Revista ARACÊ
ISSN: 2358-2472

PREDICTION OF CANCER INCIDENCE BY GENDER IN BELO HORIZONTE: INNOVATION IN THE USE OF ARTIFICIAL NEURAL NETWORKS FOR PUBLIC HEALTH PLANNING

https://doi.org/10.56238/arev7n3-289

Submitted on: 02/28/2025 **Publication date:** 03/28/2025

Celso Fabricio Correia de Souza[1], Yuzzo Iano[2], João Bosco Arbués Carneiro Junior[3], Celso Correia de Souza[4], Raul Asseff Castelão[5], Victor Rocha Pires de Oliveira[6], Gisela Maria Jorgino Crespo[7] and Juliana Maria Correia de Souza Vieira[8]

ABSTRACT
The study analyzed the incidence of cancer in Belo Horizonte from 2000 to 2020 with data provided by INCA, using the Multilayer Perceptron Neural Network (RNA_MLP) technique to predict cancer cases for the years 2021 to 2023. The city has a mostly female population

[1]Master in Environment and Regional Development (UNIDERP).
Professor at the Pontifical Catholic University of Campinas. Campinas – SP, Brazil.
E-mail: celsofabricio36@gmail.com
ORCID: https://orcid.org/0000-0002-4593-2798
LATTES: http://lattes.cnpq.br/2942671200096844
[2]Dr. in Electrical Engineering (UNICAMP).
Professor at the State University of Campinas. Campinas – SP, Brazil.
E-mail: yuzo@decom.fee.unicamp.br
ORCID: https://orcid.org/0000-0002-9843-9761
LATTES: http://lattes.cnpq.br/1819902045004845
[3]Dr. in Environment and Regional Development
Federal University of Rondonópolis
Email - jbacj@hotmail.com
ORCID: http://orcid.org/0000-0001-7094-2245
LATTES: http://lattes.cnpq.br/5073023761658236
[4]Dr. in Electrical Engineering
Anhanguera Uniderp University
E-mail: csouza939@gmail.com
ORCID: https://orcid.org/0000-0002-2689-8264
LATTES: http://lattes.cnpq.br/2881392515816773
[5]Dr. in Environment and Regional Development
Anhanguera Uniderp University
E-mail: raulassefcastelao@gmail.com
ORCID: http://orcid.org/ 0000-0002-4677-6088
LATTES: http://lattes.cnpq.br/1364663546445753
[6]Medical Oncologist
Casa Rosa Beneficent Association – Santa Casa de Campo Grande
E-mail: vrpiresoliveira@gmail.com
[7]Master's student in Sustainable Agribusiness
Anhanguera Uniderp University
Email: giselacrespo@icloud.com
ORCID: https://orcid.org/0009-0003-5947-4251
LATTES: http://lattes.cnpq.br/7460435585829383
[8]Master of Dentistry
Anhanguera Uniderp University
E-mail: : julianamaria2002@hotmail.com
ORCID: https://orcid.org/0000-0003-4478-9405
LATTES: http://lattes.cnpq.br/3801758547630854

Revista ARACÊ
ISSN: 2358-2472

(53.5%), which directly reflects the higher incidence of cancer among women, especially breast cancer. The analysis of the correlation between the population and the incidence of cancer revealed strong associations, with a correlation of 0.93 for male cases and 0.95 for female cases, in addition to 0.98 between cases in both genders, evidencing an almost synchronized growth pattern in cancer cases for both sexes. The performance of the RNA_MLP was evaluated based on quadratic errors, presenting a sum of squared errors of 1.901 in the training and 0.299 in the tests, indicating a good fit of the model. Relative errors were also lower in the tests, with 7.8% for the general model and 7.8% for both sexes. Forecasts indicated a continuous increase in cancer incidence between 2021 and 2023, with an upward trend for both males and females, reflecting not only the population increase, but also a possible improvement in the detection of the disease. Despite being a prediction based on estimates, the study highlighted the importance of the model to assist in the planning of public health policies and prevention strategies, considering the impact of cancer on the city's public health.

Keywords: Machine learning. Predictive methods. Forecast of incidence. Public health.

INTRODUCTION

The incidence of cancer has become a growing concern in several parts of the world, including Brazil, driven by factors such as population growth and longevity (JEMAL *et al*., 2011). Although improvements in diagnostics make it possible to detect cases that would previously have gone unnoticed, the main concern is the urgent need for investments in public health to keep up with this increase (DUFFY *et al*., 2013). The aging of the population has contributed to the growth of incidence, especially among the elderly, which makes the planning of public health policies even more essential (SOBRAL et al., 2022; DePINHO, 2000).

Currently, cancer research has expanded in the world and in Brazil, with a growing emphasis on the use of advanced technologies for prediction and treatment (FIOCRUZ, 2024; BARIONI et al., 2024). Computational models, such as Artificial Neural Networks (ANNs), have been a promising tool, as they allow you to simulate patterns and make predictions based on large volumes of data, such as cancer incidence histories. In Brazil, the use of these technologies is still a developing field, but with great potential to transform the way public health managers deal with cancer-related issues (BARRETO *et al*., 2018).

In Belo Horizonte, for example, there are variations in the growth of cancer cases between men and women, with emphasis on types such as breast, prostate, lung, and colon cancer, which have high incidence rates (INCA, 2024). The situation requires more effective actions in terms of prevention, treatment, and resources to meet this growing demand, as well as adequate planning to prevent the increase in incidence from further burdening the public health system.

The central question of this study is to investigate whether, in Belo Horizonte, there are significant differences in cancer incidences between genders, considering historical trends in population increase, longevity and diagnoses. The study aims to analyze whether the incidence of cancer in this city presents significant differences according to gender, in the period from 2000 to 2020. In addition, a prediction will be made on the incidence of cancer for the period from 2021 to 2023, using an Artificial Neural Network Multilayer Perceptron (RNA_MLP). The focus will be on assessing incidence trends for both males and females in order to identify patterns and provide accurate estimates for the future.

Thus, the objective is to understand how variations in incidence by gender, as well as trends observed over time, can contribute to the formulation of more targeted and effective public health policies, in addition to enabling better planning of health resources in

Revista ARACÊ
ISSN: 2358-2472

the future. This study also seeks to fill gaps in the understanding of gender differences in the prevalence of certain types of cancer, providing information for the improvement of public health policies in the city and, eventually, in other contexts.

It is expected that the results will provide an accurate predictive model with low errors that can be used by public managers to better allocate resources and prevention and treatment strategies. In addition, the study opens doors for the use of other machine learning technologies in public health, driving future research in this field (SAÚDE PÚBLICA, 2021).

In summary, the present study seeks to contribute to the strategic planning of public health in Belo Horizonte, proposing an innovative tool to predict the incidence of cancer and, thus, assist in decision-making to cope with this disease. The use of neural networks in the study of cancer incidence can contribute and guide how data can be used to anticipate cancer treatment as soon as possible, in addition to providing a model that can be applied in other cities and national and international contexts. The continuity of this research can open the way for the implementation of more advanced and accurate models, boosting the use of artificial intelligence, contributing to the well-being of society.

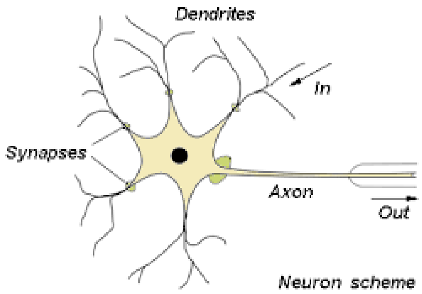
THEORETICAL FRAMEWORK

Artificial Neural Networks (ANNs) are a relevant area of Artificial Intelligence (AI), inspired by the functioning of the human brain, with the ability to learn, adapt and make decisions. Its use is broad, covering complex tasks such as pattern recognition, time series forecasting, and real-time decision-making (HAYKIN, 2009).

According to Kovácz (2006), ANNs are formed by processing units called artificial neurons, interconnected by synaptic weights that determine the strength and direction of the inputs' influences. The functioning of an artificial neuron is inspired by the biological model, with the difference that it has a simplified and mathematically more efficient form. Biological neurons, figure 1, are composed of dendrites, cell body, and axon, with the transmission of electrical impulses occurring through synapses (BARBER, 2012).

Revista
ARACÊ
ISSN: 2358-2472

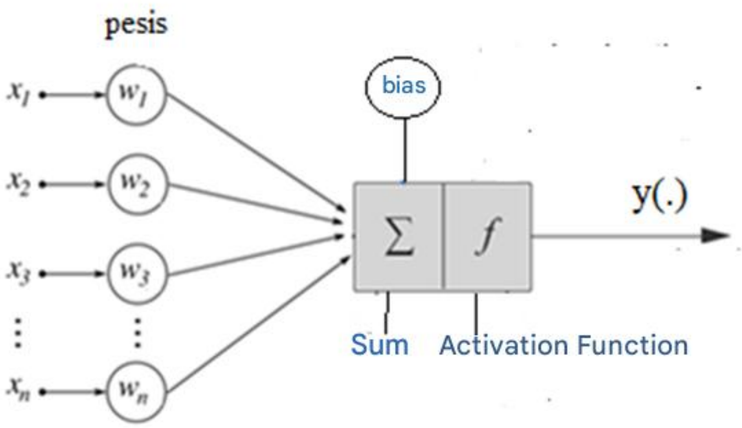
Figure 1. Representation of a biological neuron and its main components.



Source: Barber (2012).

In the 1940s, McCulloch and Pitts proposed the mathematical model for the artificial neuron, which was fundamental for the development of RNAs. This model consists of input signals, synaptic weights, a linear combinator, and an activation function. The input signals are weighted by the weights, added together, and then go through the activation function, which determines whether the neuron "fires" or not. This process introduces nonlinearity, allowing ANNs to learn and perform complex tasks. Figure 2 represents the model that was the basis for modern neural networks and continues to be an important reference in artificial intelligence (BARBER, 2012).

Figure 2. Schematic of an artificial neuron with the entry and exit points and their activation functions.



Source: Adapted from Barber (2012).

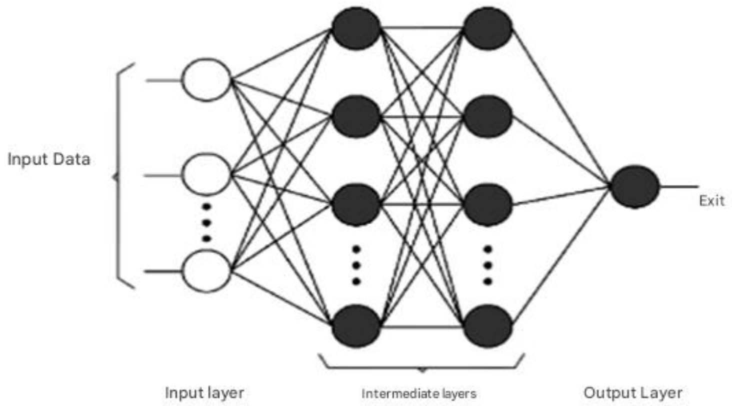
The artificial neuron is the fundamental unit in neural networks, receiving inputs that are multiplied by their synaptic weights and added with a bias. The resulting sum goes through an activation function that transforms the obtained sum into an output, which will decide whether the neuron should be activated or not. Activation functions, such as hyperbolic tangent, are essential to allow the network to capture nonlinear relationships in the data. The output of the neuron is then transmitted to the next layer or, in the case of the

Revista ARACÊ
ISSN: 2358-2472

last layer, to the final prediction. Learning occurs by optimally adjusting weights and biases during training. This process makes it possible for neural networks to solve complex problems, such as pattern recognition and trend prediction. Thus, neurons work together to perform advanced data analysis tasks.$x_1, x_2, \cdots, x_n w_1, w_2, \cdots, w_n \;\; \sum \;\; , f y$

ANNs can be structured in different ways, including feedback and recurrent networks, with variations in their topologies, such as single-layer and multilayer networks. Among these, the Multilayer Perceptron (MLP) is one of the most common and widely used to solve complex problems, such as the analysis of large volumes of data, nonlinear patterns, and trend predictions, as illustrated in figure 3.

Figure 3. Setting up a multilayer artificial neural network (MLP), showing the input layer, two hidden layers, and the output layer.



Source: Adapted from Barber (2012).

The equation of an ANN is fundamental to understand how these computational structures learn and adapt. Its mathematical modeling involves describing the algorithms that govern the transmission of information between artificial neurons. Through activation functions and training algorithms, such as *backpropagation*, it seeks to adjust the weights of the connections, minimizing errors and improving accuracy. These training processes, which include parameter optimization, are essential to ensure that ANNs can generalize and perform effective predictions.

According to Haykin (2009), the mathematical study of ANNs provides an insight into the behavior and efficiency of these systems, as each artificial neuron performs a linear combination of inputs, followed by an activation function to determine its output. The general equation for a neuron with inputs is given by equation (1).n

Revista ARACÊ
ISSN: 2358-2472

$$y = f(\sum_{i=1}^{n} x_i w_i + b) \quad (1)$$

Where: are the entrances to the neuron; are the weights associated with each entry; it is the bias (bias), which adjusts the weighted sum and; is the activation function, which turns the weighted sum into an output. x_i $w_i b$ f

The purpose of the activation function is to introduce nonlinearities, allowing the network to learn complex patterns. Activation functions determine how neuron output is calculated. Examples of activation functions are equations (2) and (3).

Sigmoid:

$$f(x) = \frac{1}{1+e^{-x_i}} \quad (2)$$

This function produces an output between 0 and 1, which is often used in binary classification tasks.

Hyperbolic tangent (tanh):

$$f(x) = tanh(x_i) = \frac{e^{x_i} - e^{-x_i}}{e^{x_i} + e^{-x_i}} \quad (3)$$

The output of this function ranges between -1 and 1, helping to centralize the data around zero.

DIRECT SPREAD

Direct propagation in a neural network refers to the process of calculating the outputs of each layer of the network from the inputs, following a sequence of mathematical transformations. This process, equation (4), involves applying activation functions and using the connection weights, allowing data to move from one layer to the next until the final output is generated.

$$z^{(l)} = W^{(l)} a^{(l-1)} + b^{(l)} \quad (4)$$

Where: $z^{(l)}$ is the weighted sum vector of layer l; It is the matrix of weights of layer L; it is the activation of the previous layer and; is the layer bias vector $W^{(l)}$ $a^{(l-1)}$ $b^{(l)} l$

Activation is calculated by equation (5).

$$a^{(l)}=f(z^{(l)}) \quad (5)$$

Revista ARACÊ
ISSN: 2358-2472

Where is the activation function applied to each element of $fz^{(l)}$.

MEAN SQUARE ERROR (MSE) FUNCTION

The mean square error (MSE) function for a neural network, used in regression problems, is calculated considering m examples and the difference between the actual outputs and the predicted outputsy_i $\hat{y}_i$. The MSE equation (6), measures the mean of squares of the differences between the observed and estimated values, providing a metric that quantifies the accuracy of the model in predicting expected outcomes. The lower the MSE value, the better the neural network adjusts to the training data. y_i $\hat{y}_i$

$$MSE = \frac{1}{2m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 \tag{6}$$

Where: are the real values; are the forecasts of the network and; is the number of examples in the dataset. y_i $\hat{y}_i$ mThe factor 2 in the formula is introduced to simplify the calculation of the gradient. When calculating the gradient of the error function with respect to the network weights during training, the square derivative of a number brings a factor of 2. That is, by dividing the MSE by 2, you avoid having this extra 2 when calculating the derivative, which makes the expression of the gradient simpler and avoids multiplying by 2 during the process of updating the weights.

BACKPROPAGATION

To optimize the matrix of the W weights and b. biases, the gradients of the E error are calculated in relation to each parameter using the backpropagation algorithm. The update of the parameters for the weights and bias is done using the descending gradient, equations (7) and (8).

$$W^{(l)} = W^{(l)} - \eta\frac{\delta E}{\delta W^{(l)}} \tag{7}$$
$$b^{(l)} = b^{(l)} - \eta\frac{\delta E}{\delta b^{(l)}} \tag{8}$$

Where: η is the learning rate; $\partial E/\partial W^{(l)}$ is the gradient of the error with respect to the weights e;

$\partial E/\partial b^{(l)}$ is the gradient of the error with respect to the biases.

Gradients are calculated using the derivative chain (chain rule), starting with the output layer and working backwards to the previous layers.

Revista ARACÊ
ISSN: 2358-2472

Gradient Descent

The update of the weights and biases in the network is performed using the descending gradient, where the idea is to minimize the error function E, equation (9).

$$\theta = \theta - \eta \nabla_{\theta} E \qquad (9)$$

Where: θ can be a weight or bias; is the gradient of the error function with respect to $\nabla_{\theta} E$θ e; η\is the rate of learning.

SUMMARY OF COMPLETE PROPAGATION TO NEURONSn

The general function for an RNA with L-layers, where the L-layer has neurons is given by the following algorithm of the propagation process:n_l

1. **Input**: The input (the first layer) is passed through the network.a^l
2. **For each layer l (from 1 to L)**:
   - Calculate pre-activation $z^{(l)} = W^{(l)} a^{(l-1)} + b^{(l)}$
   - Calculate activation $a^{(l)} = f(z^{(l)})$
3. **Output**: The last layer provides the final prediction.$\hat{y} = a^{(L)}$

This process is performed both during direct propagation and in the calculation of the gradient during the training process (backpropagation), with the aim of minimizing the MSE error function.

The backpropagation algorithm, a supervised learning technique, performs the update of synaptic weights through two phases: forward and backward. The forward phase calculates the output based on the inputs and updates the weights, while the backward phase adjusts these weights based on the calculated error between the predicted output and the desired output (SAFI and BOUROUMI, 2011). The RNA_MLP can be trained according to these principles by applying quadratic error to adjust the weights until the model performs satisfactorily (SMITH, 2018).

The application of ANNs also extends to financial forecasting, especially in stock price forecasting. As noted by Arun and Venkatalakshmi (2017), ANNs are efficient in dealing with the dynamics and non-linearity of financial markets. They are able to process large volumes of data and identify complex patterns, which makes them suitable for predicting market behavior. However, ANNs also have limitations, such as the difficulty of

Revista ARACÊ
ISSN: 2358-2472

interpreting the results due to their "black box" nature and sensitivity to data quality (YUVARAJ *et al.*, 2019).

METHODOLOGY

This section describes the methodology used in the research, which is of an applied and experimental nature, focusing on computer simulations by means of ANNs. The main variables involve the population of Belo Horizonte and cancer incidence segmented by gender, male and female, and the year of the series. The network configuration, training parameters such as the learning rate and number of iterations, as well as the error function, were also considered. Data collection was carried out from official sources of the National Cancer Institute (INCA), with an emphasis on the most common types of cancer among men and women in the city of Belo Horizonte (INCA, 2024). These data were chosen because they are, from the authors' point of view, the most complete and detailed about the disease available in Brazil.

The analyses carried out included the evaluation of cancer prevalences by gender, as well as the analysis of Pearson's correlations between cancer incidences in the male and female populations. For the analysis of the results, metrics such as mean square error, accuracy, and training time were used. The model was trained using backpropagation and gradient descent, and the performance evaluation was performed with training and test data. The results were presented through graphs and tables, with the objective of understanding the relationship between the parameters of the ANNs and the performance in supervised learning (BRAGA *et al*., 2014; HAYKIN, 2009).

The prediction methodology with ANNs, such as the one used in the MLP model, is based on a quantitative approach, with data collection, pre-processing, and division into training, validation, and testing sets. The RNA_MLP is then trained using the backpropagation method, adjusting their weights iteratively. After training, the performance of the network is evaluated using metrics such as the Mean Absolute Percentage Error (MAPE), which evaluates the accuracy of the forecast by comparing the actual and predicted values (MONTGOMERY *et al*., 2008; RIAHI *et al*., 2013).

The RNA_MLP methodology was applied to predict the number of cancer deaths in the city of Belo Horizonte, Brazil, from 2021 to 2023, with the analysis of the results. For the implementation of MLP RNA for the prediction of cancer mortality, the Statistical Package for the Social Sciences (SPSS) software was used, which was essential to adjust

Revista ARACÊ
ISSN: 2358-2472

the network parameters and obtain the predictions of mortality rates. The training process, parameter adjustments, and validation of results are important steps to ensure the effectiveness of the neural network (IBM-SPSS, 2024). While neural network techniques have shown promise in healthcare, the accuracy of the prediction model depends on the quality of the data and the network's ability to learn from observed historical patterns.

The application of the RNA_MLP method using IBM-SPSS made use of the following algorithm:

i) Data collection: Annual data on cancer deaths were obtained from public sources;
ii) Data pre-processing: The data has been adjusted to ensure an appropriate format for neural network input, including the removal of missing data and normalization;
iii) Data division: Data were divided into training, validation, and test sets;
iv) Construction of the RNA_MLP model: A RNA_MLP with hidden layers was developed, and the parameters were adjusted during training;
v) Network training: The model was trained using the training set, adjusting the weights to minimize error;
vi) Evaluation and adjustment: Performance was evaluated with the validation set, using metrics such as mean square error;
vii) Testing and prediction: The fitted model was tested and used to perform predictions of the number of deaths, and;
viii) Analysis of the results: The predictions were analyzed and patterns and limitations of the model were identified.
ix) Improvement of the model: Based on the analyses, adjustments were made to the model without changing its architecture.

The quality of the fit was evaluated by the Mean Absolute Percentage Error (MAPE), which measures the accuracy of the model, according to equation (10).

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{R_t - P_t}{R_t}\right| \times 100\ (\%) \quad (10)$$

Where N is the number of periods used, it is the actual observed value of the variable and is the predicted value of the variable by RNA_MLP. $R_t P_t$

Revista ARACÊ
ISSN: 2358-2472

According to Carneiro Junior *et al.* (2024), lower MAPE values indicate better performance of the prediction model, as shown in chart 1 of error acceptance levels, where MAPE < 5% is considered excellent.

Table 1. Error acceptance ranges for the differences between the observed values and predicted values of the MAPE.

MAPE Range	Interpretation
MAPE < 5%	Excellent – High accuracy
$5\% \leq MAPE < 10\%$	Good – Moderate accuracy
$10\% \leq MAPE < 20\%$	Regular – Reasonable accuracy
$MAPE \geq 20\%$	Poor – Low accuracy

Source: Carneiro Junior *et al*. (2024).

RESULTS AND DISCUSSION

Chart 2 presents the general data on the incidence of cancer, by gender, in Belo Horizonte, based on data provided by the National Cancer Institute (INCA), in the period from 2000 to 2020.

Table 2. Population of Belo Horizonte (Population), Cancer incidence, by gender (Male and Female), in the city of Belo Horizonte, from 2000 to 2020.

Year	Population	Male	Female
2000	2238526	2798	3169
2001	2246819	3263	3850
2002	2255292	3334	3721
2003	2264021	3473	3663
2004	2273011	3582	3604
2005	2282348	3415	3799
2006	2291969	3497	3934
2007	2301885	3914	4452
2008	2312029	3726	4468
2009	2322458	4508	4956
2010	2375151	4553	5591
2011	2384246	4352	5278
2012	2393918	4982	5587
2013	2403754	4305	5435
2014	2413813	5023	5535
2015	2424051	5163	5776
2016	2434472	4871	5744
2017	2445059	6070	7093
2018	2455879	6086	7078
2019	2466901	6536	7609
2020	2478184	5364	6239

Source: INCA (2024).

Revista ARACÊ
ISSN: 2358-2472

According to the IBGE (2025), the city of Belo Horizonte is composed of 53.5% women and 46.5% men, reflecting a slight female predominance in the population. On the other hand, this demographic factor can influence the analysis of the incidence of diseases such as cancer, since the absolute number of women in the city is higher than the absolute number of men.

Chart 2 presents data on the population (Population) and the incidence of cancer, by gender (Male and Female), in the city of Belo Horizonte, in the period from 2000 to 2020. The figures in Table 2 show an upward trend in population and cancer incidence between 2000 and 2020, with a consistently higher number of cases among women. This increase can be attributed, in part, to the higher proportion of women in the city, suggesting that, while IBGE data help to contextualize the demographic distribution of the population, INCA data (2024) show the higher incidence of cancer among women, reflecting a combination of demographic and public health factors.

Initially, the correlation matrix between the variables "Population", "Male" and "Female" in Table 2 was calculated with the objective of verifying the association between these variables, obtaining Table 1. Correlation analysis seeks to identify the intensity and direction of the relationships between these three variables over the years 2000 to 2020.

Table 1. Correlation matrix between the variables Population and cancer incidences by gender (Male and Female) in the city of Belo Horizonte, from 2000 to 2020.

Variables	Population	Male	Female
Population	1		
Male	0.93	1	
Female	0.95	0.98	1

Source: Authors (2024).

The correlation between the variables Population and Male is 0.93, while the correlation between Population and Female is 0.95, indicating that variations in one of these variables tend to follow variations in the other. The correlation between the variables Male and Female is 0.98, suggesting that the variations in cancer cases in both genders are almost perfectly synchronized. This demonstrates that, throughout this period, the city's population structure maintained a balance in cancer incidences between genders.

As the objective of this study was to predict the number of future cases of cancer incidences by gender, for the period from 2021 to 2023, a trained RNA_MLP with the historical data in chart 2 was used. The SPSS software was used as a statistical tool to

Revista ARACÊ
ISSN: 2358-2472

make this prediction. In the application of the RNA_MLP 24 cases were considered, of which 21 were valid, 16 (76.2%) for training and 5 (23.8%) for tests. The remaining three cases were used for predictions. The distribution of the data aims to ensure that most of the sample is used for network training, while a smaller part is reserved for the evaluation (tests) of the model.

The configuration of the RNA_MLP used in the analysis of cancer incidence in Belo Horizonte focused on the variable "Population" as a covariate, while the dependent variables were the genders "Male" and "Female". The network's entry layer has a unit, which represents the "Population", with standardized rescheduling. In the hidden layer, there is a single layer with four units and the activation function is the hyperbolic tangent, which is often used to model nonlinear relationships between the inputs and outputs. In the output layer, the two units represent the dependent variables, "Male" and "Female", with standardized rescheduling. The activation function for this layer was identity, which is appropriate for regression problems. The error method adopted was the sum of the MSE squares, used to minimize the difference between the predicted values and the actual values during the network training. The network structure, with its layers and functions, aims to capture the complexities and patterns in cancer-related data, segmented by gender.

Table 2 summarizes the errors in the performance outputs of the training and testing model.

Table 2. Summarization of the output errors of the performance of the RNA_MLP.

<table>
<tr><td rowspan="6">Training</td><td colspan="2">Sum of Squares (MSE)</td><td>1,901</td></tr>
<tr><td colspan="2">Average Overall Relative Error</td><td>0,127</td></tr>
<tr><td rowspan="2">Relative error for Scale dependent</td><td>Male</td><td>0,145</td></tr>
<tr><td>Female</td><td>0,108</td></tr>
<tr><td colspan="2">Stop rule used</td><td>1 step with no reduction in errors</td></tr>
<tr><td colspan="2">Training time</td><td>0:00:00,02</td></tr>
<tr><td rowspan="4">Tests</td><td colspan="2">Sum of Squares (MSE)</td><td>0,299</td></tr>
<tr><td colspan="2">Average Overall Relative Error</td><td>0,078</td></tr>
<tr><td rowspan="2">Relative error for Dependents of Scala</td><td>Male</td><td>0,078</td></tr>
<tr><td>Female</td><td>0,078</td></tr>
</table>

Source: Authors (2025).

Table 2 presents data on the performance of RNA_MLP, training and testing. During training, the sum of the quadratic errors was 1.901, which indicates a considerable margin

Revista
ARACÊ
ISSN: 2358-2472

of error, considering that it is an absolute value, but without normalization, this value alone does not provide a clear notion of the magnitude of the error in relation to the total values. The average overall relative error in training was 0.127 (12.7%), suggesting a reasonable performance.

For the data segmented by sex, the male relative error was 0.145 (14.5%), while the female error was 0.108 (10.8%), showing that the model performed slightly better for women. The model used its stop rule based on "1 consecutive step with no decrease in error", which means that training was stopped as soon as there was no further improvement. The total training time was extremely short, at just 0.02 seconds.

In the set of tests, the sum of quadratic errors was much smaller, 0.299, suggesting that the model generalized effectively. The average overall relative error in the tests was 0.078 (7.8%), indicating good performance. In addition, the relative errors for the male and female groups were identical, both 0.078 (7.8%), showing that the model had a balanced performance between the sexes.

Overall, the model performed well, with fewer errors during testing compared to training and a balance between the groups analyzed. Table 3 shows the predicted values (calculated using the SPSS software) for the three years ahead of the forecast (2021, 2022, and 2023).

Table 3. Observed values of the Population and values predicted by the RNA_MLP on the incidence of cancer in the population of Belo Horizonte, by gender (Male and Female), in the period from 2021 to 2023.

YEAR	OBSERVED VALUES*	EXPECTED VALUES**	
	Population	Male	Female
2021	2.490.783	6.135	7.211
2022	2.503.527	6.260	7.362
2023	2.516.461	6.374	7.501

Source: *IBGE (2024); **Authors (2024).

The values presented in table 3 show an increase in the population of Belo Horizonte and an upward trend in the incidence of cancer in Belo Horizonte between 2021 and 2023, both for males and females, with a gradual growth in the number of predicted cases each year. This pattern may reflect a possible increase in the incidence rate or a greater awareness and diagnosis of the disease in the population over the period. While the model predicts these increases, it is important to note that this data is only estimates based on RNA_MLP, and there is no actual data available for analyzing the accuracy of the

Revista ARACÊ
ISSN: 2358-2472

predictions. In addition, the lack of data on cancer incidence for the years 2021 to 2023 limits comparative analysis and model validation.

CONCLUSION

The analysis of the incidence of cancer in Belo Horizonte, between 2000 and 2020, revealed a significant increase in diagnosed cases, reflecting population growth and improvement in early detection of the disease. The strong correlation between male and female incidences, with an index of 0.98, suggests that both groups follow similar growth patterns, possibly influenced by common demographic and environmental factors.

The values presented in Table 3, which show both the observed values and those predicted by the RNA_MLP for the years 2021 to 2023, indicate an increasing trend in the population of Belo Horizonte and in the incidence of cancer, both for males and females. The increase in the number of cases predicted each year, with values for 2021, 2022 and 2023 reaching 6,135, 6,260 and 6,374 for men, and 7,211, 7,362 and 7,501 for women, may reflect a possible increase in the incidence rate or greater awareness and diagnosis of the disease.

While forecast data provides insight into future trends, the lack of actual data for these years prevents accurate validation of forecasts. However, the good performance of the model, demonstrated by the low relative errors, suggests that RNA_MLP is an effective tool for predicting cancer incidence. These results can be used for the planning of public policies and prevention strategies, helping to optimize public health management, especially in the fight against cancer. The research thus highlights the importance of predictive modeling to anticipate trends and guide health and policy decisions.

Revista ARACÊ
ISSN: 2358-2472

REFERENCES

1. Arun, T., & Venkatalakshmi, K. (2017). Stock market prediction using artificial neural networks with multilayer perceptron. International Journal of Computer Science and Mobile Computing, 6(1), 194–199.

2. Barber, D., Sollich, P., & Saad, D. (2012). Finite size effects in on-line learning of multi-layer neural network. In S. W. Ellacott, J. C. Mason, & I. J. Anderson (Eds.), Mathematics of neural networks: Models, algorithms and applications (pp. 84–88). Springer Science.

3. Barioni, C. T. S., Wandresen, R. P. de B., Pereira, L. F., Coimbra, A. F., Kubo, B. B. de A., & Cunha, R. C. da. (2024). Artificial intelligence for the identification of biomarkers in cancer prevention and diagnosis: Advances and perspectives. Revista Brasileira de Cancerologia, 70(2), e-254692.

4. Barreto, R. G., Marinho, G. M. G., Barreto, G. F. M., Barreto, R. G., Aversari, L. O. C., & Dantas, B. L. (2018). Utilizando redes neurais artificiais para o diagnóstico de câncer cervical. Revista Saúde e Ciência Online, 7(2), 502.

5. Braga, A. P., Carvalho, A. P. L. F., & Ludermir, T. B. (2014). Redes neurais artificiais: Teoria e aplicações (2nd ed.). LTC.

6. Carneiro Junior, J. B. A., Souza, C. C., Castelão, R. A., Souza, C. F. C. de, & Crespo, G. M. J. (2024). Uso de redes neurais artificiais para prever os valores das ações das empresas de carnes e derivados listadas na B3 após a pandemia de COVID-19. Revista de Gestão e Secretariado – GeSec, 15(11), 01–21.

7. DePinho, R. A. (2000). The age of cancer. Nature, 408(6809), 248–254. https://doi.org/10.1038/35041694

8. Duffy, S. W., Chen, T. H., Smith, R. A., & Yen, M. F. (2013). Real and artificial controversies in breast cancer screening. Breast Cancer Management, 2(6), 519–528.

9. FIOCRUZ – Ciência e Saúde pela Vida. (2024, January 5). Pesquisa mostra expansão de aplicações de inteligência artificial contra o câncer. https://portal.fiocruz.br/noticia/pesquisa-mostra-expansao-de-aplicacoes-de-inteligencia-artificial-contra-o-cancer

10. Haykin, S. (2009). Neural networks and learning machines (3rd ed.). Pearson.

11. IBGE – Instituto Brasileiro de Geografia e Estatística. (2025, January 5). Panorama Censo 2022. https://www.ibge.gov.br/

12. IBM-SPSS. (2024, March 10). IBM Forecasting 28: Previsão. Incidence Forecasting, Machine Learning. https://www.ibm.com/docs/en/SSLVMB_28.0.0/pdf/IBM_SPSS_Forecasting.pdf

Revista
ARACÊ
ISSN: 2358-2472

13. INCA – Instituto Brasileiro de Câncer. (2025, January 5). Estatísticas de câncer. https://www.gov.br/inca/pt-br/assuntos/cancer/numeros

14. Jemal, A., Bray, F., Center, M. M., Ferley, J., Elizabeth, E., & Forman, W. D. (2011). Global cancer statistics. CA: A Cancer Journal for Clinicians, 61(2), 69–90.

15. Kovács, Z. L. (2006). Redes neurais artificiais: Fundamentos e aplicações (4th ed.). Livraria da Física.

16. Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2008). Introduction to time series analysis and forecasting. John Wiley & Sons.

17. Riahi, N., Hosseini-Motlagh, S. E., & Teimourpour, B. A. (2013). A three-phase hybrid times series modeling framework for improved hospital inventory demand forecast. International Journal of Hospital Research, 2(3), 133–142.

18. Safi, Y., & Bouroumi, A. A. (2011). Neural network approach for predicting forest fires. In Conference IEEE Sensors (pp. 1–5). IEEE.

19. Saúde Pública. (2021). Saúde pública no século XXI: Uma abordagem multidisciplinar (Vol. 2, J. N. Andrade, Org.) [E-book]. Omnis Scientia.

20. Smith, J. (2018). Predicting stock prices using multilayer perceptron neural networks. Journal of Finance and Investment Analysis, 7(4), 12–25.

21. Sobral, G. S., Araújo, Y. B., Kameo, S. Y., Silva, G. M., Santos, D. K. da C., & Carvalho, L. L. M. (2022). Análise do tempo para início do tratamento oncológico no Brasil: Fatores demográficos e relacionados à neoplasia. Revista Brasileira de Cancerologia, 68(3), e-122354. https://rbc.inca.gov.br/index.php/revista/article/view/2354

22. Yuvaraj, D., Deepa, P., & Govindarajan, A. (2018). Artificial neural networks for stock market prediction: A systematic literature review. International Journal of Intelligent Systems and Applications, 10(9), 1–13.