**Medical images clustering by similarities using unsupervised learning.**

Agrupamiento de imágenes médicas por similitudes usando aprendizaje no supervisado.

Alejandra Verónica **López-Chiquito**1, Marco Antonio **Aceves-Fernandez**2, Sebastián **Colores-Salazar**3

1,2 Universidad Autónoma de Querétaro, México

<https://orcid.org/0009-0008-9295-8019> | [alopez251@alumnos.uaq.mx](mailto:alopez251@alumnos.uaq.mx)

<https://orcid.org/0000-0002-5455-0329> | [marco.aceves@gmail.com](mailto:marco.aceves@gmail.com)

2 Centro de Investigaciones en Óptica, México

<https://orcid.org/0000-0002-6353-0864> | [sebastian.salazar@cio.mx](mailto:alopez251@alumnos.uaq.mx)

**Abstract**

Typically, a chest CT scan is ordered by a specialist physician to analyze a specific organ in this patient's area, such as the heart, lungs, liver, etc. Despite knowing the organ that needs to be examined, the patient's CT will contain images from the hip to the shoulders. This article explores the implementation of clustering using a set of chest CT images from patients with fibrosis to classify slices by similarity, to reduce medical analysis time, and to exclude uninformative slices or pictures, such as initial or final images that do not provide useful information to detect diseases and complications within the thoracic cavity. This allows focusing on images that contain complete views of major organs. Image preprocessing techniques were used to highlight their features and edges, deep learning models such as convolutional networks for feature extraction, and a clustering technique to divide pictures by similarity. In the end, the goal was achieved and the use in the future for classification is wide.

**Index terms:** Deep learning, unsupervised learning, feature extraction, clustering, convolutional neural networks.

**Resumen**

Normalmente, un médico especialista solicita una tomografía computarizada (TC) de tórax para analizar un órgano específico en la zona de este paciente, como el corazón, los pulmones, el hígado, etc. A pesar de conocer el órgano que necesita ser examinado, la TC del paciente contendrá imágenes desde la cadera hasta los hombros. Este artículo explora la implementación de la agrupación utilizando un conjunto de imágenes de una TC de tórax de pacientes con fibrosis para clasificar los cortes por similitud, con el objetivo de reducir el tiempo de análisis médico y así excluir cortes o imágenes no informativos, como imágenes iniciales o finales que no brindan información útil para detectar enfermedades y complicaciones dentro de la cavidad torácica. Esto permite enfocarse en imágenes que contienen vistas completas de órganos principales. Para esto se utilizaron técnicas de preprocesamiento a las imágenes para resaltar sus características y bordes, modelos de aprendizaje profundo como redes convolucionales para extracción de características y una técnica de agrupamiento (clustering) para dividir las imágenes por similitudes. Al final se logró el objetivo y el uso en el futuro para la clasificación es amplio.

**Palabras clave:** Aprendizaje profundo, aprendizaje no supervisado, extracción de características, agrupamiento, redes neuronales convolucionales.

1. introduction

Analysis of medical images has become more relevant in recent years, especially in diagnosing various diseases. With computed tomography (CT), several conditions can be detected using artificial intelligence. However, the large volume of data generated by these scans makes the diagnostic process challenging, necessitating advanced image processing techniques to assist healthcare professionals in making more accurate and faster interpretations.

In artificial intelligence (AI), there are datasets (collections of data) with two types of data: labeled and unlabeled. The first one involves data (images, video, text, etc.) with a tag or class assigned to provide context to the complete database, and it is used for supervised learning. The second type of data is unlabeled; it doesn't contain classes, context, or additional information. This data type is used for unsupervised learning and requires more pre-processing work before being used to train an AI model [1]. It is important to note that documentation about supervised learning is more common than unsupervised learning. Unsupervised learning faces more difficulties because the investigation should be done without labels, and the results are subjective. This means some metrics can be useful, but there are no standards to follow. Also, the examples or study cases about this topic are rarer than those of supervised learning. Finally, several techniques are available for unsupervised learning; each one provides a different focus, which can create confusion about the best one in a specific situation.

This paper presents an unsupervised technique, clustering, that provides a strategy for treating unlabeled images effectively: creating groups, called clusters, with similar patterns. Clustering is very useful in areas where manual labeling is expensive or impractical, such as medical imaging, and this process can involve different steps and techniques.

The preprocessing process is usually the first step in unsupervised learning. It reduces noise and enhances image quality, which is essential for more precise analysis. There are various types of filters in computer vision, and their usage depends on the specific feature that needs to be highlighted. There are filters for noise reduction, edge detection, texture detection, etc. Specifically, these filters are used as very efficient tools to highlight small details in treating medical images.

Once the images are preprocessed, the feature extraction techniques can be applied to them. It is the process of transforming data, such as images, into measurable objects, such as vectors, that represent essential patterns or properties [2][3]. The most common technique to extract image characteristics is using a convolutional neural network (CNN), focusing on parameters such as texture, shape, and lesion density, which serve as fundamental inputs for the clustering algorithm. Convolutional neural networks have been applied to solve major problems in several areas and topics. Working with labeled databases, they have given excellent results, however, in this specific case where the aim is to classify images based on their similarities, three CNN models were used. All of them are pre-trained with the database ‘ImageNet’ [4], which means the network weights are already established because the models were trained to solve a different problem, this process is known as transfer learning. Transfer learning is usually applied to solve new problems faster or to find a better solution. [5]

Finally, the clustering approach is evaluated to group the images so the patterns can be more easily identified and classified. Clustering is an unsupervised machine learning technique used to create groups of data called clusters based on the similarities of the elements, in this way, it is not necessary to assign a label. Clustering is also useful for detecting patterns, identifying similar features, and categorizing unlabeled images into categories. K-means is one of these kinds of algorithms and it is at the top of the models used for simplicity. [1]

After all the process descriptions, all the slices from a CT image group are correctly classified by similarities and ready to be used by medical teams. Also, the most useful cluster can be selected and used in future investigations, for example, to identify a specific element inside, such as the heart, lungs, etc.

1. Background

*A. Related Work*

The number of articles available about supervised learning is huge. Supervised learning is the most common situation in real problems using artificial intelligence. The classification of medical images using labeled databases is constantly finding new ways to improve the results [6][7][8]. However, unsupervised learning has increased due to the complexity of the medical problems being harder, especially when images are involved. The specialists interested in solving those problems are finding new paths to use the technology and provide different workarounds to unsupervised learning. The CT is a non-invasive medical procedure that provides a large volume of data to the medical specialist due to the machine's high resolutions, especially the chest zone, which can be analyzed slice by slice helping to detect anomalies and patterns of different diseases. [9]

In this paper the CT images will be used, nevertheless, there are different types of medical images used to examine the human being. The 3D images are very useful and provide a large volume of data, for example, Aiham Taleb (et. al) [10] used them to train a 3D U-Net model to group the images. Due to the complexity of the problem, multiple losses, augmentation techniques, and different evaluation metrics were applied to find the solution.

Euijoon Ahn (et. al), proposed a solution to a feature augmentation scheme to improve the discriminative power of feature representation [11]. This technique was applied to 3 different databases with excellent results. The CCN used in this work was a pre-trained AlexNet, which means that transfer learning was also used to solve the problem exposed.

As it was mentioned, unsupervised learning has a lot of challenges, and every investigator finds a way to solve it. In [12], the unsupervised technique K-means was used as a helper to reduce the number of false edges and over-segmentation. The principal goal of this work is segmentation; however, the clustering method was a very good way to improve the results.

It is interesting to analyze the background of a very large field of artificial intelligence as clustering or unsupervised learning. The applications are infinitive and every day new articles are posted proposing a new problem with a new combination of methods as a solution.

*B. Challenges and Contributions*

The solution proposed in this contribution is making a system to group the slices of a chest CT by similarities to reduce the analysis time by the medical part using preprocessing work, such as filters, feature extractions, and K-means clustering by similarities.

One of the challenges to overcome in this work is the resolution of the topographies. The resolution database used for this work is not high, however, if the system can cluster this kind of image, the results with a high-resolution computed tomography (HRCT) will be extraordinary.

1. Methodology

*A. Database*

The database used for this paper comes from the Open Source Imaging Consortium (OSIC) data repository [13], It is a not-for-profit association of several medical specialists focused on detecting and diagnosing pulmonary diseases with artificial intelligence. In the repository, there are a lot of datasets related to different pulmonary diseases, such as idiopathic pulmonary fibrosis (IPF), and fibrosing interstitial lung diseases (ILDs). The database used in this paper is related to diagnosed patients with IPF and is available on the Kaggle website. [14]

The database contains computed topographies (CT) and extra information about 176 patients. The additional information file describes the general data about the patients, such as gender, age, pulmonary capacity, etc. This information was not used for this work, however, all the computed topographies helped to train a neural network to identify the objectives. All CT scan sections are axial.

*A grey circle with a black background

Description automatically generated with medium confidence*

**Fig. 1.** Examples of images from the OSIC dataset [10]. Patient ID: **ID00078637202199415319443.**

Inside the dataset, the number of slices of the patients is not balanced, which means that there are patients with less than 30 slices and patients with more than 1000 slices. Due to this, 30 patients with more than 350 slices were selected for this work. Even with this patient selection, the number of images is more than 14,000.

**Fig. 2.** Number of slices per patientA graph of a patient

Description automatically generated

*B. Preprocessing*

Using medical images, preprocessing work is essential to enhance the quality of the images. As with a lot of image classification papers the dataset was analyzed to have a database correctly balanced. Different sizes of images were identified in the database. 9778 images with the size of 512 x512 and 4813 images with size of 768 x768 were reduced to 224 x224. Additionally, the images were normalized with values between 0 and 255.

In this case, edge detection is a fundamental part of the process, applying these types of filters makes all the details in the images more easily identified. However, reducing the noise is also a technique used in computer vision works. As the best results are sought, edge detection was not the only technique applied to the images. Gaussian filter is normally used to soften the images and blur the edges [15]. In medical cases, the minimal details need to be highlighted, however the best combination of techniques needs to be found.

About the edge detections, three different filters were applied to the images. Laplacian filter helps to identify the zones where the intensity is changing suddenly using the second order derivative [16]. The Prewitt filter uses the first-order derivate to find the areas where the intensity changes abruptly, it highlights edges by calculating the gradient magnitude of an image. [17] And finally, the Sobel filter was applied to the images. It is highly regarded for its ability to reduce noise while effectively enhancing edges, offering clearer results compared to simpler gradient detection methods, and computing the gradient magnitude of an image to highlight edges. [17]

Once the filters were applied to the CT of one patient all the slices were stored in 4 different H5 files. Every file contains the images of the same patient with different filters applied.

**

*A close up of a logo

Description automatically generated*

A close up of a logo

Description automatically generated

*A close-up of a logo

Description automatically generated*

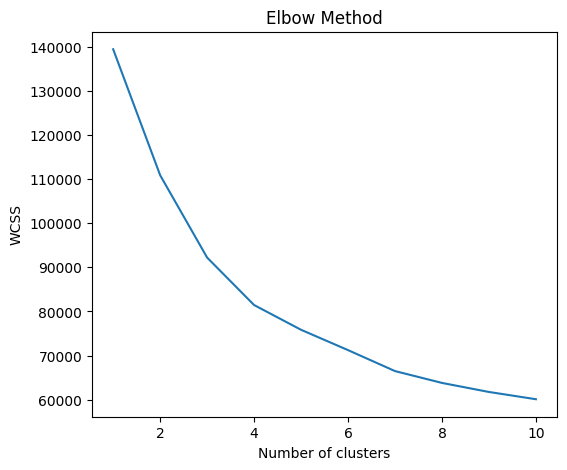
**Fig. 3.** Images after the filters apply.

*C. Feature Extractions*

As mentioned, 3 models were selected for the feature extraction using transfer learning, VGG16, VGG19, and RedNet50. The final layers were disabled because the goal of those layers is to classify the images provided, as the models were trained with the 'ImageNet' database, the model will classify the input into 1000 categories, and it is not useful in this case. The parameter that allows disabling the classification layer is ‘include top=False’. Also, a new final layer was added to all the models to reduce the characteristics of spatial dimensions. The GlobalAveragePooling2D layer is the last layer of our models. It takes an input of (W, X, Y, Z) where W is the number of the images of slices from one patient, and X, Y, and Z are the 3 channels of the images, and the result is one tensor with two dimensions (W, X) where W is the same value of the input and X is the average of the channels from the images.

Once the models are compiled, the images are passed to them by one patient at the same time. One of the filters is selected, and the images are loaded from the H5 file. Original images without filters applied also were included in the rounds. The last step of this process is the neural network models extracting the features of the complete tomography as a vector of two dimensions.

*D. Clustering*

There are a lot of models for clustering unlabeled data, in this case, the model will create groups or clusters using the similarity of the features that were extracted in the CNN from the images. The K-means is known as a technique for unsupervised learning. This model was selected and used in this work because it can find the patterns inside the features and, also find the correct k value, which means the correct number of clusters that need to be created using the elbow method. The elbow method evaluates the inertia value behavior now to increase the number of clusters since 2 and the maximum number of clusters selected, normally the maximum number selected is 10. [3]

**Fig. 4.** Example of elbow method applied to a patient with 1018 slices.

Once the optimum number is found with the elbow method, the K-means technique is applied to the features extracted by the CNN, and the images are classified in those clusters based on their similarities. The results for the clustering method depend on several things, the number of slices in a tomography in a patient, the similarities between the images, the number of optimal clusters, etc.

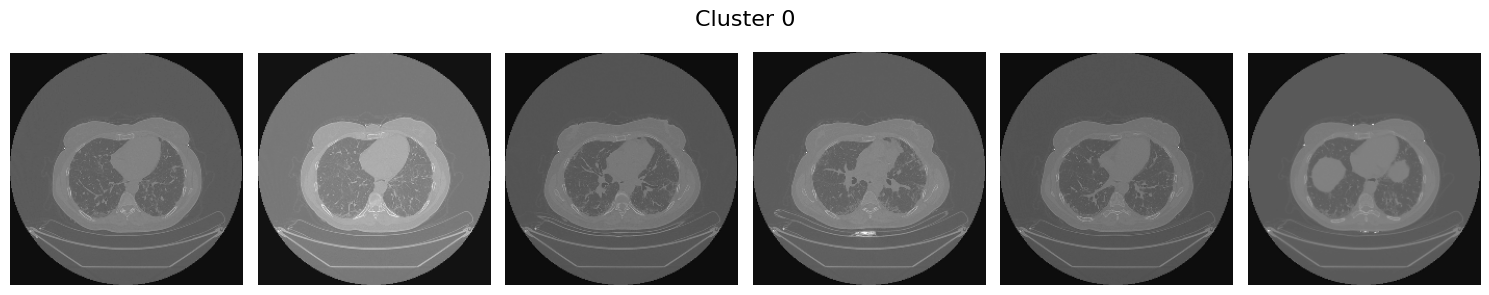
1. Results

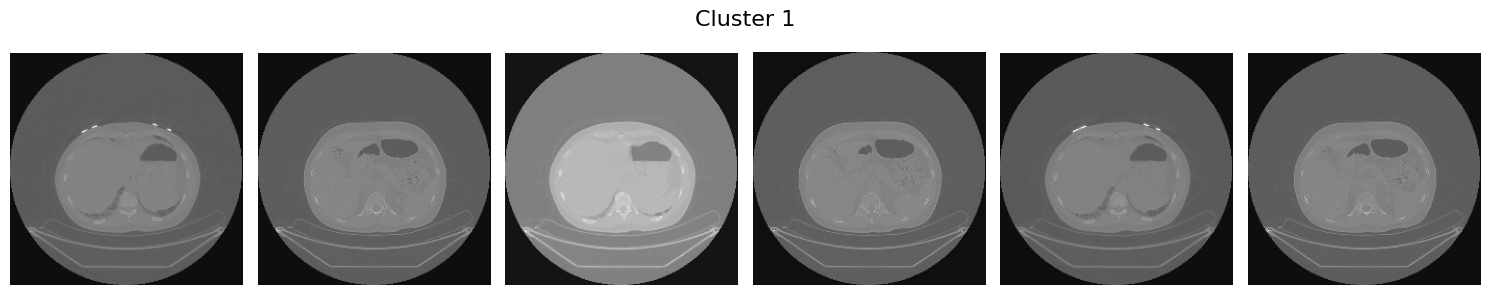
The preprocessing work, the features extractions, and clustering were applied one by one patient due to applying all the processes to all the patient images at the same time took a lot of time, also the computer cost was high. When all the process was applied to all the patient’s images, the preprocessing work was working well, however, the results of the clustering with K-means were not correct, which means the images were not classified correctly because the model seemed overfitted.

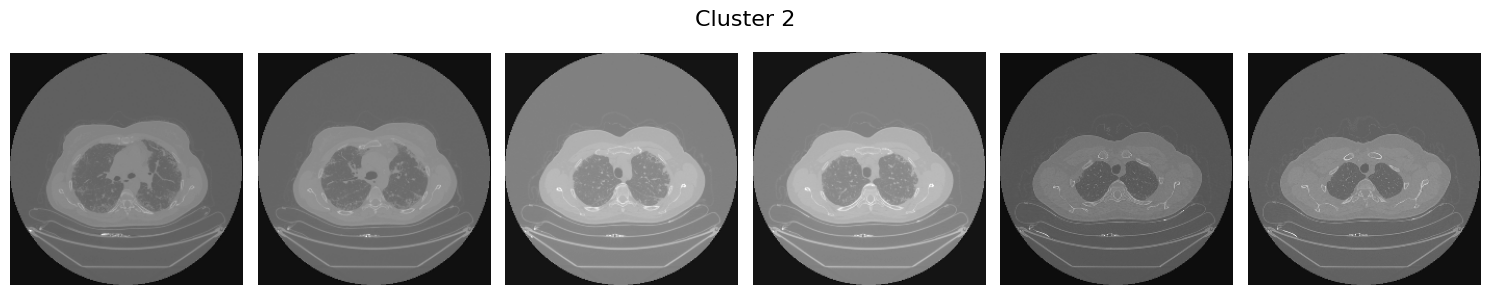
The results for the preprocessing work applied by patience separately were obtained rapidly and successfully. They are shown in Fig 3. The Gaussian filter softened the images, and with the rest of the filters, the edges were highlighted in different textures. The images that result from that were converted from 2 to 3 channels to be used for the feature extraction.

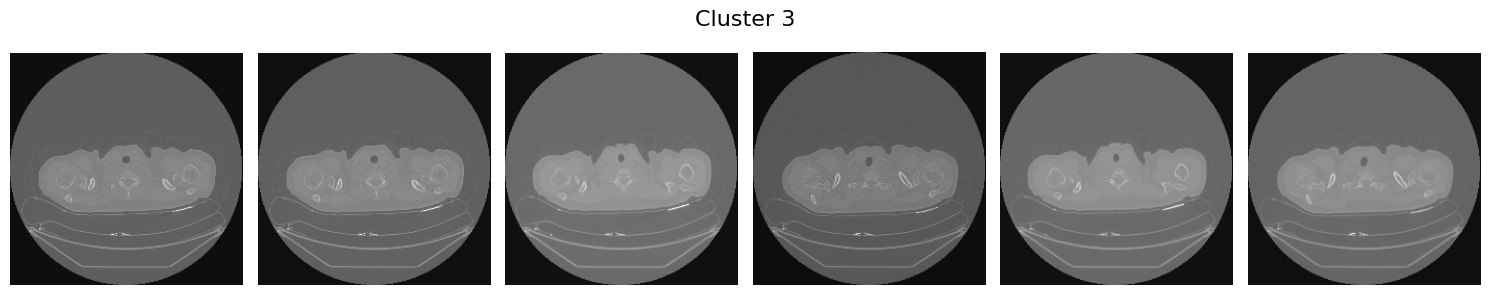
Is important to mention that the results depend on several factors, for example, the quality of the images, the number of slices by patient, etc. However, the results of the unsupervised classification using the methods mentioned in the paper were successful. In most of the patients, the number of clusters provided by the elbow method was 4, only with a few exceptions where k=3 was the optimum number of clusters. Once the images from every cluster are plotted, it is visible that the images were classified by similarities (Fig 5). The slices with more useful information were stored in a single cluster, meanwhile, the slices from the top and buttons of the thorax are in 2 different clusters. Also, when K=4, there is a cluster that is filled with slices where the lungs are visible, however, they haven’t more principal organs.

The feature extraction process made by the 3 different CNNs (VGG16, VGG19, and ResNet50) has no significant differences. The features of the images were stored in vectors of numbers apparently without sense at the human eye, nevertheless when the K-means model was applied, the results showed that the features were extracted correctly, and it was enough for the models to cluster the similarities and the number of slices without information of the main organs decreased in all the patient cases significantly.









**Fig. 5.** Result from the clustering process when K=4 from a patient with 1018 slices.

1. Conclusions

Once the clusters are plotted, notice that this unsupervised classification could be useful to the medical specialist, due to the slices without valuable information being described and the analysis work can be focused on the main organs of the patients. This work does not exclude any organ as principal, but it highlights the importance of saving time when an organ needs to be examined by a specialist and the tomography of the patient has a lot of slices.

After checking the number of images that are stored in the clusters, taking for example the patient with 1018 slices, the number of slices in the ‘correct cluster’ is 516. In a real application, it can be reflected in the number of slices the medical team needs to check from 1018 to 516 and it is a significant reduction of work from almost 50%.

A graph of a number of blue bars

Description automatically generatedAs mentioned, this work is not trying to exclude any other organ as important or principal, it knows that a chest tomography is requested to check organs such as the lung, heart, and principal arteries.

**Fig. 6.** Number of slices stored by cluster as result from K-means when K=4 from a patient with 1018 slices.

There are specific cases where these organs are not the focus of the chest CT. Knowing the latest, the results from this paper are helping to find the most important slices inside a CT reduce the analysis work from a medical team, and even diagnose a disease in the principal organs of the human being.

1. Future Work

In the end, this solution is focused only on unsupervised classification, however, there are several paths to improve the system proposed, such as using HRCT, application of new filters, improving the architecture of the CNN, etc. This is the first step in trying to reduce the medical work. Using the results from this investigation the creation of a website to find a specific organ in chest tomography is the next step.

References

1. Mahesh, B. “Machine Learning Algorithms - A Review. International Journal of Science and Research” **(IJSR), 9(1), 381–386.** 2020. <https://doi.org/10.21275/art20203995>
2. Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D., & Saeed, J. “A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction.” Journal of Applied Science and Technology Trends, 1(1), 56–70. 2020. <https://doi.org/10.38094/jastt1224>
3. Yu, S. S., Chu, S. W., Wang, C. M., Chan, Y. K., & Chang, T. C. (2018). “Two improved k-means algorithms.” Applied Soft Computing Journal, 68, 747–755. <https://doi.org/10.1016/j.asoc.2017.08.032>
4. Deng, Jia & Dong, Wei & Socher, Richard & Li, Li-Jia & Li, Kai & Li, Fei-Fei. ImageNet: a Large-Scale Hierarchical Image Database. IEEE Conference on Computer Vision and Pattern Recognition. 248-255. 2009. 10.1109/CVPR.2009.5206848.
5. S. J. Pan and Q. Yang, "A Survey on Transfer Learning," in IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345-1359, Oct. 2010, doi: 10.1109/TKDE.2009.191.
6. Gaur, L., Bhatia, U., Jhanjhi, N. Z., Muhammad, G., & Masud, M. “Medical image-based detection of COVID-19 using Deep Convolution Neural Networks.” Multimedia Systems, 29(3), 1729–1738. 2023. <https://doi.org/10.1007/s00530-021-00794-6>
7. de Sousa PM, Carneiro PC, Pereira GM, et al. “A new model for classification of medical CT images using CNN: a COVID-19 case study.” Multimed Tools Appl. Published online December 19, 2022. doi:10.1007/s11042-022-14316-7
8. Díaz-Pernas FJ, Martínez-Zarzuela M, Antón-Rodríguez M, González-Ortega D. “A Deep Learning Approach for Brain Tumor Classification and Segmentation Using a Multiscale Convolutional Neural Network.” *Healthcare*. 2021; 9(2):153. <https://doi.org/10.3390/healthcare9020153>
9. Rodriguez, K., Ashby, C. L., Varela, V. R., & Sharma, A. “High-Resolution Computed Tomography of Fibrotic Interstitial Lung Disease.” Seminars in Respiratory and Critical Care Medicine, 43(6), 764–779. 2022. https://doi.org/10.1055/s-0042-1755563
10. Aiham Taleb, Winfried Loetzsch, Noel Danz, Julius Severin, Thomas Gaertner, Benjamin Bergner, and Christoph Lippert. “3D self-supervised methods for medical imaging.” Curran Associates Inc. Article No.: 1524, Pages 18158 – 18172. 2020. <https://dl.acm.org/doi/10.5555/3495724.3497248>
11. Ahn, Euijoon & Kumar, Ashnil & Fulham, Michael & Feng, David Dagan Feng & Kim, Jinman, “Unsupervised Domain Adaptation to Classify Medical Images Using Zero-Bias Convolutional Auto-Encoders and Context-Based Feature Augmentation.” IEEE Transactions on Medical Imaging. PP. 1-1. 2020. [10.1109/TMI.2020.2971258](mailto:10.1109/TMI.2020.2971258)
12. H. P. Ng, S. H. Ong, K. W. C. Foong, P. S. Goh, and W. L. Nowinski, "Medical Image Segmentation Using K-Means Clustering and Improved Watershed Algorithm," IEEE Southwest Symposium on Image Analysis and Interpretation, Denver, CO, USA, 2006, pp. 61-65, doi: 10.1109/SSIAI.2006.1633722
13. Open Source Imaging Consortium (OSIC). <https://www.osicild.org/> (Accessed Nov. 07, 2024)
14. OSIC Pulmonary Fibrosis Progression database. Kaggel. <https://www.kaggle.com/c/osic-pulmonary-fibrosis-progression>. (Accessed Aug. 2024)
15. G. Deng and L. W. Cahill, "An adaptive Gaussian filter for noise reduction and edge detection," IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference, San Francisco, CA, USA, 1993, pp. 1615-1619 vol.3. 1993. doi: 10.1109/NSSMIC.1993.373563
16. Wilscy, M., Nair, M.S. “A New Method for Sharpening Color Images Using Fuzzy Approach.” In: Campilho, A., Kamel, M. (eds) Image Analysis and Recognition. ICIAR 2008. Lecture Notes in Computer Science, vol 5112. Springer, Berlin, Heidelberg. 2008. <https://doi.org/10.1007/978-3-540-69812-8_7>
17. Adlakha, Deepika, Devender Adlakha, and Rohit Tanwar. "Analytical comparison between Sobel and Prewitt edge detection techniques." International Journal of Scientific & Engineering Research 7.1 2016. <https://www.ijser.org/researchpaper/Analytical-Comparison-between-Sobel-and-Prewitt-Edge-Detection-Techniques.pdf>
18. Olalekan Salau, A., & Jain, S. “Feature Extraction: A Survey of the Types, Techniques, Applications.” 2019. IEEE. 10.1109/ICSC45622.2019.8938371