

Análise de modelos de *deep learning* para auxílio no pós-diagnóstico de tuberculose

Thomás Tabosa Oliveira

Elisson Rocha

Theo Lynn

Vanderson Sampaio

Patricia Takako Endo

Resumo: segundo a Organização Mundial de Saúde, a tuberculose (TB) é uma das dez maiores causas de morte no mundo, além de ser a principal causa de morte de pessoas HIV-positivas. O objetivo deste trabalho é propor e analisar modelos de *deep learning* (DL) que auxiliam no pós-diagnóstico de TB, predizendo a gravidade da doença no paciente. Dois modelos de DL são propostos, uma DL totalmente conectada e uma rede neural convolucional (do inglês *convolutional neural network* - CNN) e a técnica de *Grid-search* é aplicada nos dois modelos para definir as configurações com os melhores desempenho. Os modelos de DL apresentam resultados interessantes, com uma configuração da DL totalmente conectada atingindo 83,4% de especificidade, e uma configuração da CNN atingindo 80,5%.

Palavras-chave: Deep learning, Tuberculose; CNN; Grid-search; Predição; Agenda 2030.

Abstract: According to the World Health Organization, Tuberculosis (TB) is one of the ten leading causes of death in the world, and is the main cause of death of HIV-positive people. The objective of this paper is to propose and analyse deep learning (DL) models that support the TB post-diagnosis process. Two models are proposed, a fully connected DL and a convolutional neural network (CNN), and the Grid-search technique is performed on both models to define the configurations that achieve the best results. The models present interesting results, a fully connected DL configuration achieves 83.4% of specificity, and a CNN configuration, 80.5%.

Keywords: Deep learning; Tuberculosis; CNN; Grid-search; Prediction; Agenda 2030.

Artigo recebido em 28/10/2019 e aceito em 28/11/2019

Introdução

Em 2015, a Agenda 2030¹ foi idealizada por representantes dos Estados-membros da ONU com o propósito de erradicar a pobreza em todas as formas e dimensões para a execução do desenvolvimento sustentável do planeta. Para o cumprimento deste objetivo maior, foram elaborados 17 Objetivos de Desenvolvimento Sustentável (ODS). Dentre elas, a ODS 3 - Saúde e Bem-Estar - busca promover o bem-estar para todos, em todas as idades. A meta 3.3 da ODS 3 tem como objetivo “até 2030, acabar com as epidemias de

ANÁLISE DE MODELOS DE *DEEP LEARNING* PARA AUXÍLIO NO PÓS-DIAGNÓSTICO DE TUBERCULOSE

THOMÁS OLIVEIRA; ELISSON ROCHA; THEO LYNN; VANDERSON SAMPAIO E PATRÍCIA ENDO

AIDS, tuberculose, malária e doenças tropicais negligenciadas, e combater a hepatite, doenças transmitidas pela água, e outras doenças transmissíveis”. Dentre essas doenças tropicais negligenciadas, encontra-se a tuberculose (TB).

Segundo o *The World Health Organization* (WHO), 10 milhões de pessoas adoeceram por TB no mundo e houve 1,5 milhão de óbito pela doença em 2018. No Brasil, anualmente, ocorrem cerca de 80 mil novos casos, além de 15 mil casos 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 (Hijjar et al. 2005). Dentre os fatores que facilitam a infecção e o óbito por TB estão infecção pelo vírus do HIV, diabetes, alcoolismo, uso de drogas ilícitas, entre outros.

Segundo o relatório de 2018 do 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 não são adequados a gravidade do paciente, ocasionando um número considerável de óbitos por TB (Goldman e 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 propõem modelos de DL para detectar se o paciente está ou não com TB, analisando imagens de radiografias e cultura de escarro, respectivamente. Todavia, existem ainda 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 onde é determinada a gravidade da doença e qual o tratamento apropriado, assegurando que o paciente não receba um tratamento pouco eficiente e também evitando gastos desnecessários de

ANÁLISE DE MODELOS DE *DEEP LEARNING* PARA AUXÍLIO NO PÓS-DIAGNÓSTICO DE TUBERCULOSE

THOMÁS OLIVEIRA; ELISSON ROCHA; THEO LYNN; VANDERSON SAMPAIO E PATRÍCIA ENDO

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. Este estudo utiliza dados do Sistema de Informação de Agravos de Notificação (SINAN), referentes ao Estado do Amazonas, no período de 2007 a 2018, contendo cerca de 13 mil registros.

Vale salientar que, apesar de considerar inicialmente dados de pacientes do Estado do Amazonas, os modelos propostos neste trabalho são facilmente adaptáveis para aprender e testar dados de pacientes de outras regiões do Brasil, bem como de outros países do mundo.

Conceitos básicos:

Machine learning e deep learning

Machine learning (ML) é um método da análise de dados que tem como objetivo principal programar uma máquina para que ela possa aprender padrões. Segundo Freeman et al. (2018), a definição de aprender é “*o processo de conversão de informações em conhecimento*”, onde este conhecimento define as funcionalidades e resultados da máquina. Conforme Bulbul e Unsal (2011), “*comparado aos métodos clássicos, o processo de obtenção de informações é muito mais preciso e rápido com mineração de dados e ML*”, pois os modelos de ML auxiliam na estimativa dos resultados com o aprendizado de informações previamente inseridas.

Deep learning (DL) é uma sub-área de *ML*, que tem como objetivo projetar um modelo computacional para simular o comportamento do cérebro humano. Um modelo de DL possui neurônios e camadas. O processamento dos dados é realizado pelos neurônios, e os mesmos são organizados em camadas. Uma arquitetura de DL pode ser

ANÁLISE DE MODELOS DE *DEEP LEARNING* PARA AUXÍLIO NO PÓS-DIAGNÓSTICO DE TUBERCULOSE

THOMÁS OLIVEIRA; ELISSON ROCHA; THEO LYNN; VANDERSON SAMPAIO E PATRÍCIA ENDO

capaz de aprender automaticamente como se comportam os dados a partir da entrada bruta dos mesmos (Khan e Yairi, 2018).

Para este trabalho, serão propostos dois modelos de DL: uma DL totalmente conectada e uma Rede Neural Convolutacional (do inglês, *convolutional neural network* - CNN), descritos a seguir.

DL totalmente conectada

Uma DL totalmente conectada é uma das formas mais comuns de DL. São formadas por componentes simples, denominados neurônios. Basicamente, os neurônios são acoplados em três tipos de camadas: (a) a camada de entrada, (b) as camadas escondidas e (c) a camada de saída. A camada de entrada recebe os dados brutos que serão passados pela rede neural até serem classificados na camada de saída.

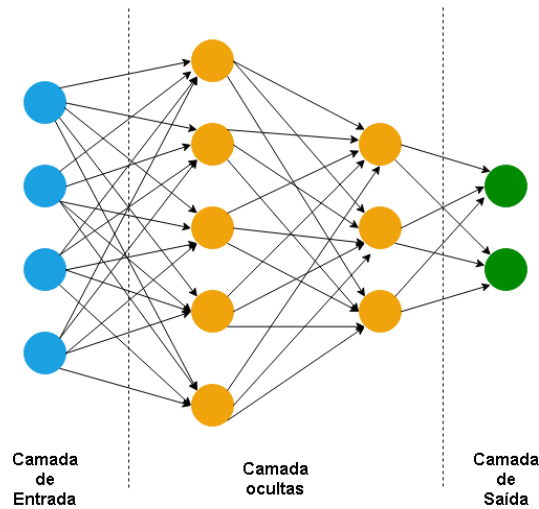
Uma rede neural composta por somente camada de entrada e saída só consegue resolver problemas lineares. Porém, os problemas do mundo real são, em quase sua totalidade, problemas não-lineares e, portanto, as camadas escondidas são responsáveis por introduzir esse fator de não-linearidade para as redes neurais. Os neurônios das camadas escondidas contêm pesos específicos que influenciam diretamente nos dados repassados a cada camada do modelo. Esses pesos são ajustados na fase de treinamento da rede (LeCun et al. 2015), (Schwing e Urtasan, 2015).

A Figura 1 apresenta um exemplo de modelo de DL totalmente conectada, onde a camada de entrada é composta por quatro neurônios, há duas camadas escondidas, com cinco e três neurônios, respectivamente, e uma camada de saída com dois neurônios. Pode-se observar que todos os neurônios estão conectados a todos os neurônios da camada posterior.

Figura 1. Exemplo de uma DL totalmente conectada

ANÁLISE DE MODELOS DE *DEEP LEARNING* PARA AUXÍLIO NO PÓS-DIAGNÓSTICO DE TUBERCULOSE

THOMÁS OLIVEIRA; ELISSON ROCHA; THEO LYNN; VANDERSON SAMPAIO E PATRÍCIA ENDO



Fonte.: Próprio Autor

CNN

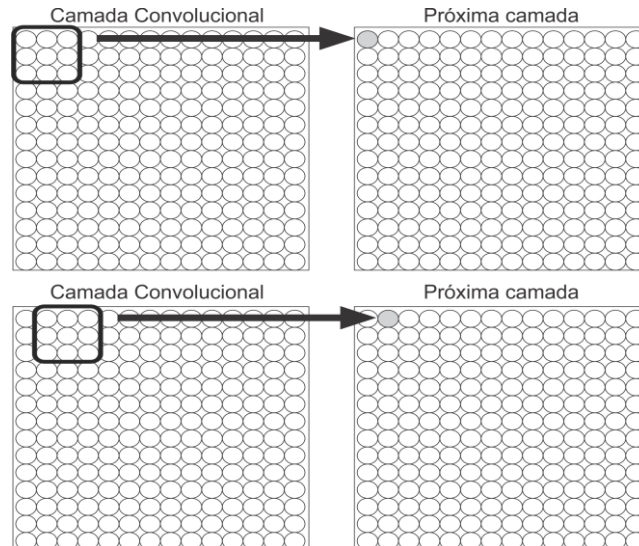
A técnica de CNN (Takahashi et al. 2018) tem sido bastante utilizada nas áreas de processamento de imagem, visão computacional, entre outros. O modelo de CNN diferencia-se pela existência de camadas convolucionais e camadas de subamostragem (*pooling*) antes das camadas densas (camadas finais do modelo). Nos modelos de CNN os neurônios são denominados de filtros.

Nas camadas convolucionais (Simonyan e Zisserman, 2014), o modelo utiliza os filtros, que criam uma unidade de área para percorrer a base, essas unidades “deslizam” por toda a base de dados e são responsáveis por detectar o elemento marcante daquela área e extrair a sua informação e enviar para a camada seguinte. O processo pode ser observado na Figura 2. Nesta camada as unidades apenas analisam os dados e encaminham os marcantes, os dados não são modificados e podem ser analisados em outra unidade de área. Ao final tem-se uma base de dados com o mesmo tamanho da anterior, porém esta base está apenas com os elementos marcantes.

ANÁLISE DE MODELOS DE *DEEP LEARNING* PARA AUXÍLIO NO PÓS-DIAGNÓSTICO DE TUBERCULOSE

THOMÁS OLIVEIRA; ELISSON ROCHA; THEO LYNN; VANDERSON SAMPAIO E PATRÍCIA ENDO

Figura 2. Filtro na camada convolucional da CNN



Fonte.: Próprio Autor

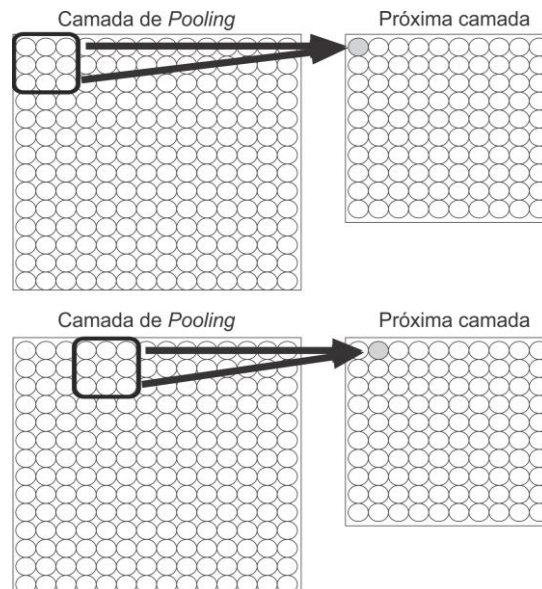
Na camada de *pooling* (Simonyan e Zisserman, 2014), os filtros utilizam as mesmas unidades de área para percorrer os dados, e as unidades são responsáveis por resumir toda a informação daquela área em um elemento menor e enviar para a camada seguinte. Após isso, o filtro parte para o próximo bloco de dados.

É importante notar que diferente da camada de convolução, nesta os dados são unidos e simplificados, não podendo ser repetidos no próximo filtro. Ao final desta camada, tem-se uma base de dados menor que a anterior pois a mesma está com os dados simplificados. A Figura 3 apresenta o processo da camada *pooling*.

Figura 3. Filtro na camada de *pooling* da CNN

ANÁLISE DE MODELOS DE *DEEP LEARNING* PARA AUXÍLIO NO PÓS-DIAGNÓSTICO DE TUBERCULOSE

THOMÁS OLIVEIRA; ELISSON ROCHA; THEO LYNN; VANDERSON SAMPAIO E PATRÍCIA ENDO



Fonte.: Próprio Autor

Por fim os dados são inseridos em uma camada totalmente conectada que terá como saída as classes definidas para classificação.

Trabalhos relacionados

A literatura atual já propõe alguns modelos de DL para diagnóstico de TB. Por exemplo, autores em Cao et al. (2016), Rajaraman et al. (2018), Hooda et al. (2017), Lakhani e Sundaram (2017), Stirenko et al. (2018) propuseram modelos que utilizam radiografia do tórax para diagnosticar TB. Em Cao et al. (2016) os autores desenvolveram um banco de dados de imagens de raios-X em larga escala, bem anotado e real e, em seguida, utilizaram o banco de dados para desenvolver modelos computacionais eficazes e eficientes (em particular os modelos baseados na CNN) para classificar imagens em diferentes categorias de manifestações de TB, atingindo 62,07% de precisão. Os autores em Hooda et al. (2017) também propuseram um modelo de DL do tipo CNN para diagnóstico de radiografia de tórax. O experimento foi realizado com classificação binária e verificou-se 82,09% de precisão na validação.

ANÁLISE DE MODELOS DE *DEEP LEARNING* PARA AUXÍLIO NO PÓS-DIAGNÓSTICO DE TUBERCULOSE

THOMÁS OLIVEIRA; ELISSON ROCHA; THEO LYNN; VANDERSON SAMPAIO E PATRÍCIA ENDO

Já os autores em Kant e Srivastava (2018), Sethi et al. (2018) propuseram modelos para o diagnóstico de bacilos a partir de imagens microscópicas de escarro. Em Kant e Srivastava (2018), a imagem passa pelo modelo de CNN que detecta se há ou não possíveis bacilos na imagem. Se negativo, o processo é encerrado e considerado negativo para bacilos no escarro. No entanto, se positivo, é processado por um segundo modelo de CNN para a confirmação do diagnóstico. Este modelo obteve 83,78% de sensibilidade. Em Sethi et al. (2018), os autores propuseram uma solução baseada em *hardware* incorporando um sistema de suporte de diagnóstico por microscopia para detecção de TB em amostras de escarro.

Os estudos acima se preocupam em diagnosticar a TB, que é uma parte fundamental do processo clínico, mas nenhum deles se concentra na previsão da gravidade por TB, como é a proposta deste trabalho. Alguns trabalhos recentes também propuseram modelos de DL para prever a mortalidade, mas em outros contextos, como por exemplo, mortalidade em idosos (Carneiro et al. 2017), morbidade por contaminação alimentar e infecções gastrointestinais (Song et al, 2017), e mortalidade intra-hospitalar pós-operatória (Lee et al. 2018).

Nenhum estudo utilizando DL foi encontrado para a fase pós-diagnóstico da TB e, especificamente, com o nosso objetivo de buscar entender a gravidade do diagnóstico do paciente e a probabilidade de morte por TB com base nas características demográficas, clínicas e/ou laboratoriais.

Método

Como apresentado anteriormente, este trabalho propõe dois modelos de DL para predição de probabilidade de morte por TB: uma DL totalmente conectada e uma CNN. Para o treinamento e teste dos modelos, utilizou-se dados referentes ao Estado do Amazonas, disponibilizados pelo SINAN. A seguir, a metodologia aplicada (tratamento dos dados, métricas de avaliação e modelos de DL) será descrita em mais detalhes e a técnica *Grid-search*, utilizada para a escolha de melhores configurações de modelos, também será apresentada.

ANÁLISE DE MODELOS DE *DEEP LEARNING* PARA AUXÍLIO NO PÓS-DIAGNÓSTICO DE TUBERCULOSE

THOMÁS OLIVEIRA; ELISSON ROCHA; THEO LYNN; VANDERSON SAMPAIO E PATRÍCIA ENDO

Base de dados

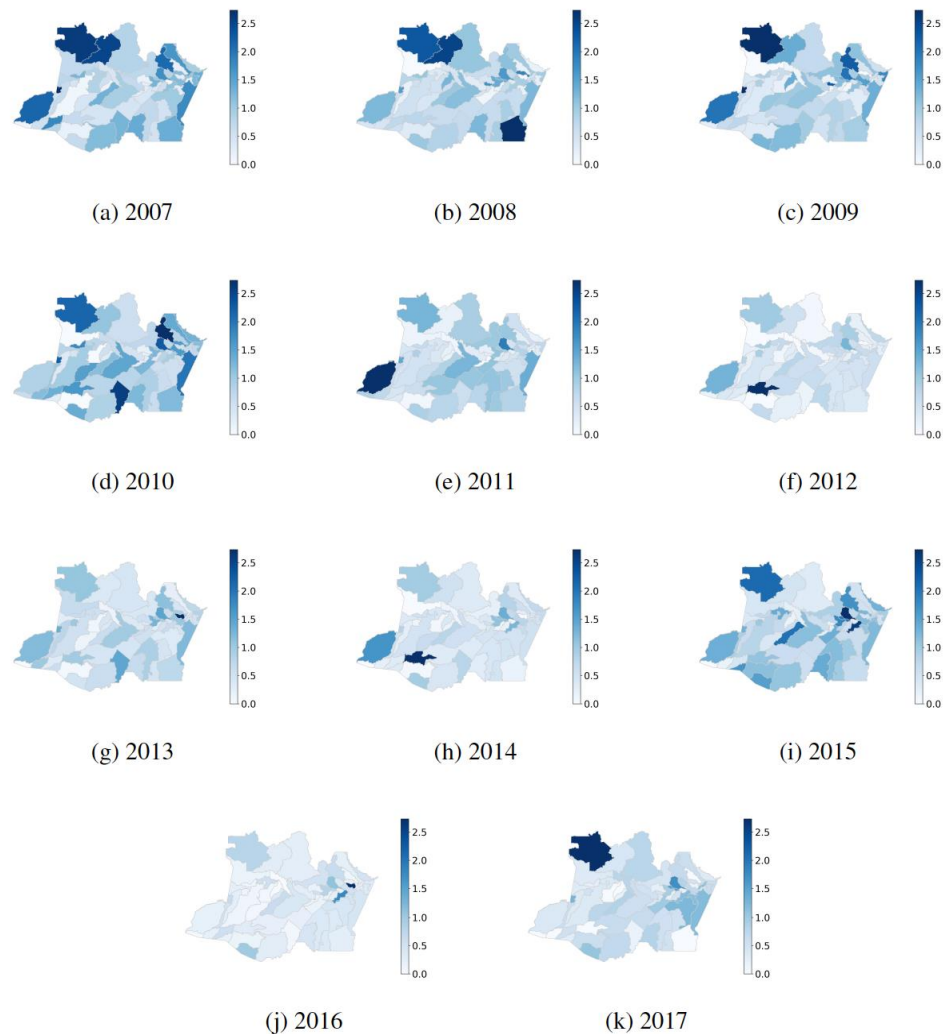
A base de dados utilizada para o treinamento e teste dos modelos de DL é proveniente do SINAN, uma base nacional contendo informações de pacientes que foram diagnosticados com alguma doença presente na lista de doenças de notificação compulsória, e a TB está incluída nesta lista. O SINAN permite o acesso a informação sobre a situação de diversas doenças no Brasil, facilitando o trabalho dos profissionais de saúde. 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 pacientes. Porém, vale ressaltar que os modelos propostos podem ser facilmente adaptáveis para receber dados de outros estados do Brasil, bem como de outros países, não se restringindo apenas ao Estado do Amazonas.

A Figura 4 mostra os incidentes de TB no Estado do Amazonas a cada mil habitantes nos de 2007 a 2017. O município com maior média de ocorrências por mil habitantes é o município de Silves, com 2,73 casos em 2016.

ANÁLISE DE MODELOS DE *DEEP LEARNING* PARA AUXÍLIO NO PÓS-DIAGNÓSTICO DE TUBERCULOSE

THOMÁS OLIVEIRA; ELISSON ROCHA; THEO LYNN; VANDERSON SAMPAIO E PATRÍCIA ENDO

Figura 4. Ocorrência de TB no estado do Amazonas a cada mil habitantes entre 2007 e 2017



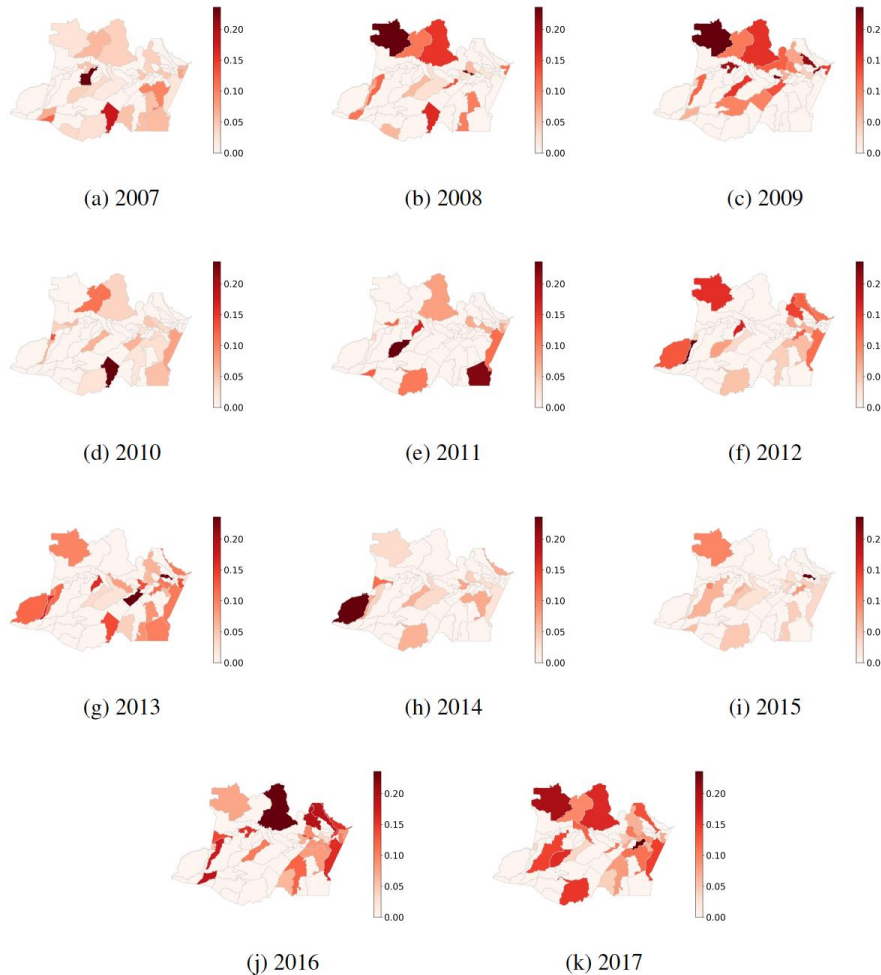
Fonte.: Próprio Autor

Já a Figura 5 apresenta os casos de óbito por TB no Estado do Amazonas a cada mil habitantes nos de 2007 a 2017. O município com maior média de ocorrências por mil habitantes é o município de Silves, com 2,73 casos em 2016. O município de Canutama registrou a maior ocorrência de óbitos por TB a cada 1000 habitantes no estado, sendo 0,23 no ano de 2010.

ANÁLISE DE MODELOS DE *DEEP LEARNING* PARA AUXÍLIO NO PÓS-DIAGNÓSTICO DE TUBERCULOSE

THOMÁS OLIVEIRA; ELISSON ROCHA; THEO LYNN; VANDERSON SAMPAIO E PATRÍCIA ENDO

Figura 5. Casos de óbito por TB no estado do Amazonas a cada mil habitantes entre 2007 e 2017



Fonte.: Próprio Autor

Preparando a base

Esta base de dados do SINAN possui originalmente 130 preditores (colunas da tabela) com informações variadas do paciente, como por exemplo nome, endereço, sexo, tipo de TB, entre outros. Em conjunto com um especialista na área de saúde, foram selecionados 19 preditores como parâmetros para os modelos de DL (ver Apêndice A).

ANÁLISE DE MODELOS DE *DEEP LEARNING* PARA AUXÍLIO NO PÓS-DIAGNÓSTICO DE TUBERCULOSE

THOMÁS OLIVEIRA; ELISSON ROCHA; THEO LYNN; VANDERSON SAMPAIO E PATRÍCIA ENDO

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. 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. A base de dados foi dividida da seguinte forma: 50% para treinamento, 25% para validação e 25% para teste.

Métricas

Para avaliar as diferentes configurações dos modelos de DL, cinco métricas serão utilizadas: F1-score, precisão, sensibilidade, especificidade e acurácia.

A métrica de acurácia é o indicador mais simples, utilizado para averiguar o quanto o modelo acertou, sendo calculado como a soma dos verdadeiros positivos (VP) com os verdadeiros negativos (VN) dividido pelo total de amostras.

$$acurácia = (VP + VN) / Total$$

Precisão é utilizada para definir quantos casos classificados como VP realmente são VP, e é calculada como o número de VP dividido pela soma de VP e falsos positivos (FP)

$$precisão = VP / (VP + FP)$$

A sensibilidade define o quanto um modelo previu corretamente os casos de VP, sendo calculado como o número de VP dividido pela soma de VP e falsos negativos (FN). A métrica de sensibilidade é considerada a mais importante para este trabalho, pois é a métrica que define o quanto que o modelo previu corretamente os casos de óbito.

$$sensibilidade = VP / (VP + FN)$$

ANÁLISE DE MODELOS DE *DEEP LEARNING* PARA AUXÍLIO NO PÓS-DIAGNÓSTICO DE TUBERCULOSE

THOMÁS OLIVEIRA; ELISSON ROCHA; THEO LYNN; VANDERSON SAMPAIO E PATRÍCIA ENDO

A métrica de especificidade determina o quanto o modelo previu corretamente os casos de VN. É calculado pelo número de VN dividido pela soma de VN e FP.

$$\text{especificidade} = VN / (VN + FP)$$

Por fim, o F1-score é a métrica que relaciona as métricas de precisão e sensibilidade, sendo calculada da seguinte forma:

$$F1 - score = (2 \times (\text{precisão} \times \text{sensibilidade}) / (\text{precisão} + \text{sensibilidade}))$$

Modelos de DL

Para este trabalho, serão propostos dois modelos de DL, uma DL totalmente conectada e uma CNN. Para detectar melhores configurações para cada modelo, a técnica *Grid-search* (Buitinck et al. 2013) será aplicada. *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 desses parâmetros e retorna o valores das métricas pré-determinadas. De acordo com Bergstra e 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, com isso diminuindo a limitação da técnica. 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.

ANÁLISE DE MODELOS DE *DEEP LEARNING* PARA AUXÍLIO NO PÓS-DIAGNÓSTICO DE TUBERCULOSE

THOMÁS OLIVEIRA; ELISSON ROCHA; THEO LYNN; VANDERSON SAMPAIO E PATRÍCIA ENDO

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

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

Resultados e análises

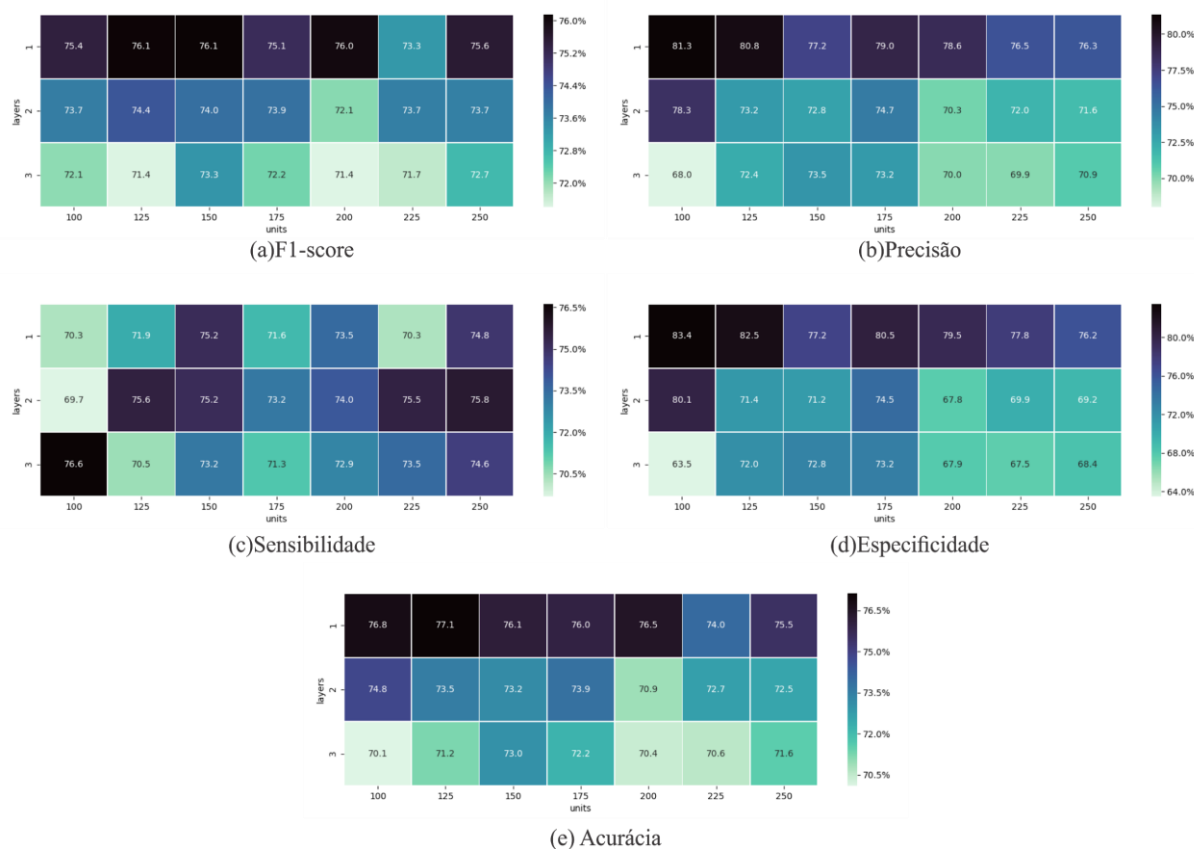
DL totalmente conectada

Os resultados para cada uma das métricas referentes às configurações do modelo de DL totalmente conectada estão presentes na Figura 6. Analisando os resultados de F1-score, duas configurações apresentaram o melhor resultado, 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 a mesma configuração também apresenta valores interessantes para as métricas de precisão (80,8%), acurácia (77,1%) e especificidade (82,5%). Já a configuração F-1-150 apresentou resultados um pouco inferiores que a F-1-125 nas métricas de acurácia (76,1%), precisão (77,2%) e especificidade (77,2%).

ANÁLISE DE MODELOS DE *DEEP LEARNING* PARA AUXÍLIO NO PÓS-DIAGNÓSTICO DE TUBERCULOSE

THOMÁS OLIVEIRA; ELISSON ROCHA; THEO LYNN; VANDERSON SAMPAIO E PATRÍCIA ENDO

Figura 6. Resultados do *Grid-search* para as configurações do modelo de DL totalmente conectada



Fonte.: Próprio Autor

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 é a configuração com 3 camadas e 100 neurônios (denominada F-3-100), com 76,6%, todavia esta 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).

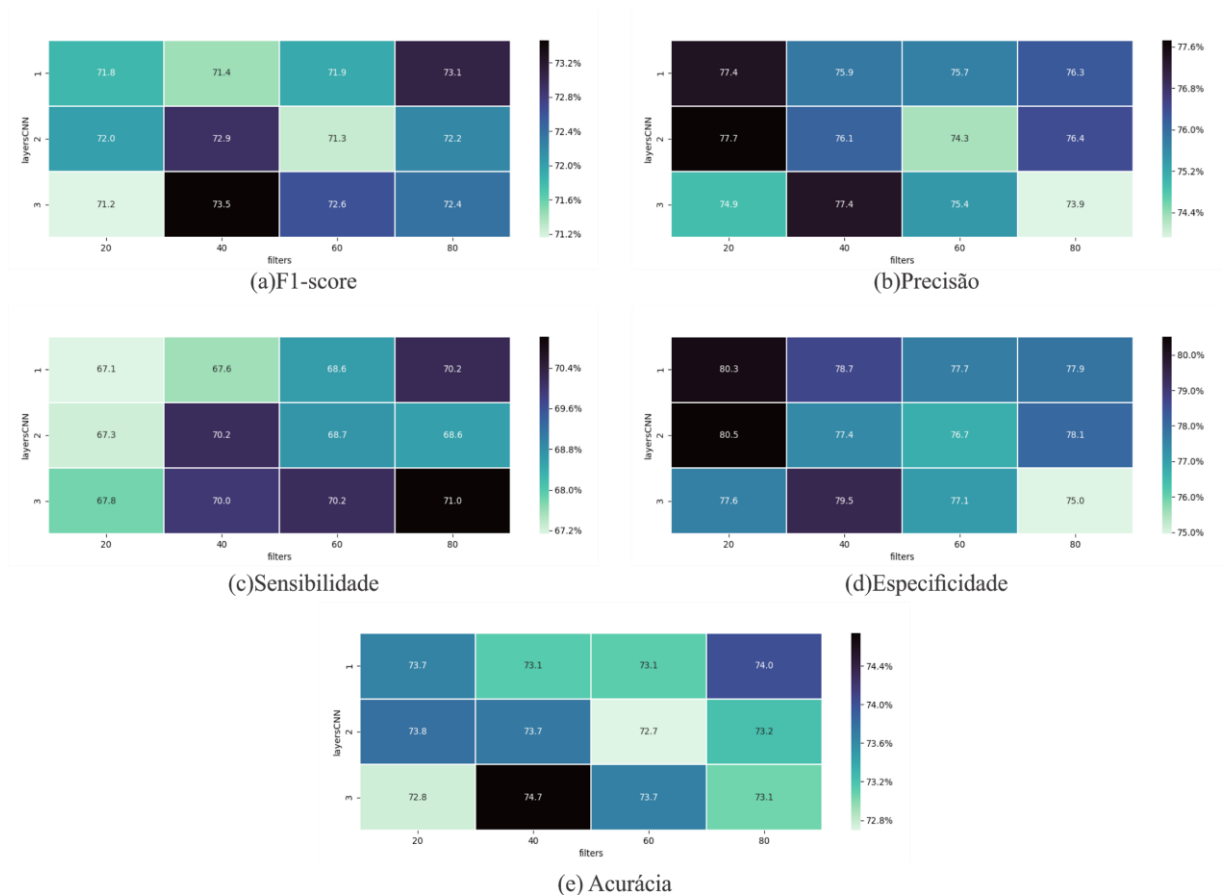
ANÁLISE DE MODELOS DE *DEEP LEARNING* PARA AUXÍLIO NO PÓS-DIAGNÓSTICO DE TUBERCULOSE

THOMÁS OLIVEIRA; ELISSON ROCHA; THEO LYNN; VANDERSON SAMPAIO E PATRÍCIA ENDO

CNN

A Figura 7 mostra os resultados do *Grid-search* do modelo de CNN. Uma configuração se destacou nos resultados de F1-score, 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.

Figura 7. Resultados do *Grid-search* dos modelos de CNN



Fonte.: Próprio Autor

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%).

ANÁLISE DE MODELOS DE *DEEP LEARNING* PARA AUXÍLIO NO PÓS-DIAGNÓSTICO DE TUBERCULOSE

THOMÁS OLIVEIRA; ELISSON ROCHA; THEO LYNN; VANDERSON SAMPAIO E PATRÍCIA ENDO

A configuração que possui a melhor métrica para sensibilidade é a configuração com 3 camadas e 80 filtros (denominada C-3-80), com uma taxa de 71%; esta configuração também apresenta valores semelhantes para as demais métricas.

Discussão comparativa entre os modelos

A Tabela 3 apresenta os valores das melhores configurações dos modelos da DL totalmente conectada e da CNN.

Tabela 3. Melhores configurações dos modelos

Configuração	F1-Score	Precisão	Sensibilidade	Especificidade	Acurácia
F-1-100	75,4%	81,3%	70,3%	83,4%	76,8%
F-1-125	76,1%	80,8%	71,9%	82,5%	77,1%
F-3-100	72,1%	68%	76,6%	63,5%	70,1%
C-2-20	72%	77,7%	67,3%	80,5%	73,8%
C-3-40	73,5%	77,4%	70%	79,5%	74,7%
C-3-80	72,4%	73,9%	71%	75%	73,1%

Analisando as configurações, pode-se notar que os modelos de DL totalmente conectada obtiveram os melhores resultados em comparação aos modelos CNN. Contudo, vale salientar que apesar de possuir valores mais baixos, os resultados são próximos.

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 melhor taxa maior de sensibilidade foi a F-3-100, alcançando 76,6%, porém o mesmo modelo apresentou uma

ANÁLISE DE MODELOS DE *DEEP LEARNING* PARA AUXÍLIO NO PÓS-DIAGNÓSTICO DE TUBERCULOSE

THOMÁS OLIVEIRA; ELISSON ROCHA; THEO LYNN; VANDERSON SAMPAIO E PATRÍCIA ENDO

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.

Para melhor entender as configurações, as Tabelas de 4 a 9 apresentam as matrizes de confusão dessas configurações.

Tabela 4. Matriz de confusão da configuração F-1-100

	Predição Óbito	Predição Não Óbito
Óbito	218	92
Não Óbito	50	252

Tabela 5. Matriz de confusão da configuração F-1-125

	Predição Óbito	Predição Não Óbito
Óbito	223	87
Não Óbito	53	249

Tabela 6. Matriz de confusão da configuração F-3-100

	Predição Óbito	Predição Não Óbito
Óbito	236	72
Não Óbito	111	193

Tabela 7. Matriz de confusão da configuração C-2-20

ANÁLISE DE MODELOS DE *DEEP LEARNING* PARA AUXÍLIO NO PÓS-DIAGNÓSTICO DE TUBERCULOSE

THOMÁS OLIVEIRA; ELISSON ROCHA; THEO LYNN; VANDERSON SAMPAIO E PATRÍCIA ENDO

	Predição Óbito	Predição Não Óbito
Óbito	209	91
Não Óbito	61	249

Tabela 8. Matriz de confusão da configuração C-3-40

	Predição Óbito	Predição Não Óbito
Óbito	217	91
Não Óbito	69	234

Tabela 9. Matriz de confusão da configuração C-3-80

	Predição Óbito	Predição Não Óbito
Óbito	218	92
Não Óbito	73	228

Confirmando os resultados anteriores, através das matrizes de confusão, pode-se observar que a configuração que obteve um acerto maior da classe de óbito é a configuração F-3-100, com 238 acertos. Em relação a taxa de acertos de não óbito, a configuração que conseguiu o maior valor é a F-1-100 com 252 acertos, e também é a configuração com menor taxa de erro da classe de não óbito, com 50 erros. Analisando a taxa de erros da classes de óbito, a configuração que apresenta o menor valor é a configuração F-3-100, com 72 erros.

A configuração F-1-100 apresenta a melhor taxa de acerto (252) e menor taxa de erro (50) da classe de não óbito, o que a torna uma configuração interessante e confiável

ANÁLISE DE MODELOS DE *DEEP LEARNING* PARA AUXÍLIO NO PÓS-DIAGNÓSTICO DE TUBERCULOSE

THOMÁS OLIVEIRA; ELISSON ROCHA; THEO LYNN; VANDERSON SAMPAIO E PATRÍCIA ENDO

para prever os casos de TB que necessitam de tratamentos mais simples. Porém, esta mesma configuração apresenta a maior taxa de erro para classe de óbito (92). Para o objetivo deste trabalho, é interessante que a taxa de erro para a classe de óbito seja a menor possível, pois essa taxa indica pacientes que precisam de um tratamento mais intensivo de TB.

A configuração F-3-100 apresenta a maior taxa de acerto para a classe de óbito (236), e a menor taxa de erro para a classe de não óbito (72), o que a torna uma configuração interessante para prever casos de TB que necessitam de um tratamento mais intensivo. Contudo, esta mesma configuração apresenta a maior taxa de erro (111) da classe de não óbito. Quanto maior essa taxa de erro, maior será o desperdício de recursos do governo na área de saúde para tratar os pacientes, já que estes pacientes poderiam ter recebido um tratamento mais simples e que utilizasse menos recursos. Além disso, esta configuração também mostrou a menor taxa de acerto (193) da classe de não óbito, não sendo interessante para prever os casos de pacientes de TB que necessitam de tratamento mais simples.

Conclusão 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 causas 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).

A partir dos resultados obtidos, percebeu-se que não há uma única configuração que atinja os melhores resultados para todas as métricas. Assim, como próximos passos, planeja-se realizar melhorias nos modelos propostos. Uma abordagem bem-sucedida é

ANÁLISE DE MODELOS DE *DEEP LEARNING* PARA AUXÍLIO NO PÓS-DIAGNÓSTICO DE TUBERCULOSE

THOMÁS OLIVEIRA; ELISSON ROCHA; THEO LYNN; VANDERSON SAMPAIO E PATRÍCIA ENDO

treinar diferentes modelos, em vez de um único, e combinar as predições desses modelos. Isso é chamado *ensemble learning*, e pode resultar em predições melhores.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, da Fundação de Amparo a Ciência e Tecnologia de Pernambuco (FACEPE) - Pernambuco - Código de Financiamento IBPG-0059-1.03/19 e do Irish Institute of Digital Bussiness (dotLAB), Irlanda.

APÊNDICE A

Preditores selecionados
Sexo
Raça
Radiografia do Tórax
Teste de TB
Forma da TB (Pulmonar, Extrapulmonar, Pulmonar +Extrapulmonar)
Tipo de TB Extrapulmonar
Agravo por Aids
Agravo por Álcool
Agravo por Diabetes
Agravo por Doença Mental
Agravo por outra doença
Diagnóstico de Baciloscopia de Escarro
Resultado da Cultura de Escarro
Resultado do Exame de HIV

ANÁLISE DE MODELOS DE *DEEP LEARNING* PARA AUXÍLIO NO PÓS-DIAGNÓSTICO DE TUBERCULOSE

THOMÁS OLIVEIRA; ELISSON ROCHA; THEO LYNN; VANDERSON SAMPAIO E PATRÍCIA ENDO

Resultado do exame histopatológico para diagnóstico de TB
Agravo por Drogas ilícitas
Agravo por Tabaco
Resultado do Teste de Sensibilidade
Resultado do exame do Teste Molecular Rápido para TB

Referências

- BERGSTRA, James; BENGIO, Yoshua. Random search for hyper-parameter optimization. **Journal of Machine Learning Research**, v. 13, n. Feb, p. 281-305, 2012.
- BULBUL, Halil Ibrahim; UNSAL, Özkan. Comparison of classification techniques used in machine learning as applied on vocational guidance data. In: **2011 10th International Conference on Machine Learning and Applications and Workshops**. IEEE, 2011. p. 298-301.
- BUITINCK, Lars et al. API design for machine learning software: experiences from the scikit-learn project. **arXiv preprint arXiv:1309.0238**, 2013.
- CAO, Yu et al. Improving tuberculosis diagnostics using deep learning and mobile health technologies among resource-poor and marginalized communities. In: **2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)**. IEEE, 2016. p. 274-281.
- CARNEIRO, Gustavo et al. Automated 5-year mortality prediction using deep learning and radiomics features from chest computed tomography. In: **2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)**. IEEE, 2017. p. 130-134.
- FREEMAN, Ion et al. What are they Researching? Examining Industry-Based Doctoral Dissertation Research through the Lens of Machine Learning. In: **2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)**. IEEE, 2018. p. 1338-1340.

ANÁLISE DE MODELOS DE *DEEP LEARNING* PARA AUXÍLIO NO PÓS-DIAGNÓSTICO DE TUBERCULOSE

THOMÁS OLIVEIRA; ELISSON ROCHA; THEO LYNN; VANDERSON SAMPAIO E PATRÍCIA ENDO

GOLDMAN, Lee; SCHAFER, Andrew I. **Goldman-Cecil Medicine E-Book**. Elsevier Health Sciences, 2015.

HIJJAR, Miguel Aiub et al. Epidemiologia da tuberculose: importância no mundo, no Brasil e no Rio de Janeiro. **Pulmão RJ**, v. 14, n. 4, p. 310-4, 2005.

HOODA, Rahul et al. Deep-learning: A potential method for tuberculosis detection using chest radiography. In: **2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)**. IEEE, 2017. p. 497-502.

KANT, Sonaal; SRIVASTAVA, Muktabh Mayank. Towards automated tuberculosis detection using deep learning. In: **2018 IEEE Symposium Series on Computational Intelligence (SSCI)**. IEEE, 2018. p. 1250-1253.

KHAN, Samir; YAIRI, Takehisa. A review on the application of deep learning in system health management. **Mechanical Systems and Signal Processing**, v. 107, p. 241-265, 2018.

LAKHANI, Paras; SUNDARAM, Baskaran. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. **Radiology**, v. 284, n. 2, p. 574-582, 2017.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. **nature**, v. 521, n. 7553, p. 436-444, 2015.

LEE, Christine K. et al. Development and validation of a deep neural network model for prediction of postoperative in-hospital mortality. **Anesthesiology: The Journal of the American Society of Anesthesiologists**, v. 129, n. 4, p. 649-662, 2018.

LIU, Xu-Ying; WU, Jianxin; ZHOU, Zhi-Hua. Exploratory undersampling for class-imbalance learning. **IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)**, v. 39, n. 2, p. 539-550, 2008.

RAJARAMAN, Sivaramakrishnan et al. A novel stacked generalization of models for improved TB detection in chest radiographs. In: **2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)**. IEEE, 2018. p. 718-721.

SCHWING, Alexander G.; URTASUN, Raquel. Fully connected deep structured networks. **arXiv preprint arXiv:1503.02351**, 2015.

ANÁLISE DE MODELOS DE *DEEP LEARNING* PARA AUXÍLIO NO PÓS-DIAGNÓSTICO DE TUBERCULOSE

THOMÁS OLIVEIRA; ELISSON ROCHA; THEO LYNN; VANDERSON SAMPAIO E PATRÍCIA ENDO

SETHI, Khushal; PARMAR, Vivek; SURI, Manan. Low-Power Hardware-Based Deep-Learning Diagnostics Support Case Study. In: **2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)**. IEEE, 2018. p. 1-4.

SIMONYAN, Karen; ZISSERMAN, Andrew. Very deep convolutional networks for large-scale image recognition. **arXiv preprint arXiv:1409.1556**, 2014.

SINANWEB. Disponível em <<http://portalsinan.saude.gov.br>> Acesso em 03 de nov. de 2019.

SONG, Qin et al. An evolutionary deep neural network for predicting morbidity of gastrointestinal infections by food contamination. **Neurocomputing**, v. 226, p. 16-22, 2017.

STIRENKO, Sergii et al. Chest X-ray analysis of tuberculosis by deep learning with segmentation and augmentation. In: **2018 IEEE 38th International Conference on Electronics and Nanotechnology (ELNANO)**. IEEE, 2018. p. 422-428.

TAKAHASHI, Kei et al. Development of generic CNN deep learning method using feature graph. In: **2018 Sixth International Symposium on Computing and Networking Workshops (CANDARW)**. IEEE, 2018. p. 235-238.

WORLD HEALTH ORGANIZATION. Tuberculosis 2019. Disponível em <<https://www.who.int/news-room/fact-sheets/detail/tuberculosis>> Acesso em nov. de 2019.

WORLD HEALTH ORGANIZATION. Global tuberculosis report 2018. Disponível em: <https://www.who.int/tb/publications/global_report/en/Fullreport.pdf> Acesso em nov. de 2019.