

# Recognition of Eyelid Movement Using Electroencephalographic Signals

Wen-Lin Chu,<sup>1</sup> Chih-Jer Lin,<sup>2\*</sup> and Ching-Hao Chen<sup>2</sup>

<sup>1</sup>Department of Mechanical Engineering, National Chin-Yi University of Technology,  
No.57, Sec. 2, Zhongshan Rd., Taiping Dist., Taichung 41170, Taiwan (R.O.C.)

<sup>2</sup>Graduate Institute of Automation Technology, National Taipei University of Technology,  
1, Sec. 3, Zhongxiao E. Rd., Taipei 10608 Taiwan (R.O.C.)

(Received January 5, 2019; accepted December 9, 2019)

**Keywords:** support vector machine, wavelet transform, electroencephalography, eyelid movement

Eyelid movement patterns are a key factor in the detection of fatigue, and in this study, electroencephalography (EEG) was used to record the brainwave patterns associated with eyelid movement in subjects during various stages of fatigue. The three movements involved were no eyelid movement, closing the eye, and opening the eye. The collected signals were processed using the wavelet transform (WT) to break down the EEG signal and obtain the main features. The support vector machine (SVM) and back propagation neural network (BPNN) were used to determine eyelid movement conditions.

## 1. Introduction

Fatigue is a normal physiological reaction that usually follows a strenuous exercise, intense mental exertion, or anxiety. Fatigue is temporary and subsides with rest, even after a very intense activity or a panic attack. Fatigue can be exacerbated by insufficient sleep, a heavy workload, or undue mental stress. Driving a vehicle, when suffering from fatigue can be very dangerous and can cause traffic accidents that result in serious injury or death.

The harm that can result from driving when fatigued can be very serious. The state and degree of fatigue are closely related to eyelid movement, and in this study, physiological signals were collected to explore eyelid movement patterns. Caton<sup>(1)</sup> detected potential differences in the cerebral cortex of animals using a galvanometer in 1875, and Berger<sup>(2)</sup> recorded the first electroencephalography (EEG) in 1931. EEG equipment has now become very sophisticated, and emotional reactions<sup>(3)</sup> and concentration<sup>(4)</sup> in humans can be studied in some considerable detail,<sup>(5)</sup> and can even be used with game programs.<sup>(6)</sup> Ramachandran and Sazali<sup>(7)</sup> induced emotions by vision and hearing, and analyzed them with much channel electro-encephalogram signals. Chu *et al.*<sup>(8)</sup> collected EEG signals from schizophrenic patients, induced emotions by vision, and correlated the EEG signals with the level of illness. Studies of concentration or attention using EEG signals are an important recent focus. Morris *et al.*<sup>(9)</sup> investigated the salience of speech contrasts in noise, in relation to how listening attention affected scalp-

\*Corresponding author: e-mail: cjlin@ntut.edu.tw  
<https://doi.org/10.18494/SAM.2020.2602>

recorded cortical responses. Kim *et al.*<sup>(10)</sup> studied the EEG signals generated before an exercise in an attempt to classify the subjects' intentions. They discovered that the EEG signal generated during preparation for an exercise could control brain-machine interfaces (BMIs). Sestito *et al.*<sup>(11)</sup> used EEG signals to explore visual perception in pilots. Abiri *et al.*<sup>(12)</sup> explored various EEGs and brain-computer interface (BCI) systems, and presented the advantages and disadvantages of various examples. Chiang *et al.*<sup>(13)</sup> explored a detection model that enhances the attention of learning with EEGs; their results indicated that having a nap in the afternoon will enhance attention in a follow-up learning process. Das *et al.*<sup>(14)</sup> used EEG signals to detect the cognitive ability of concentrated and focused attention and working memory. When classified, the levels of attention were found to be 84 and 81%, respectively. Chen *et al.*<sup>(15)</sup> decomposed an EEG into multiple bands using the wavelet packet transform (WPT) and used a combination of the likelihood of synchronization (SL) and the minimum spanning tree (MST) to evaluate the degree of drowsiness in vehicle drivers. The results enhanced the *k*-nearest neighbors algorithm (KNN) classifier to 98.6%. Khalaf *et al.*<sup>(16)</sup> used EEGs to explore different degrees of attention through auditory stimulation at 40 Hz.

Recent studies related to concentration or attention have used many different methods; one of the most frequently used methods has been the wavelet transform (WT). Stankovic and Falkowski<sup>(17)</sup> proposed the WT analysis method, their main objective being to show results unavailable through the conventional Fourier transform (FT). Mallat<sup>(18)</sup> introduced multiresolution analysis to wavelet analysis and constructed a wavelet function that could decompose and reconstruct a signal. Daubechies<sup>(19)</sup> developed a hierarchically compact orthogonal wavelet, now known as the Daubechies wavelet, which is primarily applied in the discrete wavelet transform (DWT) and widely used for digital signal analysis, signal compression, and noise filtering. Courmontagne *et al.*<sup>(20)</sup> used wavelet signals to solve the problem of noise in underwater acoustic signals. Chen *et al.*<sup>(21)</sup> used EEG signals to explore the attention level and used DWT to decompose the signal and obtain information from each frequency band. Belle *et al.*<sup>(22)</sup> used electrocardiogram (ECG) and EEG signals to analyze brain stimulation in the studies of attention, and also used DWT in processing the EEG signals. Chen *et al.*<sup>(23)</sup> used the WPT to decompose an EEG signal into multiple frequency bands to evaluate the degree of drowsiness in vehicle drivers, and then used the phase lag index (PLI) to evaluate the information from each frequency band. Hazarika *et al.*<sup>(24)</sup> also used DWT to decompose an EEG signal to each frequency band when investigating how long-term action video gaming modulates the neural processes of the inhibitory control mechanism. Wang *et al.*<sup>(25)</sup> converted EEG signals to individual frequency bands using the WPT in a driver drowsiness investigation. Artificial neural networks (ANNs) are frequently used in machine learning and cognitive science; such mathematical models imitate the structure and function of biological neural networks. At present, ANN technology is widely used in pattern recognition and classification, because it is robust and has good learning capability.

The ANN learning model and the study of EEG signals often include a back propagation neural network (BPNN),<sup>(26,27)</sup> a probabilistic neural network (PNN),<sup>(28)</sup> or a general regression neural network (GRNN).<sup>(29)</sup> Such studies collect brain wave signals, select sample features, and then use a classifier to recognize eyelid motion.

## 2. Methods

### 2.1 Materials and experimental setup

In this study, a NeuroSky Mindwave headset, which uses a dry electrode, was used to collect EEG signals (Fig. 1). The use of a dry electrode, as opposed to adhesive conductive electrodes, is less limited by the environment. The sampling rate of this device is 512. The measurement position is the left frontal lobe (Fp1) and the earlobe (A1) is used as the reference point (Fig. 2).

The learning procedure used was as follows: the subjects stood in front of the computer camera and opened and closed their eyes once per second, fifteen times. The number of times of each opening and closing of the eyes was determined from the captured images and could be correlated with the brainwaves collected (for 128 data points) simultaneously. The features for each frame were determined and classified. Nine subjects took part in the experiments and the data from five, chosen at random, were used as training samples. The features from the experiments with the remaining four subjects were used as test samples for the support vector machine (SVM) and BPNN, and their recognition rate was calculated separately. The flow chart of collecting signals to recognition classification is shown in Fig. 3. The hardware equipment and software configuration of the computer of this study are shown in Table 1.

### 2.2 Signal analysis and feature calculation

The original signal collected by the EEG headset and PC program is in the time domain. This signal includes several different types of noise, increasing the difficulty in distinguishing the eyelid movement status. Therefore, processing and analysis must be carried out to separate the feature from the original noisy signal. The most common conversion method used FT, which focuses on filtering, or compressing the periodic signals, but for these noisy signals, it is not very effective. However, WT, which primarily uses the mother wavelet, gives



Fig. 1. (Color online) NeuroSky Mindwave headset.

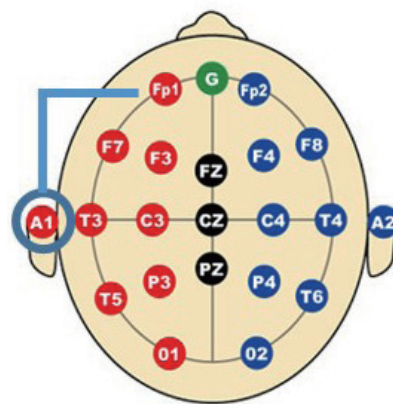


Fig. 2. (Color online) Location of electrode and reference point.

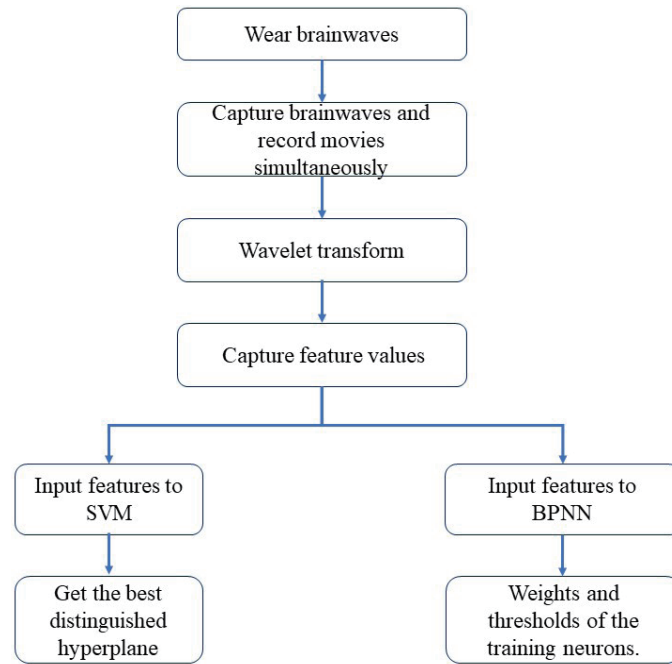


Fig. 3. (Color online) Flowchart: collection of signals, recognition, and classification.

Table 1  
Computer hardware and software.

CPU	Intel® Core™ i5-3210M
Memory	DDR3 1600 16 GB
VGA	NVIDIA GeForce GT 555M DDR3 3G
OS	Windows 7 64-bit
Analysis software	Matlab 2010a

rapid attenuation features of limited length to analyze the oscillating waveform signal. The adjustment of the coefficient allows stretching and horizontal movement, and this makes up for the deficiency of FT. WT can analyze both time domain and frequency domain signals. The basic WT formula is

$$\psi_{a,\tau}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-\tau}{a}\right), a > 0, \tag{1}$$

where  $\psi(t)$  is the mother wavelet,  $\frac{1}{\sqrt{a}}$  is the normalization factor that maintains the wavelet orthogonal base,  $\tau$  is the translation parameter, and  $\frac{1}{\sqrt{a}} \psi\left(\frac{t-\tau}{a}\right)$  is the dilation of the mother wavelet. Although a horizontal movement is achieved through  $(t - \tau)$ , different  $a$  and  $\tau$  values will have different effects on the mother wavelet, as shown in Fig. 4.

DWT is the simplified form of the continuous wavelet transform (CWT). Since CWT calculates the inner product at different times and scales of the mother wavelet, that will

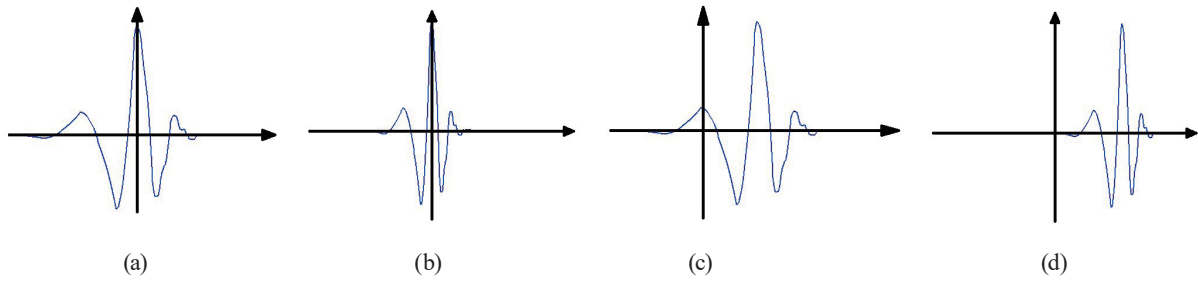


Fig. 4. (Color online) Effect of different parameters on the mother wavelet. (a)  $\Psi(t)$ , (b)  $\Psi(t/a)$ , (c)  $\Psi(t - \tau)$ , and (d)  $\Psi((t - \tau)/a)$ .

increase during calculation. DWT sends the original signal through both high-pass and low-pass filters in the wavelet and scaling functions, respectively. The brainwave signal will be divided into an approximated signal and a detailed signal after passing the high-pass and low-pass filters; the related equations are

$$\varphi_{j,k}[t] = 2^{\frac{j}{2}} \sum_k d_{j,k} \varphi[2^j t - k], \quad (2)$$

$$\psi_{j,k}[t] = 2^{\frac{j}{2}} \sum_k c_{j,k} \psi[2^j t - k]. \quad (3)$$

Here,  $\varphi(t)$  is the wavelet function,  $\psi(t)$  is the scaling function, and  $d_j$  and  $c_j$  are the wavelet and scaling coefficients of the  $j$  layer, respectively. In this study, the Daubechies<sup>(19)</sup> wavelet function is used for signal decomposition. In Matlab, the Daubechies wavelet function is expressed in the form of dbA. A is the vanishing moment of the Daubechies wavelet. The db4 of the Daubechies wavelet family is used in this study. Wave diagrams of the wavelet and scaling function waveforms are shown in Fig. 5. The horizontal axis shows the time and the vertical axis shows the amplitude. The Daubechies wavelet is mainly used for discrete wavelet conversion and frequently used in digital signal analysis

After the db4 WT, the corresponding frequency band of the original brainwave can be calculated as

$$f = 2^j \left( \frac{F_s}{N_p} \right), \quad (4)$$

where  $f$  is the upper limit of the frequency at level  $j$ ,  $F_s$  is the sampling frequency, and  $N_p$  is the number of input data points. Assuming that the brainwave signal being read is  $X_N(n)$ ,  $N$  is the  $N$ th A5 frequency band data. After capturing each frequency band via WT, the following can be derived: maximum, minimum, summation, range, standard deviation, and median absolute deviation.

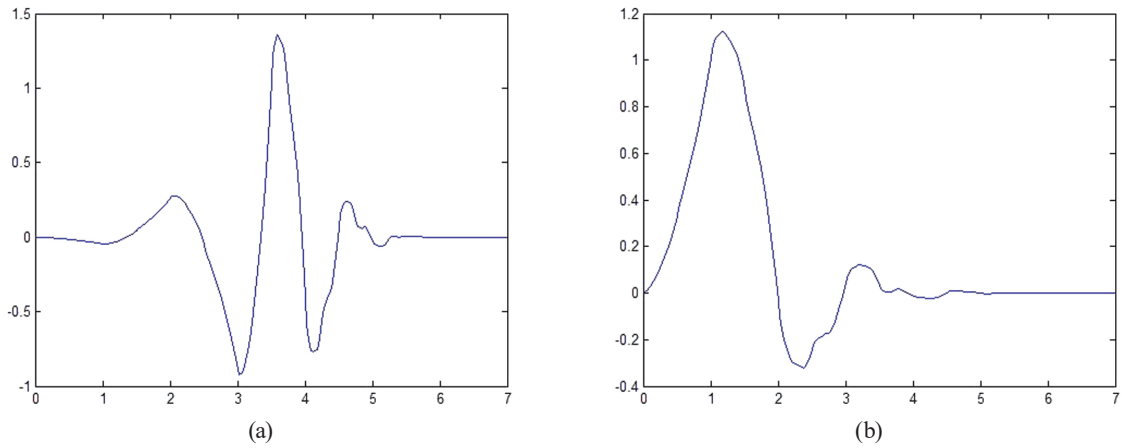


Fig. 5. (Color online) db4 wavelet function. (a) Diagram of wavelet function. (b) Diagram of scaling function.

$$X_{N,Max} = \max_{0 \leq n \leq N} (X_N(n)) \quad (5)$$

$$X_{N,min} = \min_{0 \leq n \leq N} (X_N(n)) \quad (6)$$

$$X_{N,Sun} = \sum_{n=1}^{\tilde{N}} X_N(n) \quad (7)$$

$$X_{N,Range} = X_{N,Max} - X_{N,min} \quad (8)$$

$$X_{N,\sigma} = \sqrt{\frac{1}{\tilde{N}} \sum_{n=1}^{\tilde{N}} (X_N(n) - X_{N,\mu})^2} \quad (9)$$

In Eq. (9),  $X_{N,\mu} = \frac{1}{\tilde{N}} \sum_{n=1}^{\tilde{N}} X_N(n)$ .

$$X_{N,MAD} = \text{median}(|X_N(n) - \text{median}(X_N(n))|) \quad (10)$$

### 2.3 SVM

SVM is a method proposed by Vapnik<sup>(30)</sup> in 1999. It solves many problems of classification and is very popular for machine learning. SVM is a type of supervised learning network that can establish input and output planes (a hyperplane) in the training data and predict output results of the corresponding input data via the hyperplane. SVM can be linear or nonlinear. Linear SVM finds a separating hyperplane from the input training data to maximize the margin of the two types of data. The hyperplane can be distinguished by definition as  $f: x \cdot w + b$ . If  $f$  is greater than 0, that piece of data is +1; if it is less than 0, it is classified as -1. However, according to this type of classification,  $w$  and  $b$  will have infinite combinations. The main

objective is to distinguish the hyperplane from the largest area to maximize the separation of data and use this method to reduce test errors effectively. The conditions must satisfy

$$x_i \cdot w + b \geq 1 \text{ for } y_i = +1, \quad (11)$$

$$x_i \cdot w + b \leq 1 \text{ for } y_i = -1. \quad (12)$$

Equations (11) and (12) can be combined into an inequality equation such as

$$y_i(x_i \cdot w + b) - 1 \geq 0. \quad (13)$$

The distance  $x \cdot w + b = 0$  can be calculated from Eqs. (11) and (12), where the distance is  $1/\|w\|$  and the boundary is  $2/\|w\|$ . To find the maximum boundary of the hyperplane, it is necessary to find the minimum  $w^2$  under the condition of Eq. (13). Any  $x_i$  that validates the equal sign is a support vector. Lagrange optimization can be used to find the minimum  $\|w\|^2$ . The Lagrange function is expressed as

$$L_p(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i [y_i (w \cdot x_i + b) - 1]. \quad (14)$$

In Eq. (14),  $\alpha_i$  is the Lagrange coefficient, and  $\alpha_i > 0$ . The partial differentiation of  $w$  and  $b$  in Eq. (14) making it equal to 0 allows the identities of Eqs. (15) and (16) to be obtained.

$$w = \sum_{i=1}^l y_i \alpha_i x_i \quad (15)$$

$$\sum_{i=1}^l y_i \alpha_i = 0 \quad (16)$$

Substituting Eq. (15) into Eq. (14) can yield the dual form and transforms the problem from minimization to maximization as shown in Eq. (17).

$$\begin{aligned} \text{maximize } L_D(\alpha) &= \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \\ \text{subject to } \alpha_i &\geq 0, \forall i = 1, \dots, l \text{ and } \sum_{i=1}^l \alpha_i y_i = 0 \end{aligned} \quad (17)$$

Then, by the substitution of the optimal solution ( $w^* \cdot b^*$ ) based on Karush–Kuhn–Tucker (KKT) conditions, Eq. (18) is obtained.  $\alpha_i^*$  must be greater than or equal to zero, and the result  $y_i(w_i^* \cdot x_i + b) - 1$  must be equal to zero, as shown in Eq. (18).



$$\sum_{i=1}^m \alpha_i^* [y_i^* (w^* \cdot x_i + b^*) - 1] = 0 \tag{18}$$

Equation (18) can be satisfied if  $x_i$  exists.  $x_i$  is the vector that is closest to the optimization differentiation hyperplane, that is, if there is one  $x_i$ 's  $\alpha_i^* \geq 0$ , it can be regarded as a support vector. Finding a support vector is the same as finding the maximum boundary. Finally, a function that classifies the problem can be generalized as

$$f(x) = \text{sgn} \left( \sum_{i=1}^m y_i \alpha_i \cdot (x_i \cdot x_i) + b \right). \tag{19}$$

When  $f(x) > 0$ , the classification of that piece of data is the same as the classification and label of the data; if it is different, then it is in another category. Another important point of SVM is the kernel function. SVM can map the input dataset into the feature space using the kernel function. Different kernel functions give different classifications. The common kernel functions include linear, polynomial, radial basis function (RBF), and sigmoid. In this study, the polynomial kernel function [Eq. (20)] was used for classification.

$$k(x_i, x_j) = (x_i^T x_j + 1)^d \tag{20}$$

### 2.4 BPNN

An ANN is a biomimetic neural network, which is connected by many artificial neurons to process calculation, and uses the transform function of the weighted input value product to represent the identity of the input and output values. An artificial neuron model is shown in Fig. 6.

As shown in Fig. 6, the output value of the  $j$  unit in the  $n$ th layer is the nonlinear function of the output value of the  $n - 1$  layer unit shown as

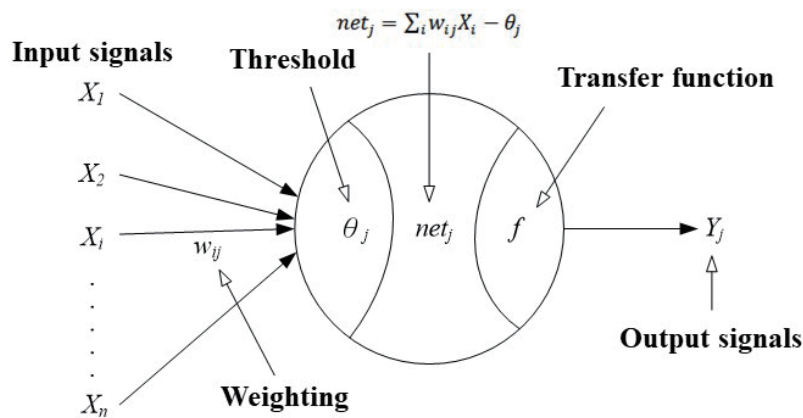


Fig. 6. Artificial neuron model.



$$Y_j^n = f\left(\text{net}_j^n \sum_{i=1}^n w_{ij} Y_i^{n-1} - \theta_j\right) = f(\text{net}_j^n). \quad (21)$$

In the above formula,  $\text{net}_j^n$  is the integrated function, while  $f$  is the transform function.

BPNN is the most common application and the current neural network learning model. The model inputs features from the input layer, transmits the data to the hidden layer with an initial set weight, and transforms the input sum to the variables of the hidden layer via the transfer function. During the process, the transform function is responsible for the summation and transformation of the input signals, and transmits the signals to the next layer. The most commonly used transform function is the S-type, which has both forward and negative convergences. The S-type transform function that was selected for use in this study is the hyperbolic tangent function shown as

$$f(\text{net}_j) = \frac{e^{\text{net}_j} - e^{-\text{net}_j}}{e^{\text{net}_j} + e^{-\text{net}_j}}. \quad (22)$$

The hidden layer can increase the complexity of the neural network by the simulation of many complex nonlinear relationships. However, if there are too many hidden layers, the memorization of the data structure of the training group and the generation of the over fitting status are not a trivial matter for the neural network. There will be further weight calculation and summation, from the hidden layer to the output layer, to enable hidden layer information to be transformed to output data. After this, the error function is used to calculate the difference between the ideal output value and the output value calculated by the neural network. The error function used in this study is the mean squared error (MSE) and is shown as

$$E = \frac{1}{2} \sum_{k=1}^n (T_k - Y_k)^2, \quad (23)$$

where  $T_k$  is the target value of the output layer and  $Y_k$  is the inference value of the output layer calculated by the neural network. After the completion of the forward pass, backward pass is entered, pushes back from the output layer, and updates the weights to minimize the error function. The learning process of the entire BPNN facilitates the minimization of the error function. This usually employs the gradient steepest descent method to minimize the error function. When a training sample is input, the weight in the network will be adjusted slightly. The sensitivity of the error function to the weighted value is proportional to the level of adjustment. The equation of the weight is shown as

$$\Delta W_{ij} = -\eta \frac{\partial E}{\partial w_{ij}}. \quad (24)$$

In Eq. (24),  $W_{ij}$  is the weight between the  $i$ th processing unit of the  $n - 1$  layer and the  $j$ th processing unit of the  $n$  layer.  $\eta$  is the learning rate that controls the step size of the error

function and is minimized each time by using the steepest descent method. After using the chain rate to expand  $\partial E/\partial w_{ij}$ , Eq. (25) is obtained:

$$\frac{\partial E}{\partial w_{ij}} = \left( \frac{\partial E}{\partial net_j^n} \right) \left( \frac{\partial net_j^n}{\partial W_{ij}} \right) = \left( \frac{\partial E}{\partial A_j^n} \right) \left( \frac{\partial A_j^n}{\partial net_j^n} \right) \left( \frac{\partial net_j^n}{\partial W_{ij}} \right). \quad (25)$$

Substituting the integrated function into Eq. (25) gives

$$\frac{\partial net_j^n}{\partial W_{ij}} = \frac{\partial}{\partial W_{ij}} \left( \sum_{k=1}^n W_{kj} A_k^{n-1} - \theta_j \right) = A_i^{n-1}. \quad (26)$$

Substituting Eq. (21) into Eq. (25) gives

$$\frac{\partial A_j^n}{\partial net_j^n} = \frac{\partial}{\partial net_j^n} f(net_j^n) = f'(net_j^n). \quad (27)$$

If the  $n$ th layer is the final layer, then the substitution of Eq. (23) gives

$$\frac{\partial E}{\partial net_j^n} = \frac{\partial}{\partial net_j^n} \left[ \frac{1}{2} \sum_{k=1}^n (T_k - A_k)^2 \right] = -(T_j - A_j^n). \quad (28)$$

If the  $n$ th layer is not the final layer, but one of the hidden layers of the network, then Eq. (29) can be derived as

$$\frac{\partial E}{\partial A_j^n} = - \sum_{k=1}^n \delta_k^{n+1} \cdot W_{jk}. \quad (29)$$

Finally,  $\partial E/\partial w_{ij}$  can be written as the general equation

$$\frac{\partial E}{\partial w_{ij}} = -\delta_j^n \cdot A_i^{n-1}. \quad (30)$$

In the above equation,  $A_i^{n-1}$  is the output value of the lower layer of the processing unit connected by  $W_{ij}$ .  $\delta_j^n$  is the difference volume of the processing unit at the upper layer connected by  $W_{ij}$ . If the processing unit at the upper layer connected by  $W_{ij}$  is the output layer, then  $\delta_j^n = (T_j - Y_j) \cdot f'(net_j^n)$ . Substituting Eq. (30) into Eq. (24) gives

$$\Delta W_{ij} = \eta \cdot \delta_j^n \cdot A_i^{n-1}. \quad (31)$$

The nonlinear transform function used in this study was a hyperbolic tangent function. If we follow Eq. (27) to differentiate  $net_j^n$  in Eq. (22), Eq. (32) can be derived as

$$f'(net_j) = [1 + f(net_j)] \cdot [1 - f(net_j)]. \quad (32)$$

The difference volume  $\delta$  of the output layer can be derived by the steepest descent method as

$$\delta_j = (T_j - Y_j) \cdot f'(net_j) = (T_j - Y_j) \cdot (1 + Y_j) \cdot (1 - Y_j). \quad (33)$$

Similarly, the equation for the change in the volume of the threshold  $\theta$  is as shown as

$$\Delta\theta = -\eta \cdot \delta_j^n. \quad (34)$$

From the above derivations, it can be seen that every time there is a backward processing pass, the difference between the weight and the threshold can be obtained. By gradually updating the weight and threshold, the error function can be converged.

### 3. Results and Discussion

The brainwave sampling rate used was 512 Hz, but the time taken by the action of opening and closing the eyes was found to be about 0.25. To obtain the waveforms associated with the actual opening and closing eyelid movements, and avoid excessive signals from eyes that were not opening or closing, the number of inputs points was reduced to 128, as shown in Fig. 7. The overlap was set to 78% ( $1 - 28/128 = 0.78$ ), and only 28 pieces of brainwave data were updated at a time. This prevented the horizontal distance between frames that captured a brainwave from becoming too large, which would cause recording failure. The brainwave obtains the bandwidth

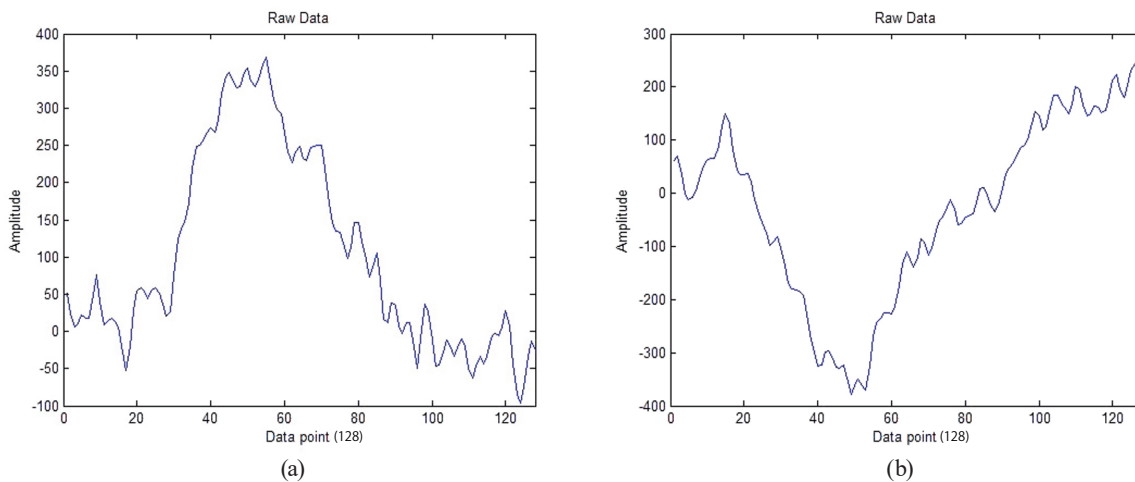


Fig. 7. (Color online) Brainwave signal waveforms: (a) Eyes closing. (b) Eyes opening.

corresponding to the wavelet of each order and the brainwave frequency band corresponding to the frequency bandwidth via Eq. (4), as shown in Table 2.

The transformation of the original brainwave using a five-hierarchy WT allows various types of brainwave to be obtained and, in this study, the original brainwaves were decomposed to allow a search for a waveform that could be used as a basis for classification. The results are shown in Fig. 8.

As can be seen in Fig. 8, the opening and closing of the eyes can be clearly distinguished in the A5 wavelet-transformed frequency band, which was chosen for classification in this study.

### 3.1 Classification results of various feature combinations

The classification correction rate equations used in this study to analyze the above-captured frequency bands and select the features that could be used as the basis for major classification are shown below:

$$\text{Correct rate of classification (\%)} = \frac{\text{correct classification}}{\text{Total number of classification data}} (\%), \quad (35)$$

$$\text{Average correct rate of classification (\%)} = \frac{\text{correct classification}}{\text{Frequency of experment implementation}} (\%). \quad (36)$$

Table 2  
db4 corresponding brainwave types after WT.

Resolution (wavelet level)	Frequency bandwidth (Hz)	Associated brainwave
0 (A5)	0–4	Delta
1 (D5)	4–8	Theta
2 (D4)	8–16	Alpha
3 (D3)	16–32	Beta
4 (D2)	32–64	Gamma
5 (D1)	64–128	Noise

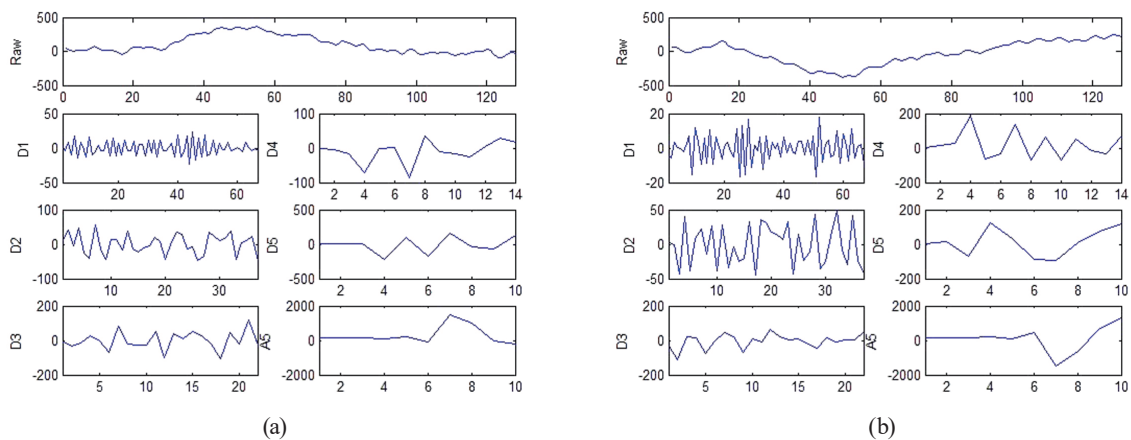


Fig. 8. (Color online) Hierarchic WTs: (a) Eyes closing. (b) Eyes opening.

Since the SVM and BPNN classifiers were used in this study, the experiments were performed separately in the feature selection section, and the features that offered good rates were used for subsequent experiments.

Three different modes were used in the following correct classification rate tests. Test A, all the data were used for training and testing; Test B, half the data were used for training and all the data were used for testing; and Test C, half the data were used for training and the other half for testing. The results of the three tests were then compared.

The signals of the A5 frequency band were used to calculate six features, namely, maximum, minimum, summation, range, standard deviation, and median absolute deviation. All the features were input into SVM and BPNN, and the classification results are shown in Table 3.

The experimental results in Table 3 show that SVM and BPNN were stable to a certain extent and had good classification accuracy. Both are suitable for the classification of brain waveforms. The selection of features is detailed in the next section and more meaningful features were selected to improve recognition.

### 3.2 Classification of features correction rate

In this part of the study, six features were classified as groups to compare the classification correction rate. Each group had two features and the test method chosen was Test C. Figure 9 shows the separation hyperplane diagram drawn from SVM classification data using maximum and minimum values. Table 4 shows the classification results obtained using SVM and BPNN in three groups covering the maximum and minimum, the summation and range, and the

Table 3  
Classification correction rate.

		Test A	Test B	Test C
Training and testing number of data	Training sample	180	90	90
	Test sample	180	180	90
SVM		97.22%	88.89%	82.22%
BPNN		93.33%	91.67%	88.89%

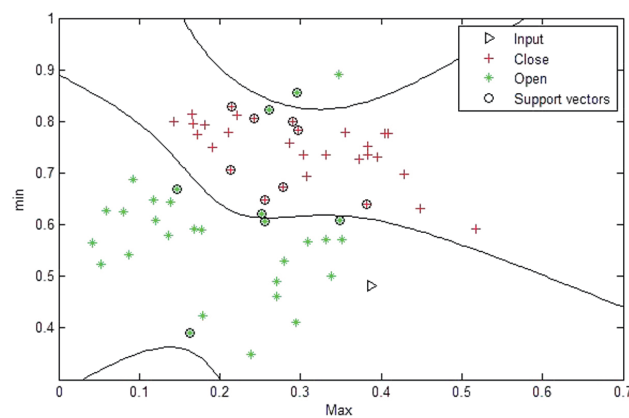


Fig. 9. (Color online) Separation hyperplane for two-feature input.

Table 4  
Classification correction of two features.

Feature	Classifier	Classification correction rate (%)
Maximum and minimum	SVM	95.83
	BPNN	95.00
Summation and range	SVM	78.33
	BPNN	83.33
Standard deviation and median absolute deviation	SVM	53.33
	BPNN	50.00

standard deviation and median absolute deviation. The results show that the classification correction rate is highest when maximum and minimum rates are used. This is because the largest difference between the maximum and the minimum is observed for the closing and opening of eyes.

In this experiment, two features were classified as one group and SVM and BPNN classifications were used to correct the rate experiments. When the number of input features was changed to three, the SVM separation hyperplane of SVM was also transformed into three-dimensional (3D) space. In Fig. 10, the  $x$ -axis is maximum, the  $y$ -axis is minimum, and the  $z$ -axis is summation. The comparative experiments of the three features are organized and shown in Table 5. The results show that the classification correction rates of maximum and minimum features are closely related and also related to the summation and range. On the other hand, the results also show that the standard deviation and median absolute deviation will reduce the classification correction rate; hence, the classification of four features in the next group will focus on the maximum and minimum features, and group with other features to find the four features that provide the highest classification correction rate.

The classification correction rate results of the SVM and BPNN classifications that adopted four features are shown in Table 6.

The experimental data in this section show that when the maximum, minimum, summation, and range are defined as input features, although the classification correction rate of SVM remains at around 95.00% and does not increase, the classification correction rate of BPNN can go up to 98.33%, which is very close to the expected correction rate for the offline classification of this study. Therefore, the follow-up online classification experiments used the maximum, minimum, summation, and range to go with BPNN to carry out the online classification correction rate experiment.

### 3.3 Online real-time detection recognition results

The features finally used in this study were the maximum, minimum, summation, and range. However, online real-time detection and recognition are not possible to confirm the data collection status and implement cutting by replaying the video clips. Therefore, this experiment used 128 original data as signal cuts and captured 28 data each time for data update; this resulted in an overlap of 78%. Each subject closed and opened their eyes 20 times, and the results of online classification are shown in Table 7. We determined from the results that online real-time recognition had an average recognition rate of 85%.

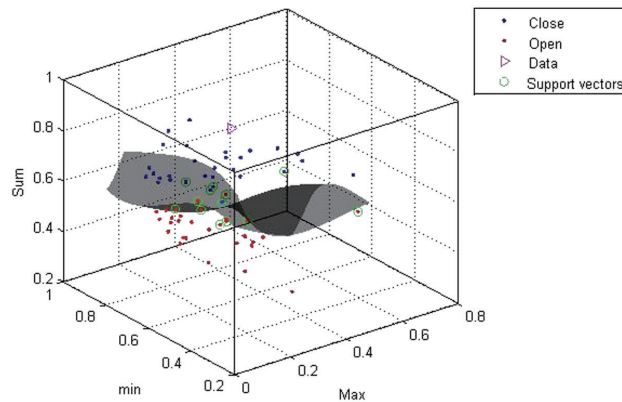


Fig. 10. (Color online) Separation hyperplane for three-feature inputs.

Table 5

Three-feature input.

Feature	Classifier	Classification correction rate (%)
Maximum, minimum, and summation	SVM	95.83
	BPNN	96.67
Minimum, summation, and range	SVM	96.67
	BPNN	96.67
Summation, standard deviation, and median absolute deviation	SVM	78.33
	BPNN	82.5

Table 6

Four-feature correction rates.

Feature	Classifier	Classification correction rate (%)
Maximum, minimum, summation, and range	SVM	95.83
	BPNN	98.33
Maximum, summation, range, and standard deviation	SVM	95.00
	BPNN	95.00
Maximum, minimum, standard deviation, and median absolute deviation	SVM	93.33
	BPNN	97.5

Table 7

Statistical results of online real-time detection recognition.

	Total data	Number of correct classifications	Classification correction rate (%)
Subject 1	40	37	92.5
Subject 2	40	36	90
Subject 3	40	30	75
Subject 4	40	34	85
Subject 5	40	33	82.5
Average recognition rate			85

## 4. Conclusions

In this study, we used single-point EEGs and focused on the closing and opening of the eyes of the subjects to analyze the level of fatigue. The signal analysis methods used the signals in the A5 frequency band after WT. After the capture of six features from the signals, SVM and



BPNN were used for training and analysis, and to determine the correlation between eye closing and opening. Experimental results showed that the calculation method using four features (maximum, minimum, summation, and range) was suitable for recognizing eye closing and opening movements, and that the maximum and minimum features were the most relevant in this respect. In the offline case, the recognition rate was up to 98.33%, and the average online recognition rate reached 85%. It is expected that these results will contribute to studies related to driving safety in the future.

### Acknowledgments

The authors would like to thank the National Science Council of the Republic of China, Taiwan, for financial support of this research under Contract Nos. MOST 106-2221-E-027-058 and MOST 108-2218-E-167-006.

### References

- 1 R. Caton: *J. Br. Med. J.* **2** (1875) 257. <https://doi.org/10.1136/bmj.2.765.257>
- 2 H. Berger: *Arch. Psychiatrie Nervenkrankheiten* **94** (1931) 16. <https://doi.org/10.1007/bf01835097>
- 3 W. L. Chu, M. W. Huang, B. L. Jian, and K. S. Cheng: *Ann. Gen. Psychiatry* **16** (2017) 34. <https://doi.org/10.1186/s12991-017-0157-z>
- 4 F. Fahimi, W. B. Goh, T.-S. Lee, and C. Guan: In: *J. Crowd Sci.* **2** (2018) 272. <https://doi.org/10.1108/ijcs-09-2018-0022>
- 5 B.-S. Lin, Y.-K. Huang, and B.-S. Lin: *Comput. Methods Programs Biomed.* **178** (2019) 41. <https://doi.org/10.1016/j.cmpb.2019.06.009>
- 6 Y. Li, J. Zhang, C. Zhang, Y. Xiao, J. Hong, M. Y. Wang, and Y. Li: 2018 15th Int. Conf. Ubiquitous Robots (UR) (2018) 749. <https://doi.org/10.1109/URAI.2018.8441816>
- 7 M. M. N. Ramachandran and Y. Sazali: *J. Bio. Sci. Eng.* **3** (2010) 390. <https://doi.org/10.4236/jbise.2010.34054>
- 8 W. L. Chu, M. W. Huang, B. L. Jian, and K. S. Cheng: *Ann. Gen. Psychiatry* **16** (2017) 34. <https://doi.org/10.1186/s12991-017-0157-z>
- 9 D. J. Morris, J. Tondering, and M. Lindgren: *Hear. Res.* **373** (2019) 1. <https://doi.org/10.1016/j.heares.2018.12.001>
- 10 H. Kim, N. Yoshimura, and Y. Koike: *Front. Hum. Neurosci.* **13** (2019) 63. <https://doi.org/10.3389/fnhum.2019.00063>
- 11 M. Sestito, A. Harel, J. Nador, and J. Flach: *Front. Hum. Neurosci.* **12** (2018) 489. <https://doi.org/10.3389/fnhum.2018.00489>
- 12 R. Abiri, S. Borhani, E. W. Sellers, Y. Jiang, and X. Zhao: *J. Neural. Eng.* **16** (2019) 011001. <https://doi.org/10.1088/1741-2552/aaf12e>
- 13 H. S. Chiang, K. L. Hsiao, and L. C. Liu: *J. Med. Biol. Eng.* **38** (2018) 847. <https://doi.org/10.1007/s40846-017-0344-z>
- 14 N. Das, A. Bertrand, and T. Francart: *J. Neural. Eng.* **15** (2018) 066017. <https://doi.org/10.1088/1741-2552/aae0a6>
- 15 J. Chen, H. Wang, and C. Hua: *Int. J. Psychophysiol.* **133** (2018) 120. <https://doi.org/10.1016/j.jpsycho.2018.07.476>
- 16 A. Khalaf, J. Kersey, S. Eldeeb, G. Alankus, E. Grattan, L. Waterstram, E. Skidmore, and M. Akcakaya: *J. Neurosci. Methods* **303** (2018) 169. <https://doi.org/10.1016/j.jneumeth.2018.03.019>
- 17 R. S. Stankovic and B. J. Falkowski: *Comput. Electr. Eng.* **29** (2003) 25. [https://doi.org/10.1016/S0045-7906\(01\)00011-8](https://doi.org/10.1016/S0045-7906(01)00011-8)
- 18 S. G. Mallat: *IEEE Trans. Pattern Anal. Mach. Intell.* **11** (1989) 674. <https://doi.org/10.1109/34.192463>
- 19 I. Daubechies: *SIAM J. Math. Anal.* **24** (1993) 499. <https://doi.org/10.1137/0524031>
- 20 P. Courmontagne, S. Ouelha, U. Moreaud, and F. Chaillan: 2013 OCEANS - San Diego (2013) 1. <https://doi.org/10.23919/OCEANS.2013.6741063>
- 21 C. M. Chen, J. Y. Wang, and C. M. Yu: *Br. J. Educ. Technol.* **48** (2017) 348. <https://doi.org/10.1111/bjet.12359>
- 22 A. Belle, R. H. Hargraves, and K. Najarian: *Comput. Math. Methods Med.* **2012** (2012) 528781. <https://doi.org/10.1155/2012/528781>

- 23 J. Chen, H. Wang, C. Hua, Q. Wang, and C. Liu: Cogn. Neurodyn. **12** (2018) 569. <https://doi.org/10.1007/s11571-018-9495-z>
- 24 J. Hazarika, P. Kant, R. Dasgupta, and S. H. Laskar: Biomed. Signal Process. Control **45** (2018) 144. <https://doi.org/10.1016/j.bspc.2018.05.023>
- 25 H. Wang, A. Dragomir, N. I. Abbasi, J. Li, N. V. Thakor, and A. Bezerianos: Cogn. Neurodyn. **12** (2018) 365. <https://doi.org/10.1007/s11571-018-9481-5>
- 26 M. Y. Liu, J. Wang, and C. X. Zheng: Non-Negative Matrix Factorizations Based Spontaneous Electroencephalographic Signals Classification Using Back Propagation Feedback Neural Networks: In Advances in Neural Networks - Isnn 2005, Pt 3, Proceedings (Springer-Verlag Berlin, Berlin, 2005) p. 731. [https://doi.org/10.1007/11427469\\_116](https://doi.org/10.1007/11427469_116)
- 27 A. Aarabi, R. Grebe, and F. Wallois: Clin. Neurophysiol. **118** (2007) 2781. <https://doi.org/10.1016/j.clinph.2007.08.012>
- 28 R. Yuvaraj, U. R. Acharya, and Y. Hagiwara: Neural Comput. Appl. **30** (2018) 1225. <https://doi.org/10.1007/s00521-016-2756-z>
- 29 P. Swami, T. K. Gandhi, B. K. Panigrahi, M. Tripathi, and S. Anand: Expert Syst. Appl. **56** (2016) 116. <https://doi.org/10.1016/j.eswa.2016.02.040>
- 30 V. N. Vapnik: IEEE Trans. Neural Netw. **10** (1999) 988. <https://doi.org/10.1109/72.788640>

### About the Authors



**Wen-Lin Chu** received her B.S. degree from the Department of Mechanical and Computer-Aided Engineering, Feng Chia University in 2009, her M.S. degree from the Graduate Institute of Automation Control from the National Taiwan University of Science and Technology in 2011, and her Ph.D. degrees from the Department of Biomedical Engineering, National Cheng Kung University in 2017. She is currently an assistant professor with the Department of Mechanical Engineering, National Chin-Yi University of Technology, Taichung, Taiwan. Her current research interests include image processing, physiological signal processing, and automatic analysis techniques. ([wlchu@ncut.edu.tw](mailto:wlchu@ncut.edu.tw))



**Chih-Jer Lin** received his B.S., M.S., and Ph.D. degrees in mechanical engineering from National Cheng Kung University, Tainan, Taiwan, in 1992, 1994, and 1998, respectively. He is currently a professor and the director of the Graduate Institute of Automation and Technology, National Taipei University of Technology, Taipei, Taiwan. His current research interests include mechatronics, vibration control, motion control, system identification, sliding mode control, robotics, and evolutionary algorithms. Professor Lin has been a guest editor of Applied Sciences and a member of the editorial board of the Journal of Chinese Society of Mechanical Engineers. ([cjlin@ntut.edu.tw](mailto:cjlin@ntut.edu.tw))



**Ching-Hao Chen** received his M.S. degree in automation technology from National Taipei University of Technology, Taiwan, in 2014. He is currently an engineer in CTCI Advanced Systems. His research interests are in mechatronics and neural networks. ([luke03142001@yahoo.com.tw](mailto:luke03142001@yahoo.com.tw))