**REDES NEURAIS ARTIFICIAIS APLICADAS AO DIAGNÓSTICO DE CÂNCER DE MAMA**

GABRIEL NAVARRO LOPES1

1Engenharia de Controle e Automação, 9º semestre, prontuário nº 1410041.

A rede neural artificial (RNA) desenvolvida nesse trabalho tem como objetivo realizar o diagnóstico de câncer de mama com base em diversas informações do paciente. As informações utilizadas para o treinamento da rede são:

1. Idade;
2. Índice de massa corporal (IMC);
3. Índice HOMA;
4. Nível de glicose;
5. Nível de insulina;
6. Nível de leptina;
7. Nível de adiponectina;
8. Nível de resistina; e
9. Nível de proteína quimioatratora de monócitos (MCP-1).

Esses elementos foram retirados de um repositório virtual (<https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Coimbra>) onde estão disponíveis os dados de 116 pacientes diagnosticados por estudantes de medicina da universidade de Coimbra que foram separados em duas classes: 0 para pacientes saudáveis (52 no total) e 1 para pacientes que apresentavam câncer de mama (64 pacientes). Destas 116, 82 amostras foram utilizadas para treinamento da rede e as 34 restantes para testes e validação.

As amostras com seus respectivos dados foram disponibilizadas em um arquivo de extensão .xlsx, onde foram organizadas em colunas de acordo com a característica à qual a informação se refere. Para realizar a leitura do arquivo de amostras foi utilizada a biblioteca *ExcelDataReader*.

Antes de iniciar o aprendizado da rede foi necessário normalizar todos os dados de entrada para uma faixa entre 0 e 1, com base nos limites máximos e mínimos de cada grupo de informação. A equação 1 apresenta a forma de normalização utilizada:

A biblioteca *Accord* foi utilizada para elaboração da RNA. Como função de ativação optou-se por uma função sigmoidal bipolar. O modelo da rede conta com 9 neurônios na camada de entrada, 1 neurônio na camada de saída e uma única camada intermediária com número de neurônios definido pelo usuário. O rebalanceamento da rede, ou aprendizado, foi realizado com base na técnica *Parallel Resilient Backpropagation* (Retropropagação Resiliente Paralela, em tradução livre).

Os pesos da rede são inicializados aleatoriamento e o aprendizado é então realizado com base em iterações sobre todo o conjunto de amostras de treinamento. O erro quadrático médio é calculado após cada iteração para checar pela condição de finalização do aprendizado, cujo valor é definido pelo usuário.

Após o fim do treinamento é possível verificar os resultados do classificador sobre o conjunto de treinamento e também o de testes. Os melhores resultados foram obtidos para uma rede apresentando 5 neurônios na camada intermediária, cujo parâmetro alfa da sigmoide foi definido como 1 e onde o erro desejado para a rede finalizar o treinamento foi estabelecido em 1. Percebeu-se que tentar minimizar esse erro tornava a rede muito eficaz na classificação das amostras de treinamento, mas incapaz de classificar de forma competente as amostras de teste, um fenômeno conhecido como sobreajuste.

Para a configuração ótima anteriormente mencionada, obteve-se uma taxa de acerto de 94,47% sobre o conjunto de treinamento e uma taxa menor correspondente a 79,95% de acerto sobre o conjunto de amostras de teste.

Não foi possível salvar o estado (modelo e os pesos) da RNA após a realização do treinamento, o que dificultou muito o processo e impossibilitou a reprodutibilidade de certos resultados. Foi possível notar também que o tempo necessário para treinar e os resultados finais da rede muito dependiam dos pesos iniciais, que eram gerados aleatoriamente e, portanto, estavam fora do controle do usuário.

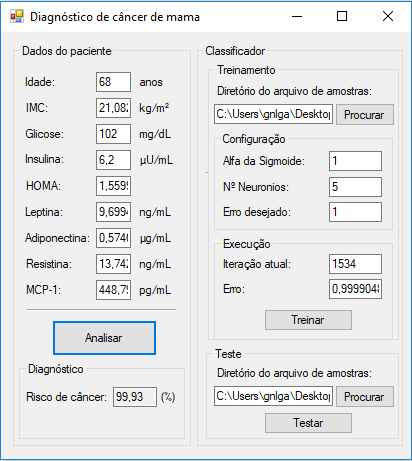
Os resultados obtidos demonstram que a técnica de RNA é aplicável à previsão de doenças e outras anomalias como o câncer de mama. Porém para isso é necessário muito cuidado e, possivelmente, um maior número de amostras para tornar o conjunto de treinamento mais representativo. Uma taxa de acerto de 79,95% pode ser encarada como alta, porém no contexto da saúde não pode ser levada como único juízo na decisão médica. Ainda assim, os resultados podem servir como uma forma de alerta para uma análise mais profunda de pacientes cujos resultados acusem uma chance significativa de câncer.

Tabela 1: Resultados obtidos para o conjunto de testes.

|  |  |  |
| --- | --- | --- |
| Amostra | Rotulo | Resultado |
| 1 | 0 | 0,931 |
| 2 | 0 | 0,791 |
| 3 | 0 | 0,000 |
| 4 | 0 | 0,000 |
| 5 | 0 | 0,000 |
| 6 | 0 | 0,000 |
| 7 | 0 | 0,000 |
| 8 | 0 | 0,460 |
| 9 | 0 | 0,000 |
| 10 | 0 | 0,986 |
| 11 | 0 | 0,000 |
| 12 | 0 | 0,679 |
| 13 | 0 | 0,549 |
| 14 | 0 | 0,000 |
| 15 | 0 | 0,000 |
| 16 | 1 | 0,000 |
| 17 | 1 | 0,999 |
| 18 | 1 | 0,843 |
| 19 | 1 | 0,998 |
| 20 | 1 | 0,999 |
| 21 | 1 | 0,999 |
| 22 | 1 | 0,240 |
| 23 | 1 | 0,842 |
| 24 | 1 | 0,727 |
| 25 | 1 | 0,999 |
| 26 | 1 | 0,999 |
| 27 | 1 | 0,998 |
| 28 | 1 | 0,999 |
| 29 | 1 | 0,999 |
| 30 | 1 | 0,931 |
| 31 | 1 | 0,999 |
| 32 | 1 | 0,995 |
| 33 | 1 | 0,998 |
| 34 | 1 | 0,999 |

Fonte: Elaboração própria.

Figura 1: Janela do usuário.



Fonte: Elaboração própria.