



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE ENGENHARIA ELÉTRICA
ENGENHARIA BIOMÉDICA

MANUELA RODRIGUES DE SOUZA E LIMA

ANÁLISE DE IMAGENS RADIOGRÁFICAS PULMONARES USANDO
ATRIBUTOS ESTATÍSTICOS E DE TEXTURA

UBERLÂNDIA
2022

MANUELA RODRIGUES DE SOUZA E LIMA

**ANÁLISE DE IMAGENS RADIOGRÁFICAS PULMONARES USANDO
ATRIBUTOS ESTATÍSTICOS E DE TEXTURA**

Trabalho de Conclusão de Curso apresentada como requisito parcial de aprovação da disciplina de Trabalho de Conclusão de Curso de Engenharia Biomédica da Faculdade de Engenharia Elétrica da Universidade Federal de Uberlândia.

Orientador: Ana Cláudia Patrocínio

UBERLÂNDIA

2022

Dedico esse trabalho à todos que me apoiaram
nessa jornada.

AGRADECIMENTOS

Primeiramente gostaria de agradecer aos meus pais Marta e Aloísio e a minha irmã Bárbara por estarem presentes durante as minhas escolhas, por me apoiarem e comemorarem comigo cada vitória. Sem eles, sem sombra de dúvidas, seria impossível a conclusão desta etapa da minha vida.

Queria agradecer, também, à todos os meus amigos. Quando estamos longe de casa sabemos a importância de encontrar pessoas que nos acolhem e cuidam da gente com carinho e preocupação. E, por isso, gostaria de deixar um pedacinho para agradecer os amigos com quem morei e ao meu namorado, pessoas maravilhosas que seguraram a minha mão nos momentos de dificuldades e que se fizeram família quando a minha estava distante.

Além de amigos de morada, tive, ainda, a sorte de encontrar companheiros no curso, pessoas que estão vivendo os mesmos momentos que nós e que entendem nossas crises. Quando as coisas apertaram, sempre tive pessoas que me deram forças e, por causa delas, consegui chegar até aqui. Em meio a várias pessoas, tive uma amiga em especial com quem pude compartilhar tudo e sempre estive ao meu lado, não é Lari?

Gostaria de agradecer, também, a todos meus professores que deram o máximo de si para transmitir conhecimento e moldaram minha educação e a de centenas de outros estudantes. E, por fim, um agradecimento especial à minha orientadora Ana Cláudia Patrocínio pela paciência e por conseguir de maneira ímpar dar o suporte necessário para este trabalho.

RESUMO

A COVID-19 é uma doença infectocontagiosa que se tornou uma pandemia em 2020 e um dos métodos utilizados para auxílio ao prognóstico foram os exames de imagem, à exemplo da radiografia de tórax. Um dos grandes problemas desse exame é a dificuldade de interpretação, agravado pelo fato de se tratar de uma doença ainda pouco conhecida e que apresenta sintomas parecidos com de outras patologias. Considerando isso, esse trabalho se propôs a analisar a média, desvio padrão, curtose, e assimetria e os principais descritores de textura de Haralick de imagens radiográficas de pacientes com COVID-19 e de exames laudados com diferentes doenças pulmonares e sem achado radiológico, a fim de contribuir com a caracterização geral de exames de radiografia de tórax da doença e investigar variáveis que auxiliem efetivamente na identificação do SARS-CoV-2. Como resultado desse trabalho foi possível perceber que os exames nos quais há presença da COVID-19 possuem características estatísticas e de textura consideravelmente diferentes das outras anomalias pulmonares, foi observado um padrão maior de uniformidade e homogeneidade locais em exames de radiografia de tórax de pacientes diagnosticados com COVID-19, o que pode ser identificado ao analisar os maiores valores de Segundo Momento Angular, Correlação e Momento de Diferença Inverso, simultaneamente com os menores valores de Contraste e Entropia dos atributos de textura de Haralick dessas imagens. Além disso, nos exames de COVID-19 foi identificado maior padrão de densidade e dispersão dos pixels demonstrado pelos discrepantes maiores valores de média e desvio padrão. Com isso, essas variáveis podem ser estudadas com o objetivo de serem utilizadas para o auxílio médico quando há incerteza do prognóstico durante a visualização do exame.

Palavras-chave: Extração de Atributos. Radiografia. COVID-19. Descritores de Textura.

ABSTRACT

COVID-19 is an infectious disease that became a pandemic in 2020 and one of the methods used to aid in the prognosis were imaging tests, such as chest radiography. One of the major problems of this exam is the difficulty of interpretation, aggravated by the fact that it is a disease still little known and that presents symptoms similar to other pathologies. Considering this, this study aimed to analyze the mean, standard deviation, kurtosis, and skewness and the main Haralick texture features of radiographic images of patients with COVID-19 and of exams reported with different pulmonary diseases and with no radiological finding, in order to contribute to the general characterization of chest radiography exams of this disease and to investigate variables that effectively help in the identification of SARS-CoV-2. As a result of this work, it was possible to notice that COVID-19 exams have statistical and texture features considerably different from those of other pulmonary anomalies, a pattern of local uniformity and homogeneity was observed in chest radiography exams of patients diagnosed with COVID-19, which can be identified by analyzing the highest values of Angular Second Moment, Correlation and Inverse Difference Moment, simultaneously with the lowest values of Contrast and Entropy of Haralick features of these images. In addition, in the COVID-19 exams, a pattern of density and dispersion of pixels was identified, demonstrated by the discrepant higher values of mean and standard deviation. Thus, these variables can be studied with the aim of being used for medical assistance when there is uncertainty in the prognosis during the examination visualization.

Keywords: Feature Extraction. Radiography. COVID-19. Texture Features.

LISTA DE FIGURAS

1	Gráfico de representação de uma distribuição normal	12
2	Processo de formação da Gray Level Co-occurrence Matrix(GLCM)	13
3	Fluxograma da metodologia	17
4	Gráfico com valores de Média e Desvio Padrão	18
5	Gráfico com valores de Curtose e respectivo desvio padrão	19
6	Gráfico com valores de Assimetria e respectivo desvio padrão	19
7	Gráfico com valores de Segundo Momento Angular e respectivo desvio padrão	20
8	Gráfico com valores de Contraste e respectivo desvio padrão	20
9	Gráfico com valores de Correlação e respectivo desvio padrão	21
10	Gráfico com valores de Variância e respectivo desvio padrão	21
11	Gráfico com valores de Momento de Diferença Inverso e respectivo desvio padrão	22
12	Gráfico com valores de Média da Soma e respectivo desvio padrão	22
13	Gráfico com valores de Variância da Soma e respectivo desvio padrão	23
14	Gráfico com valores de Entropia da Soma e respectivo desvio padrão	23
15	Gráfico com valores de Entropia e respectivo desvio padrão	23
16	Gráfico com valores de Variância da Diferença e respectivo desvio padrão	24
17	Gráfico com valores de Entropia da Diferença e respectivo desvio padrão	24

LISTA DE ABREVIATURAS E SIGLAS

BIMCV – Banco digital de Imagen Médica de la Comunidad Valenciana

COVID-19 – Coronavirus Disease of 2019

EAS – Estabelecimento Assistencial de Saúde

GLCM – Gray Level Co-occurrence Matrix

IgG – Immunoglobulin G

IgM – Immunoglobulin M

MATLAB – Matrix Laboratory

RT-PCR – Reverse Transcription Polymerase Chain Reaction

SARS-CoV-2 – Severe Acute Respiratory Syndrome Coronavirus 2

WHO – World Health Organization

SUMÁRIO

1	INTRODUÇÃO	9
1.1	OBJETIVO GERAL	10
1.2	OBJETIVOS ESPECÍFICOS	10
2	DESENVOLVIMENTO	11
2.1	FUNDAMENTAÇÃO TEÓRICA	11
2.1.1	Exame de Radiografia	11
2.1.2	Extração de Atributos	11
2.1.2.1	Atributos Estatísticos	12
2.1.2.2	Descritores de Textura de Haralick	13
2.1.3	Doenças Pulmonares	14
2.2	MATERIAIS E MÉTODOS	15
3	RESULTADOS E DISCUSSÕES	18
4	CONCLUSÕES	26
	REFERÊNCIAS	27

1 INTRODUÇÃO

A COVID-19 (Coronavirus Disease of 2019) é uma doença infectocontagiosa que se tornou uma pandemia em 2020 e ainda afeta drasticamente o planeta, registrando mais de 800 mil casos e cerca de 1500 mortes por dia em todo o mundo (World Health Organization (WHO), 2022). Essa patologia, além de afetar rapidamente o sistema cardiorrespiratório, é extremamente transmissível e desde o início da sua investigação foram encontrados e desenvolvidos diversos tipos de métodos diagnósticos com a finalidade de auxiliar na rápida identificação da doença para, assim, administrar o tratamento adequado e realizar o isolamento do paciente. Um dos métodos utilizados foram os exames de imagem, à exemplo da radiografia de tórax. (KIM *et al.*, 2020).

O equipamento de radiografia é um aparelho amplamente disponível nos estabelecimentos assistenciais de saúde (EASs), possui versões portáteis e de fácil desinfecção, além de possibilitar a visualização do acometimento do sistema respiratório, sendo assim, de grande valia para o prognóstico do Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). Porém, um dos grandes problemas desse exame é a dificuldade de interpretação, agravado pelo fato da COVID-19 se tratar de uma doença ainda pouco conhecida e sem histórico de estudo, que apresenta sintomas parecidos com outras patologias que atingem os pulmões, como as pneumonias (BRESTEL *et al.*, 2018), (ISLAM *et al.*, 2021).

Considerando isso, recentemente, diversas técnicas relacionadas com o processamento dessas imagens médicas começaram a ser estudadas aspirando facilitar a identificação precisa da doença, como nos trabalhos de (VARELA-SANTOS; MELIN, 2021), (APOSTOLOPOULOS; MPESIANA, 2020), (GOMES *et al.*, 2020), (HASSANTABAR; AHMADI; SHARIFI, 2020) e (WONG *et al.*, 2020). Um tipo de processamento largamente conhecido é a extração de características a partir da matriz de tons de cinza dos pixels que formam a imagem, a qual possui o grande benefício da utilização de baixo processamento computacional e geração de resultados rápidos e de fácil análise (LÖFSTEDT *et al.*, 2019).

Alguns dos atributos de imagem comuns de serem avaliados são variáveis estatísticas que levam em consideração os valores dos pixels e a frequência em que aparecem em uma imagem, como média, desvio padrão, curtose e assimetria. Outro tipo de atributo são os descritores de textura que representam de maneira mais detalhada os tons de cinza analisando também a interação espacial entre os pixels, à exemplo dos descritores de textura de Haralick (HARALICK, 1979).

Considerando isso, esse trabalho se propôs a analisar a média, desvio padrão, curtose, e assimetria e os principais descritores de textura de Haralick de imagens radiográficas de pacientes com COVID-19 e de exames laudados com diferentes doenças pulmonares e sem achado radiológico, a fim de contribuir com a caracterização geral de exames de radiografia de tórax da doença e investigar variáveis que auxiliem efetivamente na identificação do SARS-CoV-2.

1.1 OBJETIVO GERAL

Esse trabalho possui como objetivo geral a extração de atributos de imagens médicas em vista da análise de padrões para diferenciação de distúrbios pulmonares e identificação da COVID-19.

1.2 OBJETIVOS ESPECÍFICOS

Os objetivos específicos são:

- extração de atributos estatísticos e de textura de exames de radiografia de pacientes acometidos com COVID-19, atelectasia, consolidação pulmonar, desordem pleural, edema pulmonar, opacidade pulmonar, pneumonia e sem achado radiográfico;
- análise de atributos que podem ser utilizados para o auxílio ao prognóstico médico da patologia.

2 DESENVOLVIMENTO

2.1 FUNDAMENTAÇÃO TEÓRICA

2.1.1 Exame de Radiografia

O exame de radiografia é um dos métodos utilizados como auxílio ao prognóstico de diversas patologias médicas e é formado a partir da impressão, em um detector, do feixe de raios X que atravessa alguma parte do corpo do paciente. Esse tipo de exame é possibilitado porque os tecidos do corpo humano atenuam de maneira diferente a energia ionizante gerada pelo equipamento e, assim, é possível distinguir as estruturas e identificar alguma irregularidade ali existente.

A efetividade da técnica parte do princípio básico de que quanto maior a densidade e dureza do tecido, maior será a quantidade de raios X absorvida por ele, menor será a energia transmitida, de modo que tecidos como os ossos atenuam mais radiação e menos intensidade de fótons é transmitida sendo então menos detectada no sensor (detector), sendo identificados pelas cores mais esbranquiçadas dentro da faixa de tons de cinzas do exame. Já tecidos moles e glândulas, por serem menos densos, permitem a passagem de maior quantidade de raios e, assim, impressionam a placa com mais intensidade, sendo identificados por tons mais próximos ao preto (WOODS; GONZALEZ, 2010).

Nos exames de radiografia de tórax, o pulmão, quando saudável, é apresentado totalmente em preto, já que em seu interior os raios X atravessam apenas o ar. Porém, quando existe alguma substância mais densa que o ar dentro da cavidade, à exemplo da presença processo inflamatório por vírus ou bactérias, a imagem gerada se apresentará variação de intensidade de tons de cinza e variação na textura da imagem, caracterizando alguma anomalia (WOODS; GONZALEZ, 2010).

2.1.2 Extração de Atributos

A partir de uma imagem é possível extrair informações que possam corroborar com o reconhecimento e a caracterização de padrões. A extração de atributos, ou características, de imagens médicas é um método que permite estudar as informações que podem ser obtidas através das matrizes de pixels que as formam. Ela permite analisar imagens de um novo aspecto, diferente da investigação visual do exame em si, em busca de padrões a partir dos valores dessas variáveis que sejam relevantes para prognóstico de alguma patologia, nesse caso, a COVID-19. Para confecção deste trabalho serão exploradas as características estatísticas das imagens e os atributos de textura de Haralick.

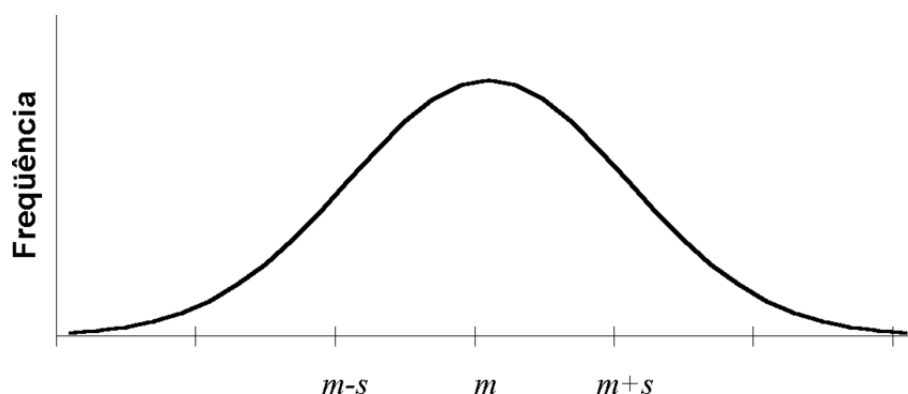
2.1.2.1 Atributos Estatísticos

As características resultantes dos atributos estatísticos possuem o objetivo de relacionar as propriedades dos valores dos tons de pixel e a suas distribuições. À exemplo desses atributos, têm-se: a média, o desvio padrão, a curtose, a assimetria, entre outros.

A média (RODRIGUES; LIMA; BARBOSA, 2017) é uma medida importante com a função de representar, de maneira geral, um extenso conjunto de dados. No estudo das características das imagens ela é um descritor relacionado com a tendência dos valores dos pixels, e, por isso, quanto maior seu valor, maiores são os valores dos tons de cinza que compõem a imagem, ou seja, aproximando-se da cor branca da escala. Para complementar a análise, estuda-se o valor do desvio padrão (RODRIGUES; LIMA; BARBOSA, 2017), uma variável que identifica como os valores dos tons de cinza da matriz de pixels da imagem estão variando em relação à média calculada, quanto maior o desvio padrão, mais pixels com tons de cinza distantes da média estão presentes na imagem estudada.

A curtose e a assimetria de acordo com (FILLIBEN; HECKERT, 2002) são atributos que dizem respeito à forma de dispersão dos valores de tons de pixel de uma imagem, de modo a descreverem como o conjunto de dados está disposto quando comparado a uma distribuição normal. A distribuição normal ou distribuição gaussiana é aquela em que a média é o valor de maior frequência no conjunto de dados e representa o centro de uma curva simétrica. Para melhor visualização, abaixo está indicado sua curva (Figura 1):

Figura 1 – Gráfico de representação de uma distribuição normal



Fonte: (RONDON; MURAKAMI; SAKAGUTI, 2002).

A figura acima representa uma distribuição normal de média m e variância s^2 .

Levando em consideração o conceito anterior, a curtose indica o achatamento da curva de distribuição dos tons da imagem em relação à distribuição normal, de maneira que o valor de curtose igual a zero é resultante de uma distribuição gaussiana. Uma curtose negativa indica maior frequência de valores de pixels distantes do centro da dispersão normal e, portanto, curva mais achatada. Já o valor positivo demonstra que o conjunto de dados está concentrado no centro do gráfico e menor o achatamento da curva.

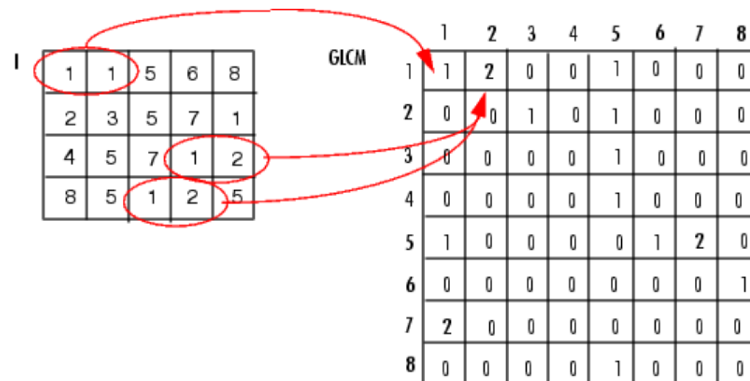
A variável assimetria descreve o quanto a imagem está simétrica ou não em relação à distribuição normal dos valores de pixels. Sua interpretação é realizada através do valor e de seu sinal, tal que o sinal positivo representa que a curva do gráfico é distorcida para a direita e o resultado negativo indica deformidade para a esquerda.

2.1.2.2 Descritores de Textura de Haralick

Os descritores de textura de Haralick são parâmetros que refletem as variações de tons de cinza de uma imagem, considerando, além de seus respectivos valores, a interação espacial entre pixels vizinhos (HARALICK, 1979). Para isso, as características em questão são calculadas a partir da matriz de co-ocorrência de tons de cinza, a Gray Level Co-occurrence Matrix (GLCM).

A GLCM é uma matriz quadrada, de tamanho definido pela quantidade de tons de cinza existentes na imagem que está sendo analisada, tal que imagens formadas por 8 bits (contendo 256 tons de cinza) formarão uma matriz de co-ocorrência de 256 linhas e colunas. Sua criação se dá através da varredura da matriz de pixels da imagem contabilizando a quantidade de vezes em que um valor de tom de cinza referência aparece vizinho a outro valor de tom de cinza específico, a uma determinada distância entre eles e um respectivo ângulo (SCHWARTZ; PEDRINI, 2003). A imagem abaixo demonstra esse processo para uma matriz de 3 bits, utilizando a distância de 1 pixel e ângulo de 0° (Figura 2):

Figura 2 – Processo de formação da Gray Level Co-occurrence Matrix (GLCM)



Fonte: (The MathWorks, Inc., 2005).

Para o cálculo das características mais conhecidas definidas por Haralick é necessário determinar a matriz de co-ocorrência nos ângulos 0° , 45° , 90° e 135° . Com o objetivo de identificar as propriedades de textura da imagem, dos 14 atributos de textura de Haralick, serão calculadas e analisadas, neste trabalho, as variáveis listadas e descritas abaixo (MATERKA; STRZELECKI *et al.*, 1998), (SANTOS *et al.*, 2015):

1. Segundo Momento Angular (Angular Second Moment - ASM): mede a uniformidade local da distribuição de pixels, também é conhecido como Energia e seu valor pode variar entre 0 e 1.

2. **Contraste (Contrast):** característica que expressa a variação de tons da imagem, definida a partir da diferença entre valores de cinza.
3. **Correlação (Correlation):** correlação é a medida da linearidade de tons de cinza em uma imagem.
4. **Variância (Variance):** calcula a dispersão da distribuição dos tons de cinza.
5. **Entropia (Entropy):** a entropia mede a quantidade de sinal da imagem de modo a analisar a desordem em sua textura, representa aleatoriedade dos pixels.
6. **Momento de Diferença Inverso (Inverse Difference Moment):** variável que indica a homogeneidade da matriz de pixels, está relacionada de maneira oposta ao contraste.
7. **Média da Soma (Sum Average):** indica a média da soma dos tons, auxilia na identificação de áreas densas e não densas de uma mesma imagem.
8. **Variância da Soma (Sum Variance):** calcula a dispersão da soma dos valores dos pixels que formam a imagem.
9. **Entropia da Soma (Sum Entropy):** variável que representa a desordem da soma dos tons de cinza a partir da GLCM.
10. **Variância da Diferença (Difference Variance):** atributo relacionado com a dispersão local da variação da distribuição da diferença dos tons de cinza da matriz.
11. **Entropia da Diferença (Difference Entropy):** está relacionado com a desordem da diferença dos tons de pixels.

2.1.3 Doenças Pulmonares

Neste trabalho para a comparação com os exames de COVID-19 foram utilizados um grupo de imagens sem nenhuma patologia (No findings) e imagens com 7 achados radiográficos pulmonares, os quais são:

1. **Atelectasia (Atelectasis):** a atelectasia é um tipo de achado radiológico relacionado com obstrução pulmonar e descreve um estado de colapso de alguma região pulmonar ou de todo o pulmão (PERONI; BONER, 2000).
2. **Cardiomegalia (Cardiomegaly):** a cardiomegalia é uma condição de aumento do coração do paciente e acometimento do bombeamento adequado do sangue. Pode ocorrer por diversas causas, como infecções e anemias (WJ, 2021).
3. **Consolidação Pulmonar (Pulmonary Consolidation):** a consolidação pulmonar é a caracterização de um acúmulo de substâncias líquidas e/ou sólidas no interior do pulmão, onde

deveria estar preenchido apenas por ar. A causa mais comum é a pneumonia (MURPHY; HACKING, 2019).

4. Desordem Pleural (Pleural Other): desordem pleural é um tipo de anomalia que acomete a pleura pulmonar, geralmente relacionada com entrada de sangue ou ar na cavidade pulmonar (National Heart, Lung, and Blood Institute (NHLBI), 2022).
5. Edema pulmonar (Pulmonary Edema): o edema pulmonar é o acúmulo de líquidos no pulmão, à exemplo do sangue. A principal causa é a hipertensão dos vasos sanguíneos que realizam a troca gasosa pulmonar (STAUB, 1974).
6. Opacidade Pulmonar (Lung Opacity): a opacidade pulmonar é um achado radiográfico que representa uma área no pulmão mais opaca que o normal. Existem diversos padrões de opacidade, como a opacidade linear e a opacidade nodular (KNIPE; JONES, 2011).
7. Pneumonia: a pneumonia é uma doença que representa alguma infecção nos pulmões causada por vírus, bactérias ou outros agentes e está relacionada com a dificuldade de realizar trocas gasosas (Drauzio Varella, 2011).

2.2 MATERIAIS E MÉTODOS

Para a realização deste trabalho foram utilizados dois bancos de imagens:

- CheXpert (IRVIN *et al.*, 2019): banco de imagens construído por um grupo de pesquisadores da Universidade de Stanford (Stanford Machine Learning Group). Esse banco de imagens possui exames de radiografia de tórax de mais de 65 mil pacientes adquiridos no Hospital da Universidade de Stanford e laudados por médicos radiologistas. Para autorização de uso foi enviado um e-mail solicitando o acesso e utilização para, então, ser realizado o download das imagens. Deste, das 14 classificações radiográficas, foram utilizadas as imagens de pacientes diagnosticados com Atelectasia, Cardiomegalia, Consolidação Pulmonar, Edema Pulmonar, Opacidade Pulmonar, Desordem Pleural, Pneumonia e nenhum achado radiográfico (no findings). O banco é formado por imagens de 8 bits em formato JPG.
- BIMCV-COVID19+ (MDLI *et al.*, 2020): banco de imagens em formato PNG de radiografia de tórax e de tomografia de pacientes diagnosticados COVID-19 da Região de Valência na Espanha. Os exames de imagem foram atrelados à resultados de Reverse Transcription Polymerase Chain Reaction (RT-PCR), Immunoglobulin G (IgG) e Immunoglobulin M (IgM) para confirmação da presença da doença. Desse banco, foram utilizadas apenas as imagens de radiografia, as quais foram convertidas para JPG a fim de normalizar os bancos no mesmo formato.

Ambos os bancos possuem imagens de pacientes em idade adulta, tanto do sexo masculino como do sexo feminino e todas as imagens foram anonimizadas.

A primeira etapa do trabalho foi a criação de uma pasta de imagens para cada doença relacionada anteriormente, sendo selecionadas apenas as imagens de radiografia de incidência anteroposterior, sendo removidas todas as outras incidências. Como critérios de exclusão também foram utilizados:

- imagens que possuíam algum tipo de sobreposição;
- imagens que possuíam aparelhos de suporte à vida e cabos como prevalência no exame radiográfico;
- imagens que não possuíam o tórax como foco.

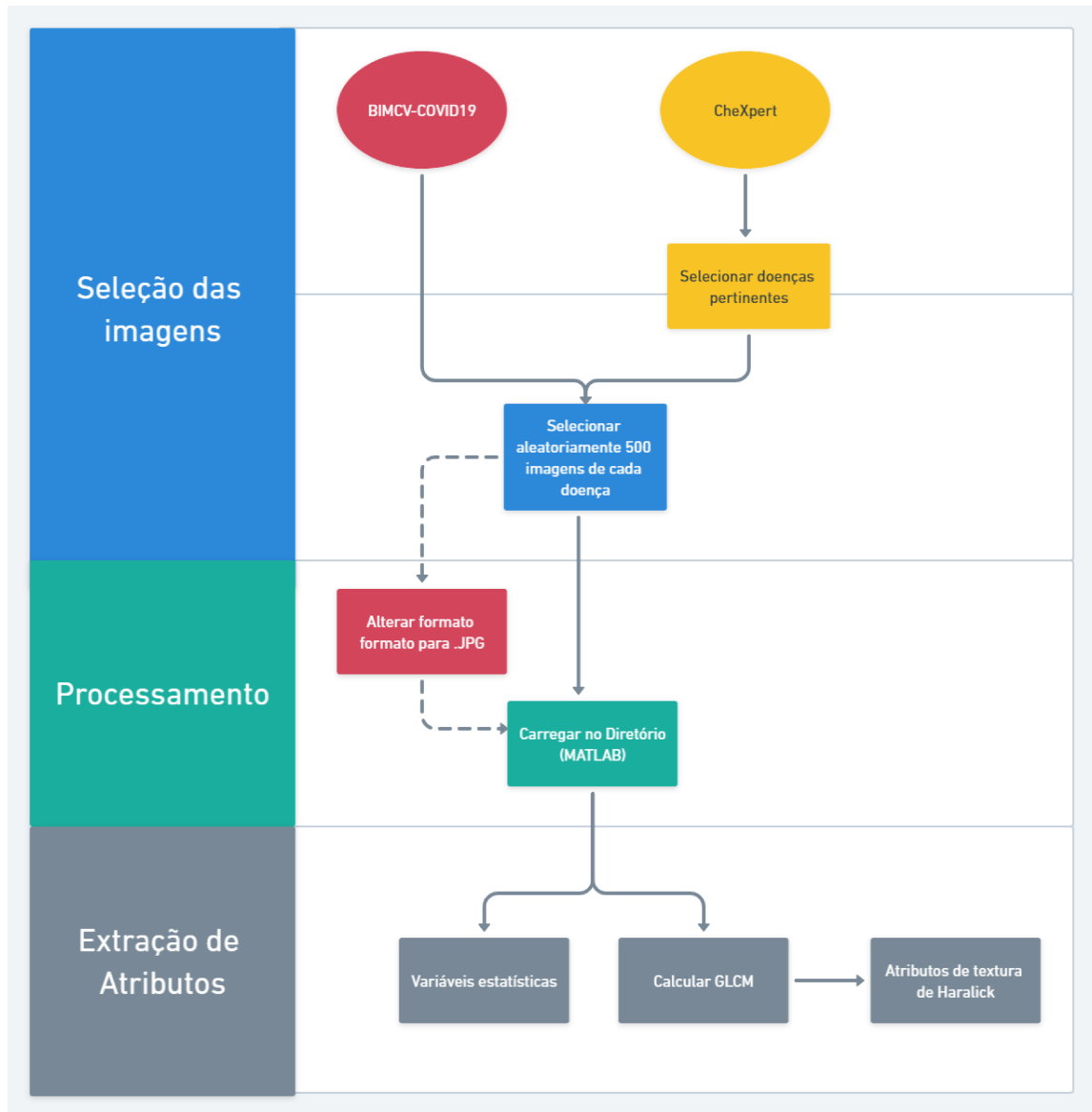
Após a eliminação, foram escolhidas, aleatoriamente, 500 imagens de cada anomalia pulmonar para serem alvo de estudo. A partir de então, foi iniciada a extração de características. Para tal, foi criado um código no programa Matlab com a função inicial de abrir a pasta com as imagens da doença selecionada, carregá-las no diretório e construir um vetor de nomes das imagens (que indicam o paciente a qual está relacionada), extrair as características de interesse e, ao final, criar uma planilha de Excel com as informações adquiridas.

A primeira série de atributos extraídos por meio do programa foram as variáveis estatísticas: média, desvio padrão, curtose e assimetria. A média e o desvio padrão foram calculados a partir do histograma de cada imagem e as duas outras características foram obtidas através da própria matriz de pixels dos exames.

A segunda série de atributos extraídos foram os descritores de textura de Haralick, para essa etapa foi necessário calcular a GLCM da distribuição de tons de cinza nos quatro ângulos definidos: 45°, 90°, 135° e 180°, sendo a distância entre pixels utilizada para varredura igual a 1. Essa matriz foi utilizada como variável de entrada da função que realiza a extração dos principais descritores de Haralick e retorna seus respectivos valores na seguinte ordem: Segundo Momento Angular, Contraste, Correlação, Variância, Momento de Diferença Inverso, Soma da Média, Soma da Variância, Soma da Entropia, Entropia, Variância da Diferença e Entropia da Diferença.

Para a melhor visualização, foi esboçado um diagrama para ilustrar a metodologia desenvolvida anteriormente (Figura 3):

Figura 3 – Fluxograma da metodologia



Fonte: Elaborado pelo autor (2022).

Como pode ser visto, a metodologia deste trabalho é dividida, basicamente, em 3 etapas: seleção das imagens, processamento e extração de atributos. Os círculos indicam os dois bancos de imagens utilizados, sendo o uso da cor vermelha associada ao BIMCV-COVID19+ e da amarela ao CheXpert.

Após a extração de atributos, foram exportadas, através do programa Matlab, uma planilha para cada anomalia pulmonar e uma para os exames sem achado radiográfico, onde estavam indicados os resultados dos atributos calculados para as imagens.

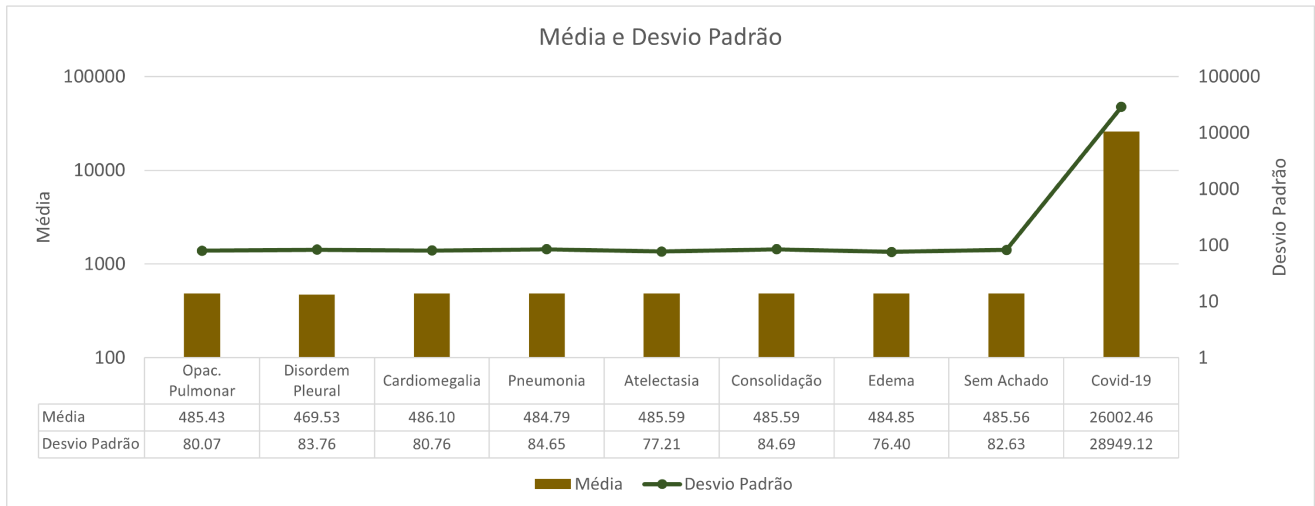
Dentro de cada planilha foi então calculada a média e o desvio padrão de cada característica, obtendo, ao final, o valor médio e o desvio padrão médio de cada atributo para cada patologia. E, por fim, foi calculada a média dos valores dos quatro ângulos dos atributos de Haralick.

3 RESULTADOS E DISCUSSÕES

Para apresentação dos resultados, os valores obtidos em cada planilha foram exibidos através de gráficos de barras facilitando a análise. As barras em marrom representam a média dos valores do atributo calculada para cada doença discutida e a linha em verde a média do desvio padrão dessa variável. A partir das informações a seguir, podem ser definidas as características nas quais os valores foram mais discrepantes nos exames de pacientes com COVID-19 e, assim, utilizá-las na contribuição do prognóstico da doença.

O primeiro gráfico é a representação dos valores dos atributos média e desvio padrão (Figura 4):

Figura 4 – Gráfico com valores de Média e Desvio Padrão

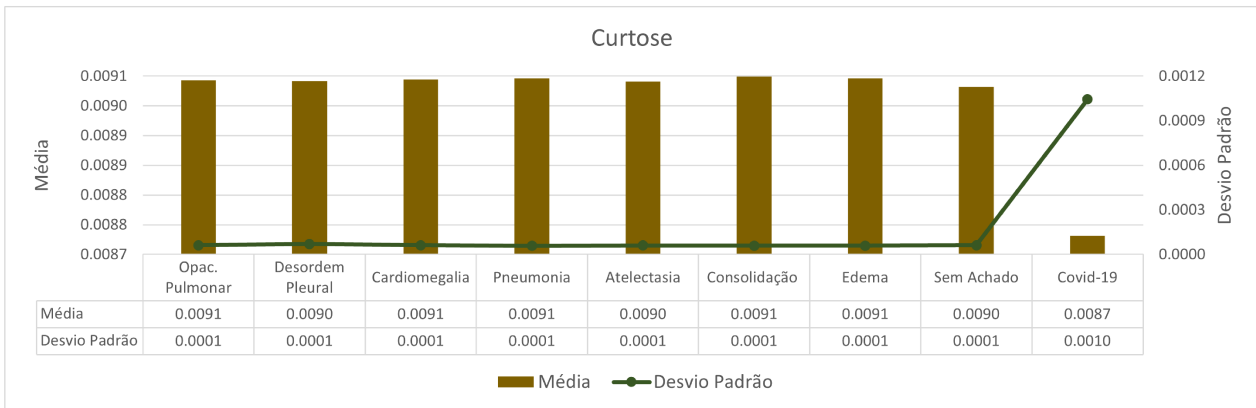


Fonte: Elaborado pelo autor (2022).

Na imagem anterior os eixos da esquerda (valores de média) e da direita (desvio padrão) são representados em escala logarítmica para melhor percepção de seus valores, já que as imagens com presença de COVID-19 resultaram em valores extremamente maiores que os das outras doenças e das imagens sem achado radiográfico, o que indica que a matriz de tons de cinza possuem cores mais claras e, portanto, existe maior densidade dentro dos pulmões do paciente. De mesmo modo, o alto desvio padrão demonstra que grande quantidade de pixels foge dos valores da média.

A seguir está indicado o gráfico dos valores de curtose e seu respectivo desvio padrão (Figura 5):

Figura 5 – Gráfico com valores de Curtose e respectivo desvio padrão

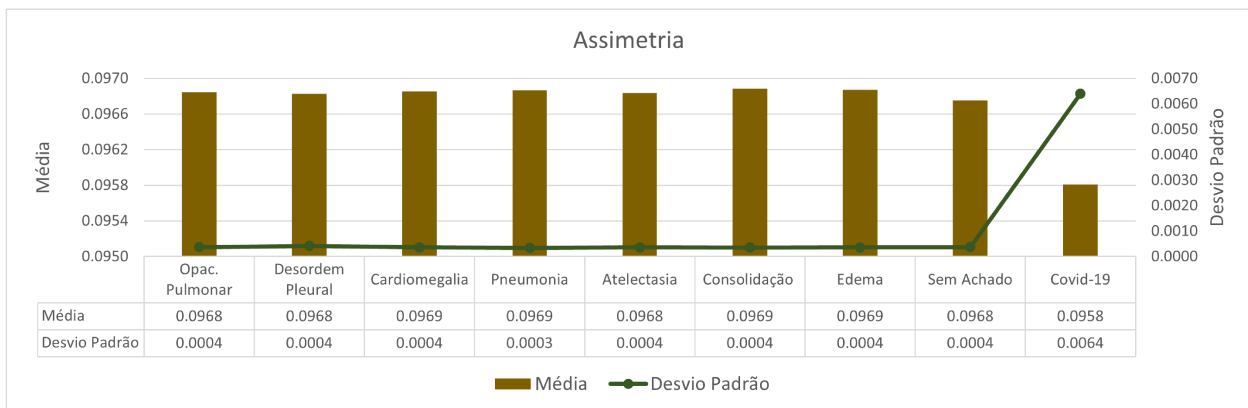


Fonte: Elaborado pelo autor (2022).

Como pode ser observado, os valores de curtose para todas as doenças pulmonares e imagens sem achado radiográfico indica valores positivos maiores que as imagens de pacientes diagnosticados com COVID-19, representando que a curva de tons de cinza de exames com presença de SARS-CoV-2 está mais próxima de uma distribuição gaussiana, enquanto as outras doenças possuem essa curva mais achatada. Já o desvio padrão elevado demonstra que apesar disso, a curva de dispersão de tons de cinza das imagens com COVID-19 possui mais valores muito distantes da média que a das outras doenças.

A curva do atributo estatístico Assimetria pode ser observada a seguir (Figura 6):

Figura 6 – Gráfico com valores de Assimetria e respectivo desvio padrão



Fonte: Elaborado pelo autor (2022).

Considerando a interpretação da variável Assimetria pode-se entender que a curva de dispersão dos valores de tons de cinza de todas as doenças representadas acima estão distorcidas para direita em relação a média, mas nas imagens radiográficas contendo infecção pelo SARS-CoV-2 essa distorção é um pouco menor.

Já para a variável Segundo Momento Angular dos atributos de textura de Haralick, os resultados estão apresentados na Figura 7:

De acordo com os valores representados na imagem acima, é possível concluir que os

Figura 7 – Gráfico com valores de Segundo Momento Angular e respectivo desvio padrão

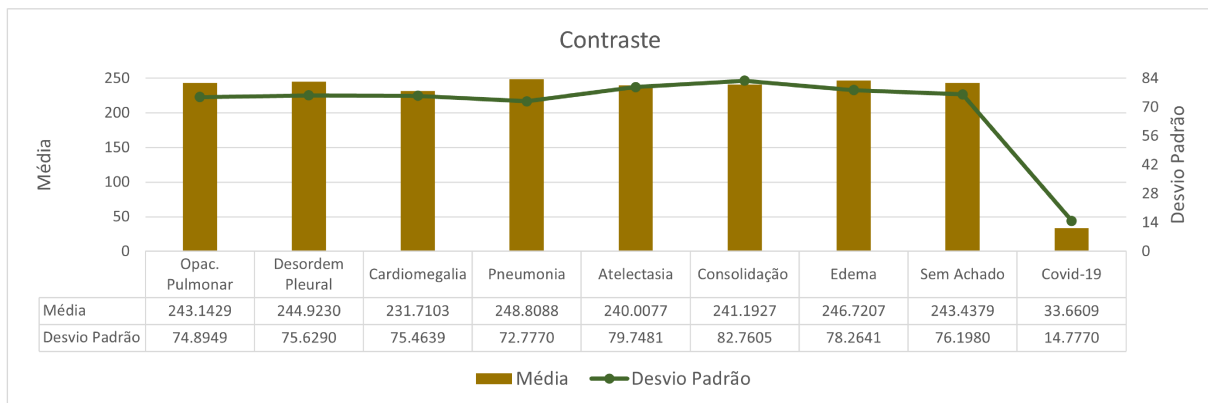


Fonte: Elaborado pelo autor (2022).

exames nos quais estão diagnosticados a COVID-19 possuem distintamente maior uniformidade local em dissonância com as outras doenças avaliadas.

Em paralelo, os números resultantes para o Contraste são descritos na Figura 8:

Figura 8 – Gráfico com valores de Contraste e respectivo desvio padrão

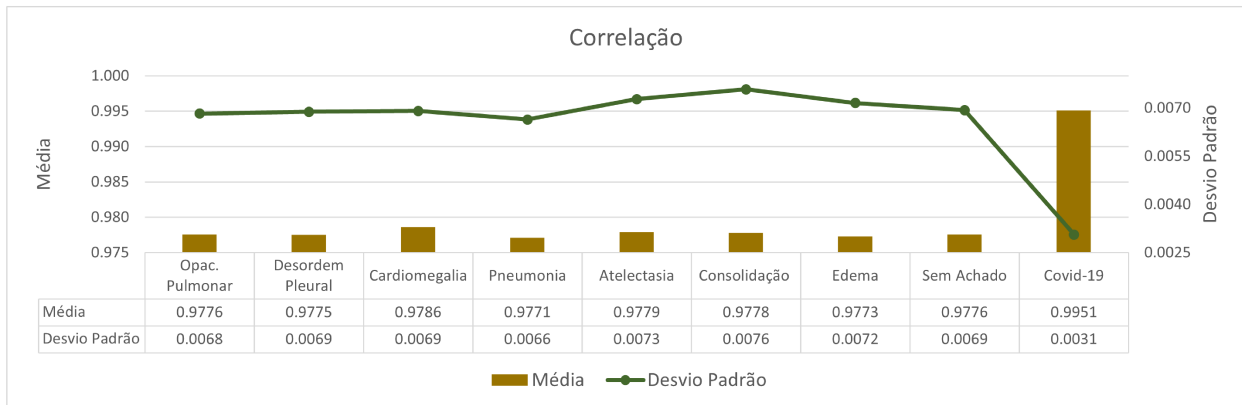


Fonte: Elaborado pelo autor (2022).

Como pode ser definido pelo gráfico anterior os exames radiográficos de pacientes com COVID-19 possuem discrepantemente menor contraste local que todas as outras anomalias pulmonares e imagens sem achado radiológico.

A Correlação é indicada a seguir (Figura 9):

Figura 9 – Gráfico com valores de Correlação e respectivo desvio padrão

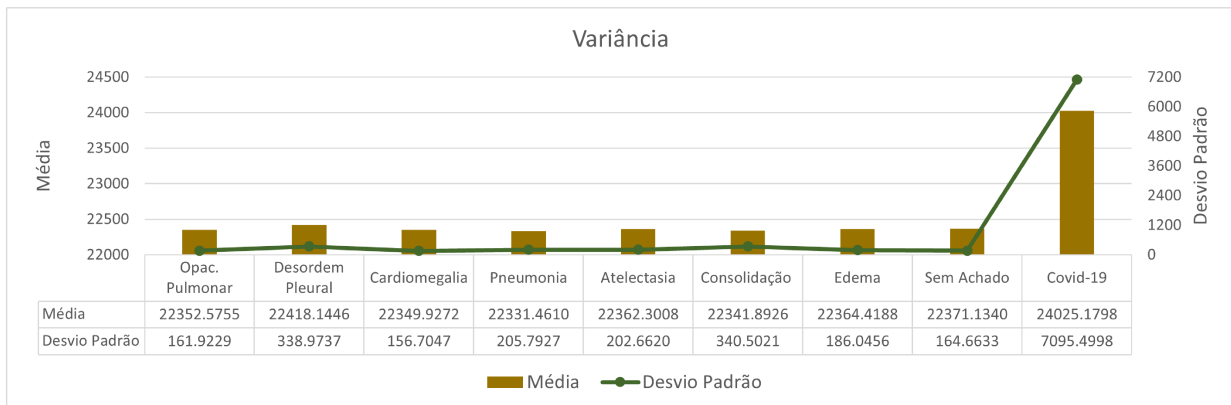


Fonte: Elaborado pelo autor (2022).

Considerando o conceito de correlação é possível interpretar que, assim, como a homogeneidade, os exames com SARS-CoV-2 possuem maior linearidade local apresentada, diferente das outras anomalias analisadas.

A avaliação dos resultados de Variância calculados estão apresentados na Figura 10:

Figura 10 – Gráfico com valores de Variância e respectivo desvio padrão

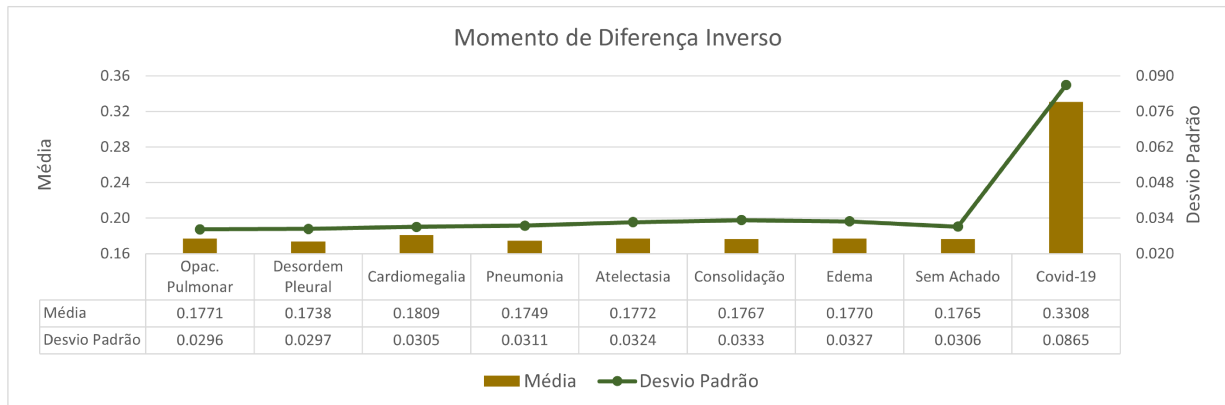


Fonte: Elaborado pelo autor (2022).

Com essa variável é possível interpretar que as imagens de exames radiográficos de pacientes com COVID-19 possuem maior dispersão de tons de pixel que outras doenças ou imagens sem achado radiológico.

Para avaliação do Momento de Diferença Inverso das características propostas por Haralick, tem-se (Figura 11):

Figura 11 – Gráfico com valores de Momento de Diferença Inverso e respectivo desvio padrão

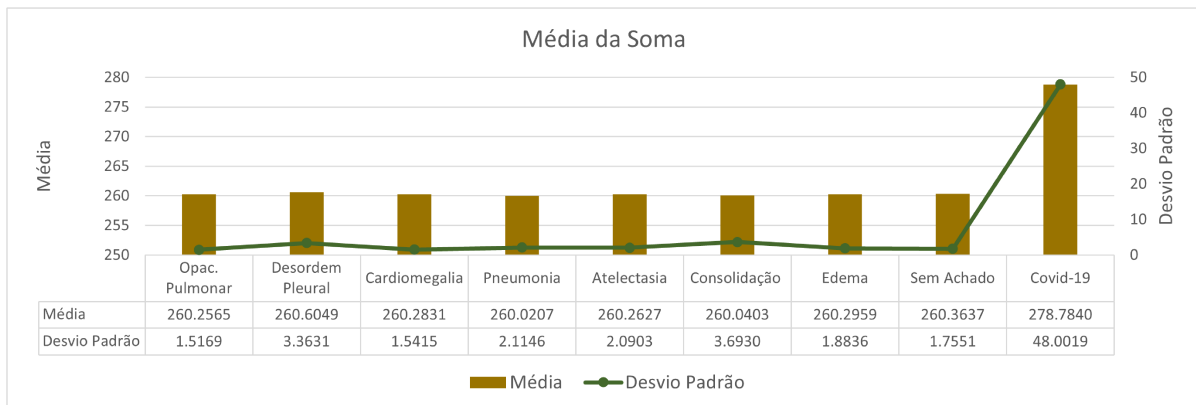


Fonte: Elaborado pelo autor (2022).

Esse atributo de textura indica nas imagens a homogeneidade da matriz de pixels e demonstra que os exames de COVID-19 chegam a possuir, em média, o dobro quando comparados com outras doenças pulmonares e imagens radiográficas sem achado.

A seguir está disposta a Soma da Média (Figura 12):

Figura 12 – Gráfico com valores de Média da Soma e respectivo desvio padrão

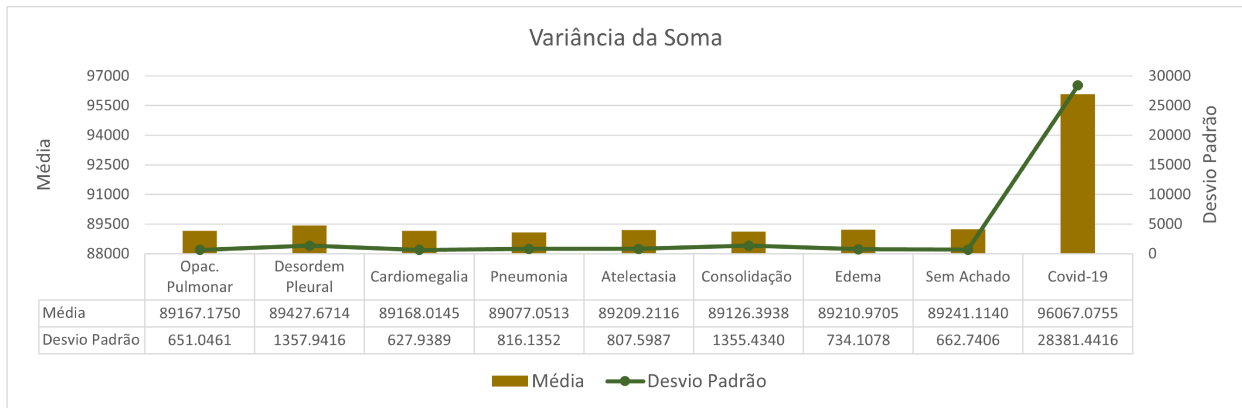


Fonte: Elaborado pelo autor (2022).

As imagens de radiografias com COVID-19 também apresentam para essa variável valor maior que para os outros distúrbios analisados, demonstrando áreas de maior densidade no exame.

A média dos valores da Variância da Soma é representada no gráfico abaixo (Figura 13):

Figura 13 – Gráfico com valores de Variância da Soma e respectivo desvio padrão

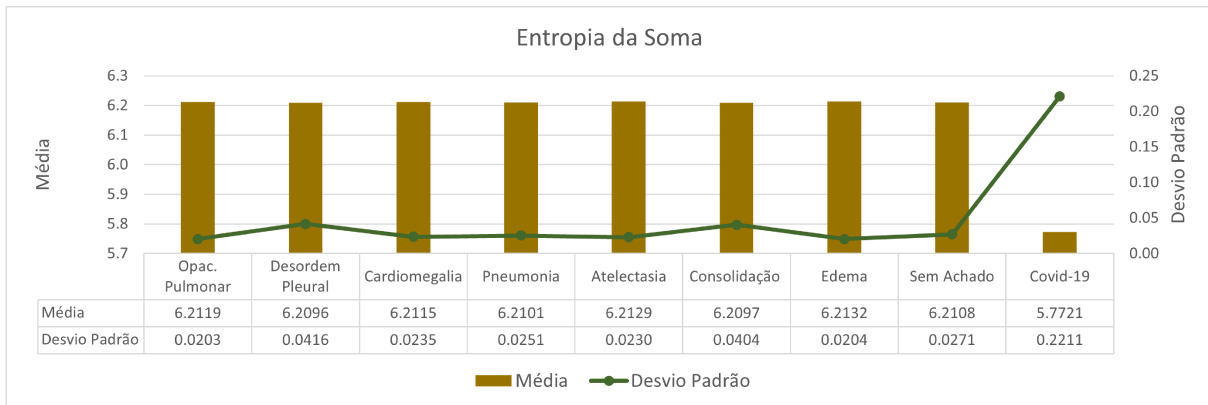


Fonte: Elaborado pelo autor (2022).

A Variância da Soma, assim, como a Variância apresentam para os exames com COVID-19 maior valor em relação às outras pastas analisadas.

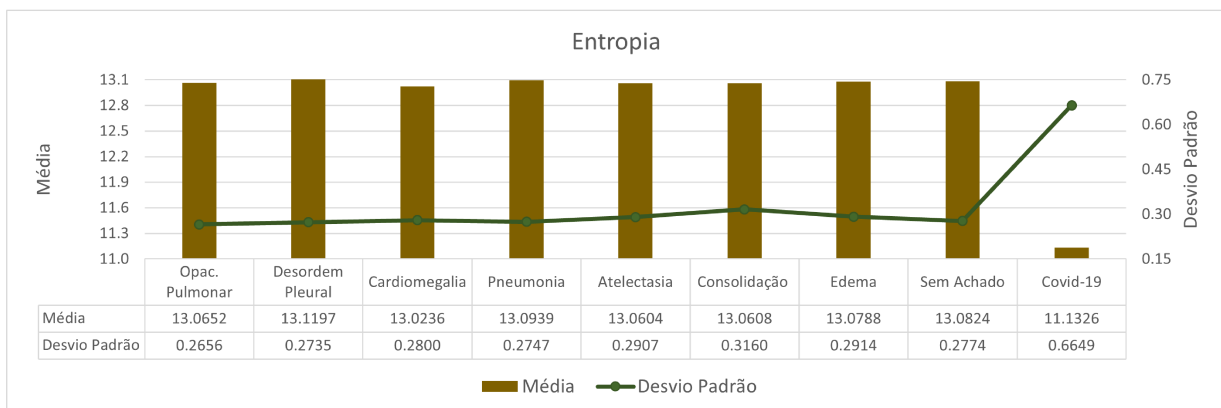
A seguir os resultados dos descritores Entropia da Soma (Figura 14) e Entropia (Figura 15):

Figura 14 – Gráfico com valores de Entropia da Soma e respectivo desvio padrão



Fonte: Elaborado pelo autor (2022).

Figura 15 – Gráfico com valores de Entropia e respectivo desvio padrão

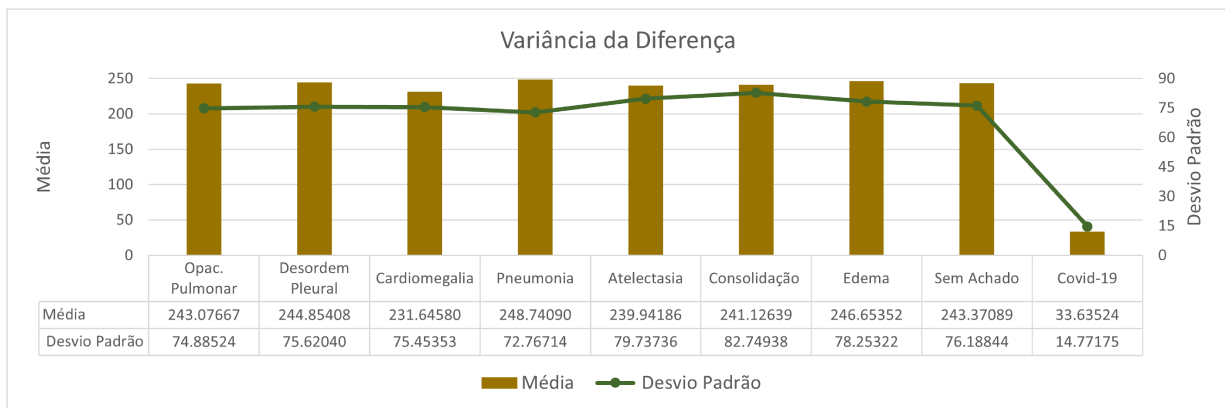


Fonte: Elaborado pelo autor (2022).

As variáveis apresentadas nas Figuras 14 e 15 estão intimamente relacionadas com a aleatoriedade dos pixels de uma imagem, como pode ser visualizado, as outras doenças pulmonares e a exames sem achado radiográfico possuem maior valor para ambas, demonstrando que imagens médicas de radiografias de pacientes com COVID-19 possuem menor desordem de textura.

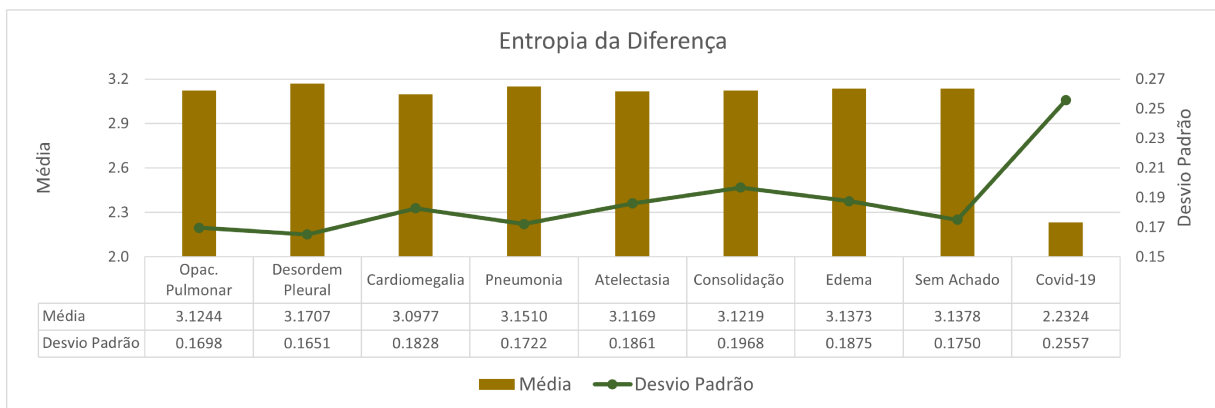
Os atributos de textura da Variância da Diferença (Figura 16) e Entropia da Diferença (Figura 17) estão dispostos abaixo:

Figura 16 – Gráfico com valores de Variância da Diferença e respectivo desvio padrão



Fonte: Elaborado pelo autor (2022).

Figura 17 – Gráfico com valores de Entropia da Diferença e respectivo desvio padrão



Fonte: Elaborado pelo autor (2022).

As variáveis representadas nos dois gráficos anteriores descrevem respectivamente a variação local de tons de cinza da imagem e a desordem dos pixels, para as imagens de pacientes diagnosticados com COVID-19 ambas possuem valores menores para essas características de textura que exames radiográficos das outras doenças pulmonares ou sem achado.

Após avaliação dos resultados apresentados anteriormente, pode-se perceber que os exames nos quais há presença do SARS-CoV-2 possuem características estatísticas e de textura consideravelmente diferentes das outras anomalias pulmonares.

É possível observar um padrão maior de uniformidade e homogeneidade locais em exames de radiografia de tórax de pacientes diagnosticados com COVID-19, o que pode ser identificado ao

analisar os maiores valores de Segundo Momento Angular, Correlação e Momento de Diferença Inverso, simultaneamente com os menores valores de Contraste, Entropia, por exemplo.

As variáveis citadas acima podem ser utilizadas para o auxílio do médico quando há incerteza do prognóstico durante a visualização do exame, pois apresentam discrepância considerável de valores em relação a outras doenças que acometem o sistema respiratório. Em consonância pode-se utilizar os valores das variáveis estatísticas média e desvio padrão dos pixels da imagem como alvo de análise por também se mostrarem extremamente distintas nesses casos.

4 CONCLUSÕES

O desenvolvimento de técnicas para auxílio ao diagnóstico de imagens médicas possui grande impacto no âmbito da saúde hoje. Um prognóstico assertivo e rápido é extremamente importante principalmente quando se trata da identificação de doenças infectocontagiosas e de rápido acometimento à saúde do paciente, como a COVID-19. A extração de atributos de imagens vem ganhando espaço e sendo bastante utilizada por utilizar baixo processamento computacional.

A partir dos resultados encontrados neste trabalho é possível concluir que a extração de características pode auxiliar de modo efetivo para determinação da presença da COVID-19 em detrimento de outras doenças pulmonares, já que sua distribuição de tons de pixels se mostrou notavelmente desigual em comparação com as demais. A análise simultânea da média e desvio padrão dos pixels da imagem radiológica do paciente com suspeita de COVID-19 e dos atributos de textura de Haralick é de grande valia para conclusão mais clara do prognóstico da doença.

Ainda é necessário maior aprofundamento do estudo dessas características e alinhamento desse desfecho com a visualização da imagem gerada pelo exame de radiografia de tórax, trabalho que deve ser realizado em uma equipe multidisciplinar, contendo a presença de radiologistas, por exemplo. Além disso, é importante investigar outros bancos de imagens, preferencialmente de locais diferentes, para corroborar com os resultados obtidos na detecção da patologia e, assim, possivelmente, adotar a análise no prognóstico da COVID-19.

REFERÊNCIAS

- APOSTOLOPOULOS, I. D.; MPESIANA, T. A. Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. **Physical and engineering sciences in medicine**, Springer, v. 43, n. 2, p. 635–640, 2020.
- BRESTEL, C. *et al.* Radbot-cxr: Classification of four clinical finding categories in chest x-ray using deep learning. 2018.
- Drauzio Varella. 2011. Disponível em: <<https://bvsmms.saude.gov.br/pneumonia-5/#:~:text=Pneumonia%20%C3%A9%20uma%20infec%C3%A7%C3%A3o%20que,entre%20um%20alv%C3%A9olo%20e%20outro>>. Acesso em: 01 Ago. de 2022.
- FILLIBEN, J. J.; HECKERT, A. Nist/sematech e-handbook of statistical methods. **NIST, Gaithersburg**, 2002.
- GOMES, J. C. *et al.* Ikonos: an intelligent tool to support diagnosis of covid-19 by texture analysis of x-ray images. **Research on Biomedical Engineering**, Springer, p. 1–14, 2020.
- HARALICK, R. M. Statistical and structural approaches to texture. **Proceedings of the IEEE**, IEEE, v. 67, n. 5, p. 786–804, 1979.
- HASSANTABAR, S.; AHMADI, M.; SHARIFI, A. Diagnosis and detection of infected tissue of covid-19 patients based on lung x-ray image using convolutional neural network approaches. **Chaos, Solitons & Fractals**, Elsevier, v. 140, p. 110170, 2020.
- IRVIN, J. *et al.* Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In: **Proceedings of the AAAI conference on artificial intelligence**. [S.l.: s.n.], 2019. v. 33, n. 01, p. 590–597.
- ISLAM, N. *et al.* Thoracic imaging tests for the diagnosis of covid-19. **Cochrane Database of Systematic Reviews**, John Wiley & Sons, Ltd, n. 3, 2021.
- KIM, H. W. *et al.* The role of initial chest x-ray in triaging patients with suspected covid-19 during the pandemic. **Emergency radiology**, Springer, v. 27, n. 6, p. 617–621, 2020.
- KNIPE, H.; JONES, J. **Pulmonary opacification**. Radiopaedia.org, 2011. Disponível em: <<https://doi.org/10.53347/rid-14499>>. Acesso em: 01 Ago. de 2022.
- LÖFSTEDT, T. *et al.* Gray-level invariant haralick texture features. **PloS one**, Public Library of Science San Francisco, CA USA, v. 14, n. 2, p. e0212110, 2019.
- MATERKA, A.; STRZELECKI, M. *et al.* Texture analysis methods—a review. **Technical university of lodz, institute of electronics, COST B11 report, Brussels**, v. 10, n. 1.97, p. 4968, 1998.
- MDLI, V. *et al.* Bimcv covid-19+: a large annotated dataset of rx and ct images from covid-19 patients. 2020.
- MURPHY, A.; HACKING, C. **Lobar consolidation**. Radiopaedia.org, 2019. Disponível em: <<https://doi.org/10.53347/rid-66160>>. Acesso em: 01 Ago. de 2022.

National Heart, Lung, and Blood Institute (NHLBI). **PLEURAL DISORDERS**. 2022. Disponível em: <<https://www.nhlbi.nih.gov/health/pleural-disorders#:~:text=Pleural%20disorders%20are%20conditions%20that,airways%2C%20pleura%2C%20and%20diaphragm>>. Acesso em: 01 Ago. de 2022.

PERONI, D.; BONER, A. Atelectasis: mechanisms, diagnosis and management. **Paediatric respiratory reviews**, Elsevier, v. 1, n. 3, p. 274–278, 2000.

RODRIGUES, C. F. d. S.; LIMA, F. J. C. d.; BARBOSA, F. T. Importância do uso adequado da estatística básica nas pesquisas clínicas. **Revista Brasileira de Anestesiologia**, SciELO Brasil, v. 67, p. 619–625, 2017.

RONDON, E. O. O.; MURAKAMI, A. E.; SAKAGUTI, E. S. Modelagem computacional para produção e pesquisa em avicultura. **Brazilian Journal of Poultry Science**, SciELO Brasil, v. 4, p. 199–207, 2002.

SANTOS, T. *et al.* Mri texture analysis reveals bulbar abnormalities in friedreich ataxia. **American Journal of Neuroradiology**, Am Soc Neuroradiology, v. 36, n. 12, p. 2214–2218, 2015.

SCHWARTZ, W. R.; PEDRINI, H. Método para classificação de imagens baseada em matrizes de coocorrência utilizando características de textura. **III Colóquio Brasileiro de Ciências Geodésicas, CuritibaPR, Brasil**, v. 1, 2003.

STAUB, N. C. Pulmonary edema. **Physiological reviews**, v. 54, n. 3, p. 678–811, 1974.

The MathWorks, Inc. **Using a Gray-Level Co-Occurrence Matrix (GLCM)**. 2005. Disponível em: <<http://matlab.izmiran.ru/help/toolbox/images/enhanc15.html#32796>>. Acesso em: 10 Jul. de 2022.

VARELA-SANTOS, S.; MELIN, P. A new approach for classifying coronavirus covid-19 based on its manifestation on chest x-rays using texture features and neural networks. **Information sciences**, Elsevier, v. 545, p. 403–414, 2021.

WJ;, A. H. **Cardiomegaly**. U.S. National Library of Medicine, 2021. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/31194436/>>.

WONG, H. Y. F. *et al.* Frequency and distribution of chest radiographic findings in patients positive for covid-19. **Radiology**, Radiological Society of North America, v. 296, n. 2, p. E72–E78, 2020.

WOODS, R. E.; GONZALEZ, R. C. Digital image processing third edition. 2010.

World Health Organization (WHO). **WHO Coronavirus (COVID-19) Dashboard**. 2022. Disponível em: <<https://covid19.who.int/>>. Acesso em: 14 Jul. de 2022.