



Analysis and Segmentation of X-ray Images of COVID-19 Patients using the k-means Algorithm

¹Ahmet SAYGILI

¹Tekirdağ Namık Kemal University, Çorlu Engineering Faculty, Department of Computer Engineering, Tekirdağ

Abstract

COVID-19 is a virus from the coronavirus family that can have deadly effects. This virus has affected the whole world since the end of 2019. Early diagnosis and treatment of this virus directly affect its spread. For this reason, many different studies in many different fields are carried out for this purpose. There are many studies with computer-aided systems for the detection of the virus of COVID-19. All these works have a common goal; It contributes to the solution of stopping the spread of this virus. This is the main focus of our study. In this direction, studies have been put forward with image processing and machine learning methods. In our study, segmentation of X-ray images and clarification of anomalies on the lungs was performed using the k-means method. Thanks to this segmentation process, an application has been realized that can support physicians in their decisions with the help of X-ray images. Segmentation is the process of separating similar structures on an image into groups. These groups are capable of distinguishing between diseased and healthy regions. The results show that the segmentation process reveals significant results in the detection of disease-induced deterioration on lung images. Particularly, there were significant differences between the lung images of healthy individuals and those who had the disease. In the study, lung X-ray images of 10 healthy, 10 pneumonia but not COVID-19, and 10 COVID-19 (+) patients were used. The future goal of this study is to use the segmentation results obtained as input in a computer-based study that will automatically detect the disease and improve the current success. In addition, it is among our future goals to conduct comparative studies with different segmentation methods.

Keywords: Image Processing, COVID-19, Segmentation, X-ray, k-Means, Medical Imaging

k-Ortalamlar Algoritmasını Kullanarak COVID-19 Hastalarının X-ışını Görüntülerinin Analizi ve Bölütlenmesi

Özet

COVID-19 korona virüs ailesinden ölümcül etkileri olabilen bir virüsdür. Bu virüs 2019 yılı sonlarından bu yana tüm dünyayı etkisi altına almıştır. Bu virüsün erken tanı ve tedavisi yayılımını doğrudan etkilemektedir. Bu nedenle çok farklı alanlarda çok farklı çalışmalar bu amaç için gerçekleştirilmektedir. COVID-19'un virüsünün tespiti için bilgisayar destekli sistemler ile yapılan çok sayıda çalışma bulunmaktadır. Bütün bu çalışmalarında ortak bir hedefi vardır; bu virüsün yayılımını durdurmak, çözümü için katkı sağlamaktadır.

Makale Bilgisi

Başvuru:
08/07/2021
Kabul:
10/11/2021

Gerçekleştirdiğimiz bu çalışmanın da ana odak noktası budur. Bu doğrultuda görüntü işleme ve makine öğrenmesi yöntemleri ile çalışmalar ortaya konmuştur. Çalışmamızda k-ortalamlar yöntemi ile X-ışını görüntülerinin bölütlenmesi ve akciğerler üzerindeki anomalilerin belirginleştirilmesi işlemi gerçekleştirilmiştir. Bu bölütleme işlemi sayesinde X-ışını görüntüleri yardımıyla hekimlere alacakları kararlarda destek olabilecek bir uygulama gerçekleştirilmiştir. Bölütleme bir görüntü üzerindeki benzer yapıların gruplara ayrılması işlemidir. Bu gruplar hastalıklı ve sağlıklı bölgeleri birbirinden ayırt edebilecek nitelikte olmaktadır. Elde edilen sonuçlar göstermektedir ki bölütleme işlemi akciğer görüntüleri üzerindeki hastalıktan kaynaklı bozulmaların tespitinde anlamlı sonuçlar ortaya koymaktadır. Özellikle sağlıklı bireyle ile hastalık geçiren kişilerin akciğer görüntüleri arasında belirgin farklar ortaya çıkmıştır. Çalışmada 10 sağlıklı, 10 COVID-19 kaynaklı olmayan pnömoni ve 10 COVID-19 geçirmiş kişilere ait akciğer X-ışını görüntüsü kullanılmıştır. Bu çalışmanın gelecekteki hedefi elde edilen bölütleme sonuçlarını hastalığın otomatik olarak tespit edilmesini sağlayacak bilgisayar temelli bir çalışmada girdi olarak kullanmak ve mevcut başarıyı yukarılara taşımaktır. Bunun yanı sıra farklı bölütleme yöntemleri ile kıyaslamalı çalışmalar yapmakta gelecek hedeflerimiz arasındadır.

Anahtar Kelimeler: Görüntü İşleme, COVID-19, Bölütleme, X-ışını, k-Ortalamlar, Tıbbi Görüntüleme

¹ Corresponding Author e-mail: asaygili@nku.edu.tr

1 Introduction

COVID-19 is a type of virus from the coronavirus family that has affected the world. Since December 2019, the whole world has been working hard to overcome this virus. Early diagnosis and treatment of the virus are important in the course of the disease. For this reason, many computer-assisted studies are carried out for the detection of the virus. The common point of all these studies is that they can support specialist physicians and health workers. A test called the Polymerase Chain Reaction (PCR) test is used to detect the COVID-19 virus. This thesis is obtained as a result of studying the swab samples taken from the mouth and nose of the individuals in the laboratory environment. The disadvantage of this test is its sensitivity rate of 60-70% [1, 2]. For this reason, radiological imaging and blood analysis are used to support the test while making the diagnosis. X-ray imaging is used in radiological imaging because it gives fast and mostly good results.

There are many academic studies done with X-ray imaging. The first of the studies conducted in this area proposed a model on X-ray with the VGG16 model [3]. In their study, they used the data set [4] images by resizing it to (224×224) . The method they proposed achieved an accuracy rate of 96.1%. In the study of Abbas et al. [5], the status of being positive or negative from COVID-19 X-ray images was evaluated. The deep transfer learning method was used in the study. The principal component analysis method was used in the study to reduce the high number of features. The success of the system was measured with validation, sensitivity, and specificity metrics, and success rates of 95.12, 97.91, and 91.87 were obtained, respectively. Nour et al., using Deep Features and Bayes Optimization, kNN, SVM, and decision trees methods, also used X-ray images and COVID-19(+), COVID-19(-) and non-COVID-19 determined the pneumonia disease [6]. A success rate of 97.25% was achieved in the study.

The importance of segmentation in the analysis of medical images cannot be denied [7-9]. Particularly, segmentation approaches are frequently used to detect anomalies on images [7-9]. Segmentation or clustering of medical images in the k-means method is an effective method that is frequently used. There are many studies in this field in the literature. You can view some of these studies in the [10-13] articles.

In this study, we have also segmented COVID-19 X-ray images using the k-means method. The second part of the study consists of materials and methods, the third part consists of experimental studies and the last part consists of results.

2 Material and Methods

The data set we used in our study consists of X-ray images. The dataset includes 125 images of COVID-19 positive, 43 of which are female and 82 are male. In addition, there are 500 clear lung X-RAY images without any finding, and 500 images with pneumonia but not COVID. This dataset was created by bringing together images from different datasets [4, 14] in the study numbered [15]. 26 of the positive ones about the data set have age information and their average is 55. In our study, images of COVID-19 (+) and healthy people in the data set were used. Sample images of this data set are shown in Figure-1.

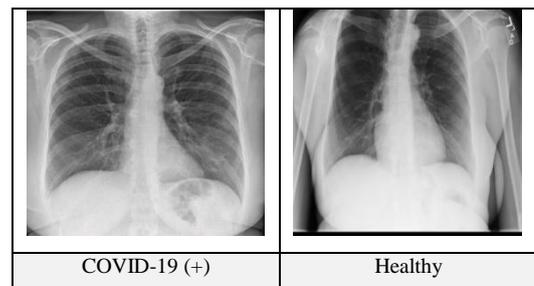


Figure 1. Sample images from the dataset

To analyze the segmentation approaches, a total of 30 images, 10 healthy, 10 pneumonia, and 10 COVID-19 (+), were used from the data set.

In our study, various preprocessing was applied before the segmentation of the images. The first of these pre-process is the resizing of images. The sizes of the images in the data set can be different from each other. For this reason, the images were resized (525×525) to have a standard structure. The next preprocessing step is the conversion of images from a multi-channel structure to a single-channel structure. Here, too, all images are converted to gray level.

After the preprocessing steps, the images were segmented using the k-means method. The K-means method is a segmentation method that has been used for many years and produces successful results. It has been successfully applied in many areas such as image processing, data mining, and machine learning. The k-means clustering algorithm was developed in 1967 by J.B. It is one of the oldest clustering algorithms developed by

MacQueen [16]. It is one of the most widely used unsupervised learning methods. The assignment mechanism of k-means allows each data to belong to only one cluster [17]. The working mechanism of the k-means algorithm is as follows;

1. First, k objects are selected, each representing the center or mean of a cluster.
2. The remaining objects are included in the clusters with which they are most similar, taking into account their distance from the mean values of the clusters.
3. Then, new cluster centers are determined by calculating the average value of each cluster, and the distances of the objects to the center are examined again.
4. The algorithm continues to shift until there is no change in the clusters.

The most common square error criterion (SSE) is used to evaluate the k-means clustering method. The clustering result with the lowest SSE value gives the most successful result. SSE calculation is calculated with the following formula [18].

$$SSE = \sum_{i=1}^k \sum_{x \in C_i} dist2(m_i, x) \quad (1)$$

x: An object in the C_i set,

m_i : center point of cluster C_i

k: Number of clusters

Here, dist is the standard Euclidean Distance between two objects, value x is an object in the set C_i , value m_i is the center point of the set C_i .

3 Experiments

In our study, healthy, pneumonia (not COVID-19 (+)) and COVID-19 (+) images were used in the segmentation of COVID-19 X-ray images. Different cluster numbers were tried for the segmentation process, but when the results were analyzed, it was decided to determine the number of clusters as 2. The results obtained in the study are given in Figures 2, 3, and 4.

In our study, the MATLAB platform was used in the coding phase. The steps of the operations performed in our study are shown in Figure 2.

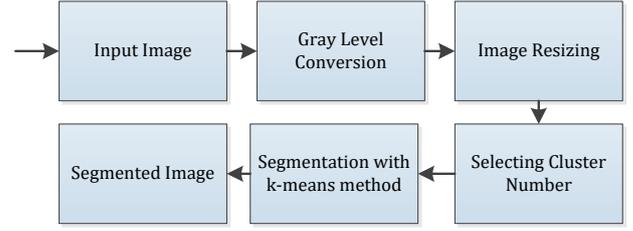


Figure-2. Flowchart of the process

When Figures-3, 4, and 5 are examined, it is seen that there are very clear differences between the images obtained as a result of segmentation of healthy and COVID-19 lung images. We can see the anomaly in the segmented images caused by the frosted glass image created by the COVID-19 virus on the lungs in Figure-3. As a result of the segmentation obtained as a result of the healthy lung images, it is seen that the region where the lungs are located is segmented as a whole, as can be seen in Figure-4, and there is no skipping between them. It can be seen from Figures-3 and 4 that automatic diagnosis systems can be created using these segmented images.

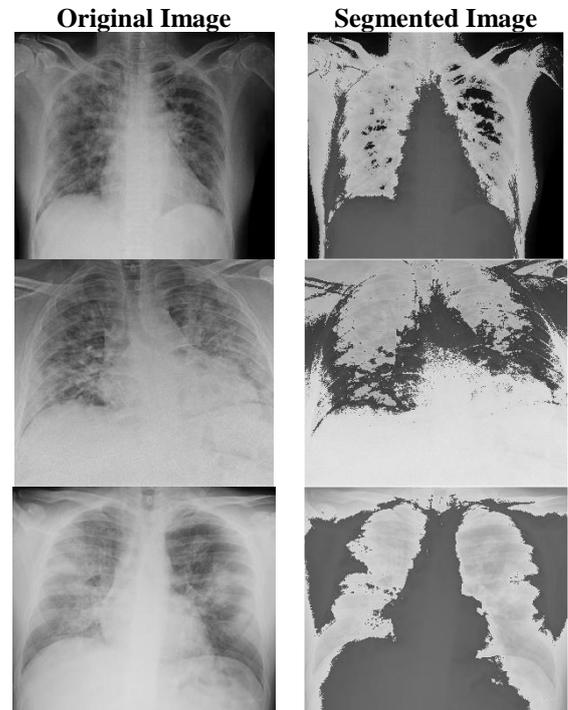


Figure-3 Segmentation of COVID-19 (+) images. The number of clusters: 2.

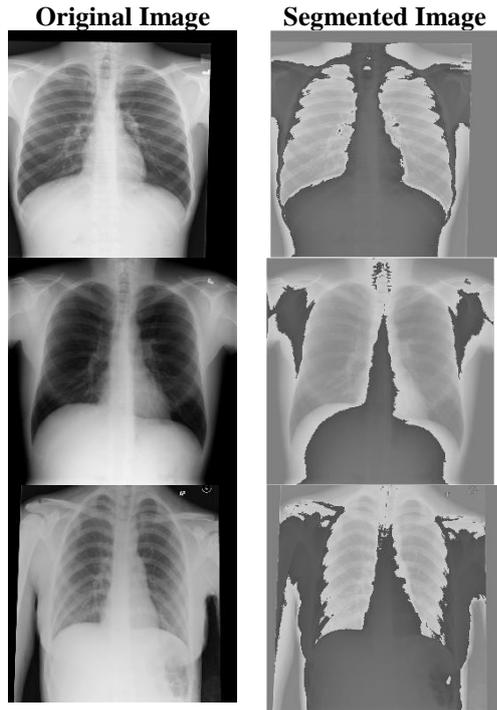


Figure-4 Segmentation of COVID-19 (-) images. Number of clusters=2.

Figure 5 shows the segmentation results obtained for images with pneumonia but without COVID-19. When Figure 3 and Figure 5 are compared, it is understood that COVID-19 has different effects from standard pneumonia and this situation can be seen more clearly as a result of segmentation. Similarly, similar differences are seen between healthy images and images with pneumonia. These differences show that the segmentation process highlights the differences in the images and the results obtained will positively affect the classification process.

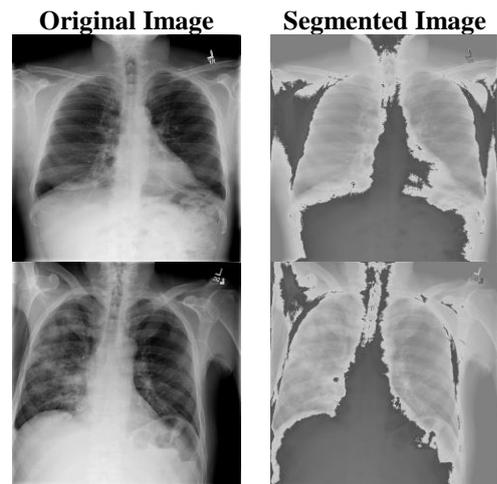


Figure-5 Segmentation of Pneumonia (not COVID-19 (+)) images. Number of clusters=2.

4 Results

In our study, the segmentation of X-ray images of healthy people and people who had COVID-19 was performed using the k-means method. The results obtained are shown on three COVID-19 (+) and three healthy lung X-ray images with different structures. Thanks to these images, it was seen that the foggy image caused by COVID-19 on the lungs, in other words, the frosted glass image, could be detected as a result of segmentation. It is aimed to make automatic diagnosis systems by using these results in classification processes in future studies. In this way, studies that can support medical professionals in early diagnosis and treatment will be revealed.

Acknowledgment

This study was supported by Tekirdağ Namık Kemal University Scientific Research Projects Coordination Unit. Project number: NKUBAP.06.GA.21.317

References

- [1] Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, Tao Q, Sun Z, Xia L. "Correlation of chest CT and RT-PCR testing for coronavirus disease". *COVID-19 in China: a report of 1014 cases, Radiology*, 296, E32-E40, 2020.
- [2] Lan L, Xu D, Ye G, Xia C, Wang S, Li Y, Xu H. "Positive RT-PCR test results in patients recovered from COVID-19". *Jama*, 323, 1502-1503, 2020.
- [3] Hall L.O., Paul R, Goldgof D.B., Goldgof G.M. "Finding covid-19 from chest x-rays using deep learning on a small dataset". *arXiv preprint arXiv:2004.02060*, 2020.
- [4] Cohen J.P., Morrison P, Dao L, Roth K, Duong T.Q., Ghassemi M. "Covid-19 image data collection: Prospective predictions are the future". *arXiv preprint arXiv:2006.11988*, 2020.
- [5] Abbas A, Abdelsamea M.M, Gaber M.M. "Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network". *Applied Intelligence*, 51, 854-864, 2021.
- [6] Nour M, Cömert Z, Polat K. "A novel medical diagnosis model for COVID-19 infection detection

- based on deep features and Bayesian optimization". *Applied Soft Computing*, 97, 106580, 2020.
- [7] Norouzi A, Rahim M, Altameem A, Saba T, Rad A.E., Rehman A, Uddin M. "Medical image segmentation methods, algorithms, and applications". *IETE Technical Review*, 31, 199-213, 2014.
- [8] Pham D.L., Xu C, Prince J.L. "Current methods in medical image segmentation". *Annual review of biomedical engineering*, 2, 315-337, 2000.
- [9] Sharma N, Aggarwal L.M. "Automated medical image segmentation techniques". *Journal of medical physics/Association of Medical Physicists of India*, 35, 3, 2010.
- [10] Arunkumar N, Mohammed M.A., Abd Ghani M.K., Ibrahim D.A., Abdulhay E, Ramirez-Gonzalez G, Albuquerque V.H.C. "K-means clustering and neural network for object detecting and identifying abnormality of brain tumor". *Soft Computing*, 23, 9083-9096, 2019.
- [11] Hrosik R.C., Tuba E, Dolicanin E, Jovanovic R, Tuba M. "Brain image segmentation based on firefly algorithm combined with k-means clustering". *Stud. Inform. Control*, 28, 167-176, 2019.
- [12] Karrar A.M.A., Sun J. "Segmentation and measurement of medical image quality using k-means clustering algorithm". *American Journal of Neural Networks and Applications*, 5, 36-44, 2019.
- [13] Sharma M, Purohit G, Mukherjee S. "Information retrieves from brain MRI images for tumor detection using hybrid technique K-means and artificial neural network (KMANN)". *Networking communication and data knowledge engineering, Springer*, 145-157, 2018.
- [14] Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers R.M. "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases". *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2097-2106, 2017.
- [15] Ozturk T, Talo M, Yildirim E.A., Baloglu U.B., Yildirim O, Acharya U.R. "Automated detection of COVID-19 cases using deep neural networks with X-ray images". *Computers in biology and medicine*, 121, 103792, 2020.
- [16] Likas A, Vlassis N, Verbeek J.J. "The global k-means clustering algorithm". *Pattern recognition*, 36, 451-461, 2003.
- [17] Davidson I. "Understanding K-means non-hierarchical clustering". *SUNY Albany Technical Report*, 2, 2-14, 2002.
- [18] Pang-Ning T, Steinbach M, Kumar V. *Introduction to data mining*, Addison-Wesley, 2005.