

PREVISÃO DA VITALIDADE DE RECÉM-NATOS

NÚBIA K. O. ALMEIDA, CARLOS E. PEDREIRA, RENAN M. V. R. ALMEIDA

*Programa de Engenharia Biomédica, COPPE, Universidade Federal do Rio de Janeiro
Caixa Postal 68510, Cidade Universitária, Ilha do Fundão – 21945-970, Rio de Janeiro, RJ
E-mails: getnkoa@vm.uff.br, carlosp@centroin.com.br, renan@peb.ufrj.br*

Abstract— An artificial neural network model was investigated with the goal of predicting the vitality level (*Apgar*) of newborns, taking into account information measured during pregnancy besides other characteristics from the mothers. We used a neural network structure (4x3x1), with the following dichotomic attributes: *place of birth* (in a health facility / in other), *income Human Development Index* of the mother's residency state (<0.7 / ≥ 0.7), *duration of pregnancy* (pre-term/term) and *birth weight* ($<1,500$ g / $\geq 1,500$ g). In the test set the overall rate of correct classification was 63.62% (68.58% sensitivity and 63.60% specificity). The classification rule generated by a NeuroRule algorithm indicated that if the birth is at term, it takes place in a health center, the mother lives in a region with high HD and the newborn weights 1,500 g or more, then the vitality of the unborn child would be estimated as high *Apgar*.

Keywords— *Apgar* Score, Artificial Neural Networks, Bayesian Regularization, Neural Interpretation Diagram, Sensibility, Garson's Algorithm, NeuroRule.

Resumo— Foi investigada uma rede neural artificial para prever o índice de vitalidade (*Apgar*) de um recém-nato, baseada em informações medidas durante a gravidez além de características da mãe. Foi utilizada a rede neural de estrutura 4x3x1 com os seguintes atributos dicotômicos: *local de nascimento* (em estabelecimento de saúde/em outro tipo de estabelecimento), *IDH-renda* do estado onde reside a mãe ($< 0,7/\geq 0,7$), *duração da gestação* (pré-termo/termo) e *peso ao nascer* ($< 1.500g/\geq 1.500g$). No conjunto de teste foram obtidos percentual de acerto global de classificação de 63,62%, sensibilidade 68,58% e especificidade de 63,60%. A regra de classificação gerada pelo NeuroRule, a partir da rede 4x3x1, indica que se o nascimento for a termo, ocorrer em estabelecimentos de saúde, a mãe residir em estado com alto IDH-renda e o recém-nascido nascer com peso maior ou igual a 1.500g, então a vitalidade do nascituro deve ser estimada como alto *Apgar*.

Palavras-chave— Índice *Apgar*, Redes Neurais Artificiais, Regularização Bayesiana, Diagrama de Interpretação Neural, Sensibilidade, Algoritmo de Garson, NeuroRule.

1 Introdução

O exame *Apgar* foi desenvolvido na década de 50, com o objetivo de avaliar, de uma forma objetiva, a vitalidade de recém-nascidos [1]. Este exame vem sendo largamente utilizado e tornou-se obrigatório aos 1 e 5 minutos após o nascimento. O resultado do *Apgar* ao 1º minuto representa um diagnóstico imediato do recém-nascido, enquanto o resultado ao 5º minuto (foco do interesse deste estudo) é um índice prognóstico de sua saúde neurológica [2]. Neste exame, uma nota menor ou igual a 6 pontos é considerada baixa, indicando alto risco de injúria hipóxica-isquêmica ou problemas neurológicos futuros para o recém-nascido.

Um modelo preditor do *Apgar* no início de uma gestação baseado em dados e suposições sobre a gravidez poderia conduzir a condutas que, durante a gestação, levassem à reversão dos fatores que determinaram um mau prognóstico. Além disso, estas previsões poderiam ser atualizadas ao longo da gestação, permitindo seu monitoramento. Atualmente não existe modelo de previsão do índice *Apgar* com essas características.

O objetivo central deste estudo foi investigar a potencialidade de métodos baseados em Redes Neurais Artificiais, como estimador do resultado (bai-

xo/alto) do *Apgar* ao 5º minuto, a partir de informações da gestante, da gestação, do parto e do nascituro. Os objetivos específicos são: projetar uma rede neural artificial, com bom desempenho, para prever o *Apgar*, principalmente no que se refere à sensibilidade e criar uma regra de classificação para o *Apgar* de fácil entendimento e aplicação, a partir da consideração de padrões gestacionais.

2 Materiais e Métodos

Neste estudo utilizou-se principalmente o banco de dados dos registros de nascimentos ocorridos no Brasil no ano de 2005, o SINASC, desenvolvido pelo DATASUS – Ministério da Saúde [3]. As informações usadas são referentes: i) à gestante: idade, estado civil, escolaridade, quantidade de filhos vivos e de filhos mortos, ocupação e município onde reside, ii) à gestação: quantidade de consultas de pré-natal e tempo de gestação, iii) ao parto: tipo de parto, local e município onde ocorreu o parto e iv) ao nascituro: sexo, peso ao nascer e *Apgar* o 5º minuto. Além destas variáveis, foi incorporado ao banco de dados o atributo IDH-renda do estado (PNUD 2005) onde reside a gestante. Foram considerados no estudo 1.952.486 registros decorrentes apenas de gestações únicas (não gêmeos).

A Tabela 1 apresenta os cinco itens para avaliação do *Apgar* [1, 4]. A cada um desses itens é atribuída nota 0, 1 ou 2, de acordo com características ou comportamento do recém-nascido. A avaliação final é dada pela soma das notas de todos os itens, resultando, portanto, em um valor de 0 a 10.

A Tabela 2 apresenta os atributos contemplados nesta investigação, com suas respectivas faixas de valores ou categorias originais, bem como as categorias binárias adotadas no presente estudo.

Tabela 1. Itens de avaliação do *Apgar*

Itens	Pontos		
	0	1	2
Frequência Cardíaca	Ausente	< 100 bpm	≥ 100 bpm
Respiração	Ausente	Fraca, Irregular	Forte / Choro
Tônus Muscular	Flácido	Flexão de pernas e braços	Movimento ativo / Boa flexão
Cor	Cianótico / Pálido	Cianose das extremidades	Rosado
Irritabilidade Reflexa	Ausente	Algum movimento	Espirro / Choro

2.1 População com Categorias Desbalanceadas

Dos registros utilizados, 32.323 (1,66%) correspondem aos nascituros com baixo valor de *Apgar*. Para lidar com a questão de classes desbalanceadas, foi usada a metodologia de *undersampling* [5]. Tal procedimento descarta registros da classe mais populosa até que a quantidade de observações de cada uma das classes seja a mesma. Desta forma, dos 1.920.163 registros de nascimentos correspondentes a recém-nascidos com alto valor de *Apgar*, foi extraída uma amostra aleatória de 32.323 registros, obtendo-se um banco de dados com 64.646 registros, 50% deles referentes a cada uma das duas categorias de *Apgar*.

2.2 Rede Neural Artificial

O conjunto de dados foi dividido em aproximadamente 60% (19.393 registros) para o treinamento da rede, 20% (6.465 registros) para a validação do treinamento e 20% (6.465 registros) para o teste do desempenho da rede neural. A distribuição dos dados nestes conjuntos foi feita aleatoriamente, mantendo dentro de cada um o percentual de 50% para cada categoria *Apgar*. Foi criado também um outro conjunto de teste (denominado de teste auxiliar) constando dos 6.465 registros de baixo *Apgar* que já participavam do primeiro conjunto de teste, somados aos

1.887.840 registros de alto *Apgar* ignorados após a seleção da amostra da subpopulação de alto *Apgar*

Tabela 2. Variáveis utilizadas

Variáveis	Categorias/Valores	
	Originais	Estudo
Atributos:		
Idade (anos)	10 a 59	18 a 35 / <18 ou >35
Estado civil	solteira/casada, viúva/separada judicialmente/ união consensual	casada ou união consensual/ solteira, viúva ou separada judicialmente
Escolaridade (anos)	0 / 1 a 3 / 4 a 7 / 8 a 11 / 12 ou mais	0 a 7 / 8 ou mais
Primípara	nº de filhos vivos e mortos	sim / não
Dona de casa	código de ocupação	sim / não
Nº consultas de pré-natal	0 / 1 a 3 / 4 a 6 / 7 ou mais	4 ou mais / menos de 4
Sexo	masculino / feminino	masculino / feminino
Tempo de gestação (semanas)	menos de 22 / 22 a 27 / 28 a 31 / 32 a 36 / 37 a 41 / 42 ou mais	menos de 37 / 37 ou mais
Peso (gramas)	180 a 6.630	< 1.500 / ≥ 1.500
Local de nascimento	hospital / outro estabelecimento de saúde / domicílio / outro	hospital ou outro estabelecimento de saúde/ domicílio ou outro
Deslocamento intermunicipal	Municípios de residência e de onde ocorreu o parto	sim / não
Tipo de parto	vaginal/ cesáreo	vaginal/ cesáreo
IDH-renda	0 a 1	≥ 0,7 / < 0,7
Desfecho:		
Apgar ao 5º minuto	0 a 10	≤ 6 / > 6

(seção *População com Categorias Desbalanceadas*). Desta forma, o conjunto de teste auxiliar ficou constituído por 0,34% de observações de baixo *Apgar* e 99,66% de observações de alto *Apgar*.

O método de treinamento foi retropropagação do erro, com regularização bayesiana [6], fazendo com que a escolha do número de neurônios na camada intermediária seja menos arbitrária. A arquitetura utilizada foi inicializada com treze variáveis de entrada, uma camada intermediária com nove neurônios com função de ativação do tipo tangente hiperbólica

e a camada de saída com apenas um neurônio com função de ativação logsmóide [7]. O desempenho da rede foi avaliado por meio do percentual de acerto global de classificação, sensibilidade e especificidade do modelo. No problema aqui tratado, estas estatísticas indicam, respectivamente, o percentual de acerto de classificação da rede considerando todas as observações, apenas as de baixo *Apgar* e apenas as de alto *Apgar*.

Com a finalidade de avaliar comparativamente o desempenho, foi também aplicada a metodologia de regressão logística, considerando os mesmos treze atributos de entrada da rede neural.

Todo o procedimento de ajuste dos modelos foi efetuado utilizando-se o software Matlab.

2.3 Aprimoramentos na Rede Neural

Para identificação de atributos mais relevantes, foi utilizado o algoritmo apresentado, através de um exemplo, no Quadro 1. Esta técnica gera um rank das relevâncias das variáveis consideradas, denominadas importâncias relativas. As etapas deste algoritmo podem ser descritas da seguinte forma: Passo 1 - Listagem dos pesos da rede ao final do treinamento; Passo 2 - Cálculo das contribuições de cada atributo para a saída da rede, através de cada um dos neurônios da camada intermediária. No exemplo, a contribuição de X_1 para a saída da rede Y , através do neurônio A, é obtida pelo produto do peso da conexão entre X_1 e A com o peso da conexão de A com Y . A contribuição final de cada atributo é dada pela soma de cada uma de suas contribuições. (No caso de X_1 , $-1,80+1,62=-0,18$) e Passo 3 - Cálculo das impor-

tâncias relativas de cada atributo.

Com os resultados do Passo 2, é possível mensurar qual foi a contribuição relativa de cada atributo para a saída da rede. Os resultados do Passo 3, no exemplo considerado, sugerem que o atributo X_2 é o mais relevante e que talvez X_1 possa ser descartado do modelo, por ter importância relativa muito inferior aos demais.

2.4 Interpretação da Rede Neural

Diagramas de interpretação neural [8 e 9] foram utilizados para tentar explicitar a relação entre os atributos de entrada da rede e sua saída. Estes diagramas visam, de uma forma gráfica, mostrar as contribuições inibitórias (negativas) ou excitatórias (positivas) de cada atributo para a variável desfecho, além de apontar quais atributos estão gerando as maiores contribuições para a saída da rede. A contribuição de um atributo de entrada para a saída da rede, através de um determinado neurônio da camada intermediária, é obtida da mesma forma que foi descrita no algoritmo. No diagrama, conexões excitatórias são indicadas pela cor azul e conexões inibitórias pela cor vermelha. Além disso, quanto mais forte a conexão (maior seu valor absoluto), mais espessa será a seta referente a esta conexão. A Figura 1 apresenta o diagrama para o exemplo da rede $3 \times 2 \times 1$.

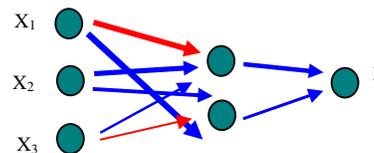


Figura 1. Diagrama de Interpretação Neural

A técnica *NeuroRule* [10] também foi utilizada neste estudo para facilitar a interpretação dos resultados. A técnica tem por objetivo criar uma árvore/regra de classificação, baseado em uma rede neural com desempenho satisfatório para explicar o comportamento do fenômeno investigado. O *NeuroRule*, no entanto, só pode ser usado em redes onde os atributos de entrada e a variável desfecho sejam categóricos, bem como a rede neural adotada não poderá ter mais de uma camada intermediária. O *NeuroRule* é aplicado aos dados do conjunto de treinamento e é composto das seguintes etapas: i) agrupa valores de ativação dos neurônios da camada intermediária da rede neural podada; ii) gera regra que explique as saídas da rede de acordo com os valores de ativação agrupados; iii) gera regras que expliquem os grupos de valores de ativação em função dos atributos de entrada e iv) concatena as duas regras obtidas em

Passo 1: Listas dos pesos da rede $3 \times 2 \times 1$		
Pesos w_{ij}		
atributos	Neurônio A	Neurônio B
X_1	-1,5	1,8
X_2	1,3	1,1
X_3	0,8	-0,6
Y: Saída da rede	1,2	0,9

Passo 2: Cálculo das contribuições			
Contribuições: w_{ij} (camada intermediária) \times w_{ij} (saída)			
atributos	Neurônio A	Neurônio B	Total
X_1	$-1,5 \times 1,2 = -1,80$	$1,8 \times 0,9 = 1,62$	0,18
X_2	1,56	0,99	2,55
X_3	0,96	- 0,54	0,42
Total	---	---	3,15

Passo 3: Cálculo das importâncias relativas	
Importância relativa: $(\text{total} / 3,15)$	
atributos	Importância relativa
X_1	$(0,18 / 3,15) = 0,0571$
X_2	0,8095
X_3	0,1334

Quadro 1. Algoritmo: exemplo em uma rede $3 \times 2 \times 1$

uma única que explique as saídas da rede em termos dos atributos de entrada.

3 Resultados

A rede inicialmente proposta (regularização bayesiana inicializada com 13x9x1) obteve, respectivamente, nos conjuntos de teste e de teste auxiliar percentual de acerto global de 67,15% e 74,70%, sensibilidade de 59,47% e 59,47% e especificidade de 74,82% e 74,76%. O algoritmo apresentado neste estudo gerou as seguintes importâncias relativas para os atributos peso: 0,307; local de nascimento: 0,301; tempo de gestação: 0,153 e IDH-renda: 0,071. Os demais nove atributos não obtiveram importância relativa superior a 0,04. Assim, a rede menos complexa (4x3x1), com os quatro atributos de entrada de maior importância relativa (que juntos totalizam importância relativa de 0,832) foi avaliada e obteve, respectivamente, nos conjunto de teste e de teste auxiliar percentual de acerto global de 66,41% e 63,62%, sensibilidade de 68,58% e 68,58% e especificidade de 64,24% e 63,60%. O diagrama de interpretação neural desta rede é observado na Figura 2.

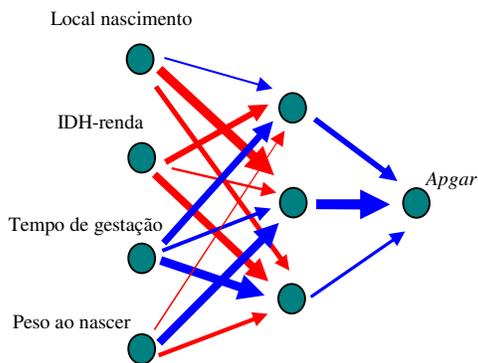


Figura 2. Diagrama de Interpretação Neural

A regra de classificação extraída da rede 4x3x1, através da técnica de NeuroRule, é:

Se o nascimento ocorrer em hospital ou em outro estabelecimento de saúde, a mãe residir em um estado com IDH-renda $\geq 0,7$, o nascimento for a termo e o bebê nascer com peso $\geq 1.500g$, então estime a vitalidade como ALTO APGAR; caso contrário, como BAIXO APGAR.

A técnica de regressão logística utilizando todos os treze atributos, listados na Tabela 2, indicou que somente a variável *dona de casa* (S/N) não era relevante (ao nível de significância de 20%) para explicar o fenômeno. Uma nova seleção sobre as doze variáveis, agora ao nível de significância de 5%, excluiu o atributo *deslocamento*. O modelo final, com as onze variáveis restantes, resultou, respectivamente,

para os conjuntos de teste e de teste auxiliar, percentual de acerto de classificação de 66,48% e 75,51%, sensibilidade de 57,65% e 57,65% e especificidade de 75,31% e 75,57%.

4 Discussão

O desempenho da rede 4x3x1, comparativamente ao modelo logístico, elevou em mais de 10% a sensibilidade, evidenciando uma real vantagem da técnica de redes neurais.

Deve ser observado que as informações sobre a gestante, utilizadas como entradas da rede, estão disponíveis logo no início da gravidez, mas outras informações referentes à gestação, ao parto e ao recém-nato, na verdade, só são de fato conhecidas ao longo ou ao término da gravidez. Para que o modelo pudesse então ser facilmente utilizado como mecanismo preditor neste problema, optou-se pelas mudanças nas categorias dos atributos existentes no banco de dados, que foram apresentadas na Tabela 2. Por exemplo, o número de consultas de pré-natal foi dicotomizado pelo critério do número mínimo (quatro) de consultas indicado pela Organização Mundial de Saúde. Assim, no início da gestação, a previsão poderia ser feita para os dois casos: em que mãe fará o mínimo de consultas e em que a mãe não fará. Se os resultados das previsões para ambos os casos fossem, respectivamente, alto *Apgar* e baixo *Apgar*, ter-se-ia uma indicação que a gestante deveria ser fortemente orientada a fazer pelo menos 4 consultas ao longo da gestação.

Apesar da classe de maior interesse na população investigada ser a menos freqüente, há de se notar que o número de observações de baixo *Apgar* (32.323) poderia ser considerado razoável para a realização de um treinamento supervisionado. Por esta razão, optou-se por utilizar a metodologia de *undersampling* em lugar da *oversampling* [5], mais comumente adotada em problemas da área de saúde, nos quais haja escassez de informações da categoria mais relevante no estudo [11, 12]. Já o conjunto de teste auxiliar foi criado no intuito de avaliar o desempenho da rede em um grupo de observações com distribuição das classes mais parecida com a população investigada, ampliando, assim, o escopo de validação da rede.

Entre as diversas técnicas que aparecem na literatura para identificação de atributos mais relevantes, podem ser citados os métodos baseados em Informação Mútua [13 e 14] e o algoritmo de Garson [8, 15 e 16]. O algoritmo de Garson, no entanto, apesar de fácil implementação, temo como sua maior crítica considerar em valor absoluto as contribuições geradas por cada atributo, através de um determinado neurônio. Com isso, a importância relativa de um

atributo na rede pode ser superestimada. O algoritmo proposto no Quadro 1, inspirado na técnica criada por Garson, tenta superar esta questão. Observa-se também que a algoritmo adotado, uma vez que identificou os atributos mais relevantes, gerou uma rede menos complexa (4x3x1) cuja sensibilidade de 68,58% foi mais de 9% superior a sensibilidade obtida pela rede 13x9x1.

Investigando a bibliografia sobre o tema que abordamos, verificamos que a sensibilidade da rede 4x3x1 foi mais de 20% superior a da rede de outro estudo [17] que possuía informações antenatais relativas à gestante e outras relacionadas à gestação e ao parto, que não estão disponíveis no SINASC.

O diagrama de interpretação neural foi utilizado neste estudo na tentativa de minimizar uma das maiores críticas a técnica de redes neurais que é a não existência de uma expressão matemática explícita indicando qual a relação entre os atributos de entrada da rede e a variável desfecho. O diagrama indicou, principalmente, que o atributo tempo de gestação gera contribuições excitatórias para a saída da rede (ou seja, quando tempo de gestação é *pré-termo*, o Apgar é *baixo*), enquanto o atributo IDH-renda fornece contribuições inibitórias (IDH-renda *alto* implica em Apgar *alto*). Além disso, observa-se que a contribuição de maior magnitude para saída da rede é alcançada (no segundo neurônio da camada intermediária) pelo atributo *local de nascimento*.

A utilização da técnica do *NeuroRule* permitiu que, através da regra de classificação gerada, qualquer avaliador do problema, tendo posse das informações de entrada para um determinado elemento ou indivíduo, pudesse identificar qual a categoria mais provável da variável de desfecho.

Agradecimentos

Este trabalho foi parcialmente financiado pelo CNPq e pela FAPERJ.

Referências Bibliográficas

- [1] National Library Medicine - National Institutes of Health (NLM/NIH) - Changing the face of medicine - biography - Dr Virginia Apgar. Disponível em http://nlm.nih.gov/changingthefaceofmedicine/physicians/biography_12.html. Acesso 17 de março 2008.
- [2] Apgar V. (1966) "The Newborn (Apgar) Scoring System" *Pediatric Clinics of North America*, v.13, p. 645-650.
- [3] SINASC – DATASUS - Disponível em http://tabnet.datasus.gov.br/tabdata/sinasc/dados/nov_indice.htm. Acesso em 18 de março. 2008.
- [4] Apgar, V. (1953) "A proposal for a new method of evaluation of the newborn infant" *Current Researches in Anesthesia and Analgesia*, v.32, n.4, p. 260-267.
- [5] Chawla, N.V (2006) "Data Mining for Imbalanced Datasets: An Overview." In: *Data Mining and Knowledge Discovery Handbook* Springer, p. 853-867.
- [6] MacKay, D.J.C. (1992) "Bayesian Interpolation" *Neural Computation*, v.4, n.3, p. 415-447
- [7] Haykin, Simon (2001), *Redes Neurais – Princípios e Prática – 2ª edição* – Bookman
- [8] Oden, J.D., Jackson, D.A. (2002) "Illuminating the "black box": a randomization approach for understanding variables contributions in artificial neural networks" *Ecological Modelling*, 154, 135-150
- [9] Özesmi, S.L., Özesmi U. (1999) "An artificial neural network approach to spatial habitat modeling with interspecific interaction" *Ecol. Model.*, 116, 15-31
- [10] Setiono, R., Lui, H. (1996) "Symbolic Representation of Neural Networks" *IEEE Comput* v.29, n.3, p.71-77
- [11] Mazurowski, M.A., Habas, P.A., Zurada, J.M., Lo, J.Y., Baker, J.A., Tourassi, G.D (2008) "Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance" *Neural Networks* v.21, n.2-3, p.427-436.
- [12] Wei, L., Yang, Y., Nishikawa, R.M., Jiang, Y. (2005) "A Study on Several Machine-learning Methods for Classification of Malignant and Benign Clustered Microcalcifications" *IEEE Transactions on Medical Imaging* v.24, n.3, p. 371-380.
- [13] Estevez, P.A., Tesmer, M., Perez, C.A., Zurada, J.M. (2009) "Normalized Mutual Information Feature Selection" *IEEE Trans. Neural Netw.*, v.20, n.2
- [14] Batiti, R. (1994) "Using mutual information for selecting features in supervised neural net learning" *IEEE Trans. Neural Netw.*, v.5, n.4
- [15] Garson, G.D. (1991) "Interpreting neural-network connection weights" *Artif. Intell. Expert* 6, 47-51
- [16] Goh, A.T.C. (1995) "Back-propagation neural networks for modeling complex systems" *Artif. Intell. Expert* 9, 143-151
- [17] Ibrahim, D., Frize, M., Walker, R. C. (2006) "Risk factors for Apgar Score using Artificial Neural Networks" *Proceedings of the 28th IEEE, EMBS Annual International Conference*, p. 6109-6112, New York City, USA, Aug 30-Sept 3, 2006