

CNN Approach for CT Scan Exam Classification

1

2 **Resumo:** Antes da pandemia do coronavírus de 2020. Segundo a Agência Nacional de Saúde
3 Suplementar (ANS), as doenças respiratórias causam a maioria das internações e levam a 506,9
4 mil internações, o que representa 5,9 % das causas. A classificação de alto risco para infecções
5 pulmonares com tomografia computadorizada está sendo implementada como ferramenta para
6 auxiliar no diagnóstico de doenças como câncer, tuberculose, COVID-19, etc. Para diagnósticos
7 de imagens, a interpretação muitas vezes sofre pela subjetividade e experiência do especialista,
8 o que pode fornecer um diagnóstico com viés. Com isso milhões de testes serão feitos, por isso
9 há muito interesse em desenvolver algoritmos computacionais para otimizar a classificação dos
10 pacientes de forma precisa e rápida. Um primeiro passo vital na análise de tomografias com-
11 putadorizadas de câncer de pulmão e na detecção de nódulos pulmonares, que podem ou não
12 representar câncer de pulmão no estágio inicial. O objetivo deste trabalho é apresentar o desen-
13volvimento de um sistema que permite a detecção de nódulos cancerígenos, com base no exame
14 tomográfico computadorizado, por meio de uma rede neural convolucional. A intenção desse
15 sistema é agilizar a triagem hospitalar e oferecer assertividade na classificação dos nódulos e de
16 COVID-19.

17

18 **Palavras-chave:** Exames pulmonares, Tomografia computadorizada, Rede eural convolucional
19 (CNN), Aprendizagem por transferência.

20

21

22 **Abstract:** Before the coronavirus 2020 pandemic starts, respiratory diseases cause the majority
23 of hospitalizations and lead to 506.9 thousand hospitalizations, which represents 5.9% of the
24 causes, according to the National Supplementary Health Agency (ANS). The high-risk classifi-
25 cation for lung infections with computed tomography scans is being implemented as a tool to
26 aid in the diagnosis of diseases such as cancer, tuberculosis, COVID-19, etc. For visual diag-
27 nostics of images, interpretation often suffers, permeated by the subjectivity and experience of
28 the specialist, which can provide a diagnosis with bias. With that millions of tests will be done,
29 so there is a lot of interest in developing computer algorithms to optimize the classification of
30 patients accurately and quickly. A vital first step in the analysis of lung cancer CT scans and
31 in the detection of lung nodules, which may or may not represent lung cancer at an early stage.
32 The aim of this work is to present the development of a system that allows the detection of
33 cancerous nodules, based on computed tomographic examination, through a convolutional and
34 anterior neural network. The intention of this system is to speed up hospital screening and offer
35 assertiveness in the classification of nodules and COVID-19.

36

37 **Keywords:** Lung procedures, Computed tomography, Convolutional neural network (CNN),
38 Transfer learning.

39 1 Introdução

40 O Conselho Federal de Medicina reconhece a radiologia e diagnóstico por imagem como
41 uma especialidade. Com isso, os equipamentos que geram imagens para diagnóstico têm sofrido
42 grandes avanços (STORM, et. al., 2020). E aos poucos, diferentes tipos de imagens médicas
43 foram desenvolvidas. As análises clínicas por imagem desempenham um papel importante na
44 antecipação de tratamentos e cuidado com a saúde, que vão além do diagnóstico como prevenção
45 e terapia (MANSOOR, et. al., 2020).

46 A rede neural convolucional (CNN) vem demonstrando altos níveis de desempenho na extra-
47 ção de características de um imagem, e assim tornado possível a classificação através de pequena
48 diferença entre as imagens médicas. Podemos ver os resultados de modelos CNN em WANG S.,
49 onde ele retrata sobre a extração de características de imagens; KRUTHIKA, em seu trabalho
50 sobre de detecção de alzheimer; em YILDIRIM, a partir de espectro de frequência cardíaca faz
51 a detecção de diabetes; e até mesmo para doenças recém descobertas como COVID-19 onde
52 ISLAM, usa imagens de raio-X para a detecção da doença.

53 O objetivo deste trabalho é desenvolver um modelo baseada em cnn capaz de ser utilizado
54 em um cenário supervisionado, onde os preditores estão disponíveis e também há um responsável
55 por guiar a análise e assim detectar nódulos cancerígenos no pulmão no pulmão e COVID-19 a
56 partir de imagens de tc.

57 2 Trabalhos Relacionados

58 Com o objetivo de entender como o COVID-19 e cancer no pulmão são detectados por meio
59 da TC do tórax, foi realizada uma busca na literatura para entender quais padrões caracterizam
60 essas doenças e as principais diferenças dos padrões encontrados, bem como, a busca pelo enten-
61 dimento do padrão evolutivo da doença e sua severidade conforme levantado por grande parte
62 dos trabalhos analisados. Na Figura 1 é possível visualizar a diferenças entre os exames de
63 tomografica e a identificação de cada doença.



Figura 1: Caracterização do COVID-19 e câncer pulmonar por meio da TC do tórax.

64 Trabalhos recentes (KOSTIS, et. al., 2003, QUISPE, et. al., 2020, SILVA, et. al., 2020,
65 GIETEMA, et. al., 2007, SETIO, et. al., 2016, SHEN, et. al., 2015, ZHU, et. al., 2018)
66 observam padrões de imagem na tomografia computadorizada em pacientes diagnosticados com
67 câncer no pulmão. A partir destas pesquisas foi observado que o uso de exames de TC para
68 rastrear pessoas com maior risco de câncer de pulmão salvou mais vidas em comparação com os
69 raios-X de tórax. Segundo a American Cancer Society, para pessoas de alto risco, fazer exames
70 anuais de TC antes do início dos sintomas ajuda a diminuir o risco de morte por câncer de pulmão.
71 Outras pesquisas mostram que com os mesmos métodos utilizado em classificação imagens com
72 câncer no pulmão pode ser utilizado em exames de COVID-19 como vemos nos trabalhos de (AI,
73 et. al., 2019, BUKHARI, et. al., 2020, GOZES, et. al., 2020, LI, et. al., 2007, MISHRA, et. al.,
74 2020).

75 Embora os modelos concentram-se em tarefas específicas, eles seguem linhas semelhantes de
76 desenvolvimento. A segmentação de dados são amplamente usados como parte dessas abordagens
77 para evitar sobreajustes. Além dos métodos tridimensionais (3D-CNN) para a segmentação
78 baseados em morfologia, análise e caracterização de pequenos nódulos, (KOSTIS, et. al., 2003,
79 QUISPE, et. al., 2020, ZHU, et. al., 2018,) demonstram uma acuracidade de mais ou menos 85%
80 de detecção de todos os nódulos. Além do básico método baseado em transformação que inclui
81 rotação, inversão, escala e com uma 2D-CNN usada em (SILVA, et. al., 2020, GIETEMA, et. al.,
82 2007), demonstraram que é possível reduzir a computação e manter as acurácias encontradas em
83 modelos mais complexos. Depois em sistema de detecção assistida por computador para nódulos
84 pulmonares usando redes convolucionais multi-view, (SETIO, et. al., 2016) onde é gera uma
85 modelo tridimensional próprio. Também foi encontrado artigos que tratam a heterogeneidade
86 do nódulo em imagem de TC, (SHEN, et. al., 2015) o que pode ser indicado ao modelo durante
87 o treinamento. E sistema automático “DeepLung” para a detecção de nódulos cancerígenos no
88 pulmão desenvolvido por (ZHU, et. al., 2018).

89 Como existem as mais variadas arquiteturas, neste tópico é realizado o aprofundamento e
90 discussões apenas nas Arquiteturas VGG-19, Inception V3, Xception e ResNet-50. Tal escolha,
91 parte do princípio que estas arquiteturas apresentaram um número maior de trabalhos aplicados
92 em relação às demais.

93 3 Fundamentação Teórica

94 A *cnn* é uma classe de rede neural utilizada para processamento e análise de imagens. Ela
95 foi proposta em 1998 por Yann LeCun, que é uma arquitetura capaz de reconhecer dígitos
96 manuscritos com precisão de 99,2%. Essa arquitetura foi inspirada em uma pesquisa de David
97 Hunter Hubel e Torsten Wiesel (1968) sobre o funcionamento do córtex visual dos mamíferos
98 (LECUN, et. al., 1999).

99 As CNN são compostas por diversas camadas que utilizam a operação de convolução para
100 realizar extração de características (SIMONYAN, et. al., 2015). Esta operação é realizada a
101 partir de uma janela de dados deslizante, chamada de filtro convolutivo, que percorre toda a
102 entrada da rede e opera de forma análoga à sobreposição dos campos receptivos. Sua arquitetura
103 consiste em convolução, dense (neurônios), pooling (acumuladores), flatten (mapeamento de
104 características) e uma camada totalmente conectada (SZEGEDY, et. al., 2020).

105 A rede neural convolucional (CNN) é uma das principais categorias de redes neurais para
106 classificar imagens. Entre as arquiteturas de CNNs, foram selecionadas quatro arquiteturas,
107 sendo elas: VGG-19, Inception V3, Xception e Resnet 50, para a extração de características
108 exames pulmonares por tc. Cada imagem passa por uma série de camadas de convoluções com
109 filtros, pooling, conexão e aplicando a função softmax para classificar o objeto.

110 Para implementar modelos de *cnn*, existem algumas estruturas para facilitar a criação da
111 mesma. Dessa forma, temos o Tensor Flow, que é uma biblioteca de código aberto para compu-
112 tação numérica e aprendizado de máquina em grande escala durante a execução desses aplicativos
113 em C++ de alto desempenho desenvolvida pela Google.

114 3.1 Transfer Learning

115 Para acelerar o processo, o modelo utilizado foi o Transfer Learning (transferência de aprendi-
116 zado). O Transfer Learning surgiu no desafio de imagem da Imagenet que possui 12 milhões de
117 imagens e o objetivo era categorizar mais de mil classes diferentes. E aos poucos foram criadas
118 as diversas redes neurais depois desse desafio. As arquiteturas de CNN são um conjunto de dados
119 multimodal que contém vinte e três classes e quatro modalidades, incluindo (Tomografia Com-
120 putadorizada (TC), Imagem por Ressonância Magnética (MRI), Mamografia (MG) e Tomografia
121 por Emissão de Pósitrons (PET)) são usados para demonstrar nosso método (AYYACHAMY,

122 et. al., 2019). Sendo assim, os modelos de transfer learning são modelos que herdam parâmetros
123 de uma rede neural já treinada (SILVA, et. al., 2020).

124 E dentre as diversas arquiteturas disponibilizadas, este artigo faz uso da:

- 125 • VGG-19 proposta por Simonyan e Zisserman do Visual Geometry Group at University of
126 Oxford, é considerada uma das mais populares redes e vem demonstrando eficiência. O
127 principal objetivo dessa arquitetura é a ideia de aumentar a profundidade da CNN. Ela
128 substitui grandes kernel por kernels menores o que a torna mais precisa na realização de
129 tarefas de visão computacional (MISHRA, et. al., 2020).
- 130 • Inception V3, A arquitetura Inception V3 proposta por Karen Simonyan e desenvolvida
131 pela google tem uma camada de normalização em lote é usada extensivamente em todo o
132 modelo e aplicada às entradas de ativação com o objetivo de suavizar caracteres da imagem
133 e propagar informações para as camadas mais profunda (SZEGEDY, et. al., 2020).
- 134 • Xception, A arquitetura Xception proposta por François Chollet, é uma versão extrema da
135 Inception V3 criada pela Google, que possui convoluções separadas e início com a máxima
136 profundidade, o que a torna mais eficiente no uso dos parâmetros. A arquitetura Xception
137 tem o mesmo número de parâmetros que a Inception (SZEGEDY, et. al., 2017).
- 138 • Resnet 50, A arquitetura ResNet proposta por Kaiming He, representa uma estrutura de
139 aprendizagem residual para facilitar o treinamento de redes que são substancialmente mais
140 profundas do que aquelas usadas anteriormente (HE, et. al., 2016). Segundo HE, quando
141 a rede neural tem muitas camadas ela rapidamente se degrada, ou seja, ela durante o
142 treinamento os sistemas não utilizam de forma igualitária.

143 3.2 Ensemble

144 A aprendizagem ensemble é um paradigma de aprendizagem de máquina em que vários mo-
145 delos são treinados para resolver o mesmo problema. Em contraste com as abordagens de aprendi-
146 zado de máquina comuns, que tentam aprender uma hipótese a partir de dados de treinamento,
147 os métodos de conjunto tentam construir um conjunto de hipóteses e combiná-las para usar com
148 resultado final (ZHOU, et. al., 2009).

149 Para a aplicação no mundo real, os métodos de ensemble são amplamente usados por de-
150 senvolvedores de IA. Esse método é mais robusto do que os estimadores únicos, melhoraram a
151 generalização. Os modelos simples podem ser combinados para construir um modelo complexo,
152 que é computacionalmente rápido. Porém sua desvantagem é que se tivermos um classificador
153 subjacente enviesado, isso levará a um modelo à ampliar esse viés. Os métodos de ensemble são
154 considerados classificadores de alta qualidade.

155 3.3 Metrics

156 As métricas utilizadas para a análise do desempenho e eficiência do modelos foram.

157 A Acurácia, que é a maneira mais direta de avaliar um modelo de rede neural: fazemos uma
158 previsão, olhamos para o rótulo previsto e depois o comparamos com o treino ou teste. Se os
159 valores concordarem, o modelo acertou. Em seguida, podemos fazer isso para todos os dados que
160 temos e ver a proporção dos exemplos previstos corretamente (ORWANT, et. al., 1999). Segue
161 a Equação 1 para o cálculo da acurácia.

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

162 No entanto, às vezes, a Acurácia não é a melhor medida do desempenho do modelo. Em
163 alguns casos, os valores de precisão enganam: eles sugerem que o modelo é bom, embora não
164 seja, principalmente quando a base de dados não têm a mesma quantidade de amostras para

165 cada classe. Existem alternativas que podem superar esse problema. Outras opções de métricas
 166 são a precision e a recall. Segue na Equação 2 o cálculo da precisão.

$$precision = \frac{TP + TN}{TP + FP + TN + FN} \quad (2)$$

167 E o recall é dado pela fração de itens positivos preditos entre todos os itens que o modelo
 168 previu com falso negativo. Segue na Equação 3 o cálculo do Recall,

$$recall = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

169 E combinando a precision e recall temos a métrica F1 de modo a trazer um número único
 170 que indique de forma mais genérica e verifica, se o modelo está bom e o quão bom ele trabalha
 171 até com conjuntos de dados que possuem classes desproporcionais, portanto, mesmo que haja
 172 um grande número de exemplos negativos, ainda podemos realizar uma avaliação significativa do
 173 modelo. Segue na Equação 4 o cálculo do F1.

$$F1 = \frac{TP + TN}{TP + FP + TN + FN} \quad (4)$$

174 4 Metodologia

175 Por meio da Figura 2 é apresentado as etapas de treinamento do modelo. A fase de pré-
 176 processamento, é feito um tratamento sobre as informações do banco de dados. Esse tratamento
 177 consiste em organizar as imagens por paciente com seus respectivos targets.

178 Feito o tratamento dos dados foi realizada a primeira divisão, treinamento, para o ajuste
 179 do modelo, e a outra parte para o teste para a generalização e assim verificar se um modelo
 180 apresenta boa performance. A aprendizagem com o dataset sempre há o risco de sobreajuste. O
 181 sobreajuste ocorre quando o modelo começa a “adivinhar” a classe em vez de detectar padrões.

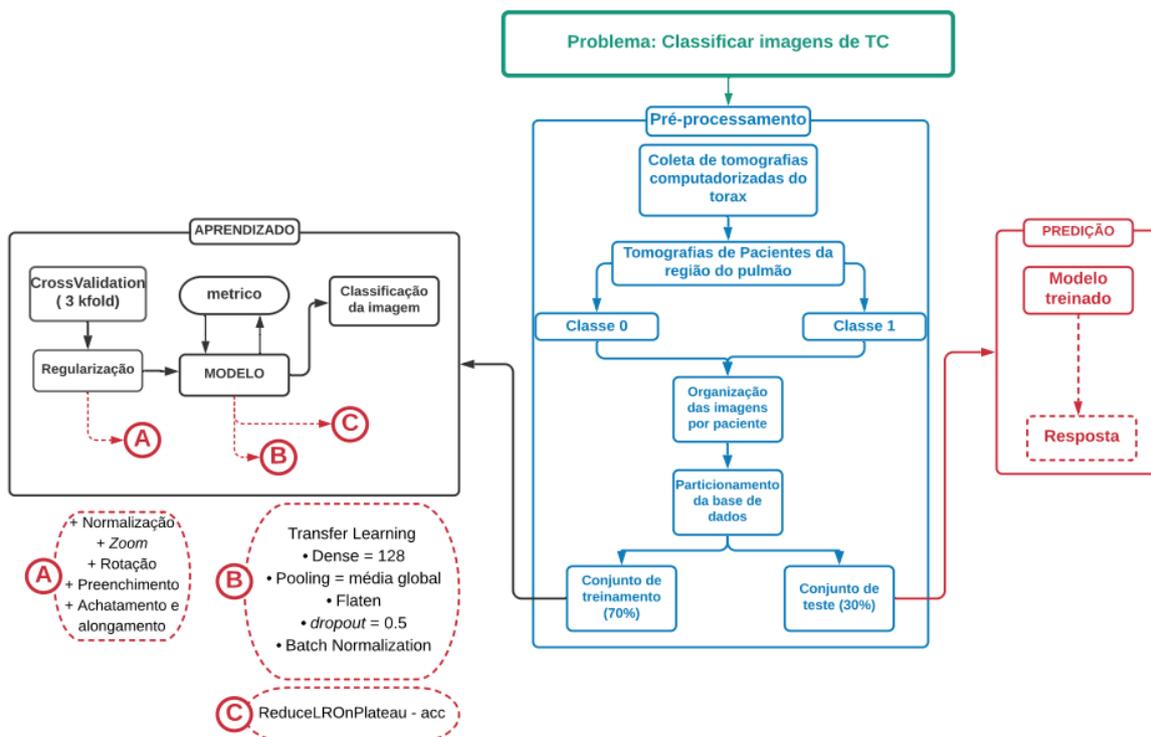


Figura 2: Diagrama da operacionalização dos dados.

182 A maneira mais prática de verificar se um modelo se adapta ou não é emular uma base de
183 dados, ou seja, separar parte dos dados disponíveis e não usá-las durante o treinamento, ou seja,
184 selecionar amostras do dado original e cria-se, de forma aleatória, dois conjuntos independentes.
185 Sendo assim, para a presente aplicação foi feita divisão de em dois do banco de dados, sendo
186 30% para teste e 70% para treinamento para que no fim uma deles seja utilizado para a avaliação
187 o modelo final.

188 Depois do pré-processo é iniciado o aprendizado. Agora os dados de treinamento precisam
189 de uma validação para poder encontrar o modelo mais generalizado. Esse método é conhecido
190 como validação cruzada K-fold. A validação cruzada fornece a estimativa de desempenho e a
191 verificação da propagação do erro (SCHREIBER, et. al., 2017). O K-fold ajuda a selecionar
192 modelos com melhor desempenho. Os dados são divididos em K partes, para cada uma dessas
193 partes, o modelo irá usar a parte K-1 para treinar, enquanto usará uma parte para validar. Ao
194 final do processo, quando o modelo iterar todas as k vezes, obtém-se um verdadeiro score de
195 como seu modelo está generalizando. Para a presente aplicação foi adotado k igual a 3. E para
196 a divisão entre os folds terem a mesma quantidade de cada classe é aplicado a estratificação
197 (ORWANT, et. al., 1999).

198 Outra técnica utilizada foi a argumentação de dados que tem como objetivo aumentar a
199 quantidade de dados com ligeiras modificações nos dados existentes (PEREZ, et. al., 2017).
200 Essa técnica funciona com regularizador, o qual irá reduzir o overfit ao executar o aprendizado
201 de máquina. As combinações de modificações realizadas no conjunto de treinamento foram:
202 Normalização dos valores dos pixels da imagem; zoom; rotação; preenchimento, achatamento
203 e alongamento. Contudo, o flip não é utilizado, nem horizontal e nem vertical, pois o pulmão
204 direito tem suas diferenças importantes com relação ao pulmão esquerdo (AI, et. al., 2019).

205 4.1 Modelo

206 Para a presente aplicação não foi incluída toda a rede neural, porque como as redes de transfer
207 learning possuem milhões de parâmetros treináveis, para reduzir o processo de computação será
208 treinado apenas as camadas mais profundas. As camadas iniciais aprendem com características
209 muito gerais e a medida que ela se aprofunda nas camadas tende a aprender padrões mais
210 específicos. Sendo assim, a partir do 5 bloco são adicionados novos blocos ao modelo para o
211 reconhecimento do COVID-19 e nódulos cancerígenos em imagens de TC do tórax.

212 O modelo conta com dense de 128 neurônios, depois adicionado a camada de pooling a qual irá
213 fazer a redução de de resolução de cada imagem pela média, também é adicionado a Flatten para
214 que a computação entenda melhor a imagem. Depois, para reduzir o sobreajuste foi utilizado
215 a técnica de método do dropout o qual irá desligar 5 neurônios e tentar treinar com esses 5
216 neurônios, e assim conseguir regularizar a rede neural. Também foi utilizado a técnica de Batch
217 Normalization que durante o treinamento é feito a normalização a cada passada por neurônios e
218 isso irá aumentar a estabilidade da rede neural.

219 E para a compilação do modelo é utilizado uma Cross Entropy, a qual tem o papel de atribuir
220 valores entre as classes, e o otimizador Adam que faz a descida gradiente estocástica mantém
221 uma única taxa de aprendizado cada vez que faz a atualização dos pesos da rede neural (BOCK,
222 et. al., 2018).

223 4.2 Implementação d Ensemble

224 Para obter uma acurácia melhor, a técnica de ensemble foi utilizada. No modelo foi aplicado
225 ao conjunto de predições e cada predição teve sua importância atribuída a uma peso e depois
226 aplicado a média ponderada. Esta combinação de preditores com pesos ajudam em problemas
227 de difícil segmentação ou de difícil classificação como em exames de TC (ZHOU, et. al., 2009).
228 Porém, existem centenas de combinações até encontrar o peso ideal para cada preditor. Para
229 isso, foi gerado um algoritmo automatizado que busca por pesos que realmente trará a melhor

230 acurácia. Isso pode gerar a exclusão de algum modelo, caso ele não esteja dando valor ao resultado
231 final.

232 Então, foi capturado todos os preditores em uma Tabela e foi colocada em um loop para
233 testar pesos de 0 a 0.5 para cada arquitetura. E entre eles, devem obter todas as combinações
234 somando um e colocá-los em três loops aninhados. E então, o valor máximo é dado como o valor
235 da acurácia final

236 5 Resultados dos experimentos

237 5.1 Base de dados

238 Esse trabalho tem duas bases de dados diferentes. O primeiro conjunto de dados de tomografia
239 computadorizada do tórax de paciente com câncer, está disponível em uma competição no site do
240 Kaggle, onde foi disponibilizado por Hamdalla F. Kareem. Nela temos o total de 1097 pacientes,
241 sendo classificadas em:

- 242 • 416 imagens de pacientes saudáveis
- 243 • 561 imagens de pacientes diagnosticados com câncer maligno
- 244 • 120 imagens de pacientes diagnosticados com câncer benigno,

245 Cada paciente tem 1 imagem, com o local exato da anomalia, todas as imagens estão em
246 formato .png, coletadas a partir de uma máquina de tomografia computadorizada.

247 Devido ao desbalanceamento dos dados de paciente com câncer benigno, o trabalho descon-
248 sidera essas imagens, ou seja, o treinamento foi realizado apenas com pacientes saudáveis e com
249 câncer maligno.

250 O segundo conjunto de dados de tomografia computadorizada do tórax foram coletados no
251 Hospital do Servidor Público Municipal de São Paulo (HSPM). A base foi disponibilizada no site
252 ResearchGate em maio de 2020. Nela temos o total de 210 pacientes sendo classificadas em:

- 253 • 80 pacientes com COVID-19,
- 254 • 80 pacientes com outras doenças (não foi especificada quais são as outras doenças)
- 255 • 50 pacientes com os pulmões saudáveis.

256 Os pacientes do conjunto de dados infectados por SARS-CoV-2 foram confirmados através
257 do teste rRT-PCR para COVID-19. Nesta base cada paciente tem em média 20 imagens em
258 formato de .png. Isso porque foi coletado o exame completo de TC.

259 E para a base de dados da COVID-19 foram desconsiderados os dados de pacientes com outras
260 doenças. Se as imagens fossem consideradas o modelo poderia gerar viés durante a análise, pois
261 na base de dados da COVID-19 não foram especificadas quais são as outras doenças, ou seja,
262 existe a possibilidade de ter imagens de pacientes com câncer.

263 5.2 Análise dos resultados

264 A aplicação desenvolvida foi testada a partir do conjunto de dados reservados para o teste, o
265 que possibilita a avaliação do modelo. A implementação foi realizada na plataforma do Google
266 Colab (COLAB, 2020), sendo possível o uso da GPU disponibilizada pela mesma e a biblioteca
267 do Keras (KERAS, 2021), a qual oferece diversas arquiteturas de transfer learning, sendo elas:
268 VGG-19, Inception, Xception e Resnet 50.

269 Todas as quatro arquiteturas foram modeladas com os mesmos parâmetros descritos na
270 Tabela 1. A escolha dos parâmetros foi baseada em testes até que o modelo atingisse o melhor

Tabela 1: Hiperparâmetros para a construção da rede neural classificadora.

Parâmetros	Valor
tamanho da imagem	224x224
<i>batch size</i>	32
<i>epochs</i>	100
<i>learning rate</i>	0,001
<i>loss</i>	binary crossentropy
<i>optimize</i>	adam

desempenho, e durante esse processo, foi observado que todos esses parâmetros têm influência ativa no processo de construção das arquiteturas. A Métrica utilizada para otimização e seleção dos melhores dos modelos durante o aprendizado foi a otimizado Adam, para medir o desempenho de acerto da predição de forma estocástica. E para evitar os mínimos locais foi adicionado o Reduce Plateau para minimizar a função objetivo.

A realização dos experimentos, como descrito previamente, foi feita utilizando o modelo pré-treinados que recebe como entrada imagens com dimensões fixas de 224 x 224 x 3 (largura, altura e profundidade, respectivamente).

Este conjunto foi formado por imagens originais da base que estão em escala de cinza. A fim de adequar à entrada da rede, foi necessária a conversão para RGB, inserindo imagens iguais em cada um dos três canais.

Para todos os conjuntos, foram realizados os seguintes procedimentos: redimensionamento das imagens para adequação à entrada da rede; pequenas alterações nas imagens para prevenir o sobreajuste; e a separação aleatória dos conjuntos em 70% das imagens para treinamento e 30% para teste. Para a realização das avaliações das bases de dados de Câncer de Pulmão e a de COVID-19. Dentro delas foram selecionados pares de classes para a análise dos experimentos. Sendo eles:

- Câncer Maligno versus Saudáveis;
- COVID-19 versus Saudáveis;
- Câncer Maligno versus COVID-19 versus Saudáveis.

5.3 Experimento com o conjunto de dados Câncer

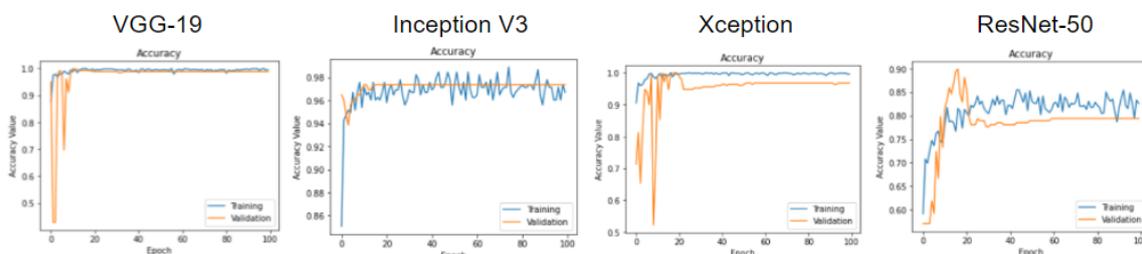


Figura 3: Gráficos de treinamento de máquina do experimento Câncer versus Saudável.

No primeiro experimento, o modelo foi treinado com o conjunto de imagens Cancer versus Saudáveis e testada com cada um das arquiteturas já citadas. A Tabela 2 exhibe os resultados obtidos.

A partir desses resultados pode-se afirmar que, o modelo conseguiu aprender com o conjunto de dado de câncer no pulmão. Como podemos ver as curvas de aprendizado de cada arquitetura no Gráfico 3. A VGG 19 apresentou o melhor desempenho com uma acurácia de 99,11%. Isso mostra

Tabela 2: Resultado de cada arquitetura para a base de dados de Câncer.

	VGG-19	Inception V3	Xception	Resnet 50
Acurácia	0,9911	0,9734	0,9823	0,7433
Precisão	0,9912	0,9735	0,9824	0,7838
F1	0,9611	0,9734	0,9822	0,7260

298 que a escolha da arquitetura pode influenciar nos resultados finais. E a com pior desempenho
299 foi a ResNet 50.

300 Quando aplicado o método de ensemble podemos ver a robustez uma pequena melhora na
301 eficiência do modelo, agora com uma acurácia de 99,55%, como mostra na Tabela 2. E se
302 comparamos as curvas de aprendizado entre o treinamento e o teste, vemos que o resultado se
303 tornou mais generalizado. E outro detalhe é que, os dois piores modelos, ResNet e Xception,
304 foram ignorados pelo ensemble por não trazer melhora no modelo geral, como podemos ver na
305 Tabela 3.

Tabela 3: Pesos de cada arquitetura, para o conjunto de dados Câncer.

	VGG-19	Inception V3	Xception	Resnet 50
Pesos	0,1	0,1	0	0

306 O resultado final para esse experimento foi de uma acurácia de 99,55% de acurácia com
307 aplicação de ensemble. Por mais que a melhora seja mínima.

308 5.4 Experimento com o conjunto de dados COVID-19

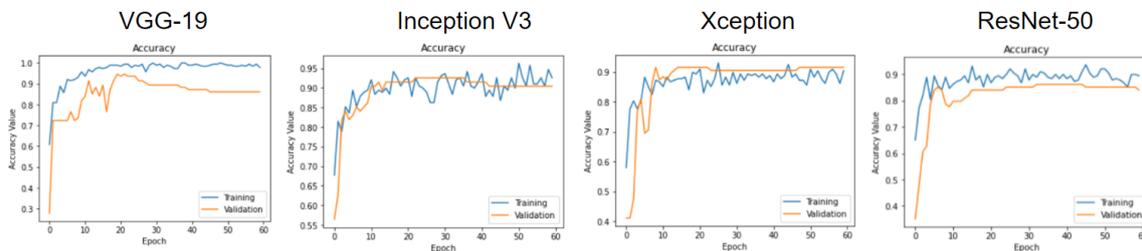


Figura 4: Gráficos de treinamento de máquina do experimento COVID-19 versus Saudável.

309 Para o segundo experimento, o modelo foi treinado com as 4 arquiteturas já citadas. E no
310 Gráfico 4 vemos as curvas de aprendizado de cada modelo seguindo com ruídos no início das
311 interações, porém, após algumas interações o aprendizado começa a ser estabilizado. O conjunto
312 de teste demonstrou uma acurácia de 93,74% com o conjunto de imagens de COVID-19 e testada
313 com a mesma base de dados. Na Tabela 4 exibe os resultados obtidos neste experimento.

314 Mais uma vez temos a VGG 19 com o melhor desempenho, com uma acurácia de 93,74%.

315 Após o treinamento das redes pré-treinadas, VGG-19, Resnet 50, Inception V3 e Xception, foi
316 implementado o método de ensemble, onde era realizado a média ponderada entre as previsões
317 das redes neurais demonstra uma precisão notavelmente eficaz. Na Tabela 5, há os pesos de cada
318 arquitetura para a formação do ensemble. O resultado final para esse experimento foi de uma
319 acurácia de 95,05% de acurácia com aplicação de ensemble. A partir desses resultados, pode-se
320 afirmar que ao aplicar o ensemble, o modelo se tornou mais robusto e consistente, o que melhora
321 da confiabilidade do modelo. E dessa vez, o modelo ignorado foi o Xception.

Tabela 4: Resultado de cada arquitetura para a base de dados de COVID-19.

	VGG-19	Inception V3	Xception	Resnet 50
Acurácia	0,9374	0,9214	0,9157	0,8404
Precisão	0,9417	0,9325	0,9156	0,8774
F1	0,9357	0,9254	0,9136	0,8473

Tabela 5: Pesos de cada arquitetura, para o conjunto de dados COVID-19.

	VGG-19	Inception V3	Xception	Resnet 50
Pesos	0,5	0,1	0	0,1

322 5.5 Experimentos com o conjunto de dados misto

323 E por fim o último experimento realizado neste trabalho, se repete os mesmo cenários de
 324 testes e análises. Dessa forma no Gráfico 5 podemos ver a curva de aprendizado para a base
 325 de dados mista teve muita variação. Isso ocorreu porque quando o pulmão está comprometido,
 326 tanto por COVID como por Câncer no pulmão, existe uma grande similaridade nas imagens. Na
 327 Tabela 8 podemos ver os valores das predições de cada arquitetura.

Tabela 6: Resultado de cada arquitetura para a base de dados de câncer no pulmão e COVID-19 e saudáveis.

	VGG-19	Inception V3	Xception	Resnet 50
Acurácia	0,9921	0,9764	0,9738	0,8403
Precisão	0,9923	0,9770	0,9745	0,8641
F1	0,9921	0,9765	0,9739	0,8291

328 A partir desses resultados pode-se dizer que para a identificação de imagens de TC do tórax
 329 com alguma doença, a classificação começa a ter dificuldade de definir um padrão para a classi-
 330 ficação das imagens. Isso ocorreu devido a semelhança entre as imagens. Da mesma forma que
 331 os experimentos um e dois foi aplicado o método de ensemble, apresentado na Tabela 6.

332 O resultado final para esse experimento foi de uma acurácia de 99,73%. Mas com a Tabela
 333 7, é possível concluir que o ensemble realiza um compartilhamento positivo, entre todas as
 334 arquiteturas, de informação entre as arquiteturas e demonstra uma melhora significativa no
 335 resultado final.

336 6 Considerações Finais

337 Houve avanços na classificação de imagens na última década que possibilitam o uso de imagens
 338 para o diagnosticoa. O diagnóstico precoce da COVID-19 e câncer de pulmão são cruciais para o
 339 tratamento e controle da doença, assim o principal desafio no campo do aprendizado de máquina
 340 é construir um modelo que seja um classificador preciso e genérico de exames de TC.

341 Para concluir, observamos que, em geral, o desempenho é considerado regular, sendo sugerido
 342 um aumento na base de dados. Desta forma, podemos concluir que os valores de acurácia
 343 significativos de acima de 95% foram obtidos apenas quando aplicado o ensemble no conjunto
 344 de teste, levando a conclusão de que é possível criar uma ferramenta eficaz para o auxílio de
 345 profissionais da saúde na detecção de câncer no pulmão e COVID-19 a partir desse modelo.

346 Embora a classificação dos exames a TC pareça ter sucesso, a classificação de casos entre
 347 COVID-19 e outras doenças gera ambigüidade, pois ambas as imagens tem uma pequena seme-
 348 lhanças que pode gerar vies na classificação final do modelo. E com um pequeno número de
 349 imagens disponíveis na base de dados encontrada agrava ainda mais a situação.

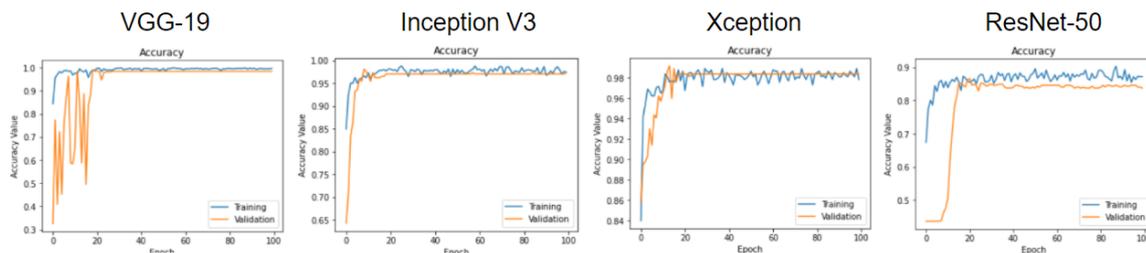


Figura 5: Gráficos de treinamento de máquina do experimento Câncer versus COVID-19 versus Saudável.

Tabela 7: Pesos de cada arquitetura, para o conjunto de dados mistos, COVID-19 e Câncer e Saudável.

	VGG-19	Inception V3	Xception	Resnet 50
Pesos	0,3	0,1	0,1	0,3

350 O uso de transfer learning ajudou bastante, pois as arquiteturas já tinham experiência com
 351 imagens de CT scan. E com a aplicação do ensemble pode ser visto como uma conquista mais
 352 eficiente, que, por sua vez, coleta o melhor de cada arquitetura. Portanto, a forma mais eficaz de
 353 treinar o modelo seria a análise de várias arquiteturas pré-treinadas e o refinamento da qualidade
 354 do conhecimento do modelo de inteligência artificial. E com isso, é possível criar uma ferramenta
 355 confiável para a classificação de doenças pulmonares a partir de TC do tórax

356 6.1 Trabalho Futuro

357 Quanto aos próximos passos deste trabalho existe um estudo no pré-processamento que realiza
 358 a segmentação de imagem a partir do processamento de imagem. Esse método pode diminuir
 359 falsos positivos, e assim, prover diagnósticos mais precisos. Vale ressaltar, que os modelos serão
 360 estendidos com explicações a fim de auxiliar os médicos e especialistas na detecção de outras
 361 doenças além de COVID-19 e câncer pulmonar.

362 References

- 363 Conselho federal de medicina.,”C. F. M., 2008.
 364
 365 P. STRÖM, “Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a
 366 population-based, diagnostic study.,”The Lancet Oncology, 2020.
 367
 368 A. MANSOOR, “Segmentation and image analysis of abnormal lungsat ct: current approaches,
 369 challenges, and future trends.,”Radiographics,2020.
 370
 371 S. WANG, “Um algoritmo de aprendizado profundo que usa imagens detc para detectar a
 372 doença do vírus corona (covid-19).,”MedRxiv, 2020.
 373
 374 K. R. KRUTHIKA, “Cbir system using capsule networks and 3d CNN for alzheimer’s disease
 375 diagnosis.,”Informatics in Medicine Unlocked,2019.
 376
 377 O. YILDIRIM, “Automated detection of diabetic subject using pre-trained 2d-cnn models with
 378 frequency spectrum images extracted from heart rate signals.,”Computers in biology and
 379 medicine, 2019.
 380

- 381 C. Z. D. LI, Xin; LI, “Covid-mobilexpert: On-device covid-19 patient triage and follow-up using
382 chest x-rays.”2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM),
383 2020.
- 384
385 W. J. KOSTIS, “Three-dimensional segmentation and growth-rate es-timation of small
386 pulmonary nodules in helical ct images.”IEEE transactions on medical imaging, 2003.
- 387
388 C. QUISPE, “Learning non-maximum suppression to improve lung nodule detection in ct
389 images: Aprendendo a suprimir não-máximospara aperfeiçoar a detecção de nódulos
390 pulmonares em imagens de ct.”Repositorio Unicamp, 2020.
- 391
392 P. SILVA, “Covid-19 detection in ct images with deep learning: A voting-based scheme and
393 cross-datasets analysis.”Informatics in MedicineUnlocked, 2020.
- 394
395 H. A. GIETEMA, “Pulmonary nodules: interscan variability of semi-automated volume
396 measurements with multisection ct—influence of in-spiration level, nodule size, and
397 segmentation performance.”Radiology,2007.
- 398
399 W. SHEN, “Multi-scale convolutional neural networks for lung nodule
400 classification.”International conference on information processing in medical imaging, 2015.
- 401
402 W. ZHU, “Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and
403 classification.”2018 IEEE Winter Conference on Applications of Computer Vision (WACV).,
404 2018.
- 405
406 T. AI, “Correlation of chest ct and rt-pcr testing for coronavirus disease2019 (covid-19) in
407 china: a report of 1014 cases,”Radiology, 2019.
- 408
409 S. U. K. BUKHARI, “The diagnostic evaluation of convolutional neuralnetwork (CNN) for the
410 assessment of chest x-ray of patients infected with covid-19.”Medical Imaging 2019: Imaging
411 Informatics for Healthcare,Research, and Applications., 2020.
- 412
413 O. GOZES, “Coronavirus detection and analysis on chest ct with deeplearning.”arXiv e-prints,
414 2020.
- 415
416 A. K. MISHRA, “Identifying Covid-19 from chest ct images: a deepconvolutional neural
417 networks based approach.”Journal of HealthcareEngineering, 2020.
- 418
419 Y. e. a. LECUN, “Reconhecimento de objetos com aprendizagem baseada em gradientes. in:
420 Forma, contorno e agrupamento em visão computacional.”Springer, Berlin, Heidelberg, 1999.
- 421
422 K. SIMONYAN and A. Zisserman, “Very deep convolutional networksfor large-scale image
423 recognition.”arXiv preprint, 2015.
- 424
425 C. SZEGEDY, “Going deeper with convolutions.”Proceedings of the IEEE conference on
426 computer vision and pattern recognition, 2020.
- 427
428 S. AYYACHAMY, “Medical image retrieval using resnet-18,”Medical Imaging 2019: Imaging
429 Informatics for Healthcare, Research, and Applications., 2019.
- 430
431 F. CHOLLET, “Xception: Deep learning with depth wise separableconvolutions.”Proceedings

- 432 of the IEEE conference on computer vision and pattern recognition., 2017.
433
- 434 K. HE, “Deep residual learning for image recognition.”Proceedings of the IEEE conference on
435 computer vision and pattern recognition, 2016.
- 436
- 437 Z. ZHOU, “Ensemble learning.” no. 1, pp. 270–273, 2009.
- 438
- 439 J. ORWANT, “Algoritmos de masterização com perl.”O’Reilly Media,Inc., 1999.
- 440
- 441 B. A. L. M. J. C. T. N. E. O. B. S. J. I. D. . R. J. W. SCHREIBER, J.N. C., “Técnicas de
442 validação de dados para sistemas inteligente: uma abordagem do software sdbayes.” 2017.
- 443
- 444 J. PEREZ, Luis; WANG, “The effectiveness of data augmentation in image classification using
445 deep learning.”arXiv preprint, 2017.
- 446
- 447 J. W. M. BOCK, Sebastian; GOPPOLD, “Medical image retrieval usingres net-18,”An
448 improvement of the convergence proof of the ADAM-optimizer., 2018.
- 449
- 450 SETIO, Arnaud Arindra Adiyoso et al. Pulmonary nodule detection in CT images: false
451 positive reduction using multi-view convolutional networks. IEEE transactions on medical
452 imaging, v. 35, n. 5, p. 1160-1169, 2016.
- 453