

PREVISÃO DA INCIDÊNCIA DE CÂNCER POR GÊNERO EM BELO HORIZONTE: INOVAÇÃO NO USO DE REDES NEURAIS ARTIFICIAIS PARA PLANEJAMENTO DE SAÚDE PÚBLICA

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

Data de submissão: 28/02/2025

Data de publicação: 28/03/2025

Celso Fabricio Correia de Souza

Mestre em Meio Ambiente e Desenvolvimento Regional (UNIDERP).
Docente na Pontifícia Universidade Católica de Campinas. Campinas – SP, Brasil.
E-mail: celsofabricio36@gmail.com
ORCID: <https://orcid.org/0000-0002-4593-2798>
LATTES: <http://lattes.cnpq.br/2942671200096844>

Yuzzo Iano

Doutorado em Engenharia Elétrica (UNICAMP).
Docente na Universidade Estadual de Campinas. Campinas – SP, Brasil.
E-mail: yuzo@decom.fee.unicamp.br
ORCID: <https://orcid.org/0000-0002-9843-9761>
LATTES: <http://lattes.cnpq.br/1819902045004845>

Joao Bosco Arbués Carneiro Junior

Doutorado em Meio Ambiente e Desenvolvimento Regional
Universidade Federal de Rondonópolis
E-mail - jbacj@hotmail.com
ORCID: <http://orcid.org/0000-0001-7094-2245>
LATTES: <http://lattes.cnpq.br/5073023761658236>

Celso Correia de Souza

Doutorado em Engenharia Elétrica
Universidade Anhanguera Uniderp
E-mail: csouza939@gmail.com
ORCID: <https://orcid.org/0000-0002-2689-8264>
LATTES: <http://lattes.cnpq.br/2881392515816773>

Raul Asseff Castelão

Doutorado em Meio Ambiente e Desenvolvimento Regional
Universidade Anhanguera Uniderp
E-mail: raulassefcastelao@gmail.com
ORCID: <http://orcid.org/0000-0002-4677-6088>
LATTES: <http://lattes.cnpq.br/1364663546445753>

Victor Rocha Pires de Oliveira

Médico Oncologista
Associação Beneficente Casa Rosa – Santa Casa de Campo Grande
E-mail: vrpiresoliveira@gmail.com

Gisela Maria Jorgino Crespo
Mestrando em Agronegócio Sustentável
Universidade Anhanguera Uniderp
E-mail: giselacrespo@icloud.com
ORCID: <https://orcid.org/0009-0003-5947-4251>
LATTES: <http://lattes.cnpq.br/7460435585829383>

Juliana Maria Correia de Souza Vieira
Mestre em Odontologia
Universidade Anhanguera Uniderp
E-mail: julianamaria2002@hotmail.com
ORCID: <https://orcid.org/0000-0003-4478-9405>
LATTES: <http://lattes.cnpq.br/3801758547630854>

RESUMO

O estudo analisou a incidência de câncer em Belo Horizonte de 2000 a 2020 com dados fornecidos pelo INCA, utilizando a técnica de Rede Neural Multilayer Perceptron (RNA_MLP) para prever casos de câncer para os anos de 2021 a 2023. A cidade possui uma população majoritariamente feminina (53,5%), o que reflete diretamente na maior incidência de câncer entre as mulheres, principalmente o câncer de mama. A análise de correlação entre a população e a incidência de câncer revelou fortes associações, com uma correlação de 0,93 para os casos masculinos e 0,95 para os femininos, além de 0,98 entre os casos nos dois gêneros, evidenciando um padrão de crescimento quase sincronizado nos casos de câncer para ambos os sexos. O desempenho da RNA_MLP foi avaliado com base nos erros quadráticos, apresentando uma soma de erros quadráticos de 1,901 no treinamento e 0,299 nos testes, indicando um bom ajuste do modelo. Os erros relativos também foram menores nos testes, com 7,8% para o modelo geral e 7,8% para ambos os sexos. As previsões indicaram um aumento contínuo na incidência de câncer entre 2021 e 2023, com uma tendência de crescimento tanto para o sexo masculino quanto feminino, refletindo não apenas o aumento populacional, mas também uma possível melhoria na detecção da doença. Apesar de ser uma previsão baseada em estimativas, o estudo destacou a importância do modelo para auxiliar no planejamento de políticas públicas de saúde e estratégias de prevenção, considerando o impacto do câncer na saúde pública da cidade.

Palavras-chave: Aprendizado de máquina. Métodos preditivos. Previsão de incidência. Saúde pública.

1 INTRODUÇÃO

A incidência de câncer tem se tornado uma crescente preocupação em diversas partes do mundo, incluindo o Brasil, impulsionada por fatores como o aumento da população e a longevidade (JEMAL *et al.*, 2011). Embora as melhorias nos diagnósticos permitam detectar casos que anteriormente passariam despercebidos, a principal preocupação está na necessidade urgente de investimentos em saúde pública para acompanhar esse aumento (DUFFY *et al.*, 2013). O envelhecimento da população tem contribuído para o crescimento da incidência, especialmente entre os mais velhos, o que torna o planejamento de políticas públicas de saúde ainda mais essencial (SOBRAL *et al.*, 2022; DePINHO, 2000).

Atualmente, as pesquisas sobre o câncer têm se expandido no mundo e no Brasil, com uma ênfase crescente no uso de tecnologias avançadas para previsão e tratamento (FIOCRUZ, 2024; BARIONI *et al.*, 2024). Modelos computacionais, como as Redes Neurais Artificiais (RNAs), têm sido uma ferramenta promissora, pois permitem simular padrões e realizar previsões baseadas em grandes volumes de dados, como os históricos de incidência de câncer. No Brasil, o uso dessas tecnologias ainda é um campo em desenvolvimento, mas com grande potencial para transformar a forma como os gestores de saúde pública lidam com as questões relacionadas ao câncer (BARRETO *et al.*, 2018).

Em Belo Horizonte, por exemplo, observam-se variações no crescimento dos casos de câncer entre homens e mulheres, com destaque para tipos como câncer de mama, próstata, pulmão e colôn, que apresentam altas taxas de incidência (INCA, 2024). A situação exige ações mais eficazes em termos de prevenção, tratamento e recursos para atender a essa demanda crescente, além de um planejamento adequado para evitar que o aumento da incidência sobrecarregue ainda mais o sistema de saúde pública.

A questão central deste estudo é investigar se, em Belo Horizonte, existem diferenças significativas nas incidências de câncer entre os gêneros, considerando as tendências históricas de aumento populacional, longevidade e diagnósticos. O estudo tem como objetivo analisar se as incidências de câncer nessa cidade apresentam diferenças significativas de acordo com o gênero, no período de 2000 a 2020. Além disso, será realizada uma previsão sobre a incidência de câncer para o período de 2021 a 2023, utilizando uma Rede Neural Artificial Multilayer Perceptron (RNA_MLP). O foco será avaliar as tendências de incidência para os gêneros masculino e feminino, a fim de identificar padrões e fornecer estimativas precisas para o futuro.

Desse modo, o objetivo é entender de que maneira as variações na incidência por gênero, bem como as tendências observadas ao longo do tempo, podem contribuir para a formulação de políticas

públicas de saúde mais direcionadas e eficazes, além de possibilitar um melhor planejamento dos recursos de saúde no futuro. Este estudo busca também preencher lacunas no entendimento das diferenças entre os gêneros na prevalência de certos tipos de câncer, fornecendo informações para o aprimoramento das políticas de saúde pública na cidade e, eventualmente, em outros contextos.

Espera-se que os resultados forneçam um modelo preditivo preciso, com baixos erros e que possa ser utilizado pelos gestores públicos para melhor alocar recursos e estratégias de prevenção e tratamento. Além disso, o estudo abre portas para o uso de outras tecnologias de aprendizado de máquina na saúde pública, impulsionando pesquisas futuras nesse campo (SAÚDE PÚBLICA, 2021).

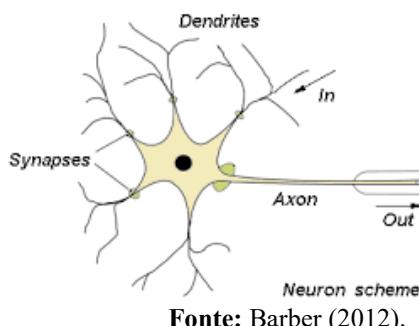
Em suma, o presente estudo busca contribuir para o planejamento estratégico de saúde pública de Belo Horizonte, propondo uma ferramenta inovadora para prever a incidência de câncer e, com isso, auxiliar na tomada de decisões para o enfrentamento dessa doença. A utilização de redes neurais no estudo da incidência de câncer pode contribuir e orientar como os dados podem ser utilizados para antecipar o quanto antes o tratamento do câncer, além de fornecer um modelo que pode ser aplicado em outras cidades e contextos nacionais e internacionais. A continuidade dessa pesquisa pode abrir caminhos para a implementação de modelos mais avançados e acurados, impulsionando o uso da inteligência artificial contribuindo com o bem-estar da sociedade.

2 REFERENCIAL TEÓRICO

As Redes Neurais Artificiais (RNAs) constituem uma área relevante da Inteligência Artificial (IA), inspirada no funcionamento do cérebro humano, com a capacidade de aprender, adaptar-se e tomar decisões. Seu uso é amplo, abrangendo tarefas complexas, como reconhecimento de padrões, previsão de séries temporais e tomada de decisões em tempo real (HAYKIN, 2009).

Segundo Kovácz (2006), as RNAs são formadas por unidades de processamento chamadas neurônios artificiais, interligados por pesos sinápticos que determinam a força e direção das influências das entradas. O funcionamento de um neurônio artificial é inspirado no modelo biológico, com a diferença de que ele possui uma forma simplificada e matematicamente mais eficiente. Os neurônios biológicos, figura 1, são compostos por dendritos, corpo celular e axônio, com a transmissão de impulsos elétricos ocorrendo através das sinapses (BARBER, 2012).

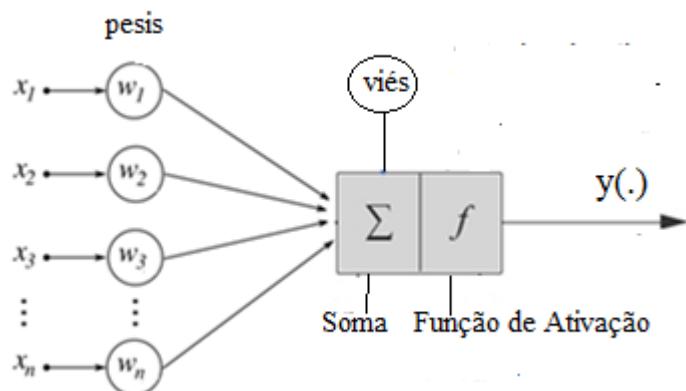
Figura 1. Representação de um neurônio biológico e seus principais componentes.



Fonte: Barber (2012).

McCulloch e Pitts propuseram, na década de 1940, o modelo matemático para o neurônio artificial, que foi fundamental para o desenvolvimento das RNAs. Esse modelo consiste em sinais de entrada, pesos sinápticos, um combinador linear e uma função de ativação. Os sinais de entrada são ponderados pelos pesos, somados e, em seguida, passam pela função de ativação, que determina se o neurônio "dispara" ou não. Esse processo introduz não-linearidade, permitindo que as RNAs aprendam e realizem tarefas complexas. Na figura 2 está representado o modelo que foi a base para as redes neurais modernas e continua sendo uma referência importante na inteligência artificial (BARBER, 2012).

Figura 2. Esquema de um neurônio artificial com os pontos de entrada e saída e suas funções de ativação.



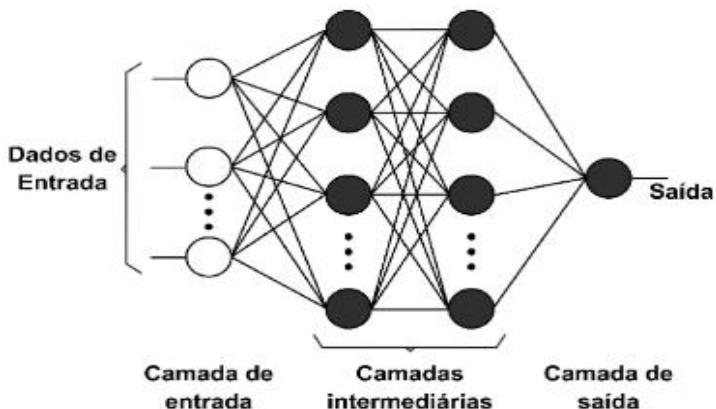
Fonte: Adaptado de Barber (2012).

O neurônio artificial é a unidade fundamental nas redes neurais, recebendo entradas x_1, x_2, \dots, x_n que são multiplicadas pelos respectivos pesos sinápticos w_1, w_2, \dots, w_n e somadas com um viés. A soma resultante \sum , passa por uma função de ativação f que transforma a soma obtida em uma saída y , que vai decidir se o neurônio deve ser ativado ou não. As funções de ativação, como a tangente hiperbólica, são essenciais para permitir que a rede capture relações não lineares nos dados. A saída do neurônio é então transmitida para a próxima camada ou, no caso da última camada, para a

previsão final. O aprendizado ocorre ao ajuste ótimo dos pesos e vieses durante o treinamento. Esse processo possibilita que as redes neurais resolvam problemas complexos, como reconhecimento de padrões e previsão de tendências. Assim, os neurônios trabalham em conjunto para realizar tarefas avançadas de análise de dados.

As RNAs podem ser estruturadas de diferentes maneiras, incluindo redes retroalimentadas e recorrentes, com variações em suas topologias, como as redes de camadas simples e multicamadas. Entre essas, o Multilayer Perceptron (MLP) é uma das mais comuns e amplamente utilizada para resolver problemas complexos, como a análise de grandes volumes de dados, padrões não lineares e previsões de tendências, como ilustrado na figura 3.

Figura 3. Configuração de uma rede neural artificial multicamadas (MLP), mostrando a camada de entrada, duas camadas ocultas e a camada de saída.



Fonte: Adaptada de Barber (2012).

O equacionamento de uma RNA é fundamental para entender como essas estruturas computacionais aprendem e se adaptam. A sua modelagem matemática envolve a descrição dos algoritmos que regem a transmissão de informações entre neurônios artificiais. Através de funções de ativação e algoritmos de treinamento, como o *backpropagation*, busca-se ajustar os pesos das conexões, minimizando erros e aprimorando a precisão. Esses processos de treinamento, que incluem otimização de parâmetros, são essenciais para garantir que as RNAs possam generalizar e realizar previsões eficazes.

Segundo Haykin (2009), o estudo matemático das RNAs proporciona uma visão sobre o comportamento e a eficiência desses sistemas, pois, cada neurônio artificial realiza uma combinação linear das entradas, seguida de uma função de ativação para determinar a sua saída. A equação geral para um neurônio com n entradas é dada pela equação (1).

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

Onde: x_i são as entradas do neurônio; w_i são os pesos associados a cada entrada; b é o viés (bias), que ajusta a soma ponderada e; f é a função de ativação, que transforma a soma ponderada em uma saída.

O propósito da função de ativação é o de introduzir não-linearidades, permitindo que a rede aprenda padrões complexos. As funções de ativação determinam como a saída do neurônio é calculada. São exemplos de funções de ativação as equações (2) e (3).

Sigmoide:

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

Essa função produz uma saída entre 0 e 1, muito usada em tarefas de classificação binária.

Tangente hiperbólica (tanh):

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

A saída dessa função varia entre -1 e 1, ajudando a centralizar os dados em torno de zero.

2.1 PROPAGAÇÃO DIRETA

A propagação direta em uma rede neural refere-se ao processo de calcular as saídas de cada camada da rede a partir das entradas, seguindo uma sequência de transformações matemáticas. Esse processo, equação (4), envolve a aplicação de funções de ativação e a utilização dos pesos das conexões, permitindo que os dados passem de uma camada para a seguinte até que a saída final seja gerada.

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

Onde: $z^{(l)}$ é o vetor de soma ponderada da camada l ; $W^{(l)}$ é a matriz de pesos da camada l ; $a^{(l-1)}$ é a ativação da camada anterior e; $b^{(l)}$ é o vetor de viés da camada l

A ativação é calculada pela equação (5).

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

Onde f é a função de ativação aplicada a cada elemento de $z^{(l)}$.

2.2 FUNÇÃO ERRO QUADRÁTICO MÉDIO (MSE)

A função de erro quadrático médio (MSE) para uma rede neural, utilizada em problemas de regressão, é calculada considerando m exemplos e a diferença entre as saídas reais y_i e as saídas previstas \hat{y}_i . O MSE equação (6), mede a média dos quadrados das diferenças entre os valores observados y_i e os estimados \hat{y}_i , fornecendo uma métrica que quantifica a precisão do modelo em prever os resultados esperados. Quanto menor o valor do MSE, melhor a rede neural ajusta-se aos dados de treinamento.

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

Onde: y_i são os valores reais; \hat{y}_i são as previsões da rede e; m é o número de exemplos no conjunto de dados. O fator 2 na fórmula é introduzido para simplificar o cálculo do gradiente. Quando se calcula o gradiente da função de erro com relação aos pesos da rede durante o treinamento, a derivada do quadrado de um número traz um fator de 2. Ou seja, ao dividir o MSE por 2, evita-se ter esse 2 extra ao calcular a derivada, o que torna a expressão do gradiente mais simples e evita a multiplicação por 2 durante o processo de atualização dos pesos.

2.3 RETROPROPAGAÇÃO

Para otimizar a matriz dos pesos W e viéses b , calcula-se os gradientes do erro E em relação a cada parâmetro usando o algoritmo de retropropagação. A atualização dos parâmetros para os pesos e o viés é feita utilizando o gradiente descendente, equações (7) e (8).

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

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

Onde: η é a taxa de aprendizado; $\partial E / \partial W^{(l)}$ é o gradiente do erro em relação aos pesos e; $\partial E / \partial b^{(l)}$ é o gradiente do erro em relação aos vieses.

Os gradientes são calculados usando a cadeia de derivadas (regra da cadeia), começando pela camada de saída e retrocedendo até as camadas anteriores.

2.3.1 Gradiente Descendente

A atualização dos pesos e vieses na rede é realizada utilizando o gradiente descendente, onde a ideia é minimizar a função de erro E , equação (9).

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

Onde: θ pode ser um peso ou viés; $\nabla_{\theta} E$ é o gradiente da função de erro em relação a θ e; η é a taxa de aprendizado.

2.4 RESUMO DA PROPAGAÇÃO COMPLETA PARA n NEURÔNIOS

A função geral para uma RNA com L camadas, onde a camada 1 tem n_l neurônios é dada pelo seguinte algoritmo do processo de propagação:

1. **Entrada:** A entrada a^l (a primeira camada) é passada pela rede.
2. **Para cada camada 1 (de 1 a L):**
 - o Calcular o pre-ativação $z^{(l)} = W^{(l)}a^{(l-1)} + b^{(l)}$
 - o Calcular a ativação $a^{(l)} = f(z^{(l)})$
3. **Saída:** A última camada fornece a previsão final $\hat{y} = a^{(L)}$.

Esse processo é realizado tanto durante a propagação direta quanto no cálculo do gradiente durante o processo de treinamento (retropropagação), com o objetivo de minimizar a função erro MSE.

O algoritmo retropropagação, uma técnica de aprendizado supervisionado, realiza a atualização dos pesos sinápticos por meio de duas fases: forward e backward. A fase forward calcula a saída com base nas entradas e atualiza os pesos, enquanto a fase backward ajusta esses pesos com base no erro calculado entre a saída prevista e a saída desejada (SAFI e BOUROUMI, 2011). A RNA_MLP pode ser treinada de acordo com esses princípios, aplicando o erro quadrático para ajustar os pesos até que o modelo tenha um desempenho satisfatório (SMITH, 2018).

A aplicação de RNAs se estende também à previsão financeira, especialmente na previsão de preços de ações. Conforme observado por Arun e Venkatalakshmi (2017), as RNAs são eficientes para lidar com a dinâmica e não linearidade dos mercados financeiros. Elas são capazes de processar grandes volumes de dados e identificar padrões complexos, o que as torna adequadas para prever o comportamento do mercado. Entretanto, as RNAs também apresentam limitações, como a dificuldade de interpretar os resultados devido à sua natureza de "caixa preta" e a sensibilidade à qualidade dos dados (YUVARAJ *et al.*, 2019).

3 METODOLOGIA

Esta seção descreve a metodologia utilizada na pesquisa, que é de caráter aplicada e experimental, com foco em simulações computacionais por meio de RNAs. As principais variáveis envolvem a população de Belo Horizonte e incidência de câncer segmentada por gênero, masculina e

feminina e o ano da série. A configuração da rede, os parâmetros de treinamento, como a taxa de aprendizado e o número de iterações, bem como a função de erro, também foram considerados. A coleta de dados foi realizada a partir de fontes oficiais do Instituto Nacional de Câncer (INCA), com ênfase nos tipos de câncer mais comuns entre homens e mulheres na cidade de Belo Horizonte (INCA, 2024). Estes dados foram escolhidos por serem, do ponto de vista dos autores, os mais completos e detalhados sobre a doença, disponíveis no Brasil.

As análises realizadas englobaram a avaliação das prevalências de câncer por gênero, bem como a análise das correlações de Pearson entre as incidências de câncer nas populações masculina e feminina. Para a análise dos resultados, foram utilizadas métricas como erro quadrático médio, precisão e tempo de treinamento. O modelo foi treinado utilizando retropropagação e gradiente descendente, e a avaliação de desempenho foi realizada com dados de treinamento e teste. Os resultados foram apresentados por meio de gráficos e tabelas, com o objetivo de compreender a relação entre os parâmetros das RNAs e o desempenho no aprendizado supervisionado (BRAGA *et al.*, 2014; HAYKIN, 2009).

A metodologia de previsão com RNAs, como a utilizada no modelo MLP, é baseada em uma abordagem quantitativa, com coleta, pré-processamento e divisão dos dados em conjuntos de treinamento, validação e teste. A RNA_MLP é então treinada utilizando o método backpropagation, ajustando seus pesos iterativamente. Após o treinamento, o desempenho da rede é avaliado usando métricas como o Erro Percentual Médio Absoluto (MAPE), que avalia a precisão da previsão comparando os valores reais e previstos (MONTGOMERY *et al.*, 2008; RIAHI *et al.*, 2013).

A metodologia de RNA_MLP foi aplicada para prever o número de mortes por câncer na cidade de Belo Horizonte, Brasil, no período de 2021 a 2023, com a análise dos resultados. Para a implementação de RNA MLP para a previsão de mortalidade por câncer utilizou o software Statistical Package for the Social Sciences (SPSS), que foi essencial para ajustar os parâmetros da rede e obter as previsões dos índices de mortalidade. O processo de treinamento, ajustes dos parâmetros e validação dos resultados são etapas importantes para garantir a eficácia da rede neural (IBM-SPSS, 2024). Embora as técnicas de redes neurais tenham se mostrado promissoras na área da saúde, a precisão do modelo de previsão depende da qualidade dos dados e da capacidade da rede em aprender com os padrões históricos observados.

A aplicação do método de RNA_MLP com o uso do IBM-SPSS fez uso do seguinte algoritmo:

- i) Coleta de dados: Dados anuais sobre as mortes por câncer foram obtidos de fontes públicas;

- ii) Pré-processamento dos dados: Os dados foram ajustados para garantir um formato adequado para a entrada na rede neural, incluindo a remoção de dados ausentes e normalização;
- iii) Divisão dos dados: Os dados foram divididos em conjuntos de treinamento, validação e teste;
- iv) Construção do modelo de RNA_MLP: Foi desenvolvida uma RNA_MLP com camadas ocultas, e os parâmetros foram ajustados durante o treinamento;
- v) Treinamento da rede: O modelo foi treinado utilizando o conjunto de treinamento, ajustando os pesos para minimizar o erro;
- vi) Avaliação e ajuste: O desempenho foi avaliado com o conjunto de validação, utilizando métricas como o erro médio quadrático;
- vii) Teste e previsão: O modelo ajustado foi testado e utilizado para realizar previsões do número de mortes, e;
- viii) Análise dos resultados: Foram analisadas as previsões e identificados padrões e limitações do modelo.
- ix) Aperfeiçoamento do modelo: Com base nas análises, ajustes foram feitos no modelo sem alterar sua arquitetura.

A qualidade do ajuste foi avaliada pelo Erro Percentual Médio Absoluto (MAPE), que mede a precisão do modelo, conforme equação (10).

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

Onde N é o número de períodos utilizados, R_t é o valor real observado da variável e P_t é o valor previsto da variável pela RNA_MLP.

Segundo Carneiro Junior *et al.* (2024), menores valores de MAPE indicam melhor desempenho do modelo de previsão, conforme apresentado no quadro 1 de níveis de aceitação do erro, onde $MAPE < 5\%$ é considerado excelente.

Quadro 1. Faixas de aceitação do erro para as diferenças entre os valores observados e valores previstos do erro MAPE.

Faixa de MAPE	Interpretação
$MAPE < 5\%$	Excelente – Precisão alta
$5\% \leq MAPE < 10\%$	Bom – Precisão moderada
$10\% \leq MAPE < 20\%$	Regular – Precisão razoável
$MAPE \geq 20\%$	Ruim – Precisão baixa

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

4 RESULTADOS E DISCUSSÃO

No quadro 2 apresenta os dados gerais sobre a incidência de câncer, por gênero, em Belo Horizonte, com base nos dados fornecidas pelo Instituto Nacional de Câncer (INCA), no período de 2000 a 2020.

Quadro 2. População de Belo Horizonte (População), Incidência de câncer, por gênero (Masculino e Feminino), na cidade de Belo Horizonte, no período de 2000 a 2020.

Ano	População	Masculino	Feminino
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

Fonte: INCA (2024).

Segundo o IBGE (2025), a cidade de Belo Horizonte é composta por 53,5% de mulheres e 46,5% de homens, refletindo uma ligeira predominância feminina na população. Por outro lado, esse fator demográfico pode influenciar a análise da incidência de doenças como o câncer, já que o número absoluto de mulheres na cidade é maior do que o número absoluto de homens.

O quadro 2 apresenta dados da população (População) e a incidência de câncer, por gênero (Masculino e Feminino), na cidade de Belo Horizonte, no período de 2000 a 2020. Os números do quadro 2 mostram uma tendência crescente população e na incidência de câncer entre 2000 e 2020, com um número de casos entre as mulheres consistentemente superior. Esse aumento pode ser atribuído, em parte, à maior proporção de mulheres na cidade, sugerindo que, enquanto os dados do

IBGE ajudam a contextualizar a distribuição demográfica da população, os dados do INCA (2024) evidenciam a maior incidência de câncer entre as mulheres, refletindo uma combinação de fatores demográficos e de saúde pública.

Inicialmente, foi calculada a matriz de correlação entre as variáveis "População", "Masculino" e "Feminino" do quadro 2 com o objetivo de verificar a associação entre essas variáveis, obtendo-se a tabela 1. A análise de correlação busca identificar a intensidade e a direção das relações entre essas três variáveis ao longo dos anos de 2000 a 2020.

Tabela 1. Matriz de correlação entre as variáveis População e as incidências de câncer por gênero (Masculino e Feminino) na cidade de Belo Horizonte, no período de 2000 a 2020.

Variáveis	População	Masculino	Feminino
População	1		
Masculino	0,93	1	
Feminino	0,95	0,98	1

Fonte: Autores (2024).

A correlação entre as variáveis População e Masculino é de 0,93, enquanto a correlação entre a População e Feminino é de 0,95, indicando que variações em uma dessas variáveis tendem a acompanhar as variações na outra. A correlação entre as variáveis Masculino e Feminino é de 0,98, sugerindo que as variações nos casos de câncer em ambos os gêneros são quase perfeitamente sincronizadas. Isso demonstra que, ao longo desse período, a estrutura populacional da cidade manteve um equilíbrio nas incidências de câncer entre os gêneros.

Como o objetivo deste estudo foi o de realizar a previsão de número de casos futuros de incidências de cânceres por gênero, para o período de 2021 a 2023, utilizou-se uma RNA_MLP treinada com os dados históricos do quadro 2. Como ferramenta estatística utilizou-se o software SPSS para realizar essa previsão. Na aplicação da RNA_MLP foram considerados 24 casos, dos quais 21 foram válidos, sendo 16 (76,2%) para treinamento e 5 (23,8%) para testes. Os três casos restantes foram utilizados para previsões. A distribuição dos dados visa garantir que a maior parte da amostra seja utilizada para o treinamento da rede, enquanto uma parte menor é reservada para a avaliação (testes) do modelo.

A configuração da RNA_MLP utilizada na análise da incidência de câncer em Belo Horizonte teve como foco a variável "População" como covariável, enquanto as variáveis dependentes foram os gêneros "Masculino" e "Feminino". A camada de entrada da rede possui uma unidade, que representa a "População", com reescalonamento padronizado. Na camada oculta, há uma única camada com quatro unidades e a função de ativação é a tangente hiperbólica, que é frequentemente utilizada para

modelar relações não lineares entre as entradas e as saídas. Na camada de saída, as duas unidades representam as variáveis dependentes, "Masculino" e "Feminino", com reescalonamento padronizado. A função de ativação para essa camada foi a identidade, que é apropriada para problemas de regressão. O método de erro adotado foi a soma dos quadrados MSE, utilizado para minimizar a diferença entre os valores previstos e os valores reais durante o treinamento da rede. A estrutura da rede, com suas camadas e funções, visa capturar as complexidades e padrões nos dados relacionados ao câncer, segmentados por gênero.

A tabela 2 apresenta a sumarização dos erros de saídas do desempenho do modelo de treinamento e testes.

Tabela 2. Sumarização dos erros de saída da do desempenho da RNA MLP.

Treinamento	Soma dos erros quadráticos (MSE)	1,901
	Erro Relativo Geral Médio	0,127
	Erro relativo para dependentes de escala	Masculino
		Feminino
	Regra de parada usada	
Testes	Tempo de treinamento	
	Soma dos erros quadráticos (MSE)	0,299
	Erro Relativo Geral Médio	0,078
	Erro relativo para dependentes de scala	Masculino
		Feminino

Fonte: Autores (2025).

A tabela 2 apresenta dados sobre o desempenho da RNA MLP, de treinamento e testes. Durante o treinamento, a soma dos erros quadráticos foi de 1,901, o que indica uma margem de erro considerável, considerando que é um valor absoluto, mas sem a normalização, esse valor por si só não fornece uma noção clara da magnitude do erro em relação aos valores totais. O erro relativo geral médio no treinamento foi 0,127 (12,7%), sugerindo um desempenho razoável.

Para os dados segmentados por sexo, o erro relativo masculino foi 0,145 (14,5%), enquanto o feminino foi 0,108 (10,8%), mostrando que o modelo teve um desempenho ligeiramente melhor para as mulheres. O modelo utilizou sua regra de parada baseada em "1 passo consecutivo sem diminuição de erro", o que significa que o treinamento foi interrompido assim que não houve mais melhorias. O tempo total de treinamento foi extremamente curto, com apenas 0,02 segundos.

No conjunto de testes, a soma dos erros quadráticos foi bem menor, 0,299, sugerindo que o modelo generalizou de forma eficaz. O erro relativo geral médio nos testes foi 0,078 (7,8%), indicando um bom desempenho. Além disso, os erros relativos para os grupos masculino e feminino foram

idênticos, ambos 0,078 (7,8%), mostrando que o modelo teve um desempenho equilibrado entre os sexos.

Em geral, o modelo teve um bom desempenho, com erros menores durante os testes em comparação com o treinamento e um equilíbrio entre os grupos analisados. Na tabela 3 estão registrados os valores previstos (calculados com o uso do software SPSS) para os três anos à frente de previsão (2021, 2022 e 2023).

Tabela 3. Valores observados da População e valores previstos pela RNA_MLP sobre a incidência de câncer na população de Belo Horizonte, por gênero (Masculino e Feminino), no período de 2021 a 2023.

ANO	VALORES OBSERVADOS*	VALORES PREVISTOS**	
		Masculino	Feminino
2021	2.490.783	6.135	7.211
2022	2.503.527	6.260	7.362
2023	2.516.461	6.374	7.501

Fonte: *IBGE (2024); **Autores (2024).

Os valores apresentados na tabela 3, mostram um aumento da população de Belo Horizonte e uma uma tendência de aumento na incidência de câncer em Belo Horizonte entre 2021 e 2023, tanto para o sexo masculino quanto feminino, com um crescimento gradual no número de casos previstos a cada ano. Esse padrão pode refletir uma possível elevação na taxa de incidência ou, uma maior conscientização e diagnóstico da doença na população ao longo do período. Embora o modelo preveja esses aumentos, é importante notar que esses dados são apenas estimativas baseadas na RNA_MLP, e não há dados reais disponíveis para a análise da precisão das previsões. Além disso, a falta de dados sobre a incidência de câncer para os anos de 2021 a 2023 limita a análise comparativa e a validação do modelo.

5 CONCLUSÃO

A análise da incidência de câncer em Belo Horizonte, entre 2000 e 2020, revelou um aumento significativo nos casos diagnosticados, refletindo o crescimento populacional e a melhoria na detecção precoce da doença. A forte correlação entre as incidências masculinas e femininas, com índice de 0,98, sugere que ambos os grupos seguem padrões de crescimento semelhantes, possivelmente influenciados por fatores demográficos e ambientais comuns.

Os valores apresentados na Tabela 3, que mostram tanto os valores observados quanto os previstos pela RNA_MLP para os anos de 2021 a 2023, indicam uma tendência crescente na população de Belo Horizonte e na incidência de câncer, tanto para o sexo masculino quanto feminino. O aumento

no número de casos previstos a cada ano, com valores para 2021, 2022 e 2023 alcançando 6.135, 6.260 e 6.374 para os homens, e 7.211, 7.362 e 7.501 para as mulheres, pode refletir uma possível elevação na taxa de incidência ou maior conscientização e diagnóstico da doença.

Embora os dados de previsão forneçam informações sobre tendências futuras, a falta de dados reais para esses anos impede uma validação precisa das previsões. No entanto, o bom desempenho do modelo, demonstrado pelos baixos erros relativos, sugere que a RNA_MLP é uma ferramenta eficaz para prever a incidência de câncer. Esses resultados podem ser utilizados para o planejamento de políticas públicas e estratégias de prevenção, ajudando a otimizar a gestão da saúde pública, especialmente no combate ao câncer. A pesquisa destaca, assim, a importância da modelagem preditiva para antecipar tendências e orientar decisões políticas e de saúde.

REFERÊNCIAS

- ARUN, T.; VENKATALAKSHMI, K. Stock Market Prediction using Artificial Neural Networks with Multilayer Perceptron. International Journal of Computer Science and Mobile Computing, 6(1), 194-199. 2017.
- BARBER, D.; SOLLICH, P.; SAAD, D. Finite Size Effects in on-line learning of multi-layer Neural Network. In: Ellacott, S. W.; Mason, J. C.; Anderson, I. J. Mathematics of Neural Networks: Models, Algorithms and Applications. Manchester: Springer Science, p. 84-88, 2012.
- BARRETO, R. G.; MARINHO, G. M. G. A.; BARRETO, G. F. M.; BARRETO, R. G.; AVERSARI, L. O. C.; DANTAS, B. L. Utilizando redes neurais artificiais para o diagnóstico de câncer cervical. Revista Saúde e Ciência online, 7(2), (maio a agosto de 2018). 502 p. 2018.
- BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDEMRIR, T. B. Redes Neurais Artificiais: Teoria e Aplicações. 2ed. Rio de Janeiro: LTC. 228p. 2014.
- CARNEIRO JUNIOR, J. B. A; SOUZA, C. C.; CASTELÃO, R. A.; SOUZA, C. F. C. de; CRESPO, G. \M. J. 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, V. 15, N. 11, P. 01-21, São José dos Pinhais, Paraná, Brasil. 2024.
- DePINHO, R. A. The age of cancer. Nature, 408(6809), 248-254. doi: 10.1038/35041694. 2000.
- DUFFY, S. W.; CHEN, T. H.; SMITH, R. A.; YEN, M. F. Real and artificial controversies in breast cancer screening. Breast Cancer Manag, 2(6):519–28. 2013.
- FIOCRUZ – Ciência e Saúde pela Vida. (2024). Pesquisa mostra expansão de aplicações de inteligência artificial contra o câncer. Disponível em: <https://portal.fiocruz.br/noticia/pesquisa-mostra-expansao-de-aplicacoes-de-inteligencia-artificial-contra-o-cancer>. Acesso em: 05.01.2024.
- HAYKIN, S. Neural networks and learning machines (3rd ed.). Pearson. 2009.
- IBGE – Instituto Brasileiro de Geografia e Estatística. (2025). Panorama Censo 2022. Disponível em: <https://www.ibge.gov.br/>. Acesso em: 05 jan.2025.
- IBM-SPSS. IBM Forecasting 28. (2024). Previsão. Incidence Forecasting, Machine Learning https://www.ibm.com/docs/en/SSLVMB_28.0.0/pdf/IBM_SPSS_Forecasting.pdf. Acesso em 10.mar.2024.
- INCA – Instituto Brasileiro de Câncer. (2024). Estatísticas de Câncer. Disponível em: <https://www.gov.br/inca/pt-br/assuntos/cancer/numeros>. Acesso em 05 jan.2025.
- JEMAL, A.; BRAY, F.; CENTER, M. M.; FERLEY, J.; ELIZABETH, E.; FORMAN, W. D. Global cancer statistics. CA: A Cancer Journal for Clinicians, 61(2), 69-90. 2011.
- KOVÁCS, Z. L. Redes Neurais Artificiais: Fundamentos e Aplicações. 4ed. São Paulo: Livraria da Física. 169p. 2006.

MONTGOMERY, D. C.; JENNINGS, C. L.; KULAHCI, M. Introduction to time series analysis and forecasting. New Jersey: John Wiley & Sons. Inc.2008.

BARIONI, C. T. S.; WANDRESEN, R. P. de B.; PEREIRA, L. F.; COIMBRA, A. F.; KUBO, B. B. de A.; CUNHA, R. C. da. Artificial Intelligence for the Identification of Biomarkers in Cancer Prevention and Diagnosis: Advances and Perspectives. Rev. Bras. Cancerol. 70(2): e-254692. 2024.

RIAHI, N.; HOSSEINI-MOTLAGH, S. E TEIMOURPOUR, B. A. Three-phase Hybrid Times Series Modeling Framework for Improved Hospital Inventory Demand Forecast. International Journal of Hospital Research, 2(3).133-142. 2013.

SAFI, Y.; BOUROUMI, A. A. Neural Network Approach for Predicting Forest Fires. In: IEEE SENSORS, Morocco. Conference IEEE Sensors. Morocco: IEEE, p. 1-5. 2011.

SAÚDE PÚBLICA. Saúde Pública no século XXI: Uma abordagem multidisciplinar. vol. 2. (livro eletrônico). Org. Andrade, J. N. – Triunfo-PE. Omnis Scientia, 225p. 2021.

SOBRAL, G. S., ARAÚJO, Y. B., KAMEO, S. Y., SILVA, G. M., SANTOS, D. K. da C., CARALHO, L. L. M. (2022). Análise do Tempo para Início do Tratamento Oncológico no Brasil: Fatores Demográficos e Relacionados à Neoplasia. ev. Bras. Cancerol. 68(3), e-122354. Disponível em: <https://rbc.inca.gov.br/index.php/revista/article/view/2354>. Acesso em: 05jan.2024.

SMITH, J. Predicting Stock Prices Using Multilayer Perceptron Neural Networks. Journal of Finance and Investment Analysis, 7(4), 12-25. 2018.

YUVARAJ, D.; DEEPA, P.; GOVINDARAJAN, A. Artificial Neural Networks for Stock Market Prediction: A Systematic Literature Review. International Journal of Intelligent Systems and Applications, 10(9), p. 1-13. 2018.