

Inteligência Médica Computadorizada: um estudo sobre técnicas de Inteligência Artificial no diagnóstico e tratamento do Câncer de Colo de Útero

Tatiane R. da Silva, Prof. Dr. David M. Fernandes

Universidade de Uberaba

Coordenação da Área de Engenharias

Av. Nenê Sabino, 1801, Campus Aeroporto

Uberaba/MG, Brasil, 38.055-500

rodriguestatiane.silva@gmail.com, davidmaikel@hotmail.com

Resumo – o câncer é uma das principais causas de morte na sociedade contemporânea. O diagnóstico tardio constitui um dos principais problemas enfrentados por portadores da doença, provenientes do desconhecimento dos sintomas por parte do paciente ou ainda por erro na descrição médica. A fim de auxiliar na prevenção do diagnóstico tardio e obter maior precisão nos resultados, a computação entra em conjunto com a medicina no manejo e na interpretação de dados obtidos de pacientes, tanto por meio de informações passadas pelos próprios, quanto por imagens obtidas por exames específicos. O propósito deste trabalho é realizar uma revisão das técnicas computacionais mais utilizadas na detecção do câncer de colo do útero, um dos tipos de câncer com maior taxa de mortalidade no Brasil, tendo enfoque nas técnicas que se baseiam no princípio da Inteligência Artificial (IA).

Palavras-chave: câncer, colo de útero, diagnóstico, Inteligência Artificial.

Abstract - Cancer is one of the main causes of death in contemporary society. Late diagnosis is one of the leading problems faced by patients that suffer from this disease. This late diagnosis is due to a patient's lack of awareness of the symptoms or even due to an error in the medical description. In order to prevent this problem and to obtain greater accuracy in test results, computation comes together with medicine in the management and interpretation of data obtained from patients, both through information given by themselves and images acquired in specific tests. The purpose of this project is to review the most popular computational techniques utilized for the detection of cervical cancer, one of the types of cancer with the highest mortality rate in

Brazil, focusing on the techniques based on the principle of Artificial Intelligence (AI).

Keywords: cancer, cervical cancer, diagnosis, Artificial Intelligence.

I. INTRODUÇÃO

O câncer de colo de útero é uma doença na qual células cancerígenas, tumorais, se formam nos tecidos do colo de útero das mulheres. De acordo com a Organização Mundial da Saúde (OMS), em 2012, o câncer de colo de útero foi o quarto tipo de câncer mais comum entre as mulheres, com 527 mil novos casos, sendo responsável pelo óbito de 265 mil mulheres no mesmo ano. Ainda de acordo com a OMS Brasil (2017), sua incidência é maior em países menos desenvolvidos, com índice de óbito de 87%, quando comparada aos países mais desenvolvidos.

Dos principais fatores que levam o câncer de colo de útero ao seu estágio final, é possível destacar:

- ❖ Atraso das mulheres em procurar ajuda médica durante os primeiros sintomas;
- ❖ Falta de equipe especializada;
- ❖ Atraso na entrega de resultados de exames;
- ❖ Interpretação errônea de resultados;
- ❖ Atraso na detecção e correção de erros de diagnóstico.

Todos esses fatores podem ser tratados ou até mesmo eliminados se for feito o uso de sistemas informatizados. O objetivo desta pesquisa é mostrar o uso da computação

associado à medicina no diagnóstico e tratamento do câncer de colo de útero. Essa junção pode ser presenciada desde o início do acompanhamento, durante os primeiros relatos de um paciente, até os processos finais de um tratamento, como será estudado neste artigo.

Inúmeras pesquisas, de ambas as áreas, relacionam as duas ciências. A computação oferece opções mais exatas, rápidas e eficazes de diagnósticos, enquanto a medicina faz com que a informática possa processar a enorme quantidade de dados obtidos e fornecer resultados exatos em todas suas aplicações.

Se tratando do tema proposto, as tecnologias aplicadas no tratamento do câncer de colo de útero envolvem, na maioria das vezes, técnicas de Inteligência Artificial (IA). Sistemas que se baseiam nesse conceito, devem imitar a capacidade de raciocínio humano sendo capazes de adquirir e integrar conhecimentos sobre a área em que são aplicados, melhorando progressivamente seu desempenho à fim de alcançar seus objetivos. Seus conceitos abordam diversas áreas derivadas principalmente da matemática aplicada, estatística, ciência e engenharia de computação.

Atualmente, a utilização de técnicas de IA em pesquisas médicas tem mostrado excelentes resultados devido ao fato de que seu funcionamento seja extremamente próximo ao comportamento humano. Os estudos e pesquisas que relacionam as duas áreas servem para ampliar a busca por especialistas na realização de trabalhos similares, aumentando as chances de novas descobertas. Essa pesquisa pondera-se com intuito semelhante; justifica-se como uma forma de esclarecer detalhes sobre o tema e então atrair mais estudos para casos similares.

Para esta pesquisa, foram realizados estudos de casos e pesquisas bibliográficas em publicações científicas e, ao fim da pesquisa, foi desenvolvido um sistema simples como o propósito de por em prática os conhecimentos técnicos adquiridos. Foram analisados três projetos: dois deles de universidades espanholas, e o último de uma universidade brasileira. Seus objetivos, metodologias e resultados foram detalhados. Foi realizada também uma discussão acerca dos trabalhos abordados e então realizadas algumas críticas e propostas de melhoria. Ao fim, depois de realizar as

considerações finais da pesquisa, encontra-se disponibilizado o código completo desenvolvido no projeto.

II. CÂNCER DE COLO DE ÚTERO

Segundo a OMS Brasil (2017), o câncer de colo de útero mata mais de 250 mil mulheres por ano, e desse total, cerca de 85% dos óbitos ocorrem em países de renda média e baixa. A implementação de programas adequados de rastreio do câncer tem diminuído a incidência e mortalidade em países desenvolvidos, como afirma a Agência Internacional de Pesquisa sobre Câncer (IARC), da OMS Brasil (2017). O quadro é distinto em países pobres, onde os programas de detecção não são implementados ou têm sido inadequados, deixando a taxa de incidência e mortalidade em constante crescimento.

A. Formas de diagnóstico

Atualmente, a principal forma de detecção precoce do câncer de colo de útero é feita por meio do exame de Papanicolau, um exame citopatológico que possibilita o diagnóstico de alterações celulares por meio de sua morfologia diante a ação de três corantes específicos. A efetividade desse exame enquadra-se em uma cobertura de 80% da população feminina e pode diminuir a taxa de mortalidade da doença em até 50% (GASPERIN, BOING, & KUPEK, 2011).

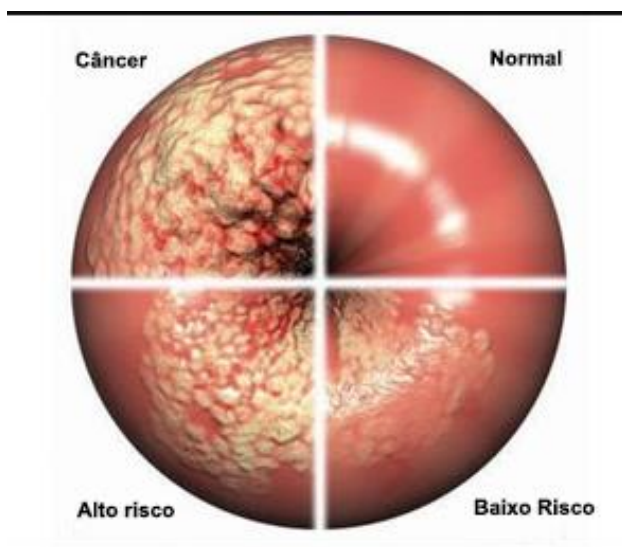
Entretanto, o exame de Papanicolau ainda não é tido como uma forma de prevenção de doenças. Sua realização muitas vezes não segue o padrão necessário de repetição determinado pela OMS Brasil (2017), com repetição a cada ano, tornando-se oportuno, ou seja, não é uma prioridade na saúde.

Além disso, apesar de sua grande eficiência na detecção de diversos tipos de doenças, o Papanicolau só serve para detectar a possibilidade do surgimento de um câncer de colo de útero. É necessário aliá-lo a outros tipos de exames de detecção. Exames de biópsia são eficientes nesses casos, entretanto podem ser aliados a exames que ofereçam um grau mais elevado de eficiência, como é o caso do exame de

colposcopia e cistoscopia (KARIMI, MOUSAVI & DEGHANI, 2011).

No exame de colposcopia, a vagina e o colo do útero são visualizados por meio de um aparelho chamado colpóscopio. Esse aparelho possui lentes que permitem a visualização de uma imagem com aumento de 10 a 40 vezes, fornecendo uma imagem mais detalhada no processo. A Fig. 1 ilustra como é visto o colo do útero por meio do colpóscopio e também a atuação de células anormais até a evolução para o câncer.

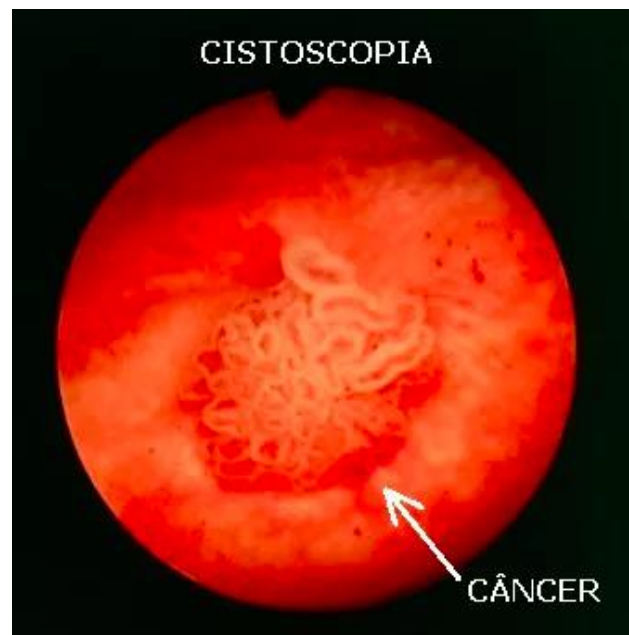
Fig. 1 - Vista frontal do colo de útero fornecida por um colpóscopio.



Fonte: Askabide (2014).

No segundo exame, cistoscopia, uma câmera é presa na extremidade de um tubo delgado e inserido através da uretra até o interior da bexiga. Nesse processo é possível verificar se o câncer se espalhou por outras regiões, como a bexiga por exemplo, por meio das imagens analisadas. A Fig. 2 exibe a imagem captada durante um exame de cistoscopia, onde é detectado um tumor na bexiga.

Fig. 2 - Imagem de tumor capturada em exame de cistoscopia.



Fonte: Magnani (2015).

Ambos exames se baseiam na análise de imagens, geralmente feita a olho nu por um médico especializado. Entretanto, novas técnicas computacionais estão sendo aplicadas nessa análise, para ampliar o grau de confiabilidade de diagnóstico do câncer e prover um resultado, muitas vezes, perto dos 100% de eficiência. As técnicas mais aplicadas nesse contexto, são técnicas capazes de reconhecer padrões nas imagens fornecidas nos resultados e analisar a evolução de tais casos de maneira eficaz.

É possível também realizar um diagnóstico precoce por meio dos sintomas apresentados pelo paciente. Relatos de dores na região pélvica, sangramentos em épocas inapropriadas (diferente do período menstrual), corrimentos com aspectos diferentes do comum (cor e/ou odores anormais), dores e hemorragia após relação sexual são características analisadas durante consultas médicas e que devem ser considerados.

Atualmente, para se obter uma análise mais eficaz dos dados obtidos, estão sendo aplicadas técnicas computacionais baseadas no contexto de Inteligência Artificial. Algumas técnicas são capazes de traçar relações entre os resultados comprovados por exames e os relatos obtidos no processo de

anamnese, gerar padrões de processamento e respostas, aprender com os resultados gerados e ainda atualizar seu aprendizado com possíveis mudanças, próprias do comportamento humano.

III. INTELIGÊNCIA ARTIFICIAL

A inteligência artificial se define como o campo da ciência e da engenharia que se ocupa com a compreensão computacional, ou seja, o comportamento inteligente e passível de aprendizagem automático de um computador. Mitchell (ref) descreve a inteligência artificial em termos técnicos como: “Um computador aprende a experiência (E) com respeito a alguma classe de tarefas (T) e medida de desempenho (P), se seu desempenho nas tarefas T, em relação a P, melhora com a experiência E”.

No centro da Inteligência Artificial, encontra-se o processo de aprendizagem automático, descrito pela representação e generalização. Os dados inseridos são representados com base em funções de avaliação. A generalização consiste no fato de que o sistema deva funcionar bem frente à criação de novos dados; é a habilidade de um programa de funcionar com exatidão frente à entrada de novos dados. Assim, seu objetivo principal é generalizar a partir da experiência obtida: extrair dos dados informações, uma função que o permita fazer previsões úteis em novos casos.

São diversas as técnicas computacionais que baseiam seu funcionamento no contexto de Inteligência Artificial. No campo da medicina, devido à grande quantidade de conhecimento necessário para analisar e resolver cada caso, algumas técnicas se enquadram melhores que outras no manejo de dados. As técnicas abordadas nesta pesquisa são os Sistemas Especialistas e as Redes Neurais.

A. Sistemas Especialistas

Segundo Cardoso et al., sistemas especialistas são “sistemas de computação que realizam funções semelhantes àquelas normalmente executadas por um especialista humano”. Ainda segundo o autor, esse tipo de programa é carregado com características prévias do problema a ser

analisado, e escolhe qual decisão tomar para chegar à solução mais eficiente.

Sua arquitetura se divide em dois componentes: um banco de dados com todas as informações relevantes sobre o problema a ser analisado e o código final com métodos baseados nos conceitos de I.A.

B. Redes Neurais Artificiais

As Redes Neurais Artificiais são amplamente utilizadas na medicina por sua enorme flexibilidade e dinamismo. De maneira generalizada, consta de uma rede de várias camadas que, antes de serem implementadas em qualquer caso, necessitam ser previamente treinadas, capacitadas. Suas camadas podem ter aspectos de inteligência compartilhados ou convergentes, que devem chegar à uma solução de maneira rápida e eficiente.

Segundo Carvalho (2009), as principais facetas de uma Rede Neural Artificial são sua arquitetura e seu algoritmo de aprendizagem. Essas redes devem ser treinadas por meio de exemplos, e seu conhecimento fica dependente disso. Em seguida, o algoritmo generaliza os dados e memoriza seu conhecimento dentro de parâmetros adaptados à rede, conhecidos também como pesos. Uma função de ativação é sempre definida nessa técnica; ela será responsável por filtrar os elementos de entradas de acordo com seu peso respectivo e então realizar ou não seu processamento. É importante ressaltar que após receberem seus valores de pesos, em geral de maneira aleatória, as entradas passam a ser chamadas de neurônios.

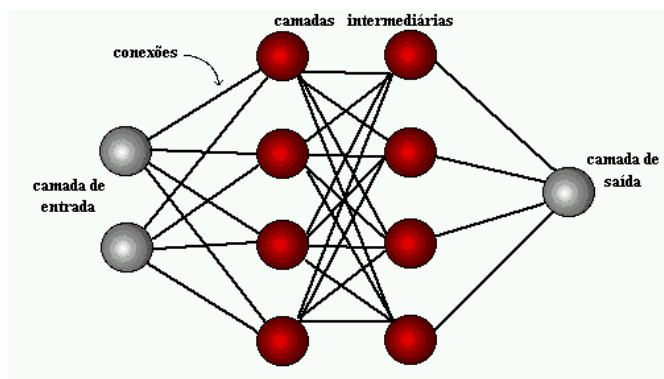
1. Perceptron

Perceptron é uma das derivações de Rede Neural. Segundo Aparicio (2015), esse tipo de algoritmo calcula um sinal de entrada filho por meio de dados de entrada e padrões neurais. Essa entrada filha, é resultante do processamento da entrada inicial por uma função de ativação. Tal função pode ser encontrada em uma única parte do programa, em uma única camada, ou separada em partes, ou seja, camadas. Sendo assim, tem-se duas classificações: *perceptron* simples e *perceptron* multicamadas.

Perceptron multicamadas é o mais utilizado em pesquisas de grande complexidade que envolvem Inteligência Artificial.

Programas que precisam ser treinados em várias etapas, geralmente utilizam essa forma de processamento, onde o aprendizado é supervisionado. A Fig. 3 ilustra o funcionamento de um algoritmo *perceptron* multicamadas.

Fig. 3 - Modelo *perceptron* multicamadas.



Fonte: Carvalho (2009).

IV. METODOLOGIA

A metodologia aplicada neste trabalho é a pesquisa descritiva de publicações que implementam técnicas de Inteligência Artificial no estudo e pesquisa do câncer de colo de útero. No processo de busca de materiais para análise, o objetivo encontrar trabalhos que utilizassem diferentes técnicas de inteligência artificial em diferentes estágios do estudo do câncer de colo de útero, desde seu diagnóstico até o tratamento final guiado por imagens.

Em seguida, após investigar toda a estrutura dos projetos, desde sua metodologia até os resultados e conclusões, as técnicas utilizadas foram avaliadas quanto sua eficácia no propósito em que se aplicaram e em alguns casos, foram sugeridos novos métodos de pesquisa, métodos mais atualizados.

V. APLICAÇÕES MÉDICAS INTELIGENTES

Neste capítulo são apresentadas 3 publicações que analisam a aplicação de técnicas de I.A. O primeiro projeto refere-se à uma tese de doutorado apresentada na Universidade de Valladolid na Espanha, no ano de 2015. A técnica aborda o uso de uma rede neural na predição do câncer de colo de útero, ou seja, é aplicação de I.A. ainda no diagnóstico do câncer, baseando-se no processo de anamnese.

O segundo trabalho analisado também foca no diagnóstico do câncer de colo de útero, porém baseia seu funcionamento em um Sistema Especialista, que atua diretamente nas informações passadas pelos pacientes. O trabalho de conclusão de curso foi realizado em uma maternidade na cidade de Chiclayo no Peru, em 2012.

A última pesquisa analisada foi realizada em Santa Catarina, Brasil, como trabalho de conclusão de curso. O estudo mostra a aplicação de Redes Neurais no processamento de imagens provenientes de exames realizados no tratamento do câncer de colo de útero, no ano de 2007.

A. Rede Neural Artificial como modelo preditivo em uma unidade de patologia cervical

O objetivo geral desta pesquisa foi criar um modelo de prognóstico baseado em Redes Neurais Artificiais que pudesse auxiliar nas decisões clínicas diárias do hospital Universitário de Valladolid.

Para isso, foram destacados os objetivos específicos da pesquisa, que foram:

- ❖ Identificar e gerar o melhor modelo de Rede Neural para o projeto;
- ❖ Elaborar seu próprio software específico para o caso;
- ❖ Elaborar padrões de acordo com a base de dados antiga do hospital (antes da implementação do software).

Para utilizar o programa, as pacientes passam primeiro por um processo de seleção. Características como idade, presença de alguma enfermidade e vida ativa sexual são analisadas, e, baseado nos critérios da pesquisa, algumas pacientes são excluídos, enquanto outros chegam à fase final de teste.

1. O algoritmo

Após o processo de filtragem e seleção de pacientes, se inicia o desenvolvimento do código capaz de atender às necessidades. O modelo de Rede Neural utilizado na pesquisa é o modelo *Perceptron* Multicamadas. Segundo a autora do projeto, esse modelo é amplamente utilizado em aplicações

médicas devido a suas boas propriedades de aproximação. Esse algoritmo é ampliado com uma camada de escalonamento em seus dados de entrada e uma camada reversa em seus dados de saída, para fazer com que seus resultados sejam normalizados.

O programa conta com 5 variáveis de saída:

- ❖ Idade (entre 25 e 54 anos);
- ❖ Citologia (sem diagnóstico de câncer confirmado);
- ❖ HPV – Human Papiloma Virus (presença do vírus, é a causa muitas vezes do câncer, com ou sem sintomas);
- ❖ Biópsia (resultado da biópsia não confirma o câncer);
- ❖ P16/KI67 (alterações nessa proteína podem confirmar a presença do câncer);

Como resultado do processamento destas 5 variáveis, temos o resultado que é mantido na variável Diagnóstico. No projeto, a Rede Neural conta com duas camadas de *perceptron*, sendo que a primeira (oculta) tem uma função de ativação senoidal e a segunda, uma função linear. O programa conta também com 9 neurônios na primeira camada, que podem ser alterados de acordo com os resultados e com a complexidade de manutenção.

a. Função de rendimento

Essa função serve para ajustar a rede neural e todo seu processamento aos dados recebidos. Para esta pesquisa, foi usada uma função de rendimento quadrática, de segundo grau, e funciona dividindo a soma dos valores de erro entre as saídas da rede e seus correspondentes objetivos. Se o erro é um valor diferente de zero, a rede precisa alterar seu processamento.

b. Função de treinamento

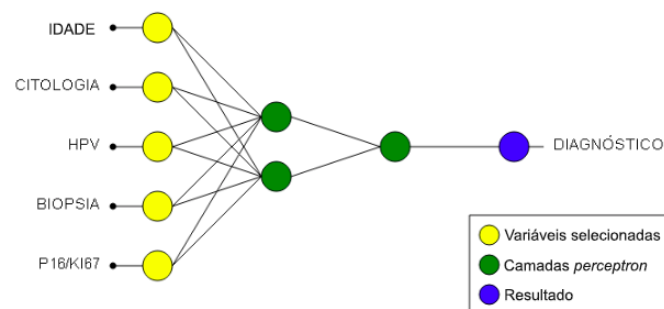
Para fazer com que uma Rede Neural simule o raciocínio humano, é necessário fazer com que a mesma aprenda sobre o conteúdo em que vai ser aplicada. O processo de

aprendizagem serve para que a Rede retorne, ao final, o melhor resultado possível.

Nesta pesquisa, o método de treinamento (ou aprendizagem) aplicado foi o método de otimização *Quase-Newton* do tipo BFGS, amplamente utilizado em softwares que necessitam de manipulação de matrizes de grande porte. A utilização desse método é padronizada, mas seus valores são relativos: estão dependentes do local de aplicação e das necessidades.

Ao final destes processos, obteve-se o seguinte esquema de organização da Rede Neural, como ilustrado na Fig. 4, a seguir.

Fig. 4 - Esquema final da Rede Neural utilizada pela Universidade de Valladolid.



Fonte: Aparicio (2015).

c. Análise de Validação

Nessa etapa, ocorre a comparação entre os resultados previstos pela rede e seus respectivos valores num conjunto de dados independentes. Se a comparação tem um resultado aceitável, a rede neural avança para a etapa final de processamento.

Em uma análise de validação, o método mais comum aplicado em Redes Neurais é o método de Regressão Linear. Nessa análise são considerados 3 parâmetros, tendo como base de análise um plano cartesiano com os resultados:

- ❖ a ordenada em relação à origem;
- ❖ o valor de inclinação da reta;
- ❖ o coeficiente entre essas duas variáveis.

Caso o esquema de validação estiver perfeito, ou seja, o valor de saída for igual ao objetivo pretendido, o valor de origem estaria em 0, a inclinação seria 1 e o coeficiente também seria 1. Para a pesquisa analisada, foram encontrados

como resultado: -0,0239 para a ordenada; 0,594 para a inclinação e 0,741 para o coeficiente. Os valores indicam que a Rede Neural, naquele momento, funcionava relativamente bem, mas é passível de melhora.

d. Modo de produção

Após passar pelo processo de validação, a Rede Neural é passa para o modo de produção, onde são definidas as fórmulas padronizadas para tornar os valores informados pelos pacientes passíveis de cálculos. Por exemplo, os valores retornados em uma citologia, não são resultados numéricos; são apresentadas descrições escritas das possíveis enfermidades encontradas. É necessário transformar tais descrições em valores números. Esses valores, a faixa de intervalo é determinada segundo os padrões utilizados no momento inicial de implementação da Rede Neural. A Fig. 5 mostra um trecho do código implementado na pesquisa analisada, onde é possível ver os cálculos aplicados no processo de conversão dos valores das 5 variáveis informadas pelo paciente.

Fig. 5 - Trecho de código da RNA.

```
scaled_IDADE=2*(IDADE-20)/(68-20)-1;
scaled_CITOLOGIA=2*(CITOLOGIA-0)/(5-0)-1;
scaled_HPVP=2*(HPV-0)/(3-0)-1;
scaled_BIOPSIA=2*(BIOPSIA-0)/(4-0)-1;
scaled_P16/K167=2*(P16/K167-0)/(1-0)-1;

y_1_1=tanh(-0.18546+0.154098*scaled_EDAD-0.0843001*scaled_CITOLOGIA-
0.326473*scaled_HPVP+1.22038*scaled_BIOPSIA-0.0590852*scaled_P16/K167);

y_1_2=tanh(0.733037+0.15708*scaled_EDAD-0.161476*scaled_CITOLOGIA-
1.03068*scaled_HPVP+0.104929*scaled_BIOPSIA+0.0370204*scaled_P16/K167);

scaled_DIAGNOSTICO=(0.309632+0.930792*y_1_1-0.696463*y_1_2);

DIAGNOSTICO=0.5*(scaled_DIAGNOSTICO+1.0)*(4-0)+0;
```

Fonte: Aparicio (2015).

Para o paciente, é necessário somente informar seus dados ao médico (variáveis de entrada), e ele será responsável por inserir os dados no programa, seguindo uma tabela de conversão. Os valores dessa tabela não foram informados na pesquisa analisada.

O formulário exibido para o médico encontra-se na Fig. 6. As telas com os resultados se encontram na Fig. 7.

Fig. 6 - Formulário de entrada de dados.

Name	Value
1 EDAD	39
2 CITOLOGIA	1
3 HPV	3
4 BIOPSIA	2
5 P16/K167	1

Fonte: Aparicio (2015).

Fig. 7 - Telas com resultado do diagnóstico.

EDAD	CITOLOGIA	HPV	BIOPSIA	P16/K167	DIAGNOSTICO
39	1	3	2	1	1.95415134

↓

EDAD	CITOLOGIA	HPV	BIOPSIA	P16/K167	DIAGNOSTICO
39	1	3	2	1	CIN II

Fonte: Aparicio (2015).

e. Resultados

O programa analisado foi testado desde 1 de janeiro de 2013 até 15 de junho de 2015. Do total de pacientes (197) 60% foram direcionados para auxiliar no processo de treinamento da rede, 20% para o processo de generalização, e 20% selecionados para realizar o processo de validação.

As diferenças entre os valores ideais e os valores encontrados nos 3 parâmetros de validação, resultaram em uma taxa de erro média de 13%, valor considerado baixo pelo autor da pesquisa. Os casos em que o erro ultrapassou essa marca foram em pequeno número e neste caso, são grandes as chances de ter ocorrido falhas durante a inserção dos dados pelo médico. Sendo assim, é possível afirmar que a Rede Neural analisada é uma rede preditiva.

B. Desenvolvimento de um sistema especialista Web para o diagnóstico precoce do câncer de colo de útero na Clínica Maternidade "Belen"

Este projeto, desenvolvido na Universidade Católica de Santo Toribio de Mogrovejo apresenta um sistema especialista baseado em IA. O objetivo desse programa é desenvolver um Sistema Especialista Web que dará suporte na realização de diagnósticos precoce de câncer de colo útero baseado em sintomas e antecedentes clínicos de pacientes de uma clínica e maternidade.

O funcionamento desse Sistema Especialista é supervisionado por um médico ginecologista, Dr. Daniel Paredes, da própria clínica, especialista em estudos de câncer de colo de útero. Tal médico forneceu de sua própria base dados e de maneira autorizada, informações sobre 113 casos de pacientes. Todos os pacientes tinham câncer de colo de útero já confirmados por exames. Ou seja, para mostrar que é eficiente, o sistema deve confirmar tais diagnósticos por meio de seu funcionamento.

1- Decisão de diagnóstico

Durante a consulta com o ginecologista, o paciente passa por uma etapa de perguntas e respostas em um formulário *online*. As respostas são inseridas pelo ginecologista e algumas respostas precisariam de confirmação. Por exemplo, quando perguntado sobre os resultados do último exame de Papanicolau realizado pela paciente, é necessária a apresentação do resultado do exame, impresso ou pela *internet*.

As questões realizadas são divididas em 3 categorias:

- ❖ Antecedentes;
- ❖ Sintomas;
- ❖ Exames Clínicos.

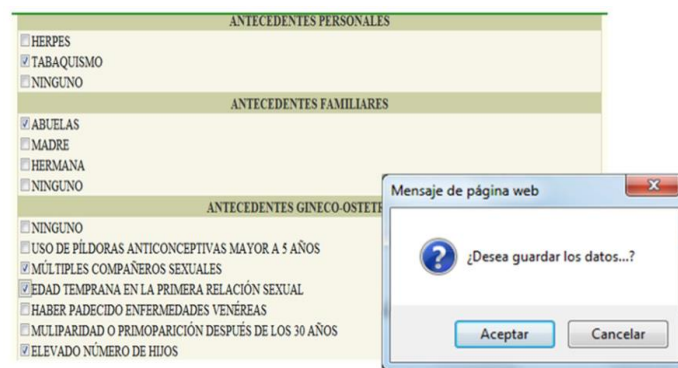
Na categoria antecedentes, dados pessoais como uso do tabaco e contágio de outras doenças, câncer em outros membros da família e assuntos ginecológicos foram coletados. Em relação aos sintomas, dores, sangramentos irregulares, perda de peso, entre outros, são questionados e ponderados de acordo com o nível em que se apresentam, por exemplo, um sangramento leve e passageiro tem menor peso de decisão se comparado ao um sangramento grave com

muitos dias de duração. E por fim, são pedidos resultados de exames clínicos, como o Papanicolau.

Após obter todas as informações necessárias do paciente, é realizada uma distribuição de valores e de pesos, para cada característica analisada. Para essa pesquisa analisada, foi desenvolvida uma tabela com 50 sintomas e antecedentes, que compõe a base de conhecimento do programa. Tais atributos são características e sintomas comuns em pacientes que apresentam o câncer de colo de útero. Destes 50 atributos, somente os relatados pelo paciente serão validados com valores e pesos. Os valores dos pesos são alterados de acordo com características e antecedentes do paciente.

A Fig. 8 mostra a tela de seleção de sintomas utilizada no sistema.

Fig. 8 – Tela para seleção de sintomas do Sistema Especialista.



Fonte: Sanchez (2012).

2- Cálculos

Para obter o resultado final do diagnóstico, foram desenvolvidas funções matemáticas de validação. Os sintomas foram divididos em primários e secundários, entretanto essa divisão não altera a forma de cálculo.

Os pesos de sintomas primários são multiplicados pelos valores de sintomas primários, como mostrado na equação (1). O mesmo ocorre nos sintomas secundários e antecedentes: seus pesos e valores são multiplicados, respectivamente.

$$sintPrimário = valor * peso; \quad (1)$$

Em seguida, uma outra variável fica responsável por realizar a soma dos resultados das multiplicações (2), ou seja,

recebe a somatória de todos os cálculos realizados anteriormente.

$$total = sintPrimario + sintSecundario + antecedente; \quad (2)$$

Depois de finalizado o cálculo total, analisa-se a porcentagem. O valor total calculado equivale à uma porcentagem do geral, ou seja, uma porcentagem dos 50 atributos, todos calculados com seus respectivos pesos e valores.

3- Resultados

Dos 113 casos passados pelo médico especialista, 25 não tinham câncer de colo de útero, enquanto o restante continha a doença. Desse total, 110 casos foram acertados pelo sistema especialista, enquanto os outros 3 não possuíam as informações necessárias, tornando a distribuição de pesos incorreta. Esses dados equivalem à uma eficiência de 97% no diagnóstico correto dos casos. A Fig.9 mostra a tela final de resultado de um dos pacientes diagnosticados.

Fig. 9 – Tela de resultado do Sistema Especialista.

Antecedentes gineco-obstetricos:

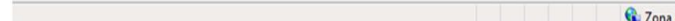
- Uso de píldoras anticonceptivas mayor a 5 años
- Edad temprana en la primera relación sexual

Sintomas:

- Dolor en las piernas
- Dolor pélvico
- Hinchazón en los ganglios inguinales
- Inflamación en una sola pierna mayor a diez días
- Leucorrea (flujo blanquecino y viscoso con mal olor)
- Sangrado o dolor después de la relación sexual
- Secreción vaginal continua que puede ser pálida, acuosa, rosada, marron, sanguinolenta o de olor fetido
- Cambios en la menstruación
- Dolor en las piernas
- Fatiga

Posible cáncer al cuello uterino, acudir a su médico.

Datos se guardaron satisfactoriamente
 Puntaje sintomas primarios: 82
 Puntaje sintomas secundario: 10
 Puntaje total : 92



Fonte: Sanchez (2012).

Além dessa alta eficiência de resultado, o sistema agiliza o diagnóstico. A comprovação de um câncer de colo de útero pode levar dias ou até meses para ser realizada. Em geral, na clínica onde o sistema foi aplicado, o resultado de uma comprovação de câncer de colo de útero leva em torno de 15

dias. Com a utilização do sistema, incluindo o tempo de entrevista com o paciente, todo o processo levou ao máximo 45 minutos para ser finalizado.

C. Diagnóstico automatizado de doenças no colo de útero baseado e Redes Neurais Artificiais e processamento de imagens digitais

Este projeto de pesquisa foi desenvolvido em conjunto por alunos e docentes da Universidade do Extremo Sul Catarinense apresenta os resultados da aplicação de um sistema de auxílio ao diagnóstico de doenças presentes no colo do útero utilizando RNA's e processamento de imagens digitais.

O projeto, assim como o primeiro estudo analisado, também utilizou o modelo de algoritmo perceptron multicamada, que teve como base de treinamento a tecnologia backpropagation. Além disso, conta com uma API (*Application Programming Interface*) que auxilia no processamento lógico das imagens utilizadas. Ao fim, foi desenvolvido também um registro computacional de cada paciente para facilitar o armazenamento e consulta de informações ginecológicas do paciente.

1- Metodologia

Em um primeiro momento, os autores relatam a realização de um estado da arte. Isso significa que foram pesquisados projetos de cunho similar com o intuito de evitar plágios. Em seguida, iniciou-se a descrição do modelo de RNA utilizado no projeto.

Os autores utilizaram um algoritmo composto por duas camadas de processamento, além das camadas de entrada e saída. As duas camadas de processamento são descritas como *forward* e *backward*, que utilizam a técnica de *backpropagation* em seus códigos.

Segundo Tissot, Camargo e Pozo (2012), os algoritmos que utilizam essa técnica têm o aprendizado feito por ajuste iterativo de seus pesos, com o objetivo de minimizar as diferenças entre os valores da saída desejada e os valores da

saída encontrada. Dessa maneira, a rede precisa ser inicializada com pesos aleatórios. O vetor de saída resultante do processamento é comparado com o vetor de saída desejado e a diferença encontrada resulta num sinal de erro, que é retropropagado através da rede para que seja possível realizar o ajuste dos pesos.

Na pesquisa analisada, a camada *forward* percorre um caminho direto entre a camada de entrada e de saída, sendo utilizada para calcular a saída da rede para um respectivo valor de entrada. Já a camada *backward* percorre o caminho contrário: ela vai da camada de saída para a camada de entrada, calculando então possíveis erros entre os valores fornecidos.

a- Processamento de imagens

As imagens utilizadas nesse projeto de pesquisa, são imagens obtidas durante a realização de exames de colposcopia realizados por um dos autores do projeto, a Prof. Esp. Sandra Aparecida Manenti, que ministra aulas no curso de medicina da universidade.

Para realizar o processamento digital das imagens, foi utilizado no projeto uma API Java que disponibiliza e implementa uma enorme gama de operações de manipulação de imagem. Para o projeto atual, foram selecionados os seguintes filtros:

- ❖ Aumento de Contraste (transferência radiométrica em cada "pixel", com o objetivo de aumentar a discriminação visual entre os objetos presentes na imagem.);
- ❖ Limiarização (processo de segmentação de imagens baseado na diferença dos níveis de cinza que compõe diferentes objetos de uma figura);
- ❖ Imagem Negativa (obtida pelo número simétrico do valor de cada canal de cada pixel no intervalo dado);
- ❖ Conversão para escala de cinza (processo facilitador de conversão de imagem para matriz numérica);
- ❖ Adição e Subtração de imagens (combinação de duas ou mais imagens e também a subtração de

partes de uma imagem, tornando as partes novas imagens);

- ❖ Erosão e dilatação (provoca efeitos de erosão das partes claras da imagem [altos níveis de cinza], gerando imagens mais escuras);
- ❖ Detecção de bordas com operadores Sobel (operador que calcula diferenças finitas, dando uma aproximação do gradiente da intensidade dos pixels da imagem, devolvendo o gradiente ou a norma deste vetor.);
- ❖ Prewitt (definido com um conjunto de 8 máscaras: a máscara que produz a resposta máxima que define a direção do gradiente.);
- ❖ Laplaciano (capaz de enfatizar regiões de descontinuidade e amenizar regiões de variação lenta de níveis de intensidade).

Para cada um dos filtros que utilizavam o conceito de borda de imagens, foram calculados os valores de médias, e também as orientações e distâncias euclidianas. A Fig.10 exhibe a tela de funcionamento da API utilizada e todos seus filtros.

Fig. 10 - Início do processamento da API Java.

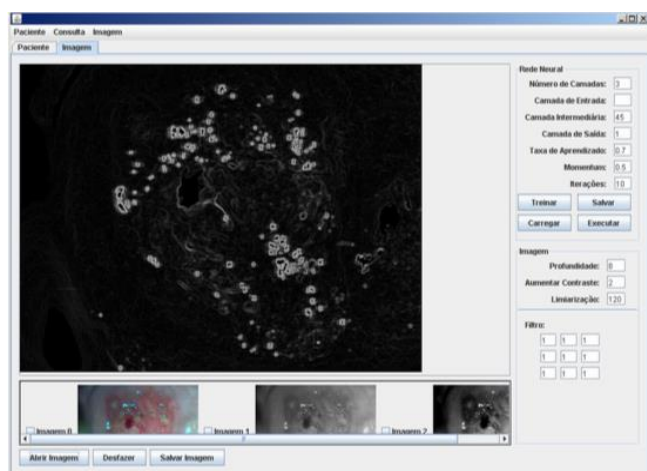


Fonte: Rocha et al. (2007).

Após a aplicação dos filtros, a imagem resultante se torna um emaranhado de pixels com cores que variam em escala de cinza (desde *pixels* completamente pretos, passando por

cinza, até *pixels* brancos). A Fig. 11 ilustra a imagem acima após ser processada pelo algoritmo.

Fig. 11 - Imagem pós processamento.



Fonte: Rocha et al. (2007).

b- Registro Eletrônico

Ainda utilizando a API Java, foi criado o sistema de registro eletrônico do paciente. O sistema nada mais é do que uma interface conectada ao banco de dados da Universidade, que armazena informações sobre o próprio paciente (nome, idade, raça, se é ou não fumante, data e motivo do último exame de colposcopia) e sobre a consulta. Na parte referente à consulta, são questionados dados característicos de anormalidade (epitélio acetobranco, leucoplasia, pontilhado, vasos atípicos, áreas negativas para iodo) e por fim, pede-se o resultado da última biópsia realizada pelo paciente.

2- Resultados

O código para o processamento das imagens deste trabalho não está finalizado. Entretanto, o módulo para RNA's foi concluído, testado e validado pelos autores do projeto. Para dar continuidade ao projeto, serão realizadas implementações no código desenvolvido, a principal é a implementação de um filtro capaz de realizar cálculos relacionados à Transformada de Fourier. Tudo isso se faz necessário para que seja possível realizar a manipulação matemática das matrizes resultantes do processamento das imagens.

VI. RESULTADOS E DISCUSSÕES

Neste projeto, foram selecionados três artigos para estudo e análise de técnicas computacionais inteligentes capazes de auxiliar no diagnóstico do câncer de colo de útero. Observando aspectos técnicos de escrita, os dois primeiros artigos fornecem uma quantidade maior de informações e detalhes de suas pesquisas. O último projeto, por ainda não obter os resultados finais de suas pesquisas, ainda não apresenta total descrição de suas atividades desenvolvidas, entretanto, a RNA do projeto já foi finalizada.

Analisando as técnicas empregadas nos três projetos, suas propostas foram cumpridas. A abrangência de resultados positivos nos dois primeiros artigos foi grande, entretanto as variáveis manipuladas durante o processo de desenvolvimento têm um grande risco de serem implementadas de maneira incorreta. A falha humana ao inserir os dados nos dois processos é capaz de gerar erros catastróficos (indicação de um câncer de colo de útero é a pior das falhas resultantes). Uma alternativa que poderia ser adotada em ambos os casos é a coleta automática de dados clínicos do paciente.

As universidades espanholas citadas possuem seu próprio Hospital Universitário. Assim, o acoplamento dos *softwares criados* junto à obtenção de resultados clínicos, de forma simultânea, poderia auxiliar no processo de prevenção de falhas. Assim que o paciente realizasse a coleta dos materiais clínicos, durante momento de análise e detecção (presença de HPV, traços de sangue, ou células cancerígenas, por exemplo), o sistema inteligente atuar simultaneamente e armazenar os resultados já encontrados em sua base de dado. Isso tornaria o processo mais automatizado, evitando falhas humanas, reduziria o tempo de processamento gasto, e evitaria o possível constrangimento de pacientes ao relatar seus dados ginecológicos.

Em relação ao último projeto analisado, o objetivo de criar uma RNA para reconhecer padrões em imagem e então analisa-los pode ser visto como um excelente método de análise. Segundo Pedrini e Schwartz (2008), uma das formas mais complexas para se analisar uma imagem, está na maneira de como se obter as características necessárias para

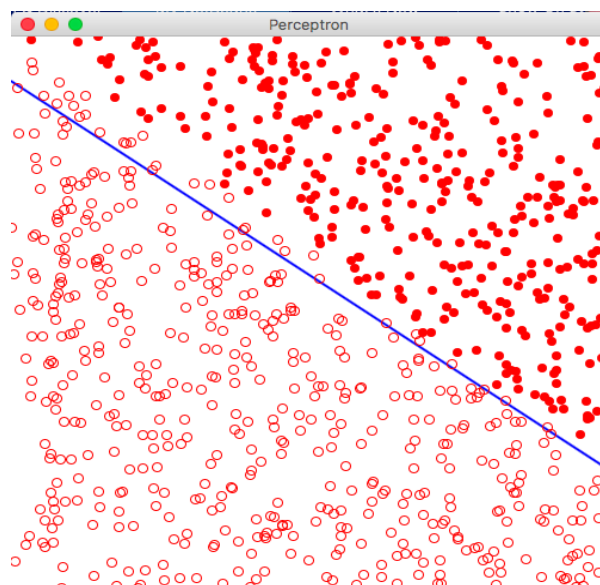
se obter um padrão. Quando comparado às características usadas na interpretação humana, observa-se que a textura é uma das características mais relevantes pelo sistema visual humano. Tal característica consegue descrever informações sobre a distribuição espacial, variação de luminosidade, arranjo estrutural das superfícies e relações entre regiões vizinhas. Assim, pode-se dizer que as informações de textura podem ser consideradas adequadas para análise de regiões em imagens. Entretanto, pode-se ainda propor sugestões para o projeto.

A vinculação do processamento das imagens junto à coleta de informações do paciente se mostra como uma escolha inteligente, mas também é passível de falha humana durante a inserção de dados. Assim como nos dois projetos anteriores, a obtenção de tais informações deve ser feita de maneira automática, durante a realização dos exames clínicos; diferentemente dos projetos anteriormente citados, as informações obtidas nesse projeto não iriam alterar o resultado da análise.

Com os estudos finalizados, como resultado, além da grande quantidade de conhecimento adquirido em relação às técnicas de IA no diagnóstico do câncer de colo de útero, foi desenvolvido um protótipo de uma Rede Neural do tipo *perceptron* unicamada. Analisando os códigos disponibilizados nas pesquisas abordadas, e a lógica abordada foi possível esquematizar um exemplo básico de aplicação da técnica de reconhecimento de padrão. O projeto não é capaz de analisar uma imagem resultante de colposcopia, ou os relatos de pacientes com enfermidades no colo do útero, entretanto é capaz de ilustrar a lógica inicial de um processamento mais complexo.

Seu objetivo é simular um espaço gráfico dividido em duas partes. Cada parte deverá conter seus respectivos itens padrões, devendo não ultrapassar o limite proposto. A Fig. 12 ilustra sua tela final de desenvolvimento. O programa foi desenvolvido em linguagem Java na plataforma IDE Netbeans.

Fig. 12 - Tela final do código Perceptron desenvolvido.



Fonte: Autor (2017).

A base para o desenvolvimento do código é similar à utilizada no terceiro algoritmo analisado: *backpropagation*. Assim como exibido na Fig. 13, o algoritmo desenvolvido baseia-se nos valores de erros encontrados entre a saída desejada e a saída resultante para então voltar ao início de seu processamento (após a entrada de dados) e refazer os cálculos com os valores de pesos sinápticos atualizados.

Fig. 13 - Função *backpropagation* do código desenvolvido.

```
void pFunction(double[] inputs, int desired) {
    int guess = feedForward(inputs);
    double error = desired - guess; //erro: valor desejado - valor obtido
    for (int i = 0; i < nValues.length; i++) {
        nValues[i] += c * error * inputs[i]; //corrigir o valor do peso,
    }
}
```

Fonte: Autor (2017).

O código completo desenvolvido encontra-se nos Apêndices, ao final deste estudo.

VII. CONSIDERAÇÕES FINAIS

Com a realização desta pesquisa realizada foi possível analisar de maneira mais profunda a aplicação de técnicas computacionais de alto nível na medicina, uma área com conceitos bem diferentes da computação, em sua totalidade. As tecnologias estudadas são bastante atuais (o último projeto analisado data do ano de 2007) e, como visto

nas pesquisas, seus resultados satisfatórios tendem a ampliar a combinação dessas duas áreas. A taxa de 13% de erro obtida pelo projeto da realizado em Valladolid foi considerada pelo autor como pequena, mostrando-se então eficiente em sua utilização. Já no segundo artigo estudado, a taxa de aproveitamento foi de 97% dos casos analisados, considerada bastante alta pelo autor.

A pesquisa também mostra o quão complexo é o desenvolvimento de técnicas de IA. Considerando a quantidade de variáveis dependentes no processo, a sua forma de processamento, que precisa ser exata, e a enorme responsabilidade daqueles que a aplicam. Quando se utiliza tais sistemas em aplicações desse tipo (diagnóstico de câncer de colo de útero), a responsabilidade tende a aumentar. Qualquer resultado obtido de maneira equivocada pode gerar problemas irreparáveis para o paciente em questão. O diagnóstico de um falso-positivo, é capaz de desestabilizar um paciente e, o diagnóstico de um falso-negativo pode atrasar um tratamento, trazendo-lhe graves complicações na saúde.

O código desenvolvido ao final, foi de grande valia para ampliar o conhecimento prático sobre as tecnologias aplicadas. É possível ampliar a área de aplicação do sistema desenvolvido, futuramente, para que o mesmo seja capaz de reconhecer padrões diferentes do previamente imposto. Sendo capaz de realizar essa atividade, o projeto pode ser aplicado em diversas outras áreas, não somente na computação ou medicina.

VIII. REFERÊNCIAS

- APARICIO, O. V. **Red neuronal artificial como modelo predictivo en una unidad de patologia cervical**. Valladolid, Espanha. 2015. Disponível em: < <https://goo.gl/VMViFn> >. Acesso em: abr. 2017.
- ASKABIDE. **Información sobre la Colposcopia**. Guipúzcoa, Espanha. 2014. Disponível em: < <https://goo.gl/vJQ5T2> >. Acesso em: jun. 2017.
- CARVALHO, A. P. L. F. **Redes Neurais Artificiais**. ICMC, USP. 2009. Disponível em: < <https://goo.gl/VmdxQb> >. Acesso em: abr. 2017.
- GASPERIN, S. I.; BOING, A. F.; KUPEK, E. **Cobertura e fatores associados à realização do exame de detecção do câncer de colo de útero em área urbana no Sul do Brasil: estudo de base populacional**. Florianópolis, Brasil. 2011. Disponível em: < <https://goo.gl/eCKghv> >. Acesso em: abr. 2017.
- KARIMI Z. M.; MOUSAVI A.; DEGHANI A. **Conservative surgery in cervical cancer: report of two radical abdominal trachelectomies and literature review**. Eur J Gynaecol Oncol. 2011. Disponível em: < <https://goo.gl/oMoaE4> >. Acesso em: abr. 2017.
- MAGMANI, D. Jr. **Câncer de Bexiga**. 2015. Disponível em: < <https://goo.gl/tTxzSK> >. Acesso em: abr. 2017.
- MITCHELL, T. **Machine Learning**. McGraw Hill. ISBN 0-07-042807-7, p.2. 1997.
- OMS BRASIL. **OMS: 85% das mortes por câncer de colo de útero ocorrem em países de média e baixa renda**. 2017. Disponível em: < <https://goo.gl/FvQORt> >. Acesso em: abr. 2017.
- PEDRINI, H.; SCHWARTZ, W.R. **Análise de imagens digitais: Princípios, algoritmos e Aplicações**. Thomson Learning. São Paulo, Brasil. 2008.
- ROCHA, E.L. et al. **Diagnóstico automatizado de doenças no colo do útero baseado em redes neurais artificiais e processamento de imagens digitais**. Santa Catarina, Brasil. 2007. Disponível em: < <https://goo.gl/pyqpQH> >. Acesso em: abr. 2017.
- SANCHEZ, L. Y. B. **Desarrollo de un sistema experto sobre web para un diagnostico temprano de câncer de**

cuello uterino en la Clinica Maternidad “Belen” – Chiclayo. Chiclayo, Peru. 2012. Disponível em: < <https://goo.gl/qAwaa8>>. Acesso em: abr. 2017.

TISSOT, H. C.; CAMARGO, L. C.; POZO, A. T. R. **Treinamento de Redes Neurais Feedforward:** comparativo dos algoritmos Backpropagation e Differential Evolution. Curitiba, Brasil. 2012. Disponível em: < <https://goo.gl/faSeuu> >. Acesso em: abr. 2017.

IX. APÊNDICE

```

package perceptronImage;

import java.awt.BasicStroke;
import java.awt.BorderLayout;
import java.awt.Color;
import javax.swing.Timer;
import java.awt.Dimension;
import java.awt.Graphics;
import java.awt.Graphics2D;
import java.awt.RenderingHints;
import java.awt.event.ActionEvent;
import java.util.Random;
import javax.swing.JFrame;
import javax.swing.JPanel;
import javax.swing.SwingUtilities;
//import javax.swing.*;

public class Perceptron extends JPanel {

    class Training {
        double[] nIn;
        int result;

        Training(double x, double y, int z) {
            nIn = new double[]{x, y, 1};
            result = z;
        }
    }

    Training[] training = new Training[2000];
    double[] nValues;
    double c = 0.00001;
    int count;

    public Perceptron(int n) {
        Random r = new Random(); //pesos aleatórios num
        primeiro momento
        Dimension dim = new Dimension(500, 500);

```

```

        setPreferredSize(dim);
        setBackground(Color.white);

        nValues = new double[n];
        for (int i = 0; i < nValues.length; i++) {
            nValues[i] = r.nextDouble() * 2 - 1;
        }

        for (int i = 0; i < training.length; i++) {
            double x = r.nextDouble() * dim.width;
            double y = r.nextDouble() * dim.height;

            int answer = 0;

            if (y < f(x)){
                answer = -1;
            }
            else{
                answer = 1;
            } //separar os lados das respostas

            training[i] = new Training(x, y, answer);
        }

        new Timer(5, (ActionEvent e) -> {
            repaint();
        }).start();
    }

    private double f(double x) {
        return x * 0.7 + 40;
    }

    int feedForward(double[] inputs) {

        double plus = 0;
        for (int i = 0; i < nValues.length; i++) {
            plus += inputs[i] * nValues[i];
        }

```

```

    return activate(plus);
}

int activate(double s) {
    return s > 0 ? 1 : -1;
}

void pFunction(double[] inputs, int desired) {
    int guess = feedForward(inputs);
    double error = desired - guess; //erro: valor desejado -
valor obtido
    for (int i = 0; i < nValues.length; i++) {
        nValues[i] += c * error * inputs[i]; //corrigir o valor
do peso de acordo com o valor do erro
    }
}

@Override
public void paintComponent(Graphics gg) {
    super.paintComponent(gg);
    Graphics2D g = (Graphics2D) gg;

g.setRenderingHint(RenderingHints.KEY_ANTIALIASING,
    RenderingHints.VALUE_ANTIALIAS_ON);
//resultados gráficos, visíveis, geralmente padrão em
aplicações simples

    int width = getWidth();
    int intWidth = (int) f(width);
    g.setStroke(new BasicStroke(2));
    g.setColor(Color.blue);
    g.drawLine(0, (int) f(0), width, intWidth);

    pFunction(training[count].nIn, training[count].result);
    count = (count + 1) % training.length;

    g.setStroke(new BasicStroke(1));
    g.setColor(Color.red);
    for (int i = 0; i < count; i++) {
        int guess = feedForward(training[i].nIn);
        width = (int) training[i].nIn[0] - 4;
        intWidth = (int) training[i].nIn[1] - 4;

        if (guess > 0)
            g.drawOval(width, intWidth, 8, 8);
        else
            g.fillOval(width, intWidth, 8, 8);
    }
}

public static void main(String[] args) {
    SwingUtilities.invokeLater(() -> {
        JFrame f = new JFrame();

        f.setDefaultCloseOperation(JFrame.EXIT_ON_CLOSE);
        f.setTitle("Recongnize Perceptron");
        f.add(new Perceptron(3), BorderLayout.CENTER);
        //perceptron 3 - x y e a linha
        f.pack();
        f.setLocationRelativeTo(null);
        f.setVisible(true);
    });
}
}

```