

# Algoritmos Não-Supervisionados para Suporte à Tomada de Decisão e Diagnóstico de Covid-19 usando Dados de Tomografia Computadorizada

Taís Aparecida Alvarenga  
*Programa de Pós-graduação em  
Engenharia de Sistemas e Automação  
Universidade Federal de Lavras*  
MG, Brasil  
tais.alvarenga@estudante.ufla.br

Luis Otávio Santos  
*Programa de Pós-graduação em  
Engenharia de Sistemas e Automação  
Universidade Federal de Lavras*  
MG, Brasil  
luis.santos3@estudante.ufla.br

Demóstenes Zegarra Rodríguez  
*Departamento de  
Ciência da Computação  
Universidade Federal de Lavras*  
MG, Brasil  
demostenes.zegarra@ufla.br

Danton Diego Ferreira  
*Departamento de Automática  
Universidade Federal de Lavras*  
MG, Brasil  
danton@ufla.br

Bruno H. G. Barbosa  
*Departamento de Automática  
Universidade Federal de Lavras*  
MG, Brasil  
brunohb@ufla.br

José Manoel de Seixas  
*Laboratório de Processamento de Sinais  
Universidade Federal do Rio de Janeiro*  
RJ, Brasil  
seixas@lps.ufrj.br

**Resumo:** —Os algoritmos de aprendizagem profundo, aliados a técnicas da teoria de conjuntos *Fuzzy*, tem sido empregados em análises de imagens. Deste modo, as aplicações de aprendizado profundo em imagens médicas pulmonares tem obtido avanços para detecção de doenças, levando a ensaios clínicos promissores no que diz respeito à triagem e diagnóstico de pacientes com o COVID-19. Assim, imagens de Tomografia Computadorizada (TC) tem sido usadas por especialistas como ferramentas para diagnóstico da COVID-19. No entanto, devido ao número elevado de pacientes e baixa quantidade de profissionais especializados, torna-se necessário um sistema de detecção por imagens para auxílio aos profissionais da saúde. Neste cenário, a inteligência computacional, mais especificamente os modelos de aprendizado profundo do tipo *Convolutional Neural Network* (CNN) e métodos de classificação não-supervisionados, entre estes em destaque o método *K-Means* e *Fuzzy C-Means*, foram utilizados para a classificação de imagens de Covid-19. Para esta finalidade, foi utilizada uma base de dados disponibilizada pela Universidade de São Paulo, contendo 4173 imagens de tomografia computadorizada. Os resultados demonstraram a importância da seleção de características através de redes convolucionais, bem como a melhoria do desempenho da classificação, advindo do uso de aprendizado profundo e seu estado da arte na área de visão computacional. Na aprendizagem não-supervisionada observou-se poucas diferenças para a clusterização utilizando os métodos *k-Means* e *Fuzzy C-Means*.

**Index Terms**—Covid-19, Aprendizado não Supervisionado, *K-Means* e *Fuzzy C-Means*.

Os autores gostariam de agradecer à CAPES, CNPq e FAPEMIG pelo suporte financeiro.

## I. INTRODUÇÃO

A COVID-19 é uma doença infecciosa do agente etiológico, a Síndrome Respiratória Aguda Grave, Coronavírus 2 (SARS-CoV-2) [1], que tem apresentado alta taxa de mortalidade se comparadas a outras influências [2]. Até o momento (24/08/2021), mais de 4,44 milhões de mortes foram registradas, sobrecarregando sistemas de saúde do mundo todo superando morbidades, tais como, SARS-CoV e MERS-CoV [3].

Em meados de 2019, com o surgimento da COVID-19, na cidade de Wuhan (China) [4] o uso de diferentes técnicas passou a ser uma ferramenta essencial no apoio à tomada de decisões de médicos e/ou especialistas. Diante da fácil disseminação e contágio, medidas como o isolamento social, têm sido impostas em muitos países, tendo como consequências quedas na economia, sobretudo doenças psicológicas e principalmente várias mortes, o que torna ainda mais significativo o uso de diversos recursos no controle desta doença.

Atualmente o diagnóstico da doença é realizado através de testes de RT-PCR, histórico clínico do paciente e imagens de tomografia computadorizada (TC). Assim, as imagens de TC desempenham um importante papel no diagnóstico da COVID-19, porém, este método apresenta certas limitações como a falta de especificidade entre as lesões que podem ter sido ocasionadas por outras doenças, bem como a falta de profissionais para a avaliação das imagens. Neste contexto, o uso de ferramentas computacionais pode auxiliar na correção desse viés técnico e falta de mão de obra qualificada [5].

Neste cenário, com um número crescente de dados (big data) tornou-se possível implementar e aprimorar técnicas

de *Deep Learning*, tais como as redes neurais convolutivas: *ResNet* [6], Rede Eficiente [7], entre outros modelos que tem mostrado resultados expressivos, principalmente para a seleção de características, uma etapa de crucial importância para o bom desempenho da etapa de classificação. Esses modelos vêm apresentando bons resultados em diversas áreas [8], bem como na tomografia computadorizada para a triagem e diagnóstico da COVID-19 [9].

O uso de técnicas computacionais na área médica tem desempenhado um papel estratégico no auxílio a diversas doenças, entre estes, observa-se resultados promissores em análises de imagens médicas [9], sendo atrativo o uso de algoritmos de aprendizagem de máquina em questões relacionadas ao COVID-19.

De modo semelhante, a utilização de modelos baseados em lógica *fuzzy*, teoria proposta inicialmente em [10], que lida com conjuntos difusos para modelar as incertezas, tem apresentado resultados importantes em diversas aplicações, entre elas a área médica [11]. Em relação à utilização de imagens de TC, modeladas pela teoria *Fuzzy* e algoritmos de *Deep Learning*, alguns estudos tem demonstrado a eficácia quando realizados para classificação de imagens de COVID-19 [12].

Neste contexto, o aprendizado profundo (do inglês *Deep Learning*) se destaca pela detecção de atributos e padrões, com destaque para o campo de visão computacional, em que os modelos são capazes de extrair as informações visuais (contornos e características) que mostram-se mais propensas a fornecer uma resposta adequada ao problema. Uma vez que, para o reconhecimento de padrões em imagens, tem-se adotado técnicas de transferência de aprendizado em redes neurais convolutivas pré-treinadas, sendo que estas se apresentam atrativas e para problemas com poucos dados e limitações de estrutura para treinamento do algoritmo.

Após extraídas, as características podem ser combinadas por meio de *ensembles*, aumentando a precisão do classificador conforme pode ser inferido por meio da análise da decomposição polarização-variância do erro do estimador [13]. Para a etapa final, ou mais especificadamente para a classificação das imagens em diferentes classes, optou-se pelo uso de modelos Fuzzy, uma vez que estes trabalham em espaços de pertinência e lidam bem com as incertezas dos dados.

Nesse sentido, o objetivo deste trabalho é propor um método de classificação de imagens de TC de COVID-19 utilizando técnicas de *Deep Learning* e teoria *Fuzzy*. Sendo assim, três modelos de redes convolucionais pré-treinados são utilizados para a extração de características e as características são por fim concatenadas para que as mesmas sejam classificadas a partir de técnicas *fuzzy* não supervisionadas.

## II. DESCRIÇÃO DOS ALGORITMOS UTILIZADOS

### A. Classificação baseada em clusters - ideia geral

Dado um conjunto de observações  $x(i) \in \mathbb{R}^{k \times 1}$ ,  $i = 1, 2, 3, \dots, N$ , deseja-se separar os dados utilizando um critério de similaridade, tais como a distância de Mahalanobis, Euclideana, Chebyshev. Ao final, espera-se que cada *cluster* represente uma estrutura natural existente nos dados. Sendo assim, a classificação está associada à ideia de tarefa não-supervisionada. Logo, o que se busca é inferir uma função que descreva de maneira suficiente a estrutura de dados não-rotulados.

### B. Método de clusterização: *k*-Means

O algoritmo *k*-Means descrito por [14], [15] tem como objetivo particionar decididamente um conjunto de dados  $X$  em uma determinada quantidade de *clusters*  $k$ . O algoritmo inicia-se com um conjunto  $k$  com pontos centrais em  $(\mu)$ , chamados protótipos, que são pontos no interior do espaço  $N$ -dimensional onde o agrupamento se concentra, eles podem ser os centros dos grupos, por exemplo. Durante cada etapa de atualização, cada uma das observações em  $x$  é atribuída ao seu ponto central mais próximo.

$$S_i^{(t)} = \{x_p : \|x_p - \mu_i^{(t)}\|^2 \leq \|x_p - \mu_j^{(t)}\|^2 \forall j, 1 \leq j \leq k\}. \quad (1)$$

Posteriormente, os pontos centrais são reposicionados calculando a média das observações atribuídas aos respectivos pontos centrais.

$$\mu_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j. \quad (2)$$

O processo de atualização ocorre novamente até que todas as observações permaneçam nos pontos centrais atribuídos e, portanto, os pontos centrais não seriam mais atualizados. Isso significa que o algoritmo *k*-Means tenta otimizar a função objetivo. Como há apenas um número finito de atribuições possíveis para a quantidade de centroides e observações disponíveis e cada iteração deve resultar em uma solução melhor, o algoritmo sempre termina em um mínimo local.

$$J = \sum_{n=1}^N \sum_{k=1}^K r_{nk} \|x_n - \mu_k\|^2, \quad (3)$$

$$\text{com } r_{nk} = \begin{cases} 1, & x_n \in S_k \\ 0, & \text{caso contrário.} \end{cases}$$

O principal problema do *k*-Means é sua dependência dos centroides escolhidos inicialmente. Os centroides podem acabar dividindo pontos de dados comuns, enquanto outros pontos separados são agrupados se alguns dos centroides forem mais atraídos por *outliers*. A abordagem mais comum é realizar vários agrupamentos com diferentes

posições iniciais. Basicamente o algoritmo *k-Means* funciona atribuindo os dados aos centroides, assim [15] propôs um algoritmo baseado o *k-Means* onde é possível tratar as incertezas utilizando “graus de pertinência” surgindo o algoritmo *Fuzzy C-Means*. Na literatura, encontra-se outros algoritmos de classificação *Fuzzy*, tal como, Algoritmo *Fuzzy C-Medoids* [16], porém, como constata [17] o *Fuzzy C-Means* é um dos algoritmos amplamente utilizado para a classificação *Fuzzy*.

### C. Método de clusterização: Fuzzy C-Means (FCM)

De modo geral, o FCM tem como objetivo particionar os dados de tal forma que as amostras intra-grupos sejam mais similares entre si, e amostras inter-grupos sejam menos similares, gerando um grau de pertinência associado a um conjunto de dados  $X$  em uma determinada quantidade de *clusters*. Esse processo ocorre via função objetivo, dada por:

$$J = \sum_{n=1}^N \sum_{k=1}^K u_{n,k}^m d(x_n, c_k), \quad (4)$$

onde  $u_{n,k}^m$  é o grau de pertinência e  $d(x_n, c_k)$  a distância entre os ponto  $x_n$  e o centro  $c_k$  definido por:

$$c_k = \frac{\sum_{n=1}^k u_{n,k}^m \cdot x_i}{\sum_u^{m_{n,k}} u_{nk}^m}, \quad (5)$$

na qual  $c$  é o número de centros, e  $U$  é a matriz de pertinência dada por:

$$U = \begin{bmatrix} u_{11} & \cdots & u_{1n} & \cdots & u_{1N} \\ \vdots & \ddots & \vdots & & \vdots \\ u_{k1} & \cdots & u_{k1} & \cdots & u_{kc} \\ \vdots & & \vdots & \ddots & \vdots \\ u_{k1} & \cdots & u_{k1} & \cdots & u_{kc} \end{bmatrix}. \quad (6)$$

sujeito as restrições:  $\sum_{i=1}^N u_{nk} = 1 \forall n$ ; e  $0 < \sum_{i=1}^K u_{nk} < 1 \forall k$ . Os elementos da matriz de partição são dados por:

$$u_{kn} = \frac{1}{\sum_{j=1}^c \frac{D_{kn}}{D_{kj}}}. \quad (7)$$

### D. Modelo de Teste: Critério de Xie-Beni e Acurácia

A fim de avaliar qual a quantidade de agrupamentos separam melhor o conjunto de dados, foi realizada uma análise variando-se a quantidade de grupos, e obtendo as métricas Xie-Beni e a acurácia final com base na matriz de pertinências resultante das iterações, o critério de Xie-Beni é dado por:

$$XB = \frac{\sum_{n=1}^N \sum_{k=1}^K u_{nk}^q \|x_n - v_k\|^2}{n \cdot \min_{n \neq j} (\|v_n - v_j\|^2)}. \quad (8)$$

O numerador do critério de Xie-Beni indica a variância intra *cluster*, enquanto que o denominador indica a separação entre os centros. Então, a melhor partição irá minimizar o valor de  $XB$ , apontando para o número de grupos

que maximiza a separação dos dados entre a quantidade de grupos de entrada. A acurácia foi determinada pela matriz de pertinências gerada pelos algoritmos, tanto no caso do *C-Means* quanto no *K-Means*, foi tomado o maior valor entre as pertinências geradas.

## III. METODOLOGIA

A metodologia consiste na concatenação de duas áreas: *Deep Learning* e *Fuzzy*. O modelo proposto denominado Deep FCK, consiste na extração das características por meio dos modelos Covid-Net CT 1L [19], Covid-Net CT 1L [19] e Covid-Net CT 1L [18], todos foram pré-treinados na base de dados ImageNet e, posteriormente, treinados na base Covidx CT com imagens de tomografia. A Figura 1, retrata o diagrama das redes propostas por [19] e [18] com a proposta do modelo Deep FCK.

Neste projeto, a base de imagens de tomografia utilizada está documentada em [20], cujas imagens estão em dimensões diversas, e como os modelos foram treinados com imagens com dimensão de  $512 \times 512$ , as imagens tiveram que ser redimensionadas para este valor, utilizando interpolação cúbica, que se caracteriza basicamente, em se fazer uso de pixels conhecidos com o objetivo de estimar valores de pixels desconhecidos. Assim algumas abordagens, como mudança de escala (redimensionamento) ou rotações podem fazer uso dessa técnica. Por vez, pode-se introduzir incerteza nos dados. Em seguida as imagens percorrem todas as camadas realizando o processo de *feedforward* no modelo até a última camada antes de percorrer a parte densa da rede.

Após realizar o processo para todas as imagens e guardar os resultados, as saídas dos três modelos foram concatenadas em um único vetor, gerando um total de 764 atributos para cada imagem. Em seguida os dados são selecionados (no caso do *K-Means*) utilizando o critério da razão discriminante de *Fisher*. Finalmente os dados são agrupados por um dos algoritmos, *C-Means* ou *K-Means*. O diagrama da figura 2 resume o fluxo dos dados desde a entrada até a obtenção dos resultados.

### A. Base de dados

O conjunto de dados é composto por tomografias computadorizadas em formato PNG, que são divididos em: 758 tomografias computadorizadas para pacientes saudáveis (15 na média tomografias por paciente). 2168 tomografias para pacientes infectados por SARS-CoV-2 (na média de 27 tomografias por paciente) e 1247 tomografias para pacientes com outras direções pulmonares (na média de 16 tomografias por paciente). Totalizando, 4173 tomografias para 210 pacientes [20].

Como a base possui um desbalanceamento entre a quantidade de amostras de cada classe, ao separar os dados em treino e teste, a quantidade de amostras de cada classe também ficaram desbalanceadas, tal como indica a Tabela I.

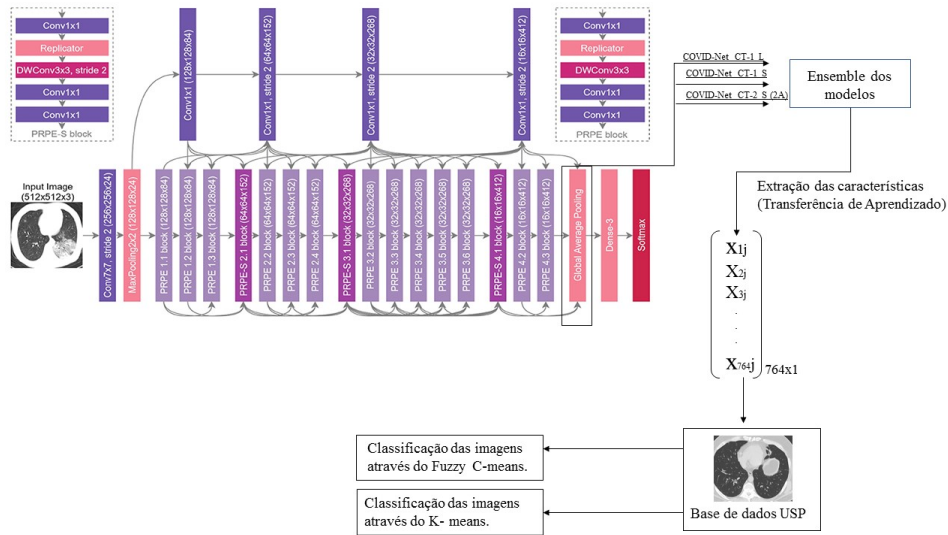


Figura 1. Deep FCK. Adaptado de [18].

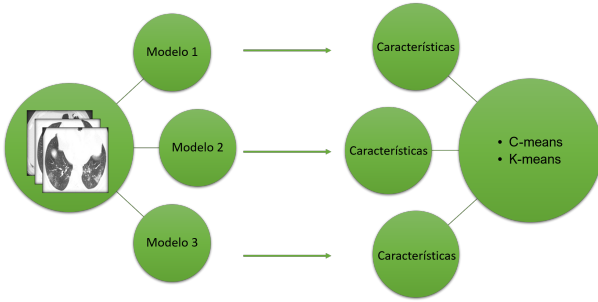


Figura 2. Diagrama Geral do método.



Figura 4. Tomografia Computadorizada de um paciente saudável

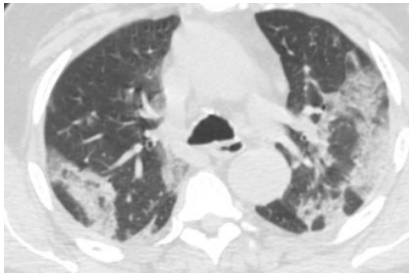


Figura 3. Tomografia Computadorizada de um paciente com COVID-19



Figura 5. Tomografia Computadorizada de um paciente com outras doenças pulmonares

Tabela I  
CONJUNTO DE DADOS UTILIZADOS

	Treino (70%)	Teste (20%)	
Covid	1517	650	
Saudáveis	530	227	
Outros	873	374	
Total	2920	1251	4171

### B. Seleção de Características

A seleção de características pode ser uma etapa fundamental para melhorar o desempenho do classificador [21]. Selecionar as características que demonstram a função densidade de probabilidade (pdf) dos dados, faz com que os padrões para cada classe possam ser identificados. Outra técnica complementar a seleção de características, visando o bom desempenho da classificação se trata de modelos especialistas, que tem o objetivo de estudar a característica para cada classe separadamente.

Assim, a fim de selecionar as melhores características extraídas pelos modelos convolutivos, foi utilizado o critério razão discriminante de *Fisher*, onde as características selecionadas apresentam maior valor desse critério. Assim, a para o cálculo dessa razão, inicia-se a partir do cálculo da média  $\mu$  e da variância  $\sigma^2$  entre os atributos  $x$  das classes tomadas 2 a 2, logo, o cálculo da razão discriminante de *Fisher* pode ser dado por:

$$J_{FDR} = \frac{(m_1(x_N) - m_2(x_N))^2}{(\sigma_1(x_N))^2 + (\sigma_2(x_N))^2} \quad (9)$$

O termo no numerador da equação 9 aplicada sobre os atributos de duas classes distintas, indica quanto estão distantes suas distribuições, enquanto o denominador indica o quão espalhadas estão estas distribuições. Assumindo uma distribuição normal, quanto maior for esta razão, maior será a capacidade de discriminação entre as classes analisadas, porém esta métrica não elimina redundâncias, [22]. De forma análoga ao discriminante de Fisher, outras técnicas para a seleção de características obtidas pelo processo de transferência de aprendizado podem ser adotadas, tais como, Algoritmos Genéticos [23] e *Autoencoders* [24].

#### IV. RESULTADOS E DISCUSSÃO

##### A. Limpeza dos dados

A extração dos atributos da imagem foi realizada tomando como base a camada anterior à camada densa da rede, totalizando um conjunto de 764 atributos, sendo 176 de dois modelos com menor quantidade de parâmetros, e 412 de outro modelo com maior quantidade de parâmetros. Uma pequena parte destes atributos eram nulos e, ao aplicar a razão de *Fisher*, os resultados foram interpretados como não numéricos, por se tratar de uma indeterminação matemática. Para solucionar este problema, os atributos foram removidos, restando 747 atributos por amostra.

##### B. Análise Baseada em Cenários

Ao analisar as imagens pelos modelos pré-treinados, as saídas foram observadas em relação à classificação original dos modelos e foi notado uma dificuldade por parte dos mesmos em distinguir as imagens das classes referentes à base em [20], exceto pelas imagens de exames de pessoas saudáveis, neste conjunto a classificação foi condizente com o a maioria das imagens. Uma observação interessante sobre este problema pode ser justificado pela qualidade das imagens, as quais não apresentam tantos problemas quanto as outras duas classes. Técnicas para o equilíbrio entre as classes podem ser adotadas, tais como a sobre-amostragem SMOTE, as quais são adotadas como perspectiva futura.

Buscando analisar esta divergência entre o modelo e a classificação nas bases de dados, foram criados três cenários diferentes para análise mais detalhada. Confrontando as classes COVID e Saudáveis, COVID e Outras Doenças, COVID e Todos.

##### C. Fuzzy C-Means

O primeiro algoritmo utilizado para análise foi o *Fuzzy C-Means*. Para verificar o melhor cenário através do critério de Xie-Beni, vários cenários foram analisados, assim variou-se a quantidade de *clusters* e manteve-se a quantidade total de atributos por amostra. Os resultados obtidos estão dispostos nas tabelas II, III:

Tabela II  
RESULTADOS DO FUZZY C-MEANS

Clusters	Covid X Saudável			Covid X Outros		
	XB(1)	XB(2)	acurácia(%)	XB	XB	acurácia(%)
2	1.88	2.18	74.2	1.60	2.19	58.4
4	1.39	1.39	66.0	1.12	1.79	58.7
6	0.84	1.56	56.1	0.86	1.54	56.5
8	0.70	1.22	63.1	0.67	1.23	51.8
12	0.46	0.81	68.2	0.39	0.67	65.7

XB(1): Xie-Beni para os dados de treinamento

XB(2): Xie-Beni para os dados de teste

Tabela III  
RESULTADOS DO FUZZY C-MEANS

Clusters	COVID X Outros X Saudável		
	XB(Treino)	XB(Teste)	acurácia(%)
3	1.57	2.69	47.1
6	0.80	1.31	44.5
9	0.61	0.90	52.0
12	0.44	0.78	58.0

Conforme observado nas Tabelas II, III, o critério de *Xie-Beni* diminui progressivamente com o aumento do número de clusters, de forma geral, a acurácia aumenta com o decréscimo do critério de Xie-Beni, resultados similares também podem ser encontrados em [14]. Assim, de acordo com os resultados obtidos, a acurácia foi melhor no primeiro cenário analisado (COVID x Saudáveis). Enquanto o critério *Xie-Beni* aponta que 12 agrupamentos separam melhor os dados.

Vê-se que por tais resultados, que a extração das características das imagens de cada classe estão confundindo entre si, e criando "conjuntos" que não se eram esperados, uma vez que, o número ideal de clusters esperado eram 3, ou seja, um para cada uma das classes (COVID, Saudável e Outras). Este fato é verificado principalmente, na III, onde para o melhor valor de Xie-Beni a acurácia não supera os 60%.

##### D. K-Means

O segundo algoritmo utilizado para análise foi o *K-Means*, variando a quantidade de *clusters* fornecidos e utilizando 222 atributos selecionados pela razão discriminante de *Fisher*. Os resultados obtidos estão dispostos na tabela IV e V:

Tabela IV  
RESULTADOS DO K-MEANS

Clusters	Covid X Saudável			Covid X Outros		
	XB(1)	XB(2)	Acurácia(%)	XB	XB	Acurácia(%)
2	0.25	0.16	73.4	0.42	0.25	68.9
4	0.62	0.41	43.0	1.13	0.78	27.3
6	1.07	0.74	30.7	2.99	2.15	38.7
8	1.44	1.0	11.2	3.69	2.65	17.5
12	2.40	1.74	16.0	5.95	4.28	19.7

XB(1): Xie-Beni para os dados de treinamento

XB(2): Xie-Beni para os dados de teste

Tabela V  
RESULTADOS DO K-MEANS

Clusters	Covid X Outros X Saudável	Acurácia(%)
	XB(Treino) XB(Teste)	
3	0.33 0.22	54.0
6	1.05 0.70	27.9
9	1.78 1.21	18.3
12	2.32 1.69	16.1

O resultado obtido pelo algoritmo de agrupamento *K-Means* obteve uma acurácia similar ao FCM, porém ao contrário da primeira análise, o critério de *Xie-Beni* retornou uma quantidade de agrupamentos ideal inversamente proporcional à acurácia. Ao variar o conjunto de atributos selecionados pelo critério da razão discriminante de Fisher para o algoritmo *C-Means*, o desempenho se mostrou melhor quando todos os atributos extraídos pelos modelos convolutivos eram utilizados (com exceção dos valores nulos). Enquanto no algoritmo *K-Means* houve a necessidade em utilizar uma quantidade menor de atributos (222), uma vez que o resultado para o conjunto total de atributos apresentou um desempenho inferior em relação ao cenário em que foi utilizado maior quantidade de atributos.

Outros trabalhos na literatura, também fazem uso de técnicas aqui estudadas, tais como *transfer learning* para extração de características e ensemble das características para aumentar a diversidade do modelo. [25] utilizou imagens de TC para classificação entre COVID e não COVID, o desempenho do modelo para a classe COVID ficou em 98.68% de precisão e 99.20% de sensibilidade. [26] utilizou transfer learning das redes convolutivas DenseNet201, ResNet50, VGG16, e Xception e o classificador para os modelos supervisionados máquina de vetor de suporte, floresta aleatória, árvore de decisão e KNN obtiveram medida de desempenho de 100% em termos de precisão, sensibilidade e acurácia usando as características extraídas da *DenseNet201*.

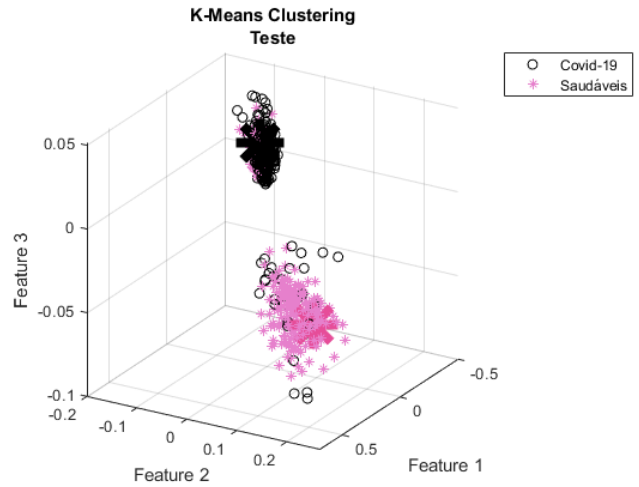


Figura 6. Centros obtidos com K-Means nos dados de teste.

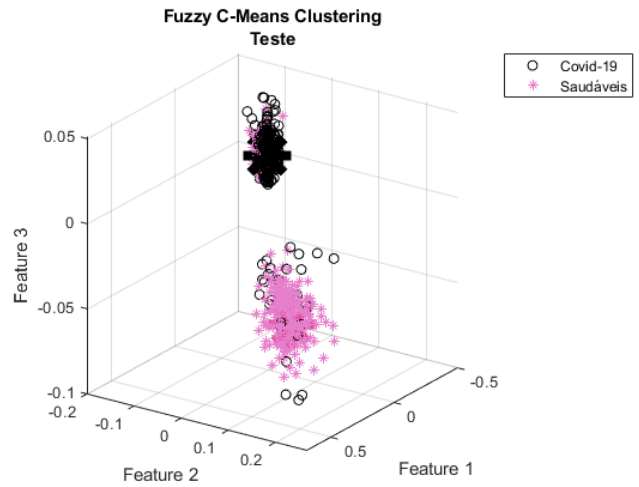


Figura 7. Centros obtidos com C-Means nos dados de teste.

Como pode ser observado, no cenário em que se compara amostras de pessoas com Covid em relação às amostras de pessoas saudáveis (Figuras 6 e 7), ao considerar apenas 3 características, alguns dos vetores de atributos de uma classe apresentam estas 3 características similares ao Covid.

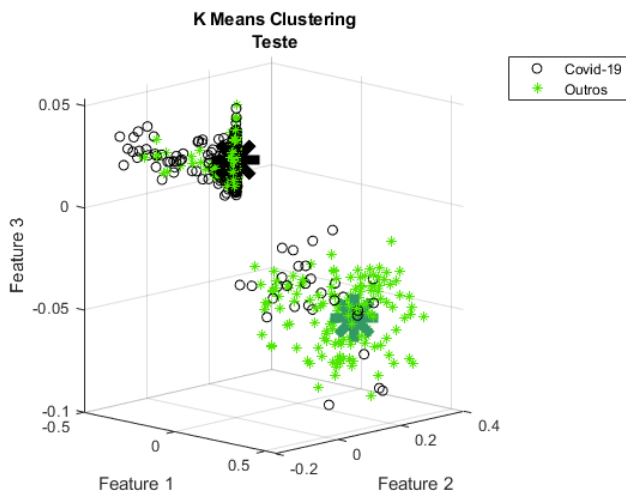


Figura 8. Centros obtidos com K-Means nos dados de teste.

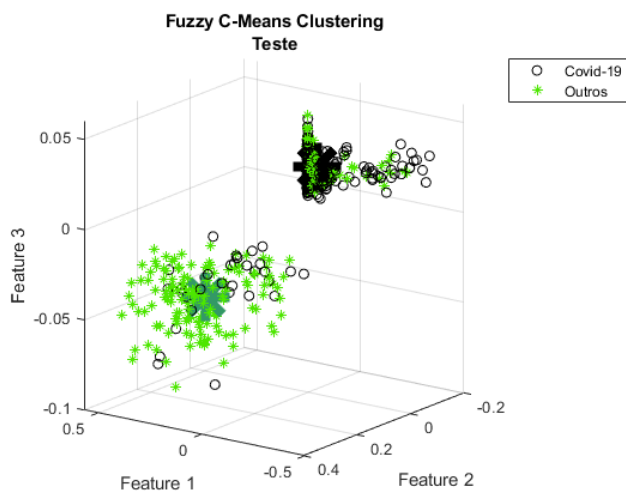


Figura 9. Centros obtidos com C-Means nos dados de teste.

Já no caso em que se compara amostras de pessoas com Covid em relação à amostras de pessoas com outros tipos de doenças (Figuras 8 e 9), a quantidade de amostras que apresentam as 3 características principais similares ao Covid é visivelmente superior que no primeiro caso.

## V. CONCLUSÃO

De acordo com os resultados obtidos, pode-se inferir que o sistema baseado em modelos previamente treinados para outras bases de dados captaram as características das imagens do banco de dados utilizado no trabalho.

Baseando-se na métrica de acurácia, observa-se que os melhores resultados obtidos com o *fuzzy C-Means* (74,2%) e com o *K-Means* (73,4%) sinalizam um desempenho razoável para a classificação de imagens de tomografia computadorizada de COVID-19, dada a característica não supervisionada dos algoritmos.

Algumas melhorias podem ser implementadas em trabalhos futuros, como a utilização de um modelo treinado com imagens em dimensão adequada à dimensão da base de dados que foi testada. Além disso, esperam-se em trabalhos futuros utilizar classificadores supervisionados.

## REFERÊNCIAS

- [1] F. ZHENG, C. LIAO, Q.-h. FAN, H.-b. CHEN, X.-g. ZHAO, Z.-g. XIE, X.-l. LI, Z.-j. WANG, J. WU, X.-y. WU, and R.-m. JIN, "Clinical Characteristics of Children with Coronavirus Disease 2019 in," vol. 40, no. 2, 2020.
- [2] T. Kobayashi, S.-m. Jung, N. M. Linton, R. Kinoshita, K. Hayashi, T. Miyama, A. Anzai, Y. Yang, B. Yuan, A. R. Akhmetzhanov, A. Suzuki, and H. Nishiura, "Communicating the Risk of Death from Novel Coronavirus Disease (COVID-19)," p. 580, 2020.
- [3] M. Park, R. S. Thwaites, and P. J. Openshaw, "COVID-19: Lessons from SARS and MERS," *European journal of immunology*, vol. 50, no. 3, pp. 308–311, 2020.
- [4] Y. Pan, H. Guan, S. Zhou, Y. Wang, Q. Li, T. Zhu, and Q. Hu, "Initial CT findings and temporal changes in patients with the novel coronavirus pneumonia ( 2019-nCoV ): a study of 63 patients in Wuhan , China," no. December 2019, pp. 3306–3309, 2020.
- [5] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song, K. Cao, D. Liu, G. Wang, Q. Xu, X. Fang, S. Zhang, J. Xia, and J. Xia, "Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy," *Radiology*, vol. 296, no. 2, pp. E65–E71, 2020.
- [6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Microsoft Research, 2015.
- [7] M. Tan and Q. V. Le, "EfficientNet : Rethinking Model Scaling for Convolutional Neural Networks," 2019.
- [8] T. T. Santos, L. Bassoi, and R. L. Martins, "Automatic grape bunch detection in vineyards based on affordable 3D phenotyping using a consumer webcam," no. October, 2017.
- [9] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, and D. Shen, "Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for COVID-19," *arXiv*, vol. 14, pp. 4–15, 2020.
- [10] L. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338 – 353, 1965.
- [11] M. Dayane, A. Araujo, and R. X. D. Brito, "Modelo de Previsão Fuzzy Como Mecanismo de Pré-diagnóstico da Esquizofrenia em Adultos," pp. 79–83, 2020.
- [12] A. M. Tripathi and A. Mishra, "Fuzzy unique image transformation: Defense against adversarial attacks on deep COVID-19 models," *arXiv*, vol. 14, no. 8, pp. 1–11, 2020.
- [13] G. Valentini and T. Dietterich, "Bias-variance analysis and ensembles of svm," vol. 2364, 06 2002, pp. 222–231.
- [14] V. C. Mota, F. A. Damasceno, E. A. Soares, and D. F. Leite, "Fuzzy clustering methods applied to the evaluation of compost bedded pack barns," *IEEE International Conference on Fuzzy Systems*, 2017.
- [15] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. Fifth Berkeley Symp. Math. Stat. Probab. Vol. 1 Stat.* Berkeley, Calif.: University of California Press, 1967, pp. 281–297. [Online]. Available: <http://projecteuclid.org/euclid.bsmsp/1200512992>
- [16] R. Krishnapuram, A. Joshi, O. Nasraoui, and L. Yi, "Low-complexity fuzzy relational clustering algorithms for Web mining," *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 4, pp. 595–607, 2001.
- [17] A. JAIN, M. MURTY, and P. FLYNN, "Data clustering," *Data Clustering: A Review*, vol. 31, no. 3, pp. 264–323, 1999.
- [18] H. Gunraj, A. Sabri, D. Koff, and A. Wong, "Covid-net ct-2: Enhanced deep neural networks for detection of covid-19 from chest ct images through bigger, more diverse learning," 2021.

- [19] H. Gunraj, L. Wang, and A. Wong, "Covidnet-ct: A tailored deep convolutional neural network design for detection of covid-19 cases from chest ct images," *Frontiers in Medicine*, vol. 7, p. 1025, 2020. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fmed.2020.608525>
- [20] E. Soares, P. Angelov, S. Biaso, M. H. Froes, and D. K. Abe, "SARS-CoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification," p. 2020.04.24.20078584, may 2020. [Online]. Available: <https://doi.org/10.1101/2020.04.24.20078584>
- [21] J. D. Guedes, D. D. Ferreira, and B. H. Barbosa, "A non-intrusive approach to classify electrical appliances based on higher-order statistics and genetic algorithm: a smart grid perspective," *Electric Power Systems Research*, vol. 140, pp. 65–69, 2016.
- [22] C. Bishop, *Pattern Recognition and Machine Learning*, ser. Information Science and Statistics. Springer New York, 2016. [Online]. Available: <https://books.google.com.br/books?id=kOXDtAEACAAJ>
- [23] R. Naves, B. H. Barbosa, and D. D. Ferreira, "Classification of lung sounds using higher-order statistics: A divide-and-conquer approach," *Computer Methods and Programs in Biomedicine*, vol. 129, pp. 12–20, 2016.
- [24] R. S. F. Nascimento, R. E. V. Vargas, B. H. G. Barbosa, and I. H. F. dos Santos, "Detecção de falhas com stacked autoencoders e técnicas de reconhecimento de padrões em poços de petróleo operados por gas lift," in *Congresso Brasileiro de Automática (CBA)*, vol. 2, no. 1, 2020.
- [25] R. Kundu, H. Basak, P. K. Singh, A. Ahmadian, M. Ferrara, and R. Sarkar, "Fuzzy rank-based fusion of CNN models using Gompertz function for screening COVID-19 CT-scans," *Scientific Reports*, vol. 11, no. 1, pp. 1–12, 2021. [Online]. Available: <https://doi.org/10.1038/s41598-021-93658-y>
- [26] S. M. Rezaei, M. Ghorvei, R. Abedi-Firouzjah, H. Mojtahedi, and H. Entezari Zarch, "Detecting COVID-19 in chest images based on deep transfer learning and machine learning algorithms," *Egyptian Journal of Radiology and Nuclear Medicine*, vol. 52, no. 1, 2021.