



Variable weight neural networks and their applications on material surface and epilepsy seizure phase classifications

H.K. Lam ^{a,*}, Udeme Ekong ^a, Bo Xiao ^a, Gaoxiang Ouyang ^b, Hongbin Liu ^a, K.Y. Chan ^c, Sai Ho Ling ^d

^a Department of Informatics, King's College London, Strand, London WC2R 2LC, United Kingdom

^b State Key Laboratory of Cognitive Neuroscience and Learning, School of Brain and Cognitive Sciences, Beijing Normal University, No. 19, XijieKoWai St., HaiDian District, Beijing 100875, PR China

^c Department of Electrical and Computer Engineering, Curtin University, Perth, Australia

^d Centre for Health Technologies, Faculty of Engineering and Information Technology, University of Technology, Sydney, NSW, Australia

ARTICLE INFO

Article history:

Received 5 May 2014

Received in revised form

21 July 2014

Accepted 6 September 2014

Communicated by Ligang Wu

Available online 16 September 2014

Keywords:

Material recognition

Epilepsy signals

Bayesian decision

KNN

Neural networks

Variable-weight neural networks

ABSTRACT

This paper presents a novel neural network having variable weights, which is able to improve its learning and generalisation capabilities, to deal with classification problems. The variable weight neural network (VWNN) allows its weights to be changed in operation according to the characteristic of the network inputs so that it can adapt to different characteristics of input data resulting in better performance compared with ordinary neural networks with fixed weights. The effectiveness of the VWNN is tested with the consideration of two real-life applications. The first application is on the classification of materials using the data collected by a robot finger with tactile sensors sliding along the surface of a given material. The second application considers the classification of seizure phases of epilepsy (seizure-free, pre-seizure and seizure phases) using real clinical data. Comparisons are performed with some traditional classification methods including neural network, k-nearest neighbours and naive Bayes classification techniques. It is shown that the VWNN classifier outperforms the traditional methods in terms of classification accuracy and robustness property when the input data is contaminated with noise.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Classification is a process that takes samples from objects and assigns each one of them to a pre-defined group or class label. This is a promising and important field of research which provides a solution to a wide range of applications e.g., classification of different investments and lending opportunities as acceptable or unacceptable risk [1], classification of electrocardiogram (ECG) arrhythmias [2], classification of ECG beat [3], facial recognition [4,5], hand-writing recognition [6–10], heart sound classification [11], human body posture classification [12], speaker verification [13], speech recognition [14,15] and text classification [16,17].

In general, a classification process usually consists of three main stages. In the first stage, data from the objects have to be collected for the design of classifiers. In the second stage, feature extraction is

performed to extract characteristics from the collected data to be classified such that redundant information is removed and representative information is extracted resulting in reduction of input dimensions and improved classification accuracy. In the third stage, a classifier is designed using the extracted feature data.

Existing methods from the literature review conducted include traditional and machine learning methods which are listed and briefly described below. An example of traditional methods is the cover linear discriminant analysis [18], logic based method (e.g., decision trees [19]), statistical approach (e.g., Bayesian classification [20]) and instance-based methods (e.g., nearest neighbour algorithm [21,22]). Machine learning methods include the support vector machines and neural network (NN) [4,9,13,17].

Bayesian decision theory [20] is one of the most important methods in statistical classification which offers a primary model for further classification procedures. Naive Bayesian classifier is based on the assumption that equal prior probabilities exists for all classes [23] this reduces the complexity of analysis and helps in resolving conflicts that occur when two or more classes are not well separable, resulting in improving the classification accuracy. Although Bayesian decision is simple and powerful, the posterior probabilities cannot be

* Corresponding author.

E-mail addresses: hak-keung.lam@kcl.ac.uk (H.K. Lam), udeme.ekong@kcl.ac.uk (U. Ekong), bo.xiao@kcl.ac.uk (B. Xiao), ouyang@bnu.edu.cn (G. Ouyang), hongbin.liu@kcl.ac.uk (H. Liu), kit.chan@curtin.edu.au (K.Y. Chan), steve.Ling@uts.edu.au (S. Ho Ling).

determined directly [24]. A recent development of the Bayesian classifier has been the proposal of a hybrid Bayesian classifier [23]. The Bayesian classifier has also been successfully implemented to many real-world applications e.g. weeds identification [20]. The k-nearest neighbour technique (kNN) [21] is easy to apply and good in dealing with text based problems such as visual category recognition [22]. However, kNN has its intrinsic limitations, the main disadvantages of the kNN technique are the large memory requirements and the lack of a logical way to choose the best 'k', this would introduce difficulties to a classification application as different data sets require an optimised value of 'k' to improve the performance of this method [25]. Furthermore, the precision accuracy of kNN will be declined when there are too many classes to deal with or when an uneven density of training samples is presented.

Machine learning methods include single layer perceptron [26], artificial neural networks [7,24,27], support vector machines, neural-fuzzy networks [2,3,8,12,15] and self-organisation map [6,11,16,28]. The first neural networks were designed based on mathematics and algorithms in the 1940 s. The McCulloch–Pitts neuron was proposed in 1943, this laid the foundation of modern neural networks [27]. However, there was no effective neural network training algorithm, so the development of neural networks was stagnated for some years. After that, a trainable network with adaptive elements, which are the building blocks, was designed [27]. A single layer perceptron [26] neural network model was first introduced by Rosenblatt in 1962, which cast a huge impact on the artificial intelligence field, and then different types of perceptron-based techniques have emerged in large numbers. A single layer perceptron has a simple structure which can be seen as a component that just weighted the inputs and then computes the sum to the output of the system. After that, the outputs are used to compare with the corresponding targets to verify the accuracy. With the information of difference between outputs and targets, the weights can be adjusted to achieve higher level of accuracy. However, there is a major limitation that restrains the applications, only linearly separable problems can be solved by the single layer perceptron. Although it has the major limitation, the single layer perceptron has also been implemented well to finger print matching [26] and image detection [29] applications. In addition to applications, the system stability of systems [30–32] invoking the NN are also investigated. Traditionally, a feedforward neural network [33] has three layers (input, hidden, and output layers) of nodes connected in a layer-to-layer manner. Neural Networks have various applications due to its favourable approximation performance and convenient modelling process. However, ordinary neural networks suffer from the 'over-fitting' and 'local optimisation' problems [34].

The Support Vector Machine (SVM) [35,36] is a kernel method that is used to map non-linear and inseparable data from an input space into a higher dimensional feature space where the data would then be linearly separable [37], this is done with the aid of the separating hyperplane [38]. The benefit of the kernel method is that the use of kernel functions enables the user to save time and computational power as the mapping is no longer compulsory [39]. The SVM algorithm aims to maximize the margin (the region separating the support vectors on either side of the hyperplane). This would result in an optimal classification accuracy of the hyperplane. Although the SVM sometimes suffers from high complexity and long computational times, it is shown to be very resistant to the problem of over-fitting the data. The SVM has a good generalisation ability and also performs well in a high dimensional feature space.

Neural-fuzzy network (NFN) involves the merging of fuzzy logic with the neural network. Neural networks are typically useful for non-linear mapping of inputs to outputs whilst fuzzy systems are designed based on the fuzzy set theory which processes data and has the ability to perform human-like reasoning when classifying data [36]. The method has been highly successful in applications due to the low-level learning and

computational efficiency of neural networks coupled with the high-level human-like reasoning of fuzzy systems [40,41]. They use the well known backpropagation algorithm for the learning of the membership functions and fuzzy rules from the training data [36]. One limitation of using this algorithm is its inability to concurrently minimise the training and test error of the system. This therefore limits the attainable classification performance at the testing phase. The method involves the combination of multiple neural network classifiers with the aim that the fusion of their output would produce a higher classification accuracy compared to just utilising a single classifier [42].

Self-organising map (SOM) is an unsupervised competitive learning technique that was proposed by Kohonen [43] in 1982. In the SOM technique, neural maps transform data from a high-dimensional input space onto a lower dimensional output space [44] in a way that would preserve the architecture. Neighbouring neurons in the output space correspond to neighbouring data points in the input space [45]. The SOM technique is competent for dimensionality reduction and topology preservation [46]. This technique however utilises a fixed network architecture which mostly has to be defined by the user before the commencement of training and thus creates a dilemma concerning the size of the pre-defined output layer. If the fixed size is too small, the model is unable to express the input data effectively. If the output size is too large, the model would take a considerably long time to converge and also produce many redundant neuron units in the output layer [47]. The growing self-organising map (GSOM) was proposed as a solution to this problem. It provides a dynamic structure to the network model instead of the previously fixed structure used in the SOM [45,48].

Compared with traditional classification techniques, the machine learning approach in general uses a black-box approach. It demonstrates an appealing advantage as the designer does not need to know much about the problem due to the fact that the characteristic and information about the problem are obtained through learning algorithm. This is a prominent factor in the motivation to employ neural networks to deal with some classification problems. However, the traditional feed-forward neural networks demonstrate a drawback as the weights are fixed after training which limits the learning and generalisation capabilities. In handling a large amount of data in a large spacial domain, a sufficiently large size of neural network is required. In this paper, we consider a variable weight neural network (VWNN) to improve the learning and generalisation capability of the traditional neural networks. A VWNN consists of two traditional neural networks, namely tuning and tuned neural networks. The tuned neural network is the one which actually classifies the input data. The tuning neural network is to provide the weights to the tuned neural network according to the characteristic of the input data. Theoretically, the VWNN can be viewed as an infinite number of traditional neural networks with fixed weights. In the operation, the VWNN will use the best traditional neural network to handle the input data for classification.

Previous works relating to this study of the VWNN include [49] where the weight coefficients are obtained by means of an adaptive genetic algorithm and [50] where the number of hidden neurons in the network is varied based on the complexity of the input pattern to the system. A VWNN in [51] is applied to a handwriting pattern recognition problem, however, where the weights are predicted in only one layer in [51] whereas the weight is predicted over multiple layers in this paper. Also this research is applied to a material surface and epilepsy seizure phase classification, these are new applications of the VWNN technique when compared to the previous paper.

In this paper, we consider two real-life applications which are surface material recognition and epilepsy seizure phases recognition. In the application of surface material recognition, we develop classifiers to recognise the surface material of an unknown object

from 18 classes. The surface information such as frictional coefficients, texture, compliance and roughness are collected using a contact sensing fingertip of a robotic hand mounted on a robot arm. In operation, the contact sensing fingertip, which is capable of identifying the normal and frictional force of an object, will slide along the object with short strokes whilst changing (increasing/decreasing) the velocity as is appropriate. Feature vectors extracted from the raw data to reduce its dimensions are used as the input of classifiers. In the application of epilepsy seizure phases recognition, classification of epilepsy signals is considered using real clinical data. We will develop classifiers which are able to classify the 3 seizure phases namely seizure-free, pre-seizure and seizure phases. Both applications demonstrate a huge potential to be applied in domestic and industrial tasks. In both of the applications, we will employ the VWNN to implement the classifiers. The classification performance will be compared with some traditional classifiers such as feedforward-neural-network, naive Bayes and kNN classifiers. Their robustness will be tested using noise-contaminated data.

The objective of the research conducted in this paper is to improve the generalisation capabilities and also robustness of the neural network by proposing the novel neural network-based classifier. One of the main limitations of the traditional neural network when dealing with large datasets is the fact that a large number of hidden neurons would be required to tackle this problem, in this paper we use a different network to predict the weights for the classifier and as a result we can have theoretically an infinite number of networks, this is the main difference between VWNN and other NNs. Particular interest is being paid to its application to the material surface and epilepsy seizure phase classification. The benefits of improving the generalisation capabilities of this classifier is very significant due to the implications that it would have in various industries such as the medical and industrial fields.

The paper is organised as follows. Section 2 gives the background of the traditional neural networks, which provide the foundation for developing the VWNNs. Section 3 introduces the VWNN and explains how it works. Section 4 presents two applications on material and epilepsy signal classification. Section 5 gives the conclusion.

2. Traditional neural networks

A 3-layer feed-forward fully connected neural network with n_{in} inputs and n_{out} outputs is shown in Fig. 1 where $w_{ji}^{(1)}$ denotes the weight between the j th hidden node and the i th input node; $w_{ji}^{(2)}$ denotes the weight between the j th output node and the i th hidden node, and $b_j^{(1)}$ and $b_j^{(2)}$ denote the weights of the biases in the j th hidden and output nodes, respectively. It has been shown that a 3-layer feed-forward fully connected neural network is a universal approximator which can approximate a smooth and continuous non-linear function in a compact domain to an arbitrary level of accuracy.

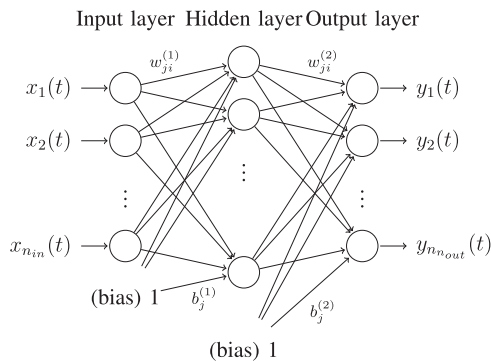


Fig. 1. A three-layer feed-forward fully connected neural network.

A multiple-layer feed-forward fully connected neural network with one input layer, n_h hidden layers and one output layer is briefly presented in this section. It takes $\mathbf{x}(t) = [x_1(t) x_2(t) \dots x_{n_{in}}(t)]$ as the t th input and produces $\mathbf{y}(t) = [y_1(t) y_2(t) \dots y_{n_{out}}(t)]$ as the outputs where n_{in} denotes the number of input nodes in the input layer and n_{out} denotes the number of output nodes in the output layer.

The output of the j th node in the input layer is given as follows:

$$f_i^{(0)}(t) = x_i(t), \quad i = 1, 2, \dots, n_{in} \tag{1}$$

and the output of the j th node in the n_i th hidden layer is given as follows:

$$f_i^{(n_i)}(t) = tf_{n_i} \left(\sum_{j=1}^{n_{n_i-1}} w_{ij}^{(n_i)} f_j^{(n_i-1)}(t) - b_j^{(n_i)} \right), \quad i = 1, 2, \dots, n_h^{(n_i)}, \tag{2}$$

where $tf_{n_i}(\cdot)$ denotes the transfer function; $n_h^{(n_i)}$ denotes the number of hidden nodes, $b_j^{(n_i)}$ denotes the bias in the n_i th hidden layer; and $w_{ij}^{(n_i)}$ denotes the weight between the j th node in the $n_h^{(n_i-1)}$ -th hidden layer and the i th node in the $n_h^{(n_i)}$ -th hidden layer.

The output of the neural network is given as follows:

$$y_i(t) = tf_{n_{out}} \left(\sum_{j=1}^{n_{n_{out}-1}} w_{ij}^{(n_{out})} f_j^{(n_{out}-1)}(t) - b_j^{(n_{out})} \right), \quad i = 1, 2, \dots, n_{out} \tag{3}$$

3. Variable-weight neural networks

A feed-forward fully connected neural network is a network with static weights which processes all input using the same connection weights between layers. Although it has been shown that it is a universal approximator, it requires a sufficiently large number of hidden nodes to offer an acceptable performance. Considering the case when the number of input data is large, the number of hidden nodes will have to be large in order to maintain the learning and generalisation capability. However, a large number of hidden layers is not favourable to both hardware and software implementations due to the increase of computational demand. Using a small number of hidden nodes will definitely offer advantages rather than disadvantages in terms of implementation costs. However, it will degrade the learning and generalisation capability of the neural network resulting in a poor performance.

A VWNN is a neural network with dynamic weights which is good in handling a large dataset. Assuming that the large dataset is divided into a number of small sub-datasets, a small neural network (with small number of hidden nodes) will work well. The VWNN works based on the concept that different connection weights are employed

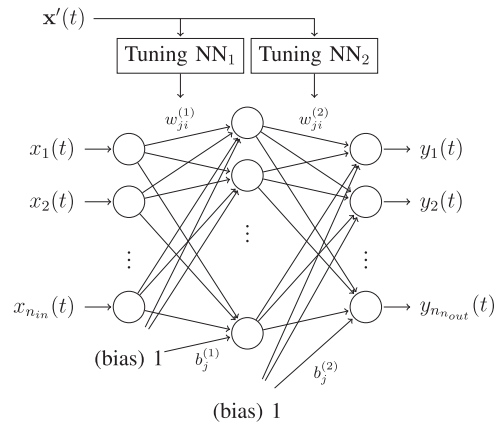


Fig. 2. A three-layer variable-weight neural network.

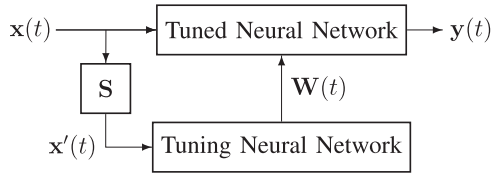


Fig. 3. A block diagram of variable-weight neural network.

by the neural network according to the network input. Consequently, the VWNN seems to consist of infinite number of neural networks and each individual input is processed by an individual neural network.

A three-layer VWNN is shown in Fig. 2. The tuning neural networks (NN₁ and NN₂) will provide connection weights $w_{ji}^{(1)}$ and $w_{ji}^{(2)}$ and bias weights $b_j^{(1)}$ and $b_j^{(2)}$ to a three-layer feed-forward fully connected neural network according to the input $\mathbf{x}'(t)$ which consists of some selected features from $\mathbf{x}(t)$. The neural network will process the input $\mathbf{x}'(t)$ according to the provided connection weights. This concept can be generalised to a VWNN with any number of hidden layers.

A block diagram of a general VWNN is shown in Fig. 3 which consists of 2 traditional neural networks, namely turning and turned neural networks. The input $\mathbf{x}(t)$ will be selected by a pre-determined constant selection matrix $\mathbf{S} \in \mathbb{R}^{n_{n_{in}}' \times n_{n_{in}}}$ such that $\mathbf{x}'_{n_{in}}(t) = \mathbf{S}\mathbf{x}_{n_{in}}(t)$ where $n_{n_{in}}' \leq n_{n_{in}}$. For example, considering

$$\mathbf{x}'_{n_{in}}(t) = \begin{bmatrix} x'_1(t) \\ x'_2(t) \end{bmatrix}, \quad \mathbf{x}_{n_{in}}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} \quad \text{and} \quad \mathbf{S} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

we have $\mathbf{x}'_{n_{in}}(t) = \mathbf{S}\mathbf{x}_{n_{in}}(t) = \begin{bmatrix} x_1(t) \\ x_3(t) \end{bmatrix}$ which selects $x_1(t)$ and $x_3(t)$ as the input of the turning neural network. The tuning neural network will produce output weight vector $\mathbf{W}(t)$ consisting of all connection weights of the tuned neural network. The tuned neural network will then use $\mathbf{W}(t)$ to process the input $\mathbf{x}(t)$. As a result, it seems like an individual input $\mathbf{x}(t)$ is processed by an individual neural network to produce an output $\mathbf{y}(t)$.

4. Applications

In this section, we employ the proposed VWNN to implement the classifier to handle two applications. The first application is the classification of materials using the data collected by a robotic finger. The second application is the classification of epilepsy using real clinical data.

4.1. Material classification

Classification problem in material surface recognition of an unknown object demonstrates a wide range of potential domestic and industrial applications, examples include robot-assisted surgery [52–55], blind grasping application [56,57], pose classification [58], prosthetic limbs [59], quality assurance [60], shape extraction and industrial inspection [61,62] and brain–machine–brain interface [63].

The VWNN is employed to implement a classifier to classify 18 materials listed in Table 1 using data collected from a robotic testing platform shown in Fig. 4, which includes a robot arm Mitsubishi RV-6SL, a 6-axis force/torque sensor ATI Nano17 (resolution=0.003 N, sampling rate=100 Hz) and a hemispherical plastic fingertip. During experiments, the fingertip which is rigidly attached to the robot arm was kept perpendicular to the material surface all the time. It was then commanded to slide on a selected object surface, keeping the normal force around 2 N. To obtain the dynamic relationship of friction and velocity, within one stroke, the sliding velocity was increased from zero to 15 mm/s with a constant acceleration rate of

Table 1

18 Materials used in the experiment.

Class label	Material
1	Un-laminated wood
2	Fine polished aluminium
3	Unpolished aluminium
4	Polished brass
5	Ceramic plate
6	Cloth liner
7	Glass
8	Artificial leather
9	Mouse pad (liner surface)
10	A4 paper
11	Laminated book cover
12	Plastic PC mouse
13	Plastic CD cover
14	Polymer composite (smooth surface)
15	Kitchen sponge
16	Stainless steel knife
17	Rubber tape
18	Un-laminated paper package

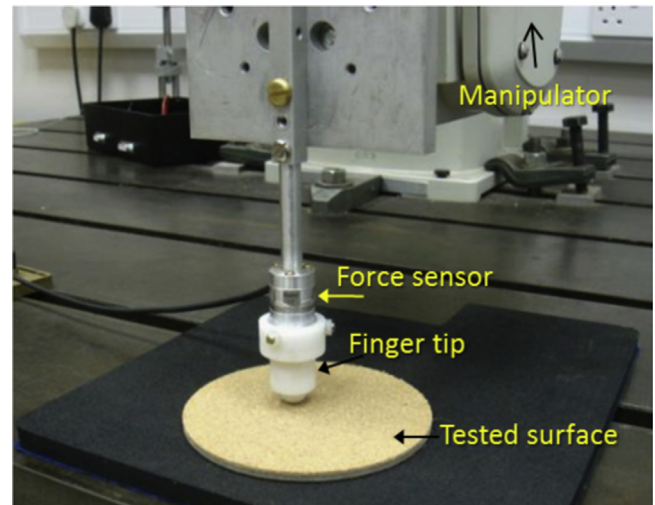


Fig. 4. The test platform.

3 mm/s². Each time the fingertip slides along a material surface, 100 numerical values (raw data of frictional force) reflecting the material characteristics are collected. The same experiment was repeated for 60 times for each of the 18 materials. In total, 60 sets of data (each set contains 100 numerical values) for each material were collected. Further detailed description of the experiment setup and data collection can be found in [64,65].

4.1.1. Feature extraction

In these experiments, the raw data of 100 points (denoted as p_1 to p_{100}) will first be reduced to feature vectors of 3, 4 and 5. The raw data of 100 numerical values of each pattern is first divided into 4 portions such that $\mathbf{p}_1 = [p_1 \ p_2 \ \dots \ p_{25}]$, $\mathbf{p}_2 = [p_{26} \ p_{27} \ \dots \ p_{50}]$, $\mathbf{p}_3 = [p_{51} \ p_{52} \ \dots \ p_{75}]$ and $\mathbf{p}_4 = [p_{76} \ p_{77} \ \dots \ p_{100}]$. Define

$$f_1(\mathbf{z}) = \frac{1}{S} \sum_{i=1}^S z_i, \quad (4)$$

$$f_2(\mathbf{p}) = \sum_{i=1}^4 |f_1(\mathbf{p}_{i+1}) - f_1(\mathbf{p}_i)|, \quad (5)$$

$$f_3(\mathbf{z}) = \frac{1}{S-1} \sum_{i=1}^S (z_i - f_1(\mathbf{z}))^2, \quad (6)$$

where $\mathbf{z} = [z_1 \ z_2 \ \dots \ z_S]$ and S is an integer representing the number of elements in \mathbf{z} .

Based on the functions in (4)–(6), we define the feature vectors of 3–5 points as follows:

Feature vector with 3 points:

$$\mathbf{x} = \left[\sum_{i=1}^4 f_1(\mathbf{p}_i) \ 50f_2(\mathbf{p}_i) \ 50 \sum_{i=1}^4 f_3(\mathbf{p}_i) \right]. \quad (7)$$

Feature vector with 4 points:

$$\mathbf{x} = \left[\sum_{i=1}^4 f_1(\mathbf{p}_i) \ 50f_2(\mathbf{p}_i) \ 50 \sum_{i=1}^4 f_3(\mathbf{p}_i) \ 20 \sum_{i=1}^4 \sqrt{f_3(\mathbf{p}_i)} \right]. \quad (8)$$

Feature vector with 5 points:

$$\mathbf{x} = \left[\sum_{i=1}^4 f_1(\mathbf{p}_i) \ 50|f_1(\mathbf{p}_2) - f_1(\mathbf{p}_1)| \ 50|f_1(\mathbf{p}_3) - f_1(\mathbf{p}_2)| \ 50|f_1(\mathbf{p}_4) - f_1(\mathbf{p}_3)| \ 50 \sum_{i=1}^4 f_3(\mathbf{p}_i) \right]. \quad (9)$$

It can be seen from (4) to (6) that $f_1(\mathbf{z})$ is the mean of \mathbf{z} , $f_2(\mathbf{z})$ is the sum of the difference of the mean of the consecutive portions of raw data, $f_3(\mathbf{z})$ is the variance of \mathbf{z} .

4.1.2. VWNN-based classifier

The proposed VWNN is employed to implement a classifier to recognise the 18 materials using the feature vectors of 3, 4, and 5 points. Fig. 5 shows the structure of classifier consisting of a VWNN with n_{in} inputs (the number of feature points) and one output.

In this experiment, the dataset is divided into training dataset consisting of 40 sets of data for each material and test dataset consisting of 20 sets of data for each material. Supervised learning was employed to train the VWNN classifier according to the class labels shown in Table 1.

We have tried different combinations of transfer functions, number of hidden nodes and hidden layers in this study. In the following, only this combination can achieve the best recognition accuracy. The overall network is 6 hidden layers structure, the number of hidden nodes is 3, 4, 6, 4, 3, 4, respectively. The first two layers's transfer function use 'tansig' function and other layers are 'linear' function, and the 3rd and 4th layers are VWNN layer, which tuned by two ordinary networks using 'tansig' transfer function. The linear transfer function is used in the output layer of all classifiers, and number of output nodes is 1. The VWNN classifier was implemented on Matlab and Levenberg–Marquardt back-propagation was used to train the classifiers by minimising the mean square error.

For comparison purposes, traditional NN, kNN and naive Bayes classifiers were employed as classifiers for this application. To test the robustness of the classifiers, the test dataset extracted from

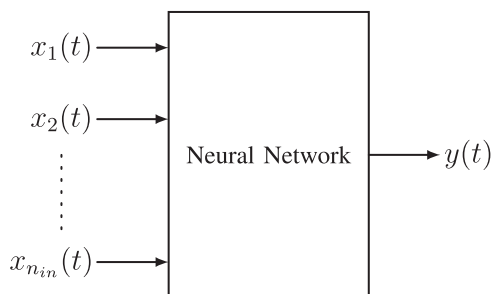


Fig. 5. NN-based classifier for materials.

Table 2

Summary of recognition performance under noise-free dataset. Classifier 1: VWNN classifier, classifier 2: traditional neural network classifier, classifier 3: K-nearest neighbour classifier, 4: Naive Bayes classifier.

# feature points	Classifier	Recognition accuracy (%)			
		Training		Testing	
		Worst	Average	Worst	Average
3	1	100	100	90	98.6111
3	2	100	100	80	96.9444
3	3	100	100	80	95.8333
3	4	100	100	90	99.4444
4	1	100	100	90	98.6661
4	2	100	100	85	96.3889
4	3	100	100	70	93.6111
4	4	100	100	100	100
5	1	100	100	95	99.1667
5	2	100	100	85	96.1111
5	3	100	100	70	89.7222
5	4	100	100	100	100

raw data contaminated by Gaussian white noise with variance of 0.005 and 0.01 was considered. Each classifier was tested 10 times using the noisy test dataset.

4.1.3. Classification results

The training and testing recognition results of the VWNN, traditional NN, kNN and naive Bayes classifiers are summarised in Table 2. In this table, the worst and average recognition accuracies for both training and test datasets are shown. The worst recognition accuracy is the worst individual accuracy in the 18 materials while the average recognition accuracy is the average individual accuracy of 18 materials.

Referring to this table, it can be seen that all classifiers perform well, achieving 100% of recognition accuracy for training dataset. For test dataset, the naive Bayes classifiers with 3, 4 and 5 feature points offer the best average recognition performance of 99.4444%, 100% and 100%, respectively. The proposed VWNN classifiers with 3, 4 and 5 come second offering the average recognition performance of 98.6111%, 98.6661% and 99.1667%. All the remaining classifiers offer an average recognition performance of less than 97%. Among all remaining classifiers, the kNN classifier with 5 feature points offers the worst average performance of 89.7222% and its worst individual recognition accuracy is 70%.

The performance of the VWNN for this particular application shows its viability and potential as a classifier, apart from the naive Bayes classifier, we observe that the VWNN outperforms all other classifiers significantly as its recognition accuracy is at least 2% greater than all the other classification techniques that are used for comparison. The distinct advantage that the VWNN has over all other techniques is its robustness and flexibility as the VWNN can be applied to any range of input data without having an adverse impact in the performance, this is due to the fact that a separate NN selects the weights used for the tuned NN. This makes the VWNN a very competent classifier for the surface recognition application as it would not just be able to recognise and classify the present materials but will have the ability to recognise additional material types that could be added to the system without having a significant reduction in its recognition accuracy.

The testing recognition results for the test dataset subject to Gaussian white noise with variance of 0.005 and 0.01 are summarised in Tables 3 and 4, respectively, which provide the statistical information including the average recognition accuracy (the average of the average recognition accuracy of the 18 materials of the 10 times of tests), worst recognition accuracy (the worst average

Table 3

Summary of recognition performance for the dataset subject to noise level of 0.005. Classifier 1: VWNN classifier, classifier 2: traditional neural network classifier, classifier 3: K-nearest neighbour classifier, 4: Naive Bayes classifier.

# feature points	Classifier	Recognition accuracy (%)				
		Worst	Average	Best	Std	Worst individual (Average)
3	1	91.3889	94.2222	97.2222	2.0916	50.0000
3	2	86.3889	92.3611	96.3889	3.1882	55.0000
3	3	88.6111	93.8889	98.3333	3.3120	84.0000
3	4	92.5000	93.5278	94.1667	0.5826	0.5000
4	1	88.6111	91.2222	93.0556	1.4722	0.0000
4	2	77.5000	82.3333	86.9444	3.1543	1.5000
4	3	81.6667	86.6667	91.1111	3.3264	27.0000
4	4	95.5556	97.5000	98.6111	0.9631	72.0000
5	1	93.6111	96.3889	98.3333	1.5817	68.0000
5	2	86.3889	92.9444	97.2222	3.5728	60.0000
5	3	83.3333	88.6389	92.7778	3.0588	39.5000
5	4	93.3333	93.9167	94.4444	0.4086	0.5000

Table 4

Summary of recognition performance for the dataset subject to noise level of 0.01. Classifier 1: VWNN classifier, classifier 2: traditional neural network classifier, classifier 3: K-nearest neighbour classifier, 4: Naive Bayes classifier.

# feature points	Classifier	Recognition accuracy (%)				
		Worst	Average	Best	Std	Worst individual (Average)
3	1	58.6111	63.9167	68.0556	3.1284	0.0000
3	2	61.1111	68.6944	74.1667	4.1574	0.0000
3	3	77.7778	86.2500	93.6111	5.2281	50.5000
3	4	74.7222	77.9722	81.1111	2.2443	0.0000
4	1	61.9444	64.7500	67.2222	1.8301	0.0000
4	2	58.6111	66.6389	72.2222	4.6941	0.0000
4	3	55.8333	62.2778	66.9444	3.7609	0.0000
4	4	67.7778	71.6111	74.7222	2.1802	0.0000
5	1	84.1667	89.5000	94.7222	3.6503	34.5000
5	2	63.8889	72.9444	80.0000	5.4622	0.0000
5	3	72.5000	82.6111	89.7222	5.8274	38.5000
5	4	79.7222	81.0185	82.5000	1.5608	0.0000

recognition accuracy of the 18 materials among the 10 times of tests) and best recognition accuracy (the best average recognition accuracy of the 18 materials among the 10 times of tests), standard deviation of the 10 times of tests and the average of the worst individual recognition accuracy among 18 materials.

It can be seen from the tables that the recognition performance of all classifiers degrades when the noise level increases. Considering the noise level of 0.005, the VWNN with 5 feature points, and naive Bayes with 4 and 5 feature points offer the best average recognition accuracy over 98%. When the noise level increases to 0.01, naive Bayes classifiers degrade their performance significantly compared with the VWNN classifier with 5 feature points. The VWNN classifier with 5 feature points is able to maintain its recognition performance offering the best of the best average recognition accuracy of 94.7222% while the kNN classifier with 3 feature points comes second offering 93.6111%.

Based on the above discussion, the VWNN classifier with 5 feature points offers the best recognition performance with noise-free raw data. Under the noisy raw data, it is able to outperform the other classifiers in terms of worst, average and best recognition accuracy suggesting that it has a comparatively superior capability to tolerating noise in the input. In a real world application there would always be an element of noise in the input dataset and this is the main reason why noise was applied to the input data to investigate what

the effect would be on recognition accuracy. It is in this instance that we are able to see the effect of the robustness and performance of the VWNN neural network as we see that it outperforms all other classification methods when the system has been subjected to a noise level of 0.01. This shows the suitability and superiority of the VWNN as a classifier when compared to the naive Bayes, kNN and Traditional NN as it has a higher performance and shows a lot of promise when applied to raw data as it has a higher level of tolerance to noise without adversely affecting its performance.

4.2. Epilepsy classification

Epilepsy is a common neurologic disorder that is a chronic disease of the brain causing sudden paradoxical discharge of cortical neurons. Abnormal, excessive or synchronous neuronal activity in the brain [66,67] will cause spontaneous and unforeseeable occurrence of seizures [68] with transient signs and/or symptoms. The duration of the absence seizure typically lasts from a few seconds up to around a minute, causing momentary lapses of consciousness to the sufferers [69]. However, in extreme cases it may recur frequently over 100 times a day [70].

These sudden and abrupt seizures will cause significant impact on the living quality of sufferers [71,72] and their carers. More importantly, it may cause life-threatening accidents when the sufferer is unconscious. Therefore, understanding of pre-seizure (the transition of brain activity toward an absence seizure) is a very demanding task [73,74]. Early detection of pre-seizure is vital to the sufferers and their carers, providing them ample time to take precautions.

Absence seizures are a form of generalised seizures accompanied by spike-and-wave complexes in the electroencephalograph (EEG) [75,76]. In general, epileptic EEG classification processes can be broken down into 2 sub-processes, namely feature extraction and classification. A wide range of feature extraction methods ranging from traditional linear methods (e.g., Fourier transforms and spectral analysis [77]) to nonlinear methods (e.g., Lyapunov exponents [78], correlation dimension [79] and similarity [80,81]) can be found in the literature. The extracted features will be used for the design of classifiers and classification in the real operation. A wide range of methodologies for epileptic EEG classification can be found in the literature such as artificial neural networks and neuro-fuzzy systems [82–84].

In this application, various classifiers (VWNN, traditional NN, kNN and naive Bayes classifiers) are employed to recognise the 3 seizure phases namely seizure-free, pre-seizure and seizure phases. To perform training and testing of classifiers, EEG recordings were collected in Peking University People's Hospital from 10 patients (6 males and 4 females) with absence epilepsy, aged from 8 to 21 years old. The study protocol has been approved by the ethics committee of Peking University People's Hospital and the patients have signed informed consent that their clinical data might be used and published for research purposes. The EEG data (sampled at a frequency of 256 Hz using a 16-bit analogue-to-digital converter and filtered within a frequency band from 0.5 to 35 Hz) were recorded by the Neurofile NT digital video EEG system from a standard international 10–20 electrode placement (Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5, T6, Fz, Cz and Pz).

Three sets of EEG signals from different seizure phases namely seizure-free, pre-seizure and seizure phases were collected where 112 2-second 19-channel EEG epochs from 10 patients were extracted for each dataset. The timing of onset and offset in spike-wave discharges (SWDs) was identified by an epilepsy neurologist, and these SWDs were defined as large-amplitude rhythmic 3–4 Hz discharges with typical spike-wave morphology lasting > 1 s. The seizure-free, pre-seizure and seizure data are determined based on the criteria of (1) the interval between the seizure-free data and the beginning point of

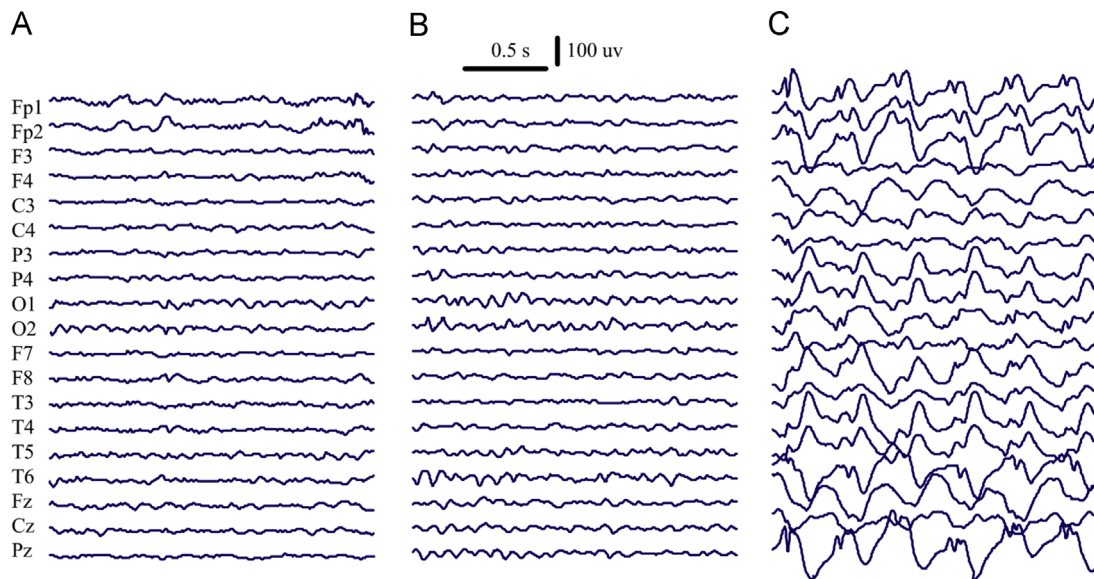


Fig. 6. Examples of raw EEG recordings for (A) seizure-free, (B) pre-seizure and (C) seizure phases.

seizures is greater than 15 s, (2) the interval is between 0 and 2 s prior to seizure onset, and (3) the interval is the first 2 s of the absence seizure, respectively. Fig. 6 shows the examples of 19-channel EEG recordings in seizure-free, pre-seizure and seizure phases. The generalised SWDs with a repetition rate of 3 Hz are typically associated with clinical absence seizures. Further details regarding the data collection can be found in [85–87].

4.2.1. Feature extraction

Feature extraction is an essential step to find and then extract the hidden characteristics and information of EEG signals, and one other important purpose of feature extraction is to reduce the redundancy of the original signals. In order to improve the classification performance, it is necessary to elect appropriate feature extraction methods and then combine them together to achieve better recognition accuracy.

In the EEG case, we have 19 columns of signals as shown in Fig. 6, which are collected from 19 EEG sensors, and each column of the EEG signals has 100 sample points. We need first to extract the useful information from the 19×100 points and then reduce its dimensions to form a feature vector, which will be taken as the input of the classifiers.

Some published results provide the evidence supporting the view of fronto-central network in absence epilepsy which suggested that not all 19 channels are of the same importance. The characteristics of early cortical activities were analysed and the spatio-temporal dynamics of interactions within and between local cortical neural networks were explored in [88]. It reveals a reproducible sequence demonstrating increased long-range desynchronisation, local synchronisation and long-range synchronisation found a multifocal fronto-central network in absence epilepsy. In the study in [89], the authors compared the functional networks in EEG background between absence epilepsy and healthy control individuals to identify which set of electrodes provide the maximum differentiation. Both studies reported a similar result that electrodes F3, Fz, F4, C3 and Cz are the most representative ones for the differentiation between control and absence patients while the rest carry useful information and patterns that can help to discriminate among different absence seizure phases.

In this study, we selected the most useful channels by considering different channel combinations. It was found that the 1st, 2nd, 3rd, 4th, 5th, 6th, 11th, 12th, 13th, 14th channels out of the 19 channels contain the most significant information for classification,

which compile with the results in [88,89] that channels F3, Fz, F4, C3 and Cz contain the most important information. From each of the chosen channels, a feature vector consisting of time-domain and frequency-domain components is formed. Features in time domain, which do not have any transformation, are straightforward and easy to comprehend [90]. The standard deviation, second order norm, third order norm, fourth order norm, the absolutely sum, the maximum value, the minimum value of each channel are computed to form part of the feature vector. The frequency or spectral domain analysis is mostly used to study and analyse the EEG signals in the frequency domain. The mean frequency, maximum frequency, minimum frequency, standard deviation of frequency, the windowing filtered mean frequency and windowing filtered maximum frequency of each chosen channel will form the rest of the feature vector. It is noted that the size of the overall feature vector formed by combining all channel feature vectors is huge. Principle component analysis (PCA) is employed to lower the dimensionality of the overall feature vector. Since each channel has its own characteristics, we choose different principle components according to different channels.

At last, we gain 45 points to form the feature vector, which will be further used in the recognition or classification stage. In the following, all NNs and classifiers take these 45 points as input to perform classification.

4.2.2. VVNN classifier

A tree-structured classifier implemented by the proposed VVNN is employed to classify 3 classes of Epilepsy signals (seizure-free, pre-seizure and seizure phases) using the feature vector achieved in the previous section. Fig. 7 shows the tree-structured VVNN classifier consisting of 2 45-input-single-output VVNNs and a class determiner. The 1st VVNN is used to determine if the testing sample belongs to class 3 (seizure phase), if not, we use the 2nd VVNN to determine if the testing sample belongs to class 1 (seizure-free phase) or class 2 (pre-seizure phase). The classifier will determine the final class according to the rules as shown in Table 5.

In this application, we have tried different combinations of transfer functions, number of hidden nodes and hidden layers in this study. The following combination can achieve the best recognition accuracy. For the 1st VVNN network as shown in Fig. 7, the tuned NN has 45 inputs, 4 hidden layers with 25, 4, 8 and 5 hidden nodes and one output node. The transfer functions corresponding

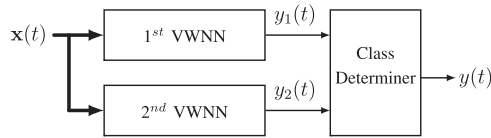


Fig. 7. Tree-structure VWNN classifier for epilepsy.

Table 5

Output classes of class determiner.

$y_1(t)$	$y_2(t)$	$y(t)$
3	1	3
3	2	3
Not 3	1	1
Not 3	2	2

to the 4 hidden layers are hyperbolic tangent sigmoid, hyperbolic tangent sigmoid, logarithm sigmoid function and logarithm sigmoid function, respectively. Linear function and logarithm sigmoid function are used in the input and output layers, respectively. The tuning NN has 45 inputs, 2 hidden layers with 25 and 4 hidden nodes and 32 output nodes. The first 3 layers, i.e., the input and the 2 hidden layers, are common to the tuned NN. The output layer uses hyperbolic tangent sigmoid function as the transfer function. The outputs of the tuning NN provide the variable weights to the connections between the 3rd and 4th hidden layers of the tuned NN.

The tuned NN of the 2nd VWNN as shown in Fig. 7 has 45 inputs, 4 hidden layers of 35, 5, 8 and 5 hidden nodes and one output node. The transfer functions corresponding to the 4 hidden layers are hyperbolic tangent sigmoid, hyperbolic tangent sigmoid, logarithm sigmoid function and logarithm sigmoid function, respectively. Linear function and logarithm sigmoid function are used in the input and output layers, respectively. The tuning NN has 45 inputs, 2 hidden layers with 35 and 5 hidden nodes and 40 output nodes. The first 3 layers, i.e., the input and the 2 hidden layers, are common to the tuned NN. The output layer uses hyperbolic tangent sigmoid function as transfer function. The outputs of the tuning NN provide the variable weights to the connections between the 3rd and 4th hidden layers of the tuned NN.

For comparison purposes, traditional NN, kNN and naive Bayes classifiers were employed as classifiers for this application. To test the robustness of the classifiers, the test dataset extracted from the raw data contaminated by Gaussian white noise with variance of 0.05–1 was considered. The NN classifier has the same structure as the VWNN classifier as shown in Fig. 7 but the VWNNs are replaced by the traditional NNs. The transfer function used in the 3rd layer of the traditional NN which has the same number of hidden nodes as that of the VWNN. Each classifier was tested 10 times using the test dataset subject to noise of different levels.

4.2.3. Classification results

The classification performance of all classifiers with the original data and data contaminated by noise level from 0.05 to 1 is shown below. The classification performance corresponding to different noise levels is summarised in Tables 6–11.

In Table 6, the worst and average recognition accuracies for both training and test datasets are shown. The worst recognition accuracy is the worst individual accuracy in the 3 classes while the average recognition accuracy is the average recognition accuracy of the individual recognition accuracy of all 3 classes. It can be seen that the VWNN offers the best performance over the other 3 traditional classifiers methods evident by average training and testing recognition accuracies of 100% and 91.111%, respectively.

Table 6

Summary of recognition performance for EEG signals with original dataset. Classifier 1: VWNN classifier, classifier 2: traditional neural network classifier, classifier 3: K-nearest neighbour classifier, 4: Naive Bayes classifier.

Classifier	Recognition accuracy (%)			
	Training		Testing	
	Worst	Average	Worst	Average
1	100	100	80.0000	91.1111
2	100	100	73.3333	86.6667
3	100	100	23.3333	56.6667
4	41.4286	77.1429	33.3333	77.7778

Table 7

Summary of testing samples recognition performance for EEG signal under dataset subject to noise level of 0.05. Classifier 1: VWNN classifier, classifier 2: traditional neural network classifier, classifier 3: K-nearest neighbour classifier, 4: Naive Bayes classifier.

Classifier	Recognition accuracy (%)				
	Worst	Average	Best	Std	Worst individual (Average)
1	85.5556	89.2222	94.4444	2.9801	75.0000
2	77.7778	85.0000	90.0000	4.0140	75.3333
3	51.1111	56.3333	61.1111	3.2735	20.3333
4	77.7778	78.3333	80.0000	0.7857	35.0000

Table 8

Summary of testing samples recognition performance for EEG signal under dataset subject to noise level of 0.1. Classifier 1: VWNN classifier, classifier 2: traditional neural network classifier, classifier 3: K-nearest neighbour classifier, 4: Naive Bayes classifier.

Classifier	Recognition accuracy (%)				
	Worst	Average	Best	Std	Worst individual (Average)
1	80.0000	86.4444	92.2222	4.1869	67.0000
2	78.8889	85.7778	90.0000	3.6501	70.6667
3	52.2222	56.5556	61.1111	2.9232	20.3333
4	76.6667	78.8889	82.2222	1.8251	37.6667

Table 9

Summary of testing samples recognition performance for EEG signal under dataset subject to noise level of 0.2. Classifier 1: VWNN classifier, classifier 2: traditional neural network classifier, classifier 3: K-nearest neighbour classifier, 4: Naive Bayes classifier.

Classifier	Recognition accuracy (%)				
	Worst	Average	Best	Std	Worst individual (Average)
1	78.8889	84.4444	90.0000	3.4978	58.6667
2	75.5556	84.1111	88.8889	3.8639	64.6667
3	53.3333	57.2222	62.2222	2.9614	20.6667
4	76.6667	79.0000	82.2222	1.8898	38.6667

The fact that the VWNN outperforms the other existing classification methods (traditional NN, kNN and naive Bayes) in all the simulations and when subjected to various levels of noise in the input data, it is fair to conclude that the VWNN is a very powerful tool with a strong suitability for this particular application and the flexibility of the VWNN to accept a wide range of input data will be very effective in dealing with the epilepsy seizure phase classification problem as this technique would enable to classifier to accept a wide range of patient data and have a good

Table 10

Summary of testing samples recognition performance for EEG signal under dataset subject to noise level of 0.5. Classifier 1: VWNN classifier, classifier 2: traditional neural network classifier, classifier 3: K-nearest neighbour classifier, 4: Naive Bayes classifier.

Classifier	Recognition accuracy (%)				
	Worst	Average	Best	Std	Worst individual (Average)
1	81.1111	83.5556	85.5556	1.5948	54.6667
2	76.6667	81.2222	86.6667	3.6429	56.0000
3	52.2222	59.0000	64.4444	3.8665	24.6667
4	75.5556	78.2222	80.0000	1.5585	36.3333

Table 11

Summary of testing samples recognition performance for EEG signal under dataset subject to noise level of 1. Classifier 1: VWNN classifier, classifier 2: traditional neural network classifier, classifier 3: K-nearest neighbour classifier, 4: Naive Bayes classifier.

Classifier	Recognition accuracy (%)				
	Worst	Average	Best	Std	Worst individual (Average)
1	77.7778	83.4444	85.5556	2.4525	54.6667
2	72.2222	78.3333	82.2222	3.9338	49.0000
3	50.0000	59.6667	66.6667	5.8187	23.3333
4	76.6667	77.7778	80.0000	1.3242	36.3333

generalisation ability that is not affected by this increase in the range of input data. The fact that the VWNN performs well under higher levels of noise further enhances its suitability as a classifier due to the fact that real-life data will not be smooth and will always be subject to noise in the system due to different factors such as levels of error in the measurements techniques.

Tables 7–11 show the testing data classification performance of all the 4 classifiers with noisy data under noise levels of 0.05, 0.1, 0.2, 0.5, and 1. From these tables, it can be seen that the recognition accuracy is in general decreasing when the noise level is increasing. Among the 4 classifiers, the VWNN classifier offers the best recognition performance with the average recognition accuracy in the range of 83.4444% of 89.2222% subject to different noise levels while kNN classifier performs the worst offering the average recognition accuracies in the range of 50% to 53.3333%.

Form the above discussion, it can be concluded that the VWNN classifier outperforms the traditional classifiers offering the best recognition performance and robustness property.

5. Conclusion

In this paper, we presented a novel neural network, the variable weights neural network, which demonstrates a great potential to cope with complicated recognition and classification problems. Different from the traditional NN, the weights of the VWNN change adaptively according to the characteristic of the input data thereby enhancing its learning and generalisation capability. We have implemented classifiers using VWNNs for 2 real-life applications, i.e., material recognition using robotic finger and epilepsy classification using clinical data, to verify the effectiveness of VWNN. From the results of these two applications, it has been shown that the VWNN classifier has demonstrated the best recognition performance over the traditional neural networks, kNN method and Naive Bayes method when original input data are considered. Moreover, the VWNN classifier has demonstrated an outstanding robustness property towards noisy input data. In the future, we will keep improving the performance of VWNN and trying to find the best way to determine the structure of VWNN, for example,

the number of hidden layers, the transfer function of each layer and the nodes of each layer.

Acknowledgements

This work was partially supported by King's College London and China Scholarship Council.

References

- [1] J. Zurada, Could decision trees improve the classification accuracy and interpretability of loan granting decisions?, in: Proceedings of the 43rd Hawaii International Conference on System Sciences (HICSS), January 2010, pp. 1–9.
- [2] Y. Özbay, R. Ceylan, B. Karlik, A fuzzy clustering neural network architecture for classification of ECG arrhythmias, *Comput. Biol. Med.* 36 (4) (2006) 376–388.
- [3] M. Engin, ECG beat classification using neuro-fuzzy network, *Pattern Recognit. Lett.* 25 (15) (2004) 1715–1722.
- [4] G. Guo, S.Z. Li, K.L. Chan, Support vector machines for face recognition, *Image Vis. Comput.* 19 (9) (2001) 631–638.
- [5] Y.-H. Liu, Y.-T. Chen, Face recognition using total margin-based adaptive fuzzy support vector machines, *IEEE Trans. Neural Netw.* 18 (1) (2007) 178–192.
- [6] Z. Chi, J. Wu, H. Yan, Handwritten numeral recognition using self-organizing maps and fuzzy rules, *Pattern Recognit.* 28 (1) (1995) 59–66.
- [7] H.K. Lam, F.H.F. Leung, Digit and command interpretation for electronic book using neural network and genetic algorithm, *IEEE Trans. Syst. Man Cybern. Part B: Cybern.* 34 (December (6)) (2004) 2273–2283.
- [8] K.F. Leung, F.H.F. Leung, H.K. Lam, S.H. Ling, On interpretation of graffiti digits and characters for eBooks: neural-fuzzy network and genetic algorithm approach, *IEEE Trans. Ind. Electron.* 51 (April (2)) (2004) 464–471.
- [9] H.K. Lam, J. Prada, Interpretation of handwritten single-stroke graffiti using support vector machines, *Int. J. Comput. Intell. Appl.* 8 (December (04)) (2009) 369–393.
- [10] J. Pradeep, E. Srinivasan, S. Himavathi, Neural network based handwritten character recognition system without feature extraction, in: Proceedings of the International Conference on Computer, Communication and Electrical Technology (ICCCET), January 2011, pp. 40–44.
- [11] Z. Dokur, T. Ölmez, Heart sound classification using wavelet transform and incremental self-organizing map, *Digital Signal Process.* 18 (6) (2008) 951–959.
- [12] C.-F. Juang, C.-M. Chang, Human body posture classification by a neural fuzzy network and home care system application, *IEEE Trans. Syst. Man Cybern. Part A: Syst. Hum.* 37 (6) (2007) 984–994.
- [13] W.M. Campbell, D.E. Sturim, D.A. Reynolds, Support vector machines using GMM supervectors for speaker verification, *IEEE Signal Process. Lett.* 13 (5) (2006) 308–311.
- [14] J.-T. Chien, Linear regression based Bayesian predictive classification for speech recognition, *IEEE Trans. Speech Audio Process.* 11 (January (1)) (2003) 70–79.
- [15] K.F. Leung, F.H.F. Leung, H.K. Lam, P.K.S. Tam, Neural fuzzy network and genetic algorithm approach for cantonese speech command recognition, in: Proceedings of the 12th IEEE International Conference on Fuzzy Systems, vol. 1. IEEE, 2003, pp. 208–213.
- [16] D. Merkl, Text classification with self-organizing maps: some lessons learned, *Neurocomputing* 21 (1) (1998) 61–77.
- [17] S. Tong, D. Koller, Support vector machine active learning with applications to text classification, *J. Mach. Learn. Res.* 2 (2002) 45–66.
- [18] S. Balakrishnama, A. Ganapathiraju, Linear discriminant analysis—a brief tutorial, *Inst. Signal Inf. Process.* (1998).
- [19] C. Vens, J. Struyf, L. Schietgat, S. Džeroski, H. Blockeel, Decision trees for hierarchical multi-label classification, *Mach. Learn.* 73 (2) (2008) 185–214.
- [20] A. Tellaeché, X.P. Burgos-Artizzu, G. Pajares, A. Ribeiro, A vision-based method for weeds identification through the Bayesian decision theory, *Pattern Recognit.* 41 (February (2)) (2008) 521–530.
- [21] M.-L. Zhang, Z.-H. Zhou, A k-nearest neighbor based algorithm for multi-label classification, in: Proceedings of the IEEE International Conference on Granular Computing, vol. 2, 2005, pp. 718–721.
- [22] H. Zhang, A.C. Berg, M. Maire, J. Malik, SVM-KNN: discriminative nearest neighbor classification for visual category recognition, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, 2006, pp. 2126–2136.
- [23] U. Kumar, S.K. Raja, C. Mukhopadhyay, T. Ramachandra, Hybrid Bayesian classifier for improved classification accuracy, *IEEE Geosci. Remote Sens. Lett.* 8 (November (3)) (2011) 474–477.
- [24] G.P. Zhang, Neural networks for classification: a survey, *IEEE Trans. Syst. Man Cybern. Part C: Appl. Rev.* 30 (November (4)) (2000) 451–462.
- [25] S. Kotsiantis, Supervised machine learning: a review of classification techniques, *Informatica* 31 (July) (2007) 249–268.
- [26] G.L. Marcialis, F. Roli, Fusion of multiple fingerprint matchers by single-layer perceptron with class-separation loss function, *Pattern Recognit. Lett.* 26 (September (12)) (2005) 1830–1839.

- [27] F. Ham, I. Kostanic, *Principles of Neurocomputing for Science and Engineering*, McGraw-Hill, New York, 2001.
- [28] C. Ambrose, G. Sèze, F. Badran, S. Thiria, Hierarchical clustering of self-organizing maps for cloud classification, *Neurocomputing* 30 (1) (2000) 47–52.
- [29] A. Santillana Fernandez, C. Delgado-Mata, R. Velazquez, Training a single-layer perceptron for an approximate edge detection on a digital image, in: *Proceedings of the International Conference on Technologies and Applications of Artificial Intelligence (TAI)*, November 2011, pp. 189–193.
- [30] L. Wu, Z. Feng, W.X. Zheng, Exponential stability analysis for delayed neural networks with switching parameters: average dwell time approach, *IEEE Trans. Neural Netw.* 21 (9) (2010) 1396–1407.
- [31] X. Su, Y. Feng, L. Wu, G. Peng, New global stability criteria for interval delayed neural networks, in: *2010 3rd International Symposium on Systems and Control in Aeronautics and Astronautics (ISSCAA)*. IEEE, 2010, pp. 977–981.
- [32] C.K. Ahn, M.K. Song, New sets of criteria for exponential 1-2 stability of Takagi-Sugeno fuzzy systems combined with Hopfield neural networks, *Int. J. Innov. Comput. Inf. Control* 9 (7) (2013) 2979–2986.
- [33] A.E. Bryson, Y.C. Ho, *Applied Optimal Control*, U.K. Blaisdell, Waltham, 1969.
- [34] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice Hall PTR, New Jersey, 1994.
- [35] C.-W. Hsu, C.-C. Chang, C.-J. Lin, et al., A practical guide to support vector classification, 2003.
- [36] H.S. Hussain, S.A. Aljunid, S. Yahya, F.H.M. Ali, A novel hybrid fuzzy-SVM image steganographic model, in: *Proceedings of the International Symposium in Information Technology (ITSim)*, vol. 1. IEEE, 2010, pp. 1–6.
- [37] S. Yaman, J. Pelecanos, W. Zhu, Unifying PLDA and polynomial kernel SVMs, in: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2013, pp. 7698–7701.
- [38] Y. Zhang, B. Yu, Face recognition using combined non-negative principal component analysis and linear discriminant analysis, in: *Proceedings of the 2013 IEEE International Congress on Signal and Image Processing*, IEEE, 2013, pp. 758–762.
- [39] Z. Shi, J. Hu, Local linear discriminant analysis with composite kernel for face recognition, in: *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2012, pp. 1–5.
- [40] R. Davtalab, M.H. Dezfoulian, M. Mansoorzadeh, Multi-level fuzzy min-max neural network classifier, *IEEE Trans. Neural Netw. Learn. Syst.* 25 (3) (2013) 470–482.
- [41] C.-F. Wu, C.-J. Lin, C.-Y. Lee, A functional neural fuzzy network for classification applications, *Expert Syst. Appl.* 38 (5) (2011) 6202–6208.
- [42] M.A. Khan, M. Nazir, A. Jaffar, A.M. Mirza, Fuzzy clustering and fuzzy entropy based classification model, in: *6th International Conference on Emerging Technologies (ICET)*, IEEE, 2010, pp. 61–64.
- [43] M.M. Soliman, A.E. Hassanien, H.M. Onsi, Robust watermarking approach for 3d triangular mesh using self organization map, in: *Proceedings of the 8th International Conference on Computer Engineering & Systems (ICCES)*, IEEE, 2013, pp. 99–104.
- [44] C. Yu, L. An, X. Zhang, “Using SOM to mine product features from free-text customer reviews, in: *Proceedings of the Fifth International Conference on Natural Computation*, vol. 3, 2009, pp. 473–477.
- [45] H.-U. Bauer, T. Villmann, Growing a hypercubical output space in a self-organizing feature map, *IEEE Trans. Neural Netw.* 8 (2) (1997) 218–226.
- [46] H.-C. Yang, C.-H. Lee, A novel self-organizing map for text document organization, in: *Proceedings of the Third International Conference on Innovations in Bio-Inspired Computing and Applications (IBICA)*, IEEE, 2012, pp. 39–44.
- [47] S. Yan, J. Xiao-xia, Remote sensing change detection based on growing hierarchical self-organization map, in: *Proceedings of the International Conference on Environmental Science and Information Application Technology (ESIAT)*, vol. 2. IEEE, 2010, pp. 60–63.
- [48] W.-S. Tai, C.-C. Hsu, A growing mixed self-organizing map, in: *Proceedings of the Sixth International Conference on Natural Computation (ICNC)*, vol. 2. IEEE, 2010, pp. 986–990.
- [49] J. Li, W. Dai, H. Pan, Variable weighted combination forecasting model based on genetic algorithm and artificial neural network, in: *Third International Conference on Natural Computation*, 2007. ICNC 2007, vol. 1. IEEE, 2007, pp. 451–458.
- [50] G.P. Liu, V. Kadirkamanathan, S.A. Billings, Variable neural networks for adaptive control of nonlinear systems, *IEEE Trans. Syst. Man Cybern. Part C: Appl. Rev.* 29 (1) (1999) 34–43.
- [51] S. Ling, H. Lam, F. Leung, A variable-parameter neural network trained by improved genetic algorithm and its application, in: *2005 IEEE International Joint Conference on Neural Networks*, vol. 1. IEEE, 2005, pp. 1343–1348.
- [52] A.M. Okamura, *Methods for haptic feedback in teleoperated robot-assisted surgery*, *Ind. Robot: Int. J.* 31 (6) (2004) 499–508.
- [53] O.A.J. Van der Meijden, M.P. Schijven, The value of haptic feedback in conventional and robot-assisted minimal invasive surgery and virtual reality training: a current review, *Surg. Endosc.* 23 (6) (2009) 1180–1190.
- [54] C.-H. King, M.O. Culjat, M.L. Franco, C.E. Lewis, E.P. Dutton, W.S. Grundfest, J.W. Bisley, Tactile feedback induces reduced grasping force in robot-assisted surgery, *IEEE Trans. Haptics* 2 (2) (2009) 103–110.
- [55] G.I. Barbash, S.A. Glied, New technology and health care costs—the case of robot-assisted surgery, *New Engl. J. Med.* 363 (8) (2010) 701–704.
- [56] R. Ozawa, J.-H. Bae, S. Arimoto, Multi-fingered dynamic blind grasping with tactile feedback in a horizontal plane, in: *Proceedings of the 2006 IEEE International Conference on Robotics and Automation (ICRA)*, 2006, pp. 1006–1011.
- [57] H. Dang, J. Weisz, P.K. Allen, Blind grasping: stable robotic grasping using tactile feedback and hand kinematics, in: *Proceedings of the 2011 IEEE International Conference on Robotics and Automation (ICRA)*, 2011, pp. 5917–5922.
- [58] H. Liu, X. Song, T. Nanayakkara, L.D. Seneviratne, K. Althoefer, A computationally fast algorithm for local contact shape and pose classification using a tactile array sensor, in: *Proceedings of 2012 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2012, pp. 1410–1415.
- [59] R.E. Fan, M.O. Culjat, C.-H. King, M.L. Franco, R. Boryk, J.W. Bisley, E. Dutton, W.S. Grundfest, A haptic feedback system for lower-limb prostheses, *IEEE Trans. Neural Syst. Rehabil. Eng.* 16 (3) (2008) 270–277.
- [60] N. Jamali, Material classification by tactile sensing using surface textures, in: *Proceedings of the 2007 IEEE International Conference on Image Processing*, May 2010, pp. 2336–2341.
- [61] V. Thilak, Material classification using passive polametric imagery, in: *Proceedings of the 2007 IEEE International Conference on Image Processing*, September 2007, pp. 121–124.
- [62] N. Charniya, Neural network based sensor for classification of material type and its surface properties, in: *Proceedings of the 2007 IEEE International Joint Conference on Neural Networks*, August 2007, pp. 424–429.
- [63] J.E. O’Doherty, M.A. Lebedev, P.J. Ifft, K.Z. Zhuang, S. Shokur, H. Bleuler, M.A.L. Nicolelis, Active tactile exploration using a brain-machine-brain interface, *Nature* 479 (7372) (2011) 228–231.
- [64] H. Liu, X. Song, J. Bimbo, L. Seneviratne, K. Althoefer, Surface material recognition through haptic exploration using an intelligent contact sensing finger, in: *Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2012, pp. 52–57.
- [65] H.K. Lam, U. Ekong, H. Liu, B. Xiao, H. Araujo, Ling, K.Y. Chan, A study of neural-network-based classifiers for material classification, *Neurocomputing* 144 (20) (2014) 367–377.
- [66] R.S. Fisher, W. van Emde Boas, W. Blume, C. Elger, P. Genton, P. Lee, Epileptic seizures and epilepsy: definitions proposed by the international league against epilepsy (ilae) and the international bureau for epilepsy (ibe), *Epilepsia* 46 (April) (2005) 22–30.
- [67] R. Mohanraj, M.J. Brodie, Early predictors of outcome in newly diagnosed epilepsy, *Seizure* 22 (June) (2013) 333–344.
- [68] B. Litt, J. Echaz, Prediction of epileptic seizures, *Lancet Neurol.* 1 (May) (2002) 22–30.
- [69] A.E. Cavanna, F. Monaco, Brain mechanisms of altered conscious states during epileptic seizures, *Nat. Rev. Neurol.* 5 (May) (2009) 267–276.
- [70] X. Bai, M. Vestal, R. Berman, M. Negishi, M. Spann, C. Vega, Dynamic time course of typical childhood absence seizures: EEG, behavior, and functional magnetic resonance imaging, *J. Neurosci.* 30 (April) (2010) 5884–5893.
- [71] B.D. Killory, X. Bai, M. Negishi, C. Vega, M.N. Spann, M. Vestal, Impaired attention and network connectivity in childhood absence epilepsy, *Neuroimage* 56 (June) (2011) 2209–2217.
- [72] H. Blumenfeld, Impaired consciousness in epilepsy, *Lancet Neurol.* 11 (September) (2012) 814–826.
- [73] D. Gupta, P. Ossenblok, G. Luijckelaar, Spacetime network connectivity and cortical activations preceding spike wave discharges in human absence epilepsy: a meg study, *Med. Biol. Eng. Comput.* 49 (2011) 555–565.
- [74] V. Crunelli, D.W. Cope, J.R. Terry, Transition to absence seizures and the role of GABA_A receptors, *Epilepsy Res.* 97 (December) (2011) 283–289.
- [75] A. Gorji, C. Mittag, P. Shahabi, T. Seidenbecher, H.C. Pape, Seizure-related activity of intralaminar thalamic neurons in a genetic model of absence epilepsy, *Neurobiol. Dis.* 43 (June) (2011) 266–274.
- [76] H.K. Meeren, J.P. Pijn, E.L.V. Luijckelaar, A.M. Coenen, F.H.L. da Silva, Cortical focus drives widespread corticothalamic networks during spontaneous absence seizures in rats, *J. Neurosci.* 22 (February) (2002) 1480–1495.
- [77] Z. Rogowski, I. Gath, E. Bental, On the prediction of epileptic seizures, *Biol. Cybern.* 42 (1981) 9–15.
- [78] L.D. Iasemidis, J.C. Sackellares, H.P. Zaveri, W.J. Williams, Phase space topography and the Lyapunov exponent of electrocorticograms in partial seizures, *Brain Topogr.* 2 (1990) 187–201, Spring.
- [79] R.G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, C.E. Elger, Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 64 (December) (2001) 061907.
- [80] V. Navarro, J. Martinerie, M.L.V. Quyen, S. Clemenceau, C. Adam, M. Baulac, Seizure anticipation in human neocortical partial epilepsy, *Brain* 125 (March) (2002) 640–655.
- [81] M. Niknazar, S.R. Mousavi, S. Motaghi, A. Dehghani, B.V. Vahdat, M.B. Shamsollahi, A unified approach for detection of induced epileptic seizures in rats using ECoG signals, *Epilepsy Behav.* 27 (March) (2013) 355–364.
- [82] I. Guler, E.D. Ubeyleli, Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients, *J. Neurosci. Methods* 148 (October) (2005) 113–121.
- [83] N. Kannathal, M.L. Choo, U.R. Acharya, P.K. Sadasivan, Entropies for detection of epilepsy in EEG, *Comput. Methods Programs Biomed.* 80 (December) (2005) 187–194.
- [84] E.D. Ubeyleli, Automatic detection of electroencephalographic changes using adaptive neuro-fuzzy inference system employing Lyapunov exponents, *Expert Syst. Appl.* 36 (July) (2009) 9031–9038.

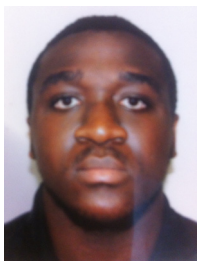
- [85] M. Thulasidas, C. Guan, J. Wu, Robust classification of EEG signal for brain-computer interface, *IEEE Trans. Neural Syst. Rehabil. Eng.* 14 (1) (2006) 24–29.
- [86] T.C. Ferree, P. Luu, G.S. Russell, D.M. Tucker, Scalp electrode impedance, infection risk, and EEG data quality, *Clin. Neurophysiol.* 112 (3) (2001) 536–544.
- [87] J. O'Keefe, M.L. Recce, Phase relationship between hippocampal place units and the EEG theta rhythm, *Hippocampus* 3 (3) (1993) 317–330.
- [88] F. Amor, S. Baillet, V. Navarro, C. Adam, J. Martinerie, M. Le Van Quyen, Cortical local and long-range synchronization interplay in human absence seizure initiation, *Neuroimage* 45 (3) (2009) 950–962.
- [89] O.A. Rosso, A. Mendes, R. Berretta, J.A. Rostas, M. Hunter, P. Moscato, Distinguishing childhood absence epilepsy patients from controls by the analysis of their background brain electrical activity (ii): a combinatorial optimization approach for electrode selection, *J. Neurosci. Methods* 181 (2) (2009) 257–267.
- [90] B. Hudgins, P. Parker, R.N. Scott, A new strategy for multifunction myoelectric control, *IEEE Trans. Biomed. Eng.* 40 (1) (1993) 82–94.



H.K. Lam received the B.Eng. (Hons.) and Ph.D. degrees from the Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong, in 1995 and 2000, respectively. During the period of 2000 and 2005, he worked with the Department of Electronic and Information Engineering at The Hong Kong Polytechnic University as a post-doctoral fellow and research fellow respectively. In 2005, he joined as a Lecturer in King's College London. Currently, he is a reader.

His current research interests include intelligent control systems and computational intelligence. He has authored/co-authored over 190 publications. He has served as a program committee member and international advisory board member for various international conferences and a reviewer for various books, international journals and international conferences. He is an associate editor for *IEEE Transactions on Fuzzy Systems*, *International Journal of Fuzzy Systems* and *Journal of Intelligent Learning Systems and Applications*; and guest editor for a number of international journals. He is in the editorial board of *IET Control Theory & Applications*, *Journal of Applied Mathematics, Modelling and Simulation in Engineering*, *Annual Review of Chaos Theory, Bifurcations and Dynamical Systems* and *The Open Cybernetics and Systemics Journal*. He is an IEEE senior member.

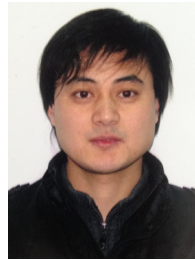
He is the coeditor for two edited volumes: *Control of Chaotic Nonlinear Circuits* (World Scientific, 2009) and *Computational Intelligence and Its Applications* (World Scientific, 2012), and the coauthor of the book *Stability Analysis of Fuzzy-Model-Based Control Systems* (Springer, 2011). His co-authored paper (J.S. Dai, H.K. Lam and S.M. Vahed, "Soil Type Identification for Autonomous Excavation Based on Dissipation Energy," *Proceedings of the Institution of Mechanical Engineers, Part I, Journal of Systems and Control Engineering*, vol. 225, no. 1, pp. 35–50, 2011) received SAGE Best Paper Award in 2011.



Udeme Essien Ekong received his BEng degree in Electronic Engineering from King's College London in 2011 and is currently in the Penultimate year of his Ph.D. at the Department of Informatics, King's College London. His research interests include neural networks, support vector machines and fuzzy logic.



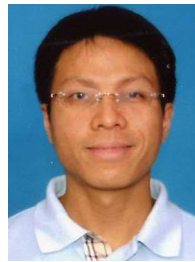
Bo Xiao received bachelor and master (Hons.) degree in Engineering from College of Communication Engineering, Chongqing University, P.R. China, in 2010 and 2013, respectively. He is currently pursuing his Ph.D. degree of Robotics at King's College London. His current research interests include: Neural Networks, Support Vector Machine, Fuzzy Logic and Fuzzy Model Based Control.



Gaoxiang Ouyang received the B.S. in automation and the M.S. in control theory and control engineering both from Yanshan University, China, in 2002 and 2004 respectively, and the Ph.D. degree at the Department of Manufacturing Engineering, City University of Hong Kong in 2010. Dr. Ouyang is an associate professor in the School of Brain and Cognitive Sciences at Beijing Normal University, Beijing, China. His research interests are bio-signal analysis, neural engineering, and dynamics system. He has published numerous refereed international journals and conference papers.



Hongbin Liu is currently a lecturer (Assistant Professor) in the Centre for Robotics Research, department of Informatics, King's College London, UK. He received the B.S. degree in 2005 from the Northwestern Polytechnique University, Xi'an, China, awarded the M.Sc. degree from King's College London in 2006. He received the Ph.D. degree in 2010 from Kings College London, UK. He received the Chinese Government Award for Outstanding Students Abroad in 2009. He is a member of IEEE. His research interests include the tactile/force perception based robotic cognition, the modelling of dynamic interaction, soft robotics and medical robotics.



Kit Yan Chan received his Ph.D. degree in Computing from London South Bank University, United Kingdom in 2006. After his Ph.D. study, he worked as a postdoctoral research fellow in the Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong, until 2009. Currently, he is a senior lecturer in the Department of Electrical and Computer Engineering, Curtin University of Technology, Australia. His research interests include new product development and manufacturing process design.



Sai Ho Ling received the BEng degree from the Department of Electrical Engineering, M.Phil. and Ph.D. degrees from the Department of Electronic and Information Engineering in the Hong Kong Polytechnic University in 1999, 2002 and 2007 respectively. Currently, he works in University of Technology, Sydney, Australia as a senior lecturer. He has authored and coauthored over 130 books, *International Journal and Conference Papers on Computational Intelligence and its Industrial Applications*. His current research interests include evolution computations, fuzzy logics, neural networks, hybrid systems and biomedical applications. Currently, he serves as Co-Editors-in-Chief for *Journal of Intelligent Learning Systems and Applications*.