International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8477
***Article Received:** 28 July 2023 **Revised:** 16 September 2023 **Accepted:** 01 October 2023*

RAU: Novel Activation Function for Deep Learning Neural Network

Prajesha T M[1], S Veni[2]
[1]Department of Computer Science
Karpagam Academy of Higher Education
Coimbatore,Thamilnadu
tmprajesha@gmail.com
[2]Department of Computer Science
Karpagam Academy of Higher Education
Coimbatore,Thamilnadu
venikarthik04@gmail.com

Abstract—In Deep learning neural networks (DNNs) activation functions perform a vital role. In each neuron activation function is responsible for generating output signals from given input signals. Hence, activation function is one of the factors that influence the performance of DNN. A novel activation unit RAU (Reciprocal activation unit) is proposed in this paper. Most of the popular algorithms given more importance to positive signals, but proposed method handles the negative and positive inputs equally. The proposed RAU tested with both multiclassification and binary classification datasets. Iris flower and Wisconsin Breast Cancer datasets are used for the analysis. In Breast cancer dataset RAU provides 99.25% and 97.08% accuracy for classification of train and test sets respectively. In Iris dataset RAU provides 99.05% and 97.78% accuracy for the classification of train and test sets. Analysis of the same datasets are performed with the existing activation functions- Sigmoid, RMAF, Swish, Tanh and ReLU. Results showed that RAU performed better than other activation functions.

Keywords- back propagation; activation function; hidden layers;artificial neural networks; deep neural networks.

I. INTRODUCTION

Artificial neural networks (ANN) developed based on the working of human brain. As biological neuron in human brain ANN contains artificial neurons. That are the computational units in neural networks. ANN consist of three layers: input, output and hidden layers. In human brain millions of neurons are interconnected by biological network. Similarly in ANN number of neurons are interconnected by Artificial Neural Network [1]. In deep neural network inputs are weighted based on its importance. Here the activation function converts the sum of weighted input signals as output [2].

$$Y(x) = Fn(w1 * x1 + w2 * x2 + w3 * x3) \qquad (1)$$

Here $x1$, $x2$, and $x3$ are the input values and $w1$, $w2$ and $w3$ are their weights respectively. Fn is the activation function that convert the sum of products of the weights and input values to an output value Y.

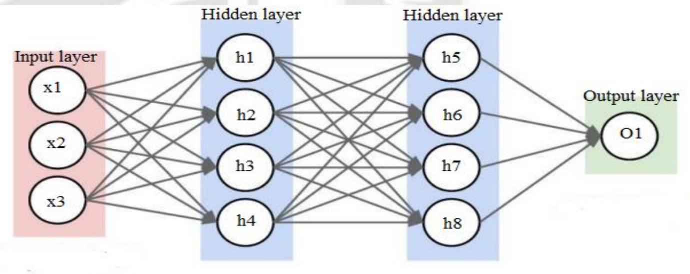


Figure 1. Neural Network

A. *Input and Output layers*

In a neural network input features are first given to the input layer. The number of neurons in the input layer is same as the number of input features in the dataset. The final output or prediction is obtained from the output layer. Number of neurons in the output layer depends upon the classes in the dataset.

B. *Hidden layers*

DNNs have one or more hidden layers. Any number of neurons and any number of hidden layers can include in neural network, there is no specific rule for deciding the same. Dense neural networks connect all the neurons in one layer to the next layer. Here repeated training is possible with the same training

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8477
***Article Received:** 28 July 2023 **Revised:** 16 September 2023 **Accepted:** 01 October 2023*

data, hence the name. The term epoch specifies the number of time training is done with same train data.

C. Back Propagation

ANN work in two phases, forward and backward phases. In forward phase outputs are generated. If there exists a difference in actual data and predicted data, signals are transmitted back to the input layers for updating the weight. This is called back propagation. MSE (Mean Squared Error) is one of the methods for calculating such difference.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - o_i)^2 \qquad (2)$$

Here predicted value is denoted by y_i and actual output is denoted by o_i. The weights are updated until the specified epochs reached or the maximum performance measure is obtained.

D. Activation Functions

In neural networks the sum of weighted input signals is converted to an output by using activation functions. So, it is an essential component in neural networks. Performance of the model highly influenced by the activation function. Hence it is the heart of deep learning networks. Nowadays many researches are going on to get better activation functions.

II. PREVIOUS WORKS

Many papers are published in this area. Here some works in recent articles are included.

A. Sigmoid Activation Function

Sigmoid is a 'S' shaped non-linear activation function. It provides big changes in output when the input closer to zero. But going to the end output becomes constant means have little change. This is because of vanishing gradient problem. The gradient becomes very small when going to the end. Sigmoid activation function one of the popular activation functions used in binary classification [3,4].

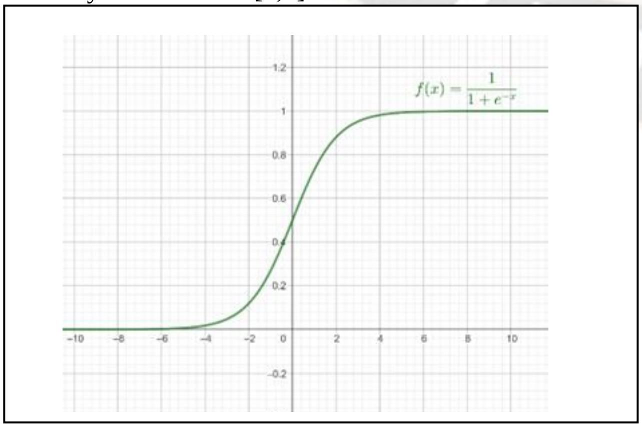


Figure 2. Sigmoid

B. Tanh Activation function

Tangent Hyperbolic Function is also called Tanh activation function. Like sigmoid, Tanh is also a nonlinear function. Vanishing gradient problem exist in Tanh but its gradients are stronger than sigmoid. As seeing in the figure 3 most of the outputs are centered to zero[3,5].

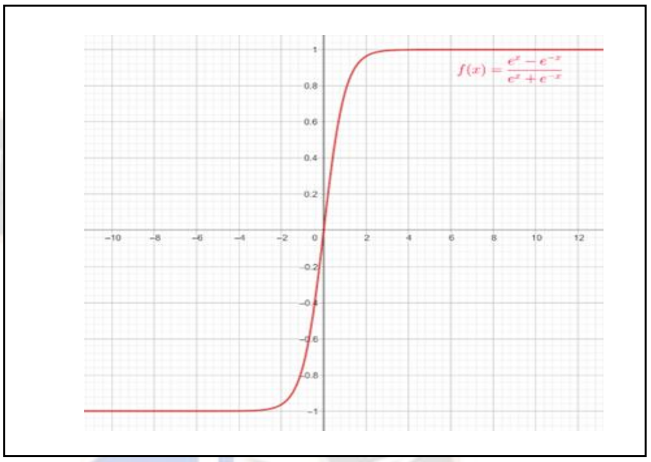


Figure 3. Tanh Activation function

C. SWISH

Ramachandran at el. [6] proposed SWISH activation function for deep learning neural network.

$$f(a) = a.sigmoid(\beta a)$$

Here a is input and β is a constant or a trainable parameter. When the value of β is 1 it behaves like sigmoid. This activation function tested and found it outperformed in several datasets. But diminishing gradient problem occurs while using this activation function [6].

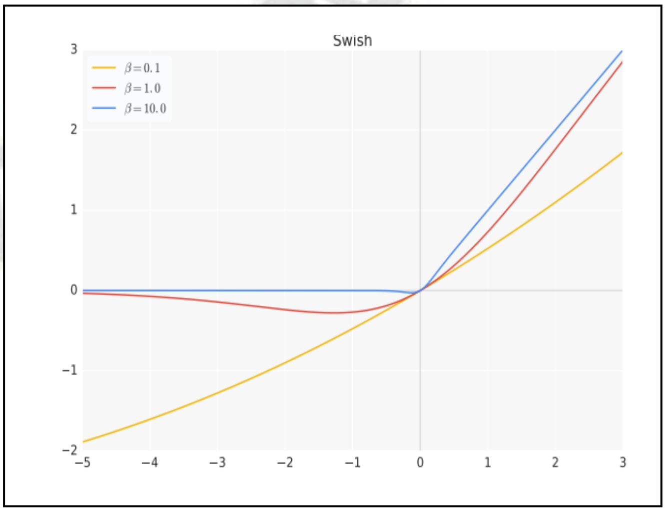


Figure 4. SWISH Function

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8477
***Article Received:** 28 July 2023 **Revised:** 16 September 2023 **Accepted:** 01 October 2023*

D. Rectified Linear Unit(ReLU)

ReLU activation function perform better in most of the datasets. If x is an input signal, ReLU returns maximum value among 0 or x. That means if x is positive value ReLU function returns x itself, otherwise it returns 0. It returns 0 for both negative and 0 inputs hence ReLU output is nondifferentiable. This activation function also suffer diminishing gradient problem[7]

Figure 5. ReLU function

E. RMAF(ReLU Memristor like Activation Function)

An activation function proposed by Yongbin Yu et al. [8], which is called RMAF activation function. Two parameters constant parameter(α) and threshold parameter(p) are introduced in RMAF for smooth functioning. It considered negative values too and experiments showed it is better than ReLU.

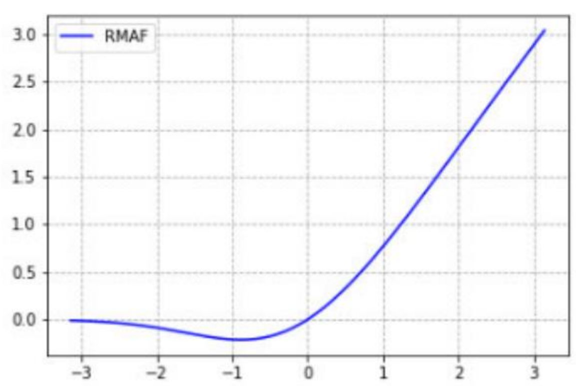


Figure 6. RMAF

F. Other Activatin Functions

ELU (Exponential Linear Unit) is another activation function which works like ReLU. In ELU negative outputs are generated for negative input values and positive output values are generated for positive inputs. Hence the mean activation is very close to zero. This function omits vanishing gradient problem. ELU is commonly used in neural networks that contains more hidden layers[9]. Another activation function is developed by adding a constant scaling parameter to ELU, which is called SELU (Scaled Exponential Linear Unit). There exists a self-normalizing function in each layer [10]. GELU is an activation function which is used for the Natural Language Processing (NLP) and image processing. It is computationally expensive [11]. Display Rectilinear unit (DReLU) is developed by enlarging the third quadrant of ReLU batch normalization [12]. Presently researches are focused to develop a nonparametric and a complex activation function for Complex valued Neural Networks (CVNNs) [13].

III. PROPOSED METHODOLOGY

Sigmoid and Tanh activation function have vanishing gradient problem and slow training speed. Hence ReLU is one of the most popularly used activation function now. But the main problem of ReLU is that it does not consider about the negative input values. ReLU activation function gives the output zero for all values which is less than or equal to zero. To solving this problem a new activation function called Reciprocal Activation Unit (RAU) is proposed. It gives the same weightage to both positive and negative input signals. RAU gives positive output values for positive input and negative output values for negative input. So mean output value is close to zero.

RAU can be defined as

$$RAU(x) = \begin{cases} \dfrac{x}{1 + |\frac{1}{x}|} & if \quad x \neq 0 \\ 0 & if \quad x = 0 \end{cases}$$

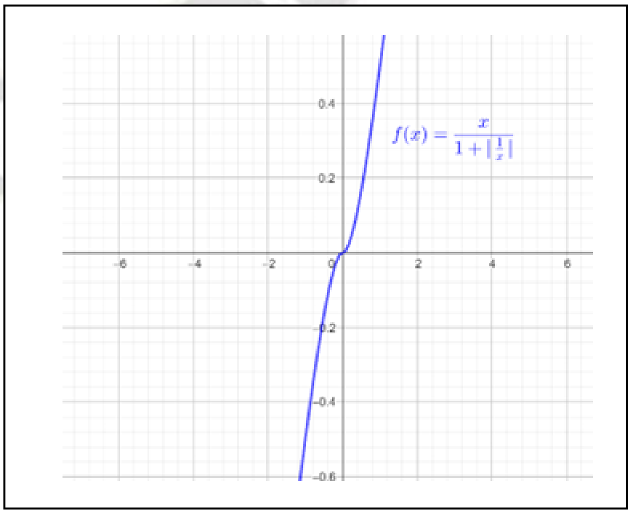


Figure 7. RAU

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8477
***Article Received:** 28 July 2023 **Revised:** 16 September 2023 **Accepted:** 01 October 2023*

In the backward phase of Neural Networks weights of the edges are updated by a process called back propagation. For this it is necessary to calculate the derivative of the activation function. The derivative of $f(x) = \frac{x}{1+|\frac{1}{x}|}$ obtained as $f^1(x) = \frac{(\frac{2}{|x|}+1)}{(1+|^1/_x|)^2}$, when $x \neq 0$

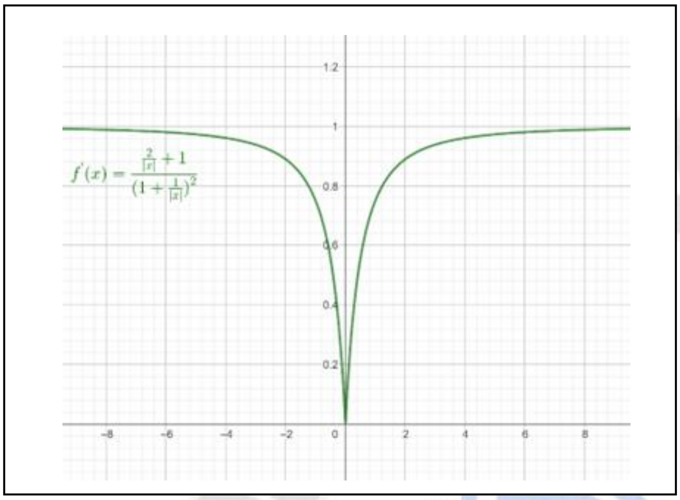


Figure 8. Derivative f of RAU

A. *Datasets*

Here two bench marked dataset is considered for the analysis. That are Wisconsin Breast Cancer dataset and Iris flower Dataset. In this Iris flower Dataset is a multiclassification dataset and cancer dataset is a binary classification dataset. Evaluation of proposed model is performed by analyzing the accuracy and loss of both training and testing data.

Iris flower dataset contains 4 features and 150 observations [14]. Dataset used to predict the variety of Iris flowers. Three varieties of Iris flowers included in the dataset. They are Iris Virginia, Iris setosa and Iris versicolor. There exist 50 observations from each classes The four features are petal length, sepal length, petal width and sepal width. Wisconsin Breast Cancer Dataset is one of the most popular datasets for cancer prediction. It contains 30 features and 569 observations. Breast Cancer dataset contains two categories of classes, no cancer (Benign) and Cancer (Malignant) [15].

B. *Experimental Setup*

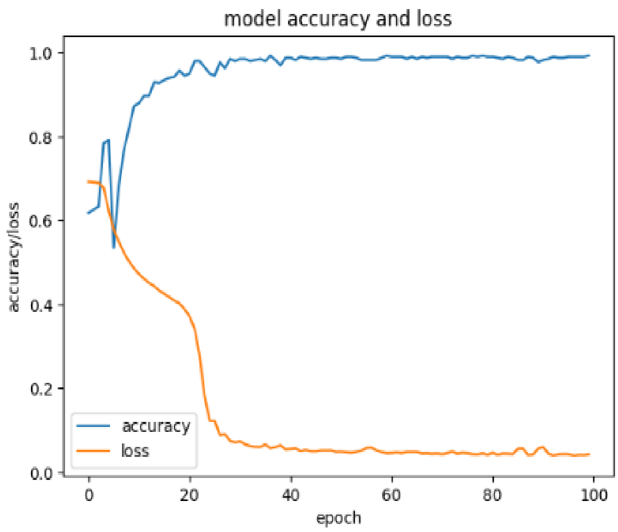
Input layer, hidden layers and output layers are the three layers in neural networks. Input layer contain the neurons which is similar to number of features. So, thirty neurons are kept in the input layer while evaluating the breast cancer dataset. Three hidden layers are used to create a model. The first hidden layer contains eight neurons second and third hidden layers contain six neurons and four neurons respectively. Cancer dataset is a binary classification dataset so the output layer contains one neurons and sigmoid activation function. For the breast cancer prediction binary cross entropy is taken as loss function, adaptive momentum optimizer is the optimizer. Performance analysis is done using loss and accuracy.

While using Iris dataset four neurons are used in input layer because dataset contains four features. Neural network contains three hidden layers and there exist five, four and four neurons in each hidden layer. In Iris dataset there are three categories to predict, so it contains three neurons in the output layer. For the analysis here accuracy, categorical cross entropy, adaptive momentum is taken as performance metrics, loss function and optimizer respectively. For both dataset 70% data used for training and 30% data used for testing. Training is performed in 100 epochs for both datasets.

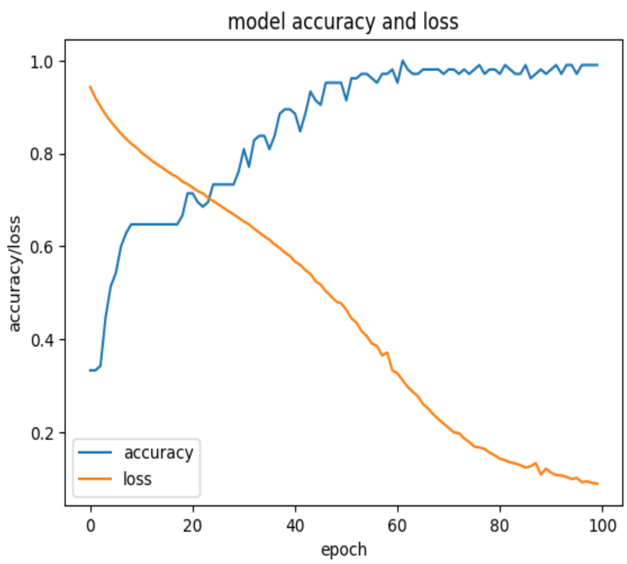
IV. RESULTS AND DISCUSSIONS

Performance of different existing activation functions and proposed method RAU are compared. In Breast Cancer dataset highest training accuracy 99.25% is obtained while using RAU. And also, it provides 97.08% accuracy for the test set which is higher than ReLU, Sigmoid, SWISH, Tanh and RMAF. The variation of loss and accuracy in each epoch are shown in Figure 9.

In the case of Iris dataset 99.05% accuracy is obtained for the training data, which is higher than all the previous activation functions reviewed. Here 97.78% is the testing accuracy. It is better than Sigmoid, ReLU, Tanh and SWISH activation functions. Performance evaluation using each activation function shown in Table 1.



(a)



(b)

Figure 9. loss and accuracy evaluation of Cancer Dataset(a) and Iris Dataset(b)

TABLE I. PERFORMANCE OF DIFFERENT ACTIVATION FUNCTIONS IN CANCER AND IRIS DATASET

Datasets	Activation Functions	Train Data		Test data	
		Accuracy	**Loss**	**Accuracy**	**Loss**
Wisconsin Breast Cancer Dataset	Tanh	96.47	0.103	94.55	0.102
	ReLU	97.44	0.054	96.31	0.053
	Swish	96.47	0.111	95.90	0.110
	Sigmoid	95.52	0.165	95.22	0.164
	RMAF	98.74	0.043	96.34	0.042
	RAU	99.25	.0427	97.08	0.079
Iris Flower Dataset	Tanh	97.46	0.094	96.26	0.090
	ReLU	98.33	0.098	96.41	0.089
	Swish	98.50	0.095	96.34	0.099
	Sigmoid	97.83	0.095	96.23	0.091
	RMAF	98.81	0.077	97.83	0.052
	RAU	99.05	0.0895	97.78	0.106

V. CONCLUSION

Activation function is a vital part of neural network. Here proposed activation function RAU used for the prediction of category of Iris flower and breast cancer. Performance of proposed RAU compared with the existing activation functions such as ReLU, Sigmoid, Swish RMAF and Tanh. In the binary and multiclassification datasets proposed RAU is found outperformed. The proposed function has the advantage of handling positive and negative input signals equally. Most of the activation functions widely used neglect the negative inputs. In the analysis using Iris and Breast cancer dataset 99.05% and 99.25% accuracy obtained for the test data. Presently the proposed model is not tested with image dataset. In future planning to do the necessary modification to use this activation function in convolutional neural networks.

REFERENCES

[1] Ravindra B V, N.Sriram,M.Geetha, “Chronic Kidney disease detection using back propagation neural network classifier”, International Confererence on Communication,Computing and Internet of Things,18528464 , Febuary 2018,pp 65-68
[2] Ibrahim M Nasser, “Predicting whether a couple is get divorced or not using neural network”,International journal of engineering and information Systems,Vol 3 ,Issue 10,pp 89-95,October 2019
[3] Jianli Feng, Shengnan Lu,”Performance Analysis of Various Activation Functions in Artificial Neural Networks” IOP Conf. Series: Journal of Physics: Conf. Series 1237 ,2019
[4] Akhilesh A Waoo, Brijesh K Soni,” Performance Analysis of Sigmoid and ReLU activation functions in Deep Neural Netwok”, Algorithms for intelliget systems book series (AIS) ,22 July 2021
[5] Junxi Feng, Xiaohai He, Oizhi Teng and Chao Ren ,” Reconstruction of Porous media from extremely limited information using unconditional generative adversarial networks”,Phuysical Riview, Vol 100, no 3, Sep 2019,
[6] Ramachandran, B. Zoph, and Q. V. Le, ‘‘Searching for activation functions,’’ 2017, arXiv:1710.05941. [Online].
[7] Maas, A. L., Hannun, A. Y., & Ng, A. Y. “Rectifier nonlinearities improve neural network acoustic models”. In Proc. Icml,2013
[8] Yongbin Yu, Kwabena Adu , Nyima Tashi, Patrick Anokye, Xiangxiang Wang,Mighty Abra Ayidzoe P,”RMAF: Relu-Memristor-Like Activation Function for Deep Learning”,IEE Access, April 2020
[9] D.A. Clevert, T. Unterthiner, and S. Hochreiter, ‘‘Fast and accurate deep network learning by exponential linear units (ELUs),’’ 2015, arXiv:1511.07289. [Online]
[10] V. Nair and G. E. Hinton, ‘‘Rectified linear units improve restricted Boltzmann machines,’’ in Proc. 27th Int. Conf. Mach. Learn. (ICML), 2010, p. 807–814.
[11] Hendrycks and K. Gimpel, ‘‘Gaussian error linear units (GELUs),’’ 2016, arXiv:1606.08415. [Online].
[12] Macêdo, C. Zanchettin, A. L. I. Oliveira, and T. Ludermir, ‘‘Enhancing batch normalized convolutional networks using displaced rectifier linear units: A systematic comparative study,’’ Expert Syst. Appl., vol. 124, pp. 271–281, Jun. 2019
[13] S. Kumar Roy, S. Manna, S. R. Dubey, and B. B. Chaudhuri, ‘‘LiSHT: Non-parametric linearly scaled hyperbolic tangent activation function for neural networks,’’ 2019, arXiv:1901.05894. [Online]. Available: http://arxiv.org/abs/1901.05894

[14] L. Zhang and P. N. Suganthan, ‘‘Random forests with ensemble of feature spaces,’’ Pattern Recognition., Vol. 47, no. 10, pp. 3429–3437, Oct. 2014
[15] Siham A. Mohammed, Sadeq Darrab ,Slah A,Noaman and Gunter Saake ,”Analysis of Breast Cancer detection using different machine learning techniques”, International conference on data mining and big data, Vol.1234, pp.108-117 ,July 2020