

# Estudo experimental do efeito de parâmetros de projeto de redes MLP na produção de comitês para o suporte ao diagnóstico de tuberculose

Errison dos Santos Alves <sup>\*</sup>, João B. O. Souza Filho <sup>†\*</sup>, Rafael Mello Galliez <sup>‡</sup>, Afrânio Kritski <sup>§</sup>

<sup>\*</sup> Programa de Pós-Graduação em Engenharia Elétrica (PPEEL),

<sup>\*</sup> Centro Federal de Educação Tecnológica Celso Suckow da Fonseca (CEFET-RJ), RJ, Brasil

<sup>†</sup> Departamento de Engenharia Eletrônica e Computação (DEL/POLI),

<sup>‡</sup> <sup>§</sup> Programa Acadêmico de Tuberculose, Faculdade de Medicina,

<sup>†</sup> <sup>‡</sup> <sup>§</sup> Universidade Federal do Rio de Janeiro, RJ, Brasil

Email: <sup>\*</sup> errison\_alves@yahoo.com.br, <sup>†</sup> joao.filho@poli.ufrj.br, <sup>‡</sup> galliez77@gmail.com, <sup>§</sup> kritskia@gmail.com

**Resumo**—A tuberculose (TB) é uma das principais doenças que acomete a sociedade, principalmente em países subdesenvolvidos. Com o objetivo de reduzir o tempo de diagnóstico, e otimizar o uso dos recursos hospitalares, este trabalho propõe um estudo experimental para o desenvolvimento de comitês visando o suporte à decisão quanto ao isolamento respiratório de pacientes suspeitos de tuberculose pulmonar. A influência dos parâmetros de projeto das redes MLP na diversidade, e seus efeitos na acurácia dos modelos são analisados. Os resultados mostraram um desempenho superior das técnicas de comitê, que atingem um valor mediano da AUC superior a 0.88, comparado a um valor de 0.87 dos modelos individuais.

**Index Terms**—Sistemas de Suporte à Decisão, Métodos de Comitê, Redes Neurais Artificiais, Diagnóstico de Tuberculose.

## I. INTRODUÇÃO

A tuberculose é uma doença infectocontagiosa causada pelo bacilo *Mycobacterium tuberculosis*. Transmitida por via aérea, ela principalmente afeta os pulmões [1]. O Brasil se encontra entre os 22 países de maior incidência de TB, apresentando, em 2013, cerca de 80.000 casos notificados, e uma mortalidade anual de 4.000 casos, aproximadamente [2].

O diagnóstico clássico desta doença é realizado com base nos sinais e sintomas do paciente, e por meio de testes clínicos, cujos principais são a baciloscopia e a cultura. O primeiro apresenta baixa sensibilidade, isto é, uma baixa capacidade de diagnosticar pacientes doentes. A cultura é um teste demorado e de disponibilidade mais restrita, levando cerca de 6 a 8 semanas para ser concluído. Tais fatos tem motivado a proposta de novos métodos diagnósticos [3].

O diagnóstico rápido e eficiente de TBP é um grande desafio, principalmente no Sistema Único de Saúde (SUS). Os Sistemas de Suporte à Decisão (SSD) baseados em Inteligência Computacional podem constituir ferramentas muito úteis para esta tarefa, em especial aqueles que envolvem modelos matemáticos utilizando redes neurais artificiais (RNAs).

Entre algumas propostas de SSD para TB, em [4], há uma comparação do desempenho das técnicas MLP e ELM no desenvolvimento de modelos de classificação, com resultados superiores da primeira. Em [5], foi avaliado o desenvolvimento de um escore de pontos, por meio de regressão linear. Em ambos, foram considerados pacientes ambulatoriais, e não internados e candidatos ao isolamento, conforme o foco deste artigo.

No complexo hospitalar do Instituto de Doenças do Tórax do Hospital Universitário Clementino Fraga Filho (IDT/HUCFF), de acordo com o Programa de Tuberculose Hospitalar (PCTH), cerca de 25% dos pacientes alocados nos quartos de isolamento respiratório, de fato, apresentaram TB ativa [6]. Este fato sinaliza a necessidade de outros métodos e ferramentas que auxiliem um isolamento mais criterioso e eficiente. Assim, os SSDs podem auxiliar a tomada de decisão médica, reduzir o número de isolamentos desnecessários, otimizar a utilização dos recursos hospitalares, bem como reduzir os custos operacionais [6].

Neste trabalho é discutido o desenvolvimento de um sistema preditivo para o diagnóstico de TB pulmonar, empregando-se comitês de redes neurais, no contexto do isolamento hospitalar. Este sistema é um esforço para se obter uma maior acurácia, quanto ao isolamento, considerando-se informações e realidade nacionais, o qual mimetiza, em parte, a prática médica.

Métodos de comitê, geralmente apresentam uma maior capacidade de generalização em problemas complexos do que modelos individuais [7], caso formados através da seleção de diversos modelos que, quando integrados, apresentam uma performance superior aos modelos isolados. Estes modelos foram desenvolvidos utilizando redes *Multilayer Perceptron* (MLP) [8], devido a sua elevada capacidade de generalização, bem como sensibilidade com relação à alteração dos seus parâmetros de projeto, possibilitando, assim, um fácil desenvolvimento de redes diversas. A seleção dos modelos membros do comitê é realizada por meio de índices de divergência,

avaliando-se a complementariedade, logo, busca selecionar modelos que, quando integrados, resultem num sistema de maior eficácia.

A estrutura do artigo é a seguinte: primeiramente, é discutida a construção de modelos baseados em métodos de comitê. Em seguida, a base de dados é descrita; e, por fim, os resultados são apresentados. Por último, apresentam-se as conclusões e os trabalhos futuros.

## II. MÉTODOS DE COMITÊ

A técnica de comitê possui inspiração na teoria da "Sabedoria das Multidões" [9], que valoriza o conhecimento individual, sem qualquer influência externa, na construção de uma decisão coletiva, realizada através de um mecanismo de integração das decisões individuais. Na área médica, faz-se a analogia com um comitê de especialistas, com opiniões individuais, o qual deriva conclusões conjuntas sobre algum procedimento, em geral, de alta gravidade ou complexidade, a ser realizado em um paciente.

Métodos de comitê costumam apresentar diversas vantagens com relação às técnicas baseadas em um único modelo, dentre as quais: (i) um comitê ótimo costuma apresentar uma melhor generalização comparado a modelos simples; (ii) comitês podem ser otimizados através do acréscimo ou retirada de modelos; (iii) diferentes formas de integração de saídas dos membros podem ser consideradas, de acordo com as performances individuais ou escopo de aplicação; (iv) uma maior robustez do sistema pode ser alcançada, caso os modelos membros atuem de forma complementar.

Para um comitê ser eficiente, isto é, apresentar uma capacidade de generalização superior a qualquer modelo individual, é necessário que seus membros apresentem erros independentes [10]. Isto significa que tais modelos precisam apresentar diversidade entre as suas saídas. Assim, a diversidade é um importante indicador a ser considerado no projeto de um comitê, onde a seleção de modelos diversos tende a resultar numa melhor acurácia [11].

### A. Diversidade

Para se desenvolver modelos classificatórios complementares, em [7], foram propostas taxonomias para a manipulação de parâmetros construtivos, que visam gerar diversidade entre os modelos:

- **Manipulação do modelo**

Este método gera diversidade através da manipulação dos parâmetros construtivos do modelo classificatório. Para as MLPs, tal efeito pode ser produzido pela variação dos pesos iniciais, da topologia e do algoritmo de treinamento.

- **Manipulação dos dados**

Neste método, cada modelo é treinado sob uma variação ou subconjunto dos dados disponível no seu desenvolvimento. Geralmente, são utilizadas técnicas de reamostragem, tais como *bagging*, *boosting*, *hold-out*, validação

cruzada, entre outras [12]. Utiliza-se, também, o particionamento dos dados através de técnicas de agrupamento (*clustering*).

### B. Seleção

Existem diversos métodos para se determinar o tamanho do comitê, bem como para selecionar os seus membros [7]. Desta forma, reduz-se o custo computacional e, principalmente, selecionam-se apenas modelos que resultem num incremento da acurácia.

A parte mais crítica desta seleção é a definição dos modelos que irão compor o comitê. Supondo modelos neurais previamente formados, faz-se necessário, portanto, encontrar aqueles que sejam complementares. Geralmente, assume-se que o aumento da diversidade pode reduzir o erro associado ao comitê [13]. Porém, por não existir uma relação direta entre a diversidade, medida por algum índice, e o melhor desempenho, esta questão permanece como um dilema nesta área [14].

Dentre os índices mais utilizados para a avaliação da diversidade, destacam-se os que relacionam pares de modelos, por serem mais simples e intuitivos. A Tabela I representa uma tabela de contingência associada a 2 modelos ( $C_i$  e  $C_j$ ) genéricos, onde  $N^{ij}$  é o quantitativo de eventos observado em cada caso.

Tabela I: Tabela de contingência

	$C_j$ acerto (1)	$C_j$ erro (0)
$C_i$ acerto (1)	$N^{11}$	$N^{10}$
$C_i$ erro (0)	$N^{01}$	$N^{00}$
$N = N^{00} + N^{01} + N^{10} + N^{11}$		

Dentre os índices mais aplicados na área de métodos de comitê, destacam-se:

- **The Disagreement Measure**

Esta medida, utilizada em [15] para se caracterizar a diversidade entre um classificador base e um complementar, é descrita pela soma dos eventos discordantes entre  $C_i$  e  $C_j$  ( $N^{01} + N^{10}$ ) sobre o total de eventos  $N$ .

$$Dis_{i,j} = \frac{N^{01} + N^{10}}{N^{00} + N^{01} + N^{10} + N^{11}} \quad (1)$$

- **The Double-Fault Measure**

Uma medida que infere a quantidade de erros simultâneos de  $C_i$  e  $C_j$  ( $N^{00}$ ) sobre o quantitativo total de  $N$ . Foi empregada em [16], para a seleção de classificadores menos correlacionados.

$$Df_{i,j} = \frac{N^{00}}{N^{00} + N^{01} + N^{10} + N^{11}} \quad (2)$$

O processo de seleção, realizado de maneira incremental, busca, inicialmente, por 2 modelos que apresentem a maior divergência entre todos os demais, de acordo com a avaliação do índice definido. A cada nova etapa, buscam-se outros modelos que sejam os mais divergentes com relação àqueles integrados anteriormente. Este procedimento é repetido até que se alcance o número desejado de modelos (membros) no comitê.

### C. Integração

Selecionados os membros do comitê, faz-se necessário integrar suas saídas, a fim de definir a decisão global do sistema. A combinação dos membros pode ocorrer de diversas formas [7], porém os métodos mais populares são a votação majoritária e a média. Na primeira, a decisão é avaliada de acordo com a classe que obtêm o maior número de votos. Já na seguinte, calcula-se a média das saídas dos membros do comitê e, com base no resultado, realiza-se a decisão quanto à classe final. Um método similar é a média ponderada, para a qual se pode definir, ainda, alguma figura de mérito associada ao desempenho dos modelos.

Para mantermos a informação sobre a distribuição das saídas, possibilitando um ajuste posterior do limiar de decisão, este trabalho se concentrou na integração por média.

## III. TRATAMENTO DOS DADOS

A base de dados abrange o período de 2001 a 2008, e contém 972 eventos (pacientes), divididos por desfecho, da seguinte maneira: 210 pacientes doentes (TB+) e 762 pacientes não-doentes (TB-). Para cada paciente, há informações gerais, e 36 variáveis com informações clínicas e sintomáticas. Todas estas variáveis, excluindo a idade, são dicotômicas, indicando sua presença, ausência ou o desconhecimento sobre o dado.

### A. Seleção de variáveis

As variáveis de entrada utilizadas no desenvolvimento dos modelos neurais foram selecionadas utilizando a técnica estatística de Regressão Logística Univariada [17], e empregando a estratégia "wrapper" [18]. Modelos logísticos são comumente utilizados na área médica para relacionar variáveis categóricas (neste caso, binárias) com o desfecho, através de um modelo não-linear simples derivado pelo método da Máxima Verossimilhança [17].

Considerando a estratégia "wrapper", adotou-se o procedimento iterativo de busca "forward floating search" [18], onde cada variável é inserida por vez no modelo, de acordo com o seu nível de significância, definido através de testes estatísticos.

A cada passo, as variáveis ainda não presentes no modelo são avaliadas quanto à inclusão, e aquelas que apresentam um menor valor de  $p$  (maior significância) são incluídas. Após a inclusão de uma variável, avalia-se a exclusão de alguma delas no modelo, neste caso, considerando-se o maior valor de  $p$ , isto é, a menor significância. O procedimento continua até que nenhuma variável seja incluída ou retirada do modelo.

Para tal procedimento, foi utilizado o *software Statistical Package for Social Sciences* (SPSS) [19], adotando-se um nível de significância para a inclusão e exclusão de 25% e 30%, respectivamente.

Ao final deste processo, dentre as 36 variáveis disponíveis no banco, 19 foram definidas como relevantes: *idade, sexo, escarro, hemoptise, hemoptóico, sudorese noturna, dispnéia, febre, desnutrição, laudo de RX (TB ativa, sequela, outra doença), tosse, dor torácica, HIV, alcoolismo, neoplasia maligna, tabagismo e dor de garganta.*

### B. Mineração de dados

Diversos procedimentos de mineração foram realizados no banco de dados, visando identificar e eliminar eventos (pacientes) que pudessem comprometer o aprendizado dos modelos neurais. Inicialmente, foram excluídos pacientes com sinais e sintomas inconsistentes e ausentes, e quando em quantidade maior que 25% do total das variáveis selecionadas.

Adicionalmente, retiraram-se os pacientes com desfechos diferentes, isto é, rotulados como TB+ e TB-, que possuíam valores idênticos em suas variáveis descritivas, com exceção da idade. Estes pacientes foram avaliados por especialistas como casos atípicos que exigem uma investigação médica mais detalhada, não se aplicando, portanto, ao escopo do sistema proposto.

Após a limpeza do banco, com o objetivo de produzir melhores conjuntos de treino, validação e teste, os pacientes TB+ e TB- foram agrupados de acordo com os critérios SAFE e DANGER [20], [21], que classificam os eventos de acordo com a sua vizinhança. De acordo com estes critérios, um evento é classificado com DANGER se um dado número de eventos vizinhos pertencerem a uma classe distinta da sua; e como SAFE, caso contrário. Estas estratégias são comumente utilizadas para se identificar erros quanto ao rótulo da classe, ou para eventos críticos, isto é, que, de fato, pertencem a classe de seu rótulo, mas que apresentam uma elevada similaridade com a outra classe.

Ao final do processo de mineração dos dados, os 818 eventos resultantes do banco foram utilizados para o desenvolvimento dos modelos MLP. A Tabela II apresenta a divisão destes eventos segundo os critérios SAFE e DANGER citados.

Tabela II: Quantitativos de eventos SAFE e DANGER

Eventos		SAFE	DANGER
TB+	161	40 (24,8%)	121 (75,2%)
TB-	657	640 (97,4%)	17 (2,6%)
TOTAL	818	680	138

## IV. RESULTADOS

Os modelos consideraram apenas redes MLP, possuindo uma camada intermediária, um neurônio na camada de saída e tangente hiperbólica como função de ativação. A diversidade foi gerada através da manipulação dos modelos, variando-se o número de neurônios na camada intermediária, as inicializações e os algoritmos de treinamento empregados.

As 19 variáveis de entrada dos 818 eventos resultantes foram codificadas como: +1, quando da presença do sinal ou sintoma; -1, na sua ausência; e 0, quando a informação é inexistente. A idade foi normalizada para valores na faixa entre -1 e 1. A presença de tuberculose foi codificada como +1; e a ausência, como -1.

Diante da complexidade do problema face ao desbalanço entre as classes, os conjuntos de desenvolvimento e teste foram formados a partir da classificação dos eventos em SAFE e DANGER. Inicialmente, foram realizadas partições em cada uma das seguintes subclassificações: TB+ SAFE, TB+

DANGER e TB- SAFE, desconsiderando-se os eventos TB-DANGER, por se tratarem de pacientes de difícil diagnóstico, segundo especialistas. Este procedimento foi seguido de alguns ensaios experimentais, visando a definição das respectivas quantidades adequadas à formação dos conjuntos de desenvolvimento e teste.

O conjunto de validação é obtido através de uma subamostragem do conjunto de desenvolvimento, realizada por uma validação cruzada (4  *folds* ) nos eventos SAFE e DANGER deste conjunto. Dessa forma, 1  *fold*  é utilizado para o conjunto de validação, enquanto os  *folds*  restantes definem o conjunto de treinamento. Este processo também insere uma diversidade oriunda da manipulação dos dados, que contribui para o estudo proposto.

O processo de partição dos eventos para os conjuntos de desenvolvimento e teste foi repetido 100 vezes, onde, de acordo com o procedimento citado, para cada conjunto de teste, existem 4 conjuntos de treinamento e validação associados. A Tabela III apresenta a quantidade aproximada de eventos nos conjuntos de treinamento, validação e teste.

Tabela III: Número aproximando de eventos nos conjuntos de treino, validação e teste

	treino	validação	teste
TB+	120	30	40
TB-	360	90	160

Para cada um dos 400 conjuntos resultantes, foram desenvolvidos modelos neurais possuindo de 1 a 15 (passo de 2) neurônios na camada intermediária, segundo 3 algoritmos de treinamento: Resilient Backpropagation (RPROP) [22], Levenberg-Marquardt (LM) [23], Broyden-Fletcher-Goldfarb-Shanno (BFGS) [24]. Para cada caso, foram produzidos 25 modelos com diferentes inicializações. Como resultado final, um total de 2400 modelos foram produzidos para cada conjunto de teste, empregando-se o critério de parada antecipada para mitigar o efeito de  *overtraining*  [8].

### A. Seleção e integração dos modelos

Com os 2400 modelos neurais desenvolvidos, selecionaram-se os membros do comitê de acordo com 2 critérios: análise da divergência e de forma aleatória. Dessa forma, além de se avaliar o desempenho do método como um todo, pode-se observar a eficácia do emprego dos índices de divergência na formação de comitês.

Para o critério de divergência, foram utilizados os 2 índices citados anteriormente ( *Disagreement Measure [Dis]* e  *Double-Fault Measure [Df]*), realizando-se a seleção por meio de sua avaliação, considerando os conjunto de validação. Em todos os casos, as saídas foram integradas pela média.

### B. Análise dos efeitos dos parâmetros de projeto

Inicialmente, o efeito dos parâmetros de projeto na acurácia do sistema considerou cada parâmetro de forma individual, fixando-se, portanto, os demais. Neste caso, considerou-se a

variação do número de neurônios na camada intermediária, as inicializações e os algoritmos de treinamento.

A avaliação foi realizada com base na área sob a curva ROC (AUC) [25], que relaciona a sensibilidade e a especificidade de um modelo para variados valores quanto ao limiar de decisão.

A primeira análise considerou o efeito da inicialização. Para isto, foi fixado o algoritmo de treinamento (LM), e o número de neurônios como 15, parâmetros sintonizados por validação cruzada para os classificadores individuais de maior desempenho.

A Figura 1 apresenta o diagrama de caixa ( *Boxplot* ) para os 400 conjuntos de treino e teste, considerando de 2 a 10 membros para o comitê. O conjunto de validação mostrou o mesmo comportamento daquele de treinamento, e, portanto, não será avaliado.

Nota-se uma melhoria no desempenho dos sistemas de comitê em relação aos melhores modelos individuais no conjunto de treino, principalmente, considerando os índices  *Dis*  e  *Df* . O mesmo ganho, porém, não se mostrou de forma tão evidente no conjunto de teste.

Independente do critério de seleção dos modelos, pode-se obter uma AUC mediana maior que 0.88, considerando-se um comitê com 4 membros. Ainda assim, os valores são superiores aos dos melhores modelos individuais, que obtêm uma AUC mediana de 0.87. Verifica-se, ainda, uma tendência de comportamento similar entre os critérios de seleção com o acréscimo de membros ao comitê, em especial para o conjunto de teste.

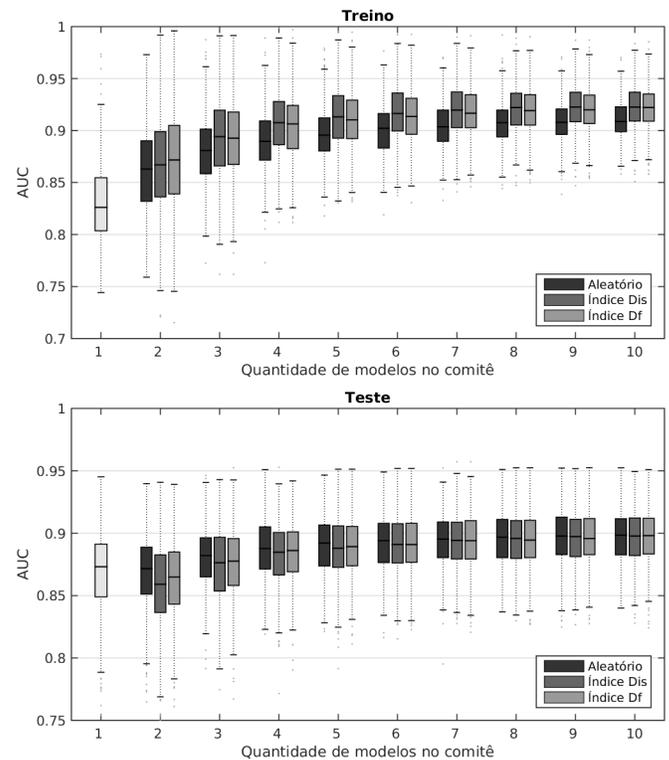


Figura 1: Boxplot do valor da AUC para os conjuntos de treino e teste relativo à influência das inicializações

No que tange a escolha do algoritmo de treinamento, adotou-se a melhor inicialização e 15 neurônios. Nesta análise, o comitê pode conter 3 modelos, devido à utilização de 3 algoritmos de treinamento distintos. Na Figura 2 são apresentados os resultados. Verifica-se, novamente, um desempenho semelhante dos critérios de seleção tanto para o conjunto de treino quanto para o teste, sendo o comitê ligeiramente melhor que o classificador individual.

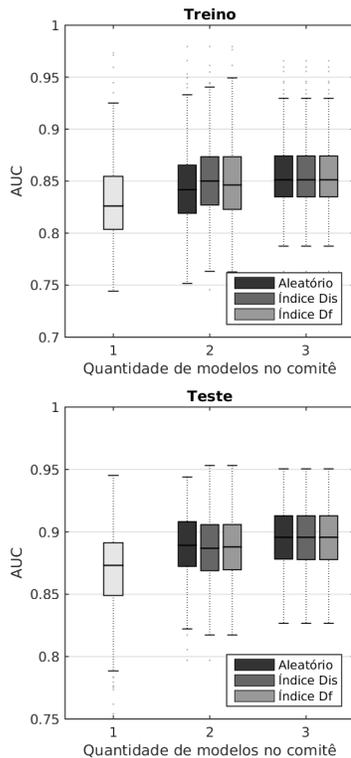


Figura 2: Boxplot do valor da AUC para os conjuntos de treino e teste relativo a influência dos algoritmos de treinamento

A última análise individual considerou a variação da quantidade de neurônios. Neste caso, o comitê pode conter 8 modelos, já que a quantidade de neurônios considerada no desenvolvimento varia de 1 a 15 com um passo de 2. A Figura 3, de forma semelhante às análises anteriores, apresenta um valor mediano para a AUC considerando o conjunto de teste superior a 0.89, considerando-se um comitê com 4 membros, para os 3 critérios de seleção.

### C. Análise global do sistema comitê

Nesta análise serão considerados, conjuntamente, a variação de todos os parâmetros de projeto. Os resultados são resumidos na Figura 4. É possível notar uma coerência com os resultados anteriores.

Entretanto, em nenhuma das análises realizadas, o melhor desempenho obtido, através da seleção pelos dos índices de divergência, para os conjuntos de treinamento foi refletido para os conjuntos de teste. Esta situação pode estar relacionada à

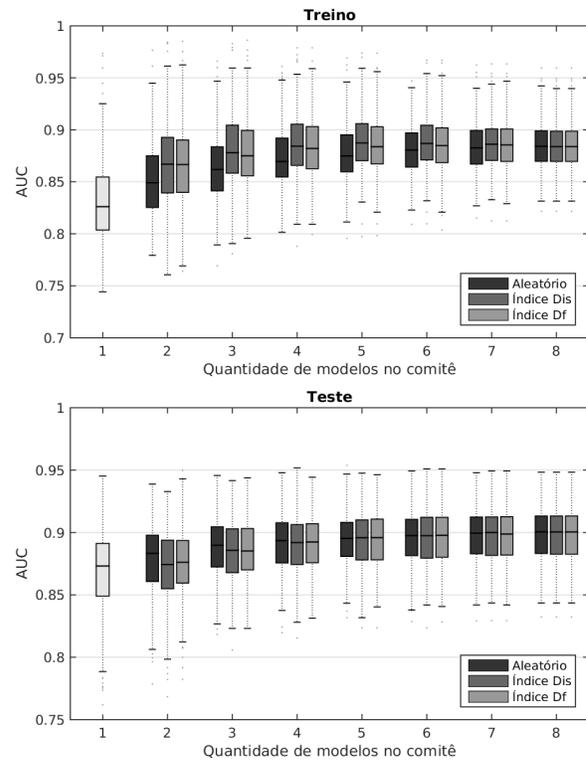


Figura 3: Boxplot do valor da AUC para os conjuntos de treino e teste avaliando a influência da quantidade de neurônios na camada intermediária

seleção dos modelos a partir do conjunto de validação e a questão do emprego destes índices como critério de seleção.

A despeito deste fato, o sistema de comitê obteve um desempenho no teste sensivelmente superior aos modelos individuais, apresentando uma AUC mediana maior que 0.88, independente do critério de seleção. Através do *t-test* para diferentes variâncias, obteve-se um valor de  $p$  baixo ( $p < 0,001$ ), o que comprova a diferença entre as distribuições das AUCs dos comitês, considerando 4 modelos, com relação aos melhores modelos individuais.

## V. CONCLUSÃO

Este trabalho propôs um estudo experimental para a formação de comitês de classificação voltados ao suporte à decisão quanto ao isolamento respiratório de pacientes suspeitos de tuberculose pulmonar. Foi avaliada a pertinência do uso dos índices de diversidade para a seleção dos membros do comitê. Os sistemas estudados apresentaram um desempenho sensivelmente superior aos melhores modelos neurais individuais para todos os critérios de seleção adotados.

No conjunto de teste, foi possível obter um valor mediano da AUC maior que 0.88, comparado a um valor de 0.87 para os modelos individuais, a despeito dos índices utilizados se mostrarem mais eficazes no aumento da acurácia para o conjunto de treinamento, onde o ganho foi mais expressivo. Neste caso, foi possível obter um valor mediano da AUC

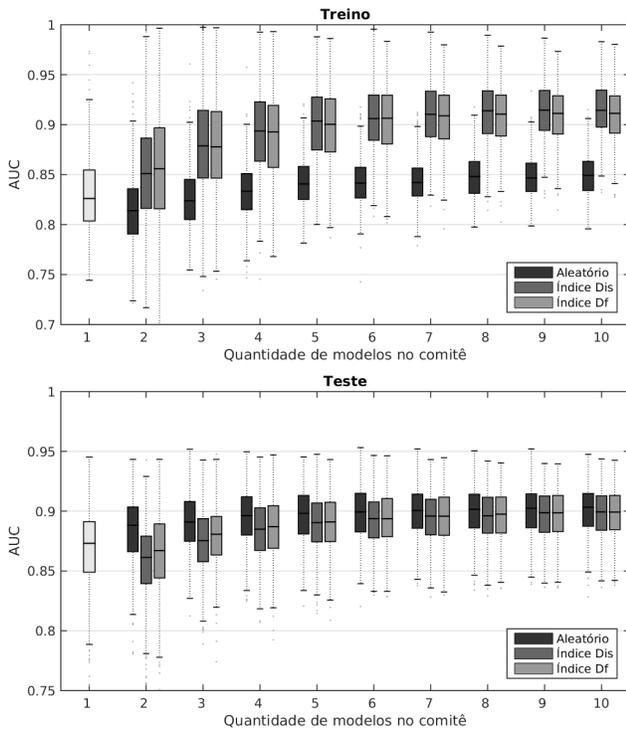


Figura 4: Boxplot do valor da AUC para os conjuntos de treino e teste considerando todos os parâmetros

superior a 0.89, considerando-se um comitê de 4 membros e a seleção através dos índices de divergência, comparado ao valor de 0.82 para os modelos individuais. Este melhor desempenho pode estar relacionado ao fato de que a seleção dos modelos foi realizada utilizando-se os conjuntos de validação, aspecto, no entanto, que deve ser aprofundado futuramente.

Como trabalhos futuros, pretende-se explorar métodos de clusterização para a formação dos conjuntos. Também serão avaliadas outras técnicas de classificação, tais como *Random Forest* e *Adaboost*.

#### REFERÊNCIAS

- [1] R. I. D. Garcia, S. B. Cecatto, R. R. Mendonça, C. E. M. Barcelos, R. O. Santos, and P. B. Rapoport, "Tuberculose e blastomicose laríngeas: relato de três casos e revisão de literatura," *Revista Brasileira de Otorrinolaringologia*, vol. 70, pp. 255 – 259, 04 2004.
- [2] WHO, in *Global tuberculosis report 2014*. Geneva: World Health Organisation, 2014.
- [3] —, *The global plan to stop TB 2006-2015*. Geneva: World Health Organisation, 2006. [Online]. Available: <http://www.stoptb.org/global/plan/main/default.asp>
- [4] F. Andrade, J. Souza Filho, R. Galliez, and A. Kritski, "Modelo neural para o suporte ao diagnóstico de tuberculose pulmonar na triagem de pacientes," in *XXIV Congresso Brasileiro de Engenharia Biomédica*, Uberlândia, MG, Brasil, 2014.
- [5] V. Muniz, J. Souza Filho, L. Almeida, F. Andrade, R. Galliez, and A. Kritski, "Escore de pontos para a identificação de tuberculose pulmonar derivado por inteligência computacional," in *XXIV Congresso Brasileiro de Engenharia Biomédica*, Uberlândia, MG, Brasil, 2014.
- [6] J. a. B. de Oliveira e Souza Filho, A. P. P. Vieira, J. M. de Seixas, F. S. Aguiar, F. C. de Queiroz Mello, and A. L. Kritski, "An intelligent system for managing the isolation of patients suspected of pulmonary tuberculosis," in *Intelligent Data Engineering and Automated Learning - IDEAL 2012*, ser. Lecture Notes in Computer Science, H. Yin, J. A.

- Costa, and G. Barreto, Eds. Springer Berlin Heidelberg, 2012, vol. 7435, pp. 818–825.
- [7] L. Rokach, *Pattern Classification Using Ensemble Methods*. River Edge, NJ, USA: World Scientific Publishing Co., Inc., 2010.
- [8] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2007.
- [9] J. Surowiecki, *The Wisdom of Crowds*. Anchor, 2005.
- [10] L. K. Hansen and P. Salamon, "Neural network ensembles," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 10, pp. 993–1001, Oct. 1990. [Online]. Available: <http://dx.doi.org/10.1109/34.58871>
- [11] X. Hu, "Using rough sets theory and database operations to construct a good ensemble of classifiers for data mining applications," in *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, 2001, pp. 233–240.
- [12] A. Krogh and J. Vedelsby, "Neural network ensembles, cross validation, and active learning," in *Advances in Neural Information Processing Systems*. MIT Press, 1995, pp. 231–238.
- [13] G. Zenobi and P. Cunningham, "Using diversity in preparing ensembles of classifiers based on different feature subsets to minimize generalization error," in *Lecture Notes in Computer Science*. Springer Verlag, 2001, pp. 576–587.
- [14] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, 2004.
- [15] D. B. Skalak, "The sources of increased accuracy for two proposed boosting algorithms," in *In Proc. American Association for Arti Intelligence, AAAI-96, Integrating Multiple Learned Models Workshop*, 1996, pp. 120–125.
- [16] G. Giacinto and F. Roli, "Design of effective neural network ensembles for image classification purposes," *Image Vision and Computing Journal*, vol. 19, pp. 699–707, 2001.
- [17] D. W. Hosmer and S. Lemeshow, *Applied logistic regression (Wiley Series in probability and statistics)*, 2nd ed. Wiley-Interscience Publication, 2000.
- [18] S. Theodoridis and K. Koutroubas, *Pattern Recognition, Fourth Edition*, 4th ed. Academic Press, 2008.
- [19] R. Levesque, *SPSS Programming and Data Management, 4th Edition: A Guide for SPSS and SAS Users*, 4th ed. SPSS, 2007.
- [20] C. Bunkhumpornpat, K. Sinapiromsaran, and C. Lursinsap, "Safe-level-smote: Safe-level-synthetic minority over-sampling technique for handling the class imbalanced problem," in *Proceedings of the 13th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining*, ser. PAKDD '09. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 475–482.
- [21] H. Han, W.-Y. Wang, and B.-H. Mao, "Borderline-smote: A new over-sampling method in imbalanced data sets learning," in *Proceedings of the 2005 International Conference on Advances in Intelligent Computing - Volume Part I*, ser. ICIC'05. Berlin, Heidelberg: Springer-Verlag, 2005, pp. 878–887.
- [22] M. Riedmiller and H. Braun, "A direct adaptive method for faster backpropagation learning: The rprop algorithm. in *ieee international conference on neural networks*," 1993.
- [23] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the marquardt algorithm," *Trans. Neur. Netw.*, vol. 5, no. 6, pp. 989–993, Nov. 1994. [Online]. Available: <http://dx.doi.org/10.1109/72.329697>
- [24] J. Dennis, John E. and R. B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations*, ser. Classics in applied mathematics. Philadelphia: Society for Industrial and Applied Mathematics, 1996, premiere publication : Englewood Cliffs (N.J.) : Prentice-Hall, cop. 1983. [Online]. Available: <http://opac.inria.fr/record=b1080807>
- [25] X. Zhou, N. Obuchowski, and D. McClish, *Statistical Methods in Diagnostic Medicine*, ser. Wiley Series in Probability and Statistics. Wiley, 2011. [Online]. Available: <http://books.google.com.br/books?id=NGWUtAdOT8kC>