International Journal of Innovative Computing, Information and Control
Volume **20**, Number **6**, December **2024**
ICIC International ©2024 ISSN 1349-4198
pp. 1587–1601

COMPLEX-VALUED NEURAL NETWORKS WITH SUPER-ACTIVATION FUNCTION

Sura Mahmood Abdullah[1], Ali Fadhil Yaseen Althabhawee[2]
Nadia Mohammed Ghanim Al-Saidi[3], Venus Wazeer Samawi[4]
and Alaa Kadhim Farhan[1]

[1]Department of Computer Sciences
[3]Department of Applied Sciences
University of Technology
Baghdad 10066, Iraq
{ Sura.m.abdullah; nadia.m.ghanim; alaa.k.farhan }@uotechnology.edu.iq

[2]Directorate General of Education in Holy
Karbala 56001, Iraq
61124@student.uotechnology.edu.iq

[4]Department of Management Information Systems/Smart Business
Isra University
Al Hezam Road, Amman 11622, Jordan
venus.samawi@iu.edu.jo

Received February 2024; revised June 2024

Abstract. *This research looks into the features of generalized transcendental functions of the super-Mittag-Leffler formula (SMLF) and their limited and analytic qualities. Based on the SMLF, we formulate a generalized activation function of a complex variable. More significantly, the complex-valued neural networks (CvNNs) are proven to be convergent without poles, and hence capable of generally approaching any nonlinear complex mapping to random accuracy. When the 2D-parameters space of the proposed activation function is within (0, 1), the simulation results for the benchmarking tasks showed that the proposed complex SMLF activation function can deal with complicated signals exceedingly well. This is attributed to minimizing the formula of the energy. The best results are obtained at the boundary of the open unit disk. This work validates the superior performance and robustness of the proposed SMLF activation function in CvNNs for complex signal processing tasks, offering thorough mathematical proofs and simulation results. All figures of the super-trigonometric functions are given by using MAPL software for the first time. Consequently, we presented a set of properties for our suggested new activation function.*
Keywords: Complex-valued neural network, Subordination, Majorization, Open unit disk, Transcendental function, Activation function, Super-Mittag-Leffler function, Analytic function

1. **Introduction.** Complex neural networks (CNNs) have complex weights, threshold values, inputs, and output signals. CNN was produced to open process data, including complex numbers. When consuming a recognized methodology to solve complex data problems, the real case of CNNs should be employed separately for real and imaginary fragments of the complicated data. However, when CNN is utilized to solve the same problem, data can be treated without demanding to differentiate real and imaginary portions. Consequently, the processing time is shorter, and the accuracy rate is higher. One

DOI: 10.24507/ijicic.20.06.1587

of CNN's key advantages is the capacity to work with data, which is critical for evaluating signals and solving various pattern recognition and classification issues [1-4]. In 1990, Clarke's work [5], which extended the real case of CAF into the complex-valued one, is one of the most primitive works to analyze complicated information utilizing NNs. He distinguished that spreading $\varphi(x) = \tanh(x)$ destabilizes the real-valued sigmoid function's boundedness condition. Because Liouville's theorem specifies that the only restricted holomorphic functions assembled for the complete complex domain are constant, he pointed out that escaping this exertion is unbearable. Partial investigations are presented by Kim and Adali to solve this problem [6]. Goh and Mandic [7] suggested complex value nonlinear gradient descent algorithms for reusable sign algorithms with minimal complexity and rapid convergence. Numerical preparation is given to get the convergence of the CNNs by using Newton's method of back-propagation [8]. Different approaches are given to imply extra accuracy of CNNs using different transcendental functions, which can be selected in [9].

The development of complex-valued neural networks (CvNNs) with the super-Mittag-Leffler function (SMLF) as an activation function has benefited from recent major advances in the fields of neural networks and complex computing. The integration of complex-valued neurons to improve spatial data processing has been demonstrated as an illustration of the usefulness of complex-valued networks in applications like cross-view geo-localization [10] and stability detection in smart grids [11]. Furthermore, studies on energy-efficient deep neural networks for EEG signal noise reduction emphasize how crucial it is to tailor neural network architectures to particular tasks; this is similar to the benefits of employing the SMLF activation function in CvNNs [12]. Moreover, developing phase-change photonic memories integrated with optical neural networks (ONNs) emphasizes the potential for high-speed, energy-efficient neural networks through advanced material and photonic integration [13].

The complexity of the CNN algorithm is determined by the activation function used. The following CAF behaviors must be considered while dealing with this issue.

- CAF should be bounded. Multilayer complex procedure systems have been demonstrated to generate mistakes in software and hardware applications when the activation function is not constrained.
- In both the real and imaginary components, CAF must not be linear. The multi-layer complex procedure systems will be useless otherwise.
- The CAF should be analytical everywhere. That is, the derivative exists and is bounded. Moreover, the CAF should be free of poles in its domain.
- The CAF with the symmetry property about the x- and y-axes taking the structure

  $$\varphi(z) = \varphi_1(x) + i\varphi_2(y)$$

  is known as the double real case of CAF. However, such a CAF is known as the not fully complex-valued CAF and should be ignored.

In this effort, we aim to present a new CAF based on the generalized transcendental functions of the MLF. The MLF is a generalization of the exponential function e^z, which has been recommended as a substitute for the fundamental transcendental functions of some CNNs. An investigational suggestion proposes that the entire exponential function achieves better than those with poles as a CAF. More importantly, CNNs are completely convergent and, hence, capable of reaching the random accuracy of any holomorphic function. Subordination and superordination concepts combined in a subset of advanced transcendental functions are advantageous to acquire a randomized close estimate of the desired mapping. In addition, we present sufficient conditions for the proposed CAF to satisfy all of the previous conditions of the activation function. We provide evidence that the function should be in the right of the z-plane with a positive real part.

The article is organized as follows. Section 2 deals with the methodology used in our investigation. The CNN technique is based on concepts like the supersine, supercosine, and super-Mittag-Leffler function. The central concept of boundedness is the subordination and superordination in the open unit disk. Section 3 involves the results and the structure, whereas the discussion of the results is presented in Section 4. Finally, Section 5 confirms the conclusion and what we discovered from this study.

2. **Mathematical Methods.**

The generalized MLF is considered as follows [14]:

$$P^{\epsilon}_{\gamma,\eta}(z) = \sum_{n=0}^{\infty} \left(\frac{(\epsilon)^n}{\Gamma(\gamma^n + \eta)} \right) \frac{z^n}{n!}, \quad (\gamma, \eta, \epsilon, z \in C,\ R(\gamma) > 0,\ R(\eta) > 0).$$

Moreover, the negative variable is defined as follows:

$$P^{\epsilon}_{\gamma,\eta}(-\mu z) = \sum_{n=0}^{\infty} \left(\frac{(\epsilon)^n}{\Gamma(\gamma^n + \eta)} \right) \frac{(-\mu)^n z^n}{n!}.$$

This feature describes the relaxing process (the return to a disturbed system's equilibrium) under the effect of an arbitrary and continuous signal with an initial jump. The relaxation process is a valuable strategy for reducing local ambiguity and achieving global consistency in various situations. It is essentially a parallel execution model that adjusts the confidence measures of the entities involved based on associated hypotheses and confidence measures. The relaxation technique is an error-correction algorithm that alternatively changes network parameters by comparing the response to a given input with the desired response to enhance performance. It includes the NN algorithm. The Prabhakar function is utilized for introducing super-trigonometric functions [15] via this section.

In this part, we present the primary definition of super-transcendental functions (as in Figure 1).

- The supersine is defined as follows:

$$[supersin]^{\epsilon}_{\gamma,\eta}(z^{\gamma}) = \frac{z^{\eta-1}}{2i} \left(P^{\epsilon}_{\gamma,\eta}(iz^{\gamma}) - P^{\epsilon}_{\gamma,\eta}(-iz^{\gamma}) \right).$$

- The supercosine is defined as follows:

$$[supercos]^{\epsilon}_{\gamma,\eta}(z^{\gamma}) = \frac{z^{\eta-1}}{2} \left(P^{\epsilon}_{\gamma,\eta}(iz^{\gamma}) + P^{\epsilon}_{\gamma,\eta}(-iz^{\gamma}) \right).$$

- The supertangent is defined as follows:

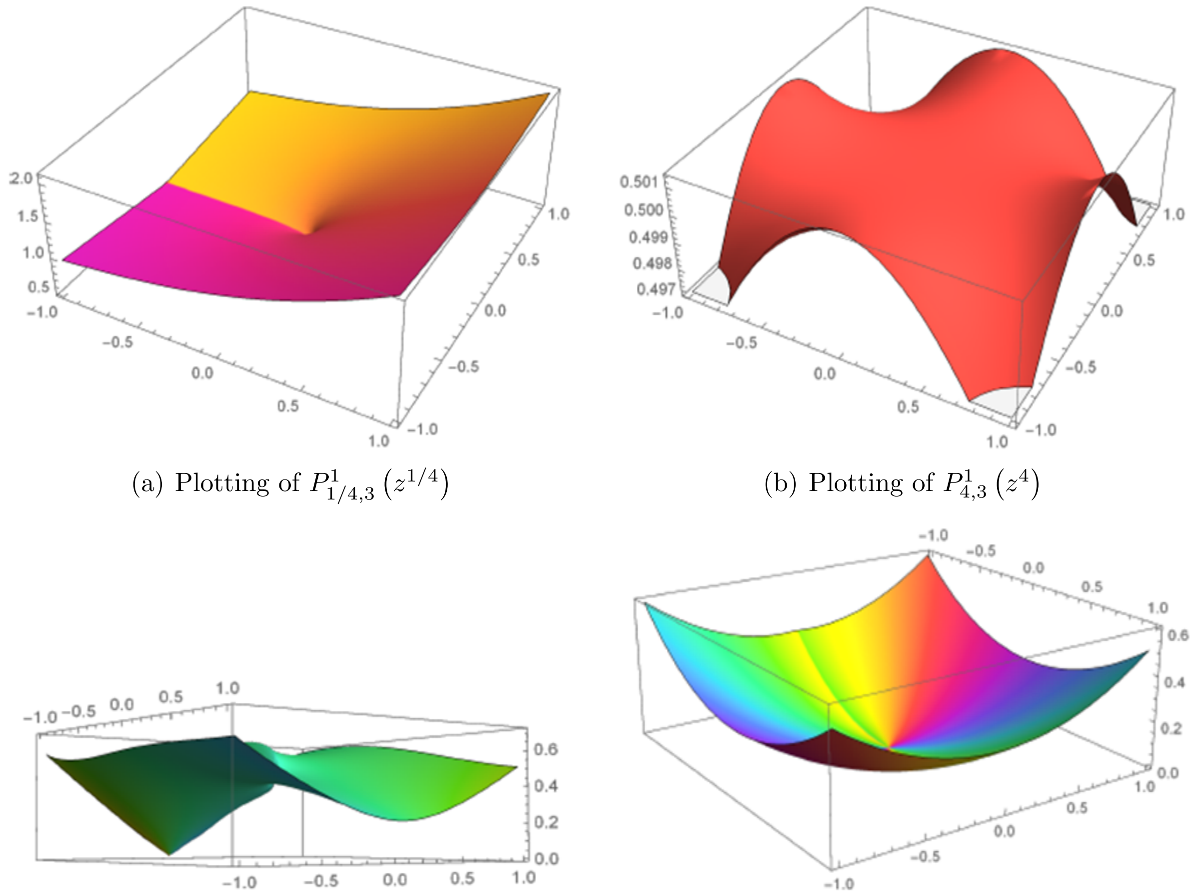
$$[supertan]^{\epsilon}_{\gamma,\eta}(z^{\gamma}) = \frac{[supersin]^{\epsilon}_{\gamma,\eta}(z^{\gamma})}{[supercos]^{\epsilon}_{\gamma,\eta}(z^{\gamma})}.$$

- The supercotangent is defined as follows:

$$[supercot]^{\epsilon}_{\gamma,\eta}(z^{\gamma}) = \frac{[supercos]^{\epsilon}_{\gamma,\eta}(z^{\gamma})}{[supersin]^{\epsilon}_{\gamma,\eta}(z^{\gamma})}.$$

- The supersecant is defined as follows:

$$[supersec]^{\epsilon}_{\gamma,\eta}(z^{\gamma}) = \frac{1}{[supercos]^{\epsilon}_{\gamma,\eta}(z^{\gamma})}.$$



(a) Plotting of $P^{1}_{1/4,3}\left(z^{1/4}\right)$ (b) Plotting of $P^{1}_{4,3}\left(z^{4}\right)$

(c) Plotting of $supersin\left(z^{1/4}\right)$ (d) Plotting of $supercos\left(z^{1/4}\right)$

FIGURE 1. Plots of some super-trigonometric functions

- The supercosecant is defined as follows:

$$[supercsc]^{\epsilon}_{\gamma,\eta}(z^{\gamma}) = \frac{1}{[supersin]^{\epsilon}_{\gamma,\eta}(z^{\gamma})}, \quad (\gamma, \eta, \epsilon, z \in C,\ R(\gamma) > 0,\ R(\eta) > 0).$$

The complex activation functions, including the hyperbolic secant, sigmoidal, and hyperbolic tangent functions, are common activation functions for CvNN.

$$\varphi(z) = \frac{1}{1+\exp(-z)},$$
$$\tanh(z) = \frac{\exp(z)-\exp(-z)}{\exp(z)+\exp(-z)},$$
$$\mathrm{sech}(z) = \frac{2}{\exp(z)+\exp(-z)}, \quad z \in C.$$

These functions are losing some analytic behaviors in the complex plane. For instance, they are no longer differentiable or restricted to close to the origin as complex value functions because their poles are close to the origin. The reality of poles in a circumscribed region near 0 is elevated once more.

We proceed to generalize the CAF by using the super-transcendental functions. Given the generalized MLF $P^{\epsilon}_{\gamma,\eta}(-z)$, we present the following sigmoidal function:

$$\varphi^{\epsilon}_{\gamma,\eta}(z) = \frac{1}{1+P^{\epsilon}_{\gamma,\eta}(-z)}, \quad z \in C. \tag{1}$$

To preserve the mathematical characteristics necessary for stable neural network operations, the generalized MLF makes sure that the activation function is analytic over the whole complex plane. The classical generalized Mittag-Leffler function's analytical characteristics result from its being the solution to the most straightforward fractional differential equation regulating relaxation processes.

The Mittag-Leffler function (MLF) is defined as

$$E_{\alpha,\beta}(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma\left(\alpha^k + \beta\right)}$$

where α and β are parameters, z is a complex variable, and Γ is the Gamma function.

To increase the flexibility of this function, additional parameters (γ and δ) are added, such as

$$E_{\alpha,\beta,\gamma,\delta}(z) = \sum_{k=0}^{\infty} \frac{\left(\gamma^k + \delta\right) z^k}{\Gamma\left(\alpha^k + \beta\right)}$$

The parameters (γ and δ) are added to make the activation function more precisely to fit particular data types or neural network architectures. This flexibility can help it perform better on a variety of tasks.

Demonstrating that the series converges for all values of z inside the unit disk $|z| < 1$ is satisfied by determining the conditions that ensure the generalized MLF has no singularity and is analytic.

The activation function derived from the generalized MLF is bounded, which helps avoid training-related issues such as exploding gradients that can occur with unbounded activation functions. This activation function maximizes the network's efficiency by minimizing the energy formula, thus reducing the amount of computing power needed for both training and inference.

The principal applications of the Mittag-Leffler function, which led us in the past to refer to it as the queen function of fractional calculus, are addressed to the reader through the various sections of the text. We expect this function will get more recognition in complex systems science over the coming years, including the ANN, CvNN (shown in Figure 2) generally, and the fractional NN specifically. Our study is based on the activation of sigmoidal function $\varphi^{\epsilon}_{\gamma,\eta}(z)$. In this study, to avoid its pole at 0, we shall use the theory of subordination and superordination.

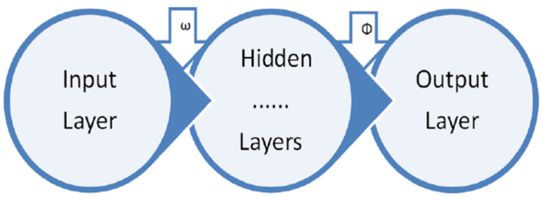


Figure 2. The CvNN architecture

The description of the proposed complex-valued neural network (CvNN) structure is as follows.

For the input layer, there are n nodes, where n represents the number of features in the input data. These data should be normalized to the values between 0 and 1 before being processed in the network.

For the second layer, there are four hidden layers, in each layer there are 4 neurons in each hidden layer. There are 100 neurons in each hidden layer for Handwriting Recognition (MNIST), and the super-Mittag-Leffler function (SMLF) is used as an activation function.

Finally in the output layer, there are two types of recognition, logical and handwriting. In the logical, a single neuron output, and in the handwriting, many neurons are in line with the number of classes (10 in the case of MNIST).

Mathematically, these points can be expressed as the following.

Let Δ be the open unit disk. Two holomorphic functions ϕ, ψ in Δ are subordinated.

$$\phi \prec \psi \text{ or } \phi(z) \prec \psi(z), \quad z \in \Delta$$

if a Schwarz holomorphic function σ, $|\sigma| \leq |z| < 1$ achieves the following equation:

$$\phi(z) = \psi(\sigma(z)), \quad z \in \Delta, \quad \phi(0) = \psi(0).$$

Moreover, ϕ is majorized by ψ ($\phi \ll \psi$) if achieved.

$$\phi(z) = \sigma(z)\psi(z), \quad z \in \Delta.$$

In other words, the coefficient inequality is maintained $|\phi_n| \leq |\psi_n|$, respectively.

Given CvNN, the function ϕ indicates the outcome, σ presents the weights propagation function, and ψ is the activation function. As a special case of the suggested CAF,

$$\varphi^1_{\gamma,\eta}(z) = \frac{1}{\frac{1}{\Gamma(\eta)}+1} + \frac{z}{\left(\frac{1}{\Gamma(\eta)}+1\right)^2 \Gamma(\gamma+\eta)} + \frac{z^2\left(\frac{1}{\left(\frac{1}{\Gamma(\eta)}+1\right)^2\Gamma(\gamma+\eta)^2} - \frac{1}{\left(\frac{1}{\Gamma(\eta)}+1\right)\Gamma(2\gamma+\eta)}\right)}{\frac{1}{\Gamma(\eta)}+1}$$
$$+ \frac{z^3\left(\frac{1}{\left(\frac{1}{\Gamma(\eta)}+1\right)\Gamma(3\gamma+\eta)} + \frac{1}{\left(\frac{1}{\Gamma(\eta)}+1\right)^3\Gamma(\gamma+\eta)^3} - \frac{2}{\left(\frac{1}{\Gamma(\eta)}+1\right)^2\Gamma(\gamma+\eta)\Gamma(2\gamma+\eta)}\right)}{\frac{1}{\Gamma(\eta)}+1} + O\left(z^4\right). \quad (2)$$

which is equivalent to

$$\begin{aligned}
\Phi_{\gamma,\eta}(z) &= \left(\frac{1}{\Gamma(\eta)}+1\right)\varphi^1_{\gamma,\eta}(z) \\
&= 1 + \frac{z}{\left(\frac{1}{\Gamma(\eta)}+1\right)\Gamma(\gamma+\eta)} \\
&\quad + z^2\left(\frac{1}{\left(\frac{1}{\Gamma(\eta)}+1\right)^2\Gamma(\gamma+\eta)^2} - \frac{1}{\left(\frac{1}{\Gamma(\eta)}+1\right)\Gamma(2\gamma+\eta)}\right) \\
&\quad + z^3\left(\frac{1}{\left(\frac{1}{\Gamma(\eta)}+1\right)\Gamma(3\gamma+\eta)} + \frac{1}{\left(\frac{1}{\Gamma(\eta)}+1\right)^3\Gamma(\gamma+\eta)^3}\right. \\
&\quad \left. - \frac{2}{\left(\frac{1}{\Gamma(\eta)}+1\right)^2\Gamma(\gamma+\eta)\Gamma(2\gamma+\eta)}\right) + O\left(z^4\right) \\
&= 1 + \varphi_1 z + \varphi_2 z^2 + \varphi_3 z^3 + O\left(z^4\right).
\end{aligned}$$

Clearly, $\Phi_{\gamma,\eta}(0) = 1$.

For example,

$$\Phi_{0.1,0.4}(z) = 1 + 0.3888z - 0.3116z^2 - z^3 + O\left(z^4\right).$$

Note that

$$\Phi_{\gamma,\eta}(z) \in P = \{p \in \Delta : p(z) = 1 + p_1 z + p_2 z^2 + \cdots R(z) > 0\}.$$

However, this class is not meromorphic. Therefore, $\Phi_{\gamma,\eta}(z)$ has no poles in the open unit disk [16].

In our discussion, we shall use the normalized CAF $\Phi_{\gamma,\eta}(z)$, which satisfies all the properties of CAF in the next proposition.

Proposition 2.1. *For some γ and η in $(0, 1)$, if $R(\Phi_{\gamma,\eta}(z)) > 1/2$, then $\Phi_{\gamma,\eta}(z)$ is bounded as well as its first derivative $[\Phi_{\gamma,\eta}(z)]'$ in the open unit disk Δ.*

Proof: Let $R(\Phi_{\gamma,\eta}(z)) > 1/2$. Then, in view of [17] (P.184), we have

$$(\Phi_{\gamma,\eta}(z)) \prec \frac{1}{1-z}, \quad z \in \Delta.$$

That is $(\Phi_{\gamma,\eta}(z)) \in S = \left\{f(z) : f(z) \prec \frac{1}{1-z}\right\}$. According to [18] (P.16), we obtain that the set of extreme points of the closed convex hull of S^v, where $S^v = \{f^v : f \in S\}$, can be defined as follows:

$$\overline{excon}(S)^v_\alpha = \left\{f(z) : f(z) = \left(\frac{1}{1-\alpha z}\right)^v, |\alpha| = 1\right\}.$$

In conclusion, we realize that $\Phi_{\gamma,\eta}(z)$ with $\alpha = v = 1$ lies in the set $\overline{excon}(S)$ for all $z \in \Delta$. The boundedness is directly obtained $\Phi_{\gamma,\eta}(z)$ by [19] (Corollary 8, P.348) in the disk.

$$|z| < \sqrt{2} - 1 \approx 0.414213$$
$$|(\Phi_{\gamma,\eta}(z))| \leq \left|\frac{1}{1-z}\right|, \quad z \in \Delta, \quad |z| < \sqrt{2} - 1$$

In addition, by the convexity of the function, $1/(1-z)$ we obtain [20] (Theorem 3), for some $|z| < \sqrt{2} - 1$

$$[(\Phi_{\gamma,\eta}(z))]' \ll \left[\frac{1}{(1-z)}\right]', \quad z \in \Delta.$$

This leads to

$$|[(\Phi_{\gamma,\eta}(z))]'| \leq \left|\left[\frac{1}{(1-z)}\right]'\right|.$$

That is, the derivative is bounded. This completes the proof.

In conclusion, the absence of poles and the function's boundedness guarantee a more reliable convergence of the neural network during training, which is crucial for complex-valued neural networks handling intricate signal-processing tasks.

Remark 2.1. *Proposition 2.1 indicates some facts regarding the suggested CAF $\Phi_{\gamma,\eta}(z)$. The result is given for a special type of analytic function in the open unit disk (normalized analytic functions $\Phi_{\gamma,\eta}(0) = 1$). Interpolation, a form of estimate, is the process of constructing (finding) new data points depending on the range of a set or domain of known data points. In the open unit disk, this class of analytical functions displays a geometric interpolation. This means that the suggested CAF $\Phi_{\gamma,\eta}(z)$ can be recognized geometrically. Moreover, the above class of analytic functions implies the univalency property (injective function). The condition $R(\Phi_{\gamma,\eta}(z)) > 1/2$ (the right half plane) is necessary to obtain the boundedness of the CAF and its first derivative, where the boundedness is the essential condition to illustrate any activation function. In our effort, this condition yields the CAF to be in the right half plane.*

The structure of CvNN is given in Figure 2. The layers of CvNN are as follows:

- The input layer contains n-nodes $\{z_1, \ldots, z_n\}$, where $|z_n| < 1$ (normalize (z_j) if not)

- Hidden layers contain h-layers:
  $\{z_1\omega_1, \ldots, z_n\omega_n\}$, where $\omega = \{\omega_1, \ldots, \omega_n\}$ is the set of vector weights.
- The following equality gives the output layer:

$$\gamma = \Phi\left(\sum_{n=1}^{N} \omega_n z^n\right), \quad |z| < 1, \tag{3}$$

  where,

$$\sum_{n=1}^{N} \omega_n z^n = \sum_{net=1}^{N=\max net} \omega_{net} z^{net}$$

  is the propagation function and Φ satisfying Equation (3) is the activation function. We utilize zero bias and look at complex-valued neurons on the open unit disk in our setup. The neuron is accomplished by minimizing the sum-of-squares error function Ξ on a preparation set covering N data points.

The output must be an element in the set [21].

$$Z = \left\{i = \pm\sqrt{-1}, \pm 1\right\},$$

with minimum energy on the total information net using the suggested CAF $\Phi_{\gamma,\eta}(z)$

$$Energy = \frac{\overline{\Phi_{\gamma,\eta}(z)}\Phi_{\gamma,\eta}(z)}{n}, \quad z \in \Delta,$$

where n indicates the number of neurons.

3. **Experimental Results.** In order to assess the effectiveness of the proposed activation function, we will use the backpropagation (BP) method as the training algorithm. In neural network applications, BP is the most commonly used supervised learning algorithm, where the activation function plays a crucial role in stabilizing the training process and mapping the input values to the appropriate output in the output layer. A variety of dynamic functions can be used as activation functions. We chose CAF to represent the output signal. The proposed CvNN will be evaluated based on its performance in solving two problems.

- The implementation of fundamental two-bit logic gates using a single complex neuron. Notably, this includes the XOR gate.
- The recognition of handwriting digits. The benchmarks are based on the MNIST dataset.

3.1. **Logic gate recognition.** In this experiment, we will utilize the consequences of gradient descent training in the exclusive-or (XOR) issue using the suggested super-Mittag-Leffler function (SMLF) as an activation function. The XOR issue is commonly used in the literature as a test case for backpropagation algorithms. Table 1 shows the training set for the XOR issue. The goal of the XOR issue is to check the weights of an NN such

TABLE 1. XOR testing collection

Boolean relation XOR	Input		Output
	X	Y	Z
	0	0	0
	0	1	1
	1	0	1
	1	1	0

that the output unit turns ON if one of the inputs is ON but not together. The proposed CAF (SMLF), which we employed in trailing, is given by Equation (2). The SMLF is assumed to be the same for the network's hidden and output layers. Algorithm 1 shows the steps of the sophisticated BP algorithm.

In this application, we employ a two-layer network with $m = 2$ input nodes, $K = 4$ hidden neurons, and $C = 1$ output neurons. A two-layered CvNN with 2^m hidden neurons can be trained from any Boolean function with variables. The gradient descent backpropagation algorithm was used to train the network. The network was trained within 0.0001 error in each example. The proposed activation function SMLF and backpropagation algorithm underwent 2000 trials, with a similar collection of unspecified foremost weights $\left(W = \left\{w : \pm\frac{i}{4}, \pm\frac{1}{4}\right\}\right)$ utilized for each collection of trials using the gradient descent back-propagation technique and a constant learning rate (μ). For CvNNs, back-propagation is a prominent approach that we used. We calculated the value of a specified error function by comparing the output numbers to the correct response.

The algorithm used this information to alter the weights of every link to minimize the error function's value by a modest amount. The network converged to a state where the computation error was modest after a few repetitions, a method for a sufficient number of trailing periods. As a result, the network has taught a specific target function. To correctly change weights, we used gradient descent, a general method for non-linear optimization. The error function's derivative corresponding to the net weights is computed, and the weights are then modified to reduce the error. The following are the steps in the BP training algorithm: Algorithm 1

1) Use small, complicated random numbers to start the complex weights.
2) Show the desired output vector and the input vector.
3) Determine the net-input values to the hidden layer units and outputs.
4) Determine the net-input values, as well as the outputs, for each output layer unit.
5) Determine the output units' error terms.
6) Change the weights on the output and hidden layers.

Table 2 indicates the outcomes of the simulations. The proposed activation function based on the SMLF and the BP training technique performed admirably in all tested input patterns.

TABLE 2. The simulation results

Simulation results: Boolean relation XOR	Input		Output
	X	Y	Z
	0	0	$\mathbf{0.9749 \approx 1}$
	0	1	$\mathbf{0.9523 \approx 1}$
	1	0	$\mathbf{-0.9865 \approx -1}$
	1	1	$\mathbf{-0.9501 \approx -1}$

Figure 3 illustrates the simulation results. The output values are in the set $Z = \{\pm 1, \pm i\}$, with the interval error of $[0.0135, 0.0499]$.

To discuss the total information, a computation implies that

$$\Phi_{0.1,0.4}(1) = 1 + 0.3888 - 0.3116 - 1 = 0.0772$$
$$\Phi_{0.1,0.4}(-1) = 1 - 0.3888 - 0.3116 + 1 = 1.2966$$
$$\Phi_{0.1,0.4}(i) = 1 + 0.3888 * i - 0.3116 i^2 - i^3 = 1.3 + 1.3i$$
$$\Phi_{0.1,0.4}(-i) = 1 + 0.3888(-i) - 0.3116 i^2 - (-i)^3 = 1.3 - 1.3i.$$

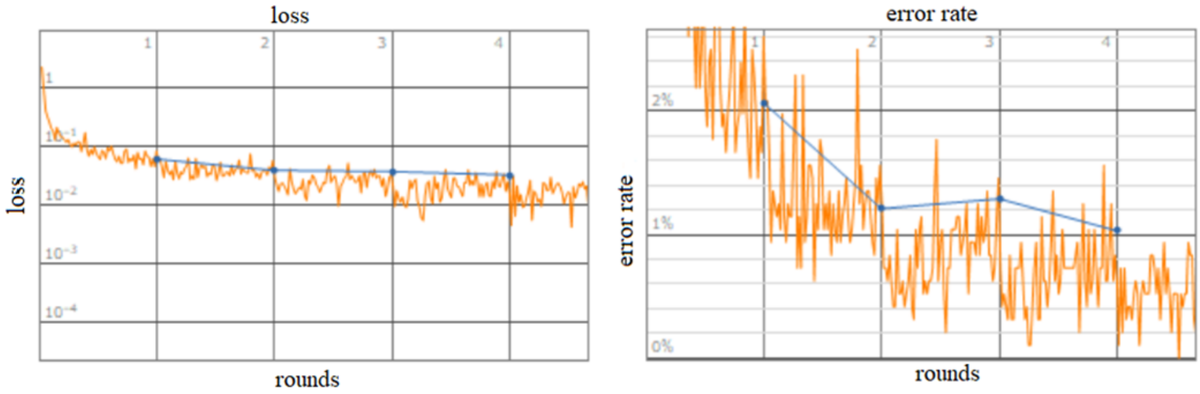


FIGURE 3. The training of CvNN using four neurons with its error rate

Estimate the energy of the net with $n = 4$, we have

$$Energy(1) = \frac{0.0772 \times 0.0772}{4} = 0.00148996$$
$$Energy(-1) = \frac{1.2966 \times 1.2966}{4} = 0.42029289$$
$$Energy(i) = \frac{\overline{1.3 + 1.3i} \times 1.3 + 1.3i}{4} = 0.845$$
$$Energy(-i) = \frac{\overline{1.3 - 1.3i} \times 1.3 - 1.3i}{4} = 0.845.$$

Hence, the total energy is

$$Energy = 2.1 > 1.$$

This means that the system needs more training. Now by consume

$$\Phi_{0.001,0.999}(1) = 1 + 0.5 = 1.5$$
$$\Phi_{0.001,0.999}(-1) = 1 - 0.5 = 0.5$$
$$\Phi_{0.001,0.999}(i) = 1 + 0.5 * i$$
$$\Phi_{0.001,0.999}(-i) = 1 - 0.5 * i.$$

Then, we have

$$Energy(1) = \frac{1.35 \times 1.35}{4} = 0.445$$
$$Energy(-1) = \frac{0.5 \times 0.5}{4} = 0.0625$$
$$Energy(i) = \frac{\overline{1 + 0.35i} \times 1 + 0.35i}{4} = 0.27$$
$$Energy(-i) = \frac{\overline{1 - 0.35i} \times 1 - 0.35i}{4} = 0.27.$$

This implies that $Energy \cong 1$; thus, the best total energy is on the boundary of the unit disk $\partial\Delta$. In conclusion, after training the system, we explore that the best energy is held when decreasing the parameter γ and increasing the parameter η. Experiments have demonstrated that the output on $\partial\Delta$ requires the least amount of energy for a CvNN with four neurons.

3.2. **Handwriting recognition by the proposed CvNN.** Automated handwriting recognition has achieved significant real-world success in targeted applications such as address recognition on mail pieces for sorting automation and reading of courtesy and legal amounts on bank checks [22]. In [23], a novel mathematical model for partially diffusive equations with boundary coupling can be used to create a new CvNN. In this work, the MNIST dataset is used to train the CvNN [24]. It contains 70000 handwritten samples of digits. Out of these, 60000 samples are used for training, while 10000 are used for testing. Each sample is presented as a 28×28-pixel image, and each pixel is considered a distinct feature. Consequently, there are 784 prediction features representing the number of input nodes. The output layer comprises ten nodes corresponding to the (0-9) digits. Moreover, the network includes four hidden layers, each with 100 neurons. The suggested SMLF is the activation function in the four layers, with a learning rate (η) of 0.01. The handwriting recognition task by the proposed CvNN achieved a recognition accuracy of 97%.

4. **Assessment of Experimental Results.** To assess and evaluate the performance of the proposed CvNN with SMLF (as an activation function), a comparison with various activation functions, including the super-Mittag-Leffler function (SMLF), Sigmoid, Hyperbolic Tangent (Tanh), Sigmoid, Rectified Linear Unit (ReLU), and Complex Exponential Linear Unit (CELU) with SMLF need to be accomplished. This comparison will be carried out in the context of handwriting recognition using the MNIST dataset. The comparison is based on accuracy, convergence rate, computational complexity, and energy efficiency. The results presented in Table 3 indicate that SMLF offers superior accuracy, especially in complex signal-processing tasks such as handwriting recognition. Regarding the convergence rate metric, SMLF showed fast convergence due to a minimized energy formula and stable learning. ReLU has fast convergence but is susceptible to dead neurons. CELU handles complex values better than ReLU. The Tanh function struggles with saturation at high values, and Sigmoid has issues with vanishing gradients in deep networks. In terms of complexity, although the ReLU, Sigmoid, and Tanh activation functions have low complexity, each has limitations. ReLU and Sigmoid have limitations concerning gradient issues, while Tanh showed limited effectiveness in capturing complex patterns. On the other hand, CELU and SMLF have moderate complexity. CELU balances complexity and performance for complex-valued inputs, while SMLF involves calculating transcendental functions but benefits from efficient convergence. Concerning energy, SMLF and ReLU are highly efficient. SMLF reduces overall energy for training due to boundedness and analyticity, while ReLU simplicity and rapid convergence minimize energy consumption. Although both SMLF and CELU are designed to manage complex values, CELU

Table 3. The performance metrics of various activation functions for digit recognition using the MNIST dataset

Activation functions	Performance metrics			
	Accuracy (%)	***Convergence rate***	***Computational complexity***	***Energy efficiency***
Tanh	92	Moderate	Low	Moderate
Sigmoid	90	Slow	Low	Low
ReLU	95	Very Fast	Very Low	High
CELU	93	Good	Moderate	Moderate
SMLF (suggested)	97	Fast	Moderate	High

is less energy efficient than SMLF. On the other hand, the Tanh and Sigmoid functions consume more energy due to slow convergence. Moreover, vanishing gradients increase energy consumption in the Sigmoid function. Based on these findings, it can be concluded that the SMLF activation function enhances the performance of complex-valued neural networks, making it a valuable contribution to the field.

The comparison results of the proposed CvNN with existing studies for handwriting recognition tasks using the MNIST database are illustrated in Table 4. In this comparison, the author in [25] applied CvNN trained by gradient descent algorithm for recognition tasks using the MNIST dataset. CvNN has 5 hidden layers, and the complex-valued error was 0.48%. Scardapane et al. [1] proposed a CvNN with non-parametric activation functions. The model achieved 95% accuracy on the MNIST dataset using three hidden layers, each with 100 neurons. Mönning and Manandhar [26] compared the performance of complex and real-valued multi-layer perceptron in benchmark recognition tasks. The model achieved an accuracy of 86% in MNIST digit recognition tasks with 4 hidden layers. The CvNN model proposed by Zhang in 2021 [27] achieved 90% accuracy in MNIST digit recognition tasks with 4 hidden layers. The proposed CvNN architecture achieved a recognition accuracy of 97% in MNIST digit recognition by using only 4 hidden layers, each with 100 neurons. This makes it a simple yet highly effective architecture for handwriting recognition. Our study shows that the super-Mittag-Leffler activation function, which has the benefits of moderate complexity and quick learning, can be effectively applied in CvNN. This makes it a straightforward but incredibly powerful architecture for handwriting recognition.and high efficiency in terms of energy.

TABLE 4. The accuracy of recognizing digits using the MNIST dataset

MNIST digit classification	**Comparison with previous studies: # of hidden layers and classification accuracy**		
	Studies	***Hidden layers***	***Accuracy (%)***
	[25]	5	52
	[1]	3	95
	[26]	4	86
	[27]	4	90
	Proposal CvNN	4	97

In conclusion, despite some drawbacks, complex-valued neural networks (CvNNs) show promise for research and application because they can handle complex data, improve convergence, and increase computational efficiency. The performance benefits demonstrated in various scenarios emphasize the importance of complex-valued activation functions, such as the SMLF, in advancing neural network capabilities.

5. **Conclusion.** This paper proposes a new activation function for complex-valued neural networks (CvNNs): the super-Mittag-Leffler function (SMLF). Compared to conventional activation functions, the SMLF-based activation function exhibits notable accuracy, convergence rate, computational efficiency, and energy efficiency improvements. The super-Mittag-Leffler function is used to suggest a new activation function. The proposed complex activation function for the net of a complex node output value was applied to the BP method to train CvNN. The simulation results for the benchmarking tasks demonstrated that the suggested complex SMLF activation function can cope with complex signals extremely well if γ and η are in $(0, 1)$ of the proposed activation function. Consequently, the complex amounts of the outputs are represented in $\partial\Delta$. The parameter

derived from the fractional CvNN is the simplest non-trivial perturbation of any unperturbed complex structure, the stable complex structure (i.e., CvNN) in which evident, necessary, and sufficient hypotheses for some different problems are recognized. All figures of the super-trigonometric functions are given by using MAPL software for the first time. Consequently, we presented a figure for our suggested new activation function.

Future research should concentrate on refining the SMLF's parameters, incorporating it with sophisticated neural network architectures, and investigating its potential applications in developing fields like bioinformatics and quantum computing.

Acknowledgement. We wish to acknowledge [University of Technology, Iraq] and [Isra University, Jordan] for providing the necessary resources and academic environment that facilitated this research.

REFERENCES

[1] S. Scardapane, S. Van Vaerenbergh, A. Hussain and A. Uncini, Complex-valued neural networks with nonparametric activation functions, *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol.4, no.2, pp.140-150, DOI: 10.1109/TETCI.2018.2872600, 2020.

[2] R. J. Al-Azawi, N. M. G. Al-Saidi, H. A. Jalab, H. Kahtan and R. W. Ibrahim, Efficient classification of COVID-19 CT scans by using q-transform model for feature extraction, *PeerJ Computer Science*, 7:e553, DOI: 10.7717/peerj-cs.553, 2021.

[3] N. M. G. Al-Saidi, H. Yahya and S. J. Obaiys, Discrete dynamic model of a disease-causing organism caused by 2D-quantum tsallis entropy, *Symmetry*, vol.14, no.8, 1677, DOI:10.3390/sym14081677, 2022.

[4] V. W. Samawi and O. A. A. Basheer, The effect of features reduction on different texture classifiers, *The 3rd IEEE Conference on Industrial Electronics and Applications*, Singapore, pp.67-72, DOI: 10.1109/ICIEA.2008.4582482, 2008.

[5] T. L. Clarke, Generalization of neural networks to the complex plane, *International Joint Conference on Neural Networks (IJCNN)*, San Diego, CA, USA, vol.2, pp.435-440, DOI: 10.1109/IJCNN.1990.137751, 1990.

[6] T. Kim and T. Adali, Approximation by fully complex MLP using elementary transcendental activation functions, *Neural Networks for Signal Processing XI: Proceedings of the 2001 IEEE Signal Processing Society Workshop (IEEE Cat. No.01TH8584)*, North Falmouth, MA, USA, pp.203-212, DOI: 10.1109/NNSP.2001.943125, 2001.

[7] S. L. Goh and D. P. Mandic, A class of low complexity and fast converging algorithms for complex-valued neural networks, *Proc. of the 2004 14th IEEE Signal Processing Society Workshop Machine Learning for Signal Processing*, Sao Luis, Brazil, pp.13-22, DOI: 10.1109/MLSP.2004.1422955, 2004.

[8] Y. L. Xiao, S. Li, G. Situ and Z. You, Optical random phase dropout in a diffractive deep neural network, *Optics Letters*, vol.46, no.20, pp.5260-5263, DOI: 10.1364/OL.428761, 2021.

[9] B. N. Örnek, S. B. Aydemir, T. Düzenli and B. Özak, Some remarks on activation function design in complex extreme learning using Schwarz Lemma, *Neurocomputing*, vol.492, pp.23-33, DOI: 10.1016/j.neucom.2022.04.010, 2022.

[10] L. Liu and H. Li, Lending orientation to neural networks for cross-view geo-localization, *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, pp.5617-5626, DOI: 10.1109/CVPR.2019.00577, 2019.

[11] S. A. Yousif, V. W. Samawi and N. M. G. Al-Saidi, Automatic machine learning classification algorithms for stability detection of smart grid, *2022 IEEE 5th International Conference on Big Data and Artificial Intelligence (BDAI)*, Fuzhou, China, pp.34-39, DOI: 10.1109/BDAI56143.2022.9862710, 2022.

[12] A. Kumar, S. Chakravarthy and A. Nanthaamornphong, Energy-efficient deep neural networks for EEG signal noise reduction in next-generation green wireless networks and industrial IoT applications, *Symmetry*, vol.15, 2129, DOI: 10.3390/sym15122129, 2023.

[13] M. Wei et al., Electrically programmable phase-change photonic memory for optical neural networks with nanoseconds in situ training capability, *Advanced Photonics*, vol.5, no.4, 046004, 2023.

[14] T. R. Prabhakar, A singular integral equation with a generalized Mittag-Leffler function in the kernel, *Yokohama Mathematical Journal*, vol.19, pp.7-15, 1971.

[15] X.-J. Yang, *Theory and Applications of Special Functions for Scientists and Engineers*, Springer Nature, New York, USA, 2021.
[16] S. S. Miller and P. T. Mocanu, *Differential Subordinations: Theory and Applications*, Marcel Dekker, 2000.
[17] O. Altintas and S. Owa, Majorizations and quasi-subordinations for certain analytic functions, *Proceedings of Japan Acad.*, vol.68, no.7, pp.181-185, DOI: 10.3792/pjaa.68.181, 1992.
[18] G. Schober, *Univalent Functions – Selected Topics*, Springer Berlin, Heidelberg, 1975.
[19] A. Szynal and J. Szynal, On some problems concerning subordination and majorization of functions, *Demonstratio Mathematica*, vol.11, no.2, pp.331-350, DOI: 10.1515/dema-1978-0206, 1978.
[20] D. M. Campbell, Majorization-subordination theorems for locally univalent functions, III, *Transactions of the American Mathematical Society*, vol.198, pp.297-306, DOI: 10.2307/1996761, 1974.
[21] K. Kreutz-Delgado, The complex gradient operator and the CR-calculus, *arXiv Preprint*, DOI: 10.48550/arXiv.0906.4835, 2009.
[22] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, vol.86, no.11, pp.2278-2324, DOI: 10.1109/5.726791, 1998.
[23] C. Phan, L. Skrzypek and Y. You, Dynamics and synchronization of complex neural networks with boundary coupling, *Analysis and Mathematical Physics*, vol.12, no.33, DOI: 10.1007/s13324-021-00613-1, 2022.
[24] L. Deng, The MNIST database of handwritten digit images for machine learning research, *IEEE Signal Processing Magazine*, vol.29, no.6, pp.141-142, 2012.
[25] C. A. Popa, Complex-valued convolutional neural networks for real-valued image classification, *International Joint Conference on Neural Networks (IJCNN)*, Anchorage, AK, USA, pp.816-822, DOI: 10.1109/IJCNN.2017.7965936, 2017.
[26] N. Mönning and S. Manandhar, Evaluation of complex-valued neural networks on real-valued classification tasks, *arXiv Preprint*, arXiv: 1811.12351, 2018.
[27] H. Zhang et al., An optical neural chip for implementing complex-valued neural network, *Nature Communications*, vol.12, no.1, 457, DOI: 10.1038/s41467-020-20719-7, 2021.

Author Biography

Sura Mahmood Abdullah is an assistant professor at the Department of Computer Sciences, University of Technology, Baghdad, Iraq. She completed her Bachelor of Software in Computer Science in 2005-2006 from the University of Technology, and her Master's degree in Science in Software Engineering in 2013 from the Institute of Informatics of Higher Studies in the Iraqi Commission for Computers and Informatics. Currently, she is teaching artificial intelligence subjects to students of preliminary studies, so she is very interested in AI topics.

Ali Fadhil Yaseen Althabhawee completed B.Sc. degree in Computer Engineering at Alhussain University College, Karbala, Iraq. He finished his M.Sc. degree in Computer Engineering at University of Technology, Baghdad, Iraq. His research interests are artificial intelligence systems, deep learning and machine learning and programming languages. He is currently working as a computer engineer at the Directorate General of Education in Holy, Karbala Province, Iraq.

Nadia Mohammed Ghanim Al-Saidi is a professor in the Department of Applied Sciences, University of Technology, Baghdad, Iraq since 2011. She completed her Bachelor of Science and Master of Science degrees in Applied Mathematics, from the Department of Applied Sciences, University of Technology, Baghdad, Iraq, in 1989, and 1995, respectively. She received her Ph.D. degree in Mathematics and Computer Applications Sciences from Al-Nahrain University, Baghdad, Iraq in 2003. In 1989, she joined the Department of Applied Sciences at University of Technology as an academic staff member. She also joined the Institute for Mathematical Research (INSPEM), University Putra Malaysia (UPM) as a post-doctorate researcher from 2008-2010 with the research project "Fractals in Cryptography". She is the author of numerous technical papers since 1994, and her research interests include cryptography, fractal geometry, and chaos theory.

Venus Wazeer Samawi is a professor at Isra University, Department of Management Information Systems/Smart Business (master program). She became a Member of the International Association of Engineers (IAENG). She received her B.Sc. degree from the University of Technology in 1987 and her M.Sc. and Ph.D. degrees from the Computer Science Department at Al-Nahrain University in 1992 and 1999, respectively. She supervises many Ph.D. and M.Sc. theses concerning system programming, pattern recognition, network security, and text classification. She also leads and teaches modules at the B.Sc. and M.Sc. computer science levels. She has been a reviewer at several conferences and journals. Her special research areas are pattern recognition, evolutionary computing, image processing, and natural language processing. Lately, Professor Samawi's primary research interest is in IoT.

Alaa Kadhim Farhan is a professor in the Department of Computer Sciences, University of Technology, Baghdad, Iraq. He completed his Bachelor of Computer Science and Master of Science degrees in Information Security from the Department of Computer Sciences, University of Technology, Baghdad, Iraq, in 2003 and 2005. He received his Ph.D. degree in Information Security from the University of Technology, Baghdad, Iraq, in 2009. In 2005, he joined the Department of Computer Sciences, University of Technology, as an academic staff member. He has been the author of numerous technical papers since 2008. His research interests include cyber security, programming languages, chaos theory, and cloud computing.