial concern in various control and automation systems where the human factor is becoming a critical component. Beyond a certain burden in cognitive processes potentially increases the possibility of accidents and results in the degradation of operator performance. The term CW is generally defined as the proportion of operator capacity or resources for information processing required to meet system demands. High CW level, or excessively low CW level, affecting human error, which is loss of attentional resources and working memory capacity over time. In vital decision-making and strategy development environments, such as air traffic control [1], medical and emergency applications [2], nuclear power plants [3].

When monitoring the level of CW, approaches are usually distributed into three classes, namely subjective scoring, secondary task performance, and neurophysiological signals. In particular, NASA Task Load Index (TLX) and Subjective Workload Assessment Technique (SWAT) have been shown to be a reliable indicator of CW, which was composed of post-hoc questionnaires. Considering the fact that subjective rating is not accessible to on-line, ongoing assessment during the time courses of human-machine (HM) tasks. It is noted that the secondary task performance can potentially impair main task performance under a certain of HM operations. Neurophysiological measures mainly include electroencephalogram (EEG), electrocardiogram (ECG), electrooculogram (EOG), functional near infrared spectroscopy (fNIRS), and electrodermal measure (EDM) [4]. Specifically, EEG is a useful biomarker in the measurement of CW level due to its noninvasive to recording instantaneous recognition via various EEG recordings of cortical activities.

The EEG has a capability to provide a precise and reliable estimation of the CW based on the power spectral density (PSD) of the EEG extracted from major frequency bands and cortical locations. There exists strong clinical evidence to show that EEG recordings allow for the measurement of neural activity because of its ease for manipulation with a spatial resolution and high temporal resolution [5]. Since activation of frontal and right parietal cerebral regions is closely related to the synchronization of the theta rhythm (4-8 Hz) and the desynchronization of the alpha rhythm (8-12 Hz), its implementation received additional working memory workload [6]. An increase of EEG theta band power may indicate increased CW [7]. EEG alpha band power in the occipital and parietal brain locations decreases with an increase of complex and multitasking environments [8].

Considering that classical machine learning estimators are a significant and reliable means to classify CW levels based on EEG datasets acquired from a single day, a single mental task and a single subject, leading to receiving high classification accuracy [9]. However, cross-day, cross-task and cross-subject issues have needed to been considerably overcome by feature selection methods and state-of-the-art classifiers. The commonly used methods applied to CW assessment are summarized as follows. Recurrent neural network (RNN), acting as the biological nervous system to capture temporal variation in brainwave patterns connected with

[*] Corresponding Author: Zhong Yin. This work is sponsored by the National Natural Science Foundation of China under Grant No. 61703277 and the Shanghai Sailing Program (17YF1427000).

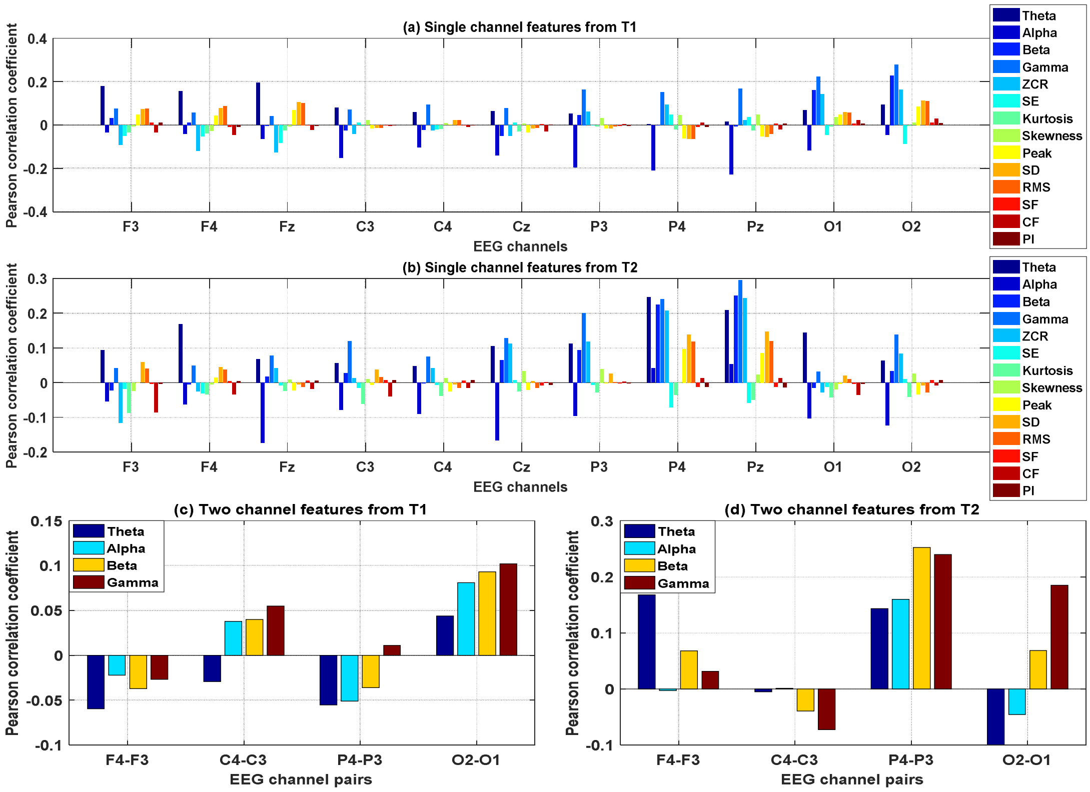

Fig. 1. Pearson correlation coefficient between the feature time courses and target CW levels in T1 and T2.

cross-day CW level recognition [10]. Usually, subject-specificity classifiers are used to learn information within a subject, yet transferring knowledge of one subject to another without a comparable performance. Therefore, the nonstationary characteristic of EEG may enhance the statistical distribution difference across subjects. More exactly, the promising solution is to develop an inter-subject model that is stable enough to track the variations of EEG data distributions in a wide variety of individuals.

The primary objective of the paper is to develop an EEG-based inter-subject CW classifier to track characteristics of features distribution across multiple individuals. The combination of feature selection and classification scheme is employed to generate a valid cross-subject CW recognition system. Common feature extraction techniques have been considerably used in theories and applications, such as differential entropy [11], principal component analysis (PCA) [12], and Fisher projection. Similarly, deep learning-based feature abstraction received attention because of its hierarchical structure to feature representation, which is crucial for extracting useful information. Furthermore, for the deep learning primitives and neural network structure varieties, i.e., autoencoder (AE), stacked autoencoder (SAE), SDAE, convolutional neural network (CNN), and deep belief network (DBN). Deep learning methods originate from the fact that its outstanding generalization capability benefitting from high level feature representation.

It is commonly known that ensemble learning methods have shown to be effective and robust classification performance in comparison with single learner. Ensemble learning techniques showed impressive performance in many real-world applications, such as imbalanced data classification, video recognition, and medical image analysis [13]. In particular, the local information preservation is used as a feature processing unit for feature space dimension. SDAE attempts to generate ensemble members by employing the advantage of deep learning among baseline learner. The critical motivation behind is that each SDAE obtains good performance under different data distribution. The Q-statistics algorithm is expected to measure the diversity of the ensemble learners pair by pair for building inputs of meta-classifier. To fuse the outputs of base learners across different subjects, the ELM has been employed as meta-classifier, due to outstanding generalization capability and learning speed. Inspired by abovementioned, we specifically develop an ELM-based deep SDAE ensemble estimator for coping with inter-subject CW level classification issue.

The remainder of this paper is outlined as follows. The materials are covered in Section 2. In Section 3, the proposed ED-SDAE model is described, followed by the derived results in Section 4. Section 5 briefly concludes the contributions of the present study.

2 Materials

2.1 EEG Datasets Description

Two EEG datasets were collected in our previous works to validate the robust of the proposed CW model [14]-[15]. Two different cognitive tasks, denoted as T1 and T2, were performed by eight (tagged by A-H) and six (tagged by I-N) healthy subjects, respectively. For each task, each subject participated two same sessions of experiments under different days. For T1 and T2, the experimental demands were set based on automation-enhanced cabin air management system

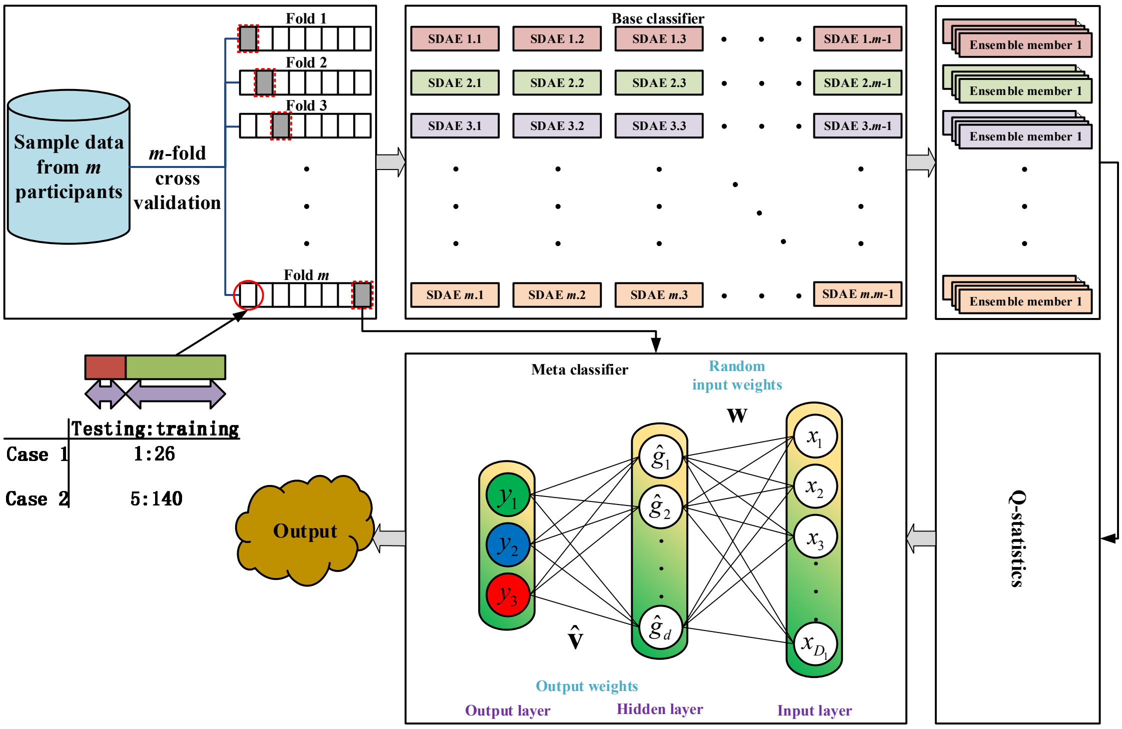

Fig. 2. Architecture of the proposed ED-SDAE for multiclass CW classification.

(AutoCAMS). Subjects were required to manually control safety-critical process control conditions in a virtual space cabin that aimed to sustain the air quality via four subsystems. By programming the number of failed subsystems and actuator sensitivity, resulting in variations of task demands. In summary, low CW (LCW), medium CW (MCW) and high CW (HCW) of each subject were recorded as indicators to show subject's CW level.

2.2 EEG Data Preprocessing and Feature Extracting

The 11 shared channels were at 500 Hz frequency sampling and placed at F3, F4, Fz, C3, C4, Cz, P3, P4, Pz, O1 and O2 positions. The raw EEG signals were preprocessed by a 4-order Butterworth filter and independent component analysis to remove the ocular and scalp-muscular noise. Then, a non-overlapping 2-s segment was obtained. Finally, 2700 and 1450 EEG segments were prepared for each session of T1 and T2, respectively. For a segment of one channel, average PSD in theta (4–8 Hz), alpha (8–13 Hz), beta (14–30 Hz), and gamma (31-40 Hz) bands, zero-crossing rate (ZCR), Shannon entropy (SE), kurtosis, skewness, peak, standard deviation (SD), root of mean square (RMS), shape factor (SF), crest factor (CF), and pulse index (PI) were computed by fast Fourier transform. Simultaneously, PSD differences between right and left scalps on channel pairs F4-F3, C4-C3, P4-P3, and O2-O1 were extracted. In total, 11*14+16=170 EEG features were extracted from each segment. The subject-average Pearson correlation coefficient between single channel features and target CW levels were shown in the Fig. 1(a) and (b). The positive correlation is found for PSD features, Peak, SD and RMS. Consistent negative correlation of the F4-F3, P4-P3, and O2-O1 power differences were shown in Fig. 1(c) and (d).

3 ELM-based Deep SDAE Ensemble

Deep SDAEs are treated as base learners, its weights between the input layer and the first hidden layer are trained by using LPP to capture the personalized feature abstractions. Given N_1 EEG instances with each feature vector denoted by $\mathbf{x} \in R^D$, mapping $\mathbf{x}_i$ to a low dimensional feature representation, $\mathbf{z}_i = \mathbf{A}^T \mathbf{x}_i$. $\mathbf{S}$ is created to evaluated the similarity between each two instances. The transformation matrix is computed by minimizing the following objective function,

$$\min \phi = \frac{1}{2} \sum_{ij} \left(\mathbf{z}_i - \mathbf{z}_j \right)^2 S_{ij}. \tag{1}$$

$$\text{s.t. } \mathbf{A}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{A} = 1$$

In following equation, $\mathbf{D}$ is a diagonal matrix with $D_{ii} = \sum S_{ij}$. By defining $\mathbf{L} = \mathbf{D} - \mathbf{S}$, the λ of matrix $\mathbf{X}$ is solved and $\mathbf{A}$ describes the corresponding eigenvector [16].

$$\mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{A} = \lambda \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{A} \tag{2}$$

The deep architecture of the SDAE is constructed by passing on the encoded outputs from an AE to the inputs of next layer. To eliminate noise-free feature abstractions, $\tilde{\mathbf{z}} \sim q_m \left(\tilde{\mathbf{z}} \middle| \mathbf{z} = \mathbf{A}^T \mathbf{X} \right)$ is employed. High level feature representation at hidden layer H is computed,

$$\tilde{\mathbf{g}}^{(H)} = f\left(\tilde{\mathbf{W}}_H^T f\left(\ldots \tilde{\mathbf{W}}_2^T f\left(\tilde{\mathbf{W}}_1^T \mathbf{A}^T \mathbf{x} + \tilde{\mathbf{b}}_1 \right) + \tilde{\mathbf{b}}_2 \ldots \right) + \tilde{\mathbf{b}}_H \right). \tag{3}$$

the final CW level is achieved by a supervise classification layer,

$$\tilde{\mathbf{y}} = f\left(\tilde{\mathbf{V}}^T \tilde{\mathbf{g}}^{(H)} + \tilde{\mathbf{b}}_v \right). \tag{4}$$

In this equation, $\tilde{\mathbf{V}}$ and $\tilde{\mathbf{b}}_v$ indicate output weights and bias, respectively. The optimal parameters of the

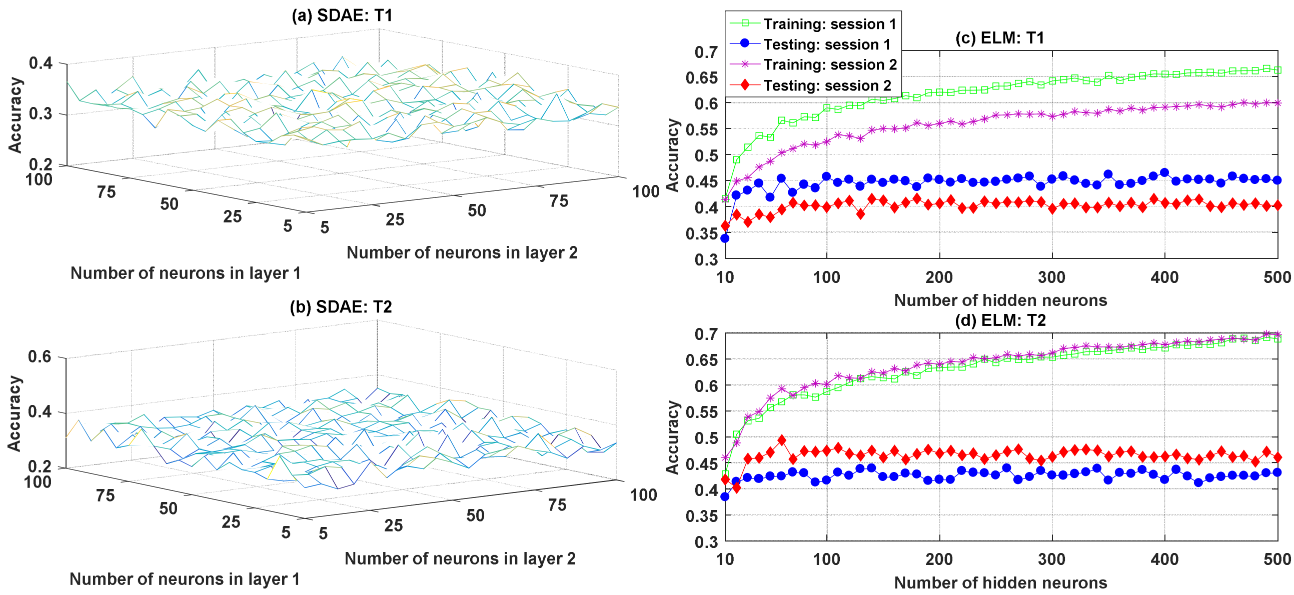

Fig. 3. Subject-average CW classification results under member classifiers across (a), (b) the SDAE learner, (c), (d) the ELM classifier.

SDAEs $\boldsymbol{\theta}_{SDAE}^{*}=\{\tilde{\mathbf{V}},\tilde{\mathbf{b}}_v,\{\tilde{\mathbf{W}}_h,\tilde{\mathbf{b}}_h\}_{h=1}^{H}\}$ are corrected through back-propagation (BP) algorithm.

we introduce Q-statistics [17] to evaluate the diversity of the member learners. Then, the Q-statistics value $Q_{i,j}$ of member E_i and E_j can be defined as,

$$Q_{i,j}=\frac{N^{11}N^{00}-N^{01}N^{10}}{N^{11}N^{00}+N^{01}N^{10}}. \quad (5)$$

N^{ij} represents the number of the instances correctly (or wrongly) classified by pair classifiers via confusion matrix. For a set of n SDAEs, the averaged values is computed as,

$$Q_{av}=\frac{2}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}Q_{i,j}. \quad (6)$$

New feature subsets S_{tr}^{*} are obtained, as seen in Eqns. (7)-(8).

$$\text{Choose } Q_{av}^{*}\rightarrow\{E(1),E(2),\ldots,E(k)\}$$
$$\text{s.t. } Q_{av}^{*}=\begin{cases}0, & if\ E(i)\neq E(j)\\ [-1,0)\cup(0,+1], & \text{otherwise}\end{cases} \quad (7)$$

$$\text{Update } S_{tr}^{*}=\left\{E(k)\Big|_{k=1}^{n^{*}},\ 2\leq n^{*}\leq n\right\}. \quad (8)$$

Where output of one base learner denotes $E(k)$.

The training of an ELM aims to reach the smallest

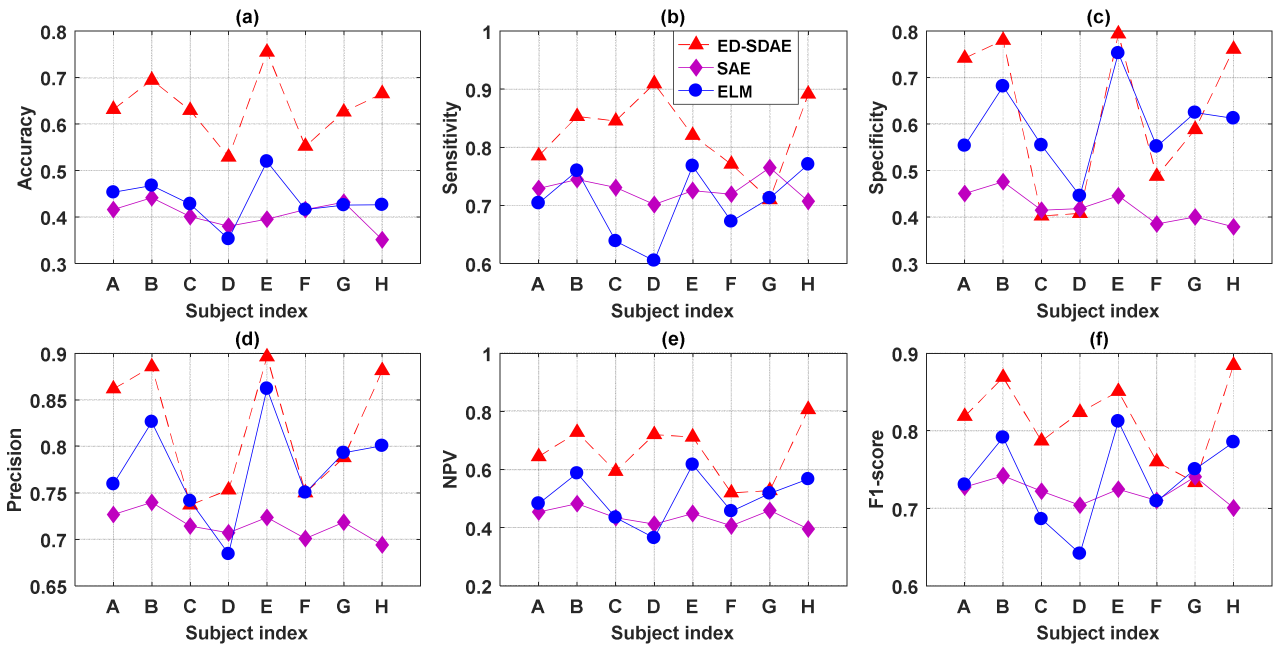

Fig. 4. Testing classification performance of 8 subjects (denoted by A-H) in T1 computed by ED-SDAE, SAE, and ELM. NPV denotes the negative predicting value. All the legends are as same as those in subfigure (b).

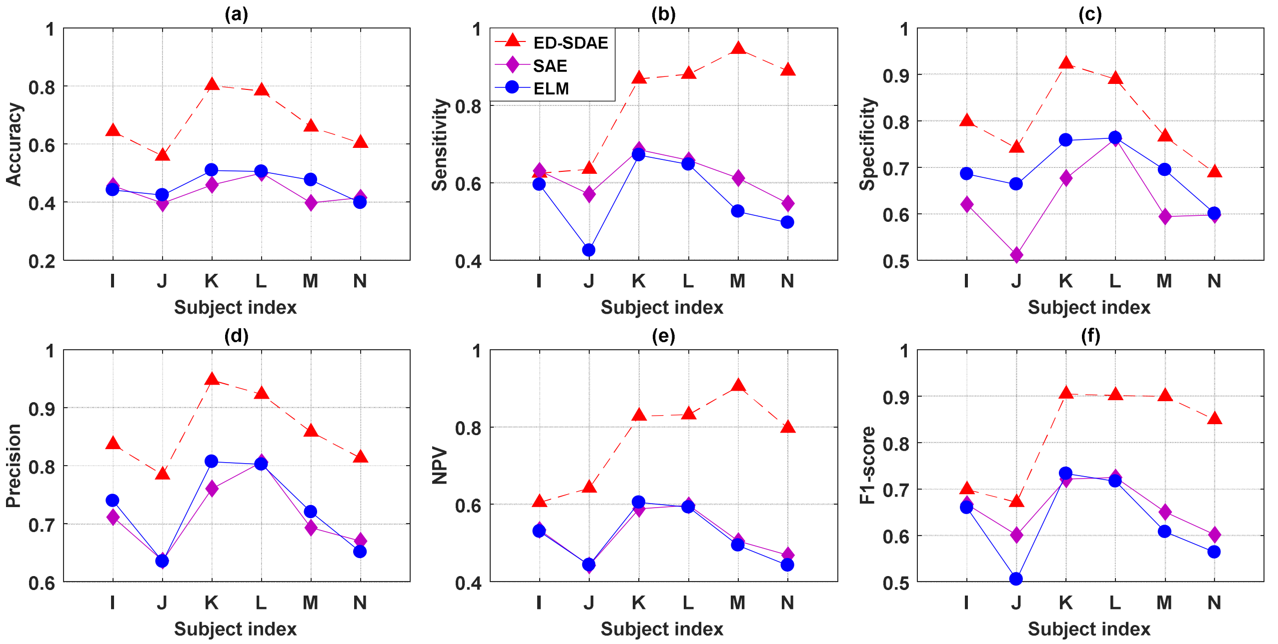

Fig. 5. Testing classification performance of 6 subjects (denoted by I-N) in T2 computed by ED-SDAE, SAE, and ELM. All the legends are as same as those in subfigure (b).

regularization error,

$$\min_{\hat{\mathbf{v}}} E(\hat{\mathbf{v}}) = \frac{1}{2}\|\hat{\mathbf{v}}\|_2^2 + C \cdot \frac{1}{2}\|\hat{\mathbf{g}}^T\hat{\mathbf{v}} - \mathbf{y}\|_2^2. \tag{9}$$

where $\hat{\mathbf{g}}$ is the sigmoid function, $\hat{\mathbf{v}}$ is the output weight. Constant C is used to balance the fitting error. Thus, the output of the final ensemble model is,

$$\mathbf{y}^* = \hat{\mathbf{g}}\left(\mathbf{w}^T S_{tr}^* + \mathbf{b}\right)\hat{\mathbf{g}}^T\left(\frac{\mathbf{I}}{C} + \hat{\mathbf{g}}\hat{\mathbf{g}}^T\right)^{-1}\mathbf{y}. \tag{10}$$

where vector $\mathbf{I}/C$ is added to the diagonal of the $\hat{\mathbf{g}}\hat{\mathbf{g}}^T$ based on ridge regression theory [18]. Only the $\hat{v}_i$ need to be tuned. The general framework of the ED-SDAE is shown in Fig. 2.

4 Cognitive Workload Classification Results

For T1, $2700 \times 8 \times 2 = 43200$ instances are prepared from eight subjects with equal number of instances for each CW level. For T2, 580, 435 and 435 instances are prepared from a subject for three CW levels, respectively. In total, $1450 \times 6 \times 2 = 17400$ instances are obtained. It is noted that division of training and testing sets on the basis of a leave-one-subject-out paradigm. That is, the EEG of one subject is used for testing, the data of

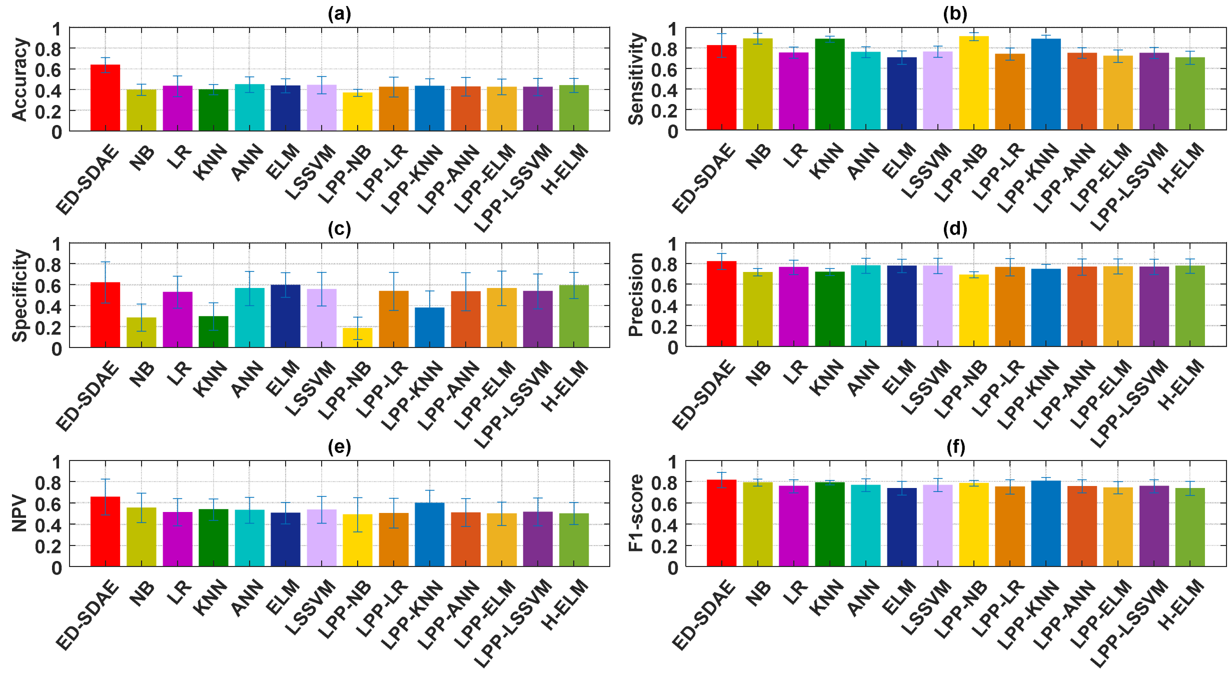

Fig. 6. Subject-average performance metric on T1 computed by ED-SDAE and 13 different classifiers. The error bar marks the standard deviation.

remaining subjects are used for training. The optimal hyper-parameters of the base learners of two sessions are separately computed. The classification performance of the SDAE under different parameters is shown via mesh plot in Fig. 3(a) and (b). The optimal parameters with 10 and 85 hidden neurons for T1, 15 and 5 hidden neurons for T2. For the ELM, the better performance is achieved with 390 and 60 hidden neurons for two Tasks in Fig. 3(c) and (d). For least square support vector machine (LSSVM), the regularization parameter is set to 2^{-8} and 2^{-12} for T1 and T2, respectively. For k values of k-nearest neighbor (KNN) classifier are fixed to 39 and 50 respectively. For Artificial neural networks (ANN), the number of hidden nodes is set to 120 and 430 respectively. For the SAE, the optimal parameter is found when 35 and 85 neurons for T1, and 95 and 40 neurons for T2. For hierarchical-ELM (H-ELM), the optimal 8 (or 9) hidden layers with each of 40 (or 10) neurons are used for T1 (or T2). Finally, the optimal parameters of above model are determined by cross-validation and grid search. Note that the naïve Bayesian (NB) and logistic regression (LR) models are also included.

The performance metrics of the accuracy, sensitivity, specificity, precision, negative predicting value (NPV) and F1-score under two Tasks are shown in Figs. 4 and 5. The LCW and remaining CW levels are acted as positive and negative classes, respectively. The ED-SDAE outperforms the conventional ELM and SAE according to classification results for all 14 subjects. The LPP is used to generate six hybrid classifiers and denoted as LPP-NB, LPP-LR, LPP-KNN, LPP-ANN, LPP-ELM and LPP-LSSVM. In Figs. 6, the performance metrics of T1 with the optimal hyper-parameters of ED-SDAE and 13 shallow classifiers are illustrated with error-bar plots. It is shown the proposed ED-SDAE achieves the highest accuracy, specificity, precision, NPV and F1-score values compared with other models. According to the t-test, we found the significant improvement in the ED-SDAE compared with other models ($p<0.05$).

5 Conclusions

In this study, the proposed an inter-subject cognitive workload estimator ED-SDAE that is modified from the deep learning primitive SDAE to filter the subject-independent variation of EEG indicators aiming at finding the dynamical properties on EEG time courses, and from ELM for pre-training quickly to improve generalizability and universality via the better measurement results of Q-statistics. The competence of the ED-SDAE has been proved on two databases, where obtained multiclass classification accuracies of 0.6353 and 0.6747 for two Tasks. The two-tailed t-test demonstrates the performance improvement is significant superior to shallow and deep classifiers with optimal hyper-parameters derived. The computational complexity of the ED-SDAE is also acceptable for high-dimensional EEG instances.

References

[1] L. Giraudet, J. P. Imbert, M. Bérenger, S. Tremblay, M. Causse, The neuroergonomic evaluation of human machine interface design in air traffic control using behavioral and EGG/ERP measures, *Behavioural Brain Research*, 2015, 294:246–253.

[2] T. Morineau, J. M. Flach, The heuristic version of cognitive work analysis: a first application to medical emergency situations, *Applied Ergonomics*, 2019, 79:98-106.

[3] L. Reinerman-Jones, G. Matthews, J. E. Mercado, Detection tasks in nuclear power plant operation: vigilance decrement and physiological workload monitoring, *Safety Science*, 2016, 88:97–107.

[4] S. Anders, M. Lotze, M. Erb, W. Grodd, N. Birbaumer, Brain activity underlying emotional valence and arousal: a response related FMRI study. *Human Brain Mapping*, 2004, 23:200-209.

[5] P. Antonenko, F. Paas, R. Grabner, T. van Gog, Using electroencephalography to measure cognitive load, *Educational Psychology Review*, 2010, 22:425–438.

[6] Z. Halim, S. Khan, A data science-based framework to categorize academic journals, *Scientometrics*, 2019, 119:393–423.

[7] X. W. Wang, D. Nie, B. L. Lu, Emotional state classification from EEG data using machine learning approach, *Neurocomputing*, 2014, 129:94–106.

[8] J. Christensen, J. Estepp, G. Wilson, C. Russell, The effects of day-to-day variability of physiological data on operator functional state classification, *NeuroImage*, 2012, 59:57–63.

[9] D. Model, M. Zibulevsky, Learning subject-specific spatial and temporal filters for single-trial EEG classification, *NeuroImage*, 2006, 32:1631- 1641.

[10] 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 Recognition Letters*, 2017, 94:96–104.

[11] R. N. Duan, J. Y. Zhu, B. L. Lu, Differential entropy feature for EEG-based emotion classification, in *6th International IEEE/EMBS Conference on Neural Engineering (NER)*, 2013: 81–84.

[12] D. Asir, S. Appavu, E. Jebamalar, Literature review on feature selection methods for high-dimensional data, *International Journal of Computer Applications*, 2016, 136:9-17.

[13] X. Zhou, Y. Shang, H. Yan, G. Guo, Ensemble similarity learning for kinship verification from facial images in the wild, *Information Fusion*, 2015, 32:40–48.

[14] Z. Yin, J. Zhang, Operator functional state classification using least square support vector machine based recursive feature elimination technique, *Computer Methods and Programs in Biomedicine*, 2014, 113:101–115.

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

[16] R. Jiang, W. Fu, L. Wen, S. Hao, R. Hong, Dimensionality reduction on anchorgraph with an efficient locality preserving projection, *Neurocomputing*, 2016, 187:109–118.

[17] L. I. Kuncheva, C. J. Whitaker, Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy, *Machine Learning*, 2003, 51:181–207.

[18] A. E. Hoerl, R. W. Kennard, Ridge regression: biased estimation for nonorthogonal problems, *Technometrics*, 1970, 12:55–67.