

Método De Treinamento De Redes Neurais Através De Otimização Evolutiva Com Restrições Para Auxílio Ao Diagnóstico Médico

Edmar Egidio P. de Souza, Eduardo F. de Simas Filho, Paulo C. M. A. Farias

Resumo—Neste trabalho é proposto um método de treinamento de Redes Neurais Artificiais para problemas de classificação binária. A metodologia apresentada utiliza algoritmos de otimização evolutiva que ajustam iterativamente os pesos da rede, permitindo a imposição de restrições que guiam o treinamento para satisfação de critérios em função de medidas de desempenho de classificação. O método proposto apresentou aumento de eficiência de classificação em problemas binários de detecção de doenças, considerando restrições para taxas de falsos positivos. Os resultados encontrados foram baseados em conjuntos de dados experimentais que representam problemas de diagnóstico médico.

Palavras-Chave—Classificação Binária, Detecção de Doenças, Otimização Evolutiva, Redes Neurais Artificiais.

Abstract—This work proposes an Artificial Neural Networks training method for binary classification problems. The presented methodology uses evolutionary optimization algorithms that iteratively adjust the network weights, allowing the imposition of constraints that guide the training procedure to satisfy criteria related to classification performance measures. The proposed method presented increased classification efficiency in disease detection problems, considering restrictions for false positive rates. The results were based on experimental datasets representing medical diagnostic problems.

Keywords—Artificial Neural Networks, Binary Classification, Evolutionary Optimization, Disease Detection.

I. INTRODUÇÃO

Sistemas de apoio ao diagnóstico são utilizados por profissionais da área médica como suporte no processo de identificação de doenças. A junção da experiência profissional à sistemas computacionais com boa performance de classificação, torna o diagnóstico mais ágil e preciso no tratamento de doenças [1]. Alguns problemas de detecção de doenças possuem complexidade elevada, devido à necessidade de análise de uma grande quantidade de variáveis [2]. Todavia, os avanços da inteligência artificial tem elevado a eficiência dos sistemas de apoio a decisão médica, onde classificadores especialistas são treinados para tarefas de diagnóstico de doenças específicas [3], caracterizadas por um grande número de atributos. Neste aspecto, as Redes Neurais Artificiais (RNA) [4] tem ganhado destaque pelos bons resultados obtidos na área de reconhecimento de padrões, aplicado ao apoio a medicina [5].

Edmar Egidio P. de Souza, Eduardo F. de Simas Filho, Paulo C. M. A. Farias, Laboratório de Sistemas Digitais, Programa de Pós Graduação em Engenharia Elétrica, UFBA, Salvador-BA, E-mails: edmar.egidio@cern.ch, eduardo.simas@ufba.br, paulo.farias@ufba.br

A eficiência do sistema de classificação depende da forma como o mesmo é treinado. Os algoritmos tradicionais baseados na minimização do erro quadrático médio geralmente são usados para treinar RNA, ajustando de forma iterativa os pesos das conexões neurais [4]. O algoritmo de Retropropagação (*Backpropagation*) [6] usa o gradiente de descida sobre a função de erro para minimizar o custo (erro quadrático médio). Este método pode encontrar dificuldades de convergência devido possível presença de mínimos locais, o que torna difícil ajustar os pesos para uma melhor eficiência de classificação. Outra limitação é que usando o erro quadrático médio como função custo a ser minimizada, não é possível configurar a RNA para atender condições operacionais específicas, por exemplo, melhorar o desempenho de detecção, para um dado valor limite de falso positivos especificado (restrições de treinamento).

Neste trabalho é apresentada uma nova metodologia de treinamento para classificadores neurais, utilizando algoritmos evolutivos. O sistema desenvolvido admite a inclusão de restrições durante o processo de treinamento, que são atreladas à parâmetros de avaliação de classificadores (taxas de verdadeiros positivos e falsos positivos), o que permite um aumento da eficiência dos classificadores nas faixas estabelecidas pelas restrições. A metodologia proposta foi aplicada em conjuntos de dados que representam problemas binários de diagnóstico médico [7] [8] e comparadas com métodos tradicionais de treinamento de RNA presentes na literatura.

A estrutura do trabalho está dividida do seguinte modo: a Seção II apresenta os conceitos de RNA e suas aplicações em sistemas para apoio ao diagnóstico médico. Apresenta-se na Seção III, o método proposto de treinamento de RNA, os aspectos dos algoritmos de otimização evolutiva e os conjuntos de dados utilizados na validação da metodologia. A Seção IV exhibe os resultados e comparações da proposta deste trabalho com algoritmos da literatura. A Seção V apresenta as conclusões obtidas com o estudo desenvolvido.

II. REDES NEURAI NO APOIO AO DIAGNÓSTICO MÉDICO

As RNA consistem em um conjunto de unidades de processamento interconectadas, inspiradas no funcionamento do neurônio biológico [4]. As unidades são dispostas em uma ou mais camadas e interligadas por conexões que simulam as sinapses biológicas e possuem pesos associados, que ponderam a entrada recebida por cada neurônio da rede. Em geral, sua boa capacidade de processar sinais multidimensionais,

tem motivado suas aplicações na medicina diagnóstica, envolvendo estudos como por exemplo em identificação de doenças do coração [5] e detecção de diabetes [9]. Sistemas de classificação para auxílio ao diagnóstico médico possuem a necessidade de fornecer respostas com baixas taxas de falsos positivos, pois no diagnóstico de um indivíduo classificado como saudável, sendo o mesmo portador da doença, o paciente doente é liberado sem o devido tratamento [10]. Em [5] um sistema neural foi desenvolvido para detecção binária de condições cardíacas, através de sinais obtidos por eletrocardiograma (ECG). O sistema proposto utilizou RNA profundas, e foi testado com 416 amostras de pacientes distintos, obtendo 85,52% de acurácia.

No trabalho [11] é apresentado um estudo para detecção da anemia, através do processamento de imagens microscópicas das células vermelhas do sangue humano. As imagens foram pré-processadas pela Transformada de Hough [12] e classificadas por RNA. A proposta atingiu cerca de 97,8% de acerto de classificação.

III. MÉTODO PROPOSTO

O método de treinamento proposto neste trabalho, inicializa uma população de RNA com uma única camada oculta, que estarão submetidas ao processo de otimização evolutiva. Para cada indivíduo da população é calculada a sua função aptidão (*fitness*). Em seguida, o algoritmo evolutivo realiza o ajuste de parâmetros de modo a maximizar a função aptidão, e ao mesmo tempo, satisfazer as restrições de operação do classificador, conforme o diagrama de blocos representado pela Figura 1.

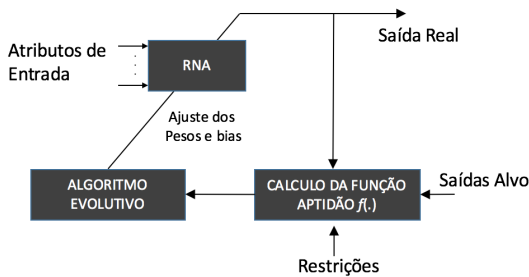


Fig. 1: Fluxo operacional da metodologia proposta.

Os pesos de cada RNA são inicializados de forma aleatória, de acordo com o método descrito em [13]. Estes valores compõem o vetor \mathbf{W} conforme ilustrado na Figura 2, que será modificado ao longo da execução do processo de otimização.

Os exemplos são selecionados através de um esquema de validação cruzada, que sorteia aleatoriamente os conjuntos de treino (40%) e validação (20%). Para a etapa de teste todo o conjunto de dados disponível foi utilizado, para obter uma maior estatística de eventos testados para comparação dos classificadores. Foram realizados 10 sorteios e para cada um deles o processo de treinamento foi reiniciado 10 vezes, totalizando assim 100 inicializações. Para o processo de otimização, dois algoritmos evolutivos diferentes foram utilizados. O Algoritmo Genético [14] a Otimização por Enxame de Partículas (*Particle Swarm Optimization* - PSO) [15].

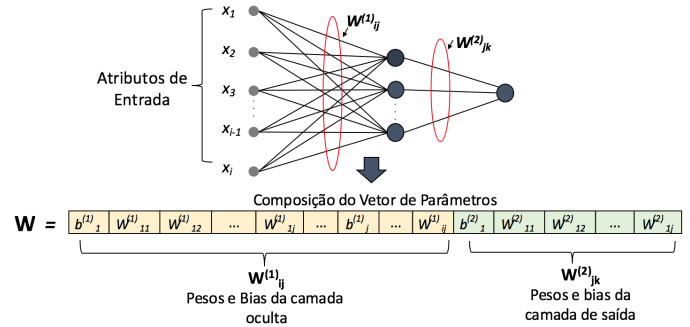


Fig. 2: Composição dos pesos da RNA nos vetores solução.

A. Problemas de Classificação Binária

Um sistema de classificação binário é um operador que realiza a escolha entre duas hipóteses: H_1 , normalmente associada à ocorrência do evento de interesse; e H_0 , que representa ocorrência de evento não relevante [16]. A probabilidade de detecção (P_D) pode ser definida como a probabilidade do classificador de identificar a hipótese positiva (H_1) corretamente:

$$P_D = P(H_1|H_1) \quad (1)$$

A probabilidade de falso alarme (P_F) está relacionada à aceitação de exemplos da hipótese negativa (H_0) à hipótese positiva (H_1):

$$P_F = P(H_1|H_0) \quad (2)$$

Neste trabalho a P_D está associada a probabilidade do classificador identificar corretamente indivíduos saudáveis. A P_F consiste na probabilidade do classificador associar um indivíduo doente como saudável. O limiar de decisão (γ) do classificador pode ser especificado através da análise da curva *ROC* (*Receiver Operating Characteristics*) [17]. Ao alterar o limiar de decisão, é possível definir o ponto de operação desejado em termos de aceitação de falsos positivos e de eficiência de detecção da classe de interesse.

B. Otimização Evolutiva

Um problema de otimização pode ser definido de acordo com a equação 3:

$$\min[f(v)|v \in \mathbf{V}] \quad \text{s.a.} \quad \begin{cases} g_1(v) \leq 0. \\ g_2(v) \leq 0. \\ \dots \\ g_k(v) \leq 0. \end{cases} \quad (3)$$

onde \mathbf{V} é o conjunto de n vetores de decisão a serem otimizados em um determinado espaço de busca e $g_k(\cdot)$ são as restrições que devem ser consideradas no problema.

A otimização evolutiva realiza o procedimento de busca de soluções para um determinado problema, inspirada no mecanismo de evolução dos seres vivos, bem como no comportamento de interação social. Em geral, tais algoritmos realizam uma busca global no espaço de variáveis. Os algoritmos evolutivos são considerados técnicas de busca meta-heurística, que não garantem o encontro da solução ideal para o problema, oferecendo soluções próximas do ótimo

através de mecanismos estocásticos e iterativos. Entretanto, apresentam vantagens em relação aos métodos de otimização convencionais, pois não necessitam da expressão da derivada da função objetivo em função dos parâmetros de entrada.

Algoritmos Genéticos

Os algoritmos genéticos (AG) são uma classe de algoritmos evolutivos que podem ser definidos como uma busca inspirada no processo biológico da evolução natural [14]. Nos AG, populações de indivíduos são criadas e submetidas a operadores genéticos, como seleção baseada em desempenho, recombinação e mutação. No treinamento de RNAs por AG, o vetor de pesos representa o cromossomo a ser recombinado no processo evolutivo, e cada peso representa o gene do indivíduo. Neste trabalho, as características do AG implementado são exibidas na Tabela I:

TABELA I: Características do AG implementado.

Codificação	Real
Taxa de Mutação	5%, 10%, 15%, 20%
Taxa de Elitismo	5%
Tamanho da População	100
Cruzamento	2pt, 4pt, 8pt, 16pt, uniforme, aleatório
Método de Seleção	Roleta Viciada/ Torneio/ Aleatório

As taxas de mutação, os operadores de seleção e cruzamento utilizados na geração, são escolhidos aleatoriamente em cada geração, com igual probabilidade de sorteio.

Otimização por Enxame de Partículas

A Otimização por Enxame de Partículas (PSO) é um método de otimização baseado no comportamento de interação social e cooperativa de grupos de espécies [15]. A PSO inicializa uma população de soluções candidatas (partículas) e move essas partículas ao redor do espaço de busca, de acordo com operações matemáticas na posição e velocidade. As partículas trocam informações sobre sua aptidão e, iterativamente, os indivíduos são orientados para a melhor solução do problema. No algoritmo PSO implementado neste trabalho, o vetor de pesos da RNA é designado como o vetor de posição, cuja velocidade é calculada iterativamente, alterando as posições para a otimização da função objetivo. As características do PSO desenvolvido são mostradas na Tabela II.

TABELA II: Características do PSO implementado.

Topologia	Global
Constante de cognição c_1	2
Constante Social c_2	1
Tamanho do Enxame	200
Constantes de Aceleração (r_1, r_2)	Aleatório [0;1]
Limites de inércia (w)	[0,4;0,9]

Função Aptidão e Restrições

Os sinais selecionados para o conjunto de treinamento (\mathbf{X}) são aplicados nos indivíduos da população. Através da equação 4 é calculada a saída da RNA y :

$$y(\mathbf{X}) = \psi_2[\mathbf{W}^{(2)}\psi_1(\mathbf{W}^{(1)} * \mathbf{X} + \mathbf{b}^{(1)}) + \mathbf{b}^{(2)}] \quad (4)$$

onde $\mathbf{W}^{(k)}$ e $\mathbf{b}^{(k)}$ são os vetores dos pesos e bias do indivíduo, respectivamente, e ψ a função de ativação tangente hiperbólica.

De posse dos valores da saída das RNA, o limiar de decisão da rede (γ) é ajustado para permitir que a restrição imposta nas condições iniciais do algoritmo seja atendida. Com o limiar de decisão definido, calcula-se a taxa de verdadeiros positivos (P_D) e então formula-se o problema de otimização, com restrições de falsos positivos P_F de acordo com a equação 5:

$$\max f = \frac{P_D(y)}{D_{KL}(y, T)} \quad s.a. \{ P_F \leq \lambda. \quad (5)$$

onde λ é a constante que determina o limite de P_F imposto ao treinamento. A $D_{KL}(y, T)$ é a divergência de Kullback-Leibler [16] entre a saída da rede (y) e as saídas alvo (T) associadas as classes, dada pela equação 6:

$$D_{KL}(y, T) = \sum_{j=1}^n y_j * \ln\left(\frac{y_j}{T_j}\right) \quad (6)$$

A medida da divergência KL é minimizada ao longo da execução do algoritmo. Esta estratégia busca um mapeamento que diminua a diferença estatística entre a entrada e a saída alvo. O procedimento do calculo da função objetivo é descrito no Algoritmo 01.

Algoritmo 1 Função Aptidão

- 1: **for** $[i] \leftarrow 1$ *to* k **do**
- 2: Calcular a saída da RNA_i (y_i) com os sinais de treinamento X ;
- 3: Ajustar limiar de decisão da RNA_i (γ) para satisfazer a restrição (P_F);
- 4: Calcular o índice P_D conforme limiar (γ) ajustado no passo 3;
- 5: Calcular a divergência KL ($D_{KL}(y_i, T)$), entre a saída da RNA (y_i) e o alvo (T);
- 6: Atribuir: $fitness = \frac{P_D(y_i)}{D_{KL}(y_i, T)}$;
- 7: $[i] \leftarrow i + 1$;
- 8: **end for**

As simulações foram realizadas com o *software* Matlab R2015b, em um PC com processador Intel Core i7-3770 CPUx8, 8 GB de RAM.

C. Conjuntos de Dados

Bases de dados públicas para aprendizado de máquina, obtidas em [18], foram utilizadas para avaliar o método proposto. Os conjuntos de dados selecionados representam problemas de classificação binária de detecção de doenças. O conjunto *Pima Diabetes* [7] apresenta exemplos de medidas biológicas corporais utilizadas para o diagnostico de diabetes. Variáveis como pressão arterial, nível de massa corporal e concentração de glicose foram obtidos através de exames realizados com indivíduos da população indígena Pima, localizada na cidade de Phoenix, Arizona, EUA.

No conjunto *Parkinson Speech* [8] são apresentados exemplos de atributos extraídos do sinal de voz de indivíduos saudáveis e de pacientes com o mal de Parkinson. Atributos presentes no conjunto como, nível de pico normalizado, número de períodos e número de quebras de voz com textos padrões, apresentam características que podem ser utilizadas para detecção da doença [8]. Na Tabela III algumas características dos conjuntos usados são exibidas, e seus detalhes podem ser consultados em [18].

TABELA III: Características dos conjuntos utilizados.

Conjuntos	Atributos	Exemplos Saudáveis	Exemplos Doentes
<i>Pima Diabetes</i>	8	268	500
<i>Parkinson Speech</i>	28	520	520

IV. RESULTADOS

A. Conjunto de Dados - *Pima Diabetes*

Inicialmente, para a definição do número de neurônios ocultos das RNA, foi adotado um procedimento iterativo com a realização de treinamentos sucessivos, variando-se a quantidade de neurônios ocultos e mantendo um único neurônio na saída. A Figura 3 apresenta o *boxplot* da P_D em função do número de neurônios ocultos, para uma taxa de $P_F = 1\%$. Os algoritmos AG e PSO, treinados com restrição nesta taxa de falso alarme, obtiveram melhor resultado com 16 e 04 neurônios na camada oculta, respectivamente. O algoritmo BP_{Rprop} obteve melhor resultado com 20 neurônios ocultos, enquanto que o BP_{GD} alcançou melhor desempenho com 08 neurônios na camada oculta. Para toda variação de neurônios ocultos estudados, nota-se um melhor desempenho dos algoritmos evolutivos, em termos de P_D para a faixa de operação de falso alarme em 1%.

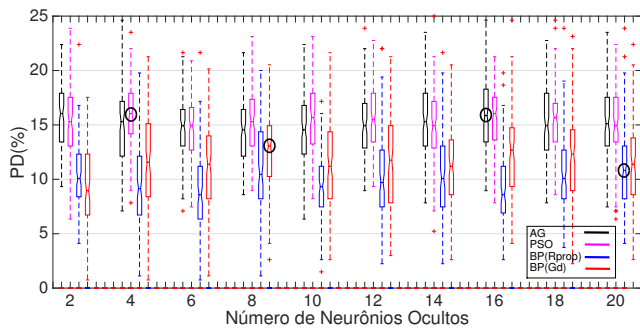


Fig. 3: *Boxplot* da P_D dos classificadores treinados por AG e PSO em função do número de neurônios ocultos.

A Tabela IV apresenta os resultados comparativos entre a metodologia proposta e os algoritmos do tipo *backpropagation*. Com os algoritmos evolutivos, foram realizadas análises do treinamento com três restrições distintas de P_F , de modo à verificar a viabilidade da proposta em aumentar a eficiência de detecção para as restrições associadas. Para $P_F = 1\%$, o AG implementado obteve superioridade para a P_D em cerca de 5 pontos percentuais em relação ao BP_{GD} . Nos pontos de operação $P_F = 2\%$ e 3% , o treinamento através do PSO, apresentou leve superioridade aos resultados encontrados pelo

AG, com redução no tempo de treinamento (tempo médio de treinamento de uma inicialização do algoritmo).

TABELA IV: Comparação das taxas de verdadeiros positivos $P_D(\%)$ - Conjunto *Pima Diabetes*.

	$PD_{PF=1\%}$	$PD_{PF=2\%}$	$PD_{PF=3\%}$	Tempo de Trein.(s)	Épocas
AG	15,85±3,25	24,40±3,40	32,74±3,34	22,05±1,12	200
PSO	15,87±2,88	24,80±3,18	33,43±3,29	5,39±0,14	200
$BP_{(GD)}$	11,84±4,28	19,86±3,13	27,41±3,30	3,65 ± 0,23	800
$BP_{(Rprop)}$	10,87±3,98	16,34±4,10	23,36±5,15	2,87 ± 0,74	600

O PSO obteve cerca de 33% de acerto de indivíduos saudáveis, admitindo apenas 3% de falsos positivos, superando em cerca de 6 pontos percentuais ao resultado obtido pelo BP_{GD} para a mesma taxa de falsos positivos. Na Figura 4 são mostradas as curvas ROC dos classificadores projetados com melhor desempenho de P_D na etapa de teste. As curvas indicam que para as faixas de P_F , onde as restrições foram configuradas, a eficiência de detecção, é superior aos classificadores treinados pelo *backpropagation*.

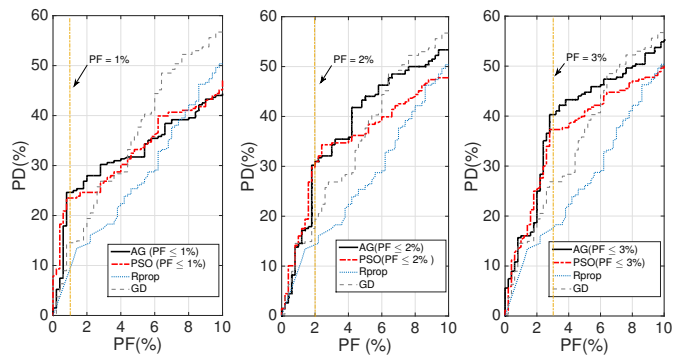


Fig. 4: Curvas ROC dos melhores classificadores treinados.

Na Figura 5, visualiza-se que nas saídas das redes treinadas pelos algoritmos evolutivos, há uma concentração maior de eventos que representam indivíduos saudáveis em relação ao BP, após o limiar de decisão que admite o limite de $P_F = 3\%$. Esta taxa foi utilizada como restrição, indicando que houve uma melhoria dos acertos da classe positiva com esse limite de falsos positivos.

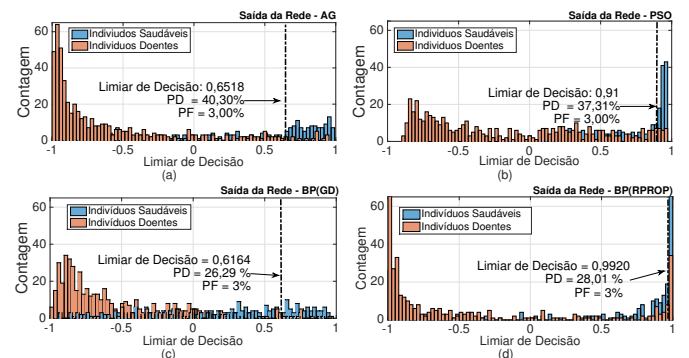


Fig. 5: Histograma de saída das RNA - (a) Trein. com AG, (b) Trein. com PSO, (c) Trein. com BP_{GD} e (d) Trein. com BP_{RPROP} .

B. Conjunto de Dados - Parkinson Disease

Neste conjunto, um procedimento semelhante ao exibido na Figura 3 foi realizado para definição do número de neurônios ocultos. Para os AG e PSO, os resultados indicaram uma melhor configuração utilizando 18 neurônios ocultos, enquanto que os algoritmos BP_{GD} e BP_{Rprop} obtiveram melhor resultado com 14 e 12 neurônios ocultos, respectivamente. A Tabela V apresenta os resultados da P_D com os demais pontos de operação em P_F analisados. Os resultados indicam um melhor desempenho através do treinamento com restrições pelo AG, apresentando aumento de cerca de 1,5 pontos percentuais de acerto de classificação da classe positiva (indivíduo saudável) em relação ao BP_{Rprop} , admitindo 1% de P_F .

TABELA V: Comparação das taxas de verdadeiros positivos - Conjunto Parkinson Disease.

	$PD(P_F=1\%)$	$PD(P_F=2\%)$	$PD(P_F=3\%)$	Tempo de Trein. (s)	Épocas
AG	6,94±2,52	12,58±2,64	14,74±3,34	82,75±1,02	200
PSO	5,12±1,88	12,30±3,18	13,43±3,29	45,89±0,14	200
$BP_{(GD)}$	4,21±2,53	9,16±3,13	10,24±2,30	33,85 ± 0,23	1000
$BP_{(Rprop)}$	5.34±4,61	10,65±4,10	11,36±3,15	27,82 ± 1,24	750

Com a restrição de 3% de P_F imposta aos algoritmos evolutivos, os resultados indicam um aumento de cerca de 3 pontos percentuais na eficiência de classificação utilizando o AG, em relação ao algoritmo BP_{Rprop} . O treinamento via PSO com restrições obteve resultados levemente inferiores ao AG, com menor tempo de treinamento. Na Figura 6 são exibidas as curvas ROC dos classificadores que tiveram melhor desempenho durante a fase de teste. Verifica-se que nas faixas de falso alarme, configuradas como restrições impostas ao treinamento pelos algoritmos AG e PSO, há um melhor desempenho em relação ao BP. Nas três restrições utilizadas, visualiza-se um melhor comportamento dos classificadores treinados com o AG e PSO com restrições em termos de aprovação de eventos da classe positiva.

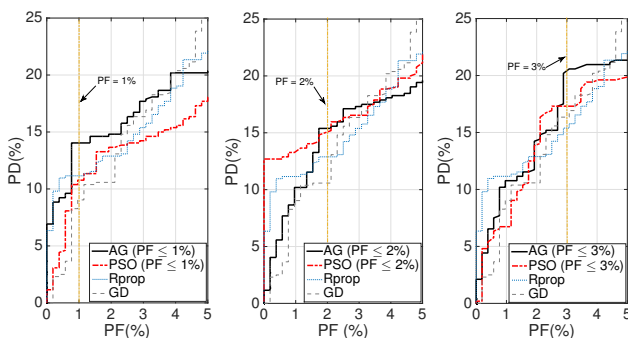


Fig. 6: Curvas ROC dos classificadores - Parkinson Speech.

V. CONCLUSÕES

Sistemas inteligentes de classificação são ferramentas importantes para apoio ao diagnóstico de doenças. O método de treinamento de classificadores neurais proposto neste trabalho, utilizando algoritmos evolutivos (AG e PSO) projetados para trabalharem com restrições de limites de aprovação de falsos positivos, permitiu guiar o treinamento para uma melhor

eficiência de classificação nos pontos de operação selecionados como restrições. Esta estratégia apresentou resultados superiores em termos de P_D em relação aos algoritmos tradicionais da literatura baseado em descida de gradiente, o BP_{GD} e BP_{Rprop} , para as bases de dados consideradas. Aplicando o método proposto na base de dados pública *Pima Indian Diabetes*, foi verificado um aumento na eficiência de detecção de indivíduos saudáveis em cerca de 5 pontos percentuais em relação ao algoritmo da classe BP que obteve melhor desempenho. A metodologia de treinamento replicada à base de dados *Parkinson Disease*, apresentou também uma melhoria em termos de eficiência de detecção, quando comparadas aos algoritmos BP, nos mesmos pontos de operação de P_F . O método proposto alcançou cerca de 1,5 pontos percentuais de melhoria no acerto de classificação de indivíduos saudáveis para $P_F = 1\%$, em comparação ao BP. Os algoritmos evolutivos obtiveram maior tempo de processamento em relação ao *backpropagation*, neste aspecto o PSO apresentou melhor desempenho em relação ao AG projetado, encontrando valores de P_D semelhantes.

REFERÊNCIAS

- [1] D. R. Nayak, R. Dash, B. Majhi, and V. Prasad, "Automated pathological brain detection system: A fast discrete curvelet transform and probabilistic neural network based approach," *Expert Systems With Applications*, pp. 152–164, Julho 2017.
- [2] C.-H. Weng, T. C.-K. Huang, and R.-P. Han, "Disease prediction with different types of neural network classifiers," *Telematics and Informatics*, pp. 277–292, Agosto 2015.
- [3] H. Elsalamony, "Anaemia cells detection based on shape signature using neural networks," *Measurement*, pp. 50–59, Julho 2017.
- [4] S. Haykin, *Neural Networks: Principles and Practice*, 2008.
- [5] V. J. R. Ripoll, A. Wojdel, E. Romero, P. Ramos, and J. Brugada, "Ecg assessment based on neural networks with pretraining," *Applied Soft Computing*, pp. 399–406, Agosto 2016.
- [6] E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Letters to Nature*, pp. 533–536, 1986.
- [7] W. C. Knowler, P. H. Bennett, R. F. Hamman, and M. Miller, "Diabetes incidence and prevalence in pima indians: a 19-fold greater incidence than in rochester, minnesota," *American Journal of Epidemiology*, pp. 497–505, 1978.
- [8] E. Sakar, B. Isenkul, and M. Sakar, "Collection and analysis of a parkinson speech dataset with multiple types of sound recordings," *Journal of Biomedical and Health Informatics*, pp. 828–834, 2013.
- [9] O. Erkamaz, M. Ozer, and M. Perc, "Performance of small-world feed-forward neural networks for diagnosis of diabetes," *Applied Mathematics and Computation*, pp. 22–28, Julho 2017.
- [10] X.-H. Zhou, N. A. Obuchowski, and D. K. McClish, *Statistical Methods in Diagnostic Medicine*. New York: John Wiley and Sons, 2002.
- [11] H. Elsalamony, "Healthy and unhealthy red blood cell detection in human blood smears using neural network," *Micron*, pp. 32–41, Fevereiro 2016.
- [12] R. C. Gonzalez and R. E. Woods, *Processamento Digital de Imagens*. Sao Paulo: Pearson Prentice Hall, 2010.
- [13] D. Nguyen and B. Widrow, "Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights," *International Joint Conference on Neural Networks*, vol. 24, pp. 21–26 vol.3, 1990.
- [14] J. H. Holland, *Adaptation in Natural and Artificial Systems*. The MIT Press, 1992, vol. 2.
- [15] J. Kennedy and R. Eberhart, "Particle swarm optimization," *Proceedings of the IEEE International Conference on Neural Networks*, pp. 1942–1945, 1995.
- [16] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. John Wiley and Sons, 2004, vol. 1.
- [17] C. Marzban, "The roc curve and the area under it as performance measures," *Weather and Forecasting*, vol. 24, pp. 1106–1114, 2004.
- [18] D. Dheeru and E. Karra Taniskidou, "UCI machine learning repository," 2017. [Online]. Available: <http://archive.ics.uci.edu/ml>