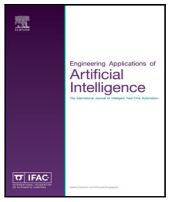
Engineering Applications of Artificial Intelligence 123 (2023) 106237

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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



# An enhanced ensemble deep random vector functional link network for driver fatigue recognition

Ruilin Li [a], Ruobin Gao [b], Liqiang Yuan [a], P.N. Suganthan [c,a,*], Lipo Wang [a], Olga Sourina [d]

[a] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
[b] School of Civil and Environmental Engineering, Nanyang Technological University, Singapore
[c] KINDI Center for Computing Research, College of Engineering, Qatar University, Doha, Qatar
[d] Fraunhofer, Nanyang Technological University, Singapore

ARTICLE INFO

*Keywords:*
Electroencephalogram (EEG)
Ensemble deep random vector functional link (edRVFL)
Feature selection
Dynamic ensemble
Cross-subject driver fatigue recognition

ABSTRACT

This work investigated the use of an ensemble deep random vector functional link (edRVFL) network for electroencephalogram (EEG)-based driver fatigue recognition. Against the low feature learning capability of the edRVFL network from raw EEG signals, two strategies were exploited in this work. Specifically, the first one was to exploit the advantages of the feature extractor module in CNNs, *i.e.*, use CNN features as the input of the edRVFL network. The second one was to improve the feature learning capability of the edRVFL network. An enhanced edRFVL network named FGloWD-edRVFL was proposed, in which four enhancements were implemented, including random forest-based **F**eature selection, **Glo**bal output layer, **W**eighting and entropy-based **D**ynamic ensemble. The proposed FGloWD-edRVFL network was evaluated on the challenging cross-subject driver fatigue recognition tasks. The results indicated that the proposed model could boost the recognition performance, significantly outperforming all strong baselines. The step-wise analysis further demonstrated the effectiveness of the proposed enhancements in the edRVFL network.

## 1. Introduction

Driving fatigue is a physiological phenomenon in which the driver's attention and alertness decrease after a long driving task (Wang et al., 2021). It is one of the main contributors to crashes and casualties (Federico et al., 2005). According to the study of Byeon (2020), driver fatigue caused at least 10%–15% of traffic accidents. For road safety, it is necessary to develop effective and efficient driver fatigue monitoring systems. In recent years, electroencephalogram (EEG), which can reflect human brain activities, has been widely used for driver fatigue detection (Koay et al., 2022; Liu et al., 2022). For the purpose of practical use, it is essential to establish a calibration-free driver fatigue recognition system with robust generalization capability and high performance. However, EEG has the disadvantages of non-stationary, high-level complexity, and subject variability. Consequently, the performance of cross-subject EEG-based classification is limited, and further investigation is still required to improve the performance of calibration-free EEG decoding in driver fatigue recognition tasks.

There are mainly two categories in the literature to perform EEG-based classification tasks. The first way is to perform hand-engineering feature extraction based on the prior knowledge of the signals, followed by classifier training. Many studies have analyzed, processed, and classified EEG-based tasks through artificially designed features such as spatio-spectral analysis (Bang et al., 2022) and other nonlinear methods. Particularly for fatigue recognition, some hand-engineering features have been investigated as well. For instance, Gao et al. (2020) exploited the traditional differential entropy and power spectral density (PSD) features to represent the characteristics associated with driver fatigue states. A coincidence-filtering-based convolutional neural network (CNN) was proposed to classify the extracted features. Fan et al. (2022) extracted numerous features including energy, entropy, rhythmic energy ratio and frontal asymmetry ratio from the EEG signals to study the use of forehead EEG for fatigue recognition. Zhang et al. (2023) extracted statistical features, PSD and entropy-related features from the decomposed EEG signals. Feature selection and feature fusion were then performed to achieve a good performance of fatigue recognition. Tuncer et al. (2021b) used binary pattern (BiPa) and statistical features of the one-dimensional discrete wavelet transform. ReliefF and neighborhood component analysis were exploited to perform feature selection. Furthermore, a dynamic center based BiPa (DCBiPa) was also proposed by Tuncer et al. (2021a). A novel textural feature extractor module was developed based on DCBiPa and a multi-threshold ternary pattern. Although the great fatigue recognition performance was presented by exploiting the hand-engineering features, the extraction of

* Corresponding author at: KINDI Center for Computing Research, College of Engineering, Qatar University, Doha, Qatar.
*E-mail addresses:* ruilin001@e.ntu.edu.sg (R. Li), gaor0009@e.ntu.edu.sg (R. Gao), liqiang002@e.ntu.edu.sg (L. Yuan), p.n.suganthan@qu.edu.qa (P.N. Suganthan), ELPWang@ntu.edu.sg (L. Wang), EOSourina@ntu.edu.sg (O. Sourina).

https://doi.org/10.1016/j.engappai.2023.106237
Received 1 December 2022; Received in revised form 9 March 2023; Accepted 27 March 2023
Available online 8 April 2023

the features and the possible feature selection followed could have a high time cost for EEG decoding. Besides, the hand-engineering features may only cover partial information and could not comprehensively reflect the human states of alert or fatigue.

With the success of deep learning in image processing, the use of deep neural networks (DNNs) for end-to-end EEG decoding has also shown great improvement compared with the conventional hand-engineering feature-based models. The automatic feature extraction from raw data provided by DNNs highlights its advantage of cost-saving. Moreover, the better learning ability of deep learning helps to extract more distinctive features. Among the different architectures of DNNs, deep CNNs have made an outstanding performance in learning from EEG signals automatically. Schirrmeister et al. (2017) first studied the design and training of CNN to classify raw EEG data. Then, a shallow and a deep CNN architecture were proposed. Better performance was obtained on motor imagery (MI) classification as compared with the filter-bank common spatial pattern-based model. Besides MI application, researchers also investigated different CNN architectures for various EEG-based recognition tasks. For instance, Thuwajit et al. (2022) proposed a multi-scale CNN, EEGWaveNet, to address epileptic seizure detection. Khare and Bajaj (2021) transformed the EEG signals into the time–frequency domain and exploited a CNN to perform emotion recognition. However, these models were only analyzed in specific domains. In 2017, Lawhern et al. (2018) proposed a compact CNN named EEGNet and performed experiments on different brain-computer interface paradigms. Comparable results were achieved in subject-dependent and subject-independent settings. Recently, Cui et al. (2022) proposed an InterpretableCNN (ICNN) which performed spatial and temporal convolution operations. The model was superior to previous models such as EEGNet in drowsiness classification based on EEG signals. Furthermore, Li et al. (2023) proposed a hybrid decomposition-based ensemble CNN framework. This framework was to deal with the high complexity of EEG signals. The ensemble CNN model could effectively decode the decomposed EEG signals and achieve better performance than ICNN. However, almost all existing deep learning-based EEG decoding methods rely heavily on conventional back-propagation (BP) to train the model. Consequently, the CNN models may suffer from the local minima problem.

Randomized neural networks (RNNs) can remedy the problem mentioned above by randomly fixing the network configurations and computing the closed-form solution in the output layers. Random vector functional link (RVFL), proposed by Pao and Takefuji (1992), has gained substantial attention due to its simplicity and outstanding performance in different fields (Elsheikh et al., 2021a,b, 2023). RVFL is proven as a universal approximator for continuous functions with bounded finite dimensions and closed-form solutions (Igelnik and Pao, 1995). In RVFL, the weights and biases of the hidden layer are randomly generated; The direct link propagates the input data directly to the output layer. This direct link serves as a form of regularization for the randomization (Zhang and Suganthan, 2016a) and helps to keep the model's complexity low (Shi et al., 2021). We refer the readers to the recent survey (Malik et al., 2022a) for more information on this topic. To boost the performance of the RVFL network, the ensemble deep variant of the RVFL network (Shi et al., 2021) was further proposed. The input features can be mapped to different feature spaces by stacking the hidden layers. The diverse features are used in ensemble learning to improve the classification performance further. This ensemble deep RVFL (edRVFL) has shown its superiority in various applications such as modeling ultrasonic welding of polymers (Elsheikh et al., 2022).

Although RVFL (Pao and Takefuji, 1992) is one of the oldest randomized single-hidden layer feedforward network, researchers have also investigated other RNN models such as extreme learning machine (ELM) (Huang et al., 2006), stochastic configuration network (SCN) (Wang and Li, 2017) and broad learning system (BLS) (Chen and Liu, 2018). ELM (Huang et al., 2006) is a variant of the RVFL without bias and direct connections. ELM has been widely used in various applications such as online sequential learning (Cao et al., 2021b). Several variants of ELM have been proposed. For example, Zou et al. (2022a) proposed a BP-ELM that could automatically assign better parameters for each coming hidden neuron by applying the BP of iterative least squares. However, recent comparative evaluations (Suganthan and Katuwal, 2021; Zhang et al., 2019) have shown that RVFL variants with direct links, in general, outperformed comparable ELM variants.

SCN (Wang and Li, 2017) is a type of RNN generated by a stochastic configuration algorithm. Cao et al. (2021a) proposed a bidirectional SCN, which proposed backward learning to configure the added neurons based on the residual error feedback. However, recent research (Hu and Suganthan, 2022) has shown that incorporating direct links to SCN improved its performance and employing Bayesian optimization to perform hyper-parameter tuning was superior to the proposed stochastic configuration algorithm. BLS (Chen and Liu, 2018) is another RNN with dense connections, which performs the broad expansion in both the feature layer and the enhancement layer. BLS is a variant of a flat neural network proposed by Pao et al. (1992). Zou et al. (2022b) proposed an improved BLS with the driving amount and optimization solution. An iterative least squares method was used to avoid the selection of the regularization parameter. Chen and Liu (2018) presented that BLS underperformed ELM variants in terms of classification accuracy on MNIST dataset as demonstrated in Table 1 of their work. Hu et al. (2022) showed that BLS underperformed RVFL variants based on pairwise statistical comparison on UCI datasets as demonstrated in Table 6 of their work. In this work, the state-of-the-art (SOTA) RVFL variant, edRVFL, was used for EEG-based cross-subject driver fatigue recognition.

In EEG processing, previous works usually employed the hand-engineering features as the input of RNNs (Yin and Zhang, 2018). However, for time-series data, RNNs lack the strong ability to extract beneficial features because of the simple architecture and the random mapping of the enhancement nodes (Cheng et al., 2021). Consequently, if the input was raw EEG signals which are high-dimensional and high-complexity, the low feature learning capability of the RNNs could deteriorate the classification performance. In this work, based on the edRVFL network, two strategies were employed to deal with this problem.

The first one was to exploit the advantages of CNNs to extract distinctive and representative features from raw EEG signals and to use them as the input of the edRVFL network. This was inspired by the work of Niu and Suen (2012), which presented that the hybrid CNN-support vector machine (SVM) model outperformed each individual classifier by exploiting the strong feature extraction ability of the CNN model and the global optimum solution of SVM. Recently, Cheng et al. (2021) also showed that the use of edRVFL was beneficial in boosting the performance of the CNN model. This type of hybrid system has also been used in different applications to help boost performance. For example, transfer learning is a popular way to generate image representations in medical image analysis. Lu et al. (2021b) exploited transfer learning to generate image representations by fine-tuning a pre-trained deep CNN model on the target small-sized dataset. Their CNN-ELM hybrid model achieved good performance in Covid-19 detection. Lu et al. (2021a) also exploited a hybrid AlexNet-ELM model for the detection of abnormal brain. A fine-tuned AlexNet was used to generate satisfactory image representations. Overall, the hybrid structure can compensate for the limitation of the CNN and edRVFL by incorporating the merits of both classifiers. Therefore, we employed this structure to boost the performance of both the CNN model and the edRVFL network in this work.

The second strategy is to enhance the feature learning capability of the edRVFL network. Specifically, four enhancements to the edRVFL network were proposed. Firstly, we proposed to exploit a random forest (RF) (Ho, 1995)-based feature selection to reduce the effects of noisy and redundant information that may be contained in the random features. The trivial features and less useful information can be

filtered out by feature selection such that the more beneficial features can be used for classification. Secondly, the weighting technique was employed in the edRVFL network, which was beneficial to diversify the trained classifiers in ensemble learning and boost the classification performance. Thirdly, a global output layer was added which used a global state that combined the features of all enhancement layers as the input. This contributed the global combination to the edRVFL network, which was neglected in the output layers previously. This further increased the diversity in ensemble learning. Lastly, instead of using a static ensemble module, we proposed to utilize an entropy-based dynamic ensemble to integrate the outputs of all output layers, including dynamic selection and dynamic weighting for each testing sample. With all the enhancements proposed, this enhanced edRVFL was named as FGloWD-edRVFL which included **F**eature selection, **Glo**bal output layer, **W**eighting and entropy-based **D**ynamic ensemble in the edRVFL network.

The contributions of this work can be summarized as follows:

- This work investigated the use of the hybrid CNN-edRVFL model for driver fatigue recognition. The merits of both CNN and the edRVFL networks were exploited to achieve better performance than the SOTA models.
- Four enhancements for the edRVFL network were further proposed, including (1) two techniques that can improve the exploitation efficiency of randomized hidden features: RF-based Feature selection and Global output layer; and (2) two techniques that can boost the generalization capability of ensemble learning: Weighting and entropy-based Dynamic ensemble.

The remaining sections of this paper are organized as follows: Section 2 introduces the proposed FGloWD-edRVFL in detail. Section 3 presents information about the datasets, comparison results, and ablation study. Then, the advantages of the proposed model are discussed in Section 4. Finally, conclusions are drawn in Section 5.

## 2. Methodology

The architecture of the proposed FGloWD-edRVFL is shown in Fig. 1. The four proposed enhancements for the edRVFL network, including RF-based feature selection, weighting, global output layer and entropy-based dynamic ensemble, are introduced in this section. Following that, the overall pipeline of the FGloWD-edRVFL input with CNN features is presented.

### 2.1. Ensemble deep random vector functional link with feature selection

The edRVFL network (Shi et al., 2021) is a deep feed-forward RNN in which the hidden layers are randomized and frozen during training. The input space of the learning task is denoted as $\boldsymbol{X} \in \mathbb{R}^{K\times d}$ (*i.e.*, extracted EEG features), the output space is denoted as $\boldsymbol{Y} \in \mathbb{R}^{K\times C}$ (*i.e.*, categories of EEG signals), where $K$, $d$ and $C$ represent the number of samples, feature dimension and categories, respectively. For an edRVFL network that has $L$ hidden layers and $N^l$ hidden nodes in the $l$th layer, the hidden features of the $l$th layer are calculated by Eq. (1).

$$\boldsymbol{H}^l = \begin{cases} \tanh(\boldsymbol{X}\boldsymbol{W}^1), & if\ l = 1, \\ \tanh(\boldsymbol{D}^{l-1}\boldsymbol{W}^l), & if\ l > 1, \end{cases} \tag{1}$$

where $\tanh(\cdot)$ is the used nonlinear activation function in this work. The term $\boldsymbol{D}^l$ is the concatenation of the input data (direct link) and the hidden features in the $l$th layer, which is represented as Eq. (2).

$$\boldsymbol{D}^l = [\boldsymbol{H}^l\ \boldsymbol{X}]. \tag{2}$$

There are $L$ output layers and the classification in each output layer is performed independently by using the calculated enhancement features $\boldsymbol{D}^l$ as the input. This is elaborated by Eq. (3).

$$Output = \boldsymbol{D}^l\boldsymbol{\beta}^l, \tag{3}$$

where $\boldsymbol{\beta}^l$ represents the weights of the $l$th output layer. In this work, a regularized least square method, Ridge regression, was exploited to solve the weights of the output layers. Specifically, the output weights $\boldsymbol{\beta}^l$ of Ridge regression are given by Eqs. (4) and (5).

$$\text{prime space: } \boldsymbol{\beta}^l = [(\boldsymbol{D}^l)^T\boldsymbol{D}^l + \lambda^l\boldsymbol{I}]^{-1}(\boldsymbol{D}^l)^T\boldsymbol{Y}, \tag{4}$$

$$\text{dual space: } \boldsymbol{\beta}^l = (\boldsymbol{D}^l)^T[\boldsymbol{D}^l(\boldsymbol{D}^l)^T + \lambda^l\boldsymbol{I}]^{-1}\boldsymbol{Y}, \tag{5}$$

where $\lambda^l$ denotes the regularization parameter in the $l$th layer. Lastly, ensemble learning integrates $L$ different outputs. Specifically, in the original edRVFL network (Shi et al., 2021), the mean of $L$ outputs is taken as the final output of the classification.

Although the hierarchical representations enlarge the space of random features, the noise generated by the nature of randomization propagates from the bottom to the top layer in the original edRVFL network. These inferior features may deteriorate performance eventually. To reduce the effect of noisy and redundant features, this work proposed to implement a RF-based feature selection in the edRVFL network due to its excellent performance for classification, fast computation and great interpretability. Specifically, the importance of the enhancement features $\boldsymbol{D}^l$ was measured in terms of their contributions to decrease the overall impurity. The criterion Gini index was exploited as the impurity function in this work. For the $i$th enhancement feature ($i \in [1, d+N^l]$) that is evaluated at the node $n$ of the $t$th tree of the RF, the impurity decrease is calculated by Eq. (6).

$$\alpha^t_{i,n} = G_n - G_e - G_r, \tag{6}$$

where $G_n$ represents the Gini index calculated at the node $n$. $G_r$ and $G_e$ represent the Gini index for the resulting right and left children nodes of the node $n$, respectively. Then, the overall feature importance of the $i$th enhancement feature in RF is calculated by Eq. (7).

$$\alpha_i = \frac{\sum_t^T \alpha^t_{i,n}}{\sum_{i=1}^{d+N}\sum_t^T \alpha^t_{i,n}}. \tag{7}$$

The numerator represents the impurity decrease of the $i$th enhancement feature in RF. The denominator means the sum of the impurity decreases of all enhancement features in RF. The term $\alpha_i$ represents the normalized impurity decrease of the $i$th enhancement feature. In the proposed edRVFL network with feature selection, a hyper-parameter $m^l$ was set to define the number of selected features. Specifically, $m^l$ features with higher feature importance were selected and then fed into the $l$th output layer. The enhancement features obtained after feature selection are denoted as $\boldsymbol{D}^l_{fs}$.

### 2.2. Weighting

The weighting algorithm is a promising method to improve the recognition performance in ensemble learning. It has been used in boosting algorithms (Hu et al., 2022; Freund and Schapire, 1997), which can force a model to better learn the hard samples during training. To take advantage of ensemble learning in the edRVFL network, the original edRVFL network was improved by performing weighting to process the enhancement features. The graphical flowchart is presented in Fig. 2. Specifically, the weights for all samples were initially set as 1 in the first output layer during training. After training, the prediction results of the training data were used to identify the wrongly classified samples which were then denoted as the hard samples. Following that, for the training in the second output layer, a new weight was assigned to the hard samples identified in the previous layer. This process was iterated for all the following layers. For simplicity, the weight computation in the $l$th hidden layer is described as follows. Assuming that the number of samples classified correctly and wrongly is $C_r$ and $C_w$, respectively, we have

$$C_r + C_w = K. \tag{8}$$

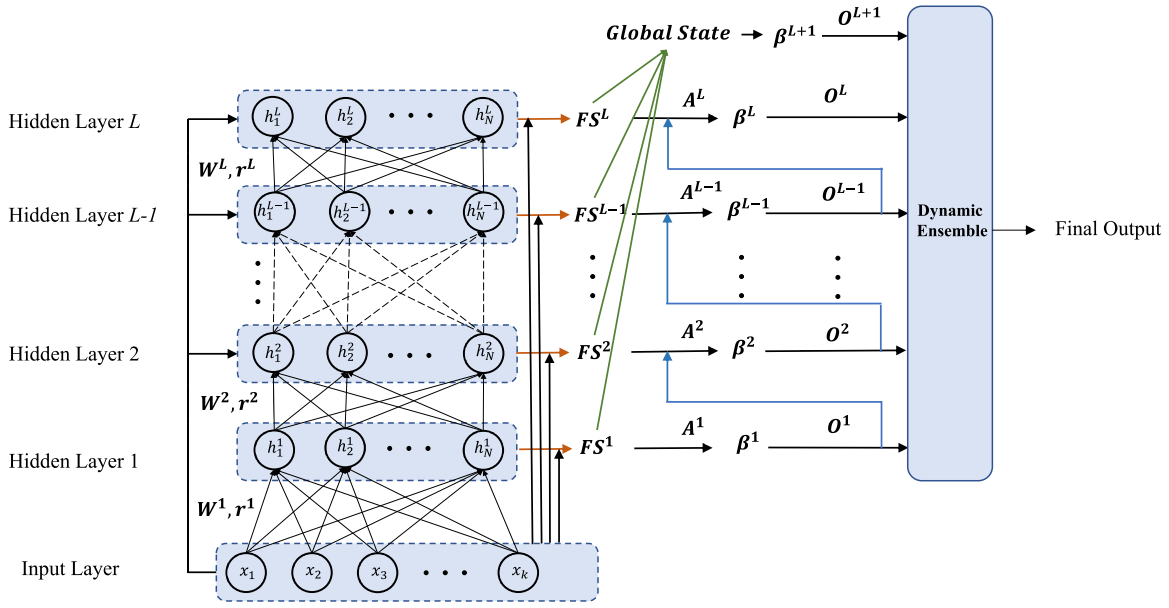



**Fig. 1.** Architecture of the FGloWD-edRVFL network. $x_k$ represents the input data. $h_N^l$ represents the $N$th hidden neuron in the $l$th hidden layer. The term $\boldsymbol{W}^l$ represents the randomly initialized hidden weights in the $l$th hidden layer. $\boldsymbol{r}^l$ represents the randomly initialized bias in the $l$th hidden layer. $\boldsymbol{FS}^l$ represents the RF-based feature selection in the $l$th output layer. $\boldsymbol{A}^l$ represents the weight matrix of the $l$th output layer. $\boldsymbol{\beta}^l$ represents the output weights of the $l$th output layer. $o^l$ represents the output probabilities of the $l$th output layer. The orange, blue and green lines represent the processes of feature selection, weighting and the combination of the local features for the global state, respectively.

The weight for the correctly classified samples $\theta_r$ and the weight for the wrongly classified samples $\theta_w$ obey the rule that the size of the samples is kept the same, which is represented as Eq. (9).

$$\theta_r C_r + \theta_w C_w = K. \tag{9}$$

Then, by setting the weight for the wrongly classified samples $\theta_w$ as the hyper-parameter, the weight of the correctly classified samples in the weighting scheme is represented as

$$\theta_r = \frac{K - C_w \theta_w}{C_r}. \tag{10}$$

To calculate the output weights $\boldsymbol{\beta}^l$, Eqs. (4) and (5) are rewritten as

$$\text{prime space: } \boldsymbol{\beta}^l = [(\boldsymbol{D}_{\boldsymbol{A}}^l)^T \boldsymbol{D}_{\boldsymbol{A}}^l + \lambda^l \boldsymbol{I}]^{-1} (\boldsymbol{D}_{\boldsymbol{A}}^l)^T \boldsymbol{Y}, \tag{11}$$

$$\text{dual space: } \boldsymbol{\beta}^l = (\boldsymbol{D}_{\boldsymbol{A}}^l)^T [\boldsymbol{D}_{\boldsymbol{A}}^l (\boldsymbol{D}_{\boldsymbol{A}}^l)^T + \lambda^l \boldsymbol{I}]^{-1} \boldsymbol{Y}, \tag{12}$$

where $\boldsymbol{D}_{\boldsymbol{A}}^l$ represents the enhancement features by multiplying the selected features and a weight matrix, which is represented as

$$\boldsymbol{D}_{\boldsymbol{A}}^l = \boldsymbol{A}^l \boldsymbol{D}_{fs}^l, \tag{13}$$

where $\boldsymbol{A}^l$ is the diagonal $(K \times K)$ weight matrix in the $l$th output layer that includes the weights for all samples.

### 2.3. Global output layer

In the original edRVFL network, the local information of each layer is exploited to compute the output weights. However, the model may neglect the function of combining different levels of features. Therefore, it was proposed to add a global output layer, of which the input was a global state that combined the features of all hidden layers obtained after feature selection. Moreover, the global state-based output layer with a different regularization parameter can increase the diversity of the obtained classifiers for the final ensemble learning. The regularization parameter in the global output layer is denoted as $\lambda_g$.

Specifically, the global state is the combination of local features obtained after feature selection, which is represented as Eq. (14).

$$\boldsymbol{D}_g = [\boldsymbol{D}_{fs}^1 \ \boldsymbol{D}_{fs}^2 \ \cdots \ \boldsymbol{D}_{fs}^l \ \cdots \ \boldsymbol{D}_{fs}^L]. \tag{14}$$

It has to be noted that the global state should be constructed after feature selection such that the global state is the combination of the beneficial enhancement features from the output layers and has proper dimensions for output weights computing. Hence, the classifier in the global output layer can positively contribute to the final ensemble learning. Otherwise, it may lead to overfitting if the model has over-high dimensional input features for training.

### 2.4. Entropy-based dynamic ensemble

For ensemble learning in the original edRVFL network, not every classifier of the output layers is an expert in classifying all unknown samples (*i.e.*, testing samples) (Cruz et al., 2018). In other words, low-competence classifiers may deteriorate the performance of ensemble learning. Therefore, in this work, a dynamic ensemble was performed that included a dynamic selection of high-competence classifiers and a dynamic weighting of the classifiers' outputs in the block of ensemble learning. The entropy of the testing samples was used as the criteria to estimate the competence level of a classifier. The two steps of the proposed dynamic ensemble are explained in the following.

Firstly, for the dynamic selection, suppose that the output probabilities of the classifier $b$ for the $j$th testing sample are represented as

$$\boldsymbol{P}_j^b = \{p_{j,1}^b, p_{j,2}^b, \ldots, p_{j,c}^b, \ldots, p_{j,C}^b\}, \tag{15}$$

where the number of $b \in [0, L+1]$, the entropy of this classifier $b$ input with the $j$th sample is formulated as

$$S_j^b = \sum_{c=1}^{C} -p_{j,c}^b \ln p_{j,c}^b. \tag{16}$$

A lower entropy output represents a higher confidence of the corresponding model to classify the testing sample correctly. Therefore, the competence level of the classifier $b$ for the $j$th testing sample is defined as $1 - S_j^b$.

A set of classifiers $\boldsymbol{B}$ with higher competence levels is selected. The number of classifiers selected $q$ is set as a hyper-parameter.

For the dynamic weighting in this entropy-based dynamic ensemble, for each testing sample, the estimated competence level of each selected classifier was exploited to weight the output probability. The weight $z_j^b$

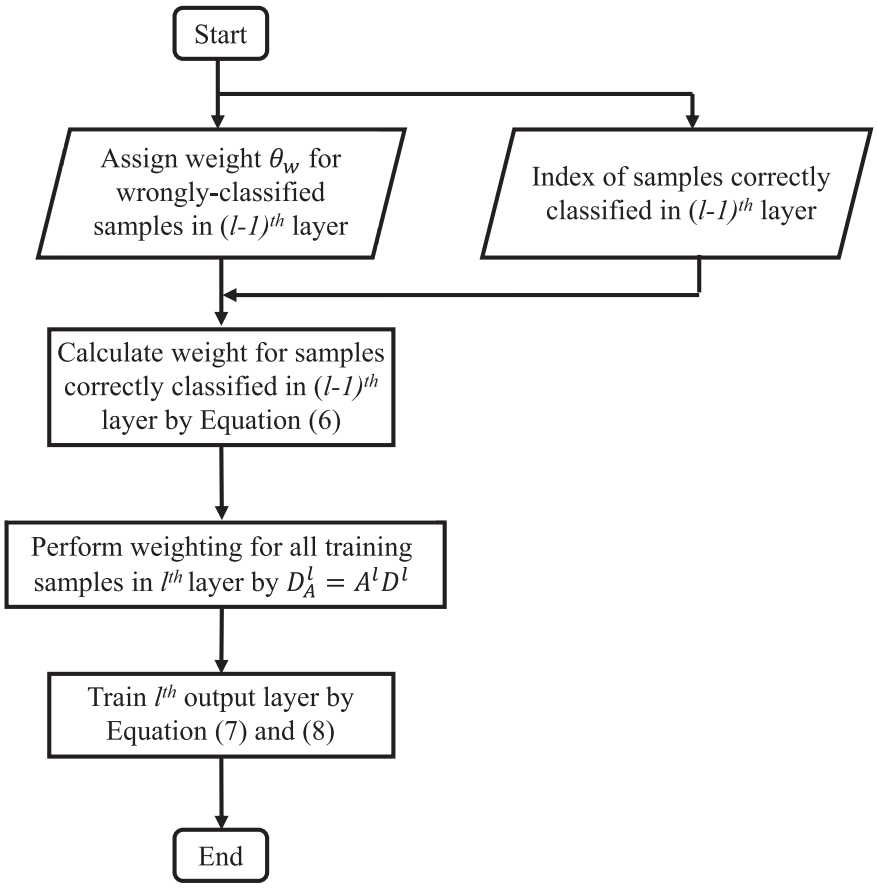



**Fig. 2.** Graphical flowchart of weighting in the $l$th output layer. $\boldsymbol{D}_{\boldsymbol{A}}^{\boldsymbol{l}} = \boldsymbol{A}^{\boldsymbol{l}}\boldsymbol{D}^{\boldsymbol{l}}$ and $\boldsymbol{A}^{\boldsymbol{l}}$ is the diagnosed $(K \times K)$ weight matrix in the $l$th output layer. $\theta_w$ was set as a hyper-parameter in the $l$th output layer.

for a classifier $b$ input with the $j$th sample is computed by normalizing the corresponding competence level:

$$z_j^b = \frac{1 - S_j^b}{\sum_{b \in B}(1 - S_j^b)}. \tag{17}$$

Finally, the weighted ensemble output of the $j$th sample is expressed as

$$\boldsymbol{P}_j = \sum_{b \in B} z^b \boldsymbol{P}_j^b. \tag{18}$$

### 2.5. FGloWD-edRVFL input with convolutional neural network (CNN) features

The graphical flowcharts of the FGloWD-edRVFL network which involved the combined implementations of the four introduced enhancements to the edRVFL network are shown in Fig. 3. Specifically, during training, in each hidden layer, the beneficial features were first selected from the obtained enhancement features. Then, a weight generated based on the classification results of the previous layer was posed on each sample. To train the output layer, the weighted samples with selected features were used to calculate the output weights. Besides, the output weights of the global output layer were also calculated. Finally, the output parameters of hidden layers and output layers were obtained. During testing, after feature selection, weighting was not performed. The outputs of each hidden layer and the global output layer were obtained. Finally, the entropy-based dynamic ensemble was used to calculate the final classification output. The algorithm for training and testing the FGloWD-edRVFL network is also presented in Algorithms 1 and 2.

To cope with the low feature learning capability of the edRVFL network, the features extracted from the last convolutional layer in the CNN were utilized as the input of the FGloWD-edRVFL network after normalization. Specifically, the extracted features $\boldsymbol{X}$ were normalized by MinMax function:

$$\boldsymbol{X}_{norm} = \frac{\boldsymbol{X} - \min(\boldsymbol{X})}{\max(\boldsymbol{X}) - \min(\boldsymbol{X})}. \tag{19}$$

**Table 1**
The hyper-parameter search space of the classification methods.

| Classification method | Hyper-parameter | Space |
|---|---|---|
| SVM | Regularization parameter $C_{SVM}$ | $[2^{-8}, 2^8]$ |
| LR | Regularization parameter $C_{LR}$ | $[2^{-8}, 2^8]$ |
| RF | The number of trees in the forest | [10, 100] |
| | The maximum depth of the tree | [3, 10] |
| | Max_samples[a] | [0.6, 1] |
| RVFL-based networks | Regularization parameter $\lambda$ | $[2^{-5}, 2^4]$ |
| | Number of selected features $m$ | [5, 100] |
| | Weighting parameter $C_w$ | [0, 2] |
| | Regularization parameter $\lambda_g$ | $[2^{-2}, 2^2]$ |
| | Number of the selected classifiers $q$ | {2, 3, 4} |

[a]The maximum number of samples to train each base estimator, which is the product of a float number and the number of training samples.

## 3. Experiments

### 3.1. Introduction of sustained-attention driving (SAD) dataset

The sustained-attention driving (SAD) dataset (Cao et al., 2019) was utilized in this work. In the experiment, lane-departure events were randomly induced to make the car drift from the original cruising lane

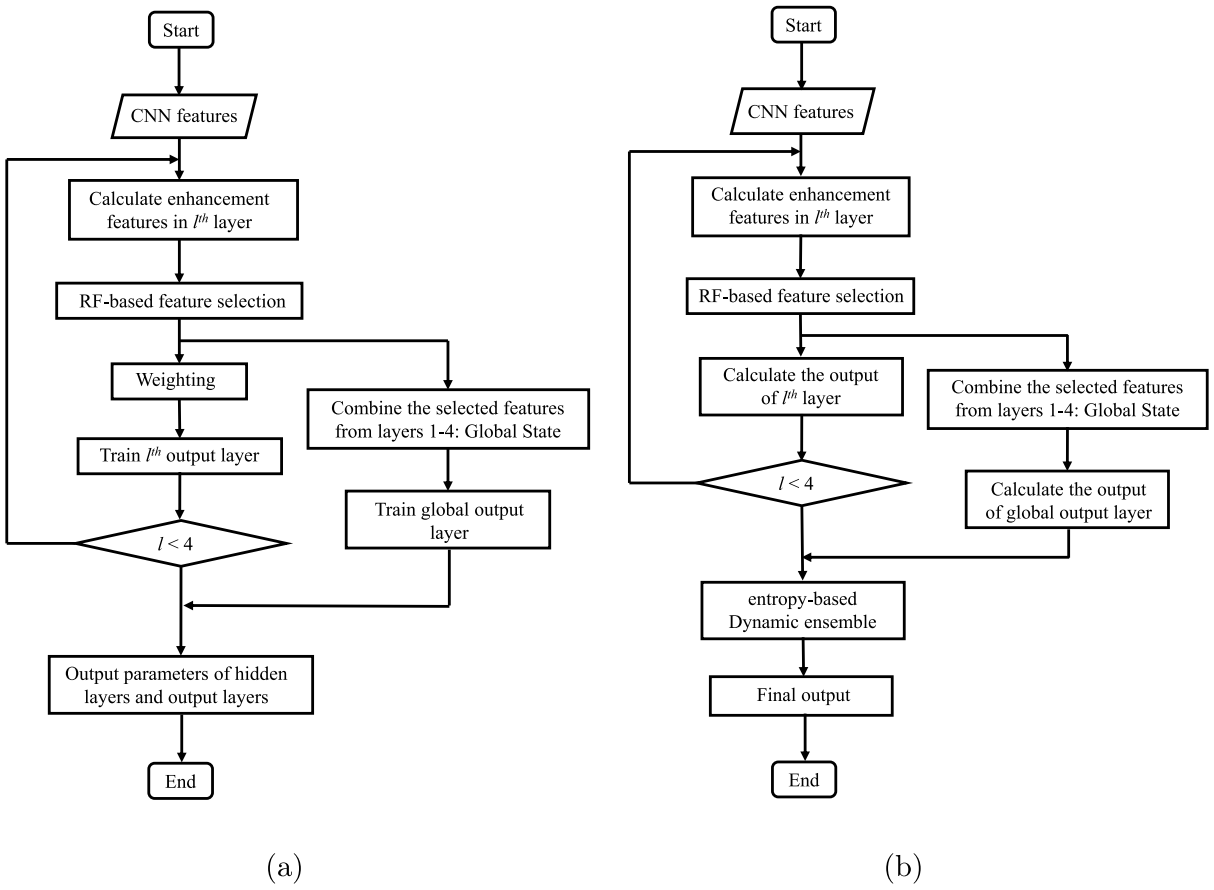



**Fig. 3.** Graphical flowchart of (a) training and (b) testing process of the FGloWD-edRVFL network.

**Algorithm 1:** Training algorithm for the FGloWD-edRVFL network

**Input:** $\boldsymbol{X} \in \mathbb{R}^{K \times d}$ which is the extracted CNN features of training data
**Output:** Output weights $[\boldsymbol{\beta}^l]_{l \in [1,L+1]}$
$l = 1$
**for** $l \leq L$ **do**
  Initialize the hidden weights $\boldsymbol{W}^l$ and bias $\mathbf{r}^l$ randomly
  **if** $l == 1$ **then**
    Set $\boldsymbol{A}^l$ as an identity matrix
  Calculate the hidden features $\boldsymbol{H}^l$ by either equation in Eq. (1)
  Generate the enhancement features $\boldsymbol{D}^l$ by concatenating $\boldsymbol{H}^l$ and $\boldsymbol{X}$
  Calculate the feature importance of $\boldsymbol{D}^l$ by Eq. (7)
  Obtain $\boldsymbol{D}^l_{fs}$ by selecting $m^l$ features with higher feature importance
  Multiply $\boldsymbol{A}^l$ with $\boldsymbol{D}^l_{fs}$ (Weighting)
  Calculate output weights $\boldsymbol{\beta}^l$ by Eqs. (11) and (12)
  Update the weight matrix $\boldsymbol{A}^{l+1}$ of correctly classified samples using (10)
  $l++$
Generate the global state $\boldsymbol{D}_g$ by combining the local features
Calculate the weights $\boldsymbol{\beta}^{L+1}$ of the global output layer by Eqs. (4) and (5)

**Algorithm 2:** Testing algorithm for the FGloWD-edRVFL network

**Input:** $\boldsymbol{X} \in \mathbb{R}^{J \times d}$ which is the extracted CNN features of test data
**Output:** Predicted probability $\boldsymbol{P}$ of test data
Load the randomly generated hidden weights $[\boldsymbol{W}^l]_{l \in [1,L+1]}$ and bias $[\mathbf{r}^l]_{l \in [1,L+1]}$
Calculate the hidden features by either equation in Eq. (1)
Select the same set of enhancement features as training data for test data
Generate the global state by combining the local features
**for** $j \leq J$ **do**
  Calculate output probabilities $\boldsymbol{P}^b_j$ of the classifier in $l^{th}$ layer
  Calculate the entropy of each classifier's output by Eq. (16)
  Select $q$ base classifiers with higher competence level
  Calculate the final output probability by weighting the probabilities of the selected classifiers by Eq. (18)
  $j++$

towards the left or right side (deviation onset). Each participant was instructed to quickly compensate for this perturbation by steering the wheel (response onset) to cause the car to move back to the original cruising lane (response offset). Deviation onset, response onset, and response offset events were all included in a complete trial. During the experiment, the EEG activity of each subject was recorded using a 32-channel Quik-Cap following the International 10–20 system of electrode placement. Processed data provided by Cao et al. (2019) were used in this work. Specifically, the pre-processing steps included bandpass filtering and artefact rejection. The bandpass finite impulse response filters of 1–50 Hz were applied to remove the low-frequency direct current drifts and power line noise. For artefact rejection, the apparent eye blink contamination in the EEG signals was manually removed by visual inspection. Following that, the artefacts were removed by the Automatic Artifact Removal plug-in for EEGLAB, which provided automatic correction of ocular and muscular artefacts in the EEG signals.

Three-second EEG data prior to the deviation onset, which was commonly exploited in previous works (Wei et al., 2018; Cui et al.,

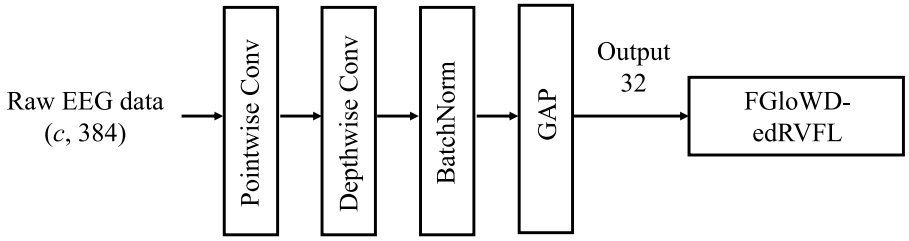



**Fig. 4.** The architecture of the FGloWD-edRVFL network input with ICNN features.

2022), was adopted to perform classification for the upcoming lane-departure event. Then, we followed Wei et al. (2018) to adopt the local reaction time (RT) and the global RT to label data. Specifically, the RT was defined as the time between the deviation onset and the response onset. For each subject, the RT in each lane-departure event was named as the local RT, while the global RT was calculated by averaging the RTs across all trials within a 90 second window before the upcoming deviation onset. The 'alert-RT' was calculated as the 5th percentile of the local RTs for each driving session. When both the local and global RT were shorter than 1.5 times the alert-RT, the corresponding extracted EEG data was labeled as 'alert'. When both the local and global RT were longer than 2.5 times the alert-RT, the data was labeled as 'fatigue'. Transitional states with moderate performance were excluded, and the neutral state was not considered in this work. When there was more than one dataset for a subject, the most balanced one was chosen to execute the filtering. Then, we further down-sampled the data to 128 Hz. Finally, we obtained a balanced driver fatigue dataset which included 2022 samples of 11 subjects. The data size of one sample was 30 (channels) × 384 (time steps).

### 3.2. Introduction of SEED–VIG dataset

The public SEED-VIG dataset recorded from a monotonous driving task in a virtual reality-based simulated driving system (Zheng and Lu, 2017) was also used in this work. The EEG data were recorded using the Neuroscan system. The electrode placement followed the International 10–20 electrode system. In parallel with the EEG recording, the percentage of eye closure (PERCLOS) (Federal Highway Administration, 1998) was measured with the Senso-Motoric-Instrument eye-tracking glasses with a window size of 60 s and a moving step of 10 s.

In this work, we further down-sampled the EEG signals to 128 Hz and processed the data with a low-pass filter of 1 Hz. EEG samples of a 3-second length were extracted prior to the PERCLOS evaluation event. The procedure adopted by Zheng and Lu (2017) was followed. The samples were labeled as 'alert' when PERCLOS was lower than 0.35. The samples were labeled as 'fatigue' when PERCLOS was higher than 0.7, while the samples in the middle range were discarded. We further discarded the sessions with less than 50 samples of either class and balanced the classes for each session by selecting the most 'alert' and 'fatigue' ones. Finally, we obtained a balanced driver fatigue dataset which included 3536 samples of 12 subjects. The data size of one sample is 17 (channels) × 384 (time steps).

### 3.3. Experiment settings and hyper-parameter optimization

The experiment was conducted using an Alienware Desktop with a 64-bit Windows 10 operation system powered by Intel(R) Core(TM) i7-6700 CPU and an NVIDIA GeForce GTX 1080 graphics card. The codes were implemented and tested on Python 3.7.0. Pytorch framework was employed in this work.

In the proposed model, the inputs were the features extracted from the ICNN (Cui et al., 2022) which has shown its great classification capability in different EEG paradigms (Li et al., 2022b). Specifically, the architecture of the proposed FGloWD-edRVFL network input with the ICNN features is shown in Fig. 4. For evaluating the performance of the FGloWD-edRVFL network on the challenging cross-subject driver fatigue recognition tasks, the leave-one-subject-out (LOSO) cross-validation (CV) was conducted. The proposed model was compared with six baselines for the cross-subject driver fatigue recognition: (1) SVM (Cervantes et al., 2020) with the extracted PSD features as the input; (2) RF (Ho, 1995) with the PSD features as the input; (3) EEGNet (Lawhern et al., 2018); (4) ShallowCNN (Schirrmeister et al., 2017); (5) Subject machine (SM) model (Li et al., 2022a); (6) ICNN (Cui et al., 2022); and (7) the original edRVFL network (Shi et al., 2021). To achieve a fair comparison, the (7) original edRVFL network also used the ICNN features as the input, *i.e.*, the FGloWD-edRVFL block in Fig. 4 was replaced with the original edRVFL network.

The PSD features used in this work were computed via Fast Fourier Transform on each EEG epoch from these four spectral bands: delta (1–4 Hz), theta (4–8 Hz), alpha (8–12 Hz) and beta (12–30 Hz). The final feature vector was a concatenation of the spectral powers extracted from the four bands and all available channels. In this study, the final feature vector was of 4 (frequency bands) × 30 (channels) = 120 dimensions.

Regarding the hyper-parameter setting of the BP-based CNN models, we used Adam optimizer with momentum $\beta_1 = 0.9$ and $\beta_2 = 0.99$. The mini-batch was 50, and the learning rate was 0.001. The model was trained for 50 epochs and the early stop was employed based on the LOSO CV on the training set. Cross-entropy was employed as the cost function for training. The mentioned hyper-parameters were kept the same for all the evaluated CNN models, and the other specific settings followed the original works of the baseline models. In addition, for the hyper-parameter setting of SVM, Logistic Regression (LR), and the RVFL-based networks, a LOSO CV was implemented on the training set to select the optimal hyper-parameters in the pre-defined search space during validation. Then, with the optimal hyper-parameters, the models were trained with all training data and tested on the testing set to obtain the testing accuracy. The search space of the hyper-parameters is presented in Table 1. For the edRVFL-based networks, the number of hidden layers $L$ was set as 4. The number of hidden nodes $N^1$ of the first hidden layer was set as 512 for the SAD dataset and 1024 for the SEED-VIG dataset. This was based on the following consideration. A higher number of hidden nodes means a higher possibility of generating more beneficial features. However, the over-high feature dimension may lead to overfitting during the classification in RF-based feature selection. This may result in the low generalization ability of the selected features for the testing data. A higher feature dimension was preferred for the SEED-VIG dataset due to its relatively larger sample size than the SAD dataset. For the second to the last hidden layer, the number of hidden nodes was tuned in a fine range in the neighborhood of the number of hidden nodes $N^1$ and kept the same across these hidden layers. This setting increased the time efficiency of hyper-parameter optimization.

### 3.4. Cross-subject driver fatigue recognition results

The proposed FGloWD-edRVFL network input with the extracted ICNN features was first evaluated on the SAD dataset. The average accuracy and standard deviation of the 11 subjects' data are shown in Table 2. The results are discussed from two aspects. The first one

**Table 2**
Comparison results (%) of LOSO driver fatigue recognition on the SAD dataset. 'ICNN+model' means that the input of the model is the extracted ICNN features. 'Avg. Acc.' represents average accuracy. 'Std.' represents standard deviation.

| Methods | Subjects | | | | | | | | | | | Avg. | Std. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | | |
| SVM (Wei et al., 2018) | 77.66 | **75.76** | 66.67 | 66.22 | 83.04 | 75.90 | 59.80 | 67.80 | 88.54 | 70.37 | 59.73 | 71.95 | 9.16 |
| RF (Ho, 1995) | 85.11 | 55.30 | 66.67 | 75.68 | 73.66 | 80.12 | **71.57** | 65.91 | 87.26 | 78.70 | 68.58 | 73.51 | 9.29 |
| EEGNet (Lawhern et al., 2018) | 56.38 | 71.97 | 46.67 | 65.54 | 53.57 | 63.86 | 67.65 | 53.79 | 71.66 | 79.63 | 58.41 | 62.65 | 9.82 |
| ShallowCNN (Schirrmeister et al., 2017) | 82.98 | 48.48 | 80.67 | 69.59 | 80.36 | 76.51 | 62.75 | 70.08 | 84.08 | 63.89 | **81.86** | 72.84 | 11.11 |
| SM model (Li et al., 2022a) | 78.72 | 68.18 | 79.33 | 68.24 | 85.27 | 83.73 | 64.71 | 57.2 | 78.03 | **82.41** | 71.68 | 74.32 | 8.94 |
| ICNN (Cui et al., 2022) | 85.11 | 45.45 | 80 | 77.7 | **91.52** | 85.54 | 65.69 | 81.06 | 88.85 | 65.74 | 69.91 | 76.05 | 13.44 |
| ICNN+edRVFL (Shi et al., 2021) | 85.11 | 52.27 | 80.00 | **79.73** | 89.73 | **87.95** | 67.65 | **81.44** | 88.54 | 64.81 | 72.12 | 77.21 | 11.78 |
| ICNN+FGloWD-edRVFL | **87.23** | 60.61 | **82.67** | 79.05 | 90.63 | 85.54 | 67.65 | 79.92 | **90.45** | 75.00 | 74.34 | **79.37** | 9.46 |

**Table 3**
Comparison results (%) of LOSO driver fatigue recognition on the SEED-VIG dataset. 'ICNN+model' means that the input of the model is the extracted ICNN features. 'Avg. Acc.' represents average accuracy. 'Std.' represents standard deviation.

| Methods | Subject | | | | | | | | | | | | Avg. Acc. | Std. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | | |
| SVM (Wei et al., 2018) | 92.11 | 74.07 | 78.76 | 60.42 | 94.85 | 68.70 | 77.57 | 94.37 | 84.41 | 87.96 | 59.35 | 88.83 | 80.12 | 12.48 |
| RF (Ho, 1995) | 92.98 | 86.42 | 84.73 | 83.59 | 96.39 | 86.48 | 81.62 | 90.14 | 86.63 | 90.88 | 67.27 | 49.51 | 83.05 | 12.80 |
| EEGNet (Lawhern et al., 2018) | 85.53 | 61.11 | 71.02 | 71.09 | 79.90 | 77.96 | 80.51 | 85.21 | 78.71 | **93.80** | 85.97 | 89.32 | 80.01 | 9.01 |
| ShallowCNN (Schirrmeister et al., 2017) | 92.54 | **93.83** | 78.10 | 78.91 | 72.16 | 92.96 | 81.25 | 90.85 | 89.85 | 93.43 | 73.02 | 85.92 | 85.24 | 8.17 |
| ICNN (Cui et al., 2022) | **95.18** | 89.51 | 89.16 | 84.64 | 85.05 | 95.00 | **84.93** | 92.25 | 93.32 | 91.97 | 87.05 | **90.29** | 89.86 | 3.82 |
| SM (Li et al., 2022a) | 92.98 | 91.98 | 87.39 | 85.42 | 92.78 | 95.37 | 71.69 | 97.18 | 91.09 | 92.34 | 91.01 | 89.81 | 89.92 | 6.55 |
| ICNN+edRVFL (Shi et al., 2021) | 93.42 | 88.89 | **89.60** | 84.38 | **96.91** | **95.56** | 82.72 | 89.44 | 90.10 | 91.97 | **91.73** | 89.32 | 90.34 | 4.07 |
| ICNN+FGloWD-edRVFL | 94.74 | **93.83** | 88.72 | **88.54** | 95.88 | 94.63 | 83.82 | **97.89** | **93.56** | 91.97 | 88.13 | 89.81 | **91.79** | 4.03 |

is to compare the original edRVFL and the proposed FGloWD-edRVFL network with the baseline methods. We can observe that the proposed FGloWD-edRVFL network outperformed all strong baseline methods for driver fatigue recognition and the improvement in average accuracy ranged from 3.32% to 16.72%. It is worth noticing that the original edRVFL network input with the ICNN features showed an approximately 1.16% increase in average accuracy compared with the SOTA method ICNN, demonstrating the superiority of the edRVFL network based on the closed-form solution. Then, comparing the original edRVFL with the FGloWD-edRVFL networks, the proposed FGloWD-edRVFL network gained an improvement of 2.16% and achieved a new SOTA average accuracy of 79.37% on the cross-subject driver fatigue recognition.

Furthermore, the proposed model was evaluated on the SEED-VIG dataset. The performance of the 12 subjects' data is presented in Table 3. From the comparison between the baseline methods and the edRVFL-based models (*i.e.*, the original edRVFL network and the FGloWD-edRVFL network), the edRVFL variants outperformed all the baseline methods. In particular, the proposed FGloWD-edRVFL network achieved improvements in average accuracy ranging from 1.87% to 11.78% over the baseline methods. Then, in the comparison with the original edRVFL network, the FGloWD-edRVFL network achieved the new SOTA average accuracy of 91.79%. The superior performance of the proposed model on both public datasets demonstrated the effectiveness of the enhancements proposed in the FGloWD-edRVFL network. The analysis of each component in the FGloWD-edRVFL network is introduced in Section 3.5.

To better understand the capability of the proposed FGloWD-edRVFL network for driver fatigue recognition, Precision, Sensitivity, Specificity, and F1-score of the proposed model and the baseline methods on the two datasets are compared in Tables 4 and 5. The class 'alert' was set as positive, while the class 'fatigue' was set as negative in the calculation of these metrics. Comparison results showed the outstanding Precision and F1-score of the FGloWD-edRVFL network over the baseline methods. In particular, the best F1-scores of 81.74% and 91.50% were achieved by the FGloWD-edRVFL network on the SAD and the SEED-VIG datasets, respectively. Furthermore, the Specificity of the proposed model was also higher than that of the baseline methods on both datasets. This demonstrated that the proposed model could ensure a higher probability that the 'fatigue' subjects are correctly classified as 'fatigue'. This is extremely meaningful in practice. Lastly, the proposed model showed higher Sensitivity than most of the baseline methods.

After presenting the classification metrics, the statistical Friedman and *post-hoc* Nemenyi tests were performed to differentiate the methods (Demšar, 2006). The Nemenyi test utilizes the critical distance (CD) to characterize the difference among the models, which is presented by:

$$CD = a\sqrt{\frac{r(r+1)}{6N_s}}, \tag{20}$$

where $a$ represents the critical value obtained from the studentized range statistic divided by $\sqrt{2}$, $r$ is the number of methods and $N_s$ is the number of subjects (Demšar, 2006). The calculated CDs for the SAD dataset and the SEED-VIG dataset were 2.90 and 2.79, respectively. Based on the CDs, the results of the statistical tests on the two datasets are shown in Fig. 5. We can observe that the proposed model had the highest ranking on both datasets.

To compare the proposed model with the baseline methods in a pair-wise manner, the one-tailed Wilcoxon paired signed-rank test was conducted. The $p$-values of the Wilcoxon test on the two datasets are shown in Tables 6 and 7. It was observed that the proposed FGloWD-edRVFL network performed significantly better than most of the baseline methods ($p < 0.05$).

Based on all the comparison results presented above, we conclude that the proposed FGloWD-edRVFL network showed an outstanding performance on the challenging cross-subject driver fatigue recognition tasks over the SOTA methods which included ICNN, SM model and the original edRVFL network. Statistically significant improvements and a higher average rank were achieved by the proposed FGloWD-edRVFL network. Lastly, the comparison on the four classification performance metrics further demonstrated the superiority of our method. Therefore, the proposed FGloWD-edRVFL network input with the ICNN features can be considered as a highly competitive classifier for the EEG-based cross-subject driver fatigue recognition tasks.

### 3.5. Ablation study

To investigate the effectiveness of each proposed enhancement in the FGloWD-edRVFL network, the LOSO average accuracy of the

**Table 4**
Precision, Sensitivity, Specificity and F1-score (%) of the proposed FGloWD-edRVFL network and the baseline methods on the SAD dataset.

| | Precision | Sensitivity | Specificity | F1-Score |
|---|---|---|---|---|
| SVM (Wei et al., 2018) | 73.76 | 73.39 | 73.89 | 73.57 |
| RF (Ho, 1995) | 76.14 | 71.02 | 77.74 | 73.49 |
| EEGNet (Lawhern et al., 2018) | 57.59 | **88.92** | 34.52 | 69.91 |
| ShallowCNN (Schirrmeister et al., 2017) | 74.21 | 76.85 | 73.29 | 75.51 |
| SM model (Li et al., 2022a) | 70.66 | 82.89 | 65.58 | 76.28 |
| ICNN (Cui et al., 2022) | 77.62 | 80.61 | 76.76 | 79.09 |
| ICNN+edRVFL (Shi et al., 2021) | 77.48 | 83.38 | 75.77 | 80.32 |
| ICNN+FGloWD-edRVFL | **79.89** | 83.68 | **78.93** | **81.74** |

**Table 5**
Precision, Sensitivity, Specificity and F1-score (%) of the proposed FGloWD-edRVFL network and the baseline methods on the SEED-VIG dataset.

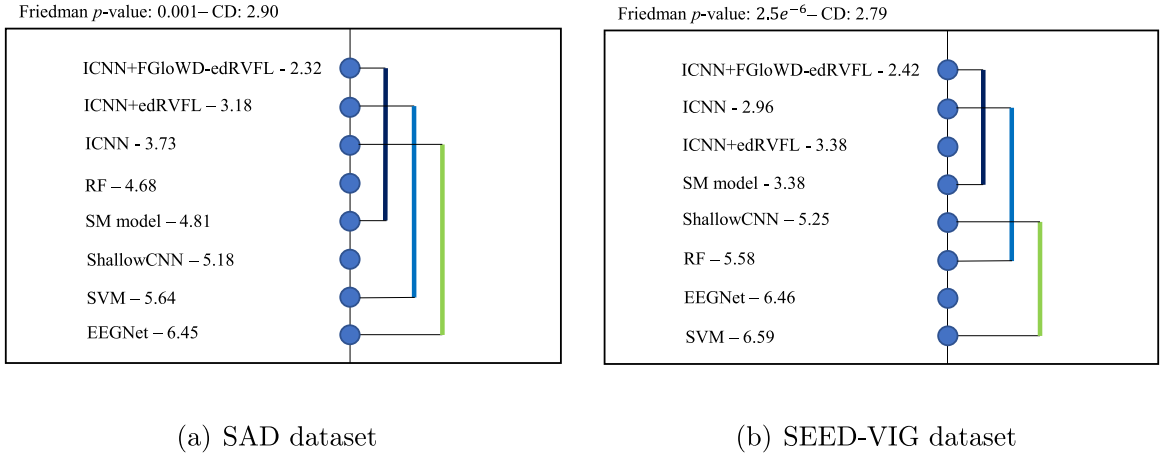
| | Precision | Sensitivity | Specificity | F-score |
|---|---|---|---|---|
| SVM (Wei et al., 2018) | 74.00 | 85.46 | 69.97 | 79.32 |
| RF (Ho, 1995) | 84.62 | 81.56 | 85.18 | 83.06 |
| EEGNet (Lawhern et al., 2018) | 74.00 | 90.16 | 68.33 | 81.29 |
| ShallowCNN (Schirrmeister et al., 2017) | 84.73 | 85.69 | 84.56 | 85.21 |
| ICNN (Cui et al., 2022) | 89.05 | 91.52 | 88.74 | 90.26 |
| SM model (Li et al., 2022a) | 88.57 | 91.12 | 88.24 | 89.82 |
| ICNN+edRVFL (Shi et al., 2021) | 87.43 | **94.46** | 86.43 | 90.81 |
| ICNN+FGloWD-edRVFL | **90.76** | 92.25 | **90.61** | **91.50** |

**Table 6**
*p*-values of the Wilcoxon results on the SAD dataset.

| | SVM | RF | EEGNet | ShallowCNN | ICNN | SM model | ICNN+edRVFL |
|---|---|---|---|---|---|---|---|
| ICNN+FGloWD-edRVFL | 0.021 | 0.016 | 0.006 | 0.005 | 0.006 | 0.042 | 0.046 |

**Table 7**
*p*-values of the Wilcoxon results on the SEED-VIG dataset.

| | SVM | RF | EEGNet | ShallowCNN | ICNN | SM model | ICNN+edRVFL |
|---|---|---|---|---|---|---|---|
| ICNN+FGloWD-edRVFL | $< 0.001$ | $< 0.001$ | $< 0.001$ | 0.001 | 0.143 | 0.031 | 0.088 |




(a) SAD dataset (b) SEED-VIG dataset

**Fig. 5.** Friedman and Nemenyi test results.

RVFL network, the edRVFL network and its variants with the proposed enhancements were compared. Specifically, the variants included (1) F-edRVFL: the edRVFL with feature selection; (2) FW-edRVFL: the edRVFL network with feature selection and weighting; (3) FGloW-edRVFL: the edRVFL network with feature selection, weighting and global output layer; (4) FGloWD-edRVFL which involved the entropy-based dynamic ensemble in the FGloW-edRVFL network. We normalized the average accuracy to $[0, 1]$ for ease of comparison. The results are shown in Fig. 6. Overall, the average accuracy was boosted in a step-wise manner. This demonstrated that the proposed implementation of all the enhancements to the edRVFL network was essential for achieving the best possible performance of the cross-subject driver fatigue recognition tasks.

Furthermore, the use of different normalization functions was also investigated. This work employed the MinMax function as shown in Eq. (19). The standardization function and RobustScaler function were used for comparison. These two functions are given by Eqs. (21) and (22), respectively.

$$\boldsymbol{X}_{norm} = \frac{\boldsymbol{X} - mean(\boldsymbol{X})}{std(\boldsymbol{X})}. \tag{21}$$

$$\boldsymbol{X}_{norm} = \frac{\boldsymbol{X} - median(\boldsymbol{X})}{IQR(\boldsymbol{X})}. \tag{22}$$

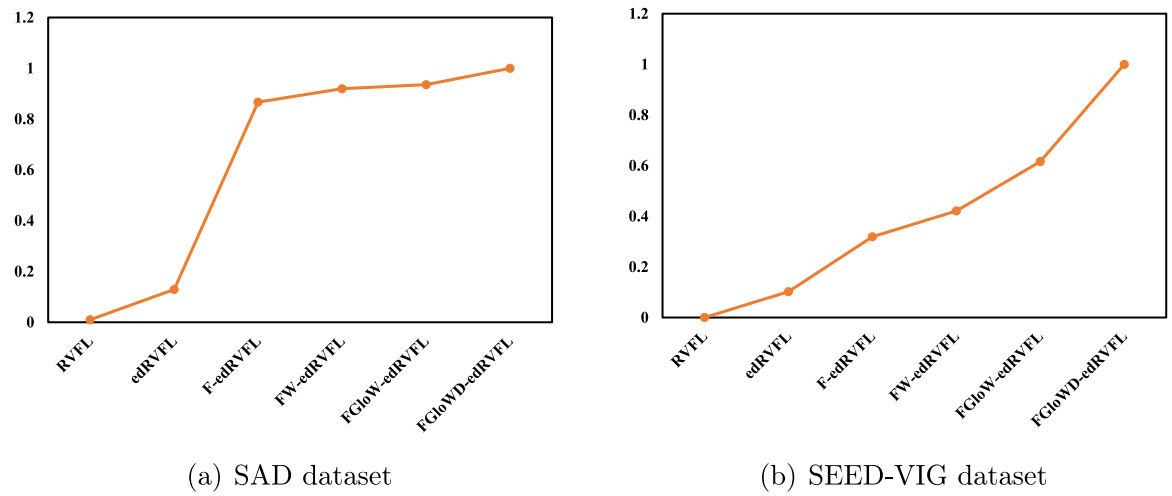



**Fig. 6.** Analysis of the proposed enhancements in the FGloWD-edRVFL network.

**Table 8**
Comparison results (%) of using different normalization functions.

| | MinMax | Standardization | RoubustScaler |
|---|---|---|---|
| SAD | 79.32 | 79.12 | 78.60 |
| SEED-VIG | 91.79 | 91.43 | 91.87 |

The terms $mean(\cdot)$ and $std(\cdot)$ represent the functions of calculating the mean and the standard deviation. The terms $median(\cdot)$ and $IQR(\cdot)$ represent the functions of calculating the median and the interquartile range. The average accuracy of using the MinMax, standardization and RobustScaler functions is shown in Table 8. As observed, there was no significant difference in the performance among these three normalization functions. The MinMax function presented the best performance for the SAD dataset, while the RobustScaler function was superior to the other two functions for the SEED-VIG dataset.

### 3.6. Further investigation on the FGloWD-edRVFL network

One may argue that the superiority of the proposed FGloWD-edRVFL network may be solely attributed to the capability of the ICNN features as the input. Therefore, to further investigate the capability of the proposed FGloWD-edRVFL network on the EEG-based driver fatigue recognition, its performance was compared with two other classical classifiers (SVM and LR) and other categories of RNNs. The input of the compared models was the same ICNN features. SVM was selected as it also has the global optimum solution and LR was selected as it is a linear model. For the RNNs, the three popular networks: ELM (Huang et al., 2006), SCN (Wang and Li, 2017) and BLS (Chen and Liu, 2018) were included for comparison in this work. For ELM, the number of hidden weights and the search space of the regularization parameter were the same as that of the RVFL-based networks as shown in Table 1. For SCN, we followed the default setting used by Hu and Suganthan (2022). For BLS, the search space of the regularization parameter of RVFL-based networks was used. The rest of the settings followed the open-source code of the original BLS work.

The performance obtained on both datasets is shown in Tables 9 and 10. Overall, with the same inputs, the FGloWD-edRVFL network showed better performance than the classical machine learning methods. Furthermore, compared with the other three popular RNNs, the outstanding performance of the proposed model could be observed. Therefore, the strong classification capability of the proposed FGloWD-edRVFL was further demonstrated.

## 4. Discussion

Compared with the SOTA CNN models for cross-subject driver fatigue recognition, using edRVFL-based networks to replace the classifier can compensate for the limitation of the classification capability of the CNN models. In edRVFL-based networks which are a type of RNNs, instead of updating the hidden layers during training, the weights of RNNs are randomly initialized and kept fixed throughout the training (Zhang and Suganthan, 2016b), which means that the hidden weights of the edRVFL networks are data independent. Furthermore, the global optimum solutions could guarantee better generalization capability of the model. Previous studies showed that using SVM (Niu and Suen, 2012) in hybrid CNN-X models could be beneficial to boost the performance of the CNN models. Therefore, based on the merits of the edRVFL networks, the hybrid CNN feature-based edRVFL model can perform better than the SOTA CNN models.

Compared with the SOTA RNN, the edRVFL network, the results presented in the previous section showed that the four proposed enhancements to the edRVFL network were beneficial to further boost the classification performance. The possible reasons can be elaborated from two aspects. The first is the higher efficiency of exploiting the randomly generated hidden features. Due to the nature of edRVFL, redundant and invalid information could be contained in the hidden features. Hence, the neuron selection techniques based on the feature importance could improve the classification performance. Moreover, the global output layer designed was based on the idea of dRVFL (Shi et al., 2021) and deep reservoir computing (Gallicchio et al., 2017) which fully utilized the features from all levels. Accordingly, adding a global state could increase the feature efficiency and the input diversity of ensemble learning.

The second reason for performance improvement is the benefits of the ensemble learning techniques used. Weighting has been shown as a promising method to improve the generalization capability of ensemble learning, *e.g.*, AdaBoost (Freund and Schapire, 1997). In this work, weighting used at different levels could implicitly assign the focus of the training phase and help the model better exploit the hard samples during training. Recently, some similar attempts have also been made. Weighting based on an intuitionistic fuzzy membership scheme (Malik et al., 2022b) and sample attention (Hu et al., 2022) brought superior classification ability for RVFL-based networks. Furthermore, entropy has been used to find the low-uncertainty samples to update the models in transfer learning (Zou et al., 2018). For a sample in the edRVFL network, a lower entropy value generated by the classifier means a higher confidence level that the sample can be correctly classified. Thus, this classifier could be recognized to have a higher competence

**Table 9**
Comparison results (%) of using different classifiers input with ICNN features on the SAD dataset.

| | Avg. | Std. | Precision | Sensitivity | Specificity | F1-Score |
|---|---|---|---|---|---|---|
| ICNN (Cui et al., 2022) | 76.05 | 13.44 | 77.62 | 80.61 | 76.76 | 79.09 |
| ICNN+SVM | 76.16 | 12.79 | 78.62 | 78.93 | 78.54 | 78.78 |
| ICNN+LR | 76.77 | 9.15 | 75.71 | 84.47 | 72.90 | 79.85 |
| ICNN+RF | 76.60 | 12.32 | 78.88 | 78.34 | **79.03** | 78.61 |
| ICNN+ELM (Huang et al., 2006) | 77.22 | 11.17 | 76.70 | 84.67 | 74.28 | 80.49 |
| ICNN+SCN (Wang and Li, 2017) | 76.17 | 12.16 | 74.91 | **84.77** | 71.61 | 79.54 |
| ICNN+BLS (Chen and Liu, 2018) | 78.00 | 11.78 | 78.18 | 83.98 | 76.56 | 80.97 |
| ICNN+FGloWD-edRVFL | **79.37** | 9.46 | **79.89** | 83.68 | 78.93 | **81.74** |

**Table 10**
Comparison results (%) of using different classifiers input with ICNN features on the SEED-VIG dataset.

| | Avg. Acc. | Std. | Precision | Sensitivity | Specificity | F1-Score |
|---|---|---|---|---|---|---|
| ICNN (Cui et al., 2022) | 89.86 | 3.82 | 89.05 | 91.52 | 88.74 | 90.26 |
| ICNN+SVM | 90.59 | 4.16 | 87.69 | 93.33 | 87.50 | 90.68 |
| ICNN+LR | 90.43 | 5.02 | 88.19 | 93.89 | 86.82 | 90.69 |
| ICNN+RF | 90.56 | 3.08 | 88.16 | 93.10 | 87.50 | 90.56 |
| ICNN+ELM (Huang et al., 2006) | 90.11 | 3.80 | 85.78 | 95.19 | 84.22 | 90.24 |
| ICNN+SCN (Wang and Li, 2017) | 88.51 | 3.94 | 84.59 | 93.16 | 83.03 | 88.67 |
| ICNN+BLS (Chen and Liu, 2018) | 90.18 | 4.00 | 85.89 | **95.36** | 84.33 | 90.38 |
| ICNN+FGloWD-edRVFL | **91.79** | 4.03 | **90.76** | 92.25 | **90.61** | **91.50** |

level. By dynamically selecting high-competence classifiers and weighting their outputs, the impact of the less beneficial classifiers could be reduced while the reliable classifiers could generate better performance in ensemble learning. Overall, the collaboration effect of the proposed enhancements was beneficial in boosting the classification capability of the edRVFL network. While the enhancements to the edRVFL network may compromise the training efficiency to some extent, it has to be emphasized that the model's superior performance was the primary goal of this work.

## 5. Conclusion

In this work, we investigated the use of the edRVFL-based networks for EEG decoding in the tasks of challenging cross-subject driver fatigue recognition. The ICNN features were used as the input of the edRVFL variants. Four enhancements were proposed to further improve the performance of the original edRVFL network, including (1) RF-based feature selection which was to reduce the impact of the inferior and redundant features generated from the random weights of the edRVFL network, (2) weighting of samples in the hidden layers which was to force the model to better learn from the hard samples and diversify the trained classifiers in ensemble learning, (3) global output layer with the input of global state which was to further increase the diversity of ensemble learning, and (4) entropy-based dynamic ensemble which was proposed to reduce the impact of the low-competence classifiers for each testing sample. The comprehensive empirical study showed that the collaboration of the proposed enhancements was beneficial to improve the classification performance of the edRVFL network. Furthermore, the proposed FGloWD-edRVFL network significantly outperformed the strong baselines, achieving the new SOTA performance on the public SAD dataset and SEED-VIG dataset. In conclusion, the proposed FGloWD-edRVFL network can substantially improve the EEG-based cross-subject driver fatigue recognition performance, indicating that using the edRVFL-based model can be a new direction for EEG-based driver fatigue recognition.

Despite the excellent performance of the proposed model, the combinative use of CNN and RNN may prolong the training time. Also, the grid-search optimization used cannot guarantee that the most suitable parameters were included in the preset values of the model during training. Hence, the following directions can be considered in future works:

- Using RNNs to learn directly from raw EEG data with high dimensions could be a research goal. Down-sampling techniques in RNNs or RNN architectures that are more suitable for processing high-dimensional data can be used.

**Table A.1**
Variables and the corresponding description.

| Variable | Description |
|---|---|
| $\boldsymbol{r}^l$ | Randomly initialized bias in the $l$th hidden layer |
| $FS^l$ | RF-based feature selection in the $l$th hidden layer |
| $\boldsymbol{o}^l$ | Output probabilities of the $l$th output layer |
| $\boldsymbol{X}$ | Extracted EEG features |
| $\boldsymbol{Y}$ | Categories of EEG signals |
| $K$ | Number of samples |
| $d$ | Number of feature dimensions |
| $C$ | Number of categories |
| $L$ | Number of hidden layers |
| $N^l$ | Number of hidden nodes in the $l$th layer |
| $H^l$ | Hidden features of the $l$th layer |
| $\boldsymbol{D}^l$ | Enhancement features of the $l$th layer |
| $\beta^l$ | Weights of the $l$th output layer |
| $\lambda^l$ | Regularization parameter in the $l$th output layer |
| $\alpha^l_{i,n}$ | Impurity decrease of the $i$th enhancement feature at the node $n$ of the $t$th tree of the RF |
| $G_n$ | Gini index calculated at the node $n$ |
| $G_r$ | Gini index for the resulting right children nodes |
| $G_e$ | Gini index for the resulting left children nodes |
| $\alpha_i$ | Normalized impurity decrease of the $i$th enhancement feature |
| $m^l$ | The number of selected features |
| $\boldsymbol{D}^l_{fs}$ | Enhancement features obtained after feature selection |
| $C_r$ | Number of samples that are classified correctly |
| $C_w$ | Number of samples that are classified and wrongly |
| $\theta_r$ | Weight for the correctly classified samples |
| $\theta_w$ | Weight for the wrongly classified samples |
| $\boldsymbol{A}^l$ | Diagonal $(K \times K)$ weight matrix in the $l$th output layer |
| $\lambda_g$ | Regularization parameter in the global output layer |
| $\boldsymbol{D}_g$ | Global state |
| $b$ | The classifier in an output layer |
| $\boldsymbol{P}^b_j$ | Output probabilities of the classifier $b$ for the $j$th testing sample |
| $p^b_{j,c}$ | Output probability of the $c$th category |
| $1 - S^b_j$ | Competence level of the classifier $b$ for the $j$th testing sample |
| $\boldsymbol{B}$ | A set of classifiers with higher competence levels |
| $q$ | Number of classifiers selected |
| $z^b_j$ | Weight for a classifier $b$ input with the $j$th sample |
| $\boldsymbol{X}_{norm}$ | Normalized CNN features |

**Table A.2**
Abbreviations and the corresponding description.

| Abbreviation | Description |
|---|---|
| BiPa | Binary pattern |
| BP | Back-propagation |
| BLS | Broad learning system |
| CD | Critical distance |
| CNN | Convolutional neural network |
| CV | Cross-validation |
| DCBiPa | Dynamic center based binary pattern |
| DE | Power spectral density |
| DNN | Deep neural network |
| edRVFL | ensemble deep random vector functional link |
| EEG | Electroencephalogram |
| ELM | Extreme learning machine |
| F-edRVFL | edRVFL with feature selection |
| FGloW-edRVFL | edRVFL network with feature selection, weighting and global output layer |
| FGloWD-edRVFL | edRVFL with **F**eature selection, **Glo**bal output layer, **W**eighting and **D**ynamic ensemble |
| FW-edRVFL | edRVFL network with feature selection and weighting |
| LOSO | leave-one-subject-out |
| LR | Logistic regression |
| MI | Motor imagery |
| ICNN | InterpretableCNN |
| PERCLOS | Percentage of eye closure |
| RF | Random forest |
| RNN | Randomized neural network |
| RT | Reaction time |
| RVFL | Random vector functional link |
| SAD | Sustained-attention driving |
| SOTA | State-of-the-art |
| SCN | Stochastic configuration network |
| SM | Subject matching |
| SVM | Support vector machine |

- Better optimization algorithms such as bayesian optimization (Pelikan et al., 1999) can be used to help further improve the performance.

## CRediT authorship contribution statement

**Ruilin Li:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Visualization, Writing – original draft. **Ruobin Gao:** Conceptualization, Methodology, Software, Validation, Formal analysis, Writing – review & editing. **Liqiang Yuan:** Formal analysis, Writing – review & editing. **P.N. Suganthan:** Conceptualization, Methodology, Writing – review & editing, Supervision. **Lipo Wang:** Formal analysis, Writing – review & editing, Supervision. **Olga Sourina:** Resources, Writing – review & editing, Supervision.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request

## Acknowledgment

Open Access funding provided by the Qatar National Library.

## Appendix

See Tables A.1 and A.2.

## References

Bang, J.-S., Lee, M.-H., Fazli, S., Guan, C., Lee, S.-W., 2022. Spatio-spectral feature representation for motor imagery classification using convolutional neural networks. IEEE Trans. Neural Netw. Learn. Syst. 33, 3038–3049. http://dx.doi.org/10.1109/TNNLS.2020.3048385.

Byeon, H., 2020. Exploring the predictors of rapid eye movement sleep behavior disorder for Parkinson's disease patients using classifier ensemble. Healthcare 8, 121. http://dx.doi.org/10.3390/healthcare8020121.

Cao, Z., Chuang, C.-H., King, J.-K., Lin, C.-T., 2019. Multi-channel EEG recordings during a sustained-attention driving task. Sci. Data 6, 19. http://dx.doi.org/10.1038/s41597-019-0027-4.

Cao, W.-P., Li, S.-D., Huang, C.-C., Wu, Y.-H., Wang, Q., Li, D.-C., Liu, Y., 2021b. An ensemble fuzziness-based online sequential learning approach and its application. In: Knowl. Sci. Eng. Manage.. pp. 255–267. http://dx.doi.org/10.1007/978-3-030-82136-4_21.

Cao, W., Xie, Z., Li, J., Xu, Z., Ming, Z., Wang, X., 2021a. Bidirectional stochastic configuration network for regression problems. Neural Netw. 140, 237–246. http://dx.doi.org/10.1016/j.neunet.2021.03.016.

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., Lopez, A., 2020. A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing 408, 189–215. http://dx.doi.org/10.1016/j.neucom.2019.10.118.

Chen, C.L.P., Liu, Z., 2018. Broad learning system: An effective and efficient incremental learning system without the need for deep architecture. IEEE Trans. Neural Netw. Learn. Syst. 29, 10–24. http://dx.doi.org/10.1109/TNNLS.2017.2716952.

Cheng, W.X., Suganthan, P.N., Katuwal, R., 2021. Time series classification using diversified ensemble deep random vector functional link and resnet features. Appl. Soft Comput. 112, 107826. http://dx.doi.org/10.1016/j.asoc.2021.107826.

Cruz, R.M.O., Sabourin, R., Cavalcanti, G.D.C., 2018. Dynamic classifier selection: Recent advances and perspectives. Inf. Fusion 41, 195–216. http://dx.doi.org/10.1016/j.inffus.2017.09.010.

Cui, J., Lan, Z., Sourina, O., Müller-Wittig, W., 2022. EEG-based cross-subject driver drowsiness recognition with an interpretable convolutional neural network. IEEE Trans. Neural Netw. Learn. Syst. 1–13, http://dx.doi.org/10.1109/TNNLS.2022.3147208.

Demšar, J., 2006. Statistical comparisons of classifiers over multiple data sets. J. Mach. Learn. Res. 7, 1–30, URL: http://jmlr.org/papers/v7/demsar06a.html.

Elsheikh, A.H., Abd Elaziz, M., Ramesh, B., Egiza, M., Al-qaness, M.A.A., 2021a. Modeling of drilling process of GFRP composite using a hybrid random vector functional link network/parasitism-predation algorithm. J. Mater. Res. Technol. 14, 298–311. http://dx.doi.org/10.1016/j.jmrt.2021.06.033.

Elsheikh, A.H., Abd Elaziz, M., Vendan, A., 2022. Modeling ultrasonic welding of polymers using an optimized artificial intelligence model using a gradient-based optimizer. Weld. World 66, 27–44. http://dx.doi.org/10.1007/s40194-021-01197-x.

Elsheikh, A.H., El-Said, E.M.S., Abd Elaziz, M., Fujii, M., El-Tahan, H.R., 2023. Water distillation tower: Experimental investigation, economic assessment, and performance prediction using optimized machine-learning model. J. Clean. Prod. 388, 135896. http://dx.doi.org/10.1016/j.jclepro.2023.135896.

Elsheikh, A.H., Elaziz, M.A., Das, S.R., Muthuramalingam, T., Lu, S., 2021b. A new optimized predictive model based on political optimizer for eco-friendly MQL-turning of AISI 4340 alloy with nano-lubricants. J. Manuf. Process. 67, 562–578. http://dx.doi.org/10.1016/j.jmapro.2021.05.014.

Fan, C., Peng, Y., Peng, S., Zhang, H., Wu, Y., Kwong, S., 2022. Detection of train driver fatigue and distraction based on forehead EEG: A time-series ensemble learning method. IEEE Trans. Intell. Transp. Syst. 23, 13559–13569. http://dx.doi.org/10.1109/TITS.2021.3125737.

Federal Highway Administration, 1998. PERCLOS: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance. URL: https://rosap.ntl.bts.gov/view/dot/113.

Federico, V., Harris, J.S., Garrison, H.G., McKay, M.P., 2005. Drowsy driving. Ann. Emerg. Med. 45, 433–434. http://dx.doi.org/10.1016/j.annemergmed.2005.01.015.

Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. System Sci. 55, 119–139. http://dx.doi.org/10.1006/jcss.1997.1504.

Gallicchio, C., Micheli, A., Pedrelli, L., 2017. Deep reservoir computing: A critical experimental analysis. Neurocomputing 268, 87–99. http://dx.doi.org/10.1016/j.neucom.2016.12.089.

Gao, Z., Li, Y., Yang, Y., Dong, N., Yang, X., Grebogi, C., 2020. A coincidence-filtering-based approach for CNNs in EEG-based recognition. IEEE Trans. Industr. Inform. 16, 7159–7167. http://dx.doi.org/10.1109/TII.2019.2955447.

Ho, T.K., 1995. Random decision forests. In: Int. Conf. Doc. Anal. Recognit.. pp. 278–282. http://dx.doi.org/10.1109/ICDAR.1995.598994.

Hu, M., Gao, R., Suganthan, P.N., Tanveer, M., 2022. Automated layer-wise solution for ensemble deep randomized feed-forward neural network. Neurocomputing 514, 137–147. http://dx.doi.org/10.1016/j.neucom.2022.09.148.

Hu, M., Suganthan, P.N., 2022. Experimental evaluation of stochastic configuration networks: Is SC algorithm inferior to hyper-parameter optimization method? Appl. Soft Comput. 126, 109257. http://dx.doi.org/10.1016/j.asoc.2022.109257.

Huang, G.-B., Zhu, Q.-Y., Siew, C.-K., 2006. Extreme learning machine: Theory and applications. Neurocomputing 70, 489–501. http://dx.doi.org/10.1016/j.neucom.2005.12.126.

Igelnik, B., Pao, Y.-H., 1995. Stochastic choice of basis functions in adaptive function approximation and the functional-link net. IEEE Trans. Neural Netw. 6, 1320–1329. http://dx.doi.org/10.1109/72.471375.

Khare, S.K., Bajaj, V., 2021. Time–frequency representation and convolutional neural network-based emotion recognition. IEEE Trans. Neural Netw. Learn. Syst. 32, 2901–2909. http://dx.doi.org/10.1109/TNNLS.2020.3008938.

Koay, H.V., Chuah, J.H., Chow, C.-O., Chang, Y.-L., 2022. Detecting and recognizing driver distraction through various data modality using machine learning: A review, recent advances, simplified framework and open challenges (2014–2021). Eng. Appl. Artif. Intell. 115, 105309. http://dx.doi.org/10.1016/j.engappai.2022.105309.

Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P., Lance, B.J., 2018. EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces. J. Neural Eng. 15, 056013. http://dx.doi.org/10.1088/1741-2552/aace8c.

Li, R., Gao, R., Suganthan, P.N., 2023. A decomposition-based hybrid ensemble CNN framework for driver fatigue recognition. Inform. Sci. 624, 833–848. http://dx.doi.org/10.1016/j.ins.2022.12.088.

Li, R., Wang, L., Sourina, O., 2022a. Subject matching for cross-subject EEG-based recognition of driver states related to situation awareness. Methods 202, 136–143. http://dx.doi.org/10.1016/j.ymeth.2021.04.009.

Li, R., Wang, L., Suganthan, P.N., Sourina, O., 2022b. Sample-based data augmentation based on electroencephalogram intrinsic characteristics. IEEE J. Biomed. Health. Inform. 26, 4996–5003. http://dx.doi.org/10.1109/JBHI.2022.3185587.

Liu, F., Chen, D., Zhou, J., Xu, F., 2022. A review of driver fatigue detection and its advances on the use of RGB-D camera and deep learning. Eng. Appl. Artif. Intell. 116, 105399. http://dx.doi.org/10.1016/j.engappai.2022.105399.

Lu, S., Wang, S.-H., Zhang, Y.-D., 2021a. Detection of abnormal brain in MRI via improved AlexNet and ELM optimized by chaotic bat algorithm. Neural Comput. Appl. 33, 10799–10811. http://dx.doi.org/10.1007/s00521-020-05082-4.

Lu, S.-Y., Zhang, Z., Zhang, Y.-D., Wang, S.-H., 2021b. CGENet: A deep graph model for COVID-19 detection based on chest CT. Biology 11, 33. http://dx.doi.org/10.3390/biology11010033.

Malik, A.K., Ganaie, M.A., Tanveer, M., Suganthan, P.N., Initiative, A.D.N.I., 2022b. Alzheimer's disease diagnosis via intuitionistic fuzzy random vector functional link network. IEEE Trans. Comput. Soc. Syst. 1–12, http://dx.doi.org/10.1109/TCSS.2022.3146974.

Malik, A.K., Gao, R., Ganaie, M.A., Tanveer, M., Suganthan, P.N., 2022a. Random vector functional link network: Recent developments, applications, and future directions. http://dx.doi.org/10.48550/arXiv.2203.11316, arXiv Preprint arXiv:2203.11316.

Niu, X.-X., Suen, C.Y., 2012. A novel hybrid CNN–SVM classifier for recognizing handwritten digits. Pattern Recognit. 45, 1318–1325. http://dx.doi.org/10.1016/j.patcog.2011.09.021.

Pao, Y.-H., Phillips, S.M., Sobajic, D.J., 1992. Neural-net computing and the intelligent control of systems. Int. J. Control 56, 263–289. http://dx.doi.org/10.1080/00207179208934315.

Pao, Y.-H., Takefuji, Y., 1992. Functional-link net computing: Theory, system architecture, and functionalities. Computer 25, 76–79. http://dx.doi.org/10.1109/2.144401.

Pelikan, M., Goldberg, D.E., Cantú-Paz, E., 1999. BOA: The Bayesian optimization algorithm. In: Proc. Genetic Evol. Comput. Conf.. pp. 525–532. http://dx.doi.org/10.5555/2933923.2933973.

Schirrmeister, R.T., Springenberg, J.T., Fiederer, L.D.J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W., Ball, T., 2017. Deep learning with convolutional neural networks for EEG decoding and visualization. Hum. Brain Mapp. 38, 5391–5420. http://dx.doi.org/10.1002/hbm.23730.

Shi, Q., Katuwal, R., Suganthan, P.N., Tanveer, M., 2021. Random vector functional link neural network based ensemble deep learning. Pattern Recognit. 117, 107978. http://dx.doi.org/10.1016/j.patcog.2021.107978.

Suganthan, P.N., Katuwal, R., 2021. On the origins of randomization-based feedforward neural networks. Appl. Soft Comput. 105, 107239. http://dx.doi.org/10.1016/j.asoc.2021.107239.

Thuwajit, P., Rangpong, P., Sawangjai, P., Autthasan, P., Chaisaen, R., Banluesombatkul, N., Boonchit, P., Tatsaringkansakul, N., Sudhawiyangkul, T., Wilaiprasitporn, T., 2022. EEGWaveNet: Multi-scale CNN-based spatiotemporal feature extraction for EEG seizure detection. IEEE Trans. Industr. Inform. 18, 5547–5557. http://dx.doi.org/10.1109/TII.2021.3133307.

Tuncer, T., Dogan, S., Ertam, F., Subasi, A., 2021a. A dynamic center and multi threshold point based stable feature extraction network for driver fatigue detection utilizing EEG signals. Cogn. Neurodyn. 15, 223–237. http://dx.doi.org/10.1007/s11571-020-09601-w.

Tuncer, T., Dogan, S., Subasi, A., 2021b. EEG-based driving fatigue detection using multilevel feature extraction and iterative hybrid feature selection. Biomed. Signal Process. Control. 68, 102591. http://dx.doi.org/10.1016/j.bspc.2021.102591.

Wang, D., Li, M., 2017. Stochastic configuration networks: Fundamentals and algorithms. IEEE Trans. Cybern. 47, 3466–3479. http://dx.doi.org/10.1109/TCYB.2017.2734043.

Wang, F., Lu, B., Kang, X., Fu, R., 2021. Research on driving fatigue alleviation using interesting auditory stimulation based on VMD-MMSE. Entropy 23, 1209. http://dx.doi.org/10.3390/e23091209.

Wei, C.-S., Wang, Y.-T., Lin, C.-T., Jung, T.-P., 2018. Toward drowsiness detection using non-hair-bearing EEG-based brain-computer interfaces. IEEE Trans. Neural Syst. Rehabilitation Eng. 26, 400–406. http://dx.doi.org/10.1109/TNSRE.2018.2790359.

Yin, Z., Zhang, J., 2018. Task-generic mental fatigue recognition based on neurophysiological signals and dynamical deep extreme learning machine. Neurocomputing 283, 266–281. http://dx.doi.org/10.1016/j.neucom.2017.12.062.

Zhang, Y., Guo, H., Zhou, Y., Xu, C., Liao, Y., 2023. Recognising drivers' mental fatigue based on EEG multi-dimensional feature selection and fusion. Biomed. Signal Process. Control 79, 104237. http://dx.doi.org/10.1016/j.bspc.2022.104237.

Zhang, L., Suganthan, P.N., 2016a. A comprehensive evaluation of random vector functional link networks. Inform. Sci. 367–368, 1094–1105. http://dx.doi.org/10.1016/j.ins.2015.09.025.

Zhang, L., Suganthan, P.N., 2016b. A survey of randomized algorithms for training neural networks. Inform. Sci. 364–365, 146–155. http://dx.doi.org/10.1016/j.ins.2016.01.039.

Zhang, Y., Wu, J., Cai, Z., Du, B., Yu, P.S., 2019. An unsupervised parameter learning model for RVFL neural network. Neural Netw. 112, 85–97. http://dx.doi.org/10.1016/j.neunet.2019.01.007.

Zheng, W.-L., Lu, B.-L., 2017. A multimodal approach to estimating vigilance using EEG and forehead EOG. J. Neural Eng. 14, 026017. http://dx.doi.org/10.1088/1741-2552/aa5a98.

Zou, W., Xia, Y., Cao, W., 2022a. Back-propagation extreme learning machine. Soft Comput. 26, 9179–9188. http://dx.doi.org/10.1007/s00500-022-07331-1.

Zou, W., Xia, Y., Cao, W., 2022b. Broad learning system based on driving amount and optimization solution. Eng. Appl. Artif. Intell. 116, 105353. http://dx.doi.org/10.1016/j.engappai.2022.105353.

Zou, Y., Yu, Z., Kumar, B.V.K.V., Wang, J., 2018. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In: Proc. Eur. Conf. Comput. Vis.. pp. 297–313. http://dx.doi.org/10.1007/978-3-030-01219-9_18.