Performance Analysis of Artificial Intelligence Techniques applied in Breast Cancer Data set.

Manoel Eric N. de Oliveira
University of Ceará
R. Cel. Estanislau Frota, 563
Sobral CE, 62010-560
+55,Brazil
manoeleric59@alu.ufc.br

Jose Claudio do Nascimento
University of Ceará
R. Cel.Estanislau Frota, 563
Sobral CE, 62010-560
+55,Brazil
claudio.nasce@gmail.com

Felipe Barros Muniz
University of Ceará
R. Cel.Estanislau Frota, 563
Sobral CE, 62010-560
+55,Brazil
felipemuniz@alu.ufc.br

Ruann C. Farrapo
University of Ceará
R. Cel.Estanislau Frota, 563
Sobral CE, 62010-560
+55,Brazil
ruann.campos_01alu.ufc.br

ABSTRACT

In this work, a comparative study was carried out between two classification methods: The Multi layer Perceptron Artificial Neural Network (MLP ANN) and the method of classification of the Nearest Neighbors, used in the classification of the diagnosis of breast cancer. The data used in this work were taken from the UCI Machine Learning Repository and contains numerical data extracted from mammography images.In addition, the results were evaluated based on the cross-validation strategy.

Keywords

Breast Cancer; Artificial Neural Neural Network; Nearest Neighbors; Classification; Machine Learning.

1. INTRODUCTION

A form of early detection of breast cancer is conventional mammography, which consists of the analysis of images by radiologists and able to identify mammographic signs. Previous studies show that exhaustive analysis of mammography images in the same period of work can be passive of errors. This fact implies that the observer may end up making mistakes for showing interest in certain areas, making other areas go unnoticed [6].

The Brazilian mortality peripheral has undergone an intense change, changing from infectious-parasitic diseases to chronic-degenerative diseases, such as cancer[10].

Breast cancer represents the main cause of death by cancer in Brazilian women and in the world it gives way only to lung cancer, representing a major public health problem worldwide. In this way, countless actions have been thought of in various spheres, such as organizations, companies and corporations in general.

However, one way out was to use new technologies as an aid tool. Thus, artificial intelligence appears as a good alternative as an aid to radiologists, thus reducing the occurrence of human errors in the diagnosis of cancer [1]. Thus, the aid to computer diagnosis is a tool for health professionals evidence probabilistic estimates the occurrence of breast cancer in certain cases. There are several studies that use artificial intelligence techniques to aid in the diagnosis of the cancer. In the work of [14], it aimed to classify nodules present in digitized mammograms, through a Neural MLP Network. [17] already carried out a work of research studies, an application that classifies breast cancer from an MLP with an analysis of the breast contour.

Thus, several studies seek to classify the diagnosis of breast cancer by analyzing various aspects collected from specific mammography data.However, an alternative that has been applied in the academic and business fields, is to make use of existing technologies such as mammography exam machines, the images generated by them, and use machine learning in this function for a better use.

Thus, this work proposes to produce a set of experimental observations obtained through the application of a database on breast cancer in classification algorithms with the task of classifying breast cancer as malignant oy beningo. In this article we will use a database of UCI Machine Learning Repository that contains several mammography data attributes. Thus, based on the use of cross-validation and the aid of the MLP network and the KNN classifier, a performance analysis will be developed for each classifier topology, alternating parameters such as the number of hidden layers in the MLP network, the number of neurons and the number of neighbors in the KNN classifier.

The work is organized in 8 specific sessions. The introduction session aims to present the whole problem of breast cancer and the purpose of this work. Sections 2,3, 4 and 5, on the other hand, basically constitute the entire theoretical framework of this work. Section 2 presents the concepts of artificial neural networks, their use and organization. Section 3 presents the concepts and state of art about K-Nearest Neighbors.

In Section 4, on the other hand, is the part of this work that presents how the previous tools were applied in the database, that is, it constitutes the methodology of the proposal. Section 5 shows how the data was pre-processing. The section 6 shows how to data validated and how the cross-validation strategy was used. Finally, section 7 shows the results obtained with the implementation of these proposals and section 8 presentation the conclusions.

2. ARTIFICIAL NEURAL NETWORKS

Building an autonomous mechanism with intelligence is a longstanding wish of researchers in science and engineering. All this yearning expanded and became an area of study called machine learning [19]. Machine learning consists of providing a good amount of relevant and processed data for training this device to perform a task based on a mathematical statistical model [13]. The study of Artificial Intelligence has increased in recent years. The intention to create a machine capable of classifying, predicting and making decisions is a current reality.

One of the most used tools lately in artificial intelligence are artificial neural networks (ANNs), which consist of algorithms that try to imitate the structure and processing capacity of the human nervous system. Such a structure allows the resolution of several problems arising from the most varied areas of knowledge [18]. According to [4] Artificial Neural Networks they can classify and make predictions of cancer based on the individual's genetic profile. Thus, this work aims to use an artificial neural network to aid in the diagnosis of breast cancer.

ANNs are a mathematical model of data processing organized in units called nodes or neurons. The way the neurons are arranged together make up the architecture of the neural network. The topology of a network, on the other hand, consists of the different forms of structure composition that architecture can assume.The main architectures of neural networks are: feed-forward, recurrent and reticulated [12]. The figure 1 exemplifies a multi-layered feed forward architecture.

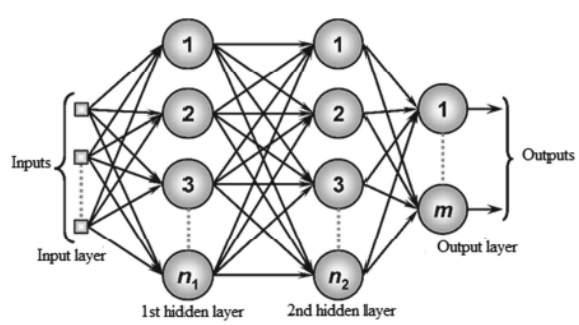

Figure 1: Multilayer feedforward architecture [19].

In this work, two methods were used for classification: a MLP-type ANN, which is a multilayer feedforward network proposed for the first time by [16] compared to the nearest neighbor classification method (KNN). The topology used in the MLP consists only of an intermediate layer of neurons that used the error back propagation method to adjust the weights.

3. K-NEAREST NEIGHBORS

The K-nearest neighbor (KNN) classification method has been an algorithm widely used in classification problems. This fact is noticeable, for example, when we look at the works of [2][3][5][20][22]. K-Nearest Neighbors is very simple method can generate increasingly competitive results, making it difficult to converge the algorithm. In addition, the performance of a classifier is mainly determined by choosing K, an attribute that corresponds to the number of neighbors to be analyzed. The most common form to be used is the form that was presented in [15]. Still, in the work of [7] it was shown that when the points are not evenly distributed, predicting the K value becomes a difficult task. In this bias, it is notable that the way to choose the parameters and the method to be used in the distance process directly influences the accuracy of the algorithm. In this sense the researchers try to show new approaches for example, Discriminant Adaptive NN [9] (DANN), Adaptive Metric NN [6] (ADAMENN), Weight Adjusted KNN [8] (WAKNN), Large Margin NN [13] (LMNN) and etc.

As in the case of the Breast Cancer database, the data can be distributed in a Linear way and the database does not have missing and scattered data, it was observed in practice that the Euclidean distance presented a good way to solve the problem described.

Summing up, KNN is a non-parametric algorithm where the structure of the model will be determined by the database. The algorithm basically works in 3 steps: find the distance, find the nearest neighbors and vote for the markers. These steps are illustrated in the figure 2

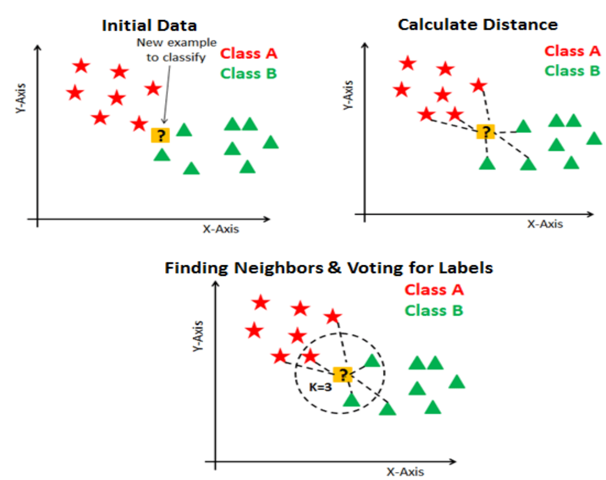

Figure 2: Operation KNN [7].

4. METHODOLOGY

This work deals with experimental research, which aims to compare two methods applied for classifying breast cancer. The method used consists of providing data for classifier algorithm , changing the settings and analyzing the results. The methodology of this work follows the flowchart shown in 3:

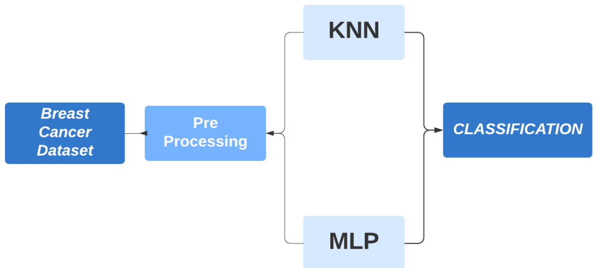

Figure 3: Methodology steps.

4.1 Data Base

A database available for free at the UCI Machine Learning Repository [8] was used, which contains resources calculated from a scanned image of a mammogram. Such data describe characteristics of the cell nuclei present in the image how to: id number, diagnosis, radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry and fractal dimension. The database has 569 instances and 32 attributes. The attributes are all numeric, except for the diagnostic output.The figure 4 shows the main characteristics of the database used in this work.

Data Set Characteristics:	Multivariate	Number of Instances:	569	Area:	Life
Attribute Characteristics:	Real	Number of Attributes:	32	Date Donated	1995-11-01
Associated Tasks:	Classification	Missing Values?	No	Number of Web Hits:	1301326

Figure 4: Breast Cancer Wisconsin (Diagnostic) Data Set (adapting of [8]).

The table 1 shows some examples of Ten real-valued features are computed for each cell nucleus:

Table 1: Attributes of database.

a) radius	mean of distances from center to points
b) texture	standard deviation of gray-scale values
c) perimeter	
d) area	
e) smoothness	local variation in radius lengths
f) compactness	perimeterˆ2 / area - 1.0
g) concavity	severity of concave portions of the contour
h) concave	number of concave portions of the contour
i) symmetry	
j) dimension	fractal dimension ”approximation” - 1

Before training, strategies had to be made so that problems such as non-converging and missing data would not cause an error in the process. For that, the data had to be normalized. Data normalization and data validation will be explained in the validation section later.

4.2 MLP training

Multilayer Perceptron ANNs (MLP) are neurons connected by connections synaptic cells that are divided into input neurons, which receive stimuli from the middle into internal neurons, responsible for making the neurons of the layers of entrance and exit; and in output neurons, which communicate with the outside [11]. An input signal xi at the input of a neuron I is multiplied by the synaptic weight wij and, after calculation, the value is sent to the input of neuron J. Each neuron J performs the sum of all signals applied to its input, according to equation (1), and apply to an activation function.

$$u = \sum w_{ij} x_i \tag{1}$$

The activation function used in this algorithm was the relu function, which has the following representation in equation (2):

$$f(x) = max(0, x) \tag{2}$$

The output y_j is equal to the value of the activation function given by equation (3):

$$y_j = f(u) \tag{3}$$

The algorithm used for learning in MLP is called descent from stochastic gradient. The stochastic gradient drop (SGD) updates the parameters using the gradient negative of the loss function in relation to a parameter that needs to be changed. Thus, as the gradient points to where the function is increasing, one seeks to walk in the opposite direction to maximize the solution [21]. The following equation (4) demonstrates how the SGD does to adjust the weights, minimizing the error.

$$E = \sum E^p \Leftrightarrow \sum_P (d_r - d_p)^2 \tag{4}$$

Where, E^P represents the error, d_r is the desired output and d_p is the output obtained. The partial derivative of the error is calculated. Subsequently, the descending gradient method is used to update the weights according to equation (5):

$$w_{ij}(t+1) = w_{ij}(t) + \frac{dE}{dw_{ij}} \tag{5}$$

In addition to the conventional error propagation method shown above, the momentum insertion technique was used. The term momentum is a device that aims to consider how much the synaptic weights have been changed between two consecutive interactions. such proposal aims that the algorithm is not stuck in local minimums thereby improving network efficiency. Equation (6) can be obtained by modifying equation (5) with the addition of the α variable, with a value between 0 and 1.

$$w_{ij}(t+1) = w_{ij}(t) + \alpha[w_{ij}(t) - w_{ij}(t-1)] + \frac{dE}{dw_{ij}} \tag{6}$$

In addition to including the term momentum, the algorithm used in this work was implemented with the strategy that if the error is not minimized between 3 consecutive iterations in 0.0001, the algorithm stop.

4.3 KNN trainig

The algorithm proceeds as follows: Given a query vector x_o and a set of N labeled instances $\{x_i, y_i\}_1^N$, he classifier's function is to predict the class label of $x0$ in the predefined P classes. The K-nearest neighbor (KNN) method tries to find the nearest neighbor to $x0$ and uses a kind of majority vote to determine the class label of $x0$.The most common and used form in KNN is to apply Euclidean distances as the distance metric as shown 7:

$$D_{pq} = \sqrt{\sum_{1}^{n}(p_i - q_i)^2} \tag{7}$$

5. PRE-PROCESSING

When analyzing a database and implementing a classification model it is a difficult task because of how the data is organized, absent or organized. For this, several strategies had to be taken before processing the algorithms. Some of these decisions can be identified in 2:

Table 2: Pre-processing strategy's .

Standardize column names
Convert all categorical variables to character
Remove target variables from the other task
Remove columns where all rows are empty
Remove rows where all columns are empty
Convert all attributes to numeric
Convert to ordinal categorical
Convert "not done" to NA (missing)
Convert logical to numeric (binary)
Remove numeric variables with cardinality less than 10
Sort categories with cardinality 1
Remove category with 1 level

5.1 Normalization

A strategy of normalization was used for avoiding computational problems. Data normalization occurs when in a large database we have many attributes of numerical values with many significant figures. Such occurrences can end up affecting the processing of the algorithms because they are repetition structures that will work in various periods of training, validation and testing.

The standardization of a data set is a common requirement for many machine learning estimators: they can behave badly if individual resources do not look more or less like normally distributed standard data. So, to make it easier, we can use the standardization strategy using StandardScaler. This method removes the average and the dimensioning of the unit, this results in data with an average equal to 0 and deviation equal to 1. The formulas for this process, follow below where they show the formula for standardization, mean and standard deviation:

$$z = \frac{x - \mu}{\sigma} \tag{8}$$

$$\mu = \frac{1}{N}\sum_{i=1}^{N}(x_i) \tag{9}$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \tag{10}$$

6. VALIDATION

For the validation of the algorithms, a cross-validation strategy was used.

The main difficulty in using classification algorithms is how to identify the best stopping point for training, as the training error tends to decrease according to the number of times of the algorithms used [11].

In addition, we may encounter a very common problem in Artificial Intelligence, which is the algorithm to decorate the data. In this way, the algorithm will give the false impression that it is being effective, when it is not what is happening. For this, seeking a better generalization of the classification algorithms, we use the cross validation strategy, where we partition the original database in sub intervals so that the algorithm is submitted to different data from the one previously trained, thus improving its generalization capacity [9].

A methodology for operating cross-validation is show in figure 5:

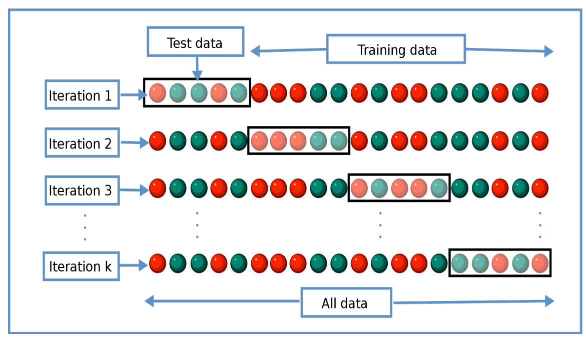

Figure 5: Cross Validation architecture [19].

The cross validation used divided the database into 5 splits, 75% being test and 25% for validation. In addition, the cross-validation was within a repetition structure that was repeated 10 times for each MLP topology and KNN classifier. Thus, at each compilation of the algorithm, we had a list with 50 results for network accuracy. The way the database was divided is illustrated in the table 3:

Table 3: Cross Validation in this work.

Validation:	25%
Training:	75%

The observation in this section is due to the fact that the division above perpetuates in all 5 splits of the cross-validation algorithm, which was repeated in a repetition structure that shuffled the data for each repetition, thus improving the generalization of the algorithm.

Accuracy can be calculated using equation (11) below, where VP corresponds to the positive truths, VN the true negatives, FP the false positives and FN the false negatives.

$$\frac{VP + VN}{VP + VN + FP + FN} \tag{11}$$

The following section will detail how these evaluation, standardization and validation metrics were used to build this work.

7. RESULTS AND DISCUSSION

7.1 MLP Results

The training of a neural network is divided into some periods, which are the times that the network executes the algorithm within a repetition structure. When finishing each loop, the network saves the weights obtained in the training of each stage, to later update them. Thus, after training, the network uses the weights obtained to train the part of the data that was not used for training, that is, the validation data.

Thus, when obtaining the predictions in the validation data, the network can show some important data such as, accuracy of the method, deviation in the data. a vector was created with 50 results for each split made without repeated K-fold.

At each iteration of the network for a given network architecture, a list was generated with all the accuracy of the training and validations of the part that was divided.Empirically, it was realized that the best organization would be the one with only an intermediate layer of neurons. Thus, successive tests were made for the amounts of neurons, where we varied from 1 to 100. The best results were used an intermediate layer and 36,38,40 and 50 neurons.

The table 4 show how the data mentioned above are associated with the evaluation metrics used in this work, so we use statistical variations such as higher value, lower value, average of values and standard deviation.

Table 4: Data of MLP training.

Number of Neurons	Mean Accuracy	Bigger Accuracy	Less Accuracy	Stan- standard Detour
36	0.9727	1.0	0.9385	0.0163
38	0.9722	1.0	0.9385	0.0158
40	0.9717	1.0	0.9298	0.0157
50	0.9724	1.0	0.9210	0.0160

The figure 6 showing the behavior of the network in each repetition of the cross validation for one layer and 40 neurons .

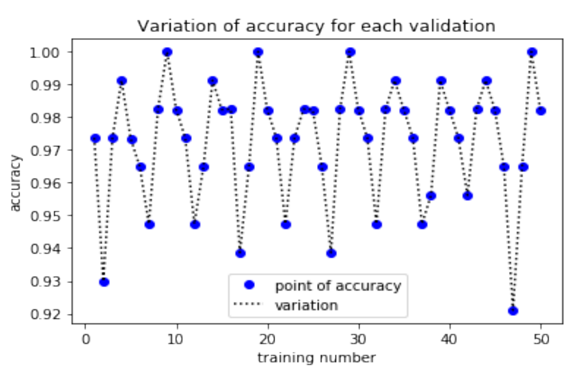

Figure 6: Graphic of MLP validation for 40 neurons.

A figure 7 that shows a matrix confusion for the accuracy 98.24561%

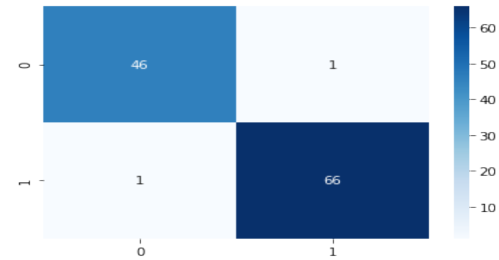

Figure 7: Matrix confusion for 40 neurons.

7.2 KNN Results

The training conducted by the KNN classifier basically follows the methodology used in MLP. However, the evaluation metrics are obtained with the variation of other parameters. As a cross-validation technique was used in this work, we obtained several results for accuracy, deviation, and algorithm precision. Unlike MLP, KNN works by drawing decision boundaries to classify data into a class. In this way, the borders are drawn according to the parameter k, which corresponds to the number of close neighbors to be analyzed. The following is an example of how the border regions are drawn, the graph shows how the KNN classifier chooses class 0 (malignant) 1 (benigno), with the exposure of the breast ray data.

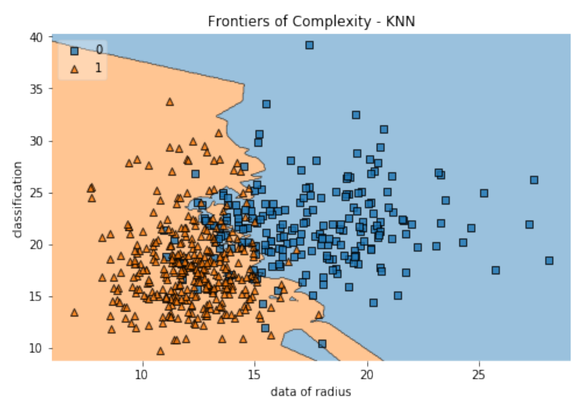

Figure 8: Graphic of border region.

Therefore, to improve the evaluation metrics and compare with the various structures that were used, the following procedure was used. At each KNN iteration for a given number of neighbors, a list was generated with all the accuracy of the training and validations of the part that was divided. For the KNN algorithm, a large variation in the number of neighbors was tested, following the form that $k = 2n + 1$. The behavior in decision regions was also observed to stipulate the parameters of the algorithm as figure 8 showed.

Thus, in an empirical way k between 1 and 100 were tested through a repetition structure that varied the k parameter and compared the relative average error in each k variation. Finally, it was noticed that the best results were obtained for k corresponding to 3,13,15,17 of according table 5:

Table 5: Data of KNN training.

Number of Neighbors	Mean Accuracy	Bigger Accuracy	Less Accuracy	Standard Detour
3	0.9366	0.9623	0.9373	0.0163
13	0.9596	0.9734	0.9473	0.0089
15	0.9578	0.9734	0.9385	0.0116
17	0.9543	0.9734	0.9298	0.0151

Next, we have a graph that illustrates the behavior of the KNN classifier for k = 13. The graph shows the accuracy in each validation of the cross-validation process, where the base was divided into 5 splits 10 times successively with the shuffling of the data for improvement the generalization of the algorithm. In this graph we can see how the classifier behaves differently from the MLP network, but these data will be discussed in the next section. Before, we must pay attention only to the spacing illustrated in the graph below, which will also influence the efficiency of the KNN method.

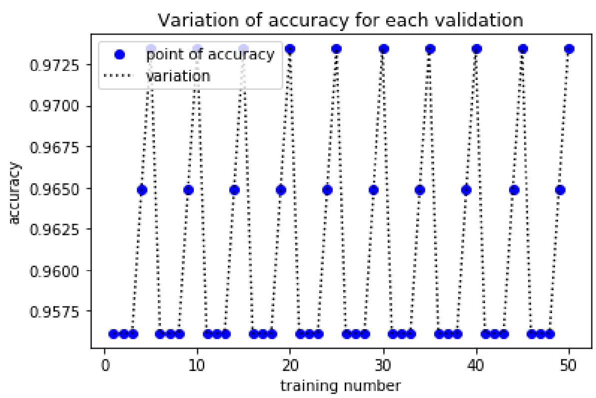

Figure 9: Graphic of KNN validation for 13 neighbors.

The figure 10 was used to show the spacing of the accuracy distributions in training for k = 13, a parameter that obtained the best results in the KNN classifier. We can see how the data are spaced, characterizing a poor convergence that will be explained in the next section.

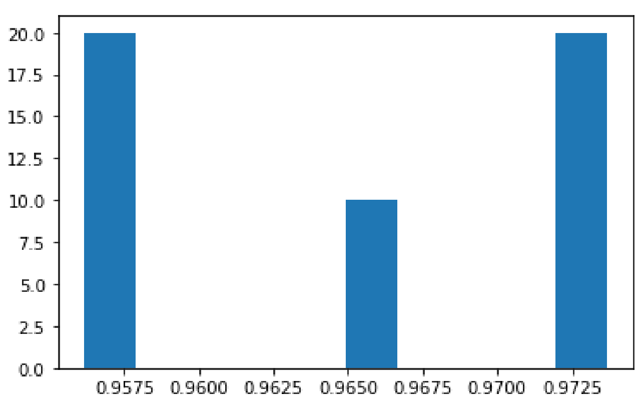

Figure 10: Histogram for 13 neighbors.

7.3 Discussion

Before discuss the problem, it's necessary show why don't was used Convolutional Neural Networks (CNN's). The non-use CNN's is fact why Data set its composed of literals data and alpha numerics already were extracted of breast image. Therefore, how work related computational vision already done , in this work is used strategies how to MLP and KNN for classification. To better discuss the data, use the following graph, which compares the best associations for each algorithm: MLP and KNN. Note that the two algorithms behave well after classification, however, note that the MLP network provided the best results. Although the KNN classifier presents low deviations from the standards, an average of them does not register an improvement with variation in the parameter of the number of neighbors.

In addition, we can observe that the KNN classifier graphically shows a generalization difficulty even in its best structure. It is observed that the algorithm falls to respective lows and takes some time to converge to a higher value, and the path is shown to be unreliable due to its minimal variation. On the other hand, the MLP neural network shows graphically that although it starts with a very low accuracy, it manages to evolve gradually to converge in higher values, and the decrease in the same way. Thus, it is shown that the MLP Neural Network shows that it is not as likely to fall in local minimums and its generalization capacity is more reliable than the KNN classifier.

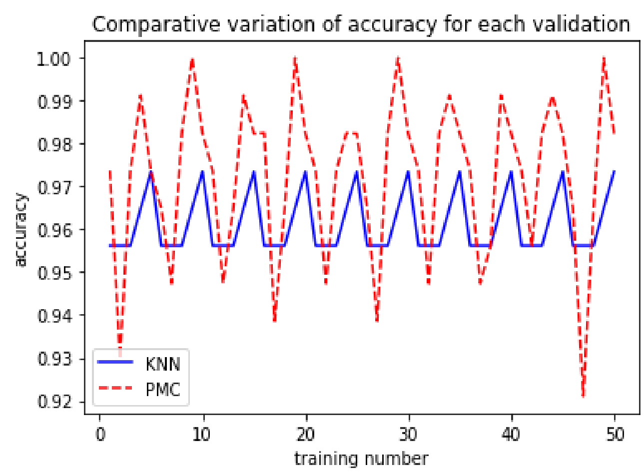

Figure 11: Comparative performance between MLP e KNN classification.

In summary, the two algorithms presented good alternatives as tools for the classification of breast cancer. However, MLP behaved in a better way, as its generalization capacity was more pronounced and reliable than the KNN method. Finally, it must be taken into account that the database is relatively small, and with that the methods may have changed due to the need for more parameters for a better classification. In the next section, we will discuss the other main conclusions observed in this work.

8. CONCLUSIONS

The purpose of this work was to compare two classification methods widely used in Machine Learning, in order to identify the most appropriate method for classifying the diagnosis of breast cancer of specific data set containing only two classes, in this case breast cancer being benign or malignant. Although there is a wide variety of methods for classifying pathology's currently available, the results of this work should serve not only for clinical uses, but also for future work in the field of computing applied to health that come to use a specific database.

When training, for example, on an artificial neural network, the biggest difficulty is how to organize the architecture of that network, for example, how to identify how many layers and quantities of neurons. Generally, these parameters are obtained with many tests carried out in an exhaustive manner that compromises the logistics of delivering a service or job. In this work it was observed that although the network topology and its respective attributes vary widely, there came a time when it did not make any difference to increase or decrease. This contribution becomes important in the academic and scientific scenario, since obtaining the topologies that provide the best results is a difficult task in this proposal.

Another difficulty with classification problems is choosing which type of algorithm to use in multi-instance and multi-class problems. Thus, in this work we present empirically that MLP Artificial Neural Networks behave better in the classification of breast cancer in that specific database than the KNN classifier. Such information can be used in future works on the same database, or in similar problems, since there are a multitude of pathology's with similar behavior. In summary, we concluded in this work that the MLP neural network method is more reliable to be used in the classification of breast cancer in this specific data set, considering that its results were superior in several aspects to the KNN classifier. However, each database has its peculiarity, in this case presented the MLP network was better than the KNN. However, in another database the result could be different. Therefore, this work can influence future research that aims to use and compare MLP and KNN in the diagnosis of other pathology's with similar behavior.

In the future, this work intends to expand and become a tool that can help the national health system, considering that breast cancer is a disease that haunts the daily lives of Brazilian women. However, many paths have yet to be followed and joint research between Medicine and Engineering must be developed to provide a broad collection for future researchers in view of the specifics of each problem and how the data is arranged and provided.

9. REFERENCES

[1] Y. Abdou, K. Attwood, T.-Y. D. Cheng, S. Yao, E. V. Bandera, G. R. Zirpoli, R. P. Ondracek, L. Stein, W. Bshara, T. Khoury, et al. Racial differences in cd8+ t cell infiltration in breast tumors from black and white women. *Breast Cancer Research*, 22(1):1–10, 2020.

[2] V. Athitsos, J. Alon, and S. Sclaroff. Efficient nearest neighbor classification using a cascade of approximate similarity measures. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 486–493. IEEE, 2005.

[3] V. Athitsos and S. Sclaroff. Boosting nearest neighbor classifiers for multiclass recognition. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops*, pages 45–45. IEEE, 2005.

[4] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber. Mitosis detection in breast cancer histology images with deep neural networks. In *International conference on medical image computing and computer-assisted intervention*, pages 411–418. Springer, 2013.

[5] T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27, 1967.

[6] P. R. Dellani and P. S. Borges. Implementação de um método para a classificação de microcalcificações pleomórficas invariante a posição, escala e orientação com rede neural de kohonen em mamografia convencional.

[7] C. Domeniconi, J. Peng, and D. Gunopulos. Locally adaptive metric nearest-neighbor classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9):1281–1285, 2002.

[8] D. Dua and C. Graff. UCI machine learning repository, 2017.

[9] A. M. Guimarães, I. M. Mathias, A. H. Dias, J. W. Ferrari, and R. d. O. Carlos. Módulo de validação cruzada para treinamento de redes neurais artificiais com algoritmos backpropagation e resilient propagation. *Publicatio UEPG: Ciências Exatas e da Terra, Agrárias e Engenharias*, 14(01), 2008.

[10] N. Haddad and M. D. SILVA. Mortalidade por neoplasmas em mulheres em idade reprodutiva-15 a 49 anos-no estado de são paulo, brasil, de 1991 a 1995. *Revista da Associação Médica Brasileira*, 47(3):221–230, 2001.

[11] S. Haykin. Redes neurais: princípios e prática, 2ª edição, tradução: Paulo martins engel. *Editora: Bookman, Porto Alegre, Cap*, 1(2):3, 2001.

[12] S. M. Jadhav, S. L. Nalbalwar, and A. A. Ghatol. Artificial neural network models based cardiac arrhythmia disease diagnosis from ecg signal data. *International Journal of Computer Applications*, 44(15):8–13, 2012.

[13] C. Kaliszyk, F. Chollet, and C. Szegedy. Holstep: A machine learning dataset for higher-order logic theorem proving. *arXiv preprint arXiv:1703.00426*, 2017.

[14] S. K. KINOSHITA, P. Marques, and A. Frère. Utilização da análise de componentes principais e redes neurais artificiais para a classificação de nódulos em mamogramas digitalizados. *Anais do CBEB*, pages 1214–1216, 2000.

[15] M. Latourrette. Toward an explanatory similarity measure for nearest-neighbor classification. In *European Conference on Machine Learning*, pages 238–245. Springer, 2000.

[16] J. L. McClelland, D. E. Rumelhart, P. R. Group, et al. Parallel distributed processing. *Explorations in the Microstructure of Cognition*, 2:216–271, 1986.

[17] P. B. Ribeiro. *Classificação por análise de contornos de nódulos mamários utilizando redes neurais*

artificiais. PhD thesis, Universidade de São Paulo, 2006.

[18] R. Rojas. Genetic algorithms. In *Neural Networks*, pages 427–448. Springer, 1996.

[19] I. d. Silva, D. H. Spatti, and R. A. Flauzino. Redes neurais artificiais para engenharia e ciências aplicadas. *São Paulo: Artliber*, 23(5):33–111, 2010.

[20] Y. Song, J. Huang, D. Zhou, H. Zha, and C. L. Giles. Iknn: Informative k-nearest neighbor pattern classification. In *European Conference on Principles of Data Mining and Knowledge Discovery*, pages 248–264. Springer, 2007.

[21] Z. Zeybek, S. Çetinkaya, H. Hapoglu, and M. Alpbaz. Generalized delta rule (gdr) algorithm with generalized predictive control (gpc) for optimum temperature tracking of batch polymerization. *Chemical engineering science*, 61(20):6691–6700, 2006.

[22] H. Zhang, A. C. Berg, M. Maire, and J. Malik. Svm-knn: Discriminative nearest neighbor classification for visual category recognition. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 2126–2136. IEEE, 2006.