

Uma Rede Neuronal Para Diagnósticos em Pneumologia

Flavio da Silva Nery

*Programa de Engenharia de Sistemas e Computação - COPPE,
Universidade Federal do Rio de Janeiro
flaviosnery@gmail.com*

Roseli S. Wedemann

*Instituto de Matemática e Estatística,
Universidade do Estado do Rio de Janeiro
roseli@ime.uerj.br*

Helio Ribeiro de Siqueira

*Hospital Universitário Pedro Ernesto,
Universidade do Estado do Rio de Janeiro
drhelio@infolink.com.br*

Resumo

Este artigo descreve como podemos aplicar uma rede neuronal de classificação como apoio à diagnósticos em pneumologia. Podemos considerar o problema de diagnóstico como sendo equivalente ao de classificação, onde sintomas são apresentados ao médico e este deve apontar uma determinada doença a partir destes dados. A rede faria o papel de auxiliar o médico neste diagnóstico, poupando seu tempo ao apontar as respostas mais prováveis. Isso é muito útil para o aprendizado de estudantes de medicina, que não possuem experiência suficiente para diagnosticar corretamente. A principal vantagem do uso das redes neurais para resolver esse tipo de problema é que ela não precisa saber as regras usadas pelo médico para realizar o diagnóstico, bastando treiná-la através da apresentação de um conjunto de diagnósticos previamente realizados.

1 Introdução

A tuberculose é uma doença que ainda hoje atinge uma parte significativa da população mundial, principalmente onde há situações de miséria e de pobreza. Um diagnóstico rápido se faz necessário, pois quanto mais cedo começa o tratamento, mais chances o paciente tem de ficar curado e de não transmitir a doença para outras pessoas.

Um programa baseado em um modelo de rede neuronal de classificação pode auxiliar o médico nesta tarefa, ao apresentar uma possível resposta a um conjunto grande de dados de entrada, conjunto este que, mesmo um médico experiente, poderia levar algum tempo para combinar e apresentar o diagnóstico. Outra possibilidade é a utilização deste programa no suporte ao aprendizado de estudantes de medicina, visto que eles poderão treinar, realizando o diagnóstico com base em dados de simulação,

comparando-o com o diagnóstico feito pela rede, sem pôr em risco a saúde de um paciente real.

A grande vantagem das redes neurais é que elas não precisam entender a lógica contida na correlação dos dados para fazer a classificação correta. Basta apresentar os dados e informar a classificação correta para um conjunto de diagnósticos conhecidos, treinar as redes e, se o conjunto de treinamento for abrangente o bastante, elas poderão generalizar, realizando diagnósticos de padrões de sintomas de pacientes que não lhe foram apresentados anteriormente.

Este artigo descreve a aplicação de um modelo de redes neurais, os *perceptrons* de múltiplas camadas, no apoio ao diagnóstico de tuberculose, utilizando um programa de classificação que implementa este modelo.

1.1 Tuberculose

A tuberculose é uma doença infectocontagiosa e endêmica, provocada pelo *Mycobacterium tuberculosis* ou bacilo de Koch [9]. Ela pode atingir quase todos os tecidos, mas ocorre em especial nos pulmões. Em 2005, haviam cerca de 9 milhões de pessoas com o bacilo ativo e um terço da população mundial infectada [4].

As áreas mais atingidas pela tuberculose são as que possuem condições sócio-econômicas piores. Países mais pobres têm mais tuberculose, sendo atingidas, em maioria, as regiões mais atrasadas, os espaços urbanos mais precários e ambientes socialmente fechados. Alguns exemplos de ambientes favoráveis ao desenvolvimento da doença são: favelas, mocambos, alagados, invasões, populações agrupadas em decorrência de calamidades públicas, escolas, creches, presídios, abrigos sociais ou de idosos etc. [6, p. 29].

Doença curável e sensível à ação dos profissionais, a tuberculose é um grave problema da Saúde Pública, destruindo, principalmente, a vida de cidadãos frágeis

socialmente em idade produtiva [6, p. 31].

Diagnosticar e tratar o mais rápido possível a tuberculose é uma grande medida prática para salvar vidas e recuperar a saúde dos enfermos. [...] Hoje o método de redução da tuberculose na sociedade é a busca de casos novos e o seu tratamento adequado [6, p. 31].

1.2 Trabalhos relacionados

Existem na literatura vários trabalhos aplicando redes neuronais artificiais (RNAs) no diagnóstico de doenças. Fizemos uma revisão desses trabalhos, selecionando os mais relacionados ao diagnóstico de tuberculose.

Ambas as referências [1] e [3] tratam de diagnóstico diferencial de doenças intersticiais do pulmão, classificando 11 doenças. A rede que faz essa classificação possui arquitetura *feed-forward*, utiliza *back-propagation* para o treinamento e é composta por 26 terminais de entrada, 18 unidades na camada escondida e 11 unidades na camada de saída. Nestes trabalhos a rede obteve um ótimo resultado, além de ajudar a melhorar o desempenho de radiologistas na tarefa de classificação.

Temos na referência [12] a detecção de tuberculose através do reconhecimento de padrões de imagem em radiografias, utilizando *perceptrons* de múltiplas camadas com o algoritmo de treinamento *back-propagation*. Imagens digitalizadas de 512×512 pixels são reduzidas para 10×10 pixels, onde cada pixel é um parâmetro de entrada para a rede, formando um conjunto de 100 terminais de entrada. Foram testadas várias configurações de rede, e a que obteve melhor desempenho (taxa média de erro de 0.08 em 800 épocas de treinamento para 10 padrões no conjunto de treinamento), possui uma camada escondida com 5 unidades e utiliza uma taxa de aprendizagem igual a 0.2 e constante *momentum* igual a 0.6.

Em [5], foi realizada a predição de tuberculose pulmonar ativa utilizando RNAs. A rede construída no artigo citado (denominada **rede neuronal de regressão**) possui uma arquitetura muito parecida com a dos *perceptrons* de múltiplas camadas (rede utilizada no presente artigo), com uma característica que difere uma da outra: a quantidade de unidades na camada intermediária, que é arbitrária nos *perceptrons* e fixa (igual ao número de padrões apresentados à rede para o treinamento) na rede neuronal de regressão. Os 21 parâmetros de entrada estão divididos entre variáveis demográficas, sintomas e achados radiológicos. A rede de regressão superou o desempenho dos médicos

A referência [2] trata do problema de diagnóstico de doenças intersticiais do pulmão. A rede utilizada

no trabalho citado classifica 9 doenças, utiliza o algoritmo de *back-propagation* para o treinamento e recebe 20 parâmetros de entrada (6 clínicos e 14 radiológicos). Foram criados 90 casos hipotéticos (10 de cada doença) e também foi realizado um método de validação em que metade dos casos, escolhidos aleatoriamente, eram utilizados no treinamento e a outra metade para testar a classificação. O treinamento desta rede com 6 unidades na camada escondida era concluído após o conjunto de treinamento ter sido apresentado a ela 200 vezes. No teste de validação, a RNA obteve o mesmo desempenho de radiologistas experientes.

Em [11] é apresentado um estudo de usabilidade das RNAs na classificação de nódulos existentes em tomografias computadorizadas de alta-resolução como malignos e benignos, usando uma rede *feed-forward* de três camadas (no artigo não é especificado se os terminais de entrada são considerados como uma camada) com algoritmo de treinamento *back-propagation*. No trabalho citado foram utilizados 155 casos pré-existentes para o treinamento, sendo 99 malignos (todos comprovados em análise patológica) e 56 benignos (27 comprovados patologicamente e 29 após acompanhamento do caso). Em cada caso foram analisadas 23 características (7 clínicas e 16 radiológicas) que serviram como dados de entrada para o treinamento. Outros 50 casos (metade malignos e metade benignos) foram utilizados como casos de teste para três grupos de radiologistas e para a RNA. Além de ter obtido um ótimo desempenho nos testes, a RNA ajudou a aumentar consideravelmente a quantidade de acertos dos três grupos de radiologistas, quando utilizada como apoio ao diagnóstico.

Todos os resultados apresentados nos referidos trabalhos se mostraram favoráveis com relação à utilização das RNAs na resolução de problemas de diagnóstico.

2 *Perceptrons* de múltiplas camadas

Para resolver o problema de diagnóstico, utilizamos o modelo de *perceptrons* de múltiplas camadas. Este modelo, descrito pela primeira vez em 1962, por Rosenblatt e seus colaboradores [13], é uma rede *feed-forward* que possui uma série de terminais de entrada, cuja única função é alimentar o restante da rede com padrões de entrada. Após estes terminais existem uma ou mais camadas intermediárias seguidas de uma camada de saída da qual o resultado do cálculo da rede é lido. Na classe de redes *feed-forward* não há conexões direcionadas de unidades de uma camada para unidades da camada anterior e também não existem conexões entre unidades da mesma camada. A Figura 1 mostra uma rede de *perceptrons* de duas camadas. Neste artigo, usaremos a notação adotada em [8, cap. 6].

As unidades de saída são denotadas por O_i , as

unidades da camada escondida por V_j e as unidades de entrada por ξ_k . Existem conexões w_{jk} das unidades de entrada para as unidades escondidas, e conexões W_{ij} das unidades escondidas para as unidades de saída. Os índices i , j e k referem-se, respectivamente, a uma unidade da camada de saída, a uma unidade da camada escondida e a uma unidade da camada de entrada. As unidades de entrada podem receber valores discretos ou contínuos. O índice μ na unidade ξ_k^μ representa o padrão do conjunto de treinamento que está sendo apresentado à rede, e $\mu = 1, 2, \dots, p$, onde p é o número de padrões de treinamento. Ressalte-se que a saída de um neurônio é dada por uma função de ativação $g(h)$ não-linear.

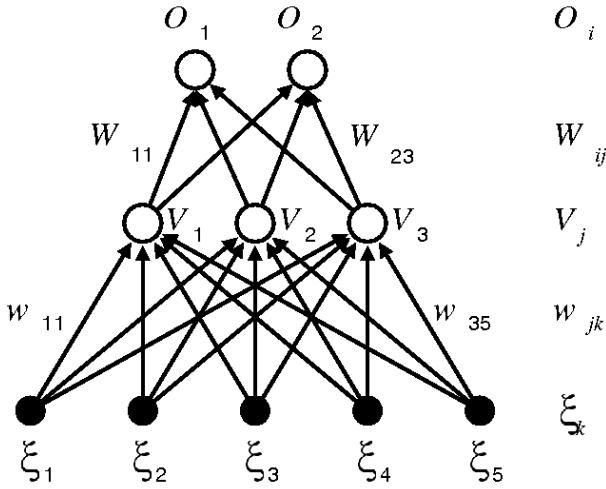


Figura 1. Perceptron de duas camadas com as notações utilizadas neste artigo.

Para um padrão de entrada μ , a unidade j da camada escondida receberá como entrada da rede

$$h_j^\mu = \sum_k w_{jk} \xi_k^\mu, \quad (1)$$

gerando a saída

$$V_j^\mu = g(h_j^\mu) = g\left(\sum_k w_{jk} \xi_k^\mu\right). \quad (2)$$

A unidade de saída i recebe como entrada

$$\begin{aligned} h_i^\mu &= \sum_j W_{ij} V_j^\mu \\ &= \sum_j W_{ij} g\left(\sum_k w_{jk} \xi_k^\mu\right), \end{aligned} \quad (3)$$

e gera como saída final

$$\begin{aligned} O_i^\mu &= g(h_i^\mu) \\ &= g\left(\sum_j W_{ij} g\left(\sum_k w_{jk} \xi_k^\mu\right)\right) \end{aligned}$$

$$= g\left(\sum_j W_{ij} g\left(\sum_k w_{jk} \xi_k^\mu\right)\right). \quad (4)$$

2.1 Aprendizagem dos perceptrons

Para a aprendizagem foi utilizado o algoritmo de *back-propagation* descrito em [8, cap. 6] com as seguintes características:

- acréscimo de um critério de parada do algoritmo e
- acréscimo da constante *momentum* (α) na regra de atualização dos pesos.

2.1.1. Critério de Parada. Há dois critérios bastante utilizados: a variação da média do erro total e o número de vezes que o conjunto de treinamento é apresentado à rede (época). No primeiro critério, verifica-se a cada época a variação da média do erro total com relação à média da época anterior. A média de erro total é definida pela equação

$$E_{AV} = \frac{1}{p} \sum_{\mu=1}^p E[w], \quad (5)$$

onde $E[w]$ é a função custo

$$E[w] = \frac{1}{2} \sum_i (\xi_i^\mu - O_i^\mu)^2, \quad (6)$$

e p é o número de padrões do conjunto de treinamento. Se a variação for menor do que um determinado valor (geralmente algum valor no intervalo $[0.01, 0.1]$) podemos considerar a rede treinada. No segundo critério, considera-se a rede treinada após determinado número de épocas. Não existe um número padrão de épocas necessárias para que a rede esteja treinada. Outro critério que pode ser adotado é a combinação dos anteriores. Neste caso considera-se a rede treinada quando ela satisfaz os dois critérios simultaneamente. O critério de parada é muito importante, pois, se usamos épocas demais, podemos aumentar desnecessariamente o tempo de treinamento e, além disso, pode haver um super-treinamento (*over-training*) da rede e até uma queda de desempenho.

2.1.2. Atualização dos Pesos. Considere a seguinte regra geral de atualização dos pesos

$$\Delta w_{pq} = \eta \delta_p^\mu V_q^\mu \quad (7)$$

onde p e q se referem, respectivamente, às unidades de destino e de origem da conexão em questão, δ_p^μ é o erro que vem da unidade p e V_q^μ é o sinal que sai da unidade q .

A velocidade da aprendizagem é determinada pelo parâmetro η que é a taxa de aprendizagem. Se essa

taxa for pequena, o treinamento fica muito lento. Mas, se essa taxa for alta demais, apesar de o treinamento ser mais rápido, a aprendizagem pode ficar instável, fazendo com que o valor dos pesos oscile [8, cap. 5, p. 106]. A solução é modificar a regra geral de atualização Δw_{pq} e acrescentar um termo que leve em conta a atualização anterior [7, cap. 4, seção 4.6]. Então esta regra transforma-se em

$$\Delta w_{pq}(t) = \eta \delta_p^\mu V_q^\mu + \alpha \Delta w_{pq}(t-1) \quad (8)$$

onde α é conhecido como **parâmetro momentum**. O valor de α deve estar no intervalo $0 \leq \alpha < 1$ e o valor 0.9 é a escolha mais comum [8, cap. 6, p. 123]. Note que, se $\alpha = 0$, voltamos a ter a equação 7. O parâmetro α acelera a descida na direção de declives regulares e tem o efeito de estabilização das direções que oscilam no tempo [10].

3 Implementação do Diagnóstico de Tuberculose

3.1 Estrutura de Dados da Implementação

O programa desenvolvido implementa uma configuração qualquer de *perceptron*, de forma que qualquer aplicação de classificação possa utilizá-lo. A rede pode ser representada por um grafo, em que cada unidade corresponde a um nó do grafo e as conexões entre as unidades correspondem às arestas do grafo. Em geral, a forma mais imediata de representarmos grafos é por meio de matrizes $n \times n$, onde n é o número total de nós do grafo. Poderíamos representar os *perceptrons* desta forma, com a desvantagem de que obteríamos matrizes esparsas já que:

- não existe comunicação entre unidades na mesma camada;
- só há comunicação entre unidades de camadas vizinhas;
- a comunicação é somente em uma direção.

A solução encontrada foi mapear os pesos por camada, isto é, em vez de somente uma matriz, temos M matrizes $a \times b$, onde M é o número de camadas, a é o número de unidades na camada $m-1$ e b é o número de unidades na camada m .

Apesar de ser a melhor solução em termos de uso de memória, foi preciso fazer algumas adaptações para manter localidade espacial de armazenamento de memória e facilidade de acesso direto a um determinado peso. A primeira foi resolvida alocando um espaço de memória contíguo, cujo tamanho é igual ao número de conexões existentes na rede. A segunda foi resolvida utilizando um vetor de M posições cujo l -ésimo elemento guarda o número total de conexões existentes até a camada $l-1$. Assim, é possível saber qual será o deslocamento necessário para chegar

àquela camada, ou seja, seu endereço base. Isso possibilitou o acesso aos pesos na forma $w(m, i, j)$, que se refere ao peso da conexão entre a i -ésima unidade da camada m e a j -ésima unidade da camada $m-1$, $m = 1, \dots, M$.

Guardando esses deslocamentos e realizando algumas contas simples, podemos acessar qualquer conexão da rede. Seja $inicio(m)$ a função que retorna o endereço base da camada m e $unidades(m)$, a função que retorna o número de unidades na camada m , temos que o endereço correspondente ao peso $w(m, i, j)$ é $inicio(m) + unidades(m-1) \times i + j$.

3.2 Levantamento de Dados de Treinamento

Para verificar a viabilidade da aplicação do método ao problema de diagnóstico da tuberculose, obtivemos dados com um médico especialista em tuberculose (o terceiro autor deste artigo). Com a ajuda do médico, realizamos um levantamento dos parâmetros de entrada para este problema. Seleccionamos todos os itens julgados necessários para o diagnóstico de tuberculose. Estes dados foram separados em quatro categorias: dados clínicos (21), dados sociais (3), dados laboratoriais (17) e dados radiológicos (5), totalizando 46 parâmetros. As duas classes nas quais a rede classifica o conjunto de entrada são: “Tuberculose” e “Outra doença ou sem diagnóstico”.

A partir deste levantamento, elaboramos um formulário de coleta para a aquisição do conjunto de treinamento da rede, que foi avaliado e aprovado pelo médico especialista. Após a aprovação, seguiu-se ao preenchimento dos formulários a partir de prontuários selecionados pelo médico. Os dados recolhidos foram transformados em padrões para o conjunto de treinamento.

Foram coletados dados de 31 prontuários, sendo 25 casos da classe “Tuberculose” e 6 da classe “Outra doença ou sem diagnóstico”. Estes foram utilizados como conjunto de treinamento do software de classificação.

4 Resultados

O próximo passo foi treinar a rede, fazendo o ajuste de seus parâmetros e testando seu desempenho. Utilizamos a função de ativação $g(h) = f_\beta = [1 + e^{-2\beta h}]^{-1}$, e portanto, os valores de saída de cada unidade estarão no intervalo $[0, 1]$. Foi atribuído o valor 0,5 para β , valor mais utilizado para este parâmetro pois determina uma inclinação suave da curva sigmóide. O valor de α foi variado conforme a necessidade ao longo dos testes.

Atribuímos valores iniciais aos pesos aleatoriamente no intervalo $[-1, 1]$ do conjunto dos números reais. A ordem de apresentação dos padrões do conjunto de treinamento também se deu de forma aleatória para cada época. Fizemos testes com os

dois critérios de parada do algoritmo, mas o critério da variação da média de erro total permitiu obter um melhor desempenho da rede. Testamos vários valores mínimos de variação e observamos que o valor 0,0001 proporcionou os melhores resultados.

Para o teste de desempenho, foram executadas 10 rodadas de treinamento e classificação. Para a classificação, foram apresentados 16 padrões à rede treinada, recolhidos com a ajuda do médico especialista, representando casos reais, onde 7 eram da classe “Tuberculose” e 9 da classe “Outra doença ou sem diagnóstico”. Foi considerada uma classificação correta aquela cuja a diferença entre a saída e a resposta correta fosse, em módulo, menor que 0,4 para todas as unidades de saída.

De acordo com a quantidade de parâmetros de entrada e as classes definidas anteriormente, determinamos o número de unidades das camadas de entrada e de saída, 46 e 2, respectivamente. Para definir o número de unidades na camada intermediária utilizamos o critério de aumentar esse número até que o desempenho começasse a cair.

Inicialmente foi fixada a taxa de aprendizagem no valor 0,3 e calculada a média da taxa de acerto para cada quantidade de unidades na camada intermediária. Ao diminuir η para 0,2 passamos a obter resultados melhores. Mantivemos então η fixo nesse valor e variamos o número de unidades na camada intermediária e o valor de α . O melhor resultado que alcançamos foi com 4 neurônios na camada intermediária e $\alpha = 0,2$, onde obtivemos uma taxa de acerto de 0,7938 (uma média de quase 13 padrões dos 16 apresentados classificados corretamente). Não satisfeitos com este resultado, seguimos para a análise dos conjuntos de treinamento e de classificação e observamos que existiam padrões no conjunto de classificação que não estavam representados pelo conjunto de treinamento.

Com a ajuda do médico especialista, analisamos os 31 prontuários para o treinamento e os 16 casos para a classificação, e selecionamos fichas de pacientes dos dois conjuntos que indicavam, através de fatores determinantes, se o diagnóstico era de “Tuberculose” ou de “Outra doença ou sem diagnóstico”. Dessa forma, definimos um novo conjunto de treinamento que contava com 22 padrões sendo 11 deles da classe “Tuberculose” e 11 da classe “Outra doença ou sem diagnóstico”.

Fizemos novos testes, desta vez com os 25 padrões restantes no conjunto de classificação buscando a melhor configuração possível. Com 5 neurônios na camada intermediária, $\eta = 0,2$, $\alpha = 0,1$ e variação da média de erro total de 0,0001, conseguimos uma taxa de acerto média de 0,6400 (16 acertos de 25 possíveis). A partir deste resultado, realizamos um procedimento de otimização do conjunto de treinamento, onde retirávamos um padrão do conjunto de classificação, adicionávamos este padrão ao conjunto

Número de padrões		Taxa de acerto	Média de erros
Treinamento	Classificação		
22	25	0,6400	9,0000
23	24	0,8167	4,3992
24	23	0,9000	2,3000
25	22	0,9273	1,5994
26	21	0,9476	1,1004

Tabela 1. Taxa de acerto dos conjuntos de treinamento. A linha em negrito indica o conjunto de treinamento ótimo.

de treinamento e executávamos o teste de desempenho utilizando o novo conjunto de treinamento e apresentando o conjunto de classificação resultante. Este teste era feito para todos os padrões do conjunto de classificação e o que obtivesse a melhor taxa de acerto era retirado do conjunto de classificação e passava a fazer parte definitivamente do conjunto de treinamento. Este procedimento se repetiu até que encontramos um conjunto de treinamento ótimo com 26 padrões, restando 21 padrões no conjunto de classificação. Observamos que 3 dos 9 padrões que eram classificados erroneamente entraram no conjunto ótimo de treinamento, indicando a importância destes padrões para o treinamento da rede. A Tabela 1 mostra os conjuntos de treinamento intermediários com suas respectivas taxas de acerto.

Em seguida, analisamos os resultados dos testes e constatamos que nas dez rodadas de treinamento e classificação executadas com o conjunto ótimo, sempre o mesmo padrão era classificado equivocadamente (em apenas uma das rodadas outro padrão além deste foi classificado errado). Verificamos, então, a ficha de paciente correspondente a este padrão e vimos que era um padrão confuso para a classificação mesmo para um médico, mantendo características típicas das duas classes, o que justifica o erro. Outro fato que observamos nos testes foi a sensibilidade do treinamento ao conjunto inicial de pesos e à aleatoriedade da apresentação dos padrões de treinamento, resultando em taxas de acerto ótimas e péssimas para um mesmo conjunto de treinamento.

5 Conclusões

Os *perceptrons* alcançaram resultados significativos no diagnóstico de tuberculose, para um conjunto de treinamento razoavelmente pequeno, mas para isso foi importante conseguirmos um conjunto de treinamento adequado representativo do problema de diagnóstico abordado. Esperamos que, com mais padrões no conjunto de treinamento, consigamos realizar o diagnóstico de outras doenças relacionadas aos dados coletados, aumentando o número de classes de resposta, ou ainda, fazer o diagnóstico diferencial de doenças do pulmão. Também é necessário aumen-

tarmos o conjunto de classificação e realizarmos mais testes neste conjunto para uma validação mais completa do treinamento.

Referências

- [1] H. Abe, K. Ashizawa, F. Li, N. Matsuyama, A. Fukushima, J. Shiraishi, H. MacMahon, and K. Doi. Artificial neural networks (anns) for differential diagnosis of interstitial lung disease: Results of a simulation test with actual clinical cases. *Academic Radiology*, 11(1), jan 2004.
- [2] N. Asada, K. Doi, H. MacMahon, S. M. Montner, M. L. Giger, C. Abe, and Y. Wu. Potential usefulness of an artificial neural network for differential diagnosis of interstitial lung diseases: Pilot study. *Radiology*, 177:857–860, 1990.
- [3] K. Ashizawa, H. MacMahon, T. Ishida, K. Nakamura, C. J. Vyborny, S. Katsuragawa, and K. Doi. Effect of an artificial neural network on radiologists performance in the differential diagnosis of interstitial lung disease using chest radiographs. *American Journal of Roentgenology*, 172:1311–1315, 1999.
- [4] M. A. Aziz and A. Wright. The world health organization/international union against tuberculosis and lung disease global project on surveillance for anti-tuberculosis drug resistance: A model for other infectious diseases. *Clinical Infectious Diseases*, 41:S258–S262, 2005.
- [5] A. A. El-Solh, C.-B. Hsiao, S. Goodnough, J. Serghani, and B. J. B. Grant. Predicting active pulmonary tuberculosis using an artificial neural network. *Chest*, 116:968–973, 1999.
- [6] FUNASA, editor. *Controle da Tuberculose - Uma Proposta de Integração Ensino-Serviço*. FUNASA/CRPHF/SBPT, Rio de Janeiro, 5 edition, 2002. Inclui bibliografia.
- [7] K. Gurney. Neural nets. Versão HTML, 1996. Disponível em: <http://www.shef.ac.uk/psychology/gurney/notes/contents.html>.
- [8] J. A. Hertz, A. S. Krogh, and R. G. Palmer. *Introduction to the theory of neural computation*, volume 1. Perseus Books, Santa Fe Institute, Studies in the Science of Complexity, Cambridge, MA, 1991.
- [9] A. Houaiss and M. S. Villar. *Dicionário Houaiss da língua portuguesa*. Objetiva, Rio de Janeiro, versão 1.0, cd-rom edition, 2001.
- [10] E. Marchiori. Notas de aula, 2006. Disponível em: http://www.cs.vu.nl/~elena/slides03/nn06_2light.ppt.
- [11] Y. Matsuki, K. Nakamura, H. Watanabe, T. Aoki, H. Nakata, S. Katsuragawa, and K. Doi. Usefulness of an artificial neural network for differentiating benign from malignant pulmonary nodules on high-resolution ct: Evaluation with receiver operating characteristic analysis. *American Journal of Roentgenology*, 178:657–663, 2002.
- [12] K. V. Ramana and S. K. Basha. Neural image recognition system with application to tuberculosis detection. In *Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC04)*, volume 2, pages 694–698. IEEE, 2004.
- [13] F. Rosenblatt. *Principles of Neurodynamics*. Spartan, Nova Iorque, 1962.