

Diagnóstico de Doenças Dermatológicas usando a Rede Neural de Kohonen

Andréa Boçois¹, André Anastácio de Oliveira¹, Paulo Henrique Siqueira¹, Flávio de Queiros Telles²

¹Programa de Pós-Graduação em Métodos Numéricos em Engenharia – Universidade Federal do Paraná (UFPR)

²Hospital de Clínicas – Universidade Federal do Paraná (UFPR)
Caixa Postal 19.011 – 81.531-980 – Curitiba – PR – Brazil

dea_mat@yahoo.com.br, {anastácio, paulohs}@ufpr.br,
queiroz.telles@uol.com.br

Abstract. *This paper proposes the diagnosis of infectious diseases of skin using the Kohonen Neural Networks. For identification of clusters the U-matrix was used as visualization method, defining the boundaries between clusters and the system RGB primary color, defining the neuron color arranged in a structure map. The network was implemented in Visual Basic language and the input data were arranged through texts and medical records. The model showed a good performance, with 94.5% accuracy on the tests set, demonstrating the possibility of software developing to aid physicians in making decision at diagnosis.*

Resumo. *Este trabalho propõe o diagnóstico de doenças infecciosas da pele utilizando-se das Redes Neurais Artificiais de Kohonen. Para a identificação dos agrupamentos foi utilizado o método de visualização da U-matriz, definindo as fronteiras entre os agrupamentos, e o sistema de cores primárias RGB, definindo a cor do neurônio disposto em uma estrutura de mapa. A rede foi implementada na linguagem Visual Basic e os dados de entrada foram arranjados através de textos e de prontuários médicos. Verificou-se que o modelo apresentou um bom desempenho, com 94,5% de acerto no conjunto de testes evidenciando a possibilidade de desenvolver um software para auxiliar o médico na tomada de decisões no momento do diagnóstico.*

1. Introdução

A utilização de computadores como ferramenta de auxílio à tomada de decisões encontra na medicina uma das principais áreas de sua aplicação. Inúmeras atividades relacionadas à informática podem ser encontradas na área médica, citando-se os sistemas de apoio ao diagnóstico como uma das mais marcantes. Esses sistemas tentam imitar o raciocínio do médico através de princípios de análises de decisões, com a capacidade de elaborar um raciocínio automático e realizar inferências lógicas, como induções e deduções.

Podemos definir o processo diagnóstico tal como Massad [2003]: um conjunto de técnicas que rotula o paciente e classifica sua doença, identifica o prognóstico provável e define o melhor tratamento disponível. É um processo de comparação e de classificação, que exige do médico um esforço para reconhecer a classe (doença) à qual determinado paciente pertence, confrontando dados reunidos através da anamnese, do exame físico e dos exames complementares, com as informações disponíveis a respeito das diversas doenças existentes.

Recentemente, a inteligência artificial vem sendo utilizada como uma ferramenta de auxílio ao diagnóstico. A Rede Neural Artificial (RNA) [HAYKIN, 2001; TAFNER *et al*, 1996] é uma das técnicas da inteligência artificial e consiste na simulação de alguns aspectos do sistema nervoso biológico, na tentativa de permitir a realização de tarefas cognitivas de forma mais eficaz do que as obtidas através do processamento serial.

Vários estudos tratam o problema do diagnóstico médico através do uso de métodos de agrupamento e classificação [DIAS, 2003; SALA, 2005; SOARES, 2006; SOUZA, 2004; VASCONCELOS, 2000; WYNS *et al*, 2004]. Visando observar o comportamento de Redes Neurais Artificiais para esta tarefa, foi escolhida a abordagem pela Rede Neural de Kohonen [KOHONEN, 1990; KOHONEN, 1995; XU *et al*, 2005; SIQUEIRA, 2005], cuja especialidade é o agrupamento de dados. Este artigo propõe avaliar os sintomas dos pacientes, detectados por meio de exames clínicos, a fim de sugerir possíveis diagnósticos.

Por motivos de simplicidade na implementação desta tarefa, restringiu-se a um subconjunto de diagnósticos da área de conhecimento da medicina denominada infectologia, em especial aquelas que acometem a pele: Esporotricose, Cromoblastomicose, Impetigo, Herpes Labial, Larva Migrans Cutânea e Leishmaniose Tegumentar Americana. O software utilizado foi programado na linguagem computacional Visual Basic, e poderá servir como uma ferramenta auxiliar para o médico, embora não tenha a intenção de ser um sistema decisivo no processo de tomada de decisão. O objetivo desta ferramenta é a redução de erros subjetivos de decisões humanas equivocadas nos diagnósticos destas doenças.

O restante do artigo é organizado da seguinte forma: a seção 2 descreve a metodologia utilizada; os resultados são apresentados na seção 3; e na seção 4 são apresentadas as discussões e conclusões deste trabalho.

2. Metodologia

2.1 Obtenção dos dados

Os dados de entrada da rede neural deste trabalho foram arranjados após leitura de textos e análise de prontuários do Hospital de Clínicas (HC) da Universidade Federal do Paraná (UFPR) do período de 1995 a 2011.

Foram analisados 60 prontuários do HC-UFPR. De cada diagnóstico foram extraídos os atributos que compõem o vetor de entradas da rede neural, juntamente com informações consultadas de textos médicos e de conversas com um dos infectologistas do HC. Não foram encontrados prontuários da doença Larva Migrans Cutânea, por isso seus atributos foram extraídos de textos médicos e de experiências relatadas por profissionais da área encontrados em textos durante a pesquisa.

Ao todo, existem 68 padrões de entrada, sendo 15 de Impetigo, 12 de Cromoblastomicose, 10 de Esporotricose, 5 de Larva Migrans Cutânea, 10 de Herpes Labial e 16 de Leishmaniose Tegumentar Americana.

2.2 Tratamento dos dados

O banco de dados é composto de 83 variáveis que podem ser organizadas basicamente do seguinte modo:

- a) *características da lesão* – purulenta; com secreção transparente; secreção purulenta; secreção fétida; há secreção somente se pressionar a lesão; pápulas; pústulas; vesículas; bolhas; nodular; verrugosa; crostas; com descamação; ulcerada, máculas; forma de uma placa; eritematosa; com edema; esbranquiçada; granulosa; amarronzada; com sangramento; bordas elevadas; bordas elevadas bem definidas; ínguas próximas a lesão; serpentiforme (linhas sinuosas); presença de pontos negros na lesão; semelhante à couve-flor; forma anelar; formou um cordão endurecido de nódulos (semelhante a um rosário);
- b) *sintomas* – febre; prurido; dor na lesão; mal estar; anorexia;
- c) *local da lesão* – tronco; membros superiores ou inferiores; lábio; perioral; língua; mucosa; face; ombro; nádegas;
- d) *início da lesão* – com pápula, mácula, eritema, vesícula, ulcera, edema, nódulo; forma de placa; início há 15 dias, entre 15 e 30 dias, entre 1 e 6 meses, entre 6 meses e um ano, mais de um ano;
- e) *progressão da lesão* – aumentou o número de lesões; aumentou a extensão da lesão; aumentou a profundidade da lesão; para pústulas, vesículas, crostas, verrugas, ulcera, secreção; ruptura espontânea (das vesículas ou bolhas); disseminação por contigüidade; para múltiplas lesões pelo corpo;
- f) *dados epidemiológicos* – teve contato com pessoa doente; viajou, morou ou mora em região com dados epidemiológicos da doença; havia trauma no local da lesão ocasionado por animal, por ferramenta agrícola, por fragmento vegetal, ou outro tipo de trauma; houve inserção em áreas florestais; teve contato com solo ou plantas; teve contato com local freqüentado por gatos; exposição solar; estresse emocional; está menstruada; imunidade baixa (por alguma doença ou imunodeficiência).

Os dados são codificados em coordenadas binárias, atribuindo o valor “0” quando o paciente não possui a característica e o valor “1” caso contrário. Tem-se um vetor de entradas com 83 coordenadas binárias, que correspondem aos seis atributos de cada paciente. A matriz de entrada de dados é de ordem 68 x 83, já que foram considerados os dados referentes a 68 pacientes.

A resposta (saída da Rede Neural) é o diagnóstico do paciente. Foram consideradas seis saídas para a RNA: Leishmaniose Cutânea, Impetigo, Cromoblastomicose, Esporotricose, Larva Migrans Cutânea e Herpes Labial, cada uma delas correspondendo a um dos diagnósticos das infecções. A resposta é determinada pelo neurônio vencedor, verificando a qual agrupamento do mapa de saída ele pertence.

2.3 Implementação da Rede Neural de Kohonen

A Rede Neural de Kohonen foi programada na linguagem computacional Visual Basic. A taxa de aprendizagem, o raio de vizinhança e a dimensão do mapa de saída devem ser definidos inicialmente. A cada iteração a taxa de aprendizagem e o raio de vizinhança decaem exponencialmente. Em todos os testes desse trabalho, o raio de vizinhança começou com valor igual ao diâmetro da rede.

Os 68 padrões que compõe a amostra foram divididos aleatoriamente em dois conjuntos: treinamento e teste, conforme tabela abaixo.

Tabela 1 – Número de padrões para treinamento e teste

	Impetigo	Cromoblastomicose	Esporotricose	Larva Migrans Cutânea	Herpes Labial	Leishmaniose Tegumentar Americana
Treinamento	10	8	7	4	7	11
Teste	5	4	3	1	3	5

Após o treinamento da rede, os neurônios são agrupados segundo as doenças pesquisadas e o agrupamento identificado por uma cor. Para determinar a fronteira entre os agrupamentos foi usada a técnica da U-matriz [ULTSCH, 2003] e para determinar a cor de cada neurônio foi usado o sistema de cores RGB (Red-Green-Blue: Vermelho-Verde-Azul), detalhado na seção seguinte, onde as cores vermelha, verde e azul são misturadas, obtendo-se tonalidades diferentes.

Os erros, topológico, de quantização e médio quadrático, foram determinados calculando-se a média dos valores encontrados dos respectivos erros após executar cinco vezes cada mapa. Para cada uma das cinco vezes que um mapa foi executado, usavam-se as mesmas configurações definidas ao programa (taxa de aprendizagem inicial, raio de vizinhança, número de iterações e a dimensão do mapa) e variavam-se aleatoriamente os valores das coordenadas dos pesos de cada neurônio.

2.3.1 Formação do mapa de saída da Rede Neural de Kohonen após treinamento

Para utilizar as cores RGB precisamos de três coordenadas, e como cada neurônio possui 83 coordenadas (variáveis), foi necessária a criação de uma combinação dessas para que no final ficássemos com apenas três. Desta forma, o número total de coordenadas do neurônio foi dividido por três, obtendo-se o valor de NVG (numero de variáveis por grupo) e as coordenadas foram separadas em três grupos: para o primeiro grupo as coordenadas do neurônio da primeira posição até a NVG-ésima posição, para o segundo grupo as coordenadas da NVG-ésima posição até a 2*NVG-ésima e as demais coordenadas ficaram no terceiro grupo.

Neurônio = $(a_1, a_2, \dots, a_{83})$ \implies 83 variáveis

NVG = $83/3 = 27,66$ \implies 27 variáveis para o 1º grupo

\implies 27 variáveis para o 2º grupo

\implies 29 variáveis para o 3º grupo

Neurônio = $(\underbrace{a_1, \dots, a_{27}}_{1^\circ \text{ grupo}}, \underbrace{a_{28}, \dots, a_{54}}_{2^\circ \text{ grupo}}, \underbrace{a_{55}, \dots, a_{83}}_{3^\circ \text{ grupo}})$

Para cada grupo encontrado, foi calculada a média de todas as variáveis estipuladas e essas médias passaram a ser as novas coordenadas dos neurônios.

$$\text{Média 1} = \text{Média 1}^\circ \text{ grupo} = (a_1 + \dots + a_{27}) / 27$$

$$\text{Média 2} = \text{Média 2}^\circ \text{ grupo} = (a_{28} + \dots + a_{54}) / 27$$

$$\text{Média 3} = \text{Média 3}^\circ \text{ grupo} = (a_{55} + \dots + a_{83}) / 29$$

Neurônio = (média 1, média 2, média 3) \Longrightarrow novas coordenadas

Assim, todos os neurônios ficam com apenas três coordenadas, número necessário para usar as cores RGB, onde a média do primeiro grupo de coordenadas (Média 1) determina a intensidade de vermelho, a segunda média (Média 2) determina a intensidade de verde e a média do último grupo (Média 3), a intensidade de azul. Mas isso não poderia ser feito de forma direta, pois o intervalo das coordenadas dos neurônios e das cores RGB eram distintos, a intensidade das cores RGB variam de 0 a 255, e as coordenadas dos neurônios variavam entre 0 e 1 (nem sempre atingindo o valor máximo um). Para que uma combinação fosse possível, optou-se trabalhar com a porcentagem.

Fazendo uma comparação entre a média 1 de todos os neurônios, ao maior valor encontrado foi estipulado 100% e aos demais valores, abaixo deste, foi determinada uma porcentagem correspondente, baseada naquela que recebeu 100%, usando uma regra de três simples. O mesmo foi feito para as médias 2 e 3. Assim, tinha-se uma porcentagem para a intensidade de vermelho, outra para a intensidade de verde e outra para a intensidade de azul, referentes as médias 1, 2 e 3, respectivamente.

Finalmente, cada neurônio recebe uma cor, baseada nos valores de suas coordenadas. Neurônios próximos que possuem coordenadas similares recebem cores também similares, formando-se assim agrupamentos que são destacados pela distinção das cores.

Após a organização em cores, os neurônios são rotulados segundo o padrão de entrada mais semelhante a ele. Os neurônios considerados vencedores recebem como rótulo o padrão de entrada que possui a menor distância Euclidiana com o neurônio.

3. Resultados

3.1 Testes

As dimensões dos mapas testados variam de 3x3 a 20x20, com vizinhança Gaussiana do tipo quadrada, e com a utilização da atualização exponencial para as taxas de aprendizagem e do raio da vizinhança dos mapas.

Para investigações sobre o número de neurônios, foram utilizados 1000, 3000 e 5000 iterações, cada mapa foi executado cinco vezes no programa tomando valores aleatórios para os pesos dos neurônios do mapa, taxa inicial de aprendizado variando entre 0,5 ou 1 e o conjunto de treinamento definitivo, com a entrada dos dados na fase de treinamento de forma aleatória.

Foram realizados 540 testes e verificou-se que os resultados para os erros de quantização com 1000, 3000 ou 5000 iterações ficavam similares e praticamente tornavam-se estáveis para mapas de dimensão maior do que 9x9. O mesmo ocorre com o erro médio quadrático, (Tabelas 2 e 3). Os mapas também foram testados variando a

taxa de aprendizado inicial entre 0,5 ou 1, mas não mostraram diferença significativa e por isso os resultados com taxa de aprendizagem inicial igual a 0,5 serão omitidos. A taxa de aprendizado inicial igual a um foi escolhida por apresentar melhores resultados para o erro topológico.

A escolha do melhor mapa para o problema em estudo deve levar em conta os três erros encontrados. Por isso, neste trabalho optou-se trabalhar com o mapa 10x10, pois apresenta valores satisfatórios para todos os erros avaliados na fase de treinamento (considerando 5000 iterações) e é o que apresenta os menores valores para os três erros analisados, durante a fase de teste.

Tabela 2 – Erros na fase de treinamento

	1000 iterações			3000 iterações			5000 iterações		
	Erro de quantização	Erro médio quadrático	Erro topológico	Erro de quantização	Erro médio quadrático	Erro topológico	Erro de quantização	Erro médio quadrático	Erro topológico
3x3	2,46	6,42	0,30	2,444	6,502	0,36	2,443	6,4658	0,39
4x4	2,036	4,838	0,34	2,058	4,753	0,31	2,040	4,8124	0,33
5x5	1,54	3,111	0,32	1,486	3,142	0,38	1,526	3,1104	0,33
6x6	0,81	1,566	0,27	0,792	1,456	0,26	0,808	1,5546	0,26
7x7	0,30	0,517	0,23	0,279	0,468	0,26	0,298	0,4984	0,21
8x8	0,12	0,065	0,14	0,065	0,117	0,10	0,105	0,174	0,10
9x9	0,008	0,009	0,04	0,018	0,021	0,04	0,026	0,02778	0,06
10x10	0,014	0,027	0,02	2,806E-07	1,01E-13	0,02	3,1E-07	0,0244	0,02
11x11	2,07E-07	4,35E-14	0,02	2,069E-07	4,35E-14	0,02	2,079E-07	4,39E-14	0,01
12x12	2,058E-07	4,30E-14	0,01	2,064E-07	4,33E-14	0,02	2,067E-07	4,34E-14	0,02
13x13	2,034E-07	4,21E-14	0,02	2,040E-07	4,23E-14	0,02	2,058E-07	4,31E-14	0,004
14x14	2,015E-07	4,23E-14	0,01	2,023E-07	4,16E-14	0,004	2,041E-07	4,23E-14	0,004
15x15	2,031E-07	4,19E-14	0,01	2,053E-07	4,29E-14	0,008	2,064E-07	4,33E-14	0,02
16x16	2,029E-07	4,19E-14	0,008	2,061E-07	4,32E-14	0	2,067E-07	4,22E-14	0,02
17x17	2,026E-07	4,18E-14	0,004	2,033E-07	4,22E-14	0,004	2,046E-07	4,26E-14	0,03
18x18	2,009E-07	4,12E-14	0	2,040E-07	4,23E-14	0,004	2,011E-07	4,16E-14	0,02
19x19	2,006E-07	4,10E-14	0,012	2,016E-07	4,14E-14	0	2,027E-07	4,19E-14	0,01
20x20	2,007E-07	4,10E-14	0,004	2,02E-07	4,15E-14	0,004	2,027E-07	4,15E-14	0,05

Como nos testes acima mencionados os valores dos erros diminuíram com o aumento do número de iterações foram feitos novos testes para o mapa 10x10 aumentando o número de iterações para 6000, 7000 e 8000, considerando taxa inicial de aprendizado igual a 1 e raio da vizinhança igual ao diâmetro da rede. Durante o treinamento com o mapa 10x10, com 5000 iterações, observou-se que os menores valores para os erros na fase de teste eram conseguidos com a quinta variação dos pesos, por isso os novos testes foram realizados usando esses mesmos pesos.

Tabela 3 – Erros na fase de teste, onde EQ é o erro de quantização, EMQ é o erro médio quadrático e ET o erro topológico

	1000 iterações			3000 iterações			5000 iterações		
	EQ	EMQ	ET	EQ	EMQ	ET	EQ	EMQ	ET
3x3	2,765	7,832	0,31	2,755	7,766	0,35	2,795	7,979	0,49
4x4	2,690	7,493	0,38	2,717	7,554	0,28	2,6984	7,468	0,32
5x5	2,661	7,325	0,47	2,722	7,624	0,39	2,6832	7,398	0,36
6x6	2,680	7,452	0,34	2,668	7,440	0,32	2,705	7,593	0,34
7x7	2,650	7,342	0,40	2,626	7,244	0,42	2,5848	7,015	0,37
8x8	2,551	6,807	0,37	2,533	6,718	0,32	2,5732	6,953	0,33
9x9	2,510	6,611	0,27	2,499	6,547	0,27	2,5068	6,595	0,31
10x10	2,481	6,647	0,25	2,495	6,518	0,25	2,455	6,309	0,27
11x11	2,476	6,425	0,33	2,476	6,443	0,31	2,4814	6,464	0,32
12x12	2,479	6,458	0,30	2,466	6,396	0,40	2,4628	6,380	0,42
13x13	2,515	6,668	0,37	2,476	6,457	0,44	2,4728	6,435	0,47
14x14	2,503	6,577	0,35	2,473	6,437	0,47	2,4638	6,397	0,43
15x15	2,519	6,673	0,49	2,486	6,518	0,48	2,4878	6,527	0,58
16x16	2,512	6,638	0,37	2,474	6,458	0,47	2,4868	6,499	0,52
17x17	2,559	6,884	0,37	2,491	6,503	0,49	2,4638	6,361	0,47
18x18	2,561	6,917	0,43	2,516	6,672	0,49	2,477	6,452	0,48
19x19	2,58	6,998	0,41	2,493	6,538	0,46	2,4534	6,320	0,54
20x20	2,598	7,087	0,36	2,516	6,647	0,35	2,4512	6,315	0,49

3.2 Resultados

Os erros encontrados após os novos testes realizados com o mapa 10x10 podem ser visualizados na Tabela 4. O menor valor para o erro topológico na fase de teste ocorre com 6000 iterações. Os demais erros são considerados aceitáveis para este número de iterações, já que estão entre os menores valores tanto para a fase de treinamento quanto para teste. Assim decidiu-se trabalhar com o mapa obtido após 6000 iterações.

Tabela 4 – Erros do mapa 10x10

Iterações	Fase de treinamento			Fase de teste		
	6000	7000	8000	6000	7000	8000
Erro de quantização	2,08E-07	2,08E-07	2,08E-07	2,511	2,509	2,505
Erro médio quadrático	4,39E-14	4,42E-14	4,41E-14	6,582	6,599	6,583
Erro topológico	0,021	0,021	0,063	0,14	0,47	0,28

Após o período de treinamento da rede, de acordo com o processo descrito na seção 2.3.1, cada neurônio recebe uma cor, do sistema de cores RGB, baseada nos valores de suas coordenadas. Aqueles de cores semelhantes possuem coordenadas similares, sendo possível formar agrupamentos destacados pela distinção das cores. Em seguida é feita a rotulagem dos neurônios, onde o neurônio vencedor é classificado segundo o padrão mais semelhante a ele, gerando o mapa da Figura 1.

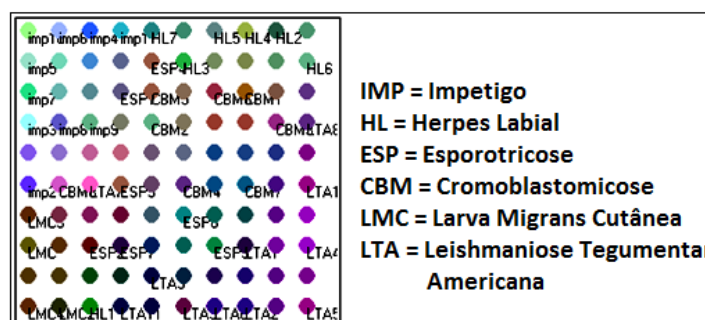


Figura 1 – Rotulagem dos neurônios após 6000 iterações

Observa-se no mapa da Figura 1 que a doença Impetigo recebeu basicamente a cor azul, a Leishmaniose Tegumentar Americana a cor roxa, variando entre tons claros e escuros, e a Larva Migrans Cutânea a cor marrom. Já o Herpes Labial mesclou tons de verde, e a Cromoblastomicose e a Esporotricose não separaram muito bem as suas cores, mesclando várias delas, entre elas o roxo, o marrom e o azul mais escuro.

Para realizar testes e verificar a eficiência da rede treinada, é necessário fazer um agrupamento dos neurônios da Figura 1 segundo a doença que representam. Para definir as fronteiras entre os grupos, usamos a U-matriz, Figura 2(A). Suas cores variam do preto ao cinza claro, de acordo com uma escala de distâncias, onde a cor escura representa as maiores distâncias entre os neurônios e indicam as fronteiras entre os agrupamentos.

Dado um mapa de tamanho $A \times B$, a U-matriz é representada pela imagem desse mapa em outro mapa de dimensão $(2A - 1) \times (2B - 1)$. Neste são representados os neurônios do mapa, além das distâncias intermediárias entre os neurônios do mapa original através de cores, de acordo com os coeficientes do vetor peso de cada neurônio. Essa técnica é utilizada na visualização do mapa para posterior separação manual dos agrupamentos, a idéia básica é utilizar a mesma métrica que foi usada durante o treinamento para calcular distâncias entre pesos sinápticos de neurônios adjacentes do mapa.

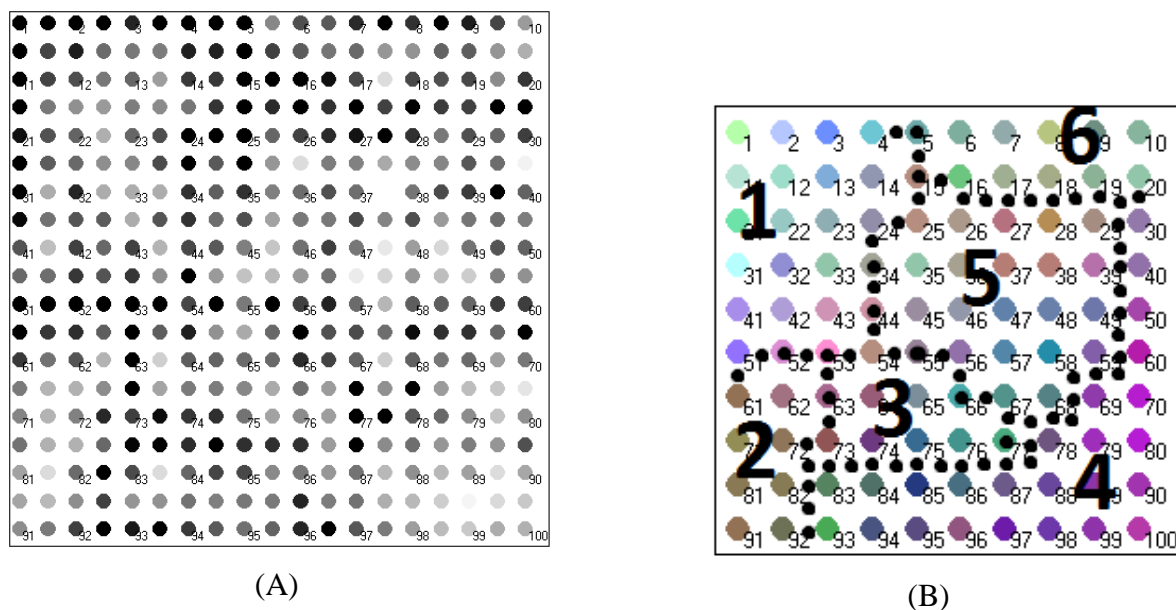


Figura 2 – (A) U-matriz, (B) separação manual dos agrupamentos

As fronteiras mostradas na figura 2(B) são determinadas após análise da figura 2(A), onde os neurônios mais escuros desta indicam as fronteiras que serão formadas na figura 1, resultando nos agrupamentos mostrados na figura 2(B). O grupo 1 refere-se à doença Impetigo, o grupo 2 a Larva Migrans Cutânea, o grupo 3 a Esporotricose, o grupo 4 a Leishmaniose Tegumentar Americana, o grupo 5 a Cromoblastomicose e o grupo 6 ao Herpes Labial.

Na figura 3 (A e B) são mostrados os neurônios vencedores após a fase de testes da rede usando os padrões reservados para essa finalidade. As fronteiras entre os agrupamentos

são constatadas pelos círculos pretos e os neurônios vencedores para os padrões de teste são identificados pelos neurônios amarelos.

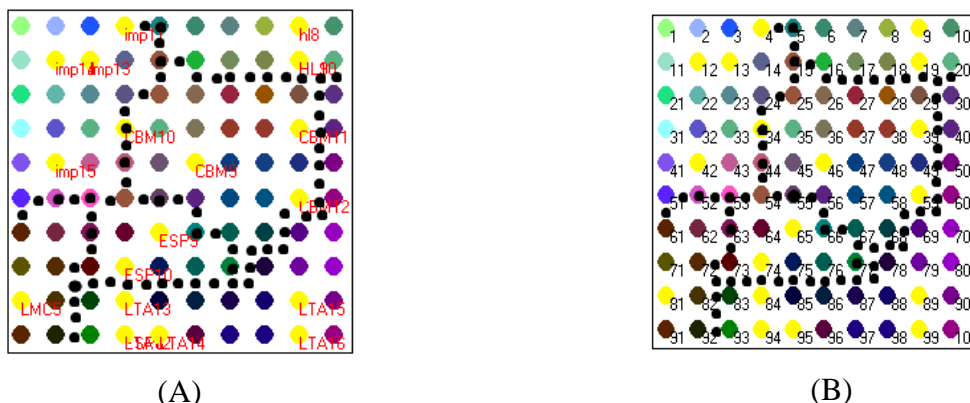


Figura 3 – Mapa de teste com (A) neurônios rotulados e (B) com neurônios enumerados

Na Tabela 5 temos o número do neurônio vencedor para cada padrão de teste.

Tabela 5 – Neurônio vencedor para cada padrão de teste

PADRÃO DE TESTE	Nº DO NEURÔNIO VENCEDOR	PADRÃO DE TESTE	Nº DO NEURÔNIO VENCEDOR
IMP11	4	ESP10	74
IMP12	12	LMC	81
IMP13	13	HL8	9
IMP14	12	HL9	19
IMP15	42	HL10	19
CBM9	46	LTA12	94
CBM10	34	LTA13	84
CBM11	39	LTA14	95
CBM12	59	LTA15	89
ESP8	94	LTA16	99
ESP9	65		

Fazendo análise dos resultados da Tabela 5 com os agrupamentos formados (Figura 2(B)) pode-se verificar que os padrões de teste de Impetigo, Herpes Labial, Larva Migrans Cutânea, Cromoblastomicose e Leishmaniose Tegumentar Americana foram classificadas corretamente, obtendo 100% de acerto, conforme mostra a matriz de confusão da Tabela 6, enquanto Esporotricose teve 66,7% de acerto, devido a sua semelhança com a Leishmaniose Tegumentar Americana e com a Cromoblastomicose, classificando-a erroneamente como uma dessas doenças. No geral, fazendo uma média entre as porcentagens de acerto de cada doença, a rede obteve 94,5% de acerto no conjunto de testes.

Tabela 6 – Matriz de confusão dos dados de teste

	Impetigo	Cromoblastomicose	Esporotricose	Herpes Labial	Larva Migrans Cutânea	Leishmaniose Tegumentar Americana
Impetigo	5	0	0	0	0	0
Cromoblastomicose	0	4	0	0	0	0
Esporotricose	0	0	2	0	0	1
Herpes Labial	0	0	0	3	0	0
Larva Migrans Cutânea	0	0	0	0	1	0
Leishmaniose Tegumentar Americana	0	0	0	0	0	5

4. Discussão e conclusões

No presente trabalho, as Redes Neurais foram exploradas para resolver o problema de diagnóstico de doenças infecciosas da pele, em particular Impetigo, Cromoblastomicose, Esporotricose, Leishmaniose Cutânea, Herpes Labial e Larva Migrans Cutânea.

Utilizou-se o algoritmo da Rede de Kohonen, empregado para o reconhecimento e classificação de padrões, baseado na competição dos neurônios pela representação da informação apresentada. Problemas encontrados na prática da clínica médica envolvem, geralmente, associação e reconhecimento de padrões, permitindo à Rede Neural de Kohonen encontrar diversas aplicações na área médica.

Verificou-se que o modelo apresentou um bom desempenho, onde os percentuais de acerto para os padrões de teste foram de 66,7 % para Esporotricose e 100% para as demais. A baixa taxa de acerto para Esporotricose deve-se ao fato de ter muitas características comuns com as doenças Cromoblastomicose e Leishmaniose Tegumentar Americana, classificando-a erroneamente como uma dessas doenças. Segundo infectologista do HC, a doença só é diagnosticada após exames laboratoriais.

Quanto à separação dos agrupamentos em cores, verificou-se que Impetigo ficou com tons de azul, Herpes Labial com tons de verde, Leishmaniose Tegumentar Americana com tons de roxo e Larva Migrans Cutânea recebeu a cor marrom. O objetivo era fazer com que cada doença recebesse uma cor diferente das demais, mas devido à semelhança entre as doenças, seus neurônios receberam cores semelhantes também, fato observado para Esporotricose e Cromoblastomicose. Nesta etapa de validação, a rede errou em sua função de extratora de atributos nos casos em que até mesmo para os médicos especialistas essa diferenciação de padrões é dificultosa e confirmada apenas com exames complementares.

Sendo assim, este trabalho mostra que é possível desenvolver um software para auxiliar o médico na tomada de decisões no momento do diagnóstico. Convém ressaltar que a decisão final será sempre do médico. Os casos avaliados neste trabalho mostram uma boa aproximação entre os resultados obtidos e aqueles esperados, no que diz respeito ao agrupamento de diagnósticos semelhantes, a partir dos sintomas apresentados à rede. Para que o sistema atue como uma verdadeira ferramenta de auxílio à tomada de decisão médica e se torne abrangente e eficiente, novos diagnósticos, exames laboratoriais e o próprio processo de inferência do software devem ser pesquisados.

Durante a realização deste estudo verificou-se a possibilidade de melhorias para possíveis estudos futuros, tais como:

- Treinar a rede ampliando os casos considerados;
- A separação dos diagnósticos em mais do que seis grupos distintos;
- Utilizar outras técnicas para visualizar os agrupamentos formados;
- O emprego de outras técnicas de Reconhecimento de Padrões para a resolução do problema, sempre objetivando a comparação dos resultados entre as diversas ferramentas, procurando pelo método que apresente um melhor desempenho;
- Estudar a viabilidade do sistema desenvolvido como ferramenta de ensino para graduandos da área de medicina.

Referências

- Dias, C. G.; Radonsky, V. (2003) “Desenvolvimento de um sistema de auxílio ao diagnóstico em pediatria com o uso de redes neurais artificiais”. In: *Exacta* (São Paulo) UNINOVE, v. 1, p. 89.
- Haykin, S. (2001) *Redes Neurais: Princípios e práticas*. 2.ed. Porto Alegre: Bookman.
- Kohonen, T. (1990) “Self-Organizing Map”. In: *Proceedings of the IEEE*, 78:1464-1480.
- Kohonen, Teuvo.(1995) *Self-Organizing Maps*. Berlin: Springer-Verlag.
- Massad, M. (2003) *O prontuário eletrônico do paciente na assistência, informação e conhecimento médico*. São Paulo: H. de F. Marin.
- Mendes, Karina Borges (2008) *O uso de redes neurais artificiais no diagnóstico preditivo dos tipos mais frequentes de cefaléia*. Dissertação de Mestrado. Universidade Federal do Paraná.
- Sala, Felipe Andrade. (2005) *Mapas de Kohonen na detecção de eventos epileptogênicos*. Dissertação de Mestrado. Universidade Federal de Santa Catarina.
- Sewo, J.; Silva, P. R. R. (2003) *Rede Neural Treinada com Algoritmo Não - Supervisionado no Contexto de Reconhecimento de Padrões*. Monografia Final de Curso. Universidade Federal de Mato Grosso do Sul.
- Siqueira, Paulo Henrique (2005) *Uma nova abordagem na resolução do problema do caixeiro viajante*. Tese de doutorado. Universidade Federal do Paraná.
- Soares, I. F.; Sant’Ana, I. B. (2006) *Classificação de padrões usando redes neurais de Kohonen*. Monografia Final de Curso. Universidade Católica de Salvador.
- Souza, C. G. e, Souza, L. de M. e. (2004) *O uso de redes neurais no diagnóstico de doenças reumatológicas*. Monografia Final de Curso. Universidade Federal de Santa Catarina.
- Tafner, M.A.; Xerez, M.; Rodrigues Filho, I.W. (1996) *Redes Neurais Artificiais: Introdução e princípios da neurocomputação*. Blumenau: Eko e FURB.
- Ultsch, A. (2003) “Maps for the visualization of high-dimensional data Spaces”. In: *Workshop on Self-Organizing Maps*. Kyushu: Proceedings of Workshop on Self-Organizing Maps - WSOM, p. 225-230.
- Vasconcelos, N. A. P. (2000) *Mapas Auto-Organizativos e suas Aplicações*. Dissertação de Mestrado em Ciências. Universidade Federal do Rio de Janeiro.

- Wyns, B.; Sette, S.; Boullart, L.; Baeten, D.; Hoffman, I.E.A.; Keyser, F. (2004) "Prediction of diagnosis in patients with early arthritis using a combined Kohonen mapping and instance-based evaluation criterion". In: *Artificial Intelligence in Medicine*, v. 31, n. 1, p. 45-55, mai.
- Xu, R.; Wunsch II, D. (2005) Survey of Clustering Algorithms. *IEEE Transactions on Neural Networks*, vol. 16, no. 3, p. 645-678.