XLI
CILAMCE
XLI Ibero-Latin American Congress on Computational Methods in Engineering
16-19 November, 2020 FOZ DO IGUAÇU
PARANÁ - BRAZIL

iagnosis of breast cancer using Artificial Neural Network

Danilo José dos Santos Costa[1], Luiz Eugênio Hoffmann Lopes[1], Matheus Pereira da Silva[1], Nildson de Castro Pinheiro Mello[1], [12]Marta de Oliveira Barreiros

[1]Dept. of Computer Engineering, State University of Maranhão – UEMA, São Luis-MA
R. Paulo VI, s/n - São Cristovão, 65055-000, São Luis/Maranhão, Brasil
danilodalessandro08@gmail.com;luiz.eugenio.hoffmann@gmail.com;matheuspsilva29@gmail.com;
nildsond@gmail.com; marta-barreiros@hotmail.com;

Abstract. Breast cancer is a common invasive cancer in women, affecting more than 10% of the world's female population, being one of the main causes of death in the world. The analysis of the biopsy procedure carried out by a qualified professional, still appears as the most effective method in the diagnosis of cancer in general, if diagnosed and treated prematurely, the patient has great chances of cure, however the process can be time-consuming and the diagnosis is not it's always assertive. In this context, automatic breast cancer classification algorithms appear, using machine learning algorithms in search of an efficient result for the final diagnosis. However, it is still necessary to establish methodologies with greater precision to guarantee the assertive result. Here, we use a multilayer artificial neural to classify breast cancer in two stages: malignant and benign. To test the proposed methodology, a Breast Cancer Wisconsin (Diagnostic) Data Set database was used, where feature extractions were made from digitalized images of breast masses. Several architectures of the neural network have been proposed. And a cross-validation was implemented with 10 k-fold. The result showed an average accuracy of 92%. Thus, with this result, a tool can be made that can expedite the diagnosis of the patient in a precise way and, consequently, enable an early treatment increasing the chances of cure.

Keywords: Cancer. Diagnosis. MLP. Algorithms.

1 Introduction

Breast cancer is a leading cause of death, with cancer being the most frequently diagnosed in the world, with about half a million cases reported each year, being the second most common cancer among women in Brazil, lagging behind only skin cancer. Breast cancer is caused by the disordered multiplication of cells in the breast, thus forming a tumor [1].

There are two main types of breast cancer risk: the likelihood that an individual will get breast cancer over a period of time and the likelihood that a mutation will occur in a high-risk gene [2]. As it does not have a single cause, there are several factors that can contribute to the development of such a disease, such as: age, genetic factors, hereditary factors, among others, age is one of the most aggravating factors in the biological changes created by the passage of time. Over time, women over 50 are more likely to develop breast cancer [2], [3].

Breast cancer has an estimated 95% chance of cure with an early diagnosis, but it has a complex diagnosis, even for professionals with years of experience [4]. However, one of the greatest difficulties reported by patients is the inefficiency of the Unified Health System (SUS), causing delay in diagnosis and harming infected patients. It is common for the population to associate breast cancer with women, but this type of cancer is not restricted to women only, it can also affect men.

The correct diagnosis, in the initial phase of the disease, can help with actions and treatment, thus enabling a higher level of efficiency. The systems and methods used in the daily life of medicine on many occasions can be laborious and slow [5], with the increase in technology and the ability to analyze information and data, techniques for image processing and data analysis have emerged that can streamline the process and assist a healthcare professional to generate a diagnosis more quickly.

In this context, computing has become a great ally in the search for a fast and reliable diagnosis, with intelligent and faster solutions has started to gain space in the search for a satisfactory result. Predicting the risk of breast cancer is vital to combat this disease. Therefore, scientists seek to apply different methods, screening at an early stage, to find the types of cancer before they cause the symptoms. Thus, many strategies are developed for the early prediction of cancer treatment outcome. With the advancement of technologies in the medical field, much data on cancer has been collected and analysis is available in the research community [6].

Even though there are already many tools for predicting the stages of breast cancer, still getting an accurate result is one of the most interesting and challenging tasks. Thus, the objective of this work is to classify breast cancer using multilayer neural networks.

2 Methodology

The program is divided into 4 stages: importing data from.csv files; parameterization of the network, giving values for the variables; instantiating the MLP classifier; using K-fold for cross-validation. In each test, only the parameters of step 2 were changed and executed 10 times to have an average of accuracy as the final result.

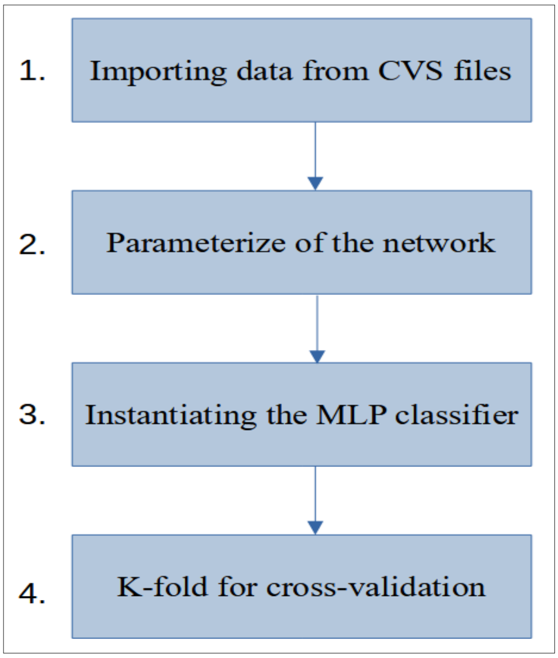
1. Importing data from CVS files
2. Parameterize of the network
3. Instantiating the MLP classifier
4. K-fold for cross-validation

Figure 1. Methodology steps.

2.1 Database

The database was taken from the UCI Machine Learning Repository [7] and is one of the ten most popular databases. The breast cancer diagnosis database in Wisconsin has 569 instances and 32 attributes, 357 of which were malignant and 212 benign. The data set was formed from images of breast mass and extracted 32 characteristics of them. Below is a figure showing these characteristics. Figure 1 shows the database attributes. Characteristics 3 to 32 (float type) were used as input to the neural network, while characteristic 2 was used as output of the supervised ANN, being of the binary type (M = malignant, B = benign) or (1 = malignant, 0 = benign). The data set was separated by 90% for training the neural network and 10% for testing using K-fold (K = 10).

Table 1. Database attributes.

Characteristics	**Attributes**
01	id_number
02	diagnosis
03	radius_mean
04	texture_mean
05	perimeter_mean
06	area_mean
07	smoothness_mean
08	compactness_mean
09	concavity_mean
10	concave_points_mean
11	symmetry_mean
12	fractal_dimension_mean
13	radius_se
14	texture_se
15	perimeter_se
16	area_se
17	smoothness_se
18	compactness_se
19	concavity_se
20	concave_points_se
21	symmetry_se
22	fractal_dimension_se
23	radius_worst
24	texture_worst
25	perimeter_worst
26	area_worst
27	smoothness_worst
28	compactness_worst
29	concavity_worst
30	concave_points_worst
31	symmetry_worst
32	fractal_dimension_worst

2.2 Parameterization of the Artificial Neural Network - MLP

A MLP (Multi-layer Perceptron) neural network was created with the following fixed parameterization:

- activation = 'logistic', sigmoid activation function for the hidden layer;
- optimization = 'lbfgs', it is an optimizer for the weights of the network;
- batch = 1, which is the size of mini-batches for stochastic optimizers;

In the table 2 shows the parameters that varied in each test.

Table 2. Parameters of the neural network architecture used.

Test	**Layers**	**Neurons**	**Learning rate**	**Epocs**
1	2	150	10^{-8}	10000
2	2	150	10^{-10}	20000
3	2	150	10^{-12}	100000
4	20	500	10^{-12}	100000
5	50	1000	10^{-12}	100000
6	100	5000	10^{-12}	100000

3 Results and Discussion

The neural network was implemented in the Python language with the aid of the sklearn library to execute the MLP classifier and to use cross validation. The test computer's configuration is an AMD A10-4600M (2.30GHz) processor, 8GB DDR3 RAM and 256GB SSD, without a video card. To determine the best efficiency of the tested MLP, the cross-validation method with k-fold equal to 10 was used, where the database is divided into 10 parts and the test part changes 10 times to go through all the data. The final result of each test is found with the average of the accuracy after it has been performed 10 times. In addition, we have the duration of each test and the maximum and minimum value of each accuracy and duration found.

Figure 2 shows the results of all simulations performed with different configurations of neural network architectures. Observing that the majority of tests with the most adequate parameters and established with 4 k-folds obtained better performance of accuracy in the data, presenting up to 92%. In addition, the configurations of tests 4 and 6 remained unstable when the amount of k-fold was increased to 6. It is also observed that increasing the number of neurons and intermediate layers had an improvement in performance, but there was only a small significance in performance. The best results were presented in the configuration of test 6, where there were 100 hidden layers and 5000 neurons, presenting better performance in k-folds 2, 4 and 6, with performance of 90%, 92% and 92%, respectively.

Similar performance, were found in a study using convolutional neural networks, the models developed predict healthy women at rates of minimum 88.2% and maximum 94.1%; and in women with breast cancer the minimum rates were 88.8% and 94. 4% [8]. With an evaluation in 800 patients with a definitive diagnosis of cancer by biopsy, the proposed neural network showed a disease prediction rate of 90.5% and the health proportion was 80.9% [9].

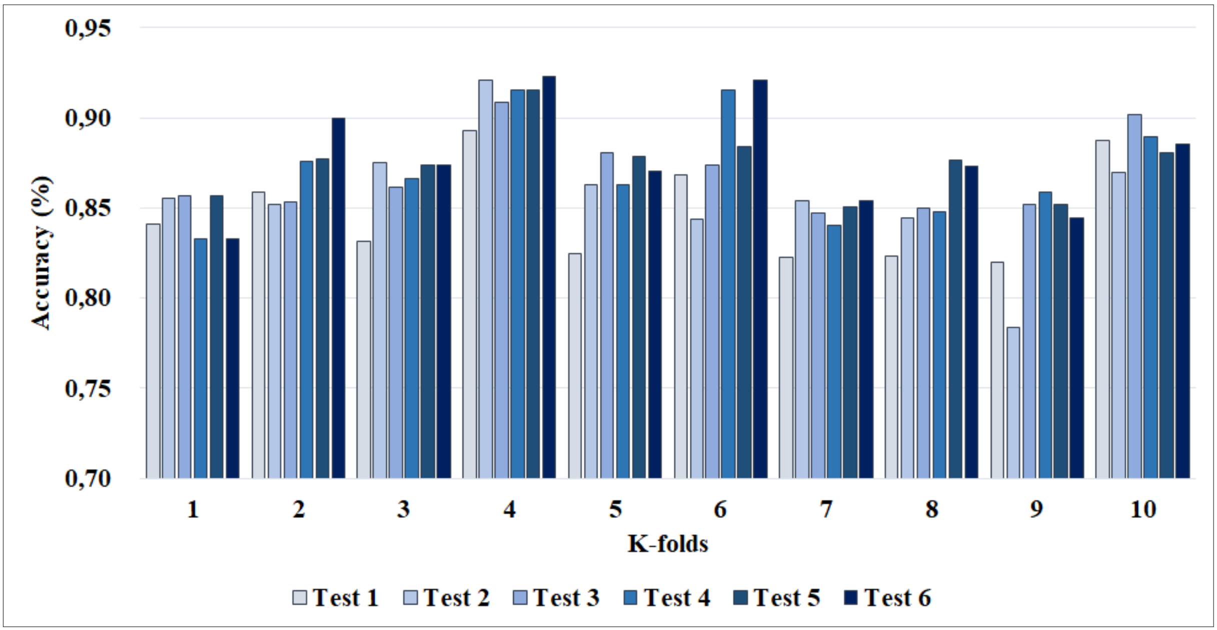

Figure 2. Average performance of the Neural Network in various tests and architectures

4 Conclusions

In this work, a multilayer perceptron artificial neural network algorithm was developed, capable of optimizing the classification as malignant or benign from the Breast Cancer Wisconsin database (Diagnostic) Data Set [7]. For this database and this ANN, the best results were presented in the configuration of test 6, where there were 100 hidden layers and 5000 neurons, presenting better performance in k-folds 2, 4 and 6, with performance of 90%, 92% and 92%, respectively, reaching the conclusion that in the MLP artificial neural network it is not interesting to increase the amount of neurons and hidden layers to improve the result, so it will only increase the complexity of the neural network.

Therefore, computing has become a great ally in the search for a quick and reliable diagnosis, offering intelligent and faster solutions, and thus, gaining more space in the search for a satisfactory result, assisting health professionals in seeking to predict risk breast cancer early.

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] R. M. Pfeiffer *et al.*, "Risk Prediction for Breast, Endometrial, and Ovarian Cancer in White Women Aged 50 y or Older: Derivation and Validation from Population-Based Cohort Studies," *PLoS Med.*, 2013.

[2] D. G. R. Evans and A. Howell, "Breast cancer risk-assessment models," *Breast Cancer Research*. 2007.

[3] G. F. Stark, G. R. Hart, B. J. Nartowt, and J. Deng, "Predicting breast cancer risk using personal health data and machine learning models," *PLoS One*, 2019.

[4] INCA. Instituto Nacional de Cancer José Alencar Gomes da Silva, *Estimativa 2016: incidência de câncer no Brasil*. 2016.

[5] K. Manikantan *et al.*, "Challenges for the future modifications of the TNM staging system for head and neck cancer: Case for a new computational model?," *Cancer Treatment Reviews*. 2009.

[6] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Computational and Structural Biotechnology Journal*. 2015.

[7] Breast Cancer Wisconsin (Diagnostic) Data Set. In: Breast Cancer Wisconsin (Diagnostic) Data Set. [S. I.], 2019 Disponível em: <https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29>. Acesso em: 19 nov. 2019.

[8] A. Karaci Predicting Breast Cancer with Deep Neural Networks. In: Hemanth D., Kose U. (eds) Artificial Intelligence and Applied Mathematics in Engineering Problems. ICAIAME 2019. Lecture Notes on Data Engineering and Communications Technologies, vol 43. Springer, Cham. https://doi.org/10.1007/978-3-030-36178-5_88

[9] I. Saritas, "Prediction of breast cancer using artificial neural networks," J. Med. Syst., 2012.