This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2020.3019849, IEEE Transactions on Instrumentation and Measurement

EEG-Based Multi-Class Workload Identification Using Feature Fusion and Selection

Zian Pei, Hongtao Wang, *Member, IEEE*, Anastasios Bezerianos, *Senior Member, IEEE*, Junhua Li*（李俊华）, *Senior Member, IEEE*

***Abstract*—The effectiveness of workload identification is one of the critical aspects in a monitoring instrument of mental state. In this field, the workload is usually recognised as binary classes. There are scarce studies towards multi-class workload identification because the challenge of the success of workload identification is much tough, even though one more workload class is added. Besides, most of the existing studies only utilized spectral power features from individual channels but ignoring abundant inter-channel features that represent the interactions between brain regions. In this study, we utilized features representing intra-channel information and inter-channel information to classify multiple classes of workload based on EEG. We comprehensively compared each category of features contributing to workload identification and elucidated the roles of feature fusion and feature selection for the workload identification. The results demonstrated that feature combination (83.12% in terms of accuracy) enhanced the classification performance compared to individual feature categories (i.e., band power features, 75.90%; connection features, 81.72%, in terms of accuracy). With the F-score feature selection, the classification accuracy was further increased to 83.47%. When the features of graph metric were fused, the accuracy was reached to 84.34%. Our study provided comprehensive performance comparisons between methods and feature categories for the multi-class workload identification and demonstrated that feature selection and fusion played an important role in the enhancement of workload identification. These results could facilitate further studies of multi-class workload identification and practical application of workload identification.**

***Index Terms*—Mental Workload Identification, Feature Fusion, Feature Selection, Graph Metric, Brain Connectivity, Power Spectral Density, EEG**

I. Introduction

WITH the increase in the pace of people's lives, their mental workload is elevated accordingly. The previous study has shown that mental overload could lead to errors during decision-making [1], which is one of the main causes of mistakes/accidents. In contrast, keeping workload always low might avoid mistakes/accidents, but it would waste mental resource and result in low work efficiency [2]. Therefore, an appropriate workload level, ensuring high efficiency but no overloading, is desired. To this end, accurate identification of workload level is prerequisite.

In general, the workload can be assessed using subjective or objective manners [3]. Subjective manner is based on individual's self-estimation of task difficulty [4]. In contrast, objective manner is to assess workload based on objective metrics such as performance score or accuracy. Another critical factor affecting workload assessment application is real-time. If an assessment is done discretely, it is not promising for practical application. Nowadays, neurophysiological signals are frequently used to monitor mental states as they can be measured continuously [5]-[8]. Using such signals, the mental workload can be assessed in real-time. To date, electroencephalogram (EEG), electrooculogram (EOG), and electrocardiogram (ECG) have been used in workload assessment [9], [10]. Among these signals, EEG is relatively better for assessing workload level as it directly reflects brain activity [11]. In addition, assessment accuracy could be higher using EEG signal compared to ECG signal, which was found in the Zhang et al.'s study [12].

As we know, band power is one of the feature categories for the investigation of mental workload. For instance, Borghini et al. found that theta band power was increased while alpha band power was decreased when drivers were under high workload [13]. In another study of driver's workload assessment, all five typical frequency bands (i.e., delta, theta, alpha, beta, and gamma) were used [14],

* indicates the corresponding author (Junhua Li, juhalee.bcmi@gmail.com, junhua.li@essex.ac.uk). This work was supported in part by the National Natural Science Foundation of China (No. 61806149), in part by the Guangdong Basic and Applied Basic Research Foundation under Grant (No. 2020A1515010991), in part by the Projects for International Scientific and Technological Cooperation (No. 2018A050506084), and in part by the Science Foundation for Young Teachers of Wuyi University (No. 2018td01).

Zian Pei and Hongtao Wang are with the Laboratory for Brain-Bionic Intelligence and Computational Neuroscience, Faculty of Intelligent Manufacturing, Wuyi University, Jiangmen 529020, China.

Anastasios Bezerianos is with the N.1 Health Institute, National University of Singapore, Singapore 117456, Singapore.

Junhua Li is with School of Computer Science and Electronic Engineering, University of Essex, Colchester, United Kingdom, also with the Laboratory for Brain-Bionic Intelligence and Computational Neuroscience, Faculty of Intelligent Manufacturing, Wuyi University, Jiangmen 529020, China, and also with the Centre for Multidisciplinary Convergence Computing, School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, China.

the results revealed that the band powers could provide high accuracy for driver mental workload classification. All typical frequency bands were used in the above studies, as well as studies in [15], [16]. In this study, we, therefore, included all typical frequency bands and compared the performance among them. Besides power features, functional connection features were recently used in workload identification [17]-[21]. The functional connection features can provide inter-channel information representing interactions between brain regions, which cannot be captured by power features that are derived from individual channels. Gupta et al. found that the EEG graph metric features were more suitable for emotion classification than traditionally used EEG features such as band powers and asymmetry index [22].

As power features and functional connection features respectively represent different information and they are complementary, we explored both of them in our study. In the other classification reports other than workload identification, feature fusion and feature selection gave a positive role in the enhancement of classification performance. In the method proposed by Chen et al. [23], significant multimodal features were selected respectively by two comparative feature selection methods: Fisher Criterion Score and Davies-Bouldin index. The comparison results showed that accuracy was significantly improved. Another study using the fusion of wavelet entropy and spectral power demonstrated the improvement of classification performance [24]. Therefore, we planned to take these two strategies (i.e., feature selection and feature fusion) to find out the role of them in workload identification. Lastly, most of the published studies performed binary classification (i.e., high workload vs. low workload) [25]-[30]. Towards practical application, it is more desirable to classify more levels of workload. To this end, we designed an aircraft operation simulation experiment to induce multiple levels of workload and performed multi-class workload identification. We compared workload identification performance among frequency bands, different individual feature categories, different combinations of feature categories, and feature selection methods. We then provided comprehensive results of workload identification and performance comparison.

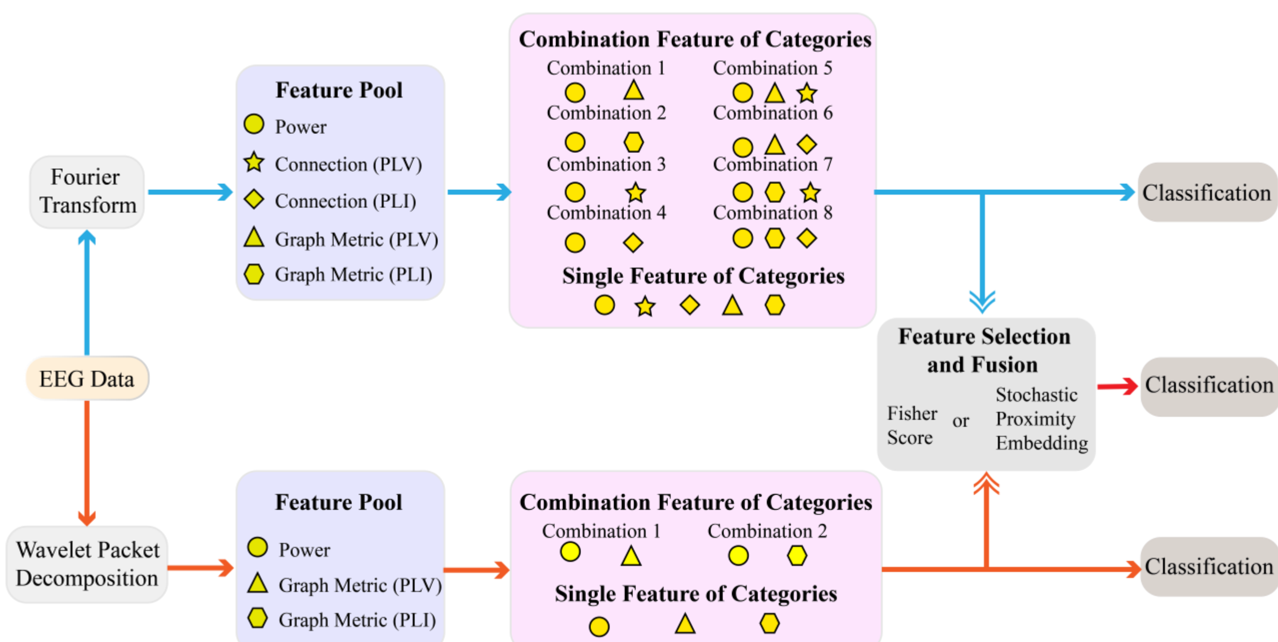

Fig. 1 The schematic of multi-class workload identification using different methods, different feature categories, and different feature combinations.

II. PARTICIPANTS AND METHOD

A. Experiment

The experiment for inducing workload is a simulated aircraft operation, where an Oculus Rift virtual reality headset was used to display virtual 3D aircraft and a joystick was provided to participants. A total of seven participants were recruited in this experiment. All of them had not had any experience of EEG experiment and the use of this aircraft simulation. They were asked to control the virtual aircraft by a joystick and performed three 2-minute long tasks, constituting a 6-minute session. They completed three identical sessions. For each session, they started a low workload task and ended with high workload task. During the low workload task, participants only monitored autonomous aircraft and were not asked to do any control actions. In the medium workload task, participants manually controlled the aircraft and had to pay more effort. In the high workload task, the effort was further increased due to more difficult manipulation for keeping aircraft balanced because the aircraft had malfunctions such as engine failure. During the experiment, 62 EEG channels were used to record brain activity with a sampling rate of 256 Hz. The protocol of the experiment was approved by the institutional review board of the National University of Singapore. All participants signed the consent form before starting the experiment.

B. Data Processing

A typical procedure was utilized to mitigate artifacts from EEG signals, including bandpass filter (0.5~48Hz) and independent component analysis (ICA). The EEG signals were partitioned into 2-second long segments, resulting in 180 segments for each level of workload and a total of 540 segments for each participant. Power features and functional connection features were then extracted for each segment. Consequently, individual categories of features and their combinations were used to identify workload. The schematic is illustrated in Fig. 1.

C. Feature Extraction

Fourier transform (FT) and wavelet packet decomposition (WPD) were respectively utilized to obtain power features in five frequency bands (i.e., delta, 1~4 Hz; theta, 4~8 Hz; alpha, 8~12 Hz; beta, 12~30 Hz; and gamma, 30~45 Hz). The wavelet Daubechies 4 (db4) was selected following the previous research [31]. There were two power features for each frequency band. These were band power and relative band power (i.e., the ratio of the band power to the total power of five bands). In our study, 62 channels were used. Therefore, there were 620 power features (62×5 band power features and 62×5 relative band power features).

The interactions between brain regions could be quantified by Phase Locking Value (PLV), which describes phase coupling. PLV method estimates the phase synchronization among channels. The PLV between channel k and channel l over time span $t = \{t_1, t_2, \cdots, t_k\}$ can be computed as follows

$$PLV_{k,l} = \langle e^{j(\varphi_k(t)-\varphi_l(t))} \rangle \qquad (1)$$

where $\langle\cdot\rangle$ stands for the arithmetic mean over the time span, φ_k and φ_l are the phases of channels k and l.

PLV is affected by volume conduction. In contrast, Phase Lag Index (PLI) is insensitive to volume conduction. The PLI is computed by

$$PLI_{k,l} = \left|\langle sign\left[sin\left(\varphi_k(t) - \varphi_l(t)\right)\right]\rangle\right| \qquad (2)$$

Where $sign$ stands for signum function and $|\cdot|$ indicates absolute value function.

PLV and PLI values are between 0 and 1. A value of 0 indicates no coupling and 1 indicates perfect phase locking. The stronger this nonzero phase locking is, the larger PLV and PLI values are. In our case, a connection matrix with the size of 62×62 was obtained by either PLV or PLI for each segment. Because the connection matrix is symmetric, the upper triangle is the same as the lower triangle. We also removed entries on the main diagonal as these entries are for self-connections. Finally, 1891 [62×(62-1)/2] connection features were obtained. Moreover, we computed the clustering coefficient and assortativity coefficient to have graph metric features. During the computation of graph metric, a sparsity threshold was applied to the connection matrix. Since there is no definitive method to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2020.3019849, IEEE Transactions on Instrumentation and Measurement

determine the sparsity threshold [32], we followed previous studies to utilize a series of thresholds to eliminate the bias due to only using one arbitrary threshold [33]-[37]. A series of thresholds ranging from 0.12 to 0.40 with an incremental step of 0.01 were used in our study and the metric values were obtained by taking integral of all values corresponding to the thresholds.

Clustering coefficient describes the connection centralization of the connection network. The clustering coefficient for channel i is defined as:

$$C_i = \frac{\sum_{k\neq i}\sum_{l\neq i, l\neq k} w_{ik} w_{il} w_{kl}}{\sum_{k\neq i}\sum_{l\neq i, l\neq k} w_{ik} w_{il}} \tag{3}$$

Where w stand for entries in the connection matrix, which were either PLI or PLV values, and i, k, l are channel indices.

The assortativity coefficient can measure the overall connecting structure of a network. Supposing a network has M edges totally and the *n-th* edge is with the degrees of α_n and β_n for each end, assortativity coefficient (r) of the network can be calculated by

$$r = \frac{\frac{1}{M}\sum_n \alpha_n\beta_n - \left[\frac{1}{M}\sum_n \frac{1}{2}(\alpha_n+\beta_n)\right]^2}{\frac{1}{M}\sum_n \frac{1}{2}(\alpha_n^2+\beta_n^2) - \left[\frac{1}{M}\sum_n \frac{1}{2}(\alpha_n+\beta_n)\right]^2} \tag{4}$$

The network is assortative if r is greater than zero and is disassortative if r is less than zero. If r is zero, the network is randomly mixed. The assortative networks are likely to consist of mutually coupled high-degree channels and to be resilient against random failures. In contrast, the disassortative networks are likely to have vulnerable high-degree nodes. For each frequency band, there were 62 clustering coefficients and one assortativity coefficient, resulting in 315 (62×5+1×5) features.

D. Feature Selection and Fusion

High computational demand is needed to process high dimensional features and there might be the curse of dimensionality. To overcome this problem, we used Fisher score (F-Score) [38] and stochastic proximity embedding (SPE) [39] to reduce the feature dimension. The desired number of features has to be set for performing these two methods. We explored different feature numbers (power features: from 20 to 620 with an incremental step of 50, graph metric features: from 5 to 315 with an incremental step of 10, connection features: from 41 to 1891 with an incremental step of 50) to obtain classification accuracies. The desired numbers for each category of features were determined when the highest accuracy was reached.

E. Classification

Random Forest (RF) is a nonlinear classifier [40], belonging to the family of ensemble methods. Such methods have good generalization [41] and are more robust to overfitting than individual trees because each node does not see all features at the same time [40]. It has been shown that random forest performed well for workload classification [42]. We, therefore, adopted random forest in this study. For the performance evaluation, 2-second long segments were considered as samples, resulting in 180 samples for each workload level and each participant. The total number of samples for each participant was 540. The accuracies were separately obtained for each participant using five-fold cross-validation. The accuracies averaged across all participants were reported in this paper.

III. Results

We first compared the performances between FT and WPD. We used FT and WPD to extract frequency bands separately and obtained classification accuracies using the features extracted from these frequency bands. The mean classification accuracy averaged across all subjects was used for performance assessment. The performance was better when using FT compared to WPD (see Fig. 2). In the cases of the single feature category, the highest accuracies under FT method was 77.99% (Graph Metric (PLV)) and the highest accuracy under WPD method was 68.60% (Band Power). The best accuracies were elevated by 0.61% and 2.60% for FT and WPD, respectively, when combining feature categories of power and graph metric (PLV). Overall, the accuracy obtained by using FT was significantly greater than that of using WPD (Wilcoxon signed rank tests, p<0.01, see Fig. 3). These results

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2020.3019849, IEEE Transactions on Instrumentation and Measurement

suggested that FT gave rise to a better performance of workload classification in our case. Therefore, we, hereafter, compared classification accuracies obtained by using FT.

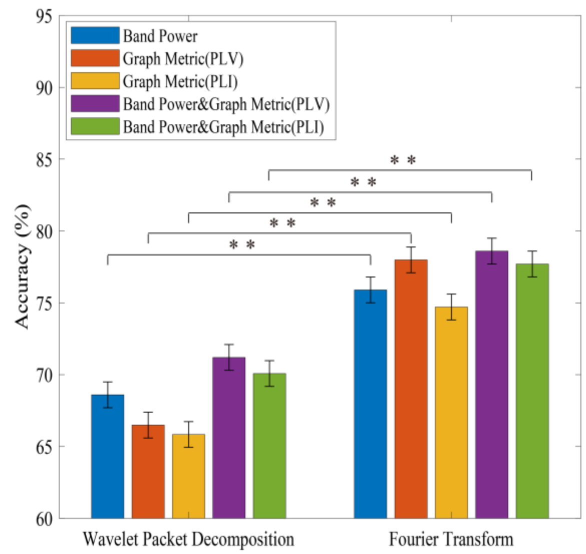

Fig. 2 Accuracies averaged across all subjects for each case. The accuracies obtained using Fourier Transform (FT) were higher than that of using Wavelet Packet Decomposition (WPD). In the cases of using the single feature category, the highest accuracies are 68.60% (Band Power) and 77.99% (Graph Metric (PLV)) for WPD and FT, respectively. When combining features of band power and graph metric (PLV), the accuracies are improved by 0.61% and 2.60% for the conditions of FT and WPD, respectively. The Wilcoxon signed rank test was utilized to check how significant the differences in the accuracies. This statistical evaluation generated p-values. The smaller p-value is, the more significantly different the accuracies are. The cases showing significant differences in the accuracies among feature categories of the same method (i.e., FT or WPD) and between FT and WPD for the same feature category are marked in the figure. * stands for p<0.05 and ** stands for p<0.01.

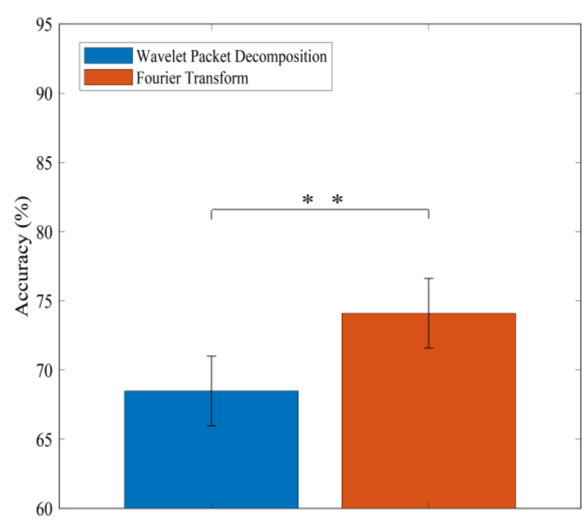

Fig. 3 The overall accuracy comparison between FT and WPD. The accuracies for WPD and FT were 68.20% and 74.77%, respectively. ** stands for p<0.01 (Wilcoxon signed rank test).

All connection feature-based classification accuracies averaged across all subjects in each frequency band and each condition are shown in Table I. Based on the results, the gamma band shows the best performance (accuracy of 80.41% averaged across all cases). Using the gamma band, the accuracy exceeded 80.00% for 5 out of 6 cases. Therefore, the connection features used in the feature combination were from this frequency band. F-score improved classification accuracies, while SPE reduced classification accuracies. The accuracy was enhanced by using feature selection of F-score.

TABLE I

Accuracies averaged across all subjects when using connection features

Band	Accuracy						Mean	Standard Deviation
	PLV			PLI				
	No Feature Selection	Stochastic Proximity Embedding	Fisher Score	No Feature Selection	Stochastic Proximity Embedding	Fisher Score		
Delta (1-4Hz)	55.45	55.00	56.69	54.81	53.47	55.26	5511	0.95
Theta (4-8Hz)	59.58	55.93	60.16	58.60	56.77	60.26	58.55	1.66
Alpha (8-12Hz)	66.61	62.83	67.17	65.87	64.05	66.51	65.51	1.55
Beta (12-30Hz)	80.56	78.04	80.63	78.47	76.27	79.02	78.83	1.50
Gamma (30-45Hz)	**81.72**	**80.24**	**82.17**	**80.26**	**77.46**	**80.61**	**80.41**	1.51

TABLE II

Accuracies averaged across all subjects for single feature categories and combinations of feature categories

Features Category	Accuracy		
	Fourier Transform		
	No Feature Selection	Stochastic Proximity Embedding	Fisher Score
Power	75.90	69.81	76.59
Graph Metric (PLV)	77.99	68.04	79.55
Graph Metric (PLI)	74.71	68.94	75.16
Power& Graph Metric (PLV)	78.60	70.69	79.10
Power& Graph Metric (PLI)	77.70	70.29	78.49
Power& Connection (PLI)	81.69	78.68	82.49
Power& Connection (PLV)	**83.12**	**80.90**	83.47
Power& Graph Metric (PLV) & Connection (PLI)	82.25	77.99	82.96
Power& Graph Metric (PLI) & Connection (PLI)	81.83	78.31	82.62
Power& Graph Metric (PLV) & Connection (PLV)	82.91	80.77	**84.34**
Power& Graph Metric (PLI) & Connection (PLV)	82.78	80.74	83.54

Table II lists the workload classification accuracies for single feature categories and combinations of feature categories. In single feature categories, the performance of graph metric features under the condition of PLV (77.99%) was higher than that of power features (75.90%). Taken Table I and Table II together, we can see that the accuracy obtained using connection features in the gamma

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2020.3019849, IEEE Transactions on Instrumentation and Measurement

band (81.72%) was higher than that of using power or graph metric features (75.90% and 77.99%, respectively) under the condition of PLV and no feature selection and fusion. In combinations of feature categories, the classification accuracies were generally improved compared to that of single feature category. The highest classification accuracy was 83.12%, which was obtained by using the combination of power and connection features (under the condition of PLV). After using feature selection and fusion (F-score), classification accuracies were improved for all cases. The highest accuracy of 84.34% was achieved when using the combination of features of band power, graph metric (PLV), and connection (PLV) and feature selection of F-score. Its confusion matrix is shown in Fig. 4. In this case, the identification of the low workload level was better than the identification of the other workload levels.

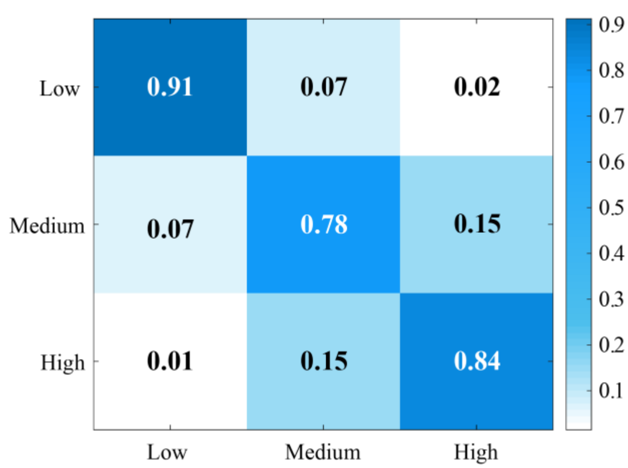

Fig. 4 Confusion matrix for the case of the best classification performance using the combination of features of band power, graph metric (PLV), and connection (PLV) and feature selection of F-score. Columns in the confusion matrix represent predicted classes and rows represent ground truth classes. The entries in the diagonal show correctly classified percentages in each class.

The detailed statistical results obtained by Wilcoxon signed rank tests are shown in Fig. 5. It depicts whether or not the accuracies were significantly different when using different categories of features. We can see that the performance was better when using connection features compared to that of using graph metric features. The combination of feature categories significantly benefited the classification of workload.

IV. Discussion

This study aimed to improve the performance of multi-class workload classification using the fusion of different kinds of features and feature selection. We comprehensively explored different cases and compared their performances in terms of accuracy. This is the first attempt to fuse single-channel features and inter-channel features for classifying three levels of workload. In the case of the single feature category, the performance was higher when using functional connection features compared to band power features. The result demonstrated that the connection features were effective for workload classification. Among the five typical frequency bands, the highest classification performance was achieved when the connection features in the gamma band were used. It has been found that the gamma rhythm originated from the interneurons with the mediation by pyramidal cells [43]. A greater number of studies using EEG recorded from either human (e.g., [44]) or animals (e.g., [45]) have shown that the gamma oscillation was related to cognitive ability. For example, Tallonbaudry and Bertrand [46] revealed that the gamma band played a key role in working memory, showing a high correlation between the enhanced gamma power and the maintenance of cognitive task. According to our study, the accuracy was lower when using graph metric features compared to connection features. We speculated that the aggregated features of graph metric might be too abstract to be as informative as the connection features. This finding informed us that high-level features might be not better than low-level features for the aim of workload classification. We were surprised to observe that the best performance was achieved when the gamma band was used, which was not accordance with our initial expectation that the theta and alpha bands should mostly contribute to the workload classification [47]-[50]. This might be partially due to that the movements during aircraft operation introduced discriminative artifacts into the gamma band of EEG signal. However, this effect should not be significant if any. Because we did not see obvious movement-related artifacts after the procedure of artifacts removal. Further studies are required to elucidate the relationship between the gamma band and mental workload.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2020.3019849, IEEE Transactions on Instrumentation and Measurement

The comparison results of classification performance demonstrated that the feature fusion of different kinds of features outperformed individual feature categories. Feature fusion enhanced classification accuracy, achieving the highest classification accuracy of 84.34% when the features of band power, PLV graph metric and PLV connection were fused. This suggested that different feature categories were complementary to each other in terms of discriminative information.

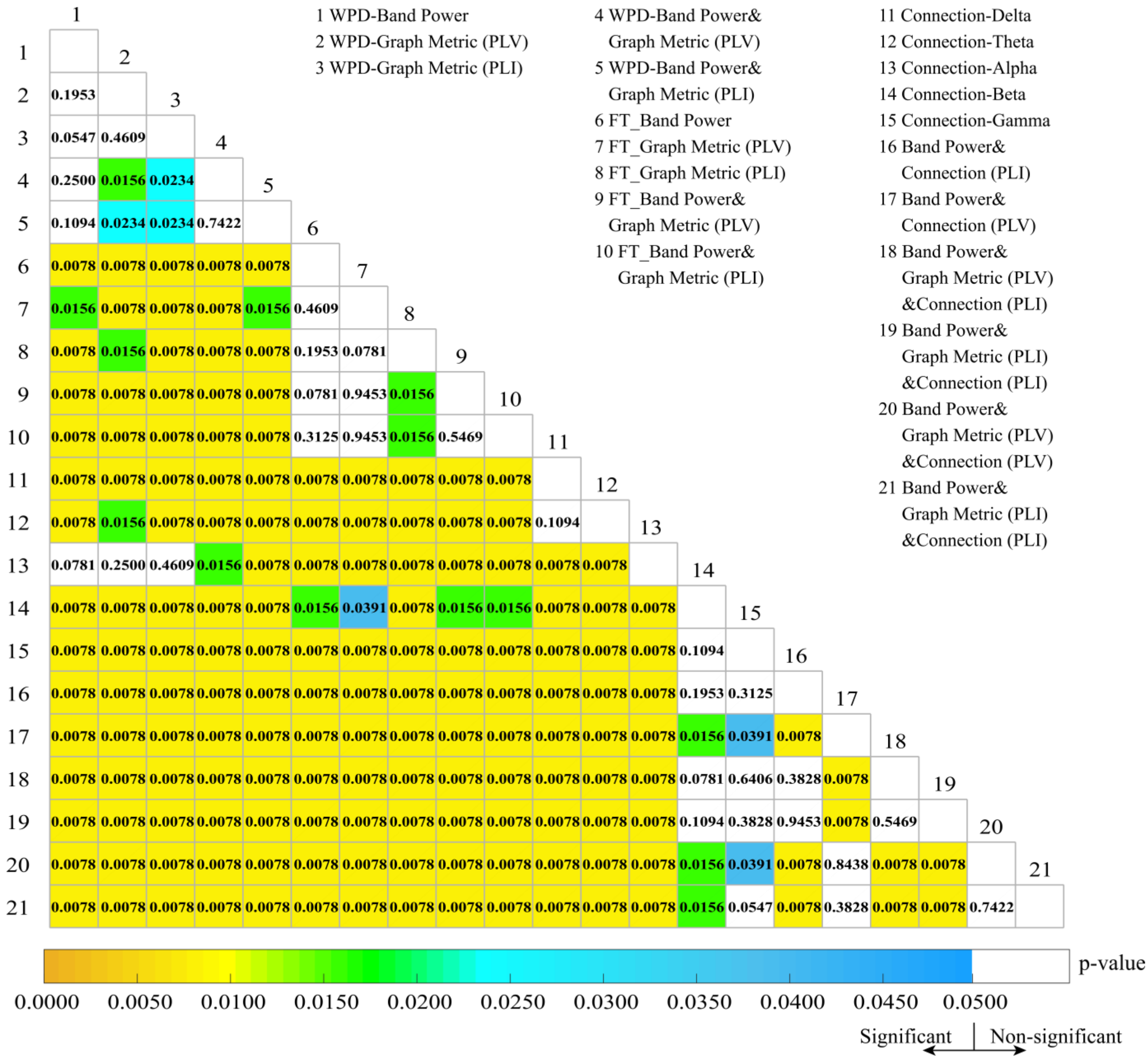

Fig. 5 The results of Wilcoxon signed rank test in performance comparisons between feature categories. Most of the compared cases were significantly different ($p < 0.05$).

According to the results of feature selection and fusion, F-score and SPE have different performances. F-score improved the classification accuracies for all cases, while SPE reduced the classification accuracies. F-score was better for the feature selection according to the obtained results. The advantage of F-score was also found in the Ren et al.'s study, showing the better performance compared to principal component analysis (PCA) [51]. Fig. 6 shows the average accuracies for

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2020.3019849, IEEE Transactions on Instrumentation and Measurement

different feature dimensions with F-Score and SPE. The results show that classification accuracies were increased quickly to a local peak and then slightly increased to a balanced level for most cases.

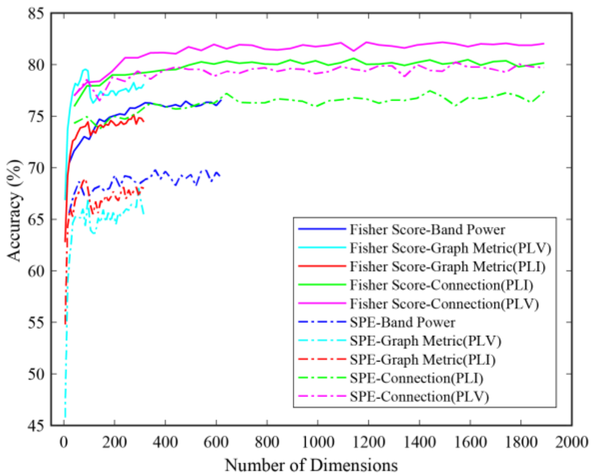

Fig. 6 The classification accuracies for different dimensions with F-Score and SPE.

Based on the current study, the SPE reduced classification accuracy, which was different from our previous results [52], indicating that the same method has different performance on the different classification tasks and different datasets. It is worth noting that the FT was better than WPD based on the results of this study, which is not in agreement with the findings in other studies. This might be due to the selection of wavelet since the wavelet dramatically affects the WPD performance. In our study, we did not explore all wavelets and selected the widely used wavelet (db4) according to previous research [31]. Therefore, the selected wavelet might not fit the data in this particular case.

This study demonstrated that workload classification was well improved using the fusion of power and functional connection features. Although the study was informative for the workload classification, there were a few limitations. First, this study constructed functional connections using PLV and PLI. Other methods such as Partial Directed Coherence (PDC) and Directed Transfer Function (DTF) [53] were not included in the study. Second, in this study, we did not discuss brain regions relevant to mental workload because the SPE compressed feature dimension as a whole, which did not enable us to trace relevant regions. Third, workload identification was not assessed in real-time. Therefore, the results reported in this paper could not reflect that derived in a real-time practical application. However, the majority of findings reported in this paper should be retained when converting to a practical application since the practical application is similar to the experiment to a large extent. Fourth, the repetition of tasks in our experiment might introduce learning effect on participants' behaviour of aircraft operation. This effect probably causes bias in the behaviour investigation, but its effect is not critical for the purposes of classification. In addition, the length of a session (a cycle of the low, medium, and high workload tasks) is only 6 minutes. The total time for the three sessions is 18 minutes. The duration is not long so that the learning effect should not be significant if any.

V. Conclusion

In summary, the current study designed an experiment of aircraft operation simulation to explore workload identification performance among frequency bands, different individual feature categories, different combinations of feature categories, and feature selection and fusion methods. The study had shown that using the connection features in the gamma band achieved the highest accuracy (81.72%) among individual features. The combination of band power features and connection features (gamma) outperformed individual feature categories, obtaining the classification accuracy of 83.12%. With feature selection using F-Score, the accuracy was further enhanced to be 83.47%. When the features of graph metric were fused with the features of band power and connection, the classification accuracy was reached to 84.34%. The results showed that feature selection and fusion gave a positive role in the multi-class workload classification.

References

[1] T. Kakizaki, "Relationship between EEG amplitude and subjective rating of task strain during performance of a calculating task," *European Journal of Applied Physiology*, vol. 53, no. 3, pp. 206-212, 1984.

[2] G. F. Wilson, "Operator functional state assessment for adaptive automation implementation," in *Biomonitoring for Physiological and Cognitive Performance during Military Operations, International Society for Optics and Photonics*, vol. 5797, pp. 100-104, 2005.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2020.3019849, IEEE Transactions on Instrumentation and Measurement

[3] Y. Yeh and C. D. Wickens, "Dissociation of Performance and Subjective Measures of Workload," *Human Factors*, vol. 30, no. 1, pp. 111-120, 1988.

[4] G. B. Reid and T. E. Nygren, "The Subjective Workload Assessment Technique: A Scaling Procedure for Measuring Mental Workload," *Advances in psychology*, vol. 52, pp. 185-218, 1988.

[5] H. Ayaz, P. A. Shewokis, S. C. Bunce, K. Izzetoglu, B. Willems, and B. Onaral, "Optical brain monitoring for operator training and mental workload assessment," *NeuroImage*, vol. 59, no. 1, pp. 36-47, 2012.

[6] Y. Wan, R. Xiao, D. Zhuang, Y. Lin, X. Xiao, and J. Song, "Influence of mental workload on detecting information varieties revealed by mismatch negativity during flight simulation," *International Journal of Industrial Ergonomics*, vol. 64, pp. 1-7, 2018.

[7] B. Wallace, F. Knoefel, R. Goubran, R. A. L. Zunini, Z. Ren, and A. Maccosham, "EEG/ERP: within episodic assessment framework for cognition, " *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 10, pp. 2525-2534, 2017.

[8] P. Arpaia, N. Moccaldi, R. Prevete, I. Sannino, and A. Tedesco, "A wearable EEG instrument for real-time frontal asymmetry monitoring in worker stress analysis, " *IEEE Transactions on Instrumentation and Measurement*, DOI: 10.1109/TIM.2020.2988744, 2020.

[9] K. Ryu and R. Myung, "Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic," *International Journal of Industrial Ergonomics*, vol. 35, no. 11, pp. 991-1009, 2005.

[10] J. B. Brookings, G. F. Wilson, and C. R. Swain, "Psychophysiological responses to changes in workload during simulated air traffic control," *Biological Psychology*, vol. 42, no. 3, pp. 361-377, 1996.

[11] M. A. Hogervorst, A. Brouwer, and J. B. F. Van Erp, "Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload," *Frontiers in Neuroscience*, vol. 8, pp. 322-322, 2014.

[12] H. Zhang, Y. Zhu, J. Maniyeri, and C. Guan, "Detection of variations in cognitive workload using multi-modality physiological sensors and a large margin unbiased regression machine," in *36th Annual International Conference of the IEEE, Engineering in Medicine and Biology Society*, pp.2985-2988, 2014.

[13] G. Borghini, L. Astolfi, G. Vecchiato, D. Mattia, and F. Babiloni, "Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness," *Neuroscience & Biobehavioral Reviews*, vol. 44, pp. 58-75, 2014.

[14] S. G. Lei, W. Sebastian, and R. Matthias, "Representation of driver's mental workload in EEG data," *Human Factors security and safety*, pp. 285-294, 2009.

[15] Z. Wang, R. M. Hope, Z. Wang, Q. Ji, and W. D. Gray, "Cross-subject workload classification with a hierarchical Bayes model," *NeuroImage*, vol. 59, no. 1, pp. 64-69, 2012.

[16] S. Chandra, K. L. Verma, G. Sharma, A. P. Mittal, and D. Jha, "EEG BASED COGNITIVE WORKLOAD CLASSIFICATION DURING NASA MATB-II MULTITASKING," *International Journal of Cognitive Research in Science, Engineering and Education*, vol. 3, no. 1, pp. 35-41, 2015.

[17] J. Gonzalez-Castillo, C. W. Hoy, D. A. Handwerker, M. E. Robinson, L. C. Buchanan, Z. S. Saad, and P. A. Bandettini, "Tracking ongoing cognition in individuals using brief, whole-brain functional connectivity patterns," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 28, pp. 8762-8767, 2015.

[18] A. Anticevic, G. Repovs, J. H. Krystal, and D. M. Barch, "A broken filter: prefrontal functional connectivity abnormalities in schizophrenia during working memory interference," *Schizophrenia Research*, vol. 141, no. 1, pp. 8-14, 2012.

[19] A. Mazaheri, S. Coffeycorina, G. R. Mangun, E. M. Bekker, A. S. Berry, and B. A. Corbett, "Functional Disconnection of Frontal Cortex and Visual Cortex in Attention-Deficit/Hyperactivity Disorder," *Biological Psychiatry*, vol. 67, no. 7, pp. 617-623, 2010.

[20] J. Li, J. Lim, Y. Chen, K. Wong, N. Thakor, A. Bezerianos, and Y. Sun, "Mid-Task Break Improves Global Integration of Functional Connectivity in Lower Alpha Band," *Frontiers in Human Neuroscience*, vol. 10, pp. 304, 2016.

[21] N. Sciaraffa, G. Borghini, P. Aricò, G. Di Flumeri, A. Colosimo, A. Bezerianos, N. V. Thakor, and F. Babiloni, "Brain Interaction during Cooperation: Evaluating Local Properties of Multiple-Brain Network," *Brain Sciences*, vol. 7, no. 7, pp. 1-13, 2017.

[22] R. Gupta, K. U. R. Laghari, and T. H. Falk, "Relevance vector classifier decision fusion and EEG graph-theoretic features for automatic affective state characterization," *Neurocomputing*, vol. 174, no. 174, pp. 875-884, 2016.

[23] J. Chen, B. Hu, L. Xu, P. Moore, and Y. Su, "Feature-level fusion of multimodal physiological signals for emotion recognition," in *IEEE International Conference on Bioinformatics and Biomedicine*, pp. 395-399, 2015.

[24] P. Bashivan, M. Yeasin, and G. M. Bidelman, "Single trial prediction of normal and excessive cognitive load through EEG feature fusion," in *IEEE signal processing in medicine and biology symposium*, pp. 1-5, 2015.

[25] S. Mathan, A. Smart, T. Ververs, and M. Feuerstein, "Towards an index of cognitive efficacy EEG-based estimation of cognitive load among individuals experiencing cancer-related cognitive decline," in *international conference of the ieee engineering in medicine and biology society*, pp. 6595-6598, 2010.

[26] C. L. Baldwin and B. N. Penaranda, "Adaptive training using an artificial neural network and EEG metrics for within- and cross-task workload classification," *NeuroImage*, vol. 59, no. 1, pp. 48-56, 2012.

[27] Y. T. Wang, K. C. Huang, C. S. Wei, T. Y. Huang, L. W. Ko, C. T. Lin, C. K. Cheng, and T. P. Jung, "Developing an EEG-based on-line closed-loop lapse detection and mitigation system," *Frontiers in Neuroscience*, vol. 8, pp. 321-321, 2014.

[28] D. De Massari, D. Pacheco, R. Malekshahi, A. Betella, P. F. Verschure, N. Birbaumer, and A. Caria, "Fast mental states decoding in mixed reality," *Frontiers in Behavioral Neuroscience*, vol. 8, pp. 415-415, 2014.

[29] M. Schultzekraft, S. Dahne, M. F. Gugler, G. Curio, and B. Blankertz, "Unsupervised classification of operator workload from brain signals," *Journal of Neural Engineering*, vol. 13, no. 3, pp. 036008-036008, 2016.

[30] G. N. Dimitrakopoulos, I. Kakkos, Z. Dai, J. Lim, J. J. deSouza, A. Bezerianos, and Y. Sun, "Task-Independent Mental Workload Classification Based Upon Common Multiband EEG Cortical Connectivity," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 11, pp. 1940-1949, 2017.

[31] L. Chen, Y. Zhao, J. Zhang, and J. Zou, "Automatic detection of alertness/drowsiness from physiological signals using wavelet-based nonlinear features and machine learning," *Expert Systems with Applications*, vol. 42, no. 21, pp. 7344-7355, 2015.

[32] S. Achard and E. T. Bullmore, "Efficiency and Cost of Economical Brain Functional Networks," *PLOS Computational Biology*, vol. 3, no. 2, pp. e17, 2007.

[33] J. Li, R. Romero-Garcia, J. Suckling, and L. Feng, "Habitual tea drinking modulates brain efficiency: evidence from brain connectivity evaluation," *Aging*, vol. 11, no. 11, pp. 3876-3890, 2019.

[34] P. Erdős and A. Renyi, "On the strength of connectedness of a random graph," *Acta Mathematica Hungarica*, vol. 12, pp. 261-267, 1961.

[35] Y. He, Z. Chen, and A. C. Evans, "Structural Insights into Aberrant Topological Patterns of Large-Scale Cortical Networks in Alzheimer's

Disease," *The Journal of Neuroscience*, vol. 28, no. 18, pp. 4756-4766, 2008.

[36] S. Achard and E. T. Bullmore, "Efficiency and Cost of Economical Brain Functional Networks," *PLOS Computational Biology*, vol. 3, no. 2, pp. e17, 2007.

[37] D. S. Bassett, E. T. Bullmore, B. A. Verchinski, V. S. Mattay, D. R. Weinberger, and A. Meyerlindenberg, "Hierarchical Organization of Human Cortical Networks in Health and Schizophrenia," *The Journal of Neuroscience*, vol. 28, no. 37, pp. 9239-9248, 2008.

[38] Y. W. Chen, and C. J. Lin, "Combining SVMs with various feature selection strategies," in *Feature extraction, Berlin Heidelberg: Springer*, pp. 315-324, 2006.

[39] D. K. Agrafiotis, "Stochastic proximity embedding," *Journal of Computational Chemistry*, vol. 24, no. 10, pp. 1215-1221, 2003.

[40] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5-32, 2001.

[41] D. Novak, M. Matjaž, and M. Marko, "A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing," *Interacting with computers*, vol. 24, no. 3, pp. 154-172, 2012.

[42] Z. Dai, B. Anastasios, S. H. Annabel, and Y. Sun, "Mental workload classification in n-back tasks based on single-trial EEG," *Chinese Journal of Scientific Instrument*, vol. 38, no. 6, pp. 1335-1344, 2017.

[43] R. D. Traub, J. G. R. Jefferys, and M. A. Whittington, "Simulation of gamma rhythms in networks of interneurons and pyramidal cells," *Journal of Computational Neuroscience*, vol. 4, no. 2, pp. 141-150, 1997.

[44] A. Von Stein and J. Sarnthein, "Different frequencies for different scales of cortical integration: from local gamma to long range alpha/theta synchronization," *International Journal of Psychophysiology*, vol. 38, no. 3, pp. 301-313, 2000.

[45] P. Fries, S. René, and O. Robert, "Finding gamma," *Neuron*, vol.58, no. 3, pp. 303-305, 2018.

[46] C. Tallonbaudry and O. F. Bertrand, "Oscillatory gamma activity in humans and its role in object representation," *Trends in Cognitive Sciences*, vol. 3, no. 4, pp. 151-162, 1999.

[47] C. L. Baldwin and B. N. Penaranda, "Adaptive training using an artificial neural network and EEG metrics for within- and cross-task workload classification," *NeuroImage*, vol. 59, no. 1, pp. 48-56, 2012.

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

[49] G. Borghini, L. Astolfi, G. Vecchiato, D. Mattia, and F. Babiloni, "Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness," *Neuroscience & Biobehavioral Reviews*, vol. 44, pp. 58-75, 2014.

[50] P. D. Antonenko, F. Paas, R. H. Grabner, and T. Van Gog, "Using Electroencephalography to Measure Cognitive Load," *Educational Psychology Review*, vol. 22, no. 4, pp. 425-438, 2010.

[51] W. Ren, M. Han, J. Wang, D. Wang, and T. Li, "Efficient feature extraction framework for EEG signals classification," in *international conference on intelligent control and information processing*, pp. 167-172, 2016.

[52] J. Harvy, E. Sigalas, N. V. Thakor, A. Bezerianos, and J. Li, "Performance Improvement of Driving Fatigue Identification Based on Power Spectra and Connectivity Using Feature Level and Decision Level Fusions," in *40th Annual international conference of the ieee engineering in medicine and biology society*, pp. 102-105, 2018.

[53] A. M. Bastos and J. Schoffelen, "A Tutorial Review of Functional Connectivity Analysis Methods and Their Interpretational Pitfalls," *Frontiers in Systems Neuroscience*, vol. 9, pp. 175-175, 2016.

Zian Pei (Student Member, IEEE) was born in Hebei, China, in 1996. He received the bachelor's degree in Building Electricity and Intelligent from Hebei University of Architecture, China, in 2018. He studied for a master's degree in pattern recognition and intelligent systems in Wuyi University, Jiangmen, China. He won P300-based BCI competition in the recent BCI Controlled Robot Contest of 2019 World Robot Conference held in Beijing. His current research interests include the fields of Brain-like Computation, Pattern Recognition, and Hybrid Intelligence.

Hongtao Wang (Member, IEEE) received the Ph.D. degree in pattern recognition and intelligent systems from the South China University of Technology, in 2015.

He is currently a Full Professor and Vice Dean with the Faculty of Intelligent Manufacturing, Wuyi University, and been selected as a Thousand-Hundred-Ten Talent of Universities in Guangdong Province. He is also the Director of the Jiangmen Brain-like Computation and Hybrid Intelligence Research and Development Center. From January 2017 to January 2019, he was a Visiting Research Fellow at the National University of Singapore. His current research interests include the fields of brain-like computation, pattern recognition, deep learning, and hybrid intelligence.

Anastasios (Tassos) Bezerianos is the head of Cognitive Engineering (COGEN) Laboratory at the N.1 Health Institute, National University of Singapore, and a Visiting Professor in the Computer Science Department of New South Wales University (NSWU), Canberra, AUS. He has been the Professor of Medical Physics at the Medical School of Patras University, Patras Greece since 2004. He studied Physics at Patras University and Telecommunications at Athens University, and he received his Ph.D. on Bioengineering from the University of Patras. His research entails diverse areas spanning from Artificial Intelligence and Robotics to Biomedical Signal Processing and Brain Imaging as well as Mathematical Biology and Systems Medicine and Bioinformatics. His work is summarized in 154 journal and 235 conference proceedings publications 1 book and 2 patents. He has research collaborations with research institutes and universities in Japan, China, and Europe, and he is Associate Editor of IEEE TNSRE and PLOS ONE Neuroscience journals and a reviewer for several international scientific journals. He is a Fellow of The European Alliance for Medical and biological Engineering &Science (EAMBES) and Senior Member of IEEE and Founder and Chairman of the biannual International Summer School on Emerging Technologies in Biomedicine.

Junhua Li (S'12-M'13-SM'18) received his PhD from the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China, in 2013.

He is currently a Lecturer in the School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK. He was a Senior Research Fellow at the National University of Singapore, Singapore, before moving to the UK. His research interests include computational neuroscience, brain-computer interface, machine learning, neurophysiological data analytics, and their practical applications.

He is an Associate Editor for IEEE Access, a Review Editor for Frontiers in Human Neuroscience, and served as a Guest Associate Editor for several journal special issues related to his research interests.