

Benchmarking of activation functions for breast cancer detection

Miguel-Angel Quiroz-Martinez¹, Ronald-Dario Montoya-Guillen¹, Galo-Enrique

Valverde-Landivar¹, Maikel-Yelandi Leyva-Vazquez²

¹ Computer Science Department, Universidad Politécnica Salesiana,
Guayaquil, Ecuador

² Facultad de Ciencias Matemáticas y Físicas , Universidad de Guayaquil,
Guayaquil, Ecuador

ABSTRACT

Breast cancer is a common disease and one of the leading causes of death global-ly; there are several methods, technologies, algorithms, or functions to detect their presence. The objective is to develop a benchmarking of activation functions in the detection of breast cancer for its selection with the purpose of increasing the effectiveness in the diagnosis of this disease. The research methodology used in this work is observation in scientific articles, experimental in the implementation of the algorithm, quantitative analysis of the results, and a descriptive approach on the activation functions and the results of the algorithm. The results of this work are an implementation of the Activation Functions Sigmoid, ReLu, Swish, Tanh, and Softmax on the Keras framework; and the realization of benchmarking in Google Colab. It was concluded that this work is an opening towards new knowledge to favor the

cooperation and cohesion of different actors; it is a way of betting on knowledge, innovation, and achieving dynamism with planning, analysis, and action of the idea to be implemented for an improvement in the field of health; ReLu has higher accuracy with 98.20% and is the first choice for pre-training and training neural networks.

Keywords: Benchmarking, Breast cancer, Artificial Neural Networks, Activation functions

INTRODUCTION

Worldwide, breast cancer is a multifactorial disease due to genetic or environmental causes that contribute to its appearance. This begins with a genetic variable that is formed with a single cell and can be caused by agents of the external area of our body or by hereditary genetics (Benavides, Fonseca, Mora, Moya, Navarro, Paniagua 2013). It is estimated that 458 thousand women die annually from this type of cancer; has an incidence rate of 39 cases per 100 thousand women; in Ecuador and other countries, breast cancer is a recurrent disease that specifically affects the female gender; every year, more people have a suspected diagnosis due to this disease called cancer, almost half of the cases are advanced; this minimizes the chances of healing and survival in people; it is estimated that by 2030 there will be 12 million deaths from cancer (Ulloa, Ronquillo Torresano 2020). Breast cancer is determined by the examinations that are performed on patients based on the interpretation of the specialist doctor; the diagnoses corresponding to the medical studies of consultation are analyzed to acquire information on the affectation of this cancerous disease; these tests are mammograms and breast tissue analysis; generally, in case of abnormal detriments, the doctor seeks the assertion of the diagnosis cytologically, by puncture-aspiration with a fine needle, or histologically, that is, the material is obtained by puncture, by thick needle or conventional surgical biopsy. This literature review article covers benchmarking as a continuous improvement strategy and technique possible to find through the neural network activation function. (Sigmoid, Relu, Swish, Tanh y Softmax) That favor the detection of breast cancer (Dasgupta, Chowdhury, and Nanda 2021). Computer-aided diagnostic (CAD) technologies help doctors interpret multimedia content obtained from tests performed on the patient, such as medical images; this assists in recognizing patterns that allow the specialist, usually a radiologist, to interpret the information with much less difficulty; the subjectivity of visual diagnostics can be minimized with digital image analysis based on machine learning techniques; these are also an aid to decision making. Convolution Neural Network (CNN) is used to analyze, identify or classify images because it facilitates the analysis in images; its advantage is that it minimizes people's efforts and pre-processing; employs backpropagation over the course of learning, and makes the network more accurate; it consists of an input layer of neurons, several hidden layers, and an output layer; the neuron in one layer is linked to neurons that belong to the next layer. The CNN architecture is used for diagnosis/classification of breast cancer because it extracts features that improve and facilitate seeing the malignant in the breast masses; their assistance maximizes the process of discovering breast cancer and treating it at an early stage, before its spread (Desai and

Shah 2021). CNN has a lot of use in image classification and feature extraction; a CNN architecture is a composition of mathematical functions; here, the convolution is a function that gets an image at the input, at the output, it gets features from a filter; there are layers that are trained with data, and these are linked by a classifier (Gupta 2020).

The objective is to develop a benchmarking of activation functions in the detection of breast cancer for its selection with the purpose of increasing the effectiveness in the diagnosis of this disease. Benchmarking is applied with the following activation functions: Sigmoid, ReLu, Swish, Tanh, and Softmax. The research methodology used in this work is observation in scientific articles, experimental in the implementation of the algorithm, quantitative analysis of the results, and a descriptive approach on the activation functions and the results of the algorithm.

PRELIMINARIES

An Activation Function (AF) confirms to the neuron whether it allows the evaluation of the weighted sum with the aggregation of a bias; the goal of AF is to implant nonlinearity in the neuron argument; neurons maneuver in correlation to weight, bias, and their AF; a process called retro propagation calculates the gradient of the error function (Dasgupta et al. 2021). The AF decides which artificial neuron is activated (Lau and Lim 2019). Implementing efficient AF is a decisive challenge in deep learning (Maguolo, Nanni, and Ghidoni 2021). There are several AF are ReLu, Logistic, Multistate (Lau and Lim 2019), Swish, Mish (Dasgupta et al. 2021), Adaptive hyperbolic tangent tanh (Lau and Lim 2019), Sigmoid (Prasath Alias Surendhar and Vasuki 2021), Leaky ReLu (Abdelrahman et al. 2021), Softmax, Sigmoid-Weighted Linear Units (SiLU), Hard Hyperbolic, Softsign, Softplus, Exponential Linear Units, Maxout, ELiSH (Nwankpa et al. 2018), (Maharjan et al. 2020).

MATERIALS Y METHODS

Data set: We used breast cancer data from the University of Wisconsin Hospital, USA (Wisconsin 2021), (Wolberg 1990), (O.L. 1990); this database describes each specimen or sample with attributes related to the characteristics of the cell nuclei of the traits found in breast tissue, which were analyzed by the Xcvt program, capable of performing analysis of cytological traits based on a digital scan. Characteristics of the set: Multivariate; Number of records: 699; Area: life; Data type: integer between 1 and 10; Number of columns/attributes: 10; Donation date: 1992-07-15; Task type: classification; Distribution: 458 are benign and 241 are malignant. Data attributes sample code, group volume, cell size balance, balance in cell shape, secondary adhesion, single-cell dimension, stripped nuclei, soft chromatin, normal nucleoli, mitosis, class: 2 is benign, and 4 is malignant.

Training and Testing: Adopt and apply the Activation Functions in Keras; Adopt and apply the functions to the data set; Implement the algorithms on the Google Colab platform (MLeyva-RMontoya 2021); The platform runs on Intel Core i5-8260U 2.00GHz CPU, 8GB RAM, 64-bit OS, and 512GB SSD. The training phase is in 5 iterations for each activation function; 80% of the dataset is for training, and 20% is for testing. Training is applied with the following FY: Sigmoid, ReLu, Swish, Tanh, Softmax.

Algorithm:

Assign the physical path to the dataset;

Classification of the set of tumors into benign and malignant (see Figure 1).

	clump_thickness	unif_cell_size	unif_cell_shape	marg_adhesion	single_epith_size	bare_nuclei	bland_chrom	norm_nucleoli	mitoses	class
0	5	4	4	5	7	10	3	2	1	0
1	3	1	1	1	2	2	3	1	1	0
2	6	8	8	1	3	4	3	7	1	0
3	4	1	1	3	2	1	3	1	1	0
4	8	10	10	8	7	10	9	7	1	1

Figure 1. Classification.

To transform the attributes/characteristics, the scale estimator is used to obtain a range between zero and one; a representation of the training data (see Figure 2).

```
array([0.44444444, 0.          , 0.11111111, 0.          , 0.11111111,
       0.99991001, 0.11111111, 0.          , 0.          ])
```

Figure 2. Scale estimation.

Activation functions are created in Keras. The two-dimensional matrix is fed into a neural network. Here the flat layer becomes a one-dimensional matrix; that is, each row is next to the other row; the hidden layers contain 27, 54, and 27 neurons each; the output layer contains one neuron for each possible value; AF returns the probability that a sample represents an integer. Passing 0.25, each hidden neuron is set to 0 with a probability of 0.25; that is, there is a change of 25% if the output of a given neuron will be forced to zero value.

The optimizer is an argument for compiling the KERAS model; to predict classes, mean squares of errors between labels/predictions are used for the loss function so that the model is minimized during training.

The lot size is 64, which calculates the gradient; after passing the 64 samples through the

neural network, it is done in 5 interactions or times as specified by the epochs.

ANALYSIS OF THE RESULTS OF THE EXPERIMENT

In a neural network, the main parameter is the activation function; in the experiment conducted among five AFs, it is a question of choosing an activation function applied to the dataset; after training the neural network with the activation functions Sigmoid, ReLu, Swish, Tanh y Softmax, this network has a congruent effect; in the experiment it was trained in 559 records and validated in 140 records; the figures represent the results of our experiment.

The set of values in training, Softmax remains the same at 0.3020; Sigmoid has a downward trend from 0.1131 to 0.0121; ReLu has a downward trend from 0.0532 to 0.0108; Swish has a downward trend from 0.0307 to 0.0098; Tanh has a minimal downward trend from 0.0168 to 0.0112; here Swish has the lowest value in interactions; the X-axis is the number of repetitions in iterations from 1 to 20; the Y-axis is the measure for training (see Figure 3).

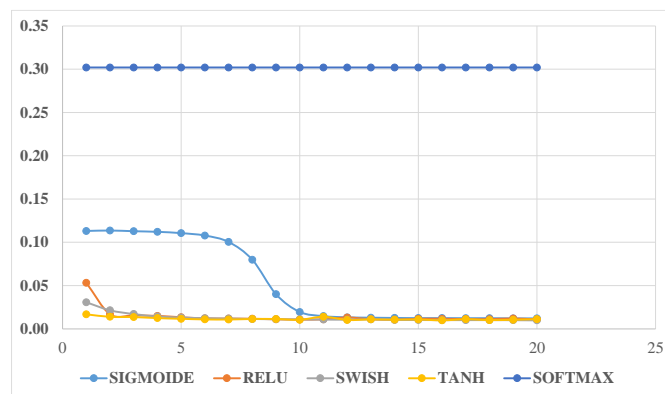


Figure 3. Training of functions in iterations.

According to the trained network, Sigmoid has an accuracy of 98.9872%, ReLu has an accuracy of 98.9194%, Swish has an accuracy of 99.0230%, Tanh has an accuracy of 98.8841%, and Softmax has an accuracy of 69.8001%; Swish has the best result in training; the X-axis is the amount of the trained AF; the Y-axis is the percentage for training (see Figure 4).

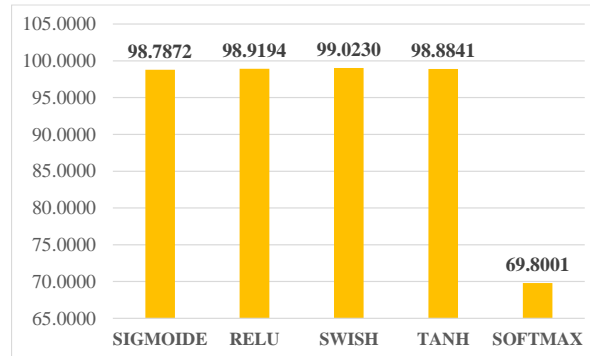


Figure 4. Results of each function in training.

Precision is the degree of closeness of the measurements of the image-based dataset, and here the precision curve is a measure between the trained value and the real one; see Figure 5 represents the precision curve, the X-axis is the number of iterations from 1 to 20; the Y-axis is the percentage for accuracy; the precision starts at 0.96 as the lowest value; then it remains between 0.98 and 1 after iteration number 10; from here the model is stable except for the Sigmoid function which then goes down; the best results are obtained with the ReLU activation function. ReLU is the best activation feature that has the highest accuracy; then, ReLU is the first choice for preparing and training neural networks.

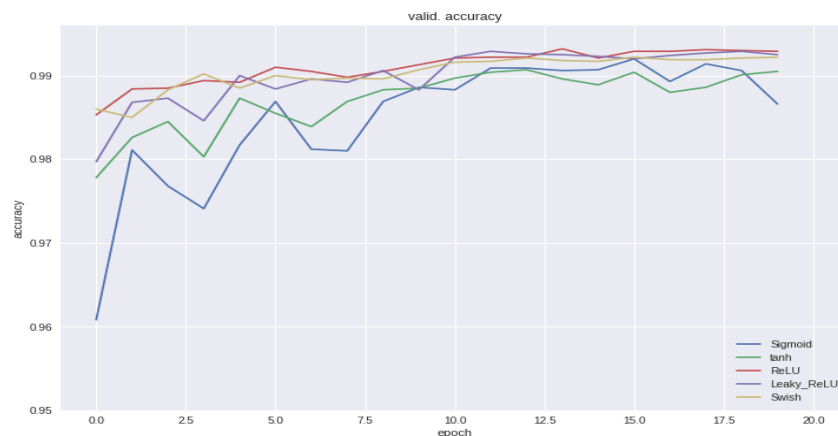


Figure 5. Accuracy of Activation Functions.

Additionally, a statistical test is performed with the Kruskal-Wallis algorithm (Mbuli et al. 2020) in the five function groups to verify whether the samples are equidistributed and to

verify whether these belong to the same distribution/population; this algorithm confirms whether the populations to be compared are asymmetric in the following code:

```
stat, p = kruskal(sigmoid, swish, tanh, softmax)
```

If the result p is greater than 0.05, then “These are probably of the same distribution” else “These are probably different distributions”. The result was: $stat = 67.116$, $p = 0.000$; the value p is less than the significance value of 5%; the answer is that the data is probably from different distributions.

CONCLUSIONS

In this scientific article, the theoretical foundations of the activation functions that assist neural networks were studied; the usefulness and importance of benchmarking for a complementary aid that contributes to diagnostic precision strategies that offer a second opinion to specialists were evidenced. AFs such as Sigmoid, ReLu, Swish, Tanh, and Softmax were addressed, which stand out among the classification methods used in computer-aided diagnostic systems; these networks are widely used in classification and pattern recognition problems in a wide variety of research areas. Diagnostic support technologies related to breast cancer are a topic of great impact worldwide; implemented a useful and effective strategy to assist specialists in designated breast cancer screening topics. The activation functions were implemented Sigmoid, ReLu, Swish, Tanh, and Softmax on framework Keras. The practice of benchmarking carried out in Google Colab presupposes an opening towards new knowledge to favor the cooperation and cohesion of different actors; it is a way to bet on knowledge, innovation and achieve a dynamism with planning, analysis, and action of the idea to be implemented for an improvement in the field of health. The results show the accuracy of the activation functions and their significant differences; ReLu has higher accuracy with 98.20% and is the first choice for preparing and training neural networks.

ACKNOWLEDGMENTS

This work has been supported by the GIIAR research group and the Universidad Politécnica Salesiana

REFERENCES

Abdelrahman, Ghamdi, Collado-Mesa, and Abdel-Mottaleb. 2021. “Convolutional Neural Networks for Breast Cancer Detection in Mammography: A Survey.” *Computers in*

- Biology and Medicine* 131(November 2020):104248. doi: 10.1016/j.compbimed.2021.104248.
- Benavides, Fonseca, Mora, Moya, Navarro, Paniagua, Quirós. 2013. “Cáncer de Mama Inflamatorio: Un Reto Diagnóstico y Terapéutico.” *Medicina Legal de Costa Rica* 30(1):99–109.
- Dasgupta, Chowdhury, and Nanda. 2021. “Performance Comparison of Benchmark Activation Function for Facial Mask Detection Using CNN.” Pp. 355–67 in. Springer Singapore.
- Desai, and Shah. 2021. “An Anatomization on Breast Cancer Detection and Diagnosis Employing MLP and CNN.” *Clinical EHealth* 4:1–11. doi: 10.1016/j.ceh.2020.11.002.
- Gupta, Chawla. 2020. “Analysis of Histopathological Images for Prediction of Breast Cancer with Pre-Trained CNN.” *Procedia Computer Science* 167(2019):878–89. doi: 10.1016/j.procs.2020.03.427.
- Lau, and Lim. 2019. “Review of Adaptive Activation Function in Deep Neural Network.” *IEEE EMBS* 686–90. doi: 10.1109/IECBES.2018.08626714.
- Maguolo, Nanni, and Ghidoni. 2021. “Ensemble of Convolutional Neural Networks Trained with Different Activation Functions.” *Expert Systems with Applications* 166(October 2020):114048. doi: 10.1016/j.eswa.2020.114048.
- Maharjan, Alsadoon, Prasad, Al-Dalain, and Alsadoon. 2020. “A Novel Enhanced Softmax Loss Function for Brain Tumour Detection Using Deep Learning.” *Journal of Neuroscience Methods* 330(August 2019):108520. doi: 10.1016/j.jneumeth.2019.108520.
- Mbuli, Mendu, Seitshiro, and Pretorius. 2020. “Statistical Analysis of Forced Outage Durations of Eskom Subtransmission Transformers.” *Energy Reports* 6(2020):321–29. doi: 10.1016/j.egy.2020.11.235.
- MLeyva-RMontoya. 2021. “Proyecto Google Colab.” 1. Retrieved October 25, 2021 (https://colab.research.google.com/drive/1KShwqNrPZScE9wR3OQTJJJeE_50uOxRg#scrollTo=J99GQ10BVDgv).
- Nwankpa, Ijomah, Gachagan, and Marshall. 2018. “Activation Functions: Comparison of Trends in Practice and Research for DL.” 1–20.
- O.L., Mangasarian and. 1990. “Multisurface Method of Pattern Separation for Medical Diagnosis Applied to Breast Cytology.” *Proceedings of the National Academy of Sciences* 87(1):9193–96.
- Prasath Alias Surendhar, S., and R. Vasuki. 2021. “Breast Cancers Detection Using Deep Learning Algorithm.” *Materials Today: Proceedings* (xxxx). doi: 10.1016/j.matpr.2020.11.600.
- Ulloa, Ronquillo Torresano, Sánchez. 2020. “Sobrevida En Pacientes Con Cáncer de Mama SOLCA.” *Revista Medica Sinergia* 5(7):e542. doi: 10.31434/rms.v5i7.542.
- Wisconsin, University of. 2021. “University of Wisconsin Hospitals Data1.” 1. Retrieved October 10, 2021 (<https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Original%29>).
- Wolberg, O. L. Mangasarian and W. H. 1990. “Cancer Diagnosis via Linear Programming.” *SIAM News* 23(5):1–18.