Fast Learning Circular Complex-valued Extreme Learning Machine (CC-ELM) for Real-valued Classification Problems

R. Savitha[a], S. Suresh[b], N. Sundararajan[a]

[a]*School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore*
[b]*School of Computer Engineering, Nanyang Technological University, Singapore*

Abstract

In this paper, we present a fast learning fully complex-valued extreme learning machine classifier, referred to as 'Circular Complex-valued Extreme Learning Machine (CC-ELM)' for handling real-valued classification problems. The CC-ELM is a single hidden layer network with a non-linear input/hidden layer and a linear output layer. A circular transformation with a translational/rotational bias that performs a unique transformation of the real-valued feature to the complex plane is used as an activation function for the neurons in the input layer. The neurons in the hidden layer employ a fully complex-valued '$sech$' activation function. The input parameters of the CC-ELM are chosen randomly and output weights are computed analytically. The circular transformation that is used as the activation function in the input layer and the orthogonal decision boundaries of the complex-valued network helps the CC-ELM to perform real-valued classification tasks efficiently.

The performance of the CC-ELM is evaluated using a set of benchmark classification problems from the UCI machine learning repository. Finally, the performance of CC-ELM is evaluated in comparison with existing methods on two practical problems, viz., the acoustic emission signal classification and a mammogram classification. The results show that the CC-ELM performs better than other existing real-valued and complex-valued classifiers, especially in the highly unbalanced data sets.

Keywords: Complex-valued extreme learning machine, orthogonal decision boundaries, complex-valued classifiers, circular function, acoustic emission and mammogram.

---

1. Introduction

Complex-valued neural networks were originally developed to solve problems using complex-valued signals. According to Liouville's theorem, an analytic and bounded function is a constant function in the Complex plane. Therefore, non-linear, analytic and *almost everywhere* (*a.e*) bounded fully complex-valued activation functions [1, 2, 3] are used as activation functions in the literature. Several applications of complex-valued neural networks operating on complex-valued signals have been reported in the literature (eg. communication channel equalization [4], adaptive array signal processing [5, 6, 7], image reconstruction [8] etc.). Recently, the complex-valued neural networks have been shown to have better computational power than the real-valued neural networks [9] and are better in performing real-valued classification tasks because of their inherent orthogonal decision boundaries [10, 11].

The orthogonal decision boundaries and the better computational power of the complex-valued neural networks motivated researchers to develop new complex-valued classifiers for solving real-valued classification tasks. In [12], Nitta showed that a complex-valued neuron with an orthogonal decision boundary can be used to solve the XOR problem and the detection of symmetry problem, both of which can not be solved using a single real-valued neuron. Further, he also showed that the decision boundary of a single complex-valued neuron consists of two hyper surfaces which intersect orthogonally and this improves the generalization ability of a single hidden layered complex-valued network [10, 11]. However, the main challenge in using complex-valued neural networks to solve real-valued classification problems is to find a unique transformation that maps the real-valued feature space to the complex-valued feature space.

Aizenberg et. al. [13, 14], was the first one to suggest a complex-valued multi-valued neuron to solve real-valued classification problems. For a data set with C classes, a multi-valued neuron maps the complex-valued input to C discrete class labels (multiple-valued threshold logic) using a piecewise continuous activation function. The complex-valued Multi Layer Multi Valued Network (MLMVN) that employs the multi-valued neurons, uses a derivative free global error correcting learning rule to update the network parameters. In MLMVN, the normalized real-valued input features (x) are mapped to a full unit circle using $exp(i2\pi x)$ and the class labels are encoded by the roots of unity in the Complex plane. However, as the input features are mapped to a full unit circle, this mapping results in the same complex-valued features for the real-valued features with values 0 and 1 (transformation are non-unique). In addition, the multi-valued neurons map the complex-valued inputs to C discrete outputs in the unit circle. As number of classes (C) increases, the region of sectors per class within the unit circle decreases, which results in higher misclassification.

Recently Amin et. al. [15, 16] have developed a Phase Encoded Complex-Valued Neural Network (PE-CVNN) to perform classification tasks. In PE-CVNN, the complex-valued input features are obtained by phase encoding the real-valued input features (x) in $[0, \pi]$, using the transformation '$exp(i\pi x)$'. This transformation considers only two quadrants of the Complex plane and does not fully exploit the advantages of the orthogonal decision boundaries offered by the complex-valued neural networks. In addition, the activation functions used in the PE-CVNN are similar to those used in the split complex-valued neural networks. Therefore, the correlation between the real and imaginary parts of the error are not considered in the network parameter update and the gradients are not fully complex-valued [1]. PE-CVNN uses a gradient-descent based batch learning algorithm that requires significant computational effort to approximate the decision surface. Hence, there is a need to develop an efficient, fast learning, fully complex-valued classifier, which completely exploits the advantages of the orthogonal decision boundaries, to solve real-valued classification problems.

In this paper, we develop a fast learning fully complex-valued neural classifier referred to as 'Circular Complex-valued Extreme Learning Machine' (CC-ELM). CC-ELM uses a non-linear transformation with a translational/rotational bias term in the input layer, that results in a unique transformation from the real-valued to the complex-valued feature space. This transformation fully exploits the advantages of the orthogonal decision boundaries offered by complex-valued neural networks in classification problems. Such a transformation, called as the 'circular transformation', uses the '$sine$' function given by $z = sin\ (\ ax + i\ bx + \alpha)$, where α is the non-zero translational/rotational bias, a and b are real-valued non-zero transformation constants and i represents the complex operator.

CC-ELM is a single hidden layer neural network with the circular transformation as an activation function in the input layer, a fully complex-valued '$sech$' activation function (presented in [3]) in the hidden layer and a linear activation function in the output layer. Similar to the Complex-valued Extreme Learning Machines (C-ELM) [17], in the CC-ELM, the transformation constants, translational/rotational bias terms and the parameters of the input/hidden layer are chosen randomly and the output weights are calculated analytically. The nonlinear circular transformation and the $sech$ activation function in the hidden layer helps in exploiting the advantages of the orthogonal decision boundaries of complex-valued neural networks to provide a better generalization performance. Moreover, as the output weights are computed analytically, the CC-ELM network results in fast learning of the decision function.

Performance of CC-ELM classifier is evaluated on a set of benchmark classification problems from the UCI machine learning repository [18]. In this performance evaluation, the effect of the orthogonal decision boundaries of the complex-valued classifiers are studied by comparing the performances of the CC-EM classifier with the best results of existing real-valued classifiers. Next, to evaluate the benefits of the circular transformation, the performance of the CC-ELM is compared against other complex-valued classifiers, viz., phase encoded complex-valued neural network (PE-CVNN) [15, 16] and multi-layered multi-valued network (MLMVN) [13]. From the results, we infer that the CC-ELM classifier outperforms the real-valued and complex-valued classifiers, reported in the literature for these problems. CC-ELM exhibits significantly better performance when the data set is highly unbalanced.

Finally, the performance of the CC-ELM classifier is also evaluated on two practical classification problems viz., the acoustic emission signal processing for health monitoring [19] and the mammogram classification for breast cancer detection [20]. The results clearly highlight that the CC-ELM classifier provides a better generalization performance than other existing real-valued/complex-valued classifiers.

The paper is organized as follows: Section 2 presents the Circular complex-valued extreme learning machine (CC-ELM). It also highlights the advantages of the circular transformation over other transformations. In Section 3, an elaborate performance comparison of the CC-ELM network on a number of benchmark and real world classification problems is done with other existing results in the literature. Finally, Section 4 presents the conclusions from this study.

2. Circular Complex-valued Extreme Learning Machines Classifier

Classification is one of the most frequently encountered decision making problem in different real world problems. In this section, a fast learning fully-complex valued classifier called "Circular Complex-valued Extreme Learning Machine" (CC-ELM) is developed to solve real-valued classification tasks. CC-ELM performs real-valued classification tasks by transforming the real-valued input features to a set of unique complex-valued features and then performing efficient classification using the orthogonal decision boundaries, that are inherent in the complex-valued classifiers. First, we present the formulation of a real-valued classification problem in the Complex domain followed by a detailed description of the CC-ELM algorithm.

2.1. Real-valued Classification Problem Definition

Suppose we have N observations $\{(\mathbf{x}_1, c_1), \cdots, (\mathbf{x}_t, c_t), \cdots, (\mathbf{x}_N, c_N)\}$, where $\mathbf{x}_t \in \Re^m$ be the m-dimensional real-valued input features of t-th observation, $c_t \in \{1, 2, \cdots, C\}$ are its class labels, and C is the number of distinct classes. The observation data $(\mathbf{x}_t, c_t)$ are random in nature and the observation $\mathbf{x}_t$ provides some useful information on the probability distribution over the observation data to predict the corresponding class label (c_t) with certain accuracy.

Since the complex-valued neural networks inherently have orthogonal decision surfaces [10, 11], it will be interesting to solve the real-valued classification problem in the complex-valued framework. For this purpose, we convert the real-valued class label c_t into coded class label in the Complex domain $(\mathbf{y}^t = [y_1^t \cdots y_k^t \cdots y_C^t]^T)$ as

$$y_k^t = \begin{cases} 1 + 1i & \text{if} \quad c_t = k \\ -1 - 1i & \text{otherwise} \end{cases} \quad k = 1, 2, \cdots, C \tag{1}$$

Therefore, the real-valued classification problem using complex-valued neural networks can be viewed as finding the decision function F that maps the real-valued input features to the complex-valued coded class labels, i.e., $F : \Re^m \rightarrow \mathbb{C}^C$. For notational convenience, the superscript t will be dropped in the rest of the text.

In the next section, we present a fast learning CC-ELM classifier to estimate the decision function (F). The detailed description of the algorithm is also presented.

2.2. Circular Complex-valued Extreme Learning Machine (CC-ELM)

The basic building block of the Circular Complex-valued Extreme Learning Machine (CC-ELM) classifier is the Complex-valued Extreme Learning Machine (C-ELM) [17]. The CC-ELM classifier is a single hidden layer network with m input neurons, K hidden neurons and C output neurons, as shown in Fig. 1. The neurons in the input layer of the CC-ELM classifier employ a non-linear circular transformation as the activation function, as shown in the inset of Fig. 1.

The circular transformation, used as the activation function in the input layer to transform the real-valued input features into the Complex domain ($\Re \rightarrow \mathbb{C}$) is given by

$$z_i = sin\left(ax_i + ibx_i + \alpha_i\right), \quad i = 1, 2, \cdots, m \tag{2}$$

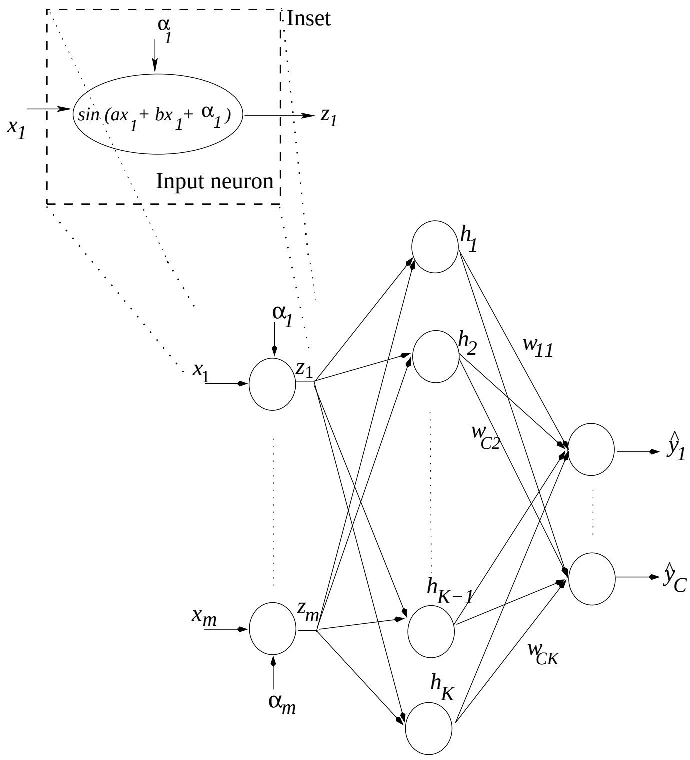

Figure 1: The architecture of a circular complex-valued extreme learning machine. The expanded box (inset figure) shows the actual circular transformation function which maps real-valued input feature to complex-valued feature.

where $a,\ b \in \Re^+$ are non-zero transformation constants and α_i is the translational/rotational bias term of i-th input neuron. Note that the real-valued input features ($\mathbf{x}$) are normalized between 0 and 1. The transformation constants a and b are chosen randomly such that $0 < a, b \leq 1$. The term α_i is a non-zero rotational/translational bias in radians (i.e., $0 < \alpha_i < 2\pi$).

Fig. 2 shows the characteristics of the circular transformation with $a, b = 1$, $x \in [0, \pi]$ and two randomly chosen values of α ($\alpha_1 = 3.7$ and $\alpha_2 = 1.32$). It can be observed from this figure that the randomly chosen bias terms α_1 and α_2 perform translation/rotation of the feature vector in different quadrants of the complex-valued feature space. Therefore, the bias term ($\alpha_i,\ i = 1,\ \cdots,\ m$) associated with each input feature ensures that all the input features are not mapped onto the same quadrant of the Complex plane. Thus, as the input features are well distributed in the Complex plane, the CC-ELM fully exploits the orthogonal decision boundaries of the fully complex-valued neural networks.

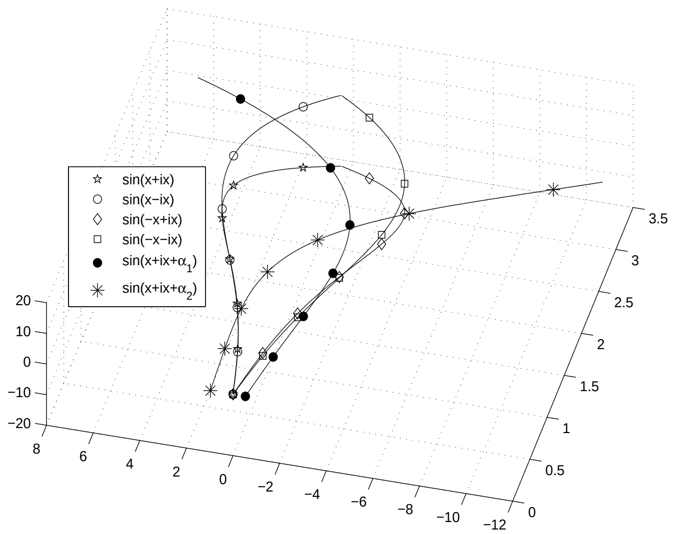

Figure 2: Significance of translational/rotational bias term (α) in the circular transformation

The essential properties of a fully complex-valued activation function given in [1] states that for a complex-valued non-linear function to be used as an activation function, the function has to be analytic and bounded *almost everywhere* (*a.e*). As the circular transformation is used as an activation function in the input layer, we need to ensure that the transformation satisfies the essential properties of a fully complex-valued activation function. The suggested '$sine$' activation function is an analytic function with an essential singularity at ∞.

$$ax_i + ibx_i + \alpha_i \quad = \quad \infty \text{ if } ax_i + \alpha_i = \infty \text{ or } bx_i = \infty \tag{3}$$

Since the transformation constants (a, b) and the translational/rotational bias (α_i) are restricted between $(0, 1]$ and $(0, 2\pi)$ respectively, the transformation becomes unbounded only when the input feature is ∞. The real-valued feature is normalized between $[0, 1]$. Therefore the circular transformation is analytic and bounded *almost everywhere*. Thus, this activation function satisfies the essential properties needed for a complex-valued activation function.

The neurons in the hidden layer of the CC-ELM employ a fully-complex valued '$sech$' activation function (Gaussian like) as developed in [3]. The hidden layer response (h_j) of the CC-ELM is given by

$$h_j = sech\left[\mathbf{u}_j^T\left(\mathbf{z}-\mathbf{v}_j\right)\right], \quad j=1,2,\cdots,K \tag{4}$$

where $\mathbf{u}_j \in \mathbb{C}^m$ is the complex-valued scaling factor of the j-th hidden neuron and $\mathbf{v}_j \in \mathbb{C}^m$ is the complex-valued center of the j-th hidden neuron.

The neurons in the output layer employ a linear activation function. The output ($\widehat{\mathbf{y}} = [\widehat{y}_1 \cdots \widehat{y}_i \cdots \widehat{y}_C]^T$) of the CC-ELM network with K hidden neurons is

$$\widehat{y}_i = \sum_{j=1}^{K} w_{ij} h_j, \quad i=1,2,\cdots,C \tag{5}$$

where w_{ij} is the output weight connecting the j-th hidden neuron and the i-th output neuron.

The estimated class label ($\widehat{c}$) is then obtained using

$$\widehat{c} = arg \max_{i=1,2,\cdots,C} \text{ real-part of } \widehat{y}_i \tag{6}$$

The output of CC-ELM given in Eq. (5) can be written in matrix form as,

$$\widehat{Y} = WH \tag{7}$$

where W is the matrix of all output weights connecting hidden and output neurons. The H is the response of hidden neurons for all training samples and is given as

$$H\left(V,U,Z\right) = \begin{bmatrix} sech\left(\mathbf{u}_1\|\mathbf{z}_1-\mathbf{v}_1\|\right) & \cdots & sech\left(\mathbf{u}_1\|\mathbf{z}_N-\mathbf{v}_1\|\right) \\ \vdots & \vdots & \vdots \\ sech\left(\mathbf{u}_K\|\mathbf{z}_1-\mathbf{v}_K\|\right) & \cdots & sech\left(\mathbf{u}_K\|\mathbf{z}_N-\mathbf{v}_K\|\right) \end{bmatrix} \tag{8}$$

Note H is $K \times N$ hidden layer output matrix. The j-th row of the H matrix represents the hidden neuron response (h_j) with respect to the inputs $\mathbf{z}_1, \cdots, \mathbf{z}_N$.

In CC-ELM, the transformation constants (a, b), the translational/rotation bias term ($\boldsymbol{\alpha}$), center (V) and scaling factor (U) are chosen randomly and the output weights W are estimated by the least squares method according to:

$$W = YH^{\dagger} \tag{9}$$

where $H^{\dagger}$ is the generalized Moore-Penrose pseudo-inverse [21] of the hidden layer output matrix and Y is the complex-valued coded class label.

In short, the CC-ELM algorithm can be summarized as:

- For a given training set (X, Y), select the appropriate number of hidden neurons K.
- Randomly select the transformation constants (a, b), the bias term $(\boldsymbol{\alpha})$, the scaling factor (U) and the neuron centers (V).
- Then calculate the output weights W analytically: $W = YH^{\dagger}$.

The performance of the CC-ELM is influenced by the selection of appropriate number of hidden neurons. Recently, an incremental constructive method to determine the appropriate number of hidden neurons for the C-ELM has been presented in [22]. This method adds hidden neurons incrementally, until a specified maximum number of neurons is reached or until the desired approximation accuracy is achieved. In this paper, we use a simple neuron incremental-decremental strategy, similar to the one presented in [23] for real-valued networks.

From this section, it can be observed that the two important properties of the CC-ELM classifier are that the CC-ELM classifier uses a unique circular transformation to transform the input features from $\Re \rightarrow \mathbb{C}$ and it completely exploits the orthogonal decision boundaries, inherent of the complex-valued neural networks. Moreover, as the CC-ELM computes the output weights analytically, the classifier requires lesser computational effort than other complex-valued classifiers.

3. Performance Evaluation of the CC-ELM Classifier

In this section, we present the performance results of the CC-ELM classifier, in comparison with other existing complex-valued and real-valued classifiers. For performance evaluation, we consider both multi-category/binary classification problems from the UCI machine learning repository [18]. Based on a wide range of the Imbalance Factor (I. F.) (as defined in [24]) of the data sets, three multi-category and four binary data sets are chosen for the study. The imbalance factor is defined as

$$\text{(I. F.)} \;=\; 1 - \left(\frac{\min_{i=1\cdots C} \;\; N_i}{\sum_{i=1}^{C} N_i} \right) * C \qquad (10)$$

where N_i is the total number of samples belonging to the class i.

The description of these data sets including the number of classes, the number of input features, the number of samples in the training/testing and the imbalance

Table 1: Description of benchmark data sets selected from [18] for performance study

Type of data set	Problem	No. of features	No. of classes	No. of samples		I. F.
				Training	Testing	
Balanced	Image Segmentation (IS)*	19	7	210	2100	0
Unbalanced	Vehicle Classification (VC)*	18	4	424	422	0.1
	Glass Identification (GI)*	9	7	109	105	0.68
	Liver Disorder	6	2	200	145	0.17
	PIMA Data	8	2	400	368	0.225
	Breast Cancer	9	2	300	383	0.26
	Ionosphere	34	2	100	251	0.28

* Multi-category classification problem

factor are presented in Table 1. From the table, it can be observed that the problems chosen for the study have both balanced and unbalanced data set and the imbalance factors of the data sets vary widely.

Finally, the CC-ELM is used to solve two real-world classification problems: the acoustic emission signal processing for health monitoring [19] and the mammogram classification for breast cancer detection [20].

The classification/confusion matrix Q is used to obtain the statistical measures for both the class-level and global performance of the various classifiers. Class-level performance is measured by the percentage classification (η_i) which is defined as:

$$\eta_i = \frac{q_{ii}}{N_i} \times 100\% \tag{11}$$

where q_{ii} is the total number of correctly classified samples in the class c_i and N_i is the total number of samples belonging to a class c_i in the training/testing data set.

The global measures used in the evaluation are the average per-class classification accuracy (η_a) and the over-all classification accuracy (η_o) defined in [25] as:

$$\eta_a = \frac{1}{C}\sum_{i=1}^{C}\eta_i$$
$$\eta_o = \frac{\sum_{i=1}^{C} q_{ii}}{\sum_{i=1}^{C} N_i} \times 100\% \quad (12)$$

The performance of the classifiers are compared using these class-level and global performance measures.

3.1. Performance Evaluation using Benchmark Classification Problems

First, the performance of the CC-ELM is evaluated on the three multi-category benchmark classification problems in Table 1, *i.e.,* the Image Segmentation (IS), the Vehicle Classification (VC) nd the Glass Identification (GI) data sets. The main aim here is to highlight the advantages of the CC-ELM classifier with respect to the following two factors, viz., :

- The orthogonal decision boundaries that are characteristic of the complex-valued neural networks and
- The unique, non-linear circular transformation used in converting the real-valued features to the Complex domain.

To study the effect of the orthogonal decision boundaries, the performance of the CC-ELM is compared against the best performing results of the existing real-valued classifiers for this problem, viz., Support Vector Machine (SVM) [26] and Self-adaptive Resource Allocation Network (SRAN) [27]. As the CC-ELM classifier is developed in the framework of C-ELM, its performance is also compared against the best results of real-valued ELM [28] and its variant, the Real-Coded Genetic Algorithm for input weight selection (RCGA-ELM) [29, 30]). The results for the SVM and SRAN are reproduced from [27]. The results for the RCGA-ELM is reproduced from [29]. Note that the SVM is implemented in C language [27] and the other algorithms are implemented in MATLAB environment on a Pentium 4 processor with 4GB RAM.

To study the effect of the circular transformation, the performance of the CC-ELM classifier is compared against other existing complex-valued classifiers *viz.*, the PE-CVNN [15, 16] and MLMVN [13] classifier. The results for the single layered PE-CVNN are reproduced from [16] and the results for the MLMVN classifier are obtained using the software simulator available in the author's web site [1].

Balanced Data Set: Image Segmentation (IS)

The image segmentation data set available in the UCI machine learning repository [18] is a multi-category well-balanced data set with 7 classes. The data set description for this problem is presented in Table 1. The data set for the IS problem has 30 samples in each class for training and 300 samples in each class for testing. Hence, as reported in Table 1, the imbalance factor (I.F.) for this problem is 0.

The structure of the CC-ELM and parameters of the network are selected using the procedure described in section 2.2. For this study, the parameters of the MLMVN are chosen according to [13]. The testing performance, the number of hidden neurons (K) and the training time for the IS problem are presented in Table 2. From the table, one can see that the CC-ELM classifier performs better than other classifiers. It can also be observed that the CC-ELM classifier requires lower computational effort, compared to all the other classifiers. The orthogonal decision boundary and the unique circular transformation of the CC-ELM classifier helps it to outperform all the real-valued classifiers in the classification of this well-balanced data set. Also, even without any self-regulation and sequence altering, the CC-ELM performs better than the SRAN classifier, with reduced computational effort.

It can also be observed that the CC-ELM classifier is better than the other complex-valued classifiers, available in the literature. Although the performance of the PE-CVNN classifier is almost equal to that of the CC-ELM classifier, the PE-CVNN classifier uses 75% of the samples in training, while the CC-ELM classifier uses only 10% of the samples for training. Also, the PE-CVNN classifier uses a computationally intensive gradient descent based learning algorithm as compared to the CC-ELM that is computationally less intensive.

Unbalanced Data Sets: Vehicle Classification (VC) and Glass Identification (GI)

Next, we consider two unbalanced multi-category data sets with varying degrees of imbalance factors for the performance study, viz., the Vehicle Classification (VC) problem with a lower imbalance factor (0.1) and the Glass Identification

[1] http://www.eagle.tamut.edu/faculty/igor/Downloads.htm

Table 2: Performance comparison on well-balanced multi-category image segmentation problem

Type	Algorithm	K	Training Time (s)	Testing Efficiency (η_o)
Real-valued	ELM	100	0.03	90.67
	RCGA-ELM	50	-	91.00
	SRAN	47	22	92.29
	SVM	96[a]	721	90.62
Complex-valued	MLMVN	80	1384	83
	PE-CVNN	-	-	93.2[b]
	CC-ELM	**60**	**0.03**	**93.52**

[a]- number of support vectors

[b]-A single layer network was used in [16]. Also, in [16], 75% of the total samples are used in training. In our work, we use only 10% of the samples in training.

(GI) problem with a higher imbalance factor (0.68).

The detailed description of these data sets including the number of input features, the number of samples in the training and testing data set and the number of classes are presented in Table 1.

The structure of the CC-ELM classifier and its parameters are chosen based on the procedure presented in section 2.2. Based on the simple constructive-destructive method [23] for selecting the number of hidden neurons, we chose 75 neurons for the VC and 100 neurons for the GI problem. The input and hidden neuron parameters of the CC-ELM (a, b, $\boldsymbol{\alpha}$, V and U) are chosen randomly and the output weights W are computed analytically using the training samples. The classification performance measures, calculated using unseen testing samples are reported in Table 3.

For these problems, the classification performance of the CC-ELM algorithm, in comparison with other real and complex-valued classifiers, is presented in Table 3. From the table, it can be observed that the average and over-all classification accuracies of the CC-ELM classifier is higher than those of the real-valued classifiers for the VC and GI problems.

The advantage in the classification performance of the CC-ELM classifier in the highly unbalanced glass identification data set can be clearly seen from the Table 3. Also, with no self-regulation, the CC-ELM classifier outperforms the best results in real-valued (SRAN algorithm) classifier by nearly 6% and 4% in the vehicle classification and the glass identification problems, respectively. Moreover,

Table 3: Performance comparison on unbalanced multi-category classification problems

Problem	Type	Algorithm	K	Training Time (s)	Testing Efficiency (η_o)	Testing Efficiency (η_{av})
Vehicle Classification Problem	Real-valued	ELM	300	0.81	78.17	78
		RCGA-ELM	75	-	74.2	74.4
		SRAN	55	113	75.12	76.86
		SVM	234[a]	550	68.72	67.99
	Complex-valued	PE-CVNN	-	-	78.7[b]	-
		MLMVN	90	1396	78	77.25
		CC-ELM	**85**	**0.1084**	**82.23**	**82.52**
Glass Identification Problem	Real-valued	ELM	60	0.21	73	65.46
		RCGA-ELM	60	-	78.1	-
		SRAN	59	28	86.21	80.95
		SVM	102	320	64.23	60.01
	Complex-valued	PE-CVNN	-	-	65.5[b]	-
		MLMVN	85	1421	73.24	66.83
		CC-ELM	**100**	**0.08**	**94.44**	**84.52**

[a]- number of support vectors

[b]-A single layer network was used in [16]. Also, in [16], 75% of the total samples are used in training. In our work, only 50% of the total samples are used in training as per the guidelines in UCI machine learning database

the CC-ELM performs significantly better than the other existing complex-valued classifiers. It can be clearly seen that the PE-CVNN, whose performance is comparable to that of the CC-ELM in the well-balanced IS data set does not perform well in classification of the unbalanced VC and GI data sets. Another interesting observation that can be made from the table is that the improvement in classification accuracy of the CC-ELM is higher in the highly unbalanced GI data set than that in the VC data set.

Comparing the performance of the CC-ELM classifier against other complex-valued classifiers on multi-category classification problems, it can be seen that the performance of the CC-ELM classifier is better than the other complex-valued classifiers, especially in unbalanced data sets. Though both the PE-CVNN and the MLMVN also use orthogonal decision boundaries, their performances may be affected by the transformation, learning algorithm and activation function.

Binary Classification Data Sets: Recently, in [31], an optimization based real-valued Extreme Learning Machine (O-ELM) is presented for solving classification problems and the performance of the O-ELM has been reported on a set of binary classification data sets. As the CC-ELM is also developed in the framework of extreme learning machines, we also compare the performance of the CC-ELM here in comparison with the O-ELM [31] on the 4 binary benchmark data sets from the UCI machine learning repository. The details of the problems considered for the study are presented in Table 1. The performance comparison of the results are tabulated in Table 4.

Table 4: Performance comparison on benchmark binary classification problems

Problem	Type	Algorithm	K	Training Time (s)	Testing Efficiency (η_o)
Liver disorders	Real-valued	O-ELM	131	0.1734	72.34
	Complex-valued	**CC-ELM**	**10**	**0.059**	**75.5**
PIMA data	Real-valued	O-ELM	217	0.2867	77.27
	Complex-valued	**CC-ELM**	**20**	**0.073**	**81.25**
Breast cancer	Real-valued	O-ELM	66	0.1423	96.32
	Complex-valued	**CC-ELM**	**15**	**0.0811**	**97.39**
Iono-sphere	Real-valued	O-ELM	32	0.0359	89.48
	Complex-valued	**CC-ELM**	**15**	**0.0312**	**92.43**

From the table, it is clearly evident that even without any optimization tools the CC-ELM outperforms the O-ELM classifier [31]. When compared to O-ELM, the network used in the classification is also compact.

Thus, from the study on both the multi-category and binary benchmark data sets considered, it may be noted that the CC-ELM classifier outperforms the existing complex-valued and real-valued classifiers. It can also be inferred that the improvement in the classification performance of the CC-ELM classifier is better in the highly unbalanced data sets. The better classification performance of the CC-ELM classifier can be associated to the orthogonal decision boundaries of the complex-valued classifiers and the unique circular transformation used as the activation function in the input layer.

Next, we study the performance of the CC-ELM classifier on two real-world classification problem data sets, viz., an acoustic emission data set for health monitoring [19] and the mammogram classification data set for breast cancer detection

[20].

3.2. Acoustic Emission Signal Classification for Health Monitoring

Acoustic emission signals are electrical version of the stress or pressure waves, produced by sensitive transducers. These waves are produced due to the transient energy release caused by irreversible deformation processes in the material [19]. There are various sources of acoustic emission, and these sources can be characterized by the acoustic signals. The classification of acoustic emission signals based on their sources is a very difficult problem, especially in the real world, where ambient noise and pseudo acoustic emission signals exist. Even in a noise free environment, superficial similarities exist between acoustic emission signals produced by different sources, making the classification task cumbersome.

In the study conducted in [19], noise free burst type acoustic emission signal from a metal surface is assumed. The data set presented in [19] uses 5 input features to classify the acoustic signals to one of the 4 sources, i.e., the pencil source, the pulse source, the spark source and the noise. A training data set with 62 samples and testing data set with 137 samples are used for the acoustic emission signal classification problem. For details of the input features and the experimental set up used in the data collection, one should refer to [19].

The performance study results of the CC-ELM classifier in comparison to best results available in the literature for the acoustic emission signal classification problem, is presented in Table 5. The results of the CC-ELM algorithm are compared against a Fuzzy K-means clustering algorithm [19] and the ant colony optimization algorithm [32]. It can be seen that the CC-ELM algorithm required only 10 neurons to achieve an over-all testing efficiency of 99.27%. Thus, the CC-ELM performs an efficient classification of the acoustic emission signals using a compact network.

3.3. Mammogram Classification for Breast Cancer Detection

Mammogram is a better means for early diagnosis of breast cancer, as tumors and abnormalities show up in mammogram much before they can be detected through physical examinations [33]. Clinically, identification of malignant tissues involves identifying the abnormal masses or tumors, if any, and then classifying the mass as either malignant or benign [34]. However, once a tumor is detected, the only method of determining whether it is benign or malignant is by conducting a biopsy, which is an invasive procedure that involves the removal of the cells or tissue from a patient. A non-invasive method of identifying the abnormalities in

Table 5: Performance comparison results for the acoustic emission problem

Classifier domain	Classifier	Train Time	K	Testing	
				η_o	η_{av}
Real-valued	Ant Colony Optimization	-	-	90.51	89.27
	Fuzzy C-Means Clustering	-	-	-	93.34
Complex-valued	**CC-ELM**	**0.031**	**10**	**99.27**	**99.17**

a mammogram can reduce the number of unnecessary biopsies, thus sparing the patients of inconvenience and saving medical costs.

In this study, the mammogram database available in [20] has been used. The 9 input features, extracted from the mammogram of the identified abnormal mass, are used to classify the tumor as either malignant or benign. Here, 97 samples are used to develop CC-ELM classifier and the performance of CC-ELM classifier is evaluated using remaining 11 samples. For further details on the input features and the data set, one should refer to [20].

The performance results of the CC-ELM classifier, in comparison with other real-valued classifiers are presented in Table 6. From the table, it is seen that the CC-ELM classifier performs a highly efficient classification with 100% classification accuracy. This is higher than the best results available in the literature for this problem by approximately 4%. Comparing the classification performance of the CC-ELM classifier to that of the real-valued ELM classifier, the classification performance is improved by nearly 9%.

Thus, from the performance study conducted with different real-valued and complex-valued classifiers for the chosen benchmark data sets and practical classification problems, it can be observed that the CC-ELM classifier performs better than other existing classifiers.

4. Conclusions

In this paper, we have presented the Circular Complex-valued Extreme Learning Machine (CC-ELM) classifier for performing real-valued classification tasks

Table 6: Performance comparison results for the mammogram problem

Classifier domain	Classifier	Train Time	K	Testing	
				η_o	η_{av}
Real-valued	ELM	0.032	30	91	91
	SVM [35]	-	-	-	95.44
Complex-valued	**CC-ELM**	**0.047**	**40**	**100**	**100**

in the Complex domain. The CC-ELM classifier uses a single hidden layer network with one non-linear input layer, one non-linear hidden layer and one linear output layer. At the input layer, a unique nonlinear circular transformation is used as the activation function to convert the real-valued input features to the Complex domain ($\mathbb{R} \rightarrow \mathbb{C}$). At the hidden layer, the complex-valued input features are mapped on to a higher dimensional Complex plane, using the '$sech$' activation function. In the CC-ELM, the input parameters and the parameters of the hidden layer are chosen randomly and the output weights are calculated analytically. Therefore, the CC-ELM requires lesser computational effort to perform classification tasks. The performance of the CC-ELM classifier is compared with other real-valued and complex-valued classifiers using benchmark classification problems from the UCI machine learning repository and two real-valued, practical classification problems. The performance of CC-ELM on these data sets clearly show that it performs better than the well-known real-valued and other complex-valued classifiers available in the literature, especially for unbalanced data sets.

References

[1] T. Kim, and T. Adali, Fully-complex multilayer perceptron network for nonlinear signal processing, *Journal of VLSI Signal Processing*, **32** (1-2) (2002), pp. 29-43.

[2] R. Savitha, S. Suresh, N. Sundararajan and P. Saratchandran, A new learning algorithm with logarithmic performance index for complex-valued neural networks, *Neurocomputing*, **72** (16-18) (2009), pp. 3771-3781.

[3] R. Savitha, S. Suresh and N. Sundararajan, A fully complex-valued radial basis function network and its learning algorithm, *International Journal of Neural Systems*, **19** (4) (2009), pp. 253 - 267.

[4] K. Burse, R. N. Yadav and S. C. Shrivastava, Channel equalization using neural networks: A review, *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, **40** (3) (2010), pp. 352-357.

[5] J. C. Bregains and F. Ares, Analysis, synthesis, and diagnostics of antenna arrays through complex-valued neural networks, *Microwave and Optical Technology Letters*, **48** (8) (2006), pp. 1512-1515.

[6] C. Shen, H. Lajos and S. Tan, Symmetric complex-valued RBF receiver for multiple-antenna-aided wireless systems, *IEEE Transactions on Neural Networks*, **19** (9) (2008), pp. 1659-1665.

[7] R. Savitha, S. Vigneshwaran, S. Suresh and N. Sundararajan, Adaptive beamforming using complex-valued radial basis function neural networks, *in: Proc. of TENCON'09, IEEE Region 10 Conference, Singapore, Nov 23-26*, 2009, pp. 1-6.

[8] N. Sinha and M. Saranathan and K. R. Ramakrishna and S. Suresh, Parallel magnetic resonance imaging using neural networks, *in: Proc. of ICIP'07, IEEE International Conference on Image Processing*, vol. 3, 2007, pp. 149-152.

[9] T. Nitta, The computational power of complex-valued neuron, *Artificial Neural Networks and Neural Information Processing ICANN/ICONIP. Lecture Notes in Computer Science*, **2714** (2003), pp. 993–1000.

[10] T. Nitta, On the inherent property of the decision boundary in complex-valued neural networks, *Neurocomputing*, **50** (2003), pp. 291-303.

[11] T. Nitta, Orthogonality of decision boundaries of complex-valued neural networks, *Neural Computation*, **16** (1) (2004), pp. 73-97.

[12] T. Nitta, Solving the XOR problem and the detection of symmetry using a single complex-valued neuron, *Neural Networks*, **16** (8) (2003), pp. 1101-1105.

[13] I. Aizenberg and C. Moraga, Multilayer feedforward neural network based on multi-valued neurons (MLMVN) and a backpropagation learning algorithm, *Soft Computing*, **11** (2) (2007), pp. 169-183.

[14] I. Aizenberg, D. V. Paliy, J. M. Zurada and J. T. Astola, Blur identification by multilayer neural network based on multivalued neurons, *IEEE Transactions on Neural Networks*, **19** (5) (2008), pp. 883-898.

[15] M. F. Amin and K. Murase, Single-layered complex-valued neural network for real-valued classification problems, *Neurocomputing*, **72** (4-6) (2009), pp. 945-955.

[16] M. F. Amin, M. M. Islam and K. Murase, Ensemble of single-layered complex-valued neural networks for classification tasks, *Neurocomputing*, **72** (10-12) (2009), pp. 2227-2234.

[17] M. B. Li, G. B. Huang, P. Saratchandran, and N. Sundararajan, Fully complex extreme learning machine, *Neurocomputing*, **68** (1-4) (2005), pp. 306-314.

[18] C. Blake and C. Merz, UCI Repository of Machine Learning Databases, Department of Information and Computer Sciences, University of California, Irvine, URL: <http://archive.ics.uci.edu/ml/>, 1998.

[19] S. N. Omkar, S. Suresh, T. R. Raghavendra, and V. Mani, Acoustic emission signal classification using fuzzy C-means clustering, *in: Proc. of the ICONIP'02, 9th International Conference on Neural Information Processing*, vol. 4, 2002, pp. 1827-1831.

[20] J. Suckling et al, The Mammographic Image Analysis Society Digital Mammogram Database, *Exerpta Medica International Congress Series*, **1069** (1994), pp. 375-378.

[21] J. M. Ortega, Matrix Theory, Plenum Press, New York, USA, 1986.

[22] G. Huang, M. Li, L. Chen and C. K. Siew, Incremental extreme learning machine with fully complex hidden nodes, *Neurocomputing*, **71** (4-6) (2008), pp. 576-583.

[23] S. Suresh, S. N. Omkar, V. Mani and T. N. Guru Prakash, Lift coefficient prediction at high angle of attack using recurrent neural network, *Aerospace Science and Technology*, **7** (8) (2003), pp. 595-602.

[24] S. Suresh, N. Sundararajan and P. Saratchandran, Risk-sensitive loss functions for sparse multi-category classification problems, *Information Sciences*, **178** (12) (2008), pp. 2621-2638.

[25] S. Suresh, N. Sundararajan and P. Saratchandran, A sequential multi-category classifier using radial basis function networks, *Neurocomputing*, **71** (7-9) (2008), pp. 1345-1358.

[26] N. Cristianini and J. S. Taylor, An Introduction to Support Vector Machines, Cambridge University Press, Cambridge, UK, 2000.

[27] S. Suresh, Keming Dong and H. J. Kim, A sequential learning algorithm for self-adaptive resource allocation network classifier, *Neurocomputing*, **73** (16-18) (2010), pp. 3012-3019.

[28] G. B. Huang, Q. Y. Zhu and C. K. Siew, Extreme learning machine: Theory and applications, *Neurocomputing*, **70** (1-3) (2006), pp. 489-501.

[29] S. Suresh, R. Venkatesh Babu and H. J. Kim, No-reference image quality assessment using modified extreme learning machine classifier, *Applied Soft Computing*, **9** (2) (2009), pp. 541-552.

[30] S. Suresh, S. Saraswathi and N. Sundararajan, Performance enhancement of extreme learning machine for multi-category sparse data classification problems, *Engineering Applications of Artificial Intelligence*, **23** (7) (2010), pp. 1149-1157.

[31] G. B. Huang, X. Ding and H. Zhou, Optimization method based extreme learning machine for classification, *Neurocomputing*, In press, 2010.

[32] S. N. Omkar and U. Raghavendra Karanth, Rule extraction for classification of acoustic emission signals using ant colony optimisation, *Engineering Applications of Artificial Intelligence*, **21** (8) (2008), pp. 1381-1388.

[33] P. Gamigami, Atlas of Mammography: New Early Signs in Breast Cancer, Blackwell Science, Cambridge, MA, USA, 1996.

[34] C. Aize, Q. Song, X. Yang, S. Liu and C. Guo, Mammographic mass detection by vicinal support vector machine, *in: Proc. of ICNN '04, International Conference on Neural Networks*, vol. 3, 2004, pp. 1953–1958.

[35] T. S. Subashini, V. Ramalingam, S. Palanivel, Automated assessment of breast tissue density in digital mammograms, *Computer Vision and Image Understanding*, **114** (1) (2010), pp. 33-43.