

CENTRO UNIVERSITÁRIO SAGRADO CORAÇÃO

LETÍCIA FERNANDA LOPES

**SISTEMA ESPECIALISTA INTELIGENTE PARA
IDENTIFICAÇÃO DE COVID-19**

BAURU
2021

LETÍCIA FERNANDA LOPES

**SISTEMA ESPECIALISTA INTELIGENTE PARA
IDENTIFICAÇÃO DE COVID-19**

Monografia de Iniciação Científica
apresentado a Pró-Reitoria de Pesquisa e
Pós-graduação do Centro Universitário
Sagrado Coração, sob orientação do Prof.
Me. Vinicius Santos Andrade.

BAURU
2021

Dados Internacionais de Catalogação na Publicação (CIP) de acordo com
ISBD

L864s	<p>Lopes, Leticia Fernanda</p> <p>Sistema especialista inteligente para identificação de Covid-19 / Leticia Fernanda Lopes. -- 2021. 33f. : il.</p> <p>Orientador: Prof. M.e Vinicius Santos Andrade</p> <p>Monografia (Iniciação Científica em Ciência da Computação) - Centro Universitário Sagrado Coração - UNISAGRADO - Bauru - SP</p> <p>1. Inteligência Artificial. 2. Covid-19. 3. Doença. I. Andrade, Vinicius Santos. II. Título.</p>
-------	---

RESUMO

O primeiro semestre de 2020 foi marcado pelo caos gerado por conta da Covid-19, também conhecida como corona vírus. Desde os primeiros relatos das pessoas infectadas com o vírus, diversos pesquisadores se uniram para a criação de algoritmos e/ou sistemas que usam Inteligência Artificial para detectar o vírus através de imagens, dados do prontuário do paciente etc. O intuito é auxiliar no diagnóstico humano além de tornar o processo de identificação e classificação da doença mais eficiente. Este trabalho apresenta um sistema que utiliza Inteligência Artificial para identificação do Covid-19.

Palavras-chave: Inteligência Artificial. Covid-19. Doença.

ABSTRACT

The first half of 2020 was marked by the chaos generated by Covid-19, also known as coronavirus. Since the first reports of people infected with the virus, several researchers have teamed up to create algorithms and/or systems that use Artificial Intelligence to detect the virus through images, data from the patient's medical record, etc. The aim is to assist in human diagnosis in addition to making the process of identifying and classifying the disease more efficient. This work presents a system that uses Artificial Intelligence to identify Covid-19.

Keywords: Artificial Intelligence. Covid-19. Illness.

LISTA DE ILUSTRAÇÕES

Figura 1 – Evolução dos casos de Covid-19 no Brasil	4
Figura 2 – O processo do aprendizado de máquina	7
Figura 3 – Impacto do valor de k no algoritmo kNN	9
Figura 4 – Uma árvore de decisão “adivinha o animal”	10
Figura 5 – Fluxograma do processamento de imagens utilizado no projeto	11
Figura 6 – Casos de COVID-19	12
Figura 7 – Raio-X tórax com região analisada demarcada	18
Figura 8 – Parte do código utilizado no front-end	19
Figura 9 – Tela de envio de imagem para identificação	19
Figura 10 – Tela com identificação negativa	20
Figura 11 – Tela com identificação positiva	20
Figura 12 – Trecho do código feito no Google Colab	21
Figura 13 – Trecho do código responsável por extrair dados analíticos.....	21
Figura 14 – Dados analíticos da imagem como um todo.....	22
Figura 15 – Dados analíticos primeiro quadrante (canto superior esquerdo)	22
Figura 16 – Dados analíticos segundo quadrante (canto superior direito)	23
Figura 17 – Dados analíticos terceiro quadrante (canto inferior esquerdo)	23
Figura 18 – Dados analíticos quarto quadrante (canto inferior direito)	24
Figura 19 – Trecho do código da API	25

LISTA DE ABREVIATURAS E SIGLAS

API	<i>Application Programming Interface</i>
IA	Inteligência Artificial
IC	Iniciação Científica
kNN	<i>k nearest neighbors</i>

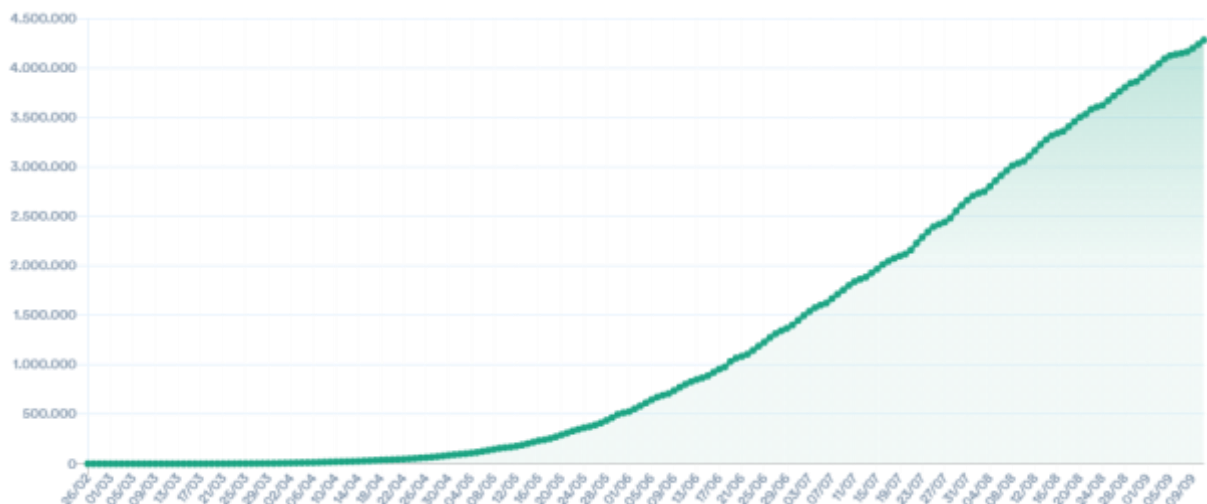
SUMÁRIO

1	INTRODUÇÃO	4
1.1	TRABALHOS RELACIONADOS	5
2	FUNDAMENTAÇÃO TEÓRICA	6
2.1	INTELIGÊNCIA ARTIFICIAL	6
2.1.1	<i>K-Vizinhos mais próximos</i>	8
2.1.2	<i>Árvore de decisão</i>	9
2.2	PROCESSAMENTO DE IMAGENS	10
2.3	COVID-19	11
3	OBJETIVOS	14
3.1	OBJETIVO GERAL	14
3.2	OBJETIVOS ESPECÍFICOS	14
4	MATERIAIS E MÉTODOS	15
5	RESULTADOS	17
6	DISCUSSÃO DOS RESULTADOS	26
7	CONSIDERAÇÕES FINAIS	26
	REFERÊNCIAS	27
8	ANEXO I - CARTA DE DISPENSA DE APRESENTAÇÃO AO CEP OU	
CEUA	28	

1 INTRODUÇÃO

Segundo G1 (2020), em dezembro de 2019 surgiram os primeiros relatos da Covid-19, na China. Atualmente, o mundo todo está vivendo em meio ao caos devido a propagação do vírus. No Brasil, a propagação do vírus está ocorrendo de forma rápida e esta realidade traz inúmeras preocupações, principalmente pelo fato do país não ter estrutura hospitalar para tratar de tantos pacientes com a doença. A Figura 1 exibe os dados estatísticos relacionados a propagação do no Brasil.

Figura 1 – Evolução dos casos de Covid-19 no Brasil.



Fonte: CORONA VÍRUS BRASIL (2020).

Nota: Por região do Brasil (12/09/2020

atualizado as 16:10).

Prontamente, pesquisadores começaram a se unir para desenvolver aplicações que utilizam Inteligência Artificial (IA) para auxiliar no diagnóstico da Covid-19.

Os testes feitos atualmente para identificar a doença sem o auxílio de IA, podem demorar até um mês, utilizando amostras de sangue, por exemplo.

Sistemas que utilizam IA, além de serem eficientes pelo fato de conseguirem processar uma grande quantidade de dados em pouco tempo, isso contando com um ambiente de hardware de alto desempenho.

Existem diversos algoritmos de Inteligência Artificial, que trazem consigo, inúmeras possibilidades de aplicações. Visto que a Covid-19 é algo recente, estudar as diversas possibilidades possíveis se torna algo trivial no processo de identificação da melhor forma de se fazer o reconhecimento da doença utilizando IA.

Portanto, este trabalho visa colaborar com esse processo de estudo, efetuando a aplicação de algoritmo(s) de IA para identificação da Covid-19.

1.1 TRABALHOS RELACIONADOS

Diversas notícias de diferentes jornais eletrônicos mostram o esforço e colaboração entre computação e medicina para a criação de soluções baseadas em IA para identificação do Covid-19. Pesquisadores do Instituto de Ciências Matemáticas e de Computação (ICMC), da Universidade de São Paulo (USP), do campus de São Carlos, estão utilizando técnicas de inteligência computacional (ramo da IA) para analisar eventos extraídos de notícias em forma de texto (Olhar Digital, 2019).

No Canadá, pesquisadores desenvolveram um algoritmo que varre milhares de fontes de dados, de documentos de autoridades a publicação médicas e relatórios de condições climáticas, em busca de informações sobre a capacidade de proliferação da doença. (Terra, 2020).

A Microsoft e a Google lançaram um desafio com o intuito de utilizar a IA para garimpar estudos de mais de 29 mil artigos médicos. (Terra, 2020).

Outro grupo de pesquisadores, utilizaram a IA para prevenção de pandemias ao Covid-19, onde criado pelos pesquisadores consegue através da revisão em mídias e redes sociais detectar a propagação de uma doença incomum com sintomas de gripe em Wuhan, na China. (UOL, 2020).

O trabalho de LI et al (2020) serviu como inspiração para este projeto. Os autores utilizam redes neurais convolucionais e, com o auxílio de técnicas de processamento de imagens, fazem análises de imagens de tórax para identificação da Covid-19.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 INTELIGÊNCIA ARTIFICIAL

Nos últimos anos, sistemas de IA ganharam espaço. Isso acontece devido ao fato do aumento de dados que muitas vezes não conseguem ser analisados de forma precisa e rápida pelos seres humanos. Independente da área, cada vez mais dependemos de dados para tomada de decisões, seja na bolsa de valores, na saúde ou educação. Na saúde, a demora para análise de um determinado dado, ou a análise de forma errônea, pode custar o bem mais precioso do ser humano, a vida.

De acordo com Faceli *et al.* (2011), técnicas de IA, em particular de Aprendizado de Máquina (AM), têm sido utilizados com sucesso em um grande número de problemas reais, como por exemplo, identificação de doenças com base em exames clínicos.

Segundo Medeiros (2018), os seguintes aspectos caracterizam a Inteligência Artificial:

- a) capacidade de resolução de problemas;
- b) aprendizado com o ambiente;
- c) desenvolvimento de estruturas cognitivas; e
- d) orientação a metas.

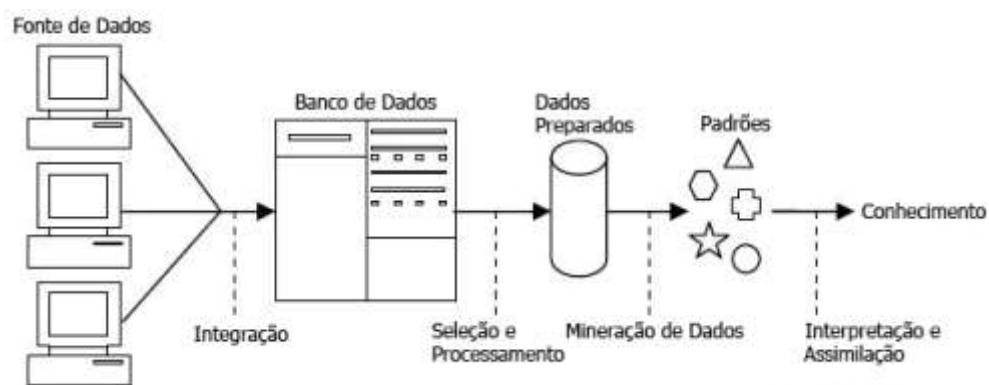
A partir da década de 1970, houve maior disseminação do uso de técnicas de computação baseadas em IA para a resolução de problemas reais. Esse avanço ocorreu devido à crescente complexidade dos problemas a serem tratados computacionalmente, além do volume de dados gerados por diferentes setores, tornando clara a necessidade de ferramentas computacionais mais sofisticadas, que fossem mais autônomas, reduzindo a necessidade de intervenção humana e dependência de especialistas. Antes disso, a área de IA era vista como uma área teórica, com aplicações apenas em pequenos problemas de pouco valor prático (FACELI *et al.*, 2011).

Embora AM seja naturalmente associado à IA, outras áreas de pesquisa têm contribuições diretas e significativas no avanço do AM, como Probabilidade e Estatística. Em AM, a ideia é utilizar dados históricos, por exemplo, para gerar modelos preditivos. Existem várias aplicações bem-sucedidas de técnicas de AM na solução de problemas reais, tais como (FACELI *et al.*, 2011):

- a) reconhecimento de palavras faladas;
- b) predição de taxas de cura de pacientes com diferentes doenças;
- c) detecção do uso fraudulento de cartões de crédito; e
- d) diagnóstico de câncer por meio de análise de dados de expressão genética.

A Figura 2 exemplifica o processo de aprendizagem de um sistema baseado em IA.

Figura 2 – O processo do aprendizado de máquina



Fonte: Bramer (2007, tradução nossa).

Antes da explicação da Figura 2, é importante definirmos o conceito de mineração de dados. Este é um termo abrangentemente usado para descrever diferentes aspectos de processamento de dados. Em outras palavras, a mineração de dados é o estudo de coleta, limpeza, processamento, análise e obtenção de informações úteis que possam ser utilizadas como dados em aplicações de AM (AGGARWAL, 2015).

Os dados que chegam, normalmente são provenientes de várias fontes. Estes dados são integrados e colocados em algum armazenamento (banco de dados). Parte é então refinada e pré-processada, levando em consideração um sistema que tem como base de dados arquivos de áudio, o pré-processamento neste caso, teria como objetivo a remoção de ruídos, por exemplo. Agora, os dados estão “preparados” para um algoritmo de mineração, que por sua vez, irá produzir uma saída na forma de regras seguindo um outro tipo de “padrão”. Agora, estes serão interpretados para, assim, gerar o “conhecimento” (BRAMER, 2007).

De acordo com Russell e Norvig (2014), existem várias formas para trabalhar-se com o AM:

- a) aprendizado a partir de exemplos;
- b) conhecimento em aprendizagem;
- c) aprendizagem de modelos probabilísticos; e
- d) aprendizagem por esforço.

A seguir, serão abordados alguns desses modelos de aprendizagem utilizados para interpretar os dados

2.1.1 K-Vizinhos mais próximos

Do inglês *k nearest neighbors* (kNN), este é o modelo mais simples para interpretar padrões. O modelo foi proposto por Fukunaga e Narendra (1975), e o aprendizado de um novo “objeto” baseia-se nos exemplos de treinamento. A seguir, a equação 1 representa o modelo.

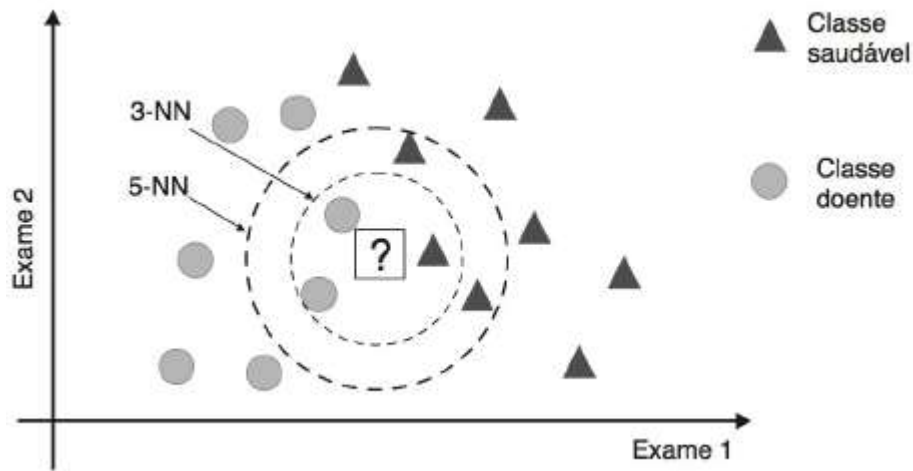
$$1) \quad d(X_i, Y_i) = \sqrt[r]{\sum_{i=1}^n |X_i - Y_i|^r}$$

Existem várias métricas para o cálculo da distância, e a escolha de qual usar varia de acordo com o problema. A mais utilizada é a distância Euclidiana, descrita pela equação 2.

$$2) \quad D_E(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

O objetivo do kNN é determinar o rótulo de classificação de uma amostra baseado nas amostras vizinhas advindas de um conjunto de treinamento. A Figura 3 exemplifica o funcionamento do modelo.

Figura 3 – Impacto do valor de k no algoritmo kNN



Fonte: Faceli *et al.*, (2011).

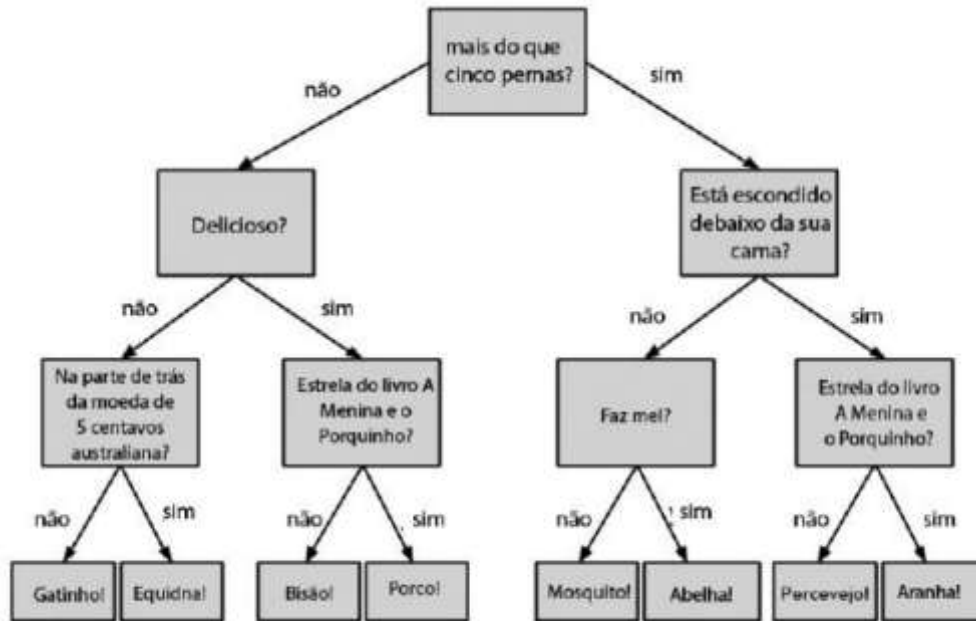
No exemplo da Figura 3, são aferidas as distâncias de uma nova amostra, representada pelo quadrado com interrogação, às demais amostras, representadas por triângulos e círculos. Em uma primeira situação a variável k , a qual representa a quantidade de vizinhos mais próximos, recebe o valor 3, o que resulta na classificação do novo objeto pertencendo à classe “doente”, enquanto para $k = 5$ o objeto de teste seria classificado como pertencendo à classe “saudável”.

Outro exemplo em que o classificador poderia ser aplicado é no sistema do famoso aplicativo Uber. Na situação em questão, a nova amostra seria o passageiro, possuindo sua localização geográfica (x, y) , representando respectivamente longitude e latitude. Ao solicitar a corrida, o classificador utiliza de tais informações para, com base nos dados dos motoristas (localização geográfica), calcular a distância euclidiana definir qual é o “vizinho” mais próximo e atribuí-lo à corrida. Diferente do exemplo da Figura 3, aqui, não existe a clusterização de dados, uma vez que os mesmos não estão sendo agrupados.

2.1.2 Árvore de decisão

Árvores de decisões são fáceis de entender e interpretar, e o processo de previsão é completamente transparente, diferente de outros modelos, como por exemplo, regressão logística. Este modelo usa a estrutura de árvore para representar um número de possíveis caminhos de decisões e um resultado para cada caminho, conforme exemplificado na Figura 4 (GRUS, 2019).

Figura 4 – Uma árvore de decisão “adivinha o animal”



Fonte: Grus (2019).

Para construir árvores de decisões, se faz necessário decidir quais perguntas fazer, assim como suas respectivas ordens. Tais perguntas, são focadas para que a árvore nos retorne informações sobre o que a árvore deveria prever. No exemplo da Figura 4, todas as perguntas se resumem a duas possibilidades de respostas, sim (*true*) e não (*false*). Tais perguntas resultam em uma quantidade pequena de informações, portanto, ao se fazer uma pergunta, é importante levar em consideração a quantidade de informação que a resposta retornará. Essa noção de “quanta informação” é intitulada entropia (GRUS, 2019).

2.2 PROCESSAMENTO DE IMAGENS

Gonzales e Woods (2010, p.1) descrevem uma imagem digital sendo:

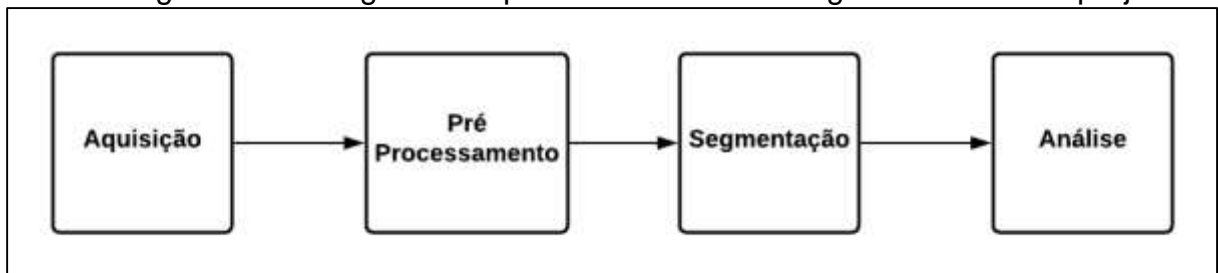
Uma função bidimensional, $f(x, y)$, em que x e y são coordenadas espaciais (plano), e a amplitude de f em qualquer par de coordenadas (x, y) é chamada de intensidade ou nível de cinza da imagem nesse ponto.

Através da imagem digital é possível ser realizado o processamento de imagens que teve sua origem no século passado por volta de 1920 onde imagens eram enviadas via cabo submarino entre Londres e Nova York, uma Impressora codificava as imagens em 5 níveis para a transmissão (cabo Bartlane) e reconstruía

no recebimento. Em 1929 essa codificação passou a ser feita em 15 níveis. Porém o grande avanço do processamento de imagens aconteceu em 1964 com os primeiros computadores de grande porte e o início do programa espacial dos Estados Unidos da América. (MARQUES FILHO, OGÊ, 1999; VIEIRA NETO, HUGO, 1999). Nos tempos atuais o processamento de imagens está presente no nosso cotidiano, utilizamos da imagem para nos comunicar, aplicativos como Intragram, Facebook, TikTok, Netflix, entre outros, utilizam do processamento de imagem como seu foco principal para seu funcionamento. Além de áreas como medicina, indústrias, pesquisas, engenharias que utilizarem também de técnicas de processamento de imagens.

O projeto irá utilizar do processamento de imagens para conseguir identificar através de raios-x se o paciente está ou não com COVID-19, para a realização disso teremos o fluxograma mostrado na Figura 5.

Figura 5 – Fluxograma do processamento de imagens utilizado no projeto.

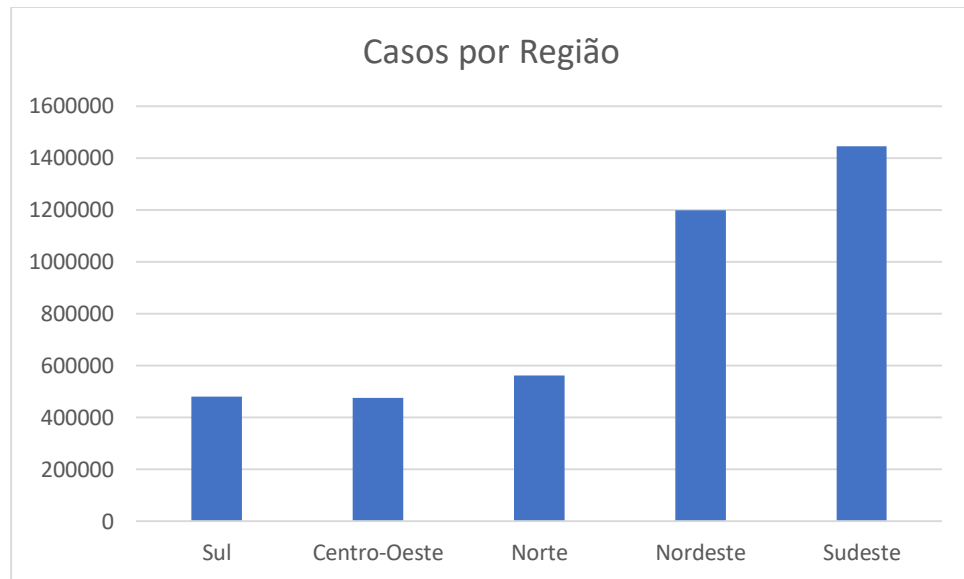


Fonte: Autoria do autor.

2.3 COVID-19

No final de 2019 surgiram os primeiros casos de COVID-19 que em pouco tempo tornou-se uma pandemia. No Brasil, o primeiro caso foi registrado em fevereiro de 2020 (CROKIDAKIS, 2020) e pouco tempo após, diversas regiões do país começaram a decretar estado de quarentena. O gráfico contido na Figura 6 exibe o número de casos de COVID-19 por regiões do Brasil (CORONA VIRUS BRASIL, 2020).

Figura 6 - Casos de COVID-19



Fonte: CORONA VÍRUS BRASIL (2020).

Nota: Por região do Brasil (09/09/2020 atualizado as 17:40).

Desde o início da pandemia no Brasil, diversos cientistas brasileiros iniciaram o desenvolvimento de pesquisas voltadas à identificação do vírus, criação de uma vacina, dentre outros. Diversas dessas pesquisas envolvem IA.

Em março deste ano, o hospital israelense Albert Einstein (2020) organizou um evento em que os participantes tinham como desafio propor soluções baseadas em IA para diagnóstico de COVID-19, com base nos dados provenientes de prontuários de pacientes.

Em parceria com o hospital Albert Einstein, a Universidade de São Paulo (USP) criou um algoritmo de Inteligência Artificial capaz de identificar o vírus. O algoritmo faz uso de 15 variáveis diferentes, tais como idade, sexo, quantidade de glóbulos vermelhos e plaquetas. Em experimentos, o algoritmo obteve 77% de assertividade para testes positivos e negativos (OLHAR DIGITAL, 2020).

Outro projeto em parceria com o Instituto de Ciências Matemáticas e de Computação da USP de São Carlos, utiliza um chatbot¹ para indicar, especialmente para o público de baixa escolaridade, qual o hospital mais próximo em que ele conseguirá atendimento. A ferramenta também fornece ao usuário a rota para o hospital, usando imagens do Google Maps (JORNAL USP, 2020).

Estes sistemas de IA são de grande importância, pois com relativamente poucos recursos, conseguem entregar inúmeros diagnósticos de pacientes. Aspecto importante das aplicações visto que, de acordo com a reportagem da BBC NEWS (2020), o Brasil é um dos países que menos realiza testes para COVID-19 devido à falta de recursos em diversas regiões do país.

Os trabalhos relatados anteriormente deixam claro o impacto e importância de sistemas baseados em IA para a área de saúde. Utilizar destes sistemas garante, principalmente, a possibilidade de análises em grande escala e com precisão aceitável.

3 OBJETIVOS

A seguir são descritos o objetivo geral e os específicos que norteiam essa pesquisa.

3.1 OBJETIVO GERAL

Desenvolvimento e avaliação de um sistema especialista em identificação de Covid-19 utilizando técnica(s) de Inteligência Artificial.

3.2 OBJETIVOS ESPECÍFICOS

- a) revisão da literatura para identificar o que já foi feito até o momento;
- b) análise e estudo das propostas obtidas na etapa anterior;
- c) seleção de algoritmos de IA;
- d) estudo da base de dados e como ela será utilizada e parametrizada;
- e) estudo e aplicação de técnicas de processamento de imagens afim de obter uma matriz de características;
- f) aplicação de algoritmos de IA para identificação da Covid-19;
- g) testes do algoritmo;
- h) geração de dados estatísticos referente a capacidade assertiva do algoritmo;

4 MATERIAIS E MÉTODOS

Este trabalho foi dividido em quatro etapas:

- a) fundamentação teórica;
- b) revisão e estudo da literatura;
- c) criação de um algoritmo utilizando IA;
- d) análise do algoritmo, geração dos dados estatísticos e conclusão do relatório final da IC.

Na fundamentação teórica, serão abordadas teorias e ferramentas computacionais necessárias ao desenvolvimento deste projeto. Este levantamento bibliográfico foi feito utilizando técnicas de revisão da literatura, baseando-se em consultas à literatura especializada e de alta relevância científica, incluindo: monografias, dissertações, teses, livros, sites de documentação e artigos científicos.

Sequencialmente, foi feito o estudo do material adquirido, com o intuito de definir parâmetros, bases de dados etc. que serão utilizadas neste projeto. Após a definição de tais pontos, foi iniciado a programação do algoritmo.

Com base nas informações obtidas com o levantamento bibliográfico feito até o momento, optou-se por utilizar a linguagem de programação Python. A linguagem é amplamente utilizada na resolução de problemas com IA havendo diversos relatos de sucesso (PYTHON, 2020).

Por fim, o algoritmo foi testado e foram gerados os dados estatísticos de desempenho de assertividade do mesmo.

A metodologia utilizada na etapa de desenvolvimento do algoritmo envolve as seguintes etapas:

- a) escolha da base de dados a ser utilizada;
- b) escolha do algoritmo de IA;
- c) definição dos parâmetros;
- d) definição do conjunto de dados para treino e teste;
- e) aplicação do algoritmo e criar sistema para detecção de COVID-19;
- f) desenvolvimento do código específico para geração de dados estatísticos; e
- g) testes no sistema.

O produto será apresentado no Fórum de Iniciação Científica do UNISAGRADO, bem como, submetido a eventos/revistas científicas da área.

5 RESULTADOS

Como já mencionado anteriormente, a linguagem escolhida para o desenvolvimento do projeto foi Python devido a inúmeros trabalhos relacionados à IA utilizarem a linguagem. Também foi possível definir as bases de dados que serão utilizadas na pesquisa, sendo elas:

a) Base 1. Disponível em: <https://github.com/ieee8023/covid-chestxray-dataset/tree/master/images>. Data do último acesso: 15 de setembro de 2020.

b) Base 2: Disponível em: <https://github.com/ieee8023/covid-chestxray-dataset>. Data do último acesso: 18 de setembro de 2020

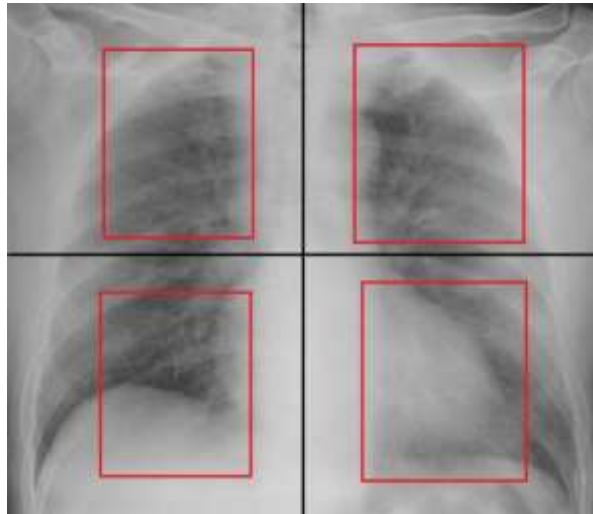
c) Base 3: Disponível em: <https://drive.google.com/uc?id=1coM7x3378f-Ou2l6Pg2wldaOI7Dntu1a>

As bases de dados são de imagens de tórax, por afinidade com a aplicação de algoritmos de IA em imagens, optou-se por utilizar esse tipo de base ao invés de base de dados oriundas de prontuários.

Com base nos estudos dos autores Rosa (2020), Chung et al. (2020) e Bai *et al.* (2020), foi definida a região de interesse que será utilizada para extrair informações e criar o vetor de características. Uma grande dificuldade encontrada é que outras doenças que atacam o pulmão, como a pneumonia, geram manchas semelhantes à da COVID-19, o que dificultaria o processo de identificação, caso o algoritmo utilizado para classificação seja muito simples – como é a proposta desse projeto. Inicialmente serão feitos testes com imagens em que a COVID-19 se encontra em estágios mais avançados e imagens onde o pulmão está saudável. A ideia é aumentar a dificuldade dos testes e criar uma tabela resumo que sintetiza os resultados.

Tendo ciência do “modelo” das imagens das bases, foi possível definir como será feita a criação do vetor de características. Para isso, a imagem foi subdividida em quatro quadrantes, conforme exhibe a Figura 7.

Figura 7 – Raio-X tórax com região analisada demarcada.



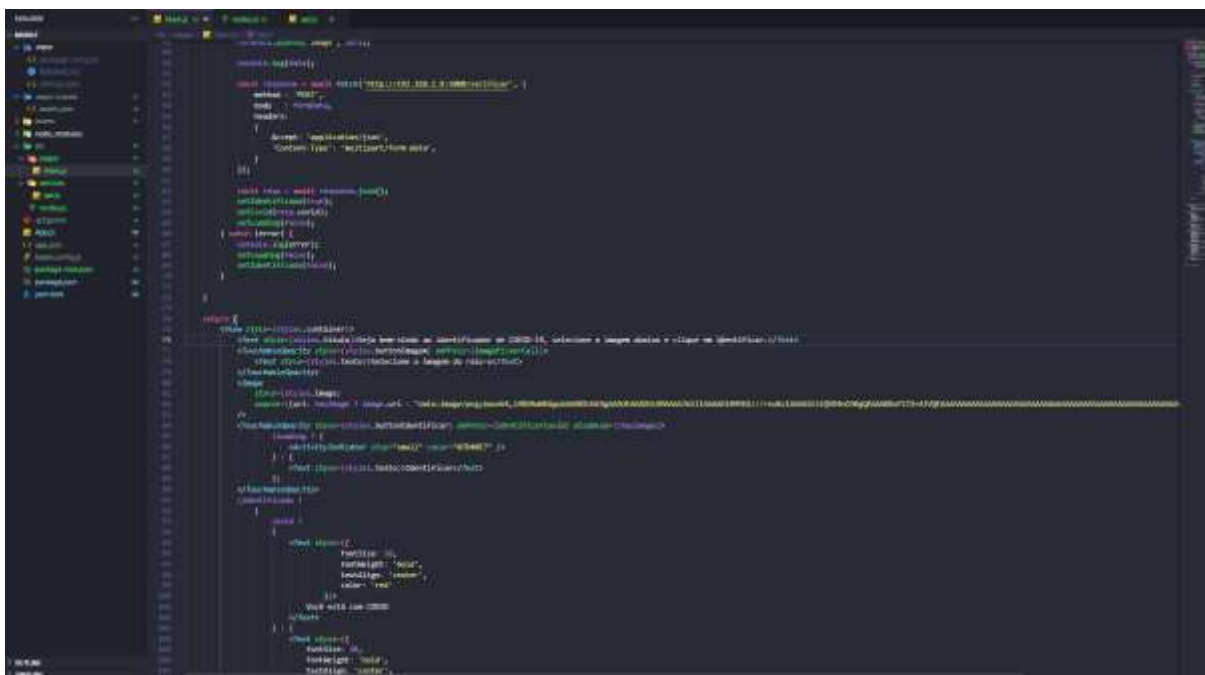
Fonte: Imagem retirada das bases citadas e demarcada pelo autor.

Gonzales e Woods (2010, p.1) descrevem uma imagem digital sendo: uma função bidimensional, $f(x, y)$, em que x e y são coordenadas espaciais (plano), e a amplitude de f em qualquer par de coordenadas (x, y) é chamada de intensidade ou nível de cinza da imagem nesse ponto.

Com base nessa definição, cada quadrante será definido de acordo com as posições x_1 e x_2 ; y_1 e y_2 , em cada quadrante. A Figura 7 exemplifica essa ideia. O vetor será criado com base nos dados do histograma de cada quadrante.

Após a realização dessas definições foi possível desenvolver um *front-end* em React Native onde o usuário pode selecionar imagem da galeria e mandar “identificar”. Ao pressionar em identificar é feita uma requisição em uma API feita em Python onde será feita a análise se está ou não com COVID-19 ambas as partes foram feitas utilizando o Visual Studio Code. A Figura 8 exibe um trecho do código que foi utilizado para criação do *front-end*. Enquanto a Figura 9, Figura 10 e Figura 11 exibe a tela para envio da imagem para identificação da COVID-19.

Figura 8 – Parte do código utilizado no *front-end*.



Fonte: Elaborado pela autora.

Figura 9 – Tela de envio de imagem para identificação.



Fonte: Elaborado pela autora.

Figura 10 – Tela com identificação negativa.



Fonte: Elaborado pela autora.

Figura 11 – Tela com identificação positiva.

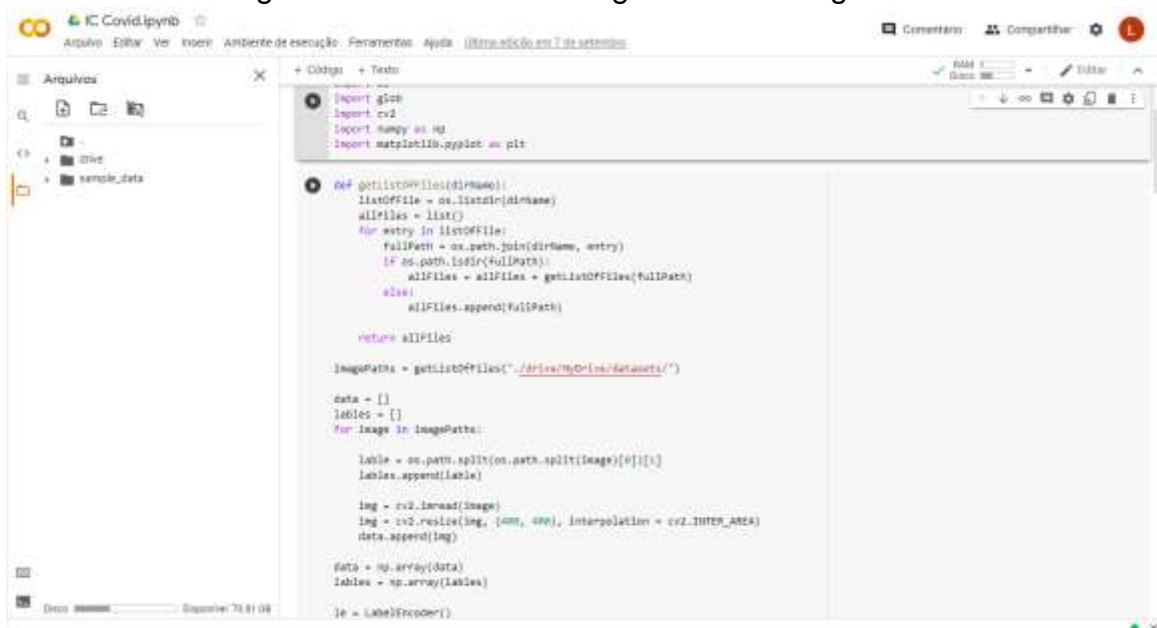


Fonte: Elaborado pela autora.

Os resultados parciais da pesquisa foram publicados e apresentados no 20º Congresso Nacional de Iniciação Científica (20º CONIC), obtendo sexto lugar no grupo Ciências Exatas e da Terra – Em Andamento.

Logo após a criação do *front-end* foi utilizado o Google Colab para fazer a primeira implementação do algoritmo, montagem do modelo e extração de dados analíticos um trecho do código é exibido na Figura 12.

Figura 12 – Trecho do código feito no Google Colab.



```

import glob
import cv2
import numpy as np
import matplotlib.pyplot as plt

def getListOfFiles(dirName):
    listoFFile = os.listdir(dirName)
    allfiles = list()
    for entry in listoFFile:
        fullPath = os.path.join(dirName, entry)
        if os.path.isdir(fullPath):
            allfiles = getListOfFiles(fullPath)
        else:
            allfiles.append(fullPath)

    return allfiles

imagePaths = getListOfFiles("../drive/MyD/Im/dataset/")

data = []
labels = []
for image in imagePaths:
    label = os.path.splitext(os.path.splitext(image)[0])[1]
    labels.append(label)

    img = cv2.imread(image)
    img = cv2.resize(img, (400, 400), interpolation = cv2.INTER_AREA)
    data.append(img)

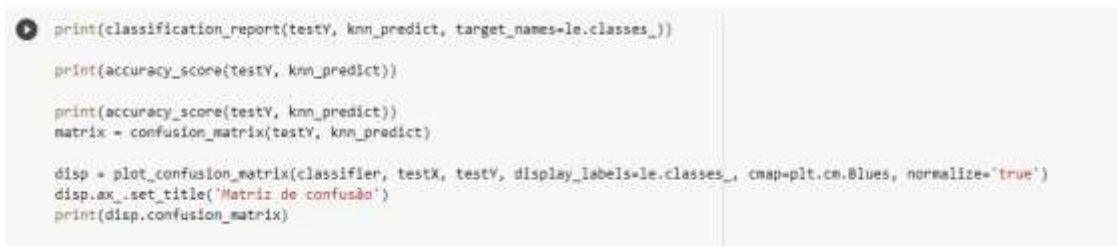
data = np.array(data)
labels = np.array(labels)

le = LabelEncoder()
  
```

Fonte: Elaborado pela autora.

Para montar o modelo do KNN, treina-lo e extrair dados analíticos do seu desempenho foi utilizada a biblioteca *scikit-learn*, juntamente com a biblioteca Matplotlib para auxiliar na exibição, dentre outras bibliotecas contidas no projeto. Para avaliar o desempenho do algoritmo foram utilizadas as métricas de avaliação: Acurácia, Precisão, Revocação e Matriz de Confusão, dando como prioridade o menor número de falso positivo, na Figura 13 é possível ver o trecho responsável por extrair tais dados.

Figura 13 – Trecho do código responsável por extrair dados analíticos.



```

print(classification_report(testY, knn_predict, target_names=le.classes_))
print(accuracy_score(testY, knn_predict))

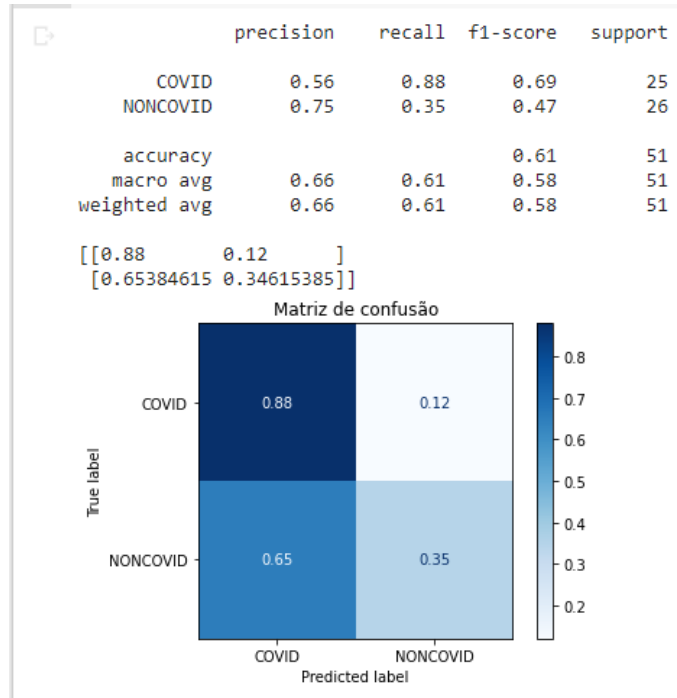
print(accuracy_score(testY, knn_predict))
matrix = confusion_matrix(testY, knn_predict)

disp = plot_confusion_matrix(classifier, testX, testY, display_labels=le.classes_, cmap=plt.cm.Blues, normalize='true')
disp.ax.set_title('Matriz de confusão')
print(disp.confusion_matrix)
  
```

Fonte: Elaborado pela autora.

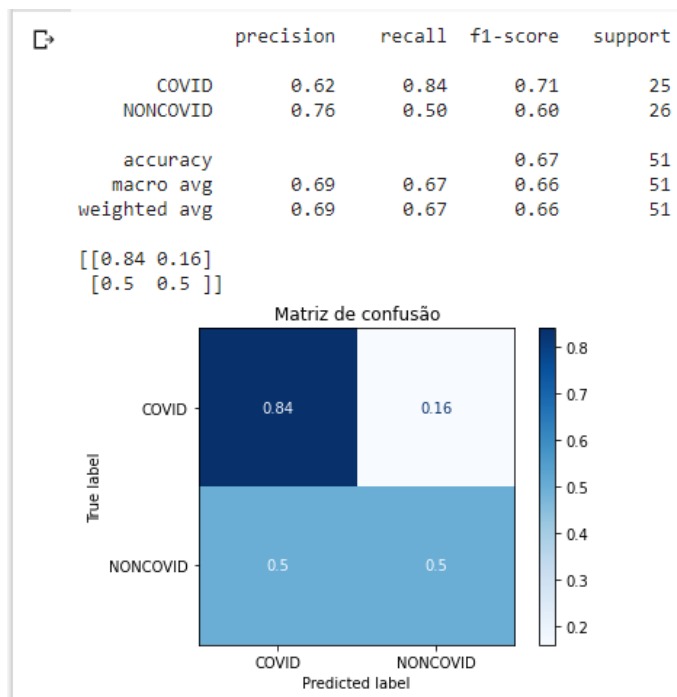
Esses dados foram extraídos de cinco situações: imagem como um todo (Figura 14), primeiro quadrante (Figura 15), segundo quadrante (Figura 16), terceiro quadrante (Figura 15) e quarto quadrante (Figura 17).

Figura 14 – Dados analíticos da imagem como um todo.



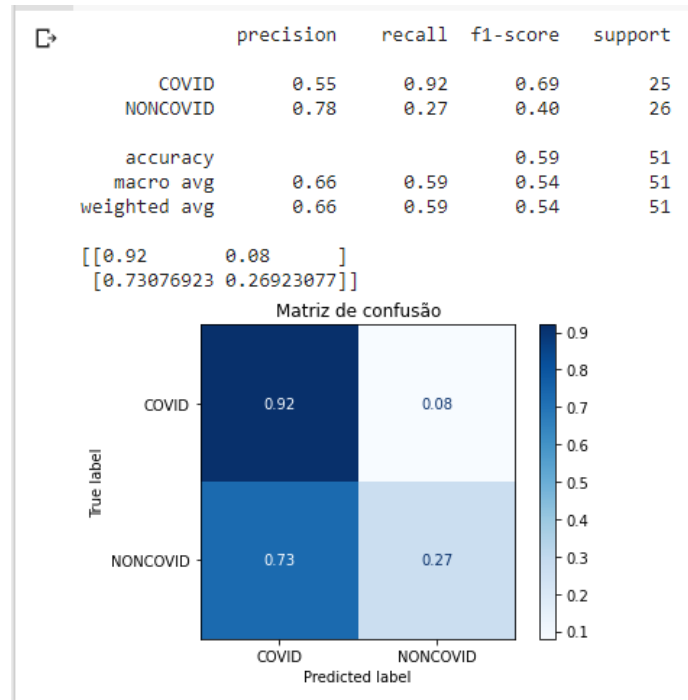
Fonte: Elaborado pela autora.

Figura 15 – Dados analíticos primeiro quadrante (canto superior esquerdo).



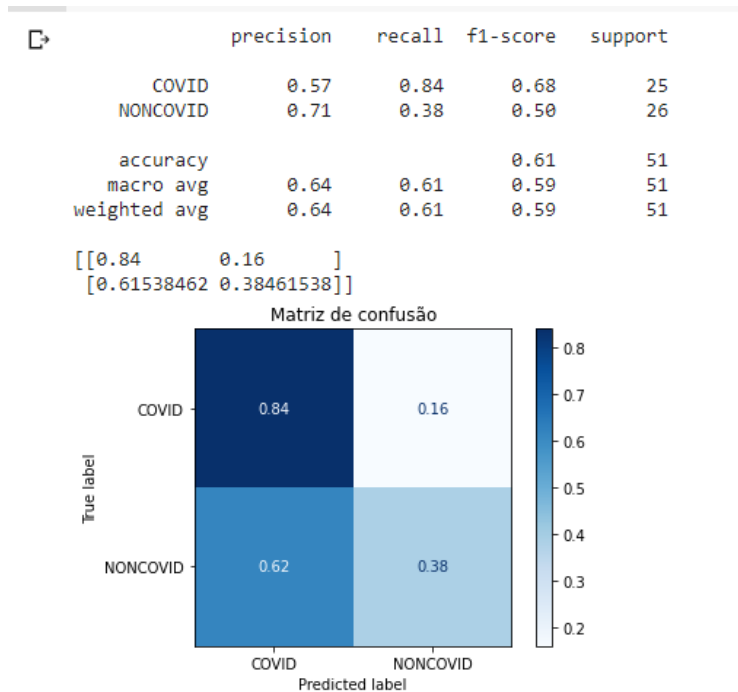
Fonte: Elaborado pela autora.

Figura 16 – Dados analíticos segundo quadrante (canto superior direito)



Fonte: Elaborado pela autora.

Figura 17 – Dados analíticos terceiro quadrante (canto inferior esquerdo).

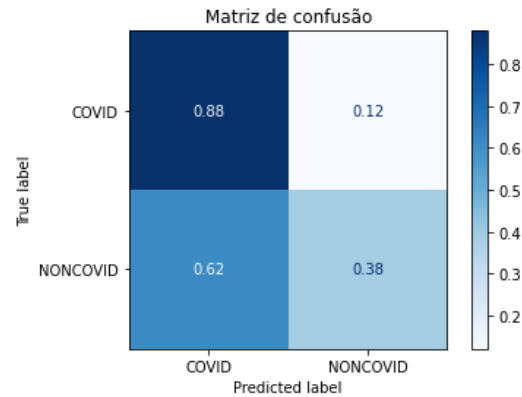


Fonte: Elaborado pela autora.

Figura 18 – Dados analíticos quarto quadrante (canto inferior direito)

	precision	recall	f1-score	support
COVID	0.58	0.88	0.70	25
NONCOVID	0.77	0.38	0.51	26
accuracy			0.63	51
macro avg	0.67	0.63	0.61	51
weighted avg	0.68	0.63	0.60	51

```
[[0.88  0.12  ]
 [0.61538462 0.38461538]]
```



Fonte: Elaborado pela autora.

Após a leitura dos dados analíticos foi escolhido trabalhar somente com o segundo quadrante, pois foi o que mais teve equilíbrio na matriz de confusão, com a análise dos dados foi feita a API usando o *framework* Flask, as Figura 19 mostra trechos do código.

Figura 19 – Trecho do código da API.

```

File Edit Selection View Go Run Terminal Help
model.py - Backend - Visual Studio Code

EXPLORER
- BACKEND
  - datasets
  - images
  - etc
  - _pycache_
  - api.py
  - model.py
  - env

model.py
1 import numpy as np
2
3
4
5 > def getListOfFiles(filename):
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

Fonte: Elaborado pela autora.

6 DISCUSSÃO DOS RESULTADOS

Trabalhar com Inteligência Artificial não é uma tarefa trivial, mesmo utilizando modelo de aprendizagem supervisionada. Ao longo do projeto foram feitos inúmeros testes até obter um resultado satisfatório.

Inicialmente a ideia era repartir a imagem em quadrantes, como foi mostrado na Figura 7, extrair os dados de cada quadrante, e montar uma matriz de característica a partir dele. O problema é que nesse formato houve muitos falsos positivos, sendo o pior cenário possível no contexto da saúde. Por isso, foram feitos outros experimentos, como mostra-se nas Figuras 14 a 18, tendo como melhor resultado trabalhar apenas com o segundo quadrante que obteve melhor resultado na matriz de confusão.

Mesmo com o projeto concluído, existem inúmeros testes a serem feitos para melhoria do projeto, como confrontar os dados com outros modelos de aprendizagem; testar outras bases de dados; melhorar o algoritmo de processamento de imagem. Essas sugestões auxiliariam na validação de algumas questões, tais como:

- a) Será que o segundo quadrante comporta todos os casos de Covid (estado inicial da doença ao estado avançado);
- b) Levando em consideração os danos no pulmão, a pneumonia possui características semelhantes quando comparada ao Covid-19 a nível de imagem. Por questões de tempo hábil, a pesquisa não efetuou testes com bases de pneumonia para verificar como o algoritmo se comportaria.

Vemos os tópicos citados anteriormente como os principais a serem feitos em trabalhos futuros, e tem-se ciência que existem outras questões a serem melhoradas – como já mencionado – sendo assim, também não se descarta tais questões em trabalhos futuros.

7 CONSIDERAÇÕES FINAIS

Ao final da pesquisa foi obtido um resultado positivo, porém há melhorias que podem ser feitas no futuro como o processamento de imagens, melhorando a forma de identificar a região somente do pulmão, pois há raio x que acabam pegando os braços o que atrapalha na análise, a comparação com outros algoritmos de IA para analisar o desempenho com a base e uma melhora na parte visual do sistema.

REFERÊNCIAS

BAI, Harrison X. et al. Performance of radiologists in differentiating COVID-19 from non-COVID-19 viral pneumonia at chest CT. **Radiology**, v. 296, n. 2, p. E46-E54, 2020.

BRAMER, Max. Principles of data mining. London: Springer, 2007.

CHUNG, Michael et al. CT imaging features of 2019 novel coronavirus (2019-nCoV). **Radiology**, v. 295, n. 1, p. 202-207, 2020.

G1. Disponível em

<<https://g1.globo.com/bemestar/coronavirus/noticia/2020/03/26/1-mes-de-coronavirus-no-brasil-compare-a-situacao-do-pais-com-china-italia-eua-e-coreia-do-sul-no-mesmo-periodo-da-epidemia.ghtml>>. Acessado em: 01 abr. 2020.

KEELE, Staffs et al. **Guidelines for performing systematic literature reviews in software engineering**. Technical report, Ver. 2.3 EBSE Technical Report. EBSE, 2007.

LI, Lin et al. Artificial Intelligence Distinguishes COVID-19 from Community Acquired Pneumonia on Chest CT. **Radiology**, p. 200905, 2020.

Olhar Digital. Disponível em: <https://olhardigital.com.br/coronavirus/noticia/brasileiros-unem-inteligencia-computacional-e-noticias-para-prever-casos-de-covid-19/98685?fbclid=IwAR24471Tixxj0OvYpaJ6CfhyuV-Egn5JJiSsgeJeHYRAZGtT_MtmPXxnGjs>. Acessado em: 01 abr. 2020.

ROSA, Marcela Emer Egypto et al. Achados da COVID-19 identificados na tomografia computadorizada de tórax: ensaio pictórico. **Einstein (São Paulo)**, v. 18, 2020.

Terra. Disponível em: <https://www.terra.com.br/noticias/tecnologia/inteligencia-artificial-vira-arma-na-luta-contra-o-coronavirus,31221fe9240a5c5f2e9122b0f8bb8714j4av4fb6.html?utm_source=NEWSSTAND&utm_medium=rss&fbclid=IwAR3vIYBV/s9qcKpNPYFeqRNjOd3Mpem1iZtACqm4c0m7XhS6QaSKK88jnw_A>. Acessado em: 01 abr. 2020.

UOL. Disponível em <<https://www.uol.com.br/tilt/noticias/afp/2020/03/26/a-inteligencia-artificial-para-a-prevencao-de-pandemias-como-a-da-covid-19.htm>> Acessado em 01 abr. 2020.

8 ANEXO I - CARTA DE DISPENSA DE APRESENTAÇÃO AO CEP OU CEUA**CARTA DE DISPENSA DE APRESENTAÇÃO AO CEP OU CEUA**

À

COORDENADORIA DO PROGRAMA DE INICIAÇÃO CIENTÍFICA DA UNISAGRADO

Informo que não é necessária a submissão do projeto de pesquisa intitulado SISTEMA ESPECIALISTA INTELIGENTE PARA IDENTIFICAÇÃO DE COVID-19, ao Comitê de Ética em Pesquisa com Seres Humanos (CEP) ou à Comissão de Ética no Uso de Animais (CEUA) devido à pesquisa não envolver seres humanos nem animais, pois só utilizará métodos de computação consagrados na literatura, programação/simulação e acesso a dados públicos da internet.

Atenciosamente,

Vinicius Santos Andrade

Vinicius Santos Andrade

Bauru, 01 de abril de 2020.