Computer Engineering and Applications Vol. 13, No. 1, February 2024

A Hybrid of Fuzzy C-Means For The Segmentation In CT Scan and X-Ray Images For Screening The COVID-19 Patients

Nitit WangNo[1*], Supailin Pichai[2], Raihan Mufid Setiadi[1]

[1] Major of Computer and Information Technology, Faculty of Information Technology, Roi Et Rajabhat University, Roi Et, Thailand
[2]Major of Computer Science, Department of Applied Science, Faculty of Science and Technology, Loei Rajabhat University, Loei, Thailand
**nitit@kkumail.com*

ABSTRACT

In this paper, using CT scan and X-ray images, we present a hybrid approach, based on combining fuzzy C-means with k-means clustering, to evaluate and determine pneumonia infection caused by the coronavirus disease (COVID-19). To achieve this objective, we introduce a hybrid method that combines fuzzy C-means clustering with K-means clustering. This hybrid approach is designed to effectively segment object boundaries within medical images, enabling the precise identification of pneumonia-related features. In addition to our hybrid method, we compare its performance with two other segmentation approaches: the Expectation Maximization (EM) algorithm and 2D Entropy segmentation. Which, the method we propose uses a comparison between the performances of the based on a database of medical imaging test. Experimental results showed that the proposed approach outperforms, it was found that the hybrid fuzzy C-means algorithm segmentation images methods give better performance in terms of accuracy, precision, and F-measure, which is effective in boundaries segmentation. Comparative results of the accuracy and image quality index demonstrate the robustness of AI. It also helps to improve work efficiency with accurate analysis of COVID-19 infection on CT scan and X-rays. In addition, the approach helps radiologists make clinical decisions for diagnosis, follow-up, and prognosis.

Keywords: Hybrid fuzzy C-means, images segmentation, diagnosis, COVID-19

1. INTRODUCTION

The world is currently grappling with the widespread impact of the COVID-19 pandemic, stemming from its rapid outbreak in Wuhan, China. Since early January 2020, as reported by the World Health Organization [1], there has been a sharp increase in the number of infected people and individuals under investigation in nearly all countries worldwide. COVID-19 symptoms include fever, acute respiratory failure, and lung infections, leading to a rapid surge in the number of infected patients and causing significant delays in the diagnosis and treatment processes. In response to this crisis, researchers worldwide have turned to machine learning methods to expedite disease diagnosis. The significance of accurate, swift, and precise COVID-19 diagnosis cannot be overstated. It not only saves lives and restricts the geographical spread of the disease but also offers an opportunity to leverage machine learning to create valuable data models. By utilizing chest radiographic datasets within the field of artificial intelligence (AI) [1], several studies have demonstrated that AI-based diagnoses are as accurate as those made by human experts, significantly reducing the time needed for radiological procedures. Furthermore, a previous study employing Deep Learning for COVID-19 diagnosis using X-ray images, as reported [2][4], emphasized that standard COVID-19 tests could be conducted using both X-rays and Computed Tomography (CT) scans. This approach offers a “fast-contact and short-lived” testing solution, readily accessible as X-ray equipment is available in hospitals. The study even suggested the use of cell phone techniques for performing CT scans.

However, despite these advancements in AI, an analysis of the research landscape over the past three years reveals a limited number of studies focusing on the segmentation of COVID-19 CT scan and X-ray images, particularly for screening severely affected patients. Nonetheless, studies such as Casey [8] have highlighted the advantages of CT scan and X-ray diagnosis reports, reinforcing physicians’ diagnoses of COVID-19, which can vary on a case-by-case basis. This approach enables physicians to predict the disease’s progression in individual patients and allocate medical resources effectively for further treatment. This effectiveness was further confirmed by the study conducted [6], which demonstrated that machine learning could aid in developing predictive algorithms to assess the mortality risk of infected patients, achieving an impressive 80% accuracy rate in disease prediction. Therefore, the primary objective of the present study is twofold: First, to develop a machine learning model capable of differentiating between COVID-19 cases and other respiratory conditions through a comparative analysis of CT scan and X-ray images. Second, to contribute significantly to the research community’s understanding of the potential applications of AI in medical diagnostics. This research aims to enhance the efficiency and accuracy of clinical COVID-19 diagnosis [7], bridging a research gap in image analysis, and ultimately advancing our understanding of AI’s role in the medical field.

2. MATERIALS AND METHODS

Artificial intelligence (AI), a new technology that allows the work or analysis of medical data more efficiently than humans. In this section, we will discuss how to extract both the CT scan and the X-ray image of Covid-19

2.1 K-MEAN CLUSTERING ALGORITHM

The K-Means algorithm is an unsupervised clustering algorithm that classifies input data points into multiple classes based on distance. The points are concentrated around the centroid $\mu_i \ \forall_i = 1\ldots$ k which is obtained by minimizing the objective [8].

$$v = \sum_{i=1}^{k} \sum_{x_j \in s_i} \left(x_j - \mu_i \right)^2 \tag{1}$$

where there are k clusters s_i , $i = 1, 2, \ldots, k$ and μ_i is the centroid or mean point of all the points $x_j \in s_i$.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

2.2 FUZZY C-MEANS

The Fuzzy C-means (FCM) method is a clustering procedure adopted by Dunn, enhanced by Bezdek, which was subsequently scaled up only by M. Matteucci during the local data segmentation process. Is considered in the FCM algorithm [12] membership functions are allowed for focusing each data item (data point) directly related to each cluster point. According to the distance between the cluster points, the membership function and group focus will be upgraded primarily from each round.

n Number of information focuses.
v_q Group focuses.
m Fuzziness index m ϵ [1,∞].
K Number of group focuses.
μ_{pq} Membership function of information focuses to group focuses.
d_{pq} The Euclidean distance between p th information focuses and q th group focuses.

The main FCM objective function is to minimize.

$$G(u,v) = \sum_{p=1}^{n} \sum_{q=1}^{k} (\mu_{p,q})^m \left|\left| \mathrm{X}_p - Vq \right|\right|^2 \tag{2}$$

where $||Xp - Vq||^2$ is the Euclidean distance between pth information focuses and qth group focuses.

2.3 EM ALGORITHM

Expectation Maximization (EM) is one of the most common algorithms used for estimating the density of data points. By the way of uncontrolled settings and, algorithms rely on searching for files Including the approximate maximum possibility of the parameter values with data replication dependent on some latent variables, the EM algorithm performs an alternating loop of Expectation (E) and Maximization (M) iterations until results converge. Procedure E calculates the expectation of expectation. It is possible by combining the latent variable as if it were observed, and the maximization procedure (M), where the maximum probability estimate of the parameter is calculated by maximizing the expected probability encountered in step E. Finally, the parameters found in step M are used to initiate another step E, and the process repeats until they converge. mathematically For a given training dataset ($x_{(1)}$, $x_{(i2, .. x (m)}$) and model $p(x, z)$, where z is the passive variable, as follow [9]:

E Step for each i:

$$Q\left(z^{(i)}\right) := p(z^{(i)}|x(i);\theta) \tag{3}$$

M Step for all z:

$$\theta := \arg \frac{max}{\theta} \sum_i \sum_{z^{(i)}} Q_{\ i}\left(z^{(i)}\right) \log \frac{p(x^{(i)}, z^{(i)}; \theta}{Q_i(z^{(i)})} \tag{4}$$

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

where Q_i is the posterior distribution of z $^{(i)}$'s given the x $^{(i)}$'s

2.4 2D ENTROPY SEGMENTATION

The 2D entropy method for segmentation of images as a result of this, many application examples have proven that the efficiency of the max 2- dimensional entropy method is better than the 1-D max entropy method. As follows:

$$g(x,y) = \frac{1}{9}\sum_{i=-1}^{1}\sum_{j=-1}^{1} f(x+i, y+j) \qquad (5)$$

Where f (x, y) is the gray value of the pixel at the point (x, y)

The pixel' s gray value f (x, y), and the average of its neighborhood g (x, y) [5], are utilized to construct a 2D histogram.

$$g(i,j) = \frac{1}{M \times N}\{n_{ij} \mid f(x,y) = i, g(x,y) = j; i,j \in (0, L-1)\} \qquad (6)$$

Where $M \times N$ indicates the size of test image, n_{ij} represents the pixel number of which the gray value is i, and the average gray value in the neighborhood is j. p (i, j) indicates the 2D histogram function of test image. L is the maximum gray values of the test images. In our test images, L is 255.

3. PROPOSED METHOD

The aim of this research was to compare a method of using machine learning to extract the edges of contour images for the detection of Covid-19 pneumonia infection using a total of 14 images with a size of 512x512 pixels. Clustering is also known as unsupervised classification technique. The classification name is not maintained because the algorithm automatically classifies the objects based on user-defined criteria. The presented procedure is shown in Figure 1. From the proposed method, we use 3-step techniques, Hybrid fuzzy C-means and K-means, Expectation-maximization algorithm (EM), and 2D Entropy segmentation, to isolate the infected area with precision. To associate with Ground-Truth-Image (GTI) to calculate from a measure. This article presents an approach to compare in terms of accuracy. The block diagram presented is as shown.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

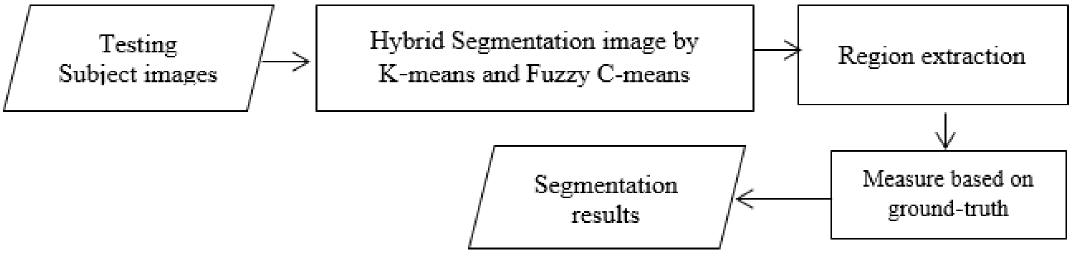


FIGURE. 1 The flow chart proposed method

The steps of segmentation method are as follows:

- Prepare infected images and was not infected with Covid- 19 in order to separate the shape of the image.
- Separate pictures by steps using combining fuzzy C-means and K- means algorithm segmentation of the data provided with free size pixels.
- Take the data obtained from the such steps to separate the region and measure the function by accuracy, precision, and F- measure parameters and ground-truth.
- Finally, results from the division of the region all three methods of the images structure.

4. RESULTS AND DISCUSSION

Several studies published recently indicated that Covid-19 is GGO, or CT lesions are often shown [3][10]Therefore, the search for abnormal areas such as GGO or lesions in CT imaging is essential to diagnosing Covid-19 for radiologists. It is also important to note that [10] many patients do not show abnormal radiology results. We have therefore researched visual images data for both Covid-19 and non-Covid-19 infection. Therefore, patients should not be overlooked in the Covid-19 screening and diagnosis process. The performance of the three methods presented in this section were compared, hybrid fuzzy C-means and K-means, Expectation-maximization algorithm (EM), and 2D Entropy segmentation. We performed all simulation in the MATLAB 2017a. All experiments were run on a 64 bit operating system with a CPU i5-3.3GHz PC with 16GB RAM, by using 14 images [11]in the experiment which have different sizes. For all the methods used on the image database with the previously mentioned of each parameter. The results confirmed that the proposed method provided the accuracy, F-measure, and precision of the infection rate for evaluation. A comparative analysis was performed in the case of Covid-19 infectious and non-infectious imaging to test the rigidity of the proposed approach for contour extraction. However, the visual margins some of the methods fail in the data image files as shown in Table 1.

4.1 QUALITATIVE COMPARISON

The hybrid fuzzy C-means and K-means elimination algorithm that we offer can work well for a variety of segmentation as region extraction images**.** In our experiments, as whole experimental methods able to get good results by Covid**-**19,

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

non-Covid-19, and chest X-rays images. Figure 2-4 show the region extraction images and the depth maps. In Figure 2, our result is comparable to the proposed method, and Figure 3-4 show that our approach outperforms. Which the experiment is equivalent to segmentation method of the images. The better boundary means that the segmentation of the three images is a perfect match, obtain the best matched images compared to the other methods.

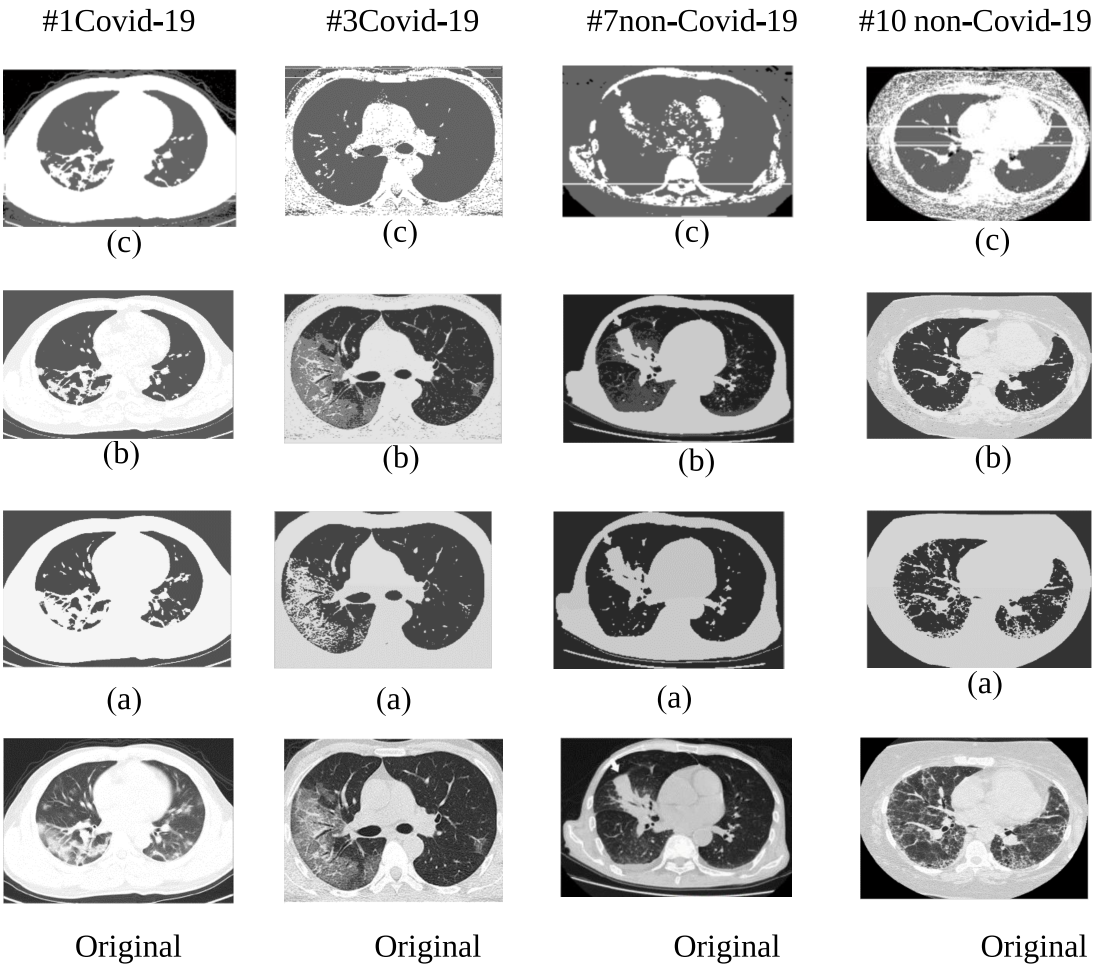
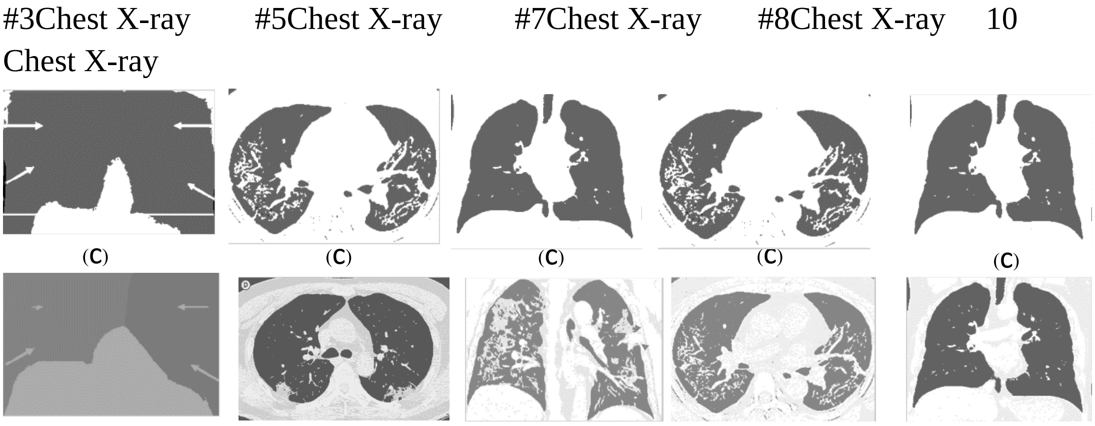


FIGURE 2. Qualitative result region extraction of Covid-19and non-Covid-19image; (a) Hybrid fuzzy C-means and K-means; (b) EM algorithm; (c) 2D Entropy segmentation



ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

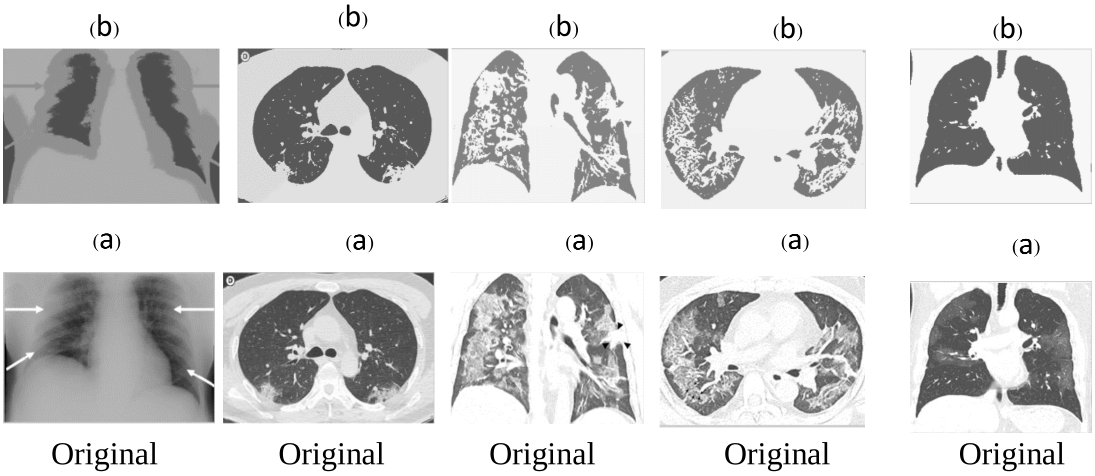


FIGURE 3. Qualitative result region extraction of Chest X-rays image; (a) Hybrid fuzzy C-means and K-means; (b) EM algorithm; (c) 2D Entropy segmentation.

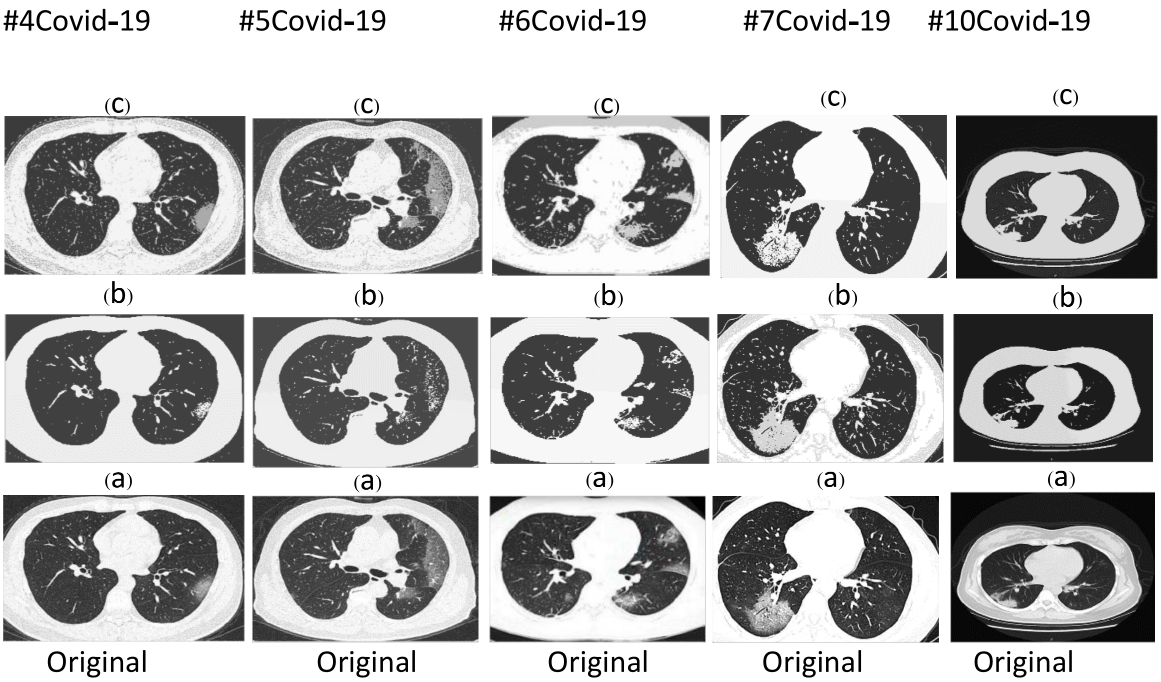


FIGURE 4. Qualitative result region extraction of Covid-19image; (a) Hybrid fuzzy C-means and K-means; (b) EM algorithm; (c) 2D Entropy segmentation

4.1 EXPERIMENT: QUANTITATIVE EVALUTION

The results in some binary sections processed on the ten sample images of the database CT image are shown in Figure 3 as shown by the results of hybrid fuzzy C-means and K-means, Expectation-maximization algorithm (EM), and 2D Entropy segmentation results in a clean binary image. However, the hybrid fuzzy C-means and K-means effect is more powerful and can effectively. produce higher image quality from dark input images shadows. In the qualitative assessment, as shown in the figure, hybrid fuzzy C-means and K-means is smooth to maintain boundaries. The hybrid fuzzy C-means and K-means, Expectation-maximization algorithm (EM), and 2D Entropy segmentation, respectively, successfully isolated the infection area that served as the foreground from the background.

Table 1 summarizes the results of the Covid-19 non-Covid, and Chest X-ray all 14 images dataset boundary segmentation method after comparing it with in terms of accuracy (ACC), F-measure, and precision. It can be seen that the efficiency of the hybrid fuzzy C-means and K-means algorithm presented for comparison is also possible. The table reveal that hybrid executes considerably better than all of other techniques in terms of ACC, iteration and time and is as sound as the better than method with respect to precision and F-measure. The images confirm that all the ground truth comparisons are recommended. Therefore, as shown in the table, hybrid obtains the first rank in terms of ACC, precision, and F-measure. Compared with other in terms, HYBRID obtains the highest score in terms of ACC and follow by precision and F-measure methods that come in second and third after the method.

From the table below the proposed method to segmentation of the Covid-19 data set achieved significant accuracy using a combination HYBRID of methods, with higher accuracy, when using the saliency-based region detection (SRIS) [13] algorithm was compared at 97.85%. The two methods have an accuracy difference of 1.87%, respectively.

TABLE 1.
The comparison of parameter values segmentation methods in dataset image

Dataset Images	Hybrid (Proposed)			EM			2D Entropy			SRIS		
	ACC	Precision	F-measure	ACC	Precision	F-measure	ACC	Precision	F-measure	ACC	Precision	F-measure
#1 Covid-19	0.97	0.98	0.92	0.95	0.97	0.83	0.67	0.64	0.78	0.97	0.98	0.97
#3 Covid-19	0.99	0.99	0.98	0.99	0.89	0.87	0.41	0.41	0.58	0.98	0.99	0.98
#7 non-Covid19	0.98	0.98	0.99	0.98	0.98	0.99	0.65	0.65	0.79	0.98	0.98	0.99
#10 non-Covid19	0.86	0.86	0.93	0.83	0.84	0.92	0.60	0.53	0.69	0.96	0.86	0.99
#4 Covid-19	0.99	0.99	0.99	0.92	0.92	0.96	0.45	0.45	0.62	0.98	0.99	0.98
#5 Covid-19	0.99	0.99	0.99	0.98	0.99	0.99	0.55	0.52	0.68	0.98	0.99	0.99
#6 Covid-19	0.99	0.99	0.99	0.99	0.99	0.99	0.58	0.56	0.72	0.98	0.97	0.98
#7 Covid-19	0.99	0.99	0.99	0.93	0.93	0.96	0.39	0.20	0.33	0.98	0.99	0.98
#10 Covid-19	0.99	0.99	0.99	0.99	0.99	0.99	0.45	0.32	0.49	0.98	0.99	0.98
#3 Chest X-ray	0.99	0.99	0.99	0.99	0.99	0.99	0.18	0.18	0.30	0.98	0.98	0.98
#5 Chest X-ray	0.99	0.99	0.99	0.98	0.98	0.98	0.53	0.53	0.69	0.98	0.99	0.99

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

#7 Chest X-ray	0.99	0.99	0.99	0.98	0.98	0.98	0.65	0.65	0.79	0.98	0.99	0.99
#8 Chest X-ray	0.99	0.99	0.99	0.98	0.98	0.98	0.66	0.66	0.79	0.98	0.99	0.98
#10 Chest X-ray	0.99	0.99	0.99	0.99	0.99	0.99	0.56	0.56	0.72	0.98	0.99	0.99

In order to evaluate the effectiveness of the proposed algorithms, synthetic reference images are used. Figure 5 (a) is #7Covid-19, the original image before segmentation in the presented method. Figure 5 (b) is an image that results from clustering segmentation in the presented way. The test multiple images, which were created with different ACC. The histograms of the reference and the above test images are shown in Figure 5 (c) and (d), respectively. The corrected histogram can be seen in Figure 5 (d). The approximate histogram has two distinct peaks that correct the ambiguity when estimating a histogram. However, the histogram, which has reduced the noise in the image when compared to other techniques, has achieved good results from the algorithm presented.

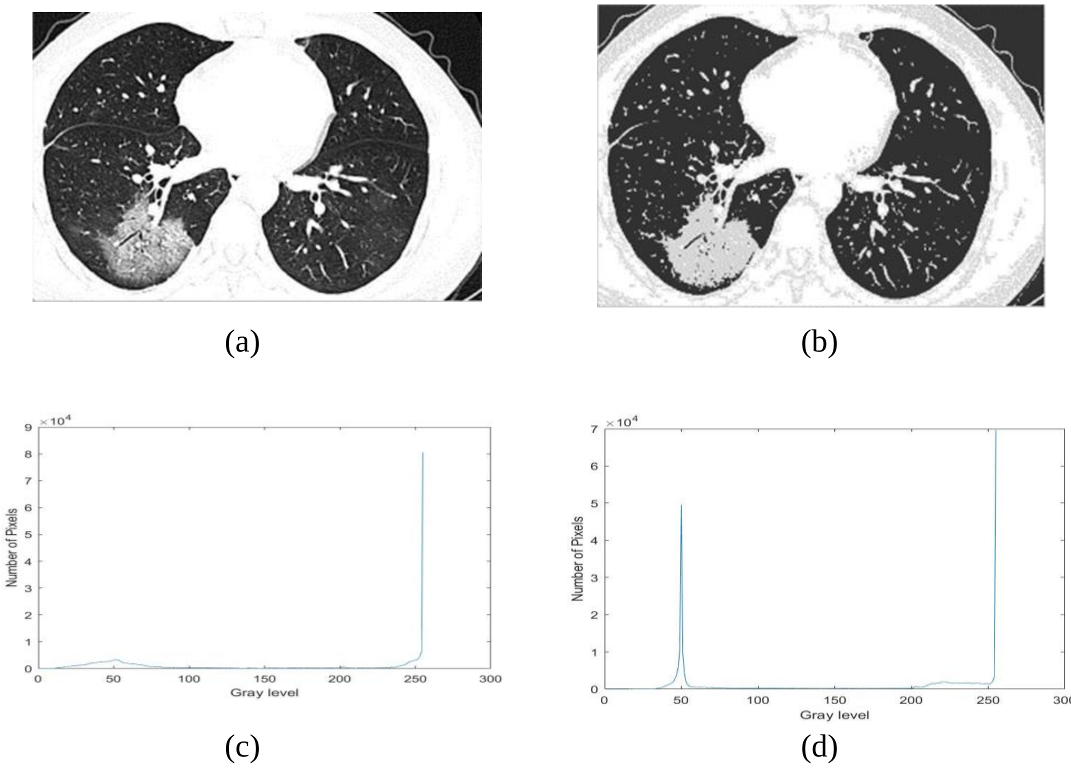


FIGURE 5. Performance evaluation of the proposed method
(a) Image #7Covid-19; (b) Image #7Covid-19 quantized on the segmented into by the hybrid algorithm; (c) Its intensity histogram. Observe that this histogram presents strong oscillations; (d) Best mixture of after segmentation image for the histogram

5. CONCLUSION

This study provides preliminary data for automated diagnosis of Covid-19 infection using AI, which is clinically difficult to distinguish between visual boundaries. Due to the rapid occurrence of infection of the disease. In this work, Active Contour was used with a lung tomography CT scan to automatically isolate the infected part. This research tested methods of hybrid, EM algorithm, and 2D Entropy segmentation. The study found that hybrid segmentation algorithms deliver satisfying and accurate results in retaining the details of the main image better than other techniques, compared to the algorithms. We offer, with two main advantages: 1) the image accuracy (ACC) of the image is effective compared to other methods, and 2) The F-measure and precision value of the image was higher than the other techniques, respectively. From comparative studies and quantitative assessments, the efficacy of the hybrid fuzzy C-means and K-means method was demonstrated. In addition, the algorithms we present work strictly on image histograms, thus resulting in the final segmentation of the image quality without affecting the image distractions. The active contour shape of the image can be clearly distinguished to aid in the diagnosis of Covid-19, both X-ray images, and CT scans. We have demonstrated that all three study methods for extracting the visual active contour of a dataset can combine AI with medical images to separate the visual boundaries. The use of medical imaging in conjunction with AI can compensate for the limitations of medical resources, as well as support the diagnosis and prognosis of Covid-19 disease. However, this detection process needs to be tested in real conditions, so future research should also segmentation other oversampling techniques, such as Statistical Region Merging (SRM) , user interactive segmentation, etc..

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] J. Bullock, A. Luccioni, K. H. Pham, C. S. N. Lam, and M. Luengo-Oroz, “Mapping the landscape of artificial intelligence applications against COVID-19,” *Journal of Artificial Intelligence Research*, vol. 69, 2020, doi: 10.1613/JAIR.1.12162.

[2] C. Casey, R. Rebecca, “Debate Flares Over Using AI to Detect Covid-19in Lung Scans,” *https://www.statnews.com/2020/03/30/debate-over-artificial-intelligence-to-detect-covid-19-in-lung-scans/*, 27 November 2020.

[3] M. Chung *et al.*, “CT imaging features of 2019 novel coronavirus (2019-NCoV),” *Radiology*, vol. 295, no. 1, 2020, doi: 10.1148/radiol.2020200230.

[4] R. Rosebrock , “Detecting COVID-19 in X-ray Images with Keras, TensorFlow, and Deep Learning,” *https://pyimagesearch.com/2020/03/16/detecting-covid-19-in-x-ray-images-with-keras-tensorflow-and-deep-learning/,* March 16, 2020**.**

[5] M. Dhieb, S. Masmoudi, M. Ben Messaoud, M. Frikha, and F. Ben Arfia, “2-D entropy image segmentation on thresholding based on particle swarm optimization (PSO),” in *2014 1st International Conference on Advanced*

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

Technologies for Signal and Image Processing, ATSIP 2014, 2014. doi: 10.1109/ATSIP.2014.6834594.

[6] A. A. Farid, G. I. Selim, and H. A. A. Khater, “A Novel Approach of CT Images Feature Analysis and Prediction to Screen for Corona Virus Disease (COVID-19),” *Int J Sci Eng Res*, vol. 11, no. 03, 2020, doi: 10.14299/ijser.2020.03.02.

[7] W. Guan *et al.*, “Clinical Characteristics of Coronavirus Disease 2019 in China,” *New England Journal of Medicine*, vol. 382, no. 18, 2020, doi: 10.1056/nejmoa2002032.

[8] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, “An efficient k-means clustering algorithms: Analysis and implementation,” *IEEE Trans Pattern Anal Mach Intell*, vol. 24, no. 7, 2002, doi: 10.1109/TPAMI.2002.1017616.

[9] S. Tatiraju et al., “Image Segmentation using k-means clustering , EM and Normalized Cuts,”*Department of EECS*, *1*(1-7), 2. 2008.

[10] H. Shi *et al.*, “Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study,” *Lancet Infect Dis*, vol. 20, no. 4, 2020, doi: 10.1016/S1473-3099(20)30086-4.

[11] 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,” *medRxiv*, 2020.

[12] G. S. Raghtate and S. S. Salankar, “Modified Fuzzy C Means with Optimized Ant Colony Algorithm for Image Segmentation,” *2015 International Conference on Computational Intelligence and Communication Networks (CICN)*, Jabalpur, India, 2015, pp. 1283-1288, doi: 10.1109/CICN.2015.246.

[13] A. Joshi, M. S. Khan, S. Soomro, A. Niaz, B. S. Han and K. N. Choi, "SRIS: Saliency-Based Region Detection and Image Segmentation of COVID-19 Infected Cases," *in IEEE Access, vol. 8*, pp. 190487-190503, 2020, doi: 10.1109/ACCESS.2020.3032288