Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Classifying Mental Workload Levels Using Semi-Supervised Learning Technique

Jianhua Zhang*, Jianrong Li**

**Dept. of Computer Science, Oslo Metropolitan University, 0166 Oslo, Norway (e-mail: jianhuaz@oslomet.no)*
***School of Info. Sci. and Eng., East China University of Science and Technology, Shanghai 200237, China (e-mail: jrli90@163.com)*

Abstract: Real-time monitoring and analysis of human operator's mental workload (MWL) is crucial for development of adaptive/intelligent human-machine cooperative systems in various safety/mission-critical application fields. Although data-driven machine learning (ML) approach has shown promise in MWL recognition, it is usually difficult to acquire sufficient labeled data to train the ML model. This paper proposes semi-supervised extreme learning machines (SS-ELM) for MWL pattern classification using solely a small number of labeled data. The experimental data analysis results are presented to show the effectiveness of the proposed SS-ELM paradigm to effectively exploit a large number of unlabeled data for the real-world 3- or 4-class MWL classification problem.

Keywords: Human factors; Mental workload; Operator functional state; EEG signals; Neuroimaging; Semi-supervised learning; Extreme learning machine.

1. INTRODUCTION

Automation, automatic control system, and artificial intelligence (AI) techniques have been widely applied to various fields, but there is still a long way for the current development of automation and AI technologies to achieve fully-automated control for many real-world complex and uncertain systems. In this connection, Human- Machine Systems (HMS) are still ubiquitous in practice in most safety-critical application domains (Lal and Craig, 2001). Compared with machines, human operators are more susceptible to external disturbances or the impact of their own psychophysiological fluctuations (Bobko et al., 1998). Therefore, it is not surprising that human factors play a significant role in the achievement of desired performance of HMS. In recent years, researchers from multiple disciplines have focused on the research of how to maintain the optimum Operator Functional State (OFS) to ensure the successful completion of the tasks in the HMS context (Hollender et al., 2010).

The operator's mental workload (MWL) is an essential dimension of the multi-dimensional construct of OFS. The MWL can be considered as a candidate variable for measuring mental status of human operator, which reflects the mental demand for operators to accomplish the tasks (Cain, 2007). For operators, too high or too low psychological load is detrimental to the performance of HMS. In order to mitigate this problem, researchers conceived Adaptive Automation (AA) strategy. The AA system can adaptively allocate the tasks between operators and the machines based on the estimated levels of operators' MWL. According to (Mahfouf and Zhang et al., 2007), MWL measurement/assessment/evaluation approaches can be roughly divided into three categories: (1) subjective assessment; (2) task performance measures; and (3) physiological data based assessment. Compared with the former two approaches, the last approach is featured by continuous on-line measurement. ElectroEncephaloGram (EEG), ElectroCardioGram (ECG) and ElectroOculoGram (EOG) have been widely used for MWL recognition (Zhang, Liu, Peng, Raisch, and Wang, 2013; Zhang, Qin, Raisch, and Wang, 2013; Zhang, Yang, and Wang, 2016). In this paper we evaluate the operators' MWL by using multi-modal psychophysiological signals and examine the potential and efficacy of semi-supervised learning technique for enhancing the accuracy and efficiency of high-risk MWL detection.

2. SEMI-SUPERVISED EXTREME LEARNING MACHINE

The SS-ELM is a semi-supervised learning algorithm based on ELM theory and manifold regularization framework, which can take advantage of the unlabeled data to improve the classification accuracy when labeled data are scarce (Huang, Song, Gupta, and Wu, 2014). It determines the output weights by minimizing the squared sum of the empirical training error of labeled data, the norm of the output weights, as well as the manifold regularization term based on both labeled and unlabeled data.

We have the SS-ELM formulation:

$$\min_{W\in\mathbb{R}^{L\times c}} \frac{1}{2}\|W\|^2 + \frac{1}{2}\sum_{i=1}^{l} C_{\mathbf{y}_i}\|\xi_i\|^2 + \frac{\lambda}{2}\mathrm{Tr}(Y^T L Y)$$
$$\text{s.t.} \quad \mathbf{h}(\mathbf{x}_i)W = \mathbf{y}_i^T - \xi_i^T, \qquad i = 1,2,\cdots,l \qquad (1)$$
$$\mathbf{y}_i = \mathbf{h}(\mathbf{x}_i)W, \qquad i = 1,2,\cdots,n$$

Where λ is a trade-off parameter, $C_{\mathbf{y}_i}$ is a penalty factor for training error of data from class y_i, $L \in \mathbb{R}^{n\times n}$ is the graph

Laplacian built from both labeled and unlabeled data, and $Y \in \mathbb{R}^{n\times c}$ is the output matrix of the network with its i-th row equal to $\mathbf{y}_i$.

Note that similar to the weighted ELM (W-ELM) algorithm, here we assign different penalty factor $C_{\mathbf{y}_i}$ to the prediction errors w.r.t. samples from different classes because when the data is unbalanced, i.e., some classes have significantly more samples than other classes, traditional ELM tend to fit the majority classes well, but fits minority classes poorly. This usually results in poor generalization to the testing set. Therefore, in order to cope with the possibly imbalanced classification problem, we reweigh examples from different classes. Suppose that $\mathbf{x}_i$ belongs to class $\mathbf{y}_i$ which has $N_{\mathbf{y}_i}$ training samples, then we assign ξ_i with a penalty of $C_{\mathbf{y}_i} = \dfrac{C_0}{N_{\mathbf{y}_i}}$, where C_0 is a user-defined parameter as in traditional ELM and $N_{\mathbf{y}_i}$ is the number of training samples in the class $\mathbf{y}_i$. In this way, the samples from the dominant classes will not be overfitted by the algorithm and the samples from a class with less samples will not be ignored.

Substituting the constraints into the objective function yields the new formulation in matrix form:

$$\min_{W\in\mathbb{R}^{L\times c}} \left[\frac{1}{2}\| W \|^2 + \frac{1}{2}\| C^{\frac{1}{2}}(\tilde{Y} - HW) \|^2 + \frac{\lambda}{2}\mathrm{Tr}(W^T H^T L H W) \right] \quad (2)$$

where $H = [\mathbf{h}(\mathbf{x}_1)^T, \mathbf{h}(\mathbf{x}_2)^T, \cdots, \mathbf{h}(\mathbf{x}_l)^T]^T \in \mathbb{R}^{l\times L}$, $\tilde{Y} \in \mathbb{R}^{n\times c}$ is the augmented training target with its first l rows equal to Y_l and the rest equal to 0, and $C \in \mathbb{R}^{n\times n}$ is a (penalty) diagonal matrix with its first l diagonal elements $c_{ii} = C_i$ and the rest equal to 0.

Now let us solve the above optimization problem. We first compute the gradient of the objective function w.r.t. W and then by setting the gradient to zero, we obtain the optimal output weights (i.e., the SS-ELM solution) if $L \le l$:

$$W^* = (\mathbf{I}_L + H^T C H + \lambda H^T L H)^{-1} H^T C \tilde{Y} \quad (3)$$

where $\mathbf{I}_L$ is an identity matrix of dimension L.

If $L > l$ (common in SSL), the optimal output weights can be solved by the alterative form:

$$W^* = H^T (\mathbf{I}_n + C H H^T + \lambda L H H^T)^{-1} C \tilde{Y} \quad (4)$$

where $\mathbf{I}_n$ is an identity matrix of dimension n.

In summary, SS-ELM training algorithm consists of two steps:

Step 1: Initialize an ELM network of L hidden neurons with random input weights and biases, and calculate the output matrix of the hidden neurons $H \in \mathbb{R}^{n\times L}$.

Step 2: Use (3) or (4) to compute the output weights W.

3. FEATURE EXTRACTION ALGORITHM

Usually EEG feature extraction algorithms can be divided into time-domain, frequency- domain, time-frequency, and nonlinear dynamics analysis. This paper uses Hilbert-Huang transform (HHT) (Huang, Long, and Shen, 1996; Huang, Shen, and Long et al., 1998; Huang and Wu, 2008; Huang and Shen, 2014; Krishna and Ramaswamy, 2017) to extract the time-frequency features of the physiological signals. HHT is a signal processing method suitable for nonlinear and nonstationary signal analysis and has been successfully applied to diverse fields, such as geophysics and biomedicine. The core idea of the HHT is to use Empirical Mode Decomposition (EMD) to decompose a time-series signal into Intrinsic Mode Functions (IMFs) with a trend and then apply the Hilbert spectral analysis (HSA) method to the IMFs to obtain instantaneous frequency data. The HHT assumes that any signal is composed of a finite and often small number of components (described as IMFs), which form a complete and nearly orthogonal basis for the original signal.

3.1 Empirical Mode Decomposition (EMD)

The general idea of the EMD method is to use the mean value of the upper and lower envelopes of the time series to determine the *instantaneous equilibrium*, and then extract the IMFs. The main steps of EMD include:

Step 1: Find the local maximum and minimum of the original sequence $x(t)$, and connect the local maximum and the minimum with the cubic curve interpolation to obtain the maximum envelope $x_{\max}(t)$ and the minimum envelope $x_{\min}(t)$.

Step 2: Obtain the mean value $m(t)$ by averaging $x_{\max}(t)$ and $x_{\min}(t)$ at each time instant.

Step 3: Calculate the difference between the original sequence $x(t)$ and the instantaneous mean $m(t)$, i.e.,

$$h(t) = x(t) - m(t) \quad (5)$$

For different datasets, $h(t)$ may or may not be an IMF, which must satisfy the following conditions:

C1: In the whole dataset, the number of extrema and the number of zero crossings must either be equal or differ at most by one.

C2: At any point, the mean value of the two envelopes defined by the local maxima and local minima is zero.

If $h(t)$ satisfies the above conditions, it is an IMF; otherwise $h(t)$ is taken as the original sequence, and Steps 1-3 are repeated until C1 and C2 are satisfied.

The first IMF $c_1(t)$ should contain the finest scale or the shortest-period oscillation in the signal, which can be subtracted from the original sequence by:

$$x(t) - c_1(t) = r_1(t) \tag{6}$$

The residue $r_1(t)$ still contains longer-period variations. This residual signal is then treated as the new data and subjected to the same sifting process of the EMD to obtain an IMF of lower frequency. This procedure is repeatedly applied to all subsequent r_j, yielding:

$$\begin{aligned} r_1(t) - c_2(t) &= r_2(t), \\ r_2(t) - c_3(t) &= r_3(t), \\ &\vdots \\ r_{n-1}(t) - c_n(t) &= r_n(t) \end{aligned} \tag{7}$$

The decomposition process finally stops when the residue $r_n(t)$ becomes a monotonic function or a function with only one extremum, from which no further IMF can be extracted. By summing up (6) and (7), we have:

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t) \tag{8}$$

Therefore, after EMD the original signal is decomposed into n IMFs $(c_1(t), c_2(t), \cdots, c_n(t))$ and a residue $r_n(t)$, which can be either the adaptive trend or a constant.

3.2 Hilbert Spectral Analysis

The Hilbert transform of any function $x(t)$ of L^P class is defined as the convolution of $x(t)$ with function $h(t) = 1/\pi t$:

$$y(t) = H[x(t)] = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau \tag{9}$$

where *P* is the Cauchy principal value of the singular integral.

With the Hilbert transform $y(t)$ of the function $x(t)$, we obtain the analytic function,

$$z(t) = x(t) + jy(t) = A(t)e^{j\theta(t)} \tag{10}$$

where $j = \sqrt{-1}$, $A(t) = \sqrt{x^2(t) + y^2(t)}$ is instantaneous amplitude, and $\theta(t) = \arctan\left[\frac{y(t)}{x(t)}\right]$ is instantaneous phase function.

An IMF after the Hilbert transform can be expressed by (10). If we perform a Fourier transform on $z(t)$, we have:

$$W(\omega) = \int_{-\infty}^{\infty} A(t) e^{j\theta(t)} e^{-j\omega t} dt = \int_{-\infty}^{\infty} A(t) e^{j(\theta(t) - \omega t)} dt \tag{11}$$

Then by the stationary phase method (Copson, 1967), the maximum contribution to $W(\omega)$ is given by the frequency satisfying the condition:

$$\frac{d}{dt}\left(\theta(t) - \omega t\right) = 0 \tag{12}$$

Therefore the best instantaneous frequency is simply $\omega(t) = \frac{d\theta(t)}{dt}$.

With both amplitude and frequency being a function of time, we can express the amplitude (or energy, the square of amplitude) in terms of a function of time and frequency, $H(\omega, t)$. The marginal spectrum can then be defined as:

$$h(\omega) = \int_0^T H(\omega, t)\, dt \tag{13}$$

where [0, *T*] is the temporal domain over which the data is defined. The marginal spectrum represents the accumulated amplitude (energy) over the entire data span in a probabilistic sense and provides a measure of the total amplitude (energy) contribution from each frequency value, serving as an alternative spectrum expression of the data to the traditional Fourier spectrum.

As pointed out in Huang, Long and Shen (1996) and Huang, Shen and Long et al. (1998), the frequency in either $H(\omega, t)$ or $H(\omega)$ has a totally different meaning from the Fourier analysis. In the Fourier representation, the existence of energy at a frequency ω means a component of sine or cosine wave persisted through the time course of the data. Here, the existence of energy at the frequency ω means only that, in the whole time course of the data, there is a higher likelihood for such a wave to have appeared locally. In fact, the Hilbert spectrum is a weighted non-normalized joint amplitude-frequency-time distribution. The weight assigned to each time-frequency cell is the local amplitude. Consequently, the frequency in the marginal spectrum indicates only the likelihood that an oscillation with such a frequency exists. The exact occurrence time of that oscillation is given in the full Hilbert spectrum.

In addition to the marginal spectrum, we can also define the Instantaneous Energy (IE) density level as:

$$IE(t) = \int_{\omega} H^2(\omega, t)\, d\omega \tag{14}$$

Obviously, IE also depends on time; it can be used to check the energy fluctuation.

In summary, the signal $x(t)$ is decomposed by EMD method into $c_i(t)$ $(i = 1, 2, \cdots, n)$:

$$c_i(t) = \mathrm{Re}[A_i(t)\exp(j\int \omega_i(t)dt)] \tag{15}$$

where $\mathrm{Re}[\cdot]$ represents the real part of terms with brackets and $A_i(t)$ is represented on the time-frequency plane.

We obtain the Hilbert spectrum of $c_i(t)$ as:

$$H_i(\omega, t) = \begin{cases} A_i(t), \text{ if } \omega = \omega_i(t) \\ 0, \quad \text{otherwise} \end{cases} \tag{16}$$

The IMFs obtained by sifting process of EMD constitute an adaptive basis. This basis usually satisfies empirically all the major mathematical requirements for a time series decomposition method, such as convergence, completeness, orthogonality, and uniqueness, as discussed in Huang, Shen and Long et al. (1998).

For an arbitrary time series of length *N*, $x(t)$, if EMD is used to obtain its IMF components and instantaneous frequencies and instantaneous amplitudes of those IMFs are obtained by using the Hilbert transform, $x(t)$ can be expressed as:

$$x(t) = \mathrm{Re}[\sum_{i=1}^{n} A_i(t)\exp(j\int \omega_i(t)dt)] \tag{17}$$

Here the residue, r_n, is not expressed in terms of a simple oscillatory form for it is either a monotonic function or a function with only one extrema not containing enough information to check if it is an oscillatory component is physically meaningful.

Equation (16) gives both amplitude and frequency of each component as functions of time. The Fourier representation of the same data is

$$x(t) = \mathrm{Re}\sum_{i=1}^{\infty} A_i e^{j\omega_i t} \tag{18}$$

where both A_i and ω_i are constants.

The difference between (17) and (18) is fundamental: the IMF represents largely a generalized Fourier expansion. The variable amplitude and instantaneous frequency not only improve the efficiency of the expansion, but enable the expansion to accommodate nonlinear and nonstationary variations in data. The IMF expansion lifts the restriction of constant amplitude and fixed frequency in the Fourier expansion, allowing for a variable amplitude and frequency representation over time. Equation (17) also enables us to represent the amplitude and the instantaneous frequency as functions of time in a 3D plot, in which the amplitude can be contoured on the frequency-time plane. This frequency-time distribution of the amplitude is designated as the Hilbert amplitude spectrum, $H(\omega, t)$, or simply Hilbert spectrum. If amplitude squared is more desired to represent energy density, then the squared values of amplitude can be substituted to produce the Hilbert energy spectrum as well.

The Hilbert amplitude spectrum $H(\omega, t)$ of the original signal $x(t)$ can be expressed as:

$$H(\omega, t) = \sum_{i=1}^{n} H_i(\omega, t) \tag{19}$$

The calculation of sample entropy requires smaller time course of either deterministic or random signal and it is insensitive to the noises. We select the first three IMF components with the largest variance contribution, compute their respective *sample entropies*, and then use them jointly with the *energy spectrum entropy* and the *Hilbert marginal spectral entropy* to constitute a 5D feature vector.

4. DATA ACQUISITION AND PREPROCESSING

4.1 Experimental Tasks

The simulated task platform used in our experiments is automated-enhanced Cabin Air Management System (aCAMS), which consists of four subsystems: concentration of oxygen (O_2), air pressure (P), concentration of carbon dioxide (CO_2), and temperature (T). In the experiment, we used the aCAMS to simulate the task environment in a closed cabin. The operator's MWL is mainly affected by the Number Of Subsystems (NOS) assigned to him for manual control and the Actuator Sensitivity (AS) in the manual control systems. The aCAMS simulation platform constitutes a complex human-machine cooperative task environment. Nihon Kohden® measurement system was used to measure physiological signals at a sampling rate of 500 Hz.

4.2 Experimental Subjects

Six subjects (22-24 y/o, male; coded by A, B, C, D, E, and F) participated in the experiments. All subjects were healthy, had normal vision and dextromanual. Before the experiments, all subjects were informed of goals and procedure of the experiment and were trained for more than 10 hrs on aCAMS-based task operations.

4.3 Experimental Data Acquisition Procedure

The aCAMS system has four subsystems, each having two control modes: automatic or manual control. The two modes of control can be switched arbitrarily. The control objective of the experiment is to maintain the output variables of the four subsystems within their target ranges by automatic control by automation systems, manual control by human operator, or a mixture of both modes. For manual control, there are two levels of actuator sensitivity (AS): Low or High. The sensitivity of the control variable under High AS is higher than Low AS (Wang, Zhang and Wang, 2016).

Each session lasts for 50 min. and consists of 10 different task-load conditions. The conditions #1, 4, 7, and 10 are under automatic control mode. Operator manually controls two subsystems (O2 and P) in the conditions #2 and 3, the only difference between the two conditions is that the AS is different. Fig. 1 illustrates the 10 task-load conditions in a session of experiment. During the last 10 s of each condition, the operator performs self-assessment of his performance in that condition, so we only consider the measured data of 290 s per condition.

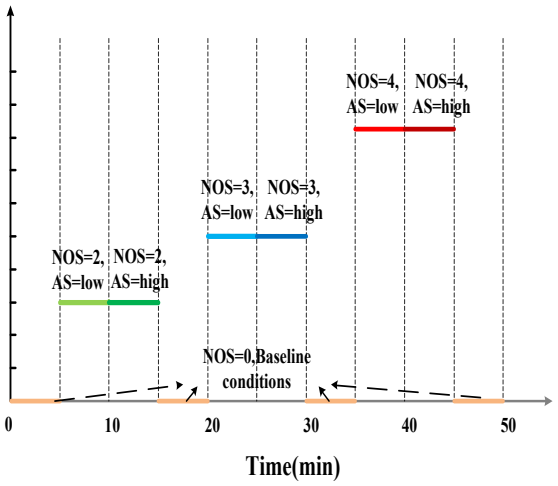

Fig. 1. The 10 task-load conditions in an experimental session.

The EEG, ECG and EOG signals for each subject were collected during the aCAMs operation by using a signal acquisition instrument (sampling rate: 500 Hz). The instrument has the function of removing the disturbance of the power frequency on the electrophysiological signals. In the international standard 10-20 system (Okamoto et al., 2004), 15 EEG electrodes that are most relevant to the MWL variations were selected, namely F3, F4, C3, C4, P3, P4, O1, O2, Fz, Cz, CPz, Pz, AFz, POz, and Oz [13,14]. In addition, the potential difference between the upper middle part of the clavicle and the lower middle part of the left rib was recorded as ECG signal. The EOG signal was measured by the potential between the electrodes above and below the left eye. The recorded raw signals is filtered by a Butterworth band-pass filter (0-40 Hz) and the coherent method is used to remove the artifacts induced by eye movement.

4.4 Determination of Target Classes

The preprocessed data is divided by the sliding time window with length of 1 s (with no overlapping), then each load condition contains 290 sample data. In addition to physiological data, the experiment also records the task performance data, i.e., the output of the subsystems under control. Performance data for subject A is shown in Fig. 2, where the area between the red lines is the target range and the area between the pink lines is the safe range.

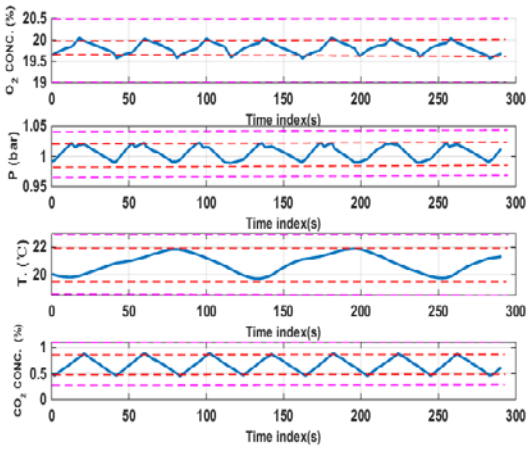

Fig. 2. The task performance data (i.e., four output time trajectories of human-machine cooperative control system under study) (subject A).

In order to quantify the MWL level, we define the Mental Workload Index (MWLI):

$$MWLI = w_{O_2} r_{O_2}(t) + w_P r_P(t) + w_{CO_2} r_{CO_2}(t) + w_T r_T(t) \qquad (20)$$

where $r(t)$ is Boolean variable of the corresponding subsystem (when the output of the corresponding subsystem is in target range at time t, $r(t)=0$; otherwise $r(t)=1$) and w represents the weight of the corresponding subsystem that can be determined by:

$$w = w_1 w_2 w_3 \qquad (21)$$

where w_1 represents the control weight of the corresponding subsystem (when the subsystem is under manual control, $w_1=1$; otherwise $w_1=0$), w_2 represents the difficulty level of the corresponding subsystem among four subsystems, and w_3 denotes the difficulty level of control of the corresponding subsystem with different level of AS. The values of w_2 and w_3 are empirically determined. The basic idea of entropy method is to determine the weight according to the indicator variability. In general, the smaller the information entropy of an indicator, the greater the variation in the indicator, the greater the amount of information provided, the greater the weight. By using (20) and (21), we

obtain the second-to-second MWLI variations, as shown in Fig. 3. We can see that there exists individual difference across the 6 subjects, but the overall trend of change is similar, for example, condition #9 has the peak (highest) level of MWL. The MWL level is higher in the conditions #3, 6 and 8, while the MWL level in the condition #2 and 5 is lower. The baseline conditions #1, 4, 7, and 10 are under automatic control, thus the MWL level in those 4 baseline conditions is zero (under-loaded). Thus we will classify MWL into three (baseline, low, high) or four classes (baseline, low, medium, high).

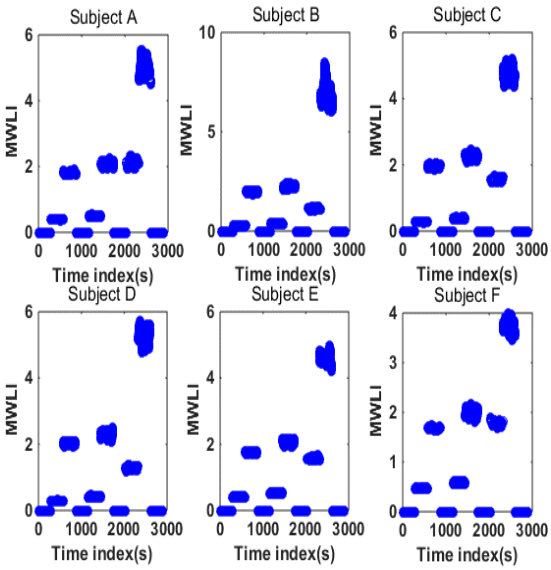

Fig. 3. The change of MWLI over time for each subject.

5. CLASSIFICATION RESULTS AND ANALYSIS

In this section, we will validate the SSL performance across the six subjects. We utilize a windowing approach with a sliding time window with length of 1 s. The methods described in Sect. 3 were applied to the windowed data to extract features based on the HHT algorithm. Fig. 4 illustrates the decomposition of a 2s ECG data into seven IMFs and a residue (trend term).

The sample entropies of the first three IMF components with the largest contribution to the variance, energy spectrum entropy and Hilbert marginal spectral entropy were taken as the features, hence we get a dataset of 2900 samples with a feature dimensionality of 85 (= 5 features/channel * 17 channels).

Usually there are much more unlabeled data than labeled data, thus we divide the dataset into labeled data and unlabeled data at the rate of 1:9. Labeled data is further divided into training and test data at the rate of 4:1. As a result, the number of training samples, test samples and unlabeled samples is 232, 58, and 2610, respectively.

The classification confusion matrices, computed on test set and unlabeled set, are presented in Figs. 5 and 6, respectively. Additionally, the 4-class confusion matrices for each subject are shown in Figs. 7 and 8.

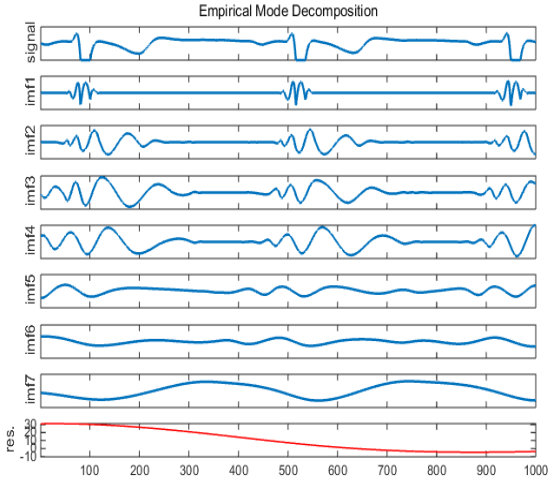

Fig. 4. The EMD of a 2s ECG data into 7 IMFs (subject A).

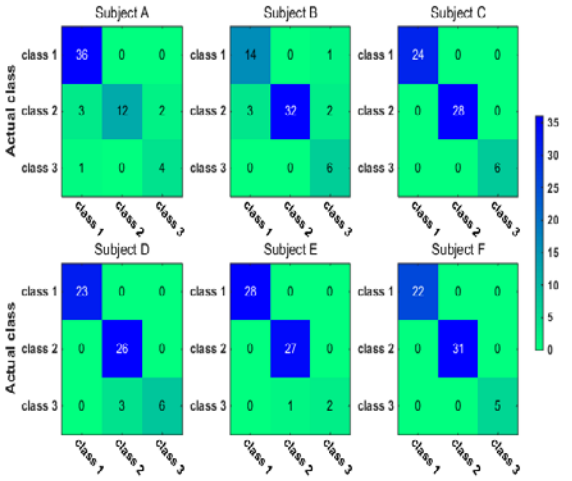

Fig. 5. Testing classification confusion matrix (3-class case).

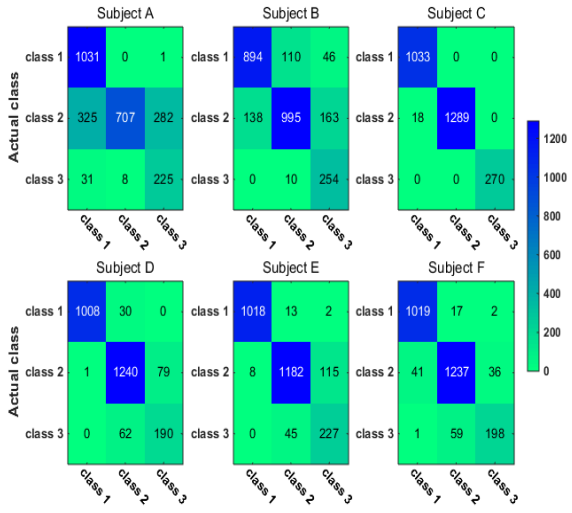

Fig. 6. The classification confusion matrix on unlabeled dataset (3-class case).

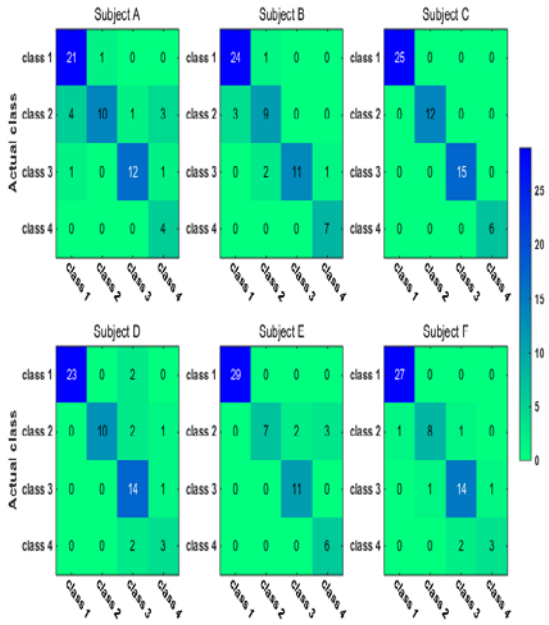

Fig. 7. Testing classification confusion matrix (4-class case).

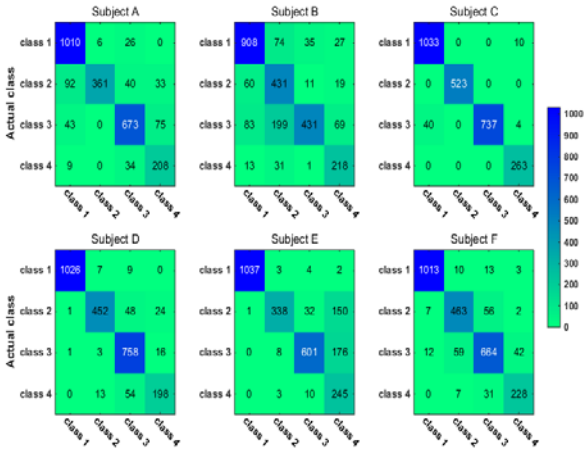

Fig. 8. The classification confusion matrix on unlabeled dataset (4-class case).

We can observe that the confusion matrix in the 4-class case is slightly lower than that in the case of 3-class. Overall, the proposed SSL algorithm leads to encouraging classification performance (i.e., testing classification accuracy is higher than 85% across the six subjects) on the MWL-related dataset. In terms of individual differences (i.e., subject-to-subject variability), there is marked individual difference between subject B and C, i.e., the worst classification performance is obtained for subject B, whereas the classification performance for subject C is the most stable in both three- and four-class cases. On the other hand, from the confusion matrix we can also find that in the case of 3-class, it is easier to classify Class1 and Class2 than Class3.

6. CONCLUSIONS

Although supervised learning techniques have shown promising performance in model-based MWL detection, a practical limitation of ML methods is the lack of sufficient number of labeled training data. Labeling massive data can be expensive or even erroneous given the little known domain knowledge about OFS state available. As the SSL method only requires small amount of labeled data, in this study the SSL paradigm is applied to real-time detection of high-risk MWL state using physiological data. We use SSL to exploit unlabeled data and improve the accuracy of high-risk MWL state detection.

The results presented show that the proposed SSL approach is a promising alternative for risky MWL detection based on physiological signals. By exploiting the information contained in unlabeled data, the graphic semi-supervised learning method can reduce the computational cost and at the same time improve the detection accuracy. It was shown that even perfect classification accuracy can be achieved sometimes by SS-ELM. Furthermore, more can be gained by using SSL method with the increase of the size of unlabeled dataset. This result suggests that by exploring the structure of those unlabeled data, we can exploit additional information to improve the performance of MWL detection.

REFERENCES

Bobko, N. et al. (1998). The mental performance of shiftworkers in nuclear and heat power plants of Ukraine. *Int. J. of Industrial Ergonomics*, vol. 21(3-4), pp. 333-340.

Cain, B. (2007). A review of the mental workload literature. Defense Research and Development Toronto, Canada.

Copson, E.T. (1967). *Asymptotic Expansions*. Cambridge University Press.

Hollender, N. et al. (2010). Integrating cognitive load theory and concepts of human–computer interaction. *Computers in Human Behavior*, vol. 26(6), pp. 1278-1288.

Huang, G., Song, S., Gupta, J.N.D. and Wu, C. (2014). Semi-supervised and unsupervised extreme learning machines. *IEEE Trans. on Cybernetics*, vol. 44(12), pp. 2405-2417.

Huang, N.E., Long, S.R. and Shen, Z. (1996). The mechanism for frequency downshift in nonlinear wave evolution. *Adv. Appl. Mech.*, vol. 32, pp. 59-111.

Huang, N.E. and Shen, S.S.P. (2014). *Hilbert-Huang Transform and Its Applications*, 2nd edn., 400 pp., World Sci. Publishing Co., Ltd., Singapore.

Huang, N.E., Shen, Z., Long S.R. et al. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis. *Proc. of the Royal Society of London A*. vol. 454, pp. 903-995.

Huang, N.E. and Wu, Z.H. (2008). A review on Hilbert-Huang transform: Method and its applications to geophysical studies. *Reviews of Geophysics*, 46, RG2006, Paper No. 2007RG000228, pp. 1-23.

Krishna, P.K.M. and Ramaswamy, K. (2017). Single Channel speech separation based on empirical mode decomposition and Hilbert Transform. *IET Signal Processing*, vol. 11, pp. 579-586.

Lal, S.K.L. and Craig, A. (2001). A critical review of the psychophysiology of driver fatigue. *Biological psychology*, vol. 55(3), pp. 173-194.

Mahfouf, M., Zhang, J. et al. (2007). Adaptive fuzzy approaches to modelling operator functional states in a human-machine process control system. in *Proc. of IEEE Int. Conf. on Fuzzy Systems (FUZZ-IEEE 2007)*, 23-26 July 2007, London, UK, pp. 1-6.

Okamoto, M. et al. (2004). Three-dimensional probabilistic anatomical cranio-cerebral correlation via the international 10–20 system oriented for transcranial functional brain mapping. *Neuroimage*, vol. 21(1), pp. 99-111.

Wang, Y., Zhang, J. and Wang, R. (2016). Mental workload recognition by combining wavelet packet transform and kernel spectral regression techniques. in *Proc. of 13th IFAC Symp. on Analysis, Design, and Evaluation of Human-Machine Systems (HMS2016)*, Kyoto, Japan, Aug. 30-Sep. 02, 2016; *IFAC-PapersOnLine*, vol. 49(19), pp. 561-566.

Zhang, J., Liu, H., Peng, X., Raisch, J. and Wang, R. (2013). Classifying human operator functional state based on electrophysiological and performance mea sures and fuzzy clustering method. *Cognitive Neurodynamics*, vol. 7, pp. 477-494.

Zhang, J., Qin, P., Raisch, J. and Wang, R. (2013). Predictive modeling of human operator cognitive state via sparse and robust support vector machines. *Cognitive Neurodynamics*, vol. 7(5), pp. 395-407.

Zhang, J., Yang, S. and Wang, R. (2016). Operator functional state estimation based on EEG-data-driven fuzzy model. *Cognitive Neurodynamics*, vol. 10(5), pp. 375-383.