

Configuração de hyper-parâmetros de modelos *deep learning* para auxílio no pós-diagnóstico de Tuberculose

Thomás Tabosa de Oliveira¹, Vanderson Sampaio^{2,3}, Patricia Takako Endo¹

¹Universidade de Pernambuco (UPE)
Caruaru, Pernambuco, Brasil

²Fundação de Medicina Tropical Doutor Heitor Vieira Dourado (FMT-AM)
Manaus, Amazonas, Brasil

³Fundação de Vigilância em Saúde do Amazonas (FVS-AM)
Manaus, Amazonas, Brasil

tto@ecomp.poli.br, vandersons@gmail.com, patricia.endo@upe.br

Abstract. *Tuberculosis (TB) is recognised as the deadliest disease in the world, according to the World Health Organization (WHO), being one of the ten leading causes of death in the world, as well as being the leading cause of death for people with HIV. Brazil is one of the countries with a high TB burden, and one of the highest mortality rates in the country is in the state of Amazonas. The aim of this paper is to analyse deep learning (DL) models to assist in the post-diagnosis of TB, predicting the severity of the disease in the patient. Two DL models are proposed and the Grid-search technique is applied to define configurations with the best performance. DL models yield interesting results, with a fully connected DL configuration reaching specificity of 83.4%.*

Resumo. *A Tuberculose (TB) é reconhecida como a doença mais mortal do mundo, segundo a World Health Organization (WHO), sendo umas das dez maiores causas de morte no mundo, além de ser a principal causa de morte de pessoas HIV-positivas. O Brasil é um dos países com alta carga de TB, e uma das maiores taxas de mortalidade do país encontra-se no estado do Amazonas. O objetivo deste trabalho é analisar modelos de deep learning (DL) para auxiliar no pós-diagnóstico de TB, predizendo a gravidade da doença no paciente. Dois modelos de DL são propostos e a técnica de Grid-search é aplicada para definir as configurações com os melhores desempenhos. Os modelos de DL apresentam resultados interessantes, com uma configuração da DL totalmente conectada atingindo 83,4% de especificidade.*

1. Introdução

A tuberculose (TB) é uma doença caracterizada como um problema de saúde mundial, sendo uma das principais causas infecciosas de morbidade e mortalidade no mundo, superando até a infecção pelo vírus da imunodeficiência humana (HIV) [Ryndak and Laal 2019]. Segundo o *The World Health Organization (WHO)*¹, 10 milhões de pessoas adoeceram por TB no mundo e houve 1,5 milhão de óbitos pela doença em 2018. No Brasil, anualmente, ocorrem cerca de 80 mil novos casos, além de 15 mil casos

¹<https://www.who.int/news-room/fact-sheets/detail/tuberculosis>

de retratamento. Esses valores fazem a TB aparecer em 4º lugar no *ranking* de mortalidade por doenças infecciosas e no 7º em despesas com internação de pacientes na saúde pública.

Segundo o relatório global de TB de 2018 [WHO], apesar da TB ser considerada um problema de saúde pública, ela é uma doença tratável, se o diagnóstico for realizado de forma correta e o paciente receber um tratamento adequado. Muitas vezes, dependendo da gravidade da doença, um tratamento domiciliar pode ser suficiente para a cura do paciente. Porém, em alguns casos, os tratamentos definidos pelos profissionais de saúde não são os mais adequados a gravidade do paciente, ocasionando um número considerável de óbitos por TB [Goldman and Schafer 2015].

Diversos estudos estão sendo desenvolvidos para auxiliar no processo de diagnóstico e pós-diagnóstico não apenas de TB, mas na área de saúde em geral; e entre as técnicas utilizadas, *deep learning* (DL) tem sido amplamente aplicada, devido sua eficiência no processo de classificação e predição. Através da técnica de DL, é possível que modelos computacionais aprendam como os dados se comportam a partir de um conjunto de informações (encontram um padrão de comportamento) [LeCun et al. 2015]

Por exemplo, em [Cao et al. 2016] e [Rajaraman et al. 2018], os autores apresentam modelos de DL para detectar se o paciente está ou não com TB, analisando imagens de radiografias e cultura de escarro, respectivamente. Todavia, existem poucos trabalhos que focam no pós-diagnóstico da doença. O pós-diagnóstico é uma fase primordial no tratamento do paciente de TB, visto que é nesta fase que a gravidade da doença é determinada, seguida pela tomada de decisão para definir o tratamento mais apropriado, assegurando que o paciente não receba um tratamento pouco eficiente e também evitando gastos desnecessários de recursos do sistema de saúde em tratamentos mais intensivos em pacientes que não necessitam.

Considerando este contexto, este trabalho tem como principal objetivo propor modelos de DL para auxiliar no pós-diagnóstico de TB, com a finalidade de prever a gravidade da doença no paciente, auxiliando assim na tomada de decisão do profissional de saúde em relação ao tratamento a ser realizado. Vale destacar que o modelo irá prever a probabilidade de óbito do paciente, com a finalidade de guiar o entendimento da gravidade da doença.

2. Materiais e métodos

2.1. Base de Dados

Os dados utilizados para o treinamento e teste dos modelos de DL são provenientes do SINAN², uma base nacional contendo informações de pacientes que foram diagnosticados com alguma doença presente na lista de doenças de notificação compulsória, entre elas a TB. Para este trabalho foram utilizadas informações de pacientes do estado do Amazonas entre os anos de 2007 e 2018, contendo cerca de 36 mil registros de pacientes³.

2.2. Pré-processamento dos dados

A base utilizada possui originalmente 130 preditores com informações variadas do paciente, como nome, endereço, sexo, tipo de TB, entre outros. Em conjunto com um

²<http://portalsinan.saude.gov.br/>

³O projeto foi aprovado pelo Comitê de Ética do SINAN. Protocolo 13142519.1.0000.0005.

especialista na área de saúde, foram selecionados 19 preditores como parâmetros para os modelos de DL. Os preditores selecionados possuem informações sobre sexo e raça do paciente, se o mesmo possui alguma doença ou situação que cause agravo da doença, como alcoolismo, AIDS, diabetes, doença mental, HIV, tabagismo ou consumo de drogas ilícitas. Por fim, temos informações de exames sobre a TB (resultado de cultura de escarro, baciloscopia, exame histopatológico, e informações sobre a forma de TB, pulmonar, extra-pulmonar ou ambos).

Após a definição dos preditores, foi necessário realizar o balanceamento da base de dados, pois a mesma possuía 34.985 casos notificados de TB com desfecho de não óbito e apenas 1.224 casos que resultaram em óbito por TB. O desbalanceamento pode fazer com que o modelo tenda a responder que o paciente estará na classe não óbito, pois o mesmo possui mais informações de treino para esta classe. Para mitigar este problema e evitar que o treinamento dos modelos de DL gerasse resultados enviesados, a técnica de re-amostragem denominada *undersampling* [Liu et al. 2008] foi aplicada. Esta técnica realiza a re-amostragem com base na classe com menor quantidade de dados (classe óbito); para isso, a classe majoritária (classe não óbito) é recortada até que a mesma obtenha o mesmo tamanho da classe menor. O recorte dos dados foi realizado por meio de um pareamento com a classe de óbito, utilizando o sexo do paciente e a idade quando o mesmo foi diagnóstico por TB, ou seja, para cada paciente da classe de óbito, pegamos um paciente da classe não óbito com a mesma idade e sexo, com o objetivo de termos essas informações mais semelhantes entre eles. A base de dados foi dividida da seguinte forma: 50% para treinamento, 25% para validação e 25% para teste.

2.3. Métricas para avaliação dos modelos

Para avaliar o desempenho dos modelos foram selecionadas cinco métricas, calculadas com base nos verdadeiros positivos (VP), verdadeiros negativos (VN), falsos positivos (FP) e falsos negativos (FN) das predições, sendo elas: F1-score, precisão, sensibilidade, especificidade e acurácia.

2.4. Modelos

Para este trabalho, serão utilizados dois modelos de DL: uma DL totalmente conectada e uma Rede Neural Convolucional (do inglês, *convolutional neural network* - CNN). A DL totalmente conectada é o modelo mais comum de DL, por esta razão, decidimos usá-lo como modelo base deste trabalho. Além disso, os modelos de DL totalmente conectada apresentam bons resultados em dados agrupados. Já a CNN é um modelo de DL muito utilizado para detecção de TB com processamento digital de imagens, como visto em [Rajaraman et al. 2018] e [Lakhani and Sundaram 2017], por exemplo. Apesar do modelo CNN ser mais utilizado para processamento de imagens e dados não agrupados, como podemos observar em [Shin et al. 2016], algumas técnicas para a utilização deste modelo em dados tabulares também estão sendo utilizadas, como em [Han et al. 2018]. Sendo assim, com objetivos de testes, resolvemos aplicar o modelo de CNN com os dados tabulares dos pacientes de TB para compararmos com o modelo de DL totalmente conectada.

3. Resultados e discussões

3.1. Grid-search

Grid-search é uma técnica onde é possível avaliar diferentes combinações de configurações de modelos de DL. Para o procedimento, são definidos dois ou mais parâmetros de configuração, e a *Grid-search* executa todas as combinações possíveis entre esses parâmetros e retorna o valores das métricas pré-determinadas. De acordo com [Bergstra and Bengio 2012], apesar da técnica de *Grid-search* possuir limitações e não ser muito recomendada para otimizações com muitos parâmetros diferentes, a mesma continua sendo, junto com a pesquisa manual, uma das técnicas mais utilizadas. Para este trabalho serão utilizados dois parâmetros diferentes para cada modelo de DL. Cada configuração de modelo será executada 10 vezes, e para obtenção dos resultados, realizar-se-á a média das métricas de F1-score, precisão, sensibilidade, especificidade e acurácia.

A Tabela 1 apresenta as configurações do *Grid-search* para o modelo de DL totalmente conectada, e a Tabela 2 apresenta as configurações utilizadas para o modelo CNN. Por limitação de espaço, apenas os resultados das melhores configurações serão apresentados e discutidos a seguir.

Tabela 1. Parâmetros do *Grid-search* para o modelo DL totalmente conectada

Parâmetros	Níveis
Quantidade de camadas	De 1 a 3, de 1 em 1
Quantidade de neurônios	De 100 a 250, de 25 em 25

Tabela 2. Parâmetros do *Grid-search* para o modelo CNN

Parâmetros	Níveis
Quantidade de camadas convolucionais	De 1 a 3, de 1 em 1
Quantidade de filtros	De 20 a 80, de 20 em 20

3.2. DL Totalmente Conectada

Analisando os resultados de F1-score, duas configurações apresentaram resultados interessantes: a configuração com 1 camada e 125 neurônios (denominada **F-1-125**) e a configuração com 1 camada e 150 neurônios (denominada de **F-1-150**), ambas com 76,1%. Averiguando a configuração F-1-125, pode-se observar que esta configuração também apresenta bons valores para as métricas de precisão (com 80,8%), acurácia (com 77,1%) e especificidade (com 82,5%). Já a configuração F-1-150 apresentou resultados um pouco inferiores que a F-1-125 nas métricas de acurácia (com 76,1%), precisão (com 77,2%) e especificidade (com 77,2%). Outra configuração que também se destacou foi a configuração com 1 camada e 100 neurônios (denominada **F-1-100**), que apresentou o melhor resultado de especificidade (83,4%), porém demonstrou uma taxa um pouco menor para a métrica de sensibilidade (70,3%), que é considerada a mais importante para este trabalho. Considerando a métrica de sensibilidade, a melhor configuração encontrada é a configuração que possui 3 camadas e 100 neurônios (denominada **F-3-100**), com 76,6%, no entanto esta mesma configuração apresenta os menores valores para as demais métricas (70,1% em acurácia, 72,1% em F1-score, 68% em precisão e 63,5% em especificidade).

3.3. CNN

Uma configuração se destacou nos resultados de F1-score do modelo CNN, foi a configuração composta por 3 camadas e 40 filtros (denominada **C-3-40**) com 73,5%. A mesma configuração também possui o melhor resultado de acurácia (74,7%) e as demais métricas apresentam valores semelhantes. Outra configuração que também se destacou foi a configuração com 2 camadas e 20 filtros (denominada **C-2-20**). Esta configuração possui os melhores valores de precisão (77,7%) e especificidade (80,5%), porém apresenta um baixo valor para sensibilidade (67,3%). A configuração que possui a melhor métrica para sensibilidade é a configuração que possui 3 camadas e 80 filtros (denominada **C-3-80**), com uma taxa de 71%. Esta configuração também apresenta valores semelhantes para as demais métricas.

3.4. Discussão comparativa dos modelos

A Tabela 3 apresenta os valores das melhores configurações dos modelos da DL totalmente conectada e da CNN. Analisando as configurações, pode-se notar que os modelos de DL totalmente conectada obtiveram os melhores resultados em comparação aos modelos CNN. O motivo deste resultado provavelmente se deve ao fato de que a CNN não é muito recomendada para dados agrupados, enquanto que a DL totalmente conectada apresenta bons resultados para estes tipos de dados. Contudo, vale salientar que apesar de possuir valores mais baixos, os resultados são próximos.

Tabela 3. Resultado das melhores configurações dos modelos a partir do *Grid-search*

Configuração	F1-Score	Precisão	Sensibilidade	Especificidade	Acurácia
F-1-100	75,40%	81,30%	70,30%	83,40%	76,80%
F-1-125	76,10%	80,80%	71,90%	82,50%	77,10%
F-3-100	72,10%	68%	76,60%	63,50%	70,10%
C-2-20	72%	77,70%	67,30%	80,50%	73,80%
C-3-40	73,50%	77,40%	70%	79,50%	74,70%
C-3-80	72,40%	73,90%	71%	75%	73,10%

De forma geral, todas as configurações de ambos os modelos apresentaram valores acima de 70% para todas as métricas, com exceção das configurações F-3-100 (em precisão e sensibilidade) e C-2-20 (em sensibilidade). A métrica de maior valor foi a de especificidade, com 83,4%, na configuração F-1-100, mostrando que essa configuração está acertando melhor os casos de pacientes que não morreram por TB.

Para o objetivo deste trabalho, uma das métricas mais importantes é a métrica de sensibilidade, que mostra o quanto o modelo está acertando os casos de pacientes que resultaram em óbito por TB. A configuração que apresentou a maior taxa de sensibilidade foi a F-3-100, alcançando 76,6%, porém o mesmo modelo apresentou uma baixa taxa de especificidade (63,5%), mostrando que este modelo está prevendo corretamente mais casos que resultam em óbito do que os casos onde o paciente sobrevive a doença.

4. Considerações finais e próximos passos

A TB ainda é uma das doenças que mais infecta pessoas no mundo e, no Brasil, dezenas de milhares de casos são registrados anualmente. Ainda, a TB está entre as principais cau-

sas de óbitos por doenças infecciosas no país. Este trabalho propôs modelos de DL para auxiliar no pós-diagnóstico de TB e prever a probabilidade de óbito do paciente e, com isso, auxiliar profissionais de saúde na tomada de decisão para a definição do tratamento mais adequado. A partir dos resultados do *Grid-search*, observou-se que o modelo F-1-100 apresentou a melhor taxa de acerto para casos de não óbito (especificidade), enquanto que o modelo F-3-100 apresentou a melhor taxa de acerto para casos de óbito (sensibilidade). Como próximos passos desta pesquisa, planeja-se realizar melhorias nos modelos propostos. Uma abordagem que pode ser bem-sucedida é treinar diferentes modelos, em vez de um único, e combinar as previsões desses modelos (*ensemble learning*).

Referências

- Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(Feb):281–305.
- Cao, Y., Liu, C., Liu, B., Brunette, M. J., Zhang, N., Sun, T., Zhang, P., Peinado, J., Garavito, E. S., Garcia, L. L., et al. (2016). Improving tuberculosis diagnostics using deep learning and mobile health technologies among resource-poor and marginalized communities. In *IEEE International Conference on Connected Health (CHASE)*, pages 274–281. IEEE.
- Goldman, L. and Schafer, A. I. (2015). *Goldman-Cecil Medicine E-Book*. Elsevier Health Sciences.
- Han, H., Zhu, X., and Li, Y. (2018). Edlt: Enabling deep learning for generic data classification. In *2018 IEEE International Conference on Data Mining (ICDM)*, pages 147–156.
- Lakhani, P. and Sundaram, B. (2017). Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology*, 284(2):574–582.
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature* 521.
- Liu, X.-Y., Wu, J., and Zhou, Z.-H. (2008). Exploratory undersampling for class-imbalance learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(2):539–550.
- Rajaraman, S., Candemir, S., Xue, Z., Alderson, P. O., Kohli, M., Abuya, J., Thoma, G. R., and Antani, S. (2018). A novel stacked generalization of models for improved tb detection in chest radiographs. In *IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 718–721. IEEE.
- Ryndak, M. B. and Laal, S. (2019). Mycobacterium tuberculosis primary infection and dissemination: A critical role for alveolar epithelial cells. *Frontiers in Cellular and Infection Microbiology*, 9:299.
- Shin, H., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., and Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, 35(5):1285–1298.