



## UTILIZANDO REDES NEURAS ARTIFICIAIS PARA O DIAGNÓSTICO DE CÂNCER CERVICAL

*Renan Gomes Barreto<sup>1</sup>, Gersica Maria Gomes Almeida Marinho<sup>2</sup>, Gabriela Ferreira Marinho Barreto<sup>3</sup>,  
Renata Gomes Barreto<sup>4</sup>, Lucas Oliveira Costa Aversari<sup>5</sup>, Brenna Lucena Dantas<sup>6</sup>*

1. Mestrando em Informática pela Universidade Federal da Paraíba, João Pessoa, PB, Brasil.  
\*Correspondência: Centro de Informática - UFPB, Endereço: Rua dos Escoteiros, s/n, Mangabeira, João Pessoa, PB. renangbarreto@ppgi.ci.ufpb.br.
2. Graduanda em Medicina pela Faculdade de Ciências Médicas da Paraíba, Cabedelo, PB, Brasil.
3. Graduanda em Nutrição pela Faculdade Maurício de Nassau, João Pessoa, PB, Brasil.
4. Terapeuta Ocupacional graduada pela Universidade Federal da Paraíba, João Pessoa, PB, Brasil.
5. Mestrando em Informática pela Universidade Federal da Paraíba, João Pessoa, PB, Brasil.
6. Graduanda em Medicina pela Faculdade de Ciências Médicas da Paraíba, Cabedelo, PB, Brasil.

### RESUMO

O câncer cervical, também conhecido como câncer de colo de útero, é a terceira maior causa de câncer em mulheres em todo o mundo. Com o aumento do poder computacional e a facilidade de coleta de informações médicas, a Inteligência Artificial tem ajudado a Oncologia no desenvolvimento de algoritmos e técnicas para o diagnóstico e estadiamento de doenças. Redes Neurais Artificiais são modelos computacionais inspirados pela biologia e podem ser utilizadas em uma grande variedade de problemas de aprendizado de máquina. Este trabalho desenvolve uma Rede Neural Artificial para o auxílio no diagnóstico de Câncer Cervical. Para tanto, foi identificada uma base de dados pública para treinamento e teste, além disso, foi apresentada um modelo de arquitetura capaz de prever o resultado de uma biópsia em um paciente. Por fim, com o objetivo de oferecer suporte a decisão clínica, o treinamento da rede foi feito em modo supervisionado utilizando dados de fatores de risco do câncer cervical. Todas as arquiteturas testadas apresentaram acurácia média maiores que 94% sendo que a melhor delas obteve 96,2% de acurácia média. Esta pesquisa contribui para o avanço do uso da Inteligência Artificial em Sistemas de Suporte a Decisão Clínica com foco na Oncologia, mostrando que Redes Neurais Artificiais podem ser utilizadas com sucesso para o diagnóstico de câncer cervical.

**Palavras-chave:** Informática Médica; Inteligência Artificial; Aprendizado de Máquina; Neoplasias do Colo do Útero.

### USING ARTIFICIAL NEURAL NETWORKS FOR CERVICAL CANCER DIAGNOSIS

### ABSTRACT

Cervical cancer is the third leading cause of cancer in women around the world. With the increase of computational power and the ease of medical information gathering, Artificial Intelligence has helped Oncology in the development of algorithms and techniques for the diagnosis and staging of diseases. Artificial Neural Networks are computational models



inspired by biology and can be used in a wide variety of machine learning problems. This work develops an Artificial Neural Network to aid in the diagnosis of Cervical Cancer. To do so, a public database for training and testing was identified. In addition, an architecture model capable of predicting the outcome of a biopsy in a patient was presented. Finally, in order to provide clinical decision support, the training of the network was done in a supervised mode using data from cervical cancer risk factors. All the architectures tested presented an average accuracy greater than 94% and the best of them obtained a 96.2% average accuracy. This research contributes to the advancement of the use of Artificial Intelligence in Clinical Decision Support systems with a focus on Oncology, showing that Artificial Neural Networks can be used successfully for the diagnosis of cervical cancer.

**Keywords:** Medical Informatics; Artificial Intelligence; Machine Learning; Uterine Cervical Neoplasms.

## INTRODUÇÃO

A quantidade de dados coletados na área da saúde aumenta periodicamente, resultando no surgimento de métodos de diagnósticos, princípios químicos, além de progressos na área da biologia molecular e da genética, entre outros avanços medicinais (1).

O câncer cervical, também conhecido como câncer de colo de útero, é a terceira maior causa de câncer em mulheres em todo o mundo. Em 2008, estima-se que ocorreram 530.000 casos, sendo registradas 275.000 mortes pela doença, perdendo apenas para o câncer de mama e o coloretal (2). No Brasil, espera-se que em 2018 serão registrados 16.370 novos casos de câncer de cervical, com um risco estimado de 15,43 casos a cada 100 mil mulheres (3).

Alguns fatores estão fortemente ligados ao acometimento do câncer cervical como: início precoce da atividade sexual, tabagismo, uso de contraceptivos orais, carência de vitaminas, múltiplos parceiros e, principalmente, infecções persistentes pelo vírus HPV (4).

Devido ao aumento do poder computacional e da facilidade de coleta de informações médicas, a Inteligência Artificial, subárea da Ciência da Computação, tem ajudado a Oncologia no desenvolvimento de algoritmos e técnicas para o diagnóstico e estadiamento de doenças (5,6).

Sistemas de Suporte à Decisão Clínica possuem potencial para reduzir a quantidade de erros médicos e melhorar a qualidade e eficiência do tratamento clínico oferecido (7). Nesse contexto, destaca-se o uso de Redes Neurais Artificiais (RNA), que são modelos computacionais inspirados pela biologia e podem ser utilizadas em uma grande variedade de problemas de aprendizado de máquina (8).



As Redes Neurais Artificiais possuem a capacidade de aproximar mapeamentos não lineares complexos, através de um conjunto exemplo de entradas e, com isso, é capaz de encontrar soluções que normalmente são difíceis de modelar utilizando uma abordagem paramétrica clássica (9). Uma Rede Neural Artificial é formada camadas de neurônios e sinapses e é capaz de adquirir conhecimento através de exemplos, ajustando os pesos de ligações entre os neurônios, a partir de um processo de aprendizagem (10).

Pretende-se assim com este trabalho, projetar e treinar uma Rede Neural Artificial, que poderá ser usada na área oncológica, visando permitir o suporte à decisão clínica por auxiliar no diagnóstico e encaminhamento apropriado de pacientes com suspeita de câncer cervical, utilizando os dados disponíveis dos mesmos e de fatores de risco da doença.

Este trabalho, tem como objetivo geral desenvolver uma Rede Neural Artificial para o auxílio no diagnóstico de Câncer de Cervical. Para tanto, definimos como objetivos específicos desta pesquisa:

- Identificar uma base de dados que será utilizada para o treinamento e validação da RNA;
- Propor uma arquitetura de Rede Neural Artificial que será treinada para o suporte ao diagnóstico de Câncer Cervical;
- Avaliar os resultados obtidos durante o treinamento e validação da Rede Neural Artificial.

## MATERIAL E MÉTODOS

Para projetar e treinar a rede neural utilizada para previsão de casos de câncer cervical, foi utilizado o framework Keras 2.1 (11), que é uma biblioteca de código aberto para aprendizado de máquina escrita na linguagem de programação Python 3.6 (12), facilitando a criação e instância de redes neurais artificiais para tarefas de classificação e regressão.

O framework Keras, durante sua execução, requer um *backend* para auxiliar na comunicação e organização das estruturas de dados no computador. Para isso, nessa pesquisa foi utilizado como *backend* o Tensorflow 1.3 (13), criado pela equipe do Google Brain em 2015, com o objetivo de ajudar na criação de sistemas de aprendizagem supervisionada. Além disso, foram utilizadas ferramentas como a Tensorboard, utilizada neste trabalho em sua versão 0.1.8, para a geração e acompanhamento dos gráficos



referentes aos erros de treinamento e validação da rede neural durante seu desenvolvimento (14).

Com a finalidade de automatizar o treinamento de diversas arquiteturas de redes, foi utilizada a biblioteca Scikit-learn (15) em sua versão 0.19, fazendo com que um número maior de arquiteturas de redes fossem treinadas e validadas paralelamente, reduzindo significativamente o tempo de execução do treinamento da rede neural.

Para a execução e testes das redes propostas, foi utilizado um servidor com as seguintes configurações: Processador Intel Core i7-6700, com 64 GB DDR4 de memória RAM, GPU GeForce GTX 1080 e 2 discos rígidos SSD de 500 GB SATA 6 Gb/s. Todas as redes neurais foram instanciadas e treinadas com ajuda da placa gráfica, contribuindo para a diminuição do tempo de execução do algoritmo.

Com o objetivo de oferecer Suporte à Decisão Clínica, o treinamento da rede foi feito em modo supervisionado, utilizando dados de fatores de risco do câncer cervical coletados por (16) e distribuídos pela Universidade da Califórnia. O banco de dados utilizado consiste nas informações demográficas, hábitos e histórico médico de 858 pacientes, juntamente às suas respostas às perguntas de um questionário sobre fatores de risco, podendo algumas das perguntas não terem sido respondidas pelos pacientes por razões de privacidade.

Dentre alguns dos fatores de risco presentes, podemos destacar: número de parceiros sexuais, número de gravidezes, se o paciente é fumante, uso de contraceptivos hormonais, uso de dispositivo intrauterino, histórico de DSTs e diagnósticos prévios de câncer cervical.

Para lidar com os atributos com muitas respostas faltantes, foi utilizada uma técnica tradicional (17) que consiste na remoção de colunas sem quantidades significativas de respostas, partindo do pressuposto que os mesmos possuem baixa influência na classificação, ou, caso contrário, evitando que os mesmos influenciem negativamente na classificação dos pacientes que não responderam a tal pergunta. A tabela 1 lista os atributos presentes no conjunto de dados, valores máximos e mínimos, suas médias e o número de instâncias válidas para aquele atributo.

Atributo	Válidos	Média	Min	Max	Atributo	Válidos	Média	Min	Max
Age	858	26,8	13	84	STDs: pelvic inflammatory disease	753	0,0	0	1
Number of sexual partners	832	2,5	1	28	STDs: genital herpes	753	0,0	0	1
First sexual intercourse	851	17,0	10	32	STDs: molluscum contagiosum	753	0,0	0	1
Number of pregnancies	802	2,3	0	11	STDs: AIDS	753	0,0	0	0



Smokes	845	0,1	0	1	STDs: HIV	753	0,0	0	1
Smokes (years)	845	1,2	0	37	STDs: Hepatitis B	753	0,0	0	1
Smokes (packs/year)	845	0,5	0	37	STDs: HPV	753	0,0	0	1
Hormonal Contraceptives	750	0,6	0	1	STDs: Number of diagnosis	858	0,1	0	3
Hormonal Contraceptives (years)	750	2,3	0	30	STDs: Time since first diagnosis	71	6,1	1	22
IUD	741	0,1	0	1	STDs: Time since last diagnosis	71	5,8	1	22
IUD (years)	741	0,5	0	19	Dx: Cancer	858	0,0	0	1
STDs	753	0,1	0	1	Dx: CIN	858	0,0	0	1
STDs (number)	753	0,2	0	4	Dx: HPV	858	0,0	0	1
STDs: condylomatosis	753	0,1	0	1	Dx	858	0,0	0	1
STDs: cervical condylomatosis	753	0,0	0	0	Hinselmann	858	0,0	0	1
STDs: vaginal condylomatosis	753	0,0	0	1	Schiller	858	0,1	0	1
STDs: vulvo-perineal condylomatosis	753	0,1	0	1	Citology	858	0,1	0	1
STDs: syphilis	753	0,0	0	1	Biopsy	858	0,1	0	1

**TABELA 1: Análise de média, valor máximo, valor mínimo e número de dados válidos dos atributos do conjunto de dados.**

Exposta a solução para o problema dos dados faltantes, optou-se por descartar as colunas “STDs: Time since first diagnosis” e “STDs: Time since last diagnosis”. Feito isso, foram excluídas todas as linhas que possuíam pelo menos um valor de atributo desconhecido. O conjunto de dados final ficou com 668 instâncias e 34 atributos, sendo o campo “Biopsy” utilizado como supervisão.

A normalização e escala dos dados foi feita removendo o valor mediano dos dados entre o primeiro e o terceiro quantil para cada atributo do conjunto de dados, forçando os seus valores a ficarem próximos a zero. Tal feito é necessário pois evita que haja interferência no grau de influência dos atributos no processo de classificação das redes, aumentando a capacidade de aprendizado da mesma (18). Além disso, o cálculo da mediana no intervalo interquantil reduz o impacto negativo que valores atípicos teriam sobre os dados (19).

Redes neurais com muitos parâmetros tem tendência a possuir relações correlacionais confusas entre os mesmos, produzindo um alto nível de ruído que pode influenciar na posterior capacidade de generalização das mesmas. Tal problema é conhecido como superadaptação (do inglês, *overfitting*) que faz com que uma rede neural já treinada não



tenha uma boa capacidade em classificar corretamente novos exemplos além dos contidos no conjunto de treinamento (20).

Para evitar a *superadaptação* foi empregada a técnica de *dropout* (20), que consiste na eliminação aleatória de alguns dos neurônios e suas conexões durante as épocas de treinamento das redes. O valor de *dropout* varia de 0 a 1 e consiste na taxa de neurônios eliminados aleatoriamente a cada rodada de aprendizado.

Como funções de ativação das camadas das arquiteturas, foram escolhidas duas opções, sendo elas: ELU (do inglês, *Exponential Linear Unit*) e Sigmoid. A função ELU, uma melhoria em relação a função RELU, é indicada para evitar uma eventual impossibilidade no treinamento das redes causada valores nulos dos pesos devido a ajustes negativos constantes para se adequar aos exemplos consumidos durante o treinamento (21). A função Sigmoid foi escolhida principalmente como função não-linear complementar alternativa devido seu amplo uso em redes neurais pois permite um treinamento geralmente rápido em redes neurais densas (22).

Com a finalidade de se encontrar os melhores parâmetros de configuração da RNA, foram treinadas 12 configurações de arquiteturas densas diferentes, variando-se o número de neurônios por camada, o valor de *dropout* e a função de ativação. Tais arquiteturas foram geradas por meio de uma técnica conhecida como *grid-search*, que facilita a criação de diversas arquiteturas de redes por meio de intervalos de parâmetros predefinidos (23). A tabela 2 mostra os parâmetros utilizados pela técnica de *grid-search* para a criação das arquiteturas.

Parâmetro	Possíveis Valores
Camadas	1; 128-1; 128-128-1
Função de Ativação	ELU; Sigmoid
Dropout	0; 0.4
Métricas Adicionais	Acurácia
Função Loss	Entropia Binária Cruzada
Otimizador	Adam

**TABELA 2: Parâmetros utilizados para a criação das arquiteturas utilizando o grid-search.**

O treinamento das RNAs foi feito em 300 épocas, ou seja, apresentando o conjunto de treinamento para cada uma delas de maneira completa 300 vezes. Por outro lado, o tamanho dos *minibatches* utilizados, quantos exemplos de treinamento eram exibidos por vez à rede antes do ajuste dos pesos, foi de 32.



Para validar o modelo proposto, foi utilizada a técnica de validação cruzada estratificada com múltiplas dobras. A validação cruzada é uma técnica para avaliar a capacidade de generalização de um modelo a partir de um conjunto de dados. Esta técnica é amplamente empregada em problemas onde o objetivo da modelagem é a predição. Busca-se então estimar o quão preciso é este modelo na prática, ou seja, o seu desempenho para um novo conjunto de dados nunca antes visto pela rede (24).

O conceito central das técnicas de validação cruzada é o particionamento do conjunto de dados em subconjuntos mutualmente exclusivos, e posteriormente, utiliza-se alguns destes subconjuntos para a estimação dos parâmetros do modelo (dados de treinamento) e o restante dos subconjuntos (dados de validação ou de teste) são empregados para determinar a capacidade de generalização de um modelo (24).

Foram utilizadas 3 dobras, fazendo com que todas as arquiteturas propostas pelo *grid-search* fossem treinadas e validadas 3 vezes, permitindo o cálculo da acurácia média de treinamento e validação.

## RESULTADOS E DISCUSSÃO

Após a criação e treinamento das arquiteturas, foram obtidos resultados referentes à acurácia do seu treinamento e validação utilizando o banco de dados proposto. O tempo total de treinamento e validação para o conjunto de redes com todas as arquiteturas propostas em todas as dobras foi de 7 minutos e 51 segundos.

Todas as RNAs treinadas apresentaram acurácia média de validação maior que 94% sendo a melhor delas com duas camadas ocultas, com 128 neurônios e uma camada de saída com um único neurônio e *dropout* de 0.4 entre as camadas. Esta, apresentou 96,2% de acurácia média nos dados de validação e 98% nos dados de treinamento.

A tabela 3 mostra um comparativo dos resultados de cada Rede Neural implementada utilizando o conjunto de dados de treinamento e validação.

Arquitetura	Função de Ativação	Dropout	Acurácia	
			Treinamento	Validação
1	Elu	0	95,5%	94,8%
128-1	Elu	0	97,6%	94,9%
128-128-1	Elu	0	99,2%	94,9%
1	Elu	0,4	95,5%	95,2%
128-1	Elu	0,4	97,3%	94,8%



128-128-1	Elu	0,4	98,0%	96,2%
1	Sigmoid	0	95,7%	95,1%
128-1	Sigmoid	0	97,0%	95,4%
128-128-1	Sigmoid	0	97,4%	95,4%
1	Sigmoid	0,4	95,5%	94,9%
128-1	Sigmoid	0,4	96,8%	95,5%
128-128-1	Sigmoid	0,4	95,5%	94,8%

**TABELA 3: Arquitetura e acurácia média apresentada por cada Rede Neural Artificial implementada nos conjuntos dados de treinamento e validação.**

Com a utilização do valor de *dropout* proposto na melhor arquitetura, não se observou *superadaptação*, tendo a rede uma boa capacidade de generalização quando são apresentados a novos exemplos, isto é, quando a rede é utilizada para prever o resultado da biópsia de uma paciente que não estava no banco de dados de treinamento.

## CONCLUSÃO

A Rede Neural Artificial projetada foi capaz de prever o resultado da biópsia para 96,2% dos pacientes presentes no conjunto de dados de validação, mostrando que Redes Neurais Artificiais podem ser utilizadas no auxílio ao diagnóstico de câncer cervical.

Esta pesquisa contribui para o avanço do uso da Inteligência Artificial em Sistemas de Suporte à Decisão Clínica com foco na Oncologia, possibilitando a futura elaboração de sistemas de diagnóstico automatizados, de maneira a auxiliar profissionais da atenção básica de saúde a realizarem de maneira mais precisa um diagnóstico e encaminhamento precoce do câncer cervical, evitando assim, eventuais casos de óbito e melhorando a eficiência do tratamento.

## REFERÊNCIAS

1. Wechsler R, Anção MS, Campos CJR de, Sigulem D. A informática no consultório médico. J Pediatr (Rio J). 2003;79:S3–12.
2. Arbyn M, Castellsague X, de Sanjose S, Bruni L, Saraiya M, Bray F, et al. Worldwide burden of cervical cancer in 2008. Ann Oncol. 2011;22(12):2675–86.
3. Instituto Nacional de Câncer José Alencar Gomes da Silva. Estimativa 2018: Incidência de Câncer no Brasil. Rio de Janeiro, RJ: Ministério da Saúde; 2018.
4. Instituto Nacional de Câncer José Alencar Gomes da Silva. Diretrizes Brasileiras para o Rastreamento do Câncer do Colo do Útero. Rio de Janeiro, RJ: Ministério da Saúde; 2016.
5. Lobo LC. Inteligência Artificial e Medicina. Rev Bras Educ Med. 2017;41(2):185–93.
6. Gyawali B. Does global oncology need artificial intelligence? Lancet Oncol.



- 2018;19(5):599–600.
7. Sim I, Gorman P, Greenes RA, Haynes RB, Kaplan B, Lehmann H, et al. Clinical decision support systems for the practice of evidence-based medicine. *J Am Med Inform Assoc.* 2001;8(6):527–34.
  8. Schmidhuber J. Deep learning in neural networks: An overview. *Neural Networks.* 2015;61:85–117.
  9. Huang G-B, Zhu Q-Y, Siew C-K. Extreme learning machine: Theory and applications. *Neurocomputing.* 2006;70(1–3):489–501.
  10. Haykin SS. *Neural networks: A comprehensive foundation.* Prentice Hall; 1999. 842 p.
  11. Chollet F. *Keras: The Python Deep Learning library.* Keras.io; 2015.
  12. Van Rossum G, Drake Jr FL. *Python tutorial.* Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands; 1995.
  13. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. TensorFlow: A System for Large-Scale Machine Learning. In: *Osd.* 2016. p. 265–83.
  14. Goldsborough P. A Tour of TensorFlow. *CoRR.* 2016;abs/1610.0.
  15. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. *J Mach Learn Res.* 2011;12(Oct):2825–30.
  16. Fernandes K, Cardoso JS, Fernandes J. Transfer Learning with Partial Observability Applied to Cervical Cancer Screening. In Springer, Cham; 2017. p. 243–50.
  17. García-Laencina Pedro J., and Sancho-Gómez J-L, R and F-VA. Pattern classification with missing data: a review. *Neural Comput Appl.* 2010;19(2):263–82.
  18. Jayalakshmi T, Santhakumaran A. Statistical normalization and back propagation for classification. *Int J Comput Theory Eng.* 2011;3(1):89.
  19. Buitinck L, Louppe G, Blondel M, Pedregosa F, Mueller A, Grisel O, et al. RobustScaler: Scikit-Learn Documentation [Internet]. 2013 [citado 7 de abril de 2018]. Available at: <http://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.RobustScaler.html>
  20. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. *J Mach Learn Res.* 2014;15(1):1929–58.
  21. Clevert D-A, Unterthiner T, Hochreiter S. Fast and Accurate Deep Network Learning by Exponential Linear Units(ELUs). *CoRR.* 2015;abs/1511.07289.
  22. Karlik B, Olgac AV. Performance analysis of various activation functions in generalized MLP architectures of neural networks. *Int J Artif Intell Expert Syst.* 2011;1(4):111–22.
  23. Buitinck L, Louppe G, Blondel M, Pedregosa F, Mueller A, Grisel O, et al. API design for machine learning software: experiences from the scikit-learn project. *CoRR.* 2013;abs/1309.0.
  24. Baker R, Isotani S, Carvalho A, Carvalho A. Mineração de Dados Educacionais: Oportunidades para o Brasil. *Rev Bras Informática na Educ.* 2011;19(02):03.