SETSCI Conference Proceedings
4 (6), 169-173, 2019

4th International Symposium on Innovative Approaches in Engineering and Natural Sciences

November 22-24, 2019, Samsun, Turkey

https://doi.org/10.36287/setsci.4.6.050
2687-5527/ © 2019 The Authors. Published by SETSCI

The New Activation Function for Complex Valued Neural Networks: Complex Swish Function

Mehmet ÇELEBİ[1*+], Murat CEYLAN [2]

[1] Department of Project/İller Bankası A.Ş, Ankara, Turkey
[2] Department of Electrical and Electronics Engineering/Konya Technical University, Konya, Turkey
[*]Corresponding author: mehmetc@ilbank.gov.tr
[+]Speaker: mehmetc@ilbank.gov.tr
Presentation/Paper Type: Oral / Full Paper

***Abstract* –** Complex-valued artificial neural network (CVANN) has been developed to process data with complex numbers directly. Weights, threshold, inputs and outputs are all complex numbers in the CVANN. The convergence of the CVANN back propagation algorithm depends on some factors such as selection of appropriate activation function, threshold values, initial weights and normalization of data. The most important of these factors is the selection of the appropriate activation function**.** The selection of activation function determines the convergence and general formation characteristics of the complex back propagation algorithm. In this study, the swish activation function discovered by Google researchers Prajit Ramachandra, Barret Zoph and Quoc V. Le is discussed in the complex domain. Swish activation function, which gives good results in real plane, has been studied in the complex plane. We have compared the performance of swish activation functions on the complex XOR and symmetry problems with other known activation functions. The simulations' results show that the proposed network using swish activation function, gives the best results when compared to other networks using the traditional complex logarithmic sigmoid and tangent sigmoid activation functions.

Keywords *–Complex-valued neural network, activation function, swish, complex XOR, complex symmetry*

I. Introduction

Artificial neural networks, (ANN) is an artificial intelligence method which is developed to realize the capabilities of learning, which is one of the characteristics of human brain, to generate new information, to create and discover new information automatically without any help. In ANN, the nerve groups come together to form the nerve layers. These interconnections are provided by lines having certain weight coefficients. Data from the nerve input are collected and form the input expression of the nerve. This input expression is passed through a function located at the nerve output to obtain the nerve output. This function at the nerve output is called the activation function [1].

Complex valued artificial neural network (CVANN), refers to neural networks whose weights, threshold values, inputs and output signals are all complex numbers. CVANN has been developed to process data with complex numbers directly. In the solution of problems involving data with complex numbers, ANN should be applied separately for real and imaginary parts of complex data when known method is used. However, when CVANN is applied for the same problem, data can be processed directly without having to separate real and imaginary parts. Thus, it has been observed that the processing time is reduced and the accuracy rate is increased [2].

One of the main advantages of CVANN is the ability to work with phase information, which is crucial for analyzing signals and solving different pattern recognition and classification problems. Even in the analysis of real-value signals, one of the most efficient approaches is frequency domain analysis involving complex numbers. By analyzing the signal properties in the frequency domain, we see that each signal is characterized by magnitude and phase, which have different information about the signal. [1].

CVANN has many areas of application such as image processing, radar, telecommunication and speech recognition which dealing with complex numbers. The convergence of the CVANN back propagation algorithm performing these processes varies depending on some factors such as selection of appropriate activation function, threshold values, initial weights and normalization of data. The most important of these factors is the selection of the appropriate activation function. The selection of an activation function determines the convergence and general formation characteristics of the complex back propagation algorithm [3].

One of the difficulties in applying back propagation algorithm to complex domain in complex valued artificial neural networks is the selection of the appropriate activation function. For a practical application of complex multi-layer perceptron, the activation function must be limited. When the activation function is not bounded, multi-layer perceptron has been shown to cause errors in software and hardware applications. In summary, the activation function clearly satisfies the following five properties:

- The activation function φ (z) should not be linear in both the real and imaginary parts of Z, Z_R and Z_I. Otherwise, the multi-layer perceptron will have no advantage. If correct, using a multi-layer perceptron would be equal to a single-layer perceptron [1].

- The function φ (z) should be bounded. The formulas described for the forward passage of the multilayered perceptron require limitation. Otherwise there will be interruptions during the training [1].

- The Partial derivatives of φ (z) should exist and be bounded. Since we use complex back-propagation, the partial derivatives of φ (z) need to be bounded [1].

- The function φ (z) must be defined as a complex function that is analytic all over the complex plane [1].

II. Materials and Method

A. Complex-Valued Artificial Neural Network (CVANN)

The difference from the real valued ANN, all the inputs, outputs, weights and biases are complex numbers in CVANN. The complex valued neuron model is given in Fig.1.

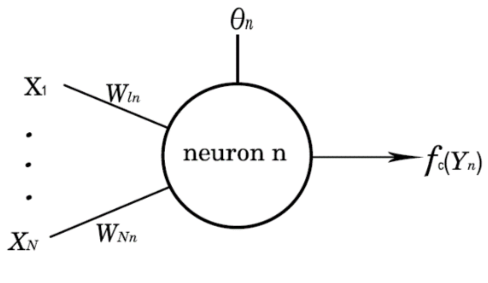

Fig. 1 CVANN model

The internal potential of a neuron Y_n with its input value, weight and threshold value are complex, is defined as given in Eq. 1 where W_{nN} is the weight connecting neuron n and m; X_N is the input signal from neuron m, Θ_n is the threshold value of neuron n.

$$Y_n = \sum_N W_{nN} X_N + \Theta_n \tag{1}$$

In this study, the complex back-propagation (CBP) algorithm have been used. Artificial neural networks with back propagation consist of two main stages; the forward-propagation and back-propagation.

<u>*Forward-propagation in CVANN*</u>

As the name suggests, the input data is fed in the forward direction through the network. An 1-input 1-output CVNN structure with an 1-neuron hidden layer CVANN model with input, hidden and output layer is given in Fig. 2.

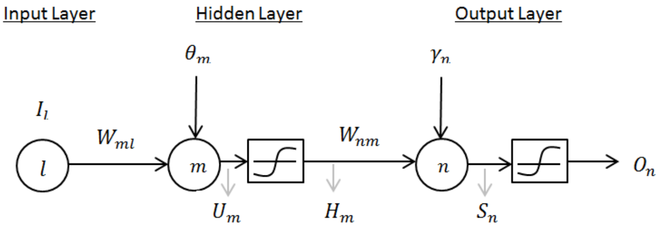

Fig. 2 Complex valued ANN model with input, hidden and output layer

The inputs (I_1) provides the initial information that then propagates to the hidden units at each layer and finally produce the output (O_n). For clarity, we will list some symbols below.

The output values of hidden layer (U_m) and output layer (S_n) can be calculated using the Eq. 2 and Eq. 3 given below where W_{ml} is the weight connecting neuron m and l; W_{nm} is the weight connecting neuron n and m; Θ_m is the threshold value of neuron m; γ_n is the threshold value of neuron n.

$$U_m = \sum_l W_{ml} I_{l+} \Theta_m \tag{2}$$

$$S_n = \sum_m V_{nm} H_{m+} \gamma_n \tag{3}$$

The resulting output values are passed through the complex activation function (f_c). The output value (H_m) for the hidden layer m neuron and the output value (O_n) for the output layer neuron n are calculated as shown in Eq. 4 and Eq. 5.

$$H_m = f_c(U_m) \tag{4}$$

$$O_n = f_c(S_n) \tag{5}$$

The output error $\delta[n]$ can be defined using the following Eq. 6.

$$\delta[n] = O_n - T_n \tag{6}$$

T_n and O_n represents the targeted output value and the output value for the output layer neuron n respectively[4].

<u>*Back-propagation in CVANN*</u>

In CVANN, just as the real valued ANN, the error is propagated from last section to the first one. The error E_p (Mean Squared Error) is calculated by Eq. 7 [4].

$$E_p = (1/2) \sum_{n=1}^{N} |T_n - O_n|^2 \tag{7}$$

Next, we define a learning rule for the CBP model described above. We can show that the weights and the thresholds should be modified according to the following Eq. 8-11 [4].

$$\Delta V_{nm} = -\varepsilon . \frac{\partial E_P}{\partial Re[V_{nm}]} - i.\varepsilon . \frac{\partial E_P}{\partial Im[V_{nm}]} \tag{8}$$

$$\Delta \gamma_n = -\varepsilon . \frac{\partial E_P}{\partial Re[\gamma_n]} - i.\varepsilon . \frac{\partial E_P}{\partial Im[\gamma_n]} \tag{9}$$

$$\Delta V_{ml} = -\varepsilon . \frac{\partial E_P}{\partial Re[W_n]} - i.\varepsilon . \frac{E_P}{\partial Im[V_{ml}]} \tag{10}$$

$$\Delta \theta_m = \varepsilon . \frac{\partial E_P}{\partial Re[\theta_m]} - i.\varepsilon . \frac{\partial E_P}{\partial Im[\theta V_m]} \tag{11}$$

B. Complex valued data

In this study, to verify the validity and applicability of the proposed network using the new activation function, we applied it to two problems: the similar XOR problem and the detection of symmetry problem.

Complex Valued XOR problem with four patterns

The input-output mapping in the XOR problem is given in Table 1.

Table 1. XOR with four patterns

Inputs		Output
X_1	X_2	Y
0	*0*	*0*
0	*1*	*1*
1	*0*	*1*
1	*1*	*0*

In order to solve the four bit XOR problem with CVANN, the input-output mapping is encoded as given in Table 2 where the real part of the output can be seen as the XOR of the input's imaginary and input's real part, and the imaginary part of the output is equal to the real part of the input [5].

Table 2. Similar XOR for CVANN

Input	Output
X_1	Y
0	*0*
i	*1*
1	*1+i*
1+i	*i*

Symmetry Detection Problem

The problem of determining symmetry aims to detect the binary activity levels of a one-dimensional array of input neurons are symmetrical about the centre point. With the increasing number of bits, the possibility of being symmetric, the event rate, decreases by half. That's why the symmetry detection problem is a very suitable problem to investigate the imbalanced data. Three input and one output symmetry detection problem is given in Table 3.

Table 3. Symmetry detection problem with 3 input 2 output

Inputs			Output
X_1	X_2	X_3	Y
0	*0*	*0*	*1*
0	*0*	*1*	*0*
0	*1*	*0*	*1*
0	*1*	*1*	*0*
1	*0*	*0*	*0*
1	*0*	*1*	*1*
1	*1*	*0*	*0*
1	*1*	*1*	*1*

When real valued problems desired to be solved with CVANN, the real valued input data need to be converted to complex valued data. This conversion can be done with sample angle-based coding. The following Eq. 12 is generally used to encode real valued data x in literature [4].

$$\varphi = \frac{\Theta(x-a)}{b-a} \tag{12}$$

With the calculated phase angle, the numbers in the real plane are moved to complex plane using the following Eq. 13

$$Z = e^{i\varphi} = cos\varphi + isin\varphi \tag{13}$$

In this study, a and b were taken as 0 and 1, respectively. The data was moved to the complex plane with a phase angle $\Theta = \pi/4$. Three input and one output symmetry detection problems in the complex plane are given Table 4.

Table 4. Complex valued symmetry detection

Inputs			Output
X_1	X_2	X_3	Y
1	*1*	*1*	*0.7+0.7i*
1	*1*	*0.7+0.7i*	*1*
1	*0.7+0.7i*	*1*	*0.7+0.7i*
1	*0.7+0.7i*	*0.7+0.7i*	*1*
0.7+0.7i	*1*	*1*	*1*
0.7+0.7i	*1*	*0.7+0.7i*	*0.7+0.7i*
0.7+0.7i	*0.7+0.7i*	*1*	*1*
0.7+0.7i	*0.7+0.7i*	*0.7+0.7i*	*0.7+0.7i*

C. Complex Swish Activation Function

Swish is a new activation function proposed by Ramachandran et al. in October 2017. According to their paper, Ramachandran et al. showed that swish tends to work better than ReLU which is the most successful and most widely used activation function in deeper models in a series of challenging datasets [6].

Swish activation function is formulized by Eq. 14.

$$y = x.sigmoid(x) = \frac{x}{1+e^{-x}} \tag{14}$$

In this study, the swish activation function has been studied in complex domain. Since the proposed CVANN will use the back propagation algorithm, the derivative of the formula is needed. The derivatives and graph of the y function is formulized by Eq. 15 and shown in Figure 3.

$$y' = y + \sigma(x)(1-y) = \frac{x.e^{-x}(x+1)+1}{(1+e^{-x})^2} \tag{15}$$

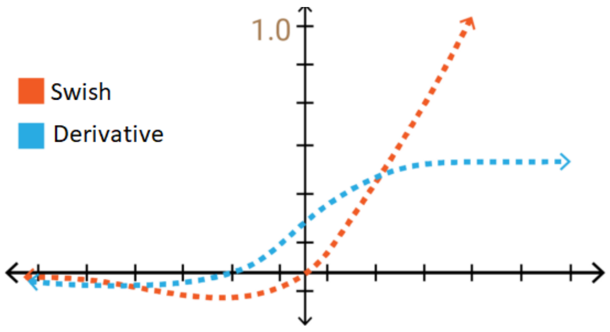

Fig. 3 Swish function and derivative

The properties of the swish function include smoothness, non-monotonic, bounded below and unbounded in the upper limits [7].

III. Results

A. Complex-Valued XOR problem with four patterns

In order to verify the validity and practicability of the proposed network using swish activation function, the similar XOR problem was used. All data used for training.

This problem has been simulated with one input, two hidden nodes in hidden layer and one output (1-2-1) CVANN to compare with other methods in the literature [8-12] for complex-valued the similar XOR problem solving. For all methods, learning rate was chosen as 0.5 and mean squared error value (MSE) was used as stopping criteria (as seen Eq.7).

When the error value (MSE) reached 0.001, the average learning epochs, target and actual output are shown in the Table 5. And the learning curve for the similar XOR problem is shown in Figure 4.

Table 5. The New CVANN Test Results for XOR Problem

Activation Function: **LOGSIG**		Activation Function: **TANSIG**		Activation Function: **SWISH**	
Iteration number with an error rate of 0.001: **3811**		Iteration number with an error rate of 0.001: **2368**		Iteration number with an error rate of 0.001: **326**	
Target	***Output***	***Target***	***Output***	***Target***	***Output***
0	*0.04+0.01i*	*0*	*0.03+0.00i*	*0*	*0.05+0.03i*
1	*0.94+0.03i*	*1*	*0.94+0.03i*	*1*	*0.97+0.02i*
1+i	*0.99+0.98i*	*1+i*	*0.98+0.99i*	*1+i*	*0.96+0.98i*
i	*0.02+0.98i*	*i*	*0.02+0.95i*	*i*	*0.02+1.00i*

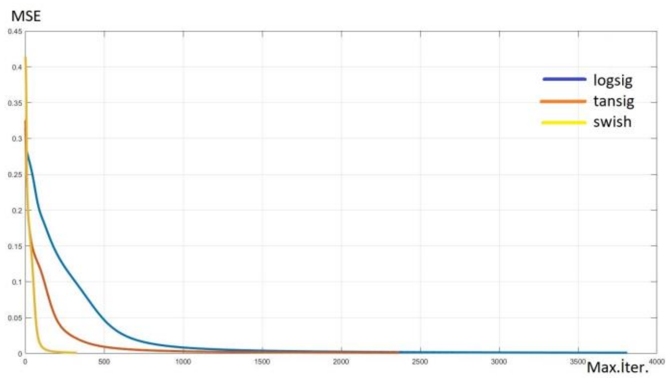

Fig. 4 The new CVANN test results for XOR problem

The proposed network using the swish activation function reached 0.001 error rate (MSE) at 375 iteration, while the other networks using logsig and tansig activation function reached at 3811 and 2368 iterations, respectively.

B. Complex-Valued Symmetry Detection Problem

We are proposed a complex valued ANN using complex valued swish activation function to the problem of the detection of symmetry.

A 3-1-1 (three input, one hidden nodes in hidden layer and one output) three-layered complex-valued ANN was used for the proposed network using swish activation function to compare with other methods in the literature [8,13] for complex-valued symmetry detection problem solving. Learning rate was chosen 0.5 and error values in Table 6 were used as stopping criteria.

The learning curve is given in Figure 5. and the average learning epochs, target and actual outputs are shown in the Table 6. when the error value (MSE) reached the stopping criteria 0.001.

Table 6. The new CVANN test results for symmetry problem

Activation Function: **LOGSIG**		Activation Function: **TANSIG**		Activation Function: **SWISH**	
Iteration number with an error rate of 0.001: **1239**		Iteration number with an error rate of 0.001: **583**		Iteration number with an error rate of 0.001: **333**	
Target	***Output***	***Target***	***Output***	***Target***	***Output***
0.7+0.7i	*0.71+0.70i*	*0.7+0.7i*	*0.72+0.71i*	*0.7+0.7i*	*0.68+0.68i*
1	*0.95+0.03i*	*1*	*0.95+0.03i*	*1*	*0.95+0.04i*
0.7+0.7i	*0.71+0.70i*	*0.7+0.7i*	*0.72+0.69i*	*0.7+0.7i*	*0.71+0.69i*
1	*0.95+0.03i*	*1*	*0.95+0.03i*	*1*	*0.96+0.04i*
1	*0.72+0.69i*	*1*	*0.95+0.03i*	*1*	*0.95+0.03i*
0.7+0.7i	*0.95+0.03i*	*0.7+0.7i*	*0.71+0.69i*	*0.7+0.7i*	*0.71+0.69i*
1	*0.72+.0.69i*	*1*	*0.95+0.03i*	*1*	*0.96+0.04i*
0.7+0.7i	*0.71+0.70i*	*0.7+0.7i*	*0.72+0.69i*	*0.7+0.7i*	*0.74+0.71i*

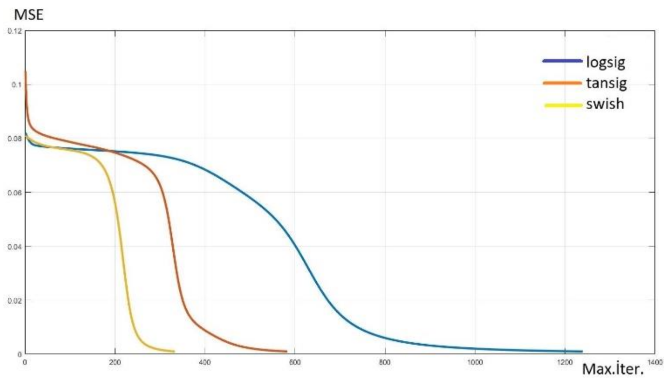

Fig. 5 The New CVANN Test Results for Symmetry Problem

A comparison of the proposed network using swish activation function and the other networks using traditional activation functions are illustrated in Figure 5, from which we can see that the proposed activation function has better stability convergence performance than the other known activation functions.

IV. DISCUSSION AND CONCLUSION

In this study, the swish activation function, which has been shown to tend to work better in real valued data sets, has been moved to the complex plane and tested with complex data. Complex swish activation function is compared with known complex activation functions (logarithmic sigmoid and tangent sigmoid) on four bit complex valued XOR problem and three input one output complex valued symmetry detection problems.

Based on the experimental results, it was seen that the CVANN using swish activation function converges to the target earlier than the other CVANN using known complex activation functions (logarithmic sigmoid and tangent sigmoid).

In future studies, modified swish activation functions, which have been shown to give better results than swish function in the real domain, can be examined on the complex domain and on more comprehensive data. These functions developed over the swish function; Modified Swish (Prajit et al., 2017) [6], E-Swish (Eric, 2018) [14], Flatten-T Swish (Hock et al., 2018) [15], Hard Swish (Avenash and Viswanath, 2018) [16].

REFERENCES

[1] M. Ceylan, Improving an algorithm with complex-valued artificial neural network and applications, MS Thesis, Selcuk University Gradute School of Natural and Applied Sciences, Konya, 13-14, 2004

[2] M. Ceylan, A new complex-valued intelligent system design on evaluating of the lung images with computerized tomography, PhD Thesis, Selcuk University Gradute School of Natural and Applied Sciences, Konya, 53, 2009

[3] R.savitha, S.Suresh, N.Sundararajan and P.Saratchandran, “Complex-Valued Function Approximation Using an Improved BP Learning Algorithm for Feed-Forward Networks”, IEEE World Congress on Computational Intelligence, 2008

[4] Y.Acar, M. Ceylan, E. Yaldız, “An examination on the effect of CVNN parameters while classifying the real-valued balanced and unbalanced data”, International Conference on Artificial Intelligence and Data Processing (IDAP), 2018

[5] M. Ceylan, “Combined complex-valued artificial neural network (CCVANN)”, Proceedings of the World Congress on Engineering. 2:955-959, 2011.

[6] P. Ramachandran, B. Zoph, “Searching for activation functions”, In International Conference on Learning Representations. Q. V. L. 2018.

[7] C. E. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, “Activation Functions: Comparison of Trends in Practice and Research for Deep Learning”, [Online]. Available: arXiv:1811.03378v1 [cs.LG] , 2018

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

[9] H. A. Jalab, R. W. Ibrahim, “New activation functions for complex-valued neural network”, International Journal of the Physical Sciences Vol. 6(7), pp. 1766-1772, 4 April, 2011

[10] T. Enomoto, K. Kakuda, S. Miura, “ New Activation Functions in CNN and Its Applications”, ICCES, vol.1, no.2, pp.36-39, 2019

[11] M. Ceylan, “Combined complex-valued artificial neural network (CCVANN)”, Proceedings of the World Congress on Engineering. 2:955-959, 2011.

[12] H. Gürüler, M. Peker, " A Software Tool for Complex-Valued Neural Network: CV-ANN", 2015, 23nd Signal Processing and Communications Applications Conference (SIU)

[13] X. Chen, Z. Tang C. Variappan, S. Li and T. Okada, “A modified error backpropagation algorithm for complex-value neural networks”, International Journal of Neural Systems, 15, 435-443, 2005.

[14] E. Alcaide. “E-swish: Adjusting activations to different network depths”, arXiv preprint arXiv:1801.07145, 2018.

[15] H. Chieng, N. Wahid, O. Pauline, S. Perla, “ Flatten-T Swish: a thresholded ReLU-Swish-like activation function for deep learning” , International Journal of Advances in Intelligent Informatics, 2018

[16] R. Avenash and P. Viswanath, “Semantic Segmentation of Satellite Images using a Modified CNN with Hard-Swish Activation Function”, VISAPP, 2019