Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Knowledge-Based Systems 231 (2021) 107432

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Ambiguous D-means fusion clustering algorithm based on ambiguous set theory: Special application in clustering of CT scan images of COVID-19

Pritpal Singh [a,*,1], Surya Sekhar Bose [b,1]

[a] Institute of Theoretical Physics, Jagiellonian University, ul.Ł,ojasiewicza 11, Kraków 30-348, Poland
[b] Department of Mathematics, Madras Institute of Technology, MIT Rd, Radha Nagar, Chromepet, Chennai, Tamil Nadu 600044, India

ARTICLE INFO

Article history:
Received 3 April 2021
Received in revised form 17 August 2021
Accepted 20 August 2021
Available online 26 August 2021

Keywords:
Coronavirus disease 2019 (COVID-19)
Computed tomography (CT)
Ambiguous set theory
Ambiguous D-means fusion clustering algorithm (ADMFCA)
Clustering

ABSTRACT

Coronavirus Disease 2019 (COVID-19) has been considered one of the most critical diseases of the 21st century. Only early detection can aid in the prevention of personal transmission of the disease. Recent scientific research reports indicate that computed tomography (CT) images of COVID-19 patients exhibit acute infections and lung abnormalities. However, analyzing these CT scan images is very difficult because of the presence of noise and low-resolution. Therefore, this study suggests the development of a new early detection method to detect abnormalities in chest CT scan images of COVID-19 patients. By this motivation, a novel image clustering algorithm, called *ambiguous D-means fusion clustering algorithm (ADMFCA)*, is introduced in this study. This algorithm is based on the newly proposed ambiguous set theory and associated concepts. The ambiguous set is used in the proposed technique to characterize the ambiguity associated with grayscale values of pixels as true, false, true-ambiguous and false-ambiguous. The proposed algorithm performs the clustering operation on the CT scan images based on the entropies of different grayscale values. Finally, a final outcome image is obtained from the clustered images by image fusion operation. The experiment is carried out on 40 different CT scan images of COVID-19 patients. The clustered images obtained by the proposed algorithm are compared to five well-known clustering methods. The comparative study based on statistical metrics shows that the proposed ADMFCA is more efficient than the five existing clustering methods.

1. Introduction

In December 2019, the first Coronavirus Disease 2019 (COVID-19) outbreak has been discovered in Wuhan, China [1]. This disease is caused by a novel virus, called *Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)* [2]. According to the World Health Organization (WHO) [3], this disease has spread around the world and reported 200,840,180 confirmed cases along with 4,265,903 deaths by the end of 6th August, 2021. Now, this disease has seriously affected the health, economic and social systems of advanced and emerging countries [4]. Therefore, the WHO has declared this disease as one of the deadliest pandemics of all time. However, the pandemic situation caused by this disease has yet to be taken seriously [5]. Due to this reason, several countries, including the United States, Brazil, India and Italy, have been seriously affected by this virus [6,7]. As a result, several research groups are collaborating in the development of strategies, vaccines and new approaches to address the pandemic [8].

COVID-19 is associated with serious respiratory symptoms that cause death in most cases [9]. This disease is observed as pneumonia, the infection of which is very contagious from one to the other [10]. Digital imaging techniques, such as X-ray [11] and computed tomography (CT) [12] have contributed significantly to the diagnosis of this disease. Computer scientists have been actively involved in developing methods for analyzing these images based on machine learning. To develop such methods, they mainly use the convolutional neural network (CNN), which is a form of deep neural network. Some of the CNN based methods devised by the researchers are summarized in Table 1.

The methodologies [11–19] discussed in Table 1 focus primarily on classifying X-ray or CT scan images in terms of infection and non-infection. Moreover, such approaches cannot identify the infection in the lungs. However, these studies indicate that X-ray and CT scan images of COVID-19 can be useful in determining the severity of lung infection. But, there are some inherent drawbacks

* Corresponding author.
E-mail addresses: drpritpalsingh82@gmail.com (P. Singh), sbose499@yahoo.com (S.S. Bose).
[1] All authors contributed equally to this article.

https://doi.org/10.1016/j.knosys.2021.107432
0950-7051/

Table 1
Machine-learning based methods for X-ray and CT scan image analysis of COVID-19 patients.

Article	Method developed	Image used	Objective
Ozturk et al. [13]	DarkNet model with 17 convolutional layers	X-ray	Binary (COVID-19 and No-finding) and multiclass (COVID-19, No-finding and Pneumonia) classification
Vaid et al. [14]	CNN with transfer learning	X-ray	Structural abnormalities and disease classification
Apostolopoulos and Mpesiana [11]	CNN with transfer learning	X-ray	Classification of abnormal X-ray images
Toraman et al. [15]	CapsNet model with capsule networks	X-ray	Binary (COVID-19 and No-finding) and multiclass (COVID-19, No-finding and Pneumonia) classification
Nour et al. [16]	CNN with k-nearest neighbor, support vector machine (SVM) and decision tree	X-ray	Classification of infection
Ardakani et al. [12]	CNN with machine learning	CT	Classification of positive COVID-19 cases
Kang et al. [17]	Multi-view representation learning	CT	Extraction of multiple features with different views
Wang et al. [18]	Weakly-supervised deep learning	CT	COVID-19 classification and lesion localization
Varela-Santos and Melin [19]	Feed-forward with CNN	X-ray	Classification of positive COVID-19 cases

associated with X-ray and CT scan images of COVID-19, which are described below:

Drawback 1: The grayscale values in these images indicate some similarities between COVID-19 infection and other forms of pneumonia, which complicates decision-making.
Drawback 2: Large dimensions of these images increase the computational complexity during their processing.
Drawback 3: These images are available in multiple slices, but only a few of them indicate infection in the lungs.
Drawback 4: These images available with low-resolution and contain dense dark pixels, making it difficult to distinguish the infected lung regions.

The drawbacks mentioned in (1)–(4) can be overcome by various image processing techniques that help in detecting abnormalities in X-ray and CT scan images. Data clustering is one such method, whose main objective is to make a group of similar and dissimilar objects based on certain distance criteria [2,20,21]. However, clustering an image is one of the most difficult tasks in image processing, because in most cases it is difficult to form clusters of pixels by distinguishing similar grayscale values from dissimilar grayscale values. In the case of an image K, clustering is a means of partitioning it into non-overlapping regions of clusters $K_1, K_2, \ldots, K_n$, such that:

$$\bigcap_{i=1}^{n} K_i = \emptyset \tag{1}$$

$$\bigcup_{i=1}^{n} K_i = K \tag{2}$$

Based on the developments in clustering algorithms, they can be classified into two categories: (a) crisp clustering and (b) soft clustering [22]. Crisp and soft clusterings differ widely in their approach to assigning group members. In crisp clustering, the grayscale values of an image belong exclusively to a fixed group, i.e., their assignment is completely binary. That is, they completely belong to a cluster (true) or not (false). On the other hand, a soft clustering approach allows grayscale values to share a degree of membership in several groups. One of the most popular algorithms in the category of crisp clustering is K-means clustering (KMC) algorithm [23,24]. Another example in this category of clustering algorithm is multi-view information-theoretic co-clustering (MV-ITCC), which is based on information-theoretic co-clustering approach [25]. According to recent advances in the development of clustering algorithms, soft clustering algorithms can be divided into fuzzy set [26] based clustering, intuitionistic fuzzy set (IFS) [27] based clustering and neutrosophic set [28] based clustering. One of the most frequently used methods, developed using fuzzy sets, is fuzzy C-means (FCM) [29,30]. Subsequently, researchers introduce several variants of the FCM algorithm. For example, Chen et al. [31] propose multiple-kernel FCM (MKFCM) algorithm using the composite kernel concept. Ji et al. [32] propose the weighted image patch-based FCM (WIPFCM) algorithm, which considers the spatial information of pixels during image clustering. Wang et al. [33] incorporate information-theoretic concept into the FCM algorithm to improve its performance. To deal with the noise in grayscale images, Zhao et al. [34] propose a new version of the FCM algorithm, called generalized fuzzy c-means clustering (GFCM) algorithm. Chaira [35] introduce a novel intuitionistic FCM (IFCM) clustering algorithm by adopting the concept of IFS. Verma et al. [36] propose an improved intuitionistic FCM (IIFCM) clustering algorithm by including the local spatial information during the clustering. To overcome the drawback of FCM algorithm, Deng et al. [37] propose a new transfer prototype-based fuzzy clustering method. Singh [38] introduces neutrosophic-entropy based clustering algorithm (NEBCA) for performing clustering operation on magnetic resonance imaging (MRI) of Parkinson's disease.

The above discussed clustering methods [29,31–36,38] are based on the concepts of fuzzy set, IFS and neutrosophic set, and they are able to deal with inherent uncertainties of grayscale values, but they have certain limitations, such as:

(a) The clustering methods based on fuzzy set (i.e., FCM [29] and its variants [31–34]) support the representation of uncertainties of grayscale values only by considering a true degree of membership. These methods define the true degree of membership of a grayscale value in [0, 1].
(b) IFS based clustering methods (i.e., IFCM [35] and IIFCM 1u [36]) model the uncertainties of grayscale values with respect to hesitancy. These methods define the hesitancy in terms of true and false degree of memberships.
(c) The clustering method based on neutrosophic sets (i.e. NEBCA [38]) defines the uncertainties of grayscale values with three degree of memberships, called true, false and indeterministic [39,40].

The above discussion indicates the need of a more robust clustering method that can deal with the drawbacks mentioned in (1)–(4) as well as inherent uncertainties in the CT scan images of COVID-19 patients precisely. Recently, Singh et al. [41] proposed a novel theory to deal with uncertainties, called *ambiguous set theory*. Ambiguous set theory is better than fuzzy set, IFS and neutrosophic set in terms of its capability of modeling the inherent ambiguities very effectively with four degree of memberships, as true, false, true-ambiguous and false-ambiguous. Singh et al. [41] show the application of this theory in segmenting MRI of human brain. However, the application of this theory has not been extended to the analysis of other types of medical images, such as X-rays and CT scan images. A recent study by Singh and Bose [2] shows that the clustering approach is useful to identify infected regions in CT scan images of COVID-19 patients. With

this motivation and considering the severity of COVID-19, this study extends the application of this theory in the development of a clustering algorithm that could be helpful in the analysis of CT scan images of COVID-19 patients. Hence, the main contributions of this study are five-fold as:

(1) **First**, we first introduce the notion of ambiguous set theory with a mathematical representation.
(2) **Second**, this study presents *ambiguous membership functions (AMFs)* to define the true, false, true-ambiguous and false-ambiguous memberships of an event using ambiguous set.
(3) **Third**, to quantify the inherent ambiguities associated with the four degree of memberships of ambiguous set (i.e., true, false, true-ambiguous and false-ambiguous), four different entropy formulas are proposed.
(4) **Fourth**, a new image clustering algorithm is proposed in this study using four memberships of the ambiguous set and their respective entropies, called *ambiguous D-means fusion clustering algorithm (ADMFCA)*. The proposed algorithm generates four different clustered images by performing clustering of grayscale values of CT scan images. For clustering, a distance metric, called *ambiguous-entropy distance function* is introduced, whose main objective is to assign the grayscale values to the different clusters based on the minimum distance criterion. Finally, in order to incorporate all the features of the four different clustered images and obtain a final image, they are aggregated with the help of image fusion [40]. This final image is referred to as *final clustered image (FCI)*. The main reason for using image fusion operation in this algorithm is to integrate the best features into the resultant image [42]. A major problem with most crisp clustering algorithms (as discussed above) is that data points are often assigned to incorrect clusters by ignoring their correlation with some other subset of data points in the problem space [43]. Therefore, the main goal of ADMFCA is to cluster the grayscale values of CT scan images of COVID-19 patients in such a way that correlated features of the infected regions can be easily identified and formed clusters with highly correlated features.
(5) **Fifth**, in support of ambiguous set theory, various definitions, set-theoretical operations, theorems and properties are discussed.

The proposed ADMFCA is validated with chest CT scan images of COVID-19 patients with the corresponding ground truth. The performance of the proposed ADMFCA is compared with five existing clustering algorithms, including KMC [23], FCM [29], GFCM [34], IIFCM [36] and NEBCA [38]. The performance of the proposed ADMFCA and existing algorithms [23,29,34,36,38] is compared using statistical metrics, such as mean squared error (MSE), peak signal-to-noise ratio (PSNR), Dice similarity coefficient (DSC), Jaccard similarity coefficient (JSC) and correlation coefficient (CC). These statistical analyses show the effectiveness of the proposed ADMFCA over the selected clustering algorithms [23,29,34,36,38].

The remainder of this article is organized as follows. Background for the study is presented in Section 2. Section 3 introduces the proposed ambiguous set theory. The proposed ADMFCA for image segmentation is presented in Section 4. Various properties of ambiguous-entropy distance function are discussed in Section 5. Experimental results are described in Section 6. Finally, conclusions and future directions are presented in Section 7.

2. Background for the study

This section presents an overview of the fuzzy set, intuitionistic fuzzy set and neutrosophic set.

2.1. Fuzzy Set

The fuzzy set $\tilde{F}$ for any event $Q_i (i = 1, 2, \ldots, n)$ in the discrete and finite universe of discourse $\mathbb{S}$ can be described as [26]:

$$\tilde{F} = \left\{ \frac{\mu_{\tilde{F}}(Q_1)}{Q_1} + \frac{\mu_{\tilde{F}}(Q_2)}{Q_2} + \cdots + \frac{\mu_{\tilde{F}}(Q_n)}{Q_n} \right\} = \left\{ \sum_{i=1}^{n} \frac{\mu_{\tilde{F}}(Q_i)}{Q_i} \right\} \tag{3}$$

For the continuous and infinite universe of discourse $\mathbb{S}$, the fuzzy set $\tilde{F}$ for any event $Q_i (i = 1, 2, \ldots, n)$ can be defined as:

$$\tilde{F} = \left\{ \frac{\mu_{\tilde{F}}(Q_i)}{Q_i} \right\} \tag{4}$$

In Eqs. (3)–(4), $\mu_{\tilde{F}}(Q_i)$ represents the degree of membership for each $Q_i \in \mathbb{S}$. In this theory, the degree of membership of each Q_i, i.e., $\mu_{\tilde{F}}(Q_i)$ always belongs to the range [0, 1]. In Eqs. (3)–(4), the horizontal bar represents a delimiter. The numerator of each term reflects the degree of membership of each Q_i in the fuzzy set $\tilde{F}$. In Eq. (3), the summation symbol "+" represents the aggregation of each Q_i, called an aggregation operator. In Eq. (4), the integral sign indicates a continuous function-theoretic aggregation operator for continuous events [26].

2.2. Intuitionistic fuzzy set

Atanassov [27] proposed the concept of intuitionistic fuzzy set (IFS). It helps to represent the hesitancy involved in each $Q_i \in \mathbb{S}$ with respect to two membership functions, called *degree of membership* and *degree of non-membership*.

For a fixed crisp set $\mathcal{C}$, an IFS is denoted as $\mathcal{C}^I(Q_i)$, and defined as:

$$\mathcal{C}^I(Q_i) = \langle Q_i, \mu_{\tilde{F}}(Q_i), \theta_{\mathcal{C}}(Q_i) \rangle \tag{5}$$

Here, $\mu_{\tilde{F}}(Q_i) \in [0, 1]$ denotes the membership of Q_i, and $\theta_{\mathcal{C}}(Q_i) \in [0, 1]$ denotes the non-membership of Q_i in the $\mathcal{C}$.

In the IFS, the boundaries of membership and non-membership of Q_i must satisfy the following condition as:

$$0 \leq \mu_{\tilde{F}}(Q_i) + \theta_{\mathcal{C}}(Q_i) \leq 1 \tag{6}$$

The non-membership of Q_i is defined in terms of the fuzzy set as: $\theta_{\mathcal{C}}(Q_i) = 1 - \mu_{\tilde{F}}(Q_i)$. However, this leads to a loss of information when Q_i changes its state from one to another. Therefore, this loss of information is shown using IFS with respect to the membership and non-membership functions as:

$$\triangle(Q_i) = 1 - (\mu_{\tilde{F}}(Q_i) + \theta_{\mathcal{C}}(Q_i)) \tag{7}$$

Here, $\triangle(Q_i)$ indicates the degree of loss. It can be defined only for the IFS, because here $\triangle(Q_i) \neq 0$. However, for ordinary fuzzy set, $\theta_{\mathcal{C}}(Q_i) = 1 - \mu_{\tilde{F}}(Q_i)$, so $\triangle(Q_i) = 0$.

For Q_i, an IFS can also be represented in terms of loss as:

$$\mathcal{C}^I(Q_i) = \langle \mu_{\tilde{F}}(Q_i), \theta_{\mathcal{C}}(Q_i), \triangle(Q_i) \rangle \tag{8}$$

2.3. Neutrosophic set

Smarandache [28] introduced the neutrosophic set theory that can model the inherent uncertainty of each $Q_i \in \mathbb{S}$ in terms of three degree of memberships, namely truth (N_T), indeterminacy (N_I) and falsity (N_F).

Assume that a neutrosophic set $\mathbb{N}$ is defined for the Q_i on the universe of discourse $\mathbb{S}$. Here, N_T, N_I and N_F for the $Q_i \in \mathbb{S}$ can be expressed as: $N_T, N_I, N_F : \mathbb{S} \rightarrow]^-0, 1^+[$, $Q_i \equiv$

$Q_i(N_T(Q_i), N_I(Q_i), N_F(Q_i)) \in \mathbb{N}$, and $^-0 \leq N_T(Q_i)+N_I(Q_i)+N_F(Q_i) \leq 3^+$.

A neutrosophic set $\mathbb{N}$ can be expressed as a *single-valued neutrosophic set (SVNS)* [44]. The SVNS can be expressed for discrete and finite $\mathbb{S}$ as:

$$\begin{aligned}\mathbb{N} &= \frac{\langle N_T(Q_1), N_I(Q_1), N_F(Q_1)\rangle}{Q_1} + \frac{\langle N_T(Q_2), N_I(Q_2), N_F(Q_2)\rangle}{Q_2} + \cdots \\ &\quad + \frac{\langle N_T(Q_n), N_I(Q_n), N_F(Q_n)\rangle}{Q_n} \\ &= \left\{ \sum_{i=1}^{n} \frac{\langle N_T(Q_i), N_I(Q_i), N_F(Q_i)\rangle}{Q_i} \right\}, \ \forall Q_i \in \mathbb{S} \end{aligned} \tag{9}$$

3. The proposed ambiguous set theory

This section presents the philosophy of the ambiguous set, its various definitions followed by related properties.

3.1. Philosophy of ambiguous set

According to the Oxford Dictionary, the word *ambiguous* is an adjective that means "open to more than one interpretation". Some information reflects different interpretations, which leads to ambiguity and incompleteness. As a result of this problem, decision-making becomes difficult in most of the cases.

Consider this proposition: "Mr. X is lying". This statement is either true or false; however, in view of the proposition, the human cognitive process can have the following perceptions:

Perception 1: Mr. X is lying.

Perception 2: Mr. X is not lying.

Perceptions 1 and 2 can have definite true and false values, respectively. However, there is often uncertainty and incompleteness between truth and falsity. According to human cognitive processes, both uncertainty and incompleteness can be interpreted in terms of the following perceptions:

Perception 3: Mr. X is a little lying.

Perception 4: It's a little false that Mr. X is lying.

Perceptions 3 and 4 cannot have distinct true and false values, respectively. Perception 3 is very close to Perception 1, and it inherits the ambiguity from Perception 1, so it can be categorized as having a true-ambiguous value. Similarly, Perception 4 is very close to Perception 2, but it develops ambiguity from Perception 2, so it can be represented with a false-ambiguous value.

The above discussion indicates that any event $\langle\mathbb{E}\rangle$ can be viewed in terms of the following four different perceptions listed in **P1**-**P4** as:

P1: $\langle\mathbb{E}\rangle$ is completely true, i.e., $\langle$True: $\mathbb{E}\rangle$.

P2: $\langle\mathbb{E}\rangle$ is completely false, i.e., $\langle$False: $\mathbb{E}\rangle$.

P3: $\langle\mathbb{E}\rangle$ is a little true, i.e., $\langle$True-ambiguous: $\mathbb{E}\rangle$.

P4: $\langle\mathbb{E}\rangle$ is a little false, i.e., $\langle$False-ambiguous: $\mathbb{E}\rangle$.

By integrating the four different perceptions of an event, a novel theory is proposed, called ambiguous set theory [41]. According to this theory, the initial perceptions of the event $\langle\mathbb{E}\rangle$ are characterized as $\langle$True: $\mathbb{E}\rangle$ and $\langle$False: $\mathbb{E}\rangle$. The ambiguities in $\langle$True: $\mathbb{E}\rangle$ and $\langle$False: $\mathbb{E}\rangle$ are characterized by the perceptions of $\langle$True-ambiguous: $\mathbb{E}\rangle$ and $\langle$False-ambiguous: $\mathbb{E}\rangle$, respectively. The ambiguities in $\langle$True: $\mathbb{E}\rangle$, $\langle$False: $\mathbb{E}\rangle$, $\langle$True-ambiguous: $\mathbb{E}\rangle$ and $\langle$False-ambiguous: $\mathbb{E}\rangle$ are defined with respect to four degree of membership functions, viz., true, false, true-ambiguous and false-ambiguous, respectively. Since the ambiguities of $\langle$True: $\mathbb{E}\rangle$ and $\langle$False: $\mathbb{E}\rangle$ are contradictory, their true and false degree of membership functions are considered complementary to one another. True-ambiguous and false-ambiguous degree of membership functions, on the other hand, are determined by the true and false degree of membership functions, respectively. Thus, the individual values of true, false, true-ambiguous and false-ambiguous, which result from the respective degree of membership functions, indicate the ambiguities of $\langle\mathbb{E}\rangle$ in terms of degree-of-true, degree-of-false, degree-of-ambiguity-in-true and degree-of-ambiguity-in-false, respectively. In this way, this theory addresses the problems associated with ambiguous features of the information in the data.

3.2. Proposed ambiguous set

The inherent uncertainty for any event x in the universe of discourse $\mathbb{S}$ can be defined using ambiguous set theory [41], which represents the uncertainty in terms of four degree of membership functions, namely, true (T), false (F), true-ambiguous (TA) and false-ambiguous (FA). Mathematically, the ambiguous set can be defined as follows.

Definition 1 (*Ambiguous Set [41]*)**.** An *ambiguous set* Å on the universe of discourse $\mathbb{S}$ can be defined based on the four membership functions $T, F, TA, FA : \mathbb{S} \to \ ]^-0, 1^+[$. These membership functions must satisfy the condition $^-0 \leq T(x) + F(x) + TA(x) + FA(x) \leq 2^+$ for all $x \in \mathbb{S}$. Here, T, F, TA and FA are real standard values or non-standard subsets of $]^-0, 1^+[$.

From a philosophical point of view, the ambiguous set takes the value $]^-0, 1^+[$ on real standard or non-standard subsets. Thus, for engineering applications, instead of taking $]^-0, 1^+[$, it is useful to take the interval [0, 1], since it is difficult to use $]^-0, 1^+[$ in real applications such as engineering and science problems.

The AMFs consist of four membership functions, namely, T, F, TA and FA. The generalized form of the ambiguous set is called *single-valued ambiguous set (SVAS)*, if the AMFs are singleton subintervals/subsets of the standard real unit interval [0, 1]. The SVAS can be defined as:

Definition 2 (*SVAS*)**.** An SVAS Å in $\mathbb{S}$ is represented by four membership functions: $T : \mathbb{S} \to [0, 1]$, $F : \mathbb{S} \to [0, 1]$, $TA : \mathbb{S} \to [0, 1]$ and $FA : \mathbb{S} \to [0, 1]$. Such SVAS Å can be designated as:

$$Å = \{\langle x, T(x), F(x), TA(x), FA(x)\rangle | x \in \mathbb{S}\}, \tag{10}$$

with the condition $^-0 \leq T(x)+F(x)+TA(x)+FA(x) \leq 2^+, \forall x \in \mathbb{S}$.

An ambiguous set can be defined for the discrete case as follows.

Definition 3 (*Discrete Ambiguous Set*)**.** An ambiguous set Å for the discrete and finite universe of discourse $\mathbb{S} = \{x_1, x_2, \ldots, x_n\}$ can be represented as:

$$\begin{aligned} Å &= \frac{\langle T(x_1), F(x_1), TA(x_1), FA(x_1)\rangle}{x_1} + \\ &\quad \frac{\langle T(x_2), F(x_2), TA(x_2), FA(x_2)\rangle}{x_2} + \\ &\quad \cdots + \\ &\quad \frac{\langle T(x_n), F(x_n), TA(x_n), FA(x_n)\rangle}{x_n} \\ &= \bigcup_{i=1}^{n} \frac{\langle T(x_i), F(x_i), TA(x_i), FA(x_i)\rangle}{x_i} \end{aligned} \tag{11}$$

In Eq. (11), both symbols "+" and "$\bigcup$" are termed as aggregation operators. For the continuous and infinite $\mathbb{S}$, the ambiguous set Å can be denoted as:

$$Å = \left\{ \frac{\langle T(x), F(x), TA(x), FA(x)\rangle}{x} \right\} \tag{12}$$

The AMFs can be defined as follows for the SVAS.

Definition 4 (*AMFs*)**.** The four degree of membership functions, namely, T, F, TA and FA for a SVAS Å in $\mathbb{S}$ can be mathematically defined as:

$$T(x) = \frac{x - \min(\mathbb{S})}{\max(\mathbb{S}) - \min(\mathbb{S})} \tag{13}$$

$$F(x) = 1 - T(x) \tag{14}$$

$$TA(x) = \frac{T(x)}{T(x) + A_F(x)} \tag{15}$$

$$FA(x) = \frac{F(x)}{F(x) + A_F(x)} \tag{16}$$

In Eqs. (15) and (16), A_F is termed as the *ambiguous distance function*. Mathematically, it can be formulated as:

$$A_F(x) = \sqrt{T(x)^2 + F(x)^2} \tag{17}$$

Example 1. A CT scan image consists of grayscale values within the range $[0, 255]$. Fig. 1(a) and (b) show a CT scan image of a COVID-19 patient and its three different grayscale values at pixel positions P_{15}, P_{17} and P_{112}, which are 215, 228 and 240, respectively. These three grayscale values create the illusion effect as well as other perceptual problems in terms of their individual intensity. The different grayscale intensities also pose the challenge of distinguishing one region from another. Consequently, users cannot confidently use the linguistic terms "dark gray", "gray" and "light gray" to describe these three grayscale values. However, this difficulty can be resolved by using ambiguous sets, where inherent imprecision or approximation of grayscale intensities are expressed by AMFs. In this respect, three different ambiguous sets $Å_1$, $Å_2$ and $Å_3$ can be defined for the grayscale values at pixel positions P_{15}, P_{17} and P_{112} on the universe of discourse $\mathbb{Z} = [0, 255]$ using the AMFs (Eqs. (13)–(16)), respectively, as:

$$Å_1 = \left\{ \frac{\langle T(P_{15}), F(P_{15}), TA(P_{15}), FA(P_{15})\rangle}{P_{15}} \right\} = \left\{ \frac{\langle 0.85, 0.15, 0.50, 0.15\rangle}{215} \right\} \tag{18}$$

$$Å_2 = \left\{ \frac{\langle T(P_{17}), F(P_{17}), TA(P_{17}), FA(P_{17})\rangle}{P_{17}} \right\} = \left\{ \frac{\langle 0.90, 0.10, 0.50, 0.10\rangle}{228} \right\} \tag{19}$$

$$Å_3 = \left\{ \frac{\langle T(P_{112}), F(P_{112}), TA(P_{112}), FA(P_{112})\rangle}{P_{112}} \right\} = \left\{ \frac{\langle 0.95, 0.05, 0.50, 0.05\rangle}{240} \right\} \tag{20}$$

In Eq. (18), $T(P_{15})$, $F(P_{15})$, $TA(P_{15})$ and $FA(P_{15})$ indicate the ambiguousness belonging to the *white pixel*, *non-white pixel*, *ambiguous white pixel* and *ambiguous non-white pixel*, respectively. Similar explanations can be provided for Eqs. (19) and (20). Graphical representations of four degree of memberships of $Å_1$, $Å_2$ and $Å_3$ are shown in Fig. 1(c)–(e), respectively.

In Fig. 1(c)–(e), each of the shaded regions is called an *ambiguous region (AR)*. This AR is extremely useful as it clearly describes the inherent ambiguity measured by AMFs. The two-dimensional regions, as shown in Fig. 1(c)–(e), obtain from the presence of ambiguous features of the grayscale values at pixel positions P_{15}, P_{17} and P_{112}, respectively. Here, the AR provides two valuable information:

1. linguistic description of all uncertainties associated with the effect of AMFs, and
2. the distribution of ambiguity in the two-dimensional plane.

3.3. Related concepts of ambiguous set

Entropy can be used to measure the individual ambiguousness represented by the AMFs, namely, T, F, TA and FA. Such measurements of ambiguousness with respect to T, F, TA and FA are called *true entropy (TE)*, *false entropy (FE)*, *true-ambiguous entropy (TAE)* and *false-ambiguous entropy (FAE)*, respectively. These four entropies can be defined as follows.

Definition 5 (*Measurements of Ambiguousness*)**.** The four different entropies, viz., TE, FE, TAE and FAE of a SVAS Å at $x \in \mathbb{S}$ are denoted as a measure $E_T(Å, x)$, $E_F(Å, x)$, $E_{TA}(Å, x)$ and $E_{FA}(Å, x)$, respectively, where $Å : \{\langle x, T(x), F(x), TA(x), FA(x)\rangle | x \in \mathbb{S}\}$, which can be defined as follows:

$$E_T(Å, x) = -T(x) \cdot \ln(T(x)) \tag{21}$$

$$E_F(Å, x) = -F(x) \cdot \ln(F(x)) \tag{22}$$

$$E_{TA}(Å, x) = -TA(x) \cdot \ln(TA(x)) \tag{23}$$

$$E_{FA}(Å, x) = -FA(x) \cdot \ln(FA(x)) \tag{24}$$

Definition 6 (*Operations on Ambiguous Sets*)**.** Let

$$Å_1 = \{\langle x, T_1(x), F_1(x), TA_1(x), FA_1(x)\rangle | x \in \mathbb{S}\}, \text{ and}$$
$$Å_2 = \{\langle x, T_2(x), F_2(x), TA_2(x), FA_2(x)\rangle | x \in \mathbb{S}\}$$

be two ambiguous sets. Some operations on ambiguous sets are given below:

1. $Å_1 \subseteq Å_2$ if and only if $T_1(x) \leq T_2(x)$, $F_1(x) \geq F_2(x)$, $TA_1(x) \geq TA_2(x)$, and $FA_1(x) \geq FA_2(x)$.
2. $Å_1^c = \{\langle x, T_1^c(x), F_1^c(x), TA_1^c(x), FA_1^c(x)\rangle | x \in \mathbb{S}\}$, where $T_1^c(x) = F_1(x)$, $F_1^c(x) = T_1(x)$, $TA_1^c(x) = 1 - TA_1(x)$, $FA_1^c(x) = 1 - FA_1(x)$.
3. $Å_1 \cap Å_2 = \{\langle x, (T_1(x) \wedge T_2(x)), (F_1(x) \vee F_2(x)), (TA_1(x) \vee TA_2(x)), (FA_1(x) \vee FA_2(a))\rangle | x \in \mathbb{S}\}$.
4. $Å_1 \cup Å_2 = \{\langle x, (T_1(x) \vee T_2(x)), (F_1(x) \wedge F_2(x)), (TA_1(x) \wedge TA_2(x)), (FA_1(x) \wedge FA_2(x))\rangle | x \in \mathbb{S}\}$.

Definition 7 (*Ambiguous Vector and Its Complement*)**.** Let $\Theta = (\Theta_1, \Theta_2, \ldots, \Theta_n)$ be a vector, where for each $j = 1, 2, \ldots, n$, $\Theta_j = \{\langle x, T_{\Theta_j}(x), F_{\Theta_j}(x), TA_{\Theta_j}(x), FA_{\Theta_j}(x)\rangle | x \in \mathbb{S}\}$ is an ambiguous set on the universe $\mathbb{S}$. Then, Θ is called an *ambiguous vector* on $\mathbb{S}$. We define the *complement* of Θ as $\Theta^c = (\Theta_1^c, \Theta_2^c, \ldots, \Theta_n^c)$. Θ^T denotes the transpose of Θ. If $n = 1$, we do not distinguish between the ambiguous vector $\Theta = (\Theta_1)$ and the ambiguous set Θ_1.

Definition 8 (*Inner Product*)**.** For each $j = 1, 2, \ldots, n$, let

$$\Theta_j = \{\langle x, T_{\Theta_j}(x), F_{\Theta_j}(x), TA_{\Theta_j}(x), FA_{\Theta_j}(x)\rangle | x \in \mathbb{S}\}, \text{ and}$$
$$\theta_j = \{\langle x, T_{\theta_j}(x), F_{\theta_j}(x), TA_{\theta_j}(x), FA_{\theta_j}(x)\rangle | x \in \mathbb{S}\}$$

be ambiguous sets on the universe $\mathbb{S}$; let $\Theta = (\Theta_1, \Theta_2, \ldots, \Theta_n)$, $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$. We call $\Theta \cdot \theta = \{\langle x, y, z, u, v\rangle | x \in \mathbb{S}\}$ the *inner product* of Θ and θ, where $\vee$ and $\wedge$ denote the max and min operations, respectively; and $y = \bigvee_{j=1}^n (T_{\Theta_j}(x) \wedge T_{\theta_j}(x))$, $z = \bigwedge_{j=1}^n (F_{\Theta_j}(x) \vee F_{\theta_j}(x))$, $u = \bigwedge_{j=1}^n (TA_{\Theta_j}(x) \vee TA_{\theta_j}(x))$, and $v = \bigwedge_{j=1}^n (FA_{\Theta_j}(x) \vee FA_{\theta_j}(x))$. Note that $\Theta \cdot \theta$ is an ambiguous set; also, when $n = 1$, we have $\Theta \cdot \theta = \Theta_1 \cap \theta_1$.

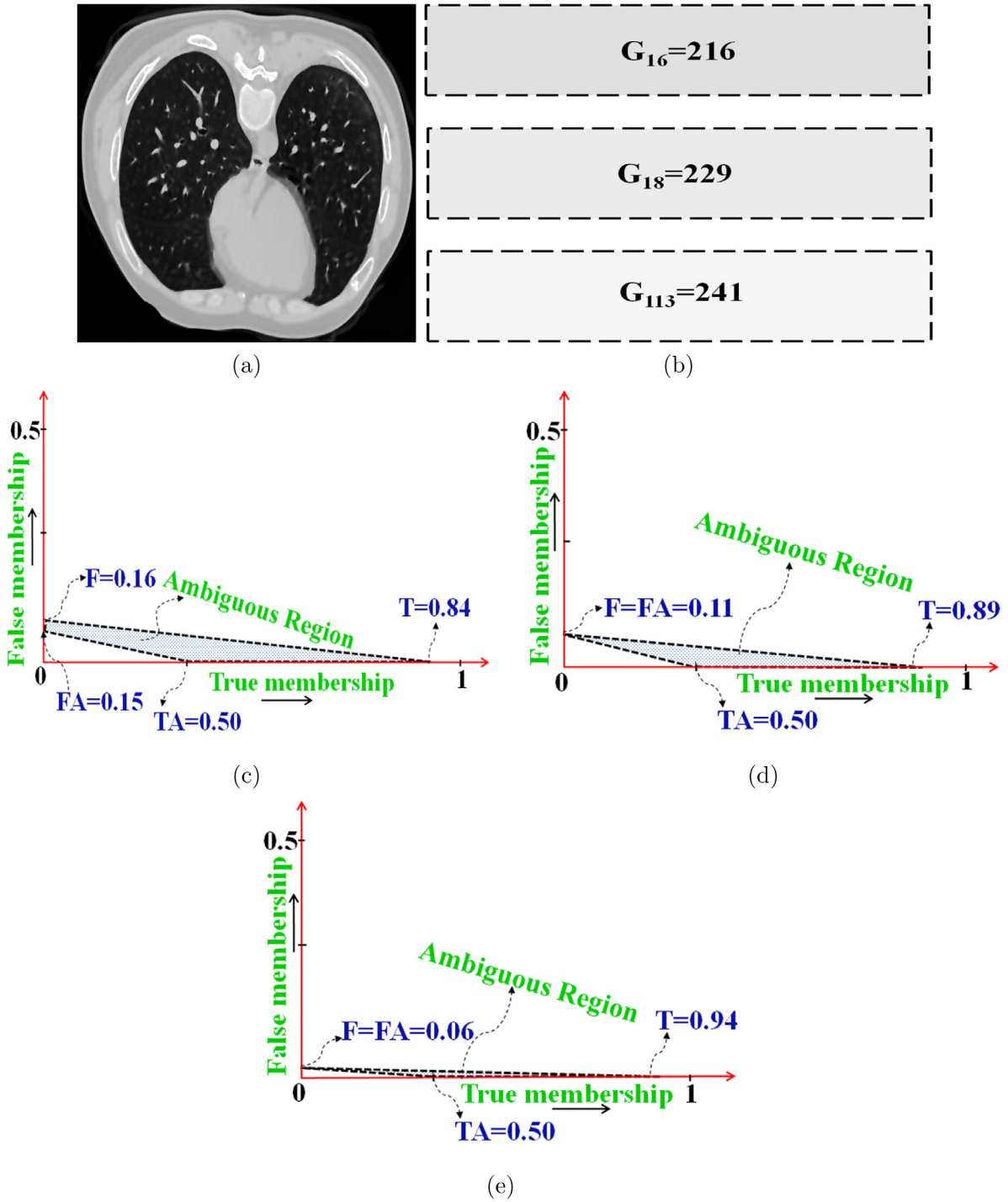

Fig. 1. Representation of grayscale values using ambiguous sets: (a) a chest CT scan image of COVID-19, (b) three different grayscale values at pixel positions $P_{15} = 215$, $P_{17} = 228$ and $P_{112} = 240$, (c) AMF values for $P_{15} = 215$, (d) AMF values for $P_{17} = 228$, and (e) AMF values for $P_{112} = 240$.

Definition 9 (*Outer Product*)**.** For each $j = 1, 2, \ldots, n$, let

$\Theta_j = \{\langle x, T_{\Theta_j}(x), F_{\Theta_j}(x), TA_{\Theta_j}(x), FA_{\Theta_j}(x)\rangle | x \in \mathbb{S}\}$, and
$\theta_j = \{\langle x, T_{\theta_j}(x), F_{\theta_j}(x), TA_{\theta_j}(x), FA_{\theta_j}(x)\rangle | x \in \mathbb{S}\}$

be ambiguous sets on the universe $\mathbb{S}$; let $\Theta = (\Theta_1, \Theta_2, \ldots, \Theta_n)$, $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$. We call $\Theta \circ \theta = \{\langle x, y', z', u', v'\rangle | x \in \mathbb{S}\}$ the *outer product* of Θ and θ, where $y' = \bigwedge_{j=1}^{n}(T_{\Theta_j}(x) \vee T_{\theta_j}(x))$, $z' = \bigvee_{j=1}^{n}(F_{\Theta_j}(x) \wedge F_{\theta_j}(x))$, $u = \bigvee_{j=1}^{n}(TA_{\Theta_j}(x) \wedge TA_{\theta_j}(x))$, and $v' = \bigvee_{j=1}^{n}(FA_{\Theta_j}(x) \wedge FA_{\theta_j}(x))$. Note that $\Theta \circ \theta$ is an ambiguous set; also, when $n = 1$, we have $\Theta \circ \theta = \Theta_1 \cup \theta_1$.

Theorem 1. *For each $j = 1, 2, \ldots, n$, let*

$\Theta_j = \{\langle x, T_{\Theta_j}(x), F_{\Theta_j}(x), TA_{\Theta_j}(x), FA_{\Theta_j}(x)\rangle | x \in \mathbb{S}\}$, *and*
$\theta_j = \{\langle x, T_{\theta_j}(x), F_{\theta_j}(x), TA_{\theta_j}(x), FA_{\theta_j}(x)\rangle | x \in \mathbb{S}\}$

be ambiguous sets on the universe $\mathbb{S}$; let $\Theta = (\Theta_1, \Theta_2, \ldots, \Theta_n)$, $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$. Then, $(\Theta \cdot \theta)^c = \Theta^c \circ \theta^c$, $(\Theta \circ \theta)^c = \Theta^c \cdot \theta^c$.

Proof. According to Definitions 6, 8 and 9, we have:

$$
\begin{aligned}
&(\Theta \cdot \theta)^c \\
&= \Big\{\Big\langle x, \bigvee_{j=1}^{n}(T_{\Theta_j}(x) \wedge T_{\theta_j}(x)), \bigwedge_{j=1}^{n}(F_{\Theta_j}(x) \vee F_{\theta_j}(x)), \\
&\quad \bigwedge_{j=1}^{n}(TA_{\Theta_j}(x) \vee TA_{\theta_j}(x)), \\
&\quad \bigwedge_{j=1}^{n}(FA_{\Theta_j}(x) \vee FA_{\theta_j}(x))\Big\rangle | x \in \mathbb{S}\Big\}^c \\
&= \Big\{\Big\langle x, \bigwedge_{j=1}^{n}(F_{\Theta_j}(x) \vee F_{\theta_j}(x)), \bigvee_{j=1}^{n}(T_{\Theta_j}(x) \wedge T_{\theta_j}(x)), \\
&\quad 1 - \bigwedge_{j=1}^{n}(TA_{\Theta_j}(x) \vee TA_{\theta_j}(x)), \\
&\quad 1 - \bigwedge_{j=1}^{n}(FA_{\Theta_j}(x) \vee FA_{\theta_j}(x))\Big\rangle | x \in \mathbb{S}\Big\} \\
&= \Big\{\Big\langle x, \bigwedge_{j=1}^{n}(F_{\Theta_j}(x) \vee F_{\theta_j}(x)), \bigvee_{j=1}^{n}(T_{\Theta_j}(x) \wedge T_{\theta_j}(x)), \\
&\quad \bigvee_{j=1}^{n}\big[(1 - TA_{\Theta_j}(x)) \wedge (1 - TA_{\theta_j}(x))\big], \\
&\quad \bigvee_{j=1}^{n}\big[(1 - FA_{\Theta_j}(x)) \wedge (1 - FA_{\theta_j}(x))\big]\Big\rangle | x \in \mathbb{S}\Big\} \\
&= \Theta^c \circ \theta^c. \\
&(\Theta \circ \theta)^c \\
&= \Big\{\Big\langle x, \bigwedge_{j=1}^{n}(T_{\Theta_j}(x) \vee T_{\theta_j}(x)), \\
&\quad \bigvee_{j=1}^{n}(F_{\Theta_j}(x) \wedge F_{\theta_j}(x)), \bigvee_{j=1}^{n}(TA_{\Theta_j}(x) \wedge TA_{\theta_j}(x)), \\
&\quad \bigvee_{j=1}^{n}(FA_{\Theta_j}(x) \wedge FA_{\theta_j}(x))\Big\rangle | x \in \mathbb{S}\Big\}^c \\
&= \Big\{\Big\langle x, \bigvee_{j=1}^{n}(F_{\Theta_j}(x) \wedge F_{\theta_j}(x)), \\
&\quad \bigwedge_{j=1}^{n}(T_{\Theta_j}(x) \vee T_{\theta_j}(x)), 1 - \bigvee_{j=1}^{n}(TA_{\Theta_j}(x) \wedge TA_{\theta_j}(x)), \\
&\quad 1 - \bigvee_{j=1}^{n}(FA_{\Theta_j}(x) \wedge FA_{\theta_j}(x))\Big\rangle | x \in \mathbb{S}\Big\} \\
&= \Big\{\Big\langle x, \bigvee_{j=1}^{n}(F_{\Theta_j}(x) \wedge F_{\theta_j}(x)), \bigwedge_{j=1}^{n}(T_{\Theta_j}(x) \vee T_{\theta_j}(x)), \\
&\quad \bigwedge_{j=1}^{n}\big[(1 - TA_{\Theta_j}(x)) \vee (1 - TA_{\theta_j}(x))\big], \\
&\quad \bigwedge_{j=1}^{n}\big[(1 - FA_{\Theta_j}(x)) \vee (1 - FA_{\theta_j}(x))\big]\Big\rangle | x \in \mathbb{S}\Big\} \\
&= \Theta^c \cdot \theta^c. \quad \blacksquare
\end{aligned}
$$

Theorem 2. *For each $j = 1, 2, \ldots, n$, let*

$\Theta_j = \{\langle x, T_{\Theta_j}(x), F_{\Theta_j}(x), TA_{\Theta_j}(x), FA_{\Theta_j}(x)\rangle | x \in \mathbb{S}\}$, *and*
$\theta_j = \{\langle x, T_{\theta_j}(x), F_{\theta_j}(x), TA_{\theta_j}(x), FA_{\theta_j}(x)\rangle | x \in \mathbb{S}\}$

be ambiguous sets on the universe $\mathbb{S}$; let $\Theta = (\Theta_1, \Theta_2, \ldots, \Theta_n)$, $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$. Then, $\Theta \cdot \theta = \theta \cdot \Theta$, $\Theta \circ \theta = \theta \circ \Theta$.

Proof. According to Definitions 8 and 9:

$$
\begin{aligned}
&\Theta \cdot \theta \\
&= \Big\{\Big\langle x, \bigvee_{j=1}^{n}(T_{\Theta_j}(x) \wedge T_{\theta_j}(x)), \bigwedge_{j=1}^{n}(F_{\Theta_j}(x) \vee F_{\theta_j}(x)), \\
&\quad \bigwedge_{j=1}^{n}(TA_{\Theta_j}(x) \vee TA_{\theta_j}(x)), \\
&\quad \bigwedge_{j=1}^{n}(FA_{\Theta_j}(x) \vee FA_{\theta_j}(x))\Big\rangle | x \in \mathbb{S}\Big\} \\
&= \Big\{\Big\langle x, \bigvee_{j=1}^{n}(T_{\theta_j}(x) \wedge T_{\Theta_j}(x)), \bigwedge_{j=1}^{n}(F_{\theta_j}(x) \vee F_{\Theta_j}(x)), \\
&\quad \bigwedge_{j=1}^{n}(TA_{\theta_j}(x) \vee TA_{\Theta_j}(x)), \\
&\quad \bigwedge_{j=1}^{n}(FA_{\theta_j}(x) \vee FA_{\Theta_j}(x))\Big\rangle | x \in \mathbb{S}\Big\} \\
&= \theta \cdot \Theta. \\
&\Theta \circ \theta \\
&= \Big\{\Big\langle x, \bigwedge_{j=1}^{n}(T_{\Theta_j}(x) \vee T_{\theta_j}(x)), \bigvee_{j=1}^{n}(F_{\Theta_j}(x) \wedge F_{\theta_j}(x)), \\
&\quad \bigvee_{j=1}^{n}(TA_{\Theta_j}(x) \wedge TA_{\theta_j}(x)), \\
&\quad \bigvee_{j=1}^{n}(FA_{\Theta_j}(x) \wedge FA_{\theta_j}(x))\Big\rangle | x \in \mathbb{S}\Big\} \\
&= \Big\{\Big\langle x, \bigwedge_{j=1}^{n}(T_{\theta_j}(x) \vee T_{\Theta_j}(x)), \\
&\quad \bigvee_{j=1}^{n}(F_{\theta_j}(x) \wedge F_{\Theta_j}(x)), \bigvee_{j=1}^{n}(TA_{\theta_j}(x) \wedge TA_{\Theta_j}(x)), \\
&\quad \bigvee_{j=1}^{n}(FA_{\theta_j}(x) \wedge FA_{\Theta_j}(x))\Big\rangle | x \in \mathbb{S}\Big\} \\
&= \theta \circ \Theta. \quad \blacksquare
\end{aligned}
$$

Definition 10 (*Jaccard Similarity Measure*)**.** Let $\eta = (\eta_1, \eta_2, \ldots, \eta_n)$ and $\varsigma = (\varsigma_1, \varsigma_2, \ldots, \varsigma_n)$ be two vectors of length n, where all coordinates are positive. The *Jaccard similarity measure* of these two vectors is defined as:

$$
\begin{aligned}
J(\eta, \varsigma) &= \frac{\eta \cdot \varsigma}{\|\eta\|^2 + \|\varsigma\|^2 + \eta \cdot \varsigma} \\
&= \frac{\sum_{i=1}^{n} \eta_i \varsigma_i}{\sum_{i=1}^{n} \eta_i^2 + \sum_{i=1}^{n} \varsigma_i^2 - \sum_{i=1}^{n} \eta_i \varsigma_i}
\end{aligned} \tag{25}
$$

where, $\eta \cdot \varsigma = \sum_{i=1}^{n} \eta_i \varsigma_i$ is the inner product of the vectors η and ς.

Definition 11 (*Dice Similarity Measure*)**.** Let $\eta = (\eta_1, \eta_2, \ldots, \eta_n)$ and $\varsigma = (\varsigma_1, \varsigma_2, \ldots, \varsigma_n)$ be two vectors of length n, where all coordinates are positive. The *Dice similarity measure* of these two vectors is defined as:

$$
D(\eta, \varsigma) = \frac{2\eta \cdot \varsigma}{\|\eta\|^2 + \|\varsigma\|^2} = \frac{2\sum_{i=1}^{n} \eta_i \varsigma_i}{\sum_{i=1}^{n} \eta_i^2 + \sum_{i=1}^{n} \varsigma_i^2} \tag{26}
$$

Definition 12 (*Cosine Similarity Measure*)**.** Let $\eta = (\eta_1, \eta_2, \ldots, \eta_n)$ and $\varsigma = (\varsigma_1, \varsigma_2, \ldots, \varsigma_n)$ be two vectors of length n, where all the coordinates are positive. The *cosine similarity measure* of these two vectors is defined as:

$$C(\eta, \varsigma) = \frac{\eta \cdot \varsigma}{\|\eta\|^2 \|\varsigma\|^2} = \frac{\sum_{i=1}^{n} \eta_i \varsigma_i}{[\sum_{i=1}^{n} \eta_i^2][\sum_{i=1}^{n} \varsigma_i^2]} \tag{27}$$

Property 1. *The Jaccard, Dice and cosine similarity measures satisfy the following properties as:*

$P_1 : J(\eta, \varsigma), D(\eta, \varsigma), C(\eta, \varsigma) \in [0, 1]$.
$P_2 : J(\eta, \varsigma) = J(\varsigma, \eta), D(\eta, \varsigma) = D(\varsigma, \eta), C(\eta, \varsigma) = C(\varsigma, \eta)$.
$P_3 :$ *If* $\eta = \varsigma$, *then* $J(\eta, \varsigma) = D(\eta, \varsigma) = C(\eta, \varsigma) = 1$.

The above similarity measures motivate the following definition.

Definition 13. Let

$$\mathring{A}_1 = \{\langle x, T_1(x), F_1(x), TA_1(x), FA_1(x)\rangle | x \in \mathbb{S}\},$$
$$\mathring{A}_2 = \{\langle x, T_2(x), F_2(x), TA_2(x), FA_2(x)\rangle | x \in \mathbb{S}\}$$

be two ambiguous sets on the universe $\mathbb{S}$. Let $\eta = (\eta_1, \eta_2, \ldots, \eta_n)$ and $\varsigma = (\varsigma_1, \varsigma_2, \ldots, \varsigma_n)$ be two vectors of length n, where $\eta_i, \varsigma_i \in \mathbb{S}$ for $i = 1, 2, \ldots, n$. Let $w_1, w_2, \ldots, w_n$ be non-negative real numbers, called *weights*. The *ambiguous weighted Jaccard similarity measure*, *ambiguous weighted Dice similarity measure* and *ambiguous weighted cosine similarity measure* of these two ambiguous sets for the vectors η, ς are defined, respectively, as:

$$\begin{aligned}
&AWJ(\mathring{A}_1, \mathring{A}_2; \eta, \varsigma)\\
&= \sum_{i=1}^{n} w_i J\big(\big(T_1(x_i), F_1(x_i), TA_1(x_i), FA_1(x_i)\big),\\
&\quad \big(T_2(x_i), F_2(x_i), TA_2(x_i), FA_2(x_i)\big)\big),\\
&AWD(\mathring{A}_1, \mathring{A}_2; \eta, \varsigma)\\
&= \sum_{i=1}^{n} w_i D\big(\big(T_1(x_i), F_1(x_i), TA_1(x_i), FA_1(x_i)\big),\\
&\quad \big(T_2(x_i), F_2(x_i), TA_2(x_i), FA_2(x_i)\big)\big),\\
&AWC(\mathring{A}_1, \mathring{A}_2; \eta, \varsigma)\\
&= \sum_{i=1}^{n} w_i C\big(\big(T_1(x_i), F_1(x_i), TA_1(x_i), FA_1(x_i)\big),\\
&\quad \big(T_2(x_i), F_2(x_i), TA_2(x_i), FA_2(x_i)\big)\big).
\end{aligned}$$

4. The proposed ADMFCA

This section introduces the proposed ADMFCA for clustering grayscale images. The proposed ADMFCA is based on ambiguous set theory, entropies and image fusion operation. Each step of the proposed ADMFCA is explained next.

Step 1. Define the grayscale domain of image: The grayscale value G_{ij} associated with each pixel $P_{ij}(i = 1, 2, \ldots, m)(j = 1, 2, \ldots, n)$ of an input gray image I_{GI} can be expressed in a grayscale domain as:

$$I_{GI} = \begin{bmatrix} G_{11} & G_{12} & \ldots & G_{1n} \\ G_{21} & G_{22} & \ldots & G_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ G_{m1} & G_{m2} & \ldots & G_{mn} \end{bmatrix} \tag{28}$$

Here, $m \times n$ represents the total number of grayscale values in the I_{GI}. In Eq. (28), each grayscale value $G_{ij} \in P_{ij}$ is defined in the range $[0, G]$ with $G = 255$. Hence, the universe of discourse $\mathbb{S}$ for each $G_{ij} \in I_{GI}$ is defined as $\mathbb{S} = [0, G]$.

Step 2. Define the ambiguous domain of image: The ambiguous domain for the grayscale image I_{GI} is defined by representing the grayscale value G_{ij} of each pixel in the ambiguous set. The ambiguous set of each G_{ij} is denoted as $\mathring{A}_{ij}$, and can be expressed in the following matrix $\mathring{A}_A$ as:

$$\mathring{A}_A = \begin{bmatrix} \mathring{A}_{11} & \mathring{A}_{12} & \ldots & \mathring{A}_{1n} \\ \mathring{A}_{21} & \mathring{A}_{22} & \ldots & \mathring{A}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathring{A}_{m1} & \mathring{A}_{m2} & \ldots & \mathring{A}_{mn} \end{bmatrix} \tag{29}$$

In Eq. (29), each $\mathring{A}_{ij}$ is defined as:

$$\mathring{A}_{ij} = \{G_{ij}, \langle T(G_{ij}), F(G_{ij}), TA(G_{ij}), FA(G_{ij})\rangle | G_{ij} \in \mathbb{S}\} \tag{30}$$

In Eq. (30), the four AMFs, namely, T, F, TA and FA for $G_{ij} \in \mathbb{S}$ can be defined as:

$$T(G_{ij}) = \frac{G_{ij} - \min(\mathbb{S})}{\max(\mathbb{S}) - \min(\mathbb{S})} \tag{31}$$

$$F(G_{ij}) = 1 - T(G_{ij}) \tag{32}$$

$$TA(G_{ij}) = \frac{T(G_{ij})}{T(G_{ij}) + A_F(G_{ij})} \tag{33}$$

$$FA(G_{ij}) = \frac{F(G_{ij})}{F(G_{ij}) + A_F(G_{ij})} \tag{34}$$

In Eq. (31), min and max represent the minimum and maximum functions, respectively. In Eqs. (33) and (34), the ambiguous distance function A_F can be defined as:

$$A_F(G_{ij}) = \sqrt{T(G_{ij})^2 + F(G_{ij})^2} \tag{35}$$

Step 3. Measurements of ambiguousness for ambiguous set: The ambiguousness of the AMFs (Eqs. (31)–(34)) can be measured individually by TE, FE, TAE and FAE, which are expressed as $E_T(\mathring{A}_{ij}, G_{ij})$, $E_F(\mathring{A}_{ij}, G_{ij})$, $E_{TA}(\mathring{A}_{ij}, G_{ij})$ and $E_{FA}(\mathring{A}_{ij}, G_{ij})$, defined in Eqs. (36)–(39), respectively, as:

$$E_T(\mathring{A}_{ij}, G_{ij}) = -T(G_{ij}) \cdot \ln(T(G_{ij})) \tag{36}$$

$$E_F(\mathring{A}_{ij}, G_{ij}) = -F(G_{ij}) \cdot \ln(F(G_{ij})) \tag{37}$$

$$E_{TA}(\mathring{A}_{ij}, G_{ij}) = -TA(G_{ij}) \cdot \ln(TA(G_{ij})) \tag{38}$$

$$E_{FA}(\mathring{A}_{ij}, G_{ij}) = -FA(G_{ij}) \cdot \ln(FA(G_{ij})) \tag{39}$$

Step 4. Selection of clusters for the entropies: Choose D initial number of clusters at random for the $E_T(\mathring{A}_{ij}, G_{ij})$, $E_F(\mathring{A}_{ij}, G_{ij})$, $E_{TA}(\mathring{A}_{ij}, G_{ij})$ and $E_{FA}(\mathring{A}_{ij}, G_{ij})$ (Eqs. (36)–(39), respectively) as C_d^T, C_d^F, C_d^{TA} and C_d^{FA}, respectively; where $d = 1, 2, \ldots, D$. Here, C_d^T, C_d^F, C_d^{TA} and C_d^{FA} occupy each $E_T(\mathring{A}_{ij}, G_{ij})$, $E_F(\mathring{A}_{ij}, G_{ij})$, $E_{TA}(\mathring{A}_{ij}, G_{ij})$ and $E_{FA}(\mathring{A}_{ij}, G_{ij})$, respectively.

Step 5. Define the set of centers for each of the clusters: Define a set of random initialized centers for each of the clusters C_d^T, C_d^F, C_d^{TA} and C_d^{FA} as:

$$W(0) = [W_1(0), W_2(0), \ldots, W_D(0)] \in C_d^T \tag{40}$$

$$X(0) = [X_1(0), X_2(0), \ldots, X_D(0)] \in C_d^F \tag{41}$$

$$Y(0) = [Y_1(0), Y_2(0), \ldots, Y_D(0)] \in C_d^{TA} \tag{42}$$

$$Z(0) = [Z_1(0), Z_2(0), \ldots, Z_D(0)] \in C_d^{FA} \tag{43}$$

Here, 0 indicates the 1st epoch of the algorithm. From Eq. (40), it can be assumed that C_d^T cluster has W_i center, where $W_i \subseteq E_T(\mathring{A}_{ij}, G_{ij})(1 \le d \le D)$. Similar assumptions can be made for C_d^F, C_d^{TA} and C_d^{FA} in terms of Eqs. (41)–(43), respectively.

Step 6. Set the epochs: For individual clustering of $E_T(\mathring{A}_{ij}, G_{ij})$, $E_F(\mathring{A}_{ij}, G_{ij})$, $E_{TA}(\mathring{A}_{ij}, G_{ij})$ and $E_{FA}(\mathring{A}_{ij}, G_{ij})$, the epoch e from 0 to *Epoch* is set as $e = 0, 1, \ldots,$ *Epoch*, where *Epoch* denotes the maximum number of epochs.

Step 7. Computation of distances between entropies and centers: Each of the entropies $E_T(\mathring{A}_{ij}, G_{ij})$, $E_F(\mathring{A}_{ij}, G_{ij})$, $E_{TA}(\mathring{A}_{ij}, G_{ij})$ and $E_{FA}(\mathring{A}_{ij}, G_{ij})$ is assigned to the individual clusters C_d^T, C_d^F, C_d^{TA} and C_d^{FA} with respect to the nearest centers $W_i(0)$, $X_i(0)$, $Y_i(0)$ and $Z_i(0)$, respectively. The determination of the nearest center vectors $W_i(0)$, $X_i(0)$, $Y_i(0)$ and $Z_i(0)$ is done by employing ambiguous-entropy distance function. The proposed function computes the distance between $E_T(\mathring{A}_{ij}, G_{ij})$ and $W_i(0)$ as:

$$Dist[E_T(\mathring{A}_{ij}, G_{ij}), W_i(0)] = [E_T(\mathring{A}_{ij}, G_{ij})]^2 + [W_i(0)]^2 - 2 \cdot [E_T(\mathring{A}_{ij}, G_{ij})] \cdot [W_i(0)] \quad (44)$$

Similarly, the proposed metric computes the distances between $E_F(\mathring{A}_{ij}, G_{ij})$ and $X_i(0)$, $E_{TA}(\mathring{A}_{ij}, G_{ij})$ and $Y(0)$, and $E_{FA}(\mathring{A}_{ij}, G_{ij})$ and $Z_i(0)$, defined in Eqs. (45)–(47), respectively, as:

$$Dist[E_F(\mathring{A}_{ij}, G_{ij}), X_i(0)] = [E_F(\mathring{A}_{ij}, G_{ij})]^2 + [X_i(0)]^2 - 2 \cdot [E_F(\mathring{A}_{ij}, G_{ij})] \cdot [X_i(0)] \quad (45)$$

$$Dist[E_{TA}(\mathring{A}_{ij}, G_{ij}), Y_i(0)] = [E_{TA}(\mathring{A}_{ij}, G_{ij})]^2 + [Y_i(0)]^2 - 2 \cdot [E_{TA}(\mathring{A}_{ij}, G_{ij})] \cdot [Y_i(0)] \quad (46)$$

$$Dist[E_{FA}(\mathring{A}_{ij}, G_{ij}), Z_i(0)] = [E_{FA}(\mathring{A}_{ij}, G_{ij})]^2 + [Z_i(0)]^2 - 2 \cdot [E_{FA}(\mathring{A}_{ij}, G_{ij})] \cdot [Z_i(0)] \quad (47)$$

In Eqs. (44)–(47), $Dist[\cdot]$ denotes the ambiguous-entropy distance metric. In Eq. (44), if $W_i(0)$ is the closest center to $E_T(\mathring{A}_{ij}, G_{ij})$, then it is assigned to the cluster C_d^T. A similar explanation can be given for Eqs. (45)–(47).

Step 8. Selection criterion of clusters: The selection of each cluster by $E_T(\mathring{A}_{ij}, G_{ij})$, $E_F(\mathring{A}_{ij}, G_{ij})$, $E_{TA}(\mathring{A}_{ij}, G_{ij})$ and $E_{FA}(\mathring{A}_{ij}, G_{ij})$ depends on the minimum values of the ambiguous-entropy distances (Eqs. (44)–(47), respectively). For example, let $W_i(0)$ and $W_j(0)$ be the two randomly defined centers for the clusters C_i^T and C_j^T with respect to the clustering $E_T(\mathring{A}_{ij}, G_{ij})$. Now, $E_T(\mathring{A}_{ij}, G_{ij}) \in W_i(0)$ if it satisfies the following condition as:

$$|Dist[E_T(\mathring{A}_{ij}, G_{ij}), W_i(0)]| < |Dist[E_T(\mathring{A}_{ij}, G_{ij}), W_j(0)]| \quad (48)$$

where, $W_i(0) \neq W_j(0)$. In Eq. (48), $Dist[E_T(\mathring{A}_{ij}, G_{ij}), W_i(0)]$ and $Dist[E_T(\mathring{A}_{ij}, G_{ij}), W_j(0)]$ can be obtained by employing Eq. (44). Eq. (48) indicates that $W_i(0)$ is the nearest center for $E_T(\mathring{A}_{ij}, G_{ij})$, so it is assigned to the cluster C_i^T. A similar explanation can be given for clustering $E_F(\mathring{A}_{ij}, G_{ij})$, $E_{TA}(\mathring{A}_{ij}, G_{ij})$ and $E_{FA}(\mathring{A}_{ij}, G_{ij})$.

Step 9. Update the centers: After each epoch, the proposed algorithm updates their centers. This process continues until it reaches the maximum epoch *Epoch*. During the individual clustering of $E_T(\mathring{A}_{ij}, G_{ij})$, $E_F(\mathring{A}_{ij}, G_{ij})$, $E_{TA}(\mathring{A}_{ij}, G_{ij})$ and $E_{FA}(\mathring{A}_{ij}, G_{ij})$, the respective new centers are denoted as $W_i(e+1)$, $X_i(e+1)$, $Y_i(e+1)$ and $Z_i(e+1)$, defined in Eqs. (49)–(52), respectively, as:

$$W_i(e+1) = \frac{1}{A \times B} \sum_{i=1}^{A} \sum_{j=1}^{B} E_T(\mathring{A}_{ij}, G_{ij}) \quad (49)$$

$$X_i(e+1) = \frac{1}{I \times J} \sum_{i=1}^{I} \sum_{j=1}^{J} E_F(\mathring{A}_{ij}, G_{ij}) \quad (50)$$

$$Y_i(e+1) = \frac{1}{P \times Q} \sum_{i=1}^{P} \sum_{j=1}^{Q} E_{TA}(\mathring{A}_{ij}, G_{ij}) \quad (51)$$

$$Z_i(e+1) = \frac{1}{S \times T} \sum_{i=1}^{S} \sum_{j=1}^{T} E_{FA}(\mathring{A}_{ij}, G_{ij}) \quad (52)$$

In Eqs. (49)–(52), $A \times B$, $I \times J$, $P \times Q$ and $S \times T$ represent the size of the clusters C_d^T, C_d^F, C_d^{TA} and C_d^{FA}, respectively.

Step 10. Stop the clustering process: Go to **Step 6** and proceed from epoch $e = 0$ to the next epoch $e = e + 1$. This process continues until the centers stop changing or the algorithm reaches the maximum epoch *Epoch*.

Step 11. Generate the clustered images: Individual clustering of $E_T(\mathring{A}_{ij}, G_{ij})$, $E_F(\mathring{A}_{ij}, G_{ij})$, $E_{TA}(\mathring{A}_{ij}, G_{ij})$ and $E_{FA}(\mathring{A}_{ij}, G_{ij})$ generates the four different clustered images, called TE clustered image (TECI), FE clustered image (FECI), TAE clustered image (TAECI) and FAE clustered image (FAECI). TECI, FECI, TAECI and FAECI are denoted as T_{ECI}, F_{ECI}, TA_{ECI} and FA_{ECI}, respectively.

Step 12. Obtain the final clustered image: The final clustered image (FCI) is generated by applying the image fusion operation [40] on four clustered images, viz., T_{ECI}, F_{ECI}, TA_{ECI} and FA_{ECI} as:

$$F_{CI} = \frac{1}{4}[T_{ECI} + F_{ECI} + TA_{ECI} + FA_{ECI}] \quad (53)$$

Here, F_{CI} denotes the FCI.

The pseudocode of the proposed ADMFCA is summarized in Algorithm 1.

Algorithm 1 PROCEDURE ADMFC().

Input: an image $\hat{G}_I$ with grayscale value $G_{ij} \in P_{ij}$ is defined between the range $\mathbb{S} = [0, G]$ with $G = 255$, where each $P_{ij}(i = 1, 2, \ldots, m)(j = 1, 2, \ldots, n)$ denotes the pixel of $\hat{G}_I$.
Output: final clustered image (FCI).
1: Represent $\hat{G}_I$ as grayscale domain denoted by I_{GI} (Eq. (28)).
2: Represent $\hat{G}_I$ into ambiguous domain denoted by $\mathring{A}_A$ (Eq. (29)).
3: Measure the ambiguousness of the AMFs of $\mathring{A}_A$, and expressed as $E_T(\mathring{A}_{ij}, G_{ij})$, $E_F(\mathring{A}_{ij}, G_{ij})$, $E_{TA}(\mathring{A}_{ij}, G_{ij})$ and $E_{FA}(\mathring{A}_{ij}, G_{ij})$ (Eqs. (36)–(39), respectively).
4: Choose D initial clusters at random for the $E_T(\mathring{A}_{ij}, G_{ij})$, $E_F(\mathring{A}_{ij}, G_{ij})$, $E_{TA}(\mathring{A}_{ij}, G_{ij})$ and $E_{FA}(\mathring{A}_{ij}, G_{ij})$ as C_d^T, C_d^F, C_d^{TA} and C_d^{FA}, respectively; where $d = 1, 2, \ldots, D$.
5: Define a set of random initialized centers for each of the clusters C_d^T, C_d^F, C_d^{TA} and C_d^{FA} (Eqs. (40)–(43), respectively).
while *e=0* **do**
  a: Compute the distances between entropies and centers (Eqs. (44)–(47)).
  b: Select each cluster by employing condition given (Eq. (48)).
  c: Update each of the centers (Eqs. (49)–(52)).
end
e=e+1;
6: Generate the four different clustered images as TECI, FECI, TAECI and FAECI.
7: Apply the image fusion operation on four clustered images to obtain the final clustered image as FCI.

5. Properties of ambiguous-entropy distance function

This section presents various properties of ambiguous-entropy distance function. This function is used in the proposed ADMFCA to compute the distance between entropy and center of the clusters. Consider the following generalized form of ambiguous-entropy distance function that computes the distance between $E_T(\mathring{A}_{ij}, G_{ij})$ and W_i as:

$$Dist[E_T(\mathring{A}_{ij}, G_{ij}), W_i] = [E_T(\mathring{A}_{ij}, G_{ij})]^2 + [W_i]^2 - 2 \cdot [E_T(\mathring{A}_{ij}, G_{ij})] \cdot [W_i] \quad (54)$$

Here, ambiguous-entropy distance function $Dist[E_T(\mathring{A}_{ij}, G_{ij}), W_i]$ is used to compute the distance between $E_T(\mathring{A}_{ij}, G_{ij})$ and W_i. For ease of explanation of various properties of this function, we only

consider the vectors $E_T(\mathring{A}_{ij}, G_{ij})$ and W_i. However, these properties are also valid in the case of the computation of the distances between $E_F(\mathring{A}_{ij}, G_{ij})$ and X_i, $E_{TA}(\mathring{A}_{ij}, G_{ij})$ and Y_i, and $E_{FA}(\mathring{A}_{ij}, G_{ij})$ and Z_i. In the following, we have discussed various properties of ambiguous-entropy distance function in terms of Eq. (54).

Property 2. *$Dist[E_T(\mathring{A}_{ij}, G_{ij}), W_i] : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is equal to 0 if $E_T(\mathring{A}_{ij}, G_{ij}) = W_i$.*

Proof. If $E_T(\mathring{A}_{ij}, G_{ij}) = W_i$, then $Dist[E_T(\mathring{A}_{ij}, G_{ij}), W_i] = [E_T(\mathring{A}_{ij}, G_{ij})]^2 + [W_i]^2 - 2 \cdot [E_T(\mathring{A}_{ij}, G_{ij})] \cdot [W_i] = [E_T(\mathring{A}_{ij}, G_{ij}) - W_i]^2 = 0$. ■

Property 3. *As the distance between $E_T(\mathring{A}_{ij}, G_{ij})$ and W_i increases, $Dist[E_T(\mathring{A}_{ij}, G_{ij}), W_i]$ increases.*

Proof. Assume two centers W_i and W_j, where $W_i, W_j \in W$ and $W_i > W_j$. From Eq. (54), it is clear that $Dist[E_T(\mathring{A}_{ij}, G_{ij}), W_i] > Dist[E_T(\mathring{A}_{ij}, G_{ij}), W_j]$. It indicates that as W_i increases, $Dist[E_T(\mathring{A}_{ij}, G_{ij}), W_i]$ increases.

Similarly, assume two TEs as $E_T(\mathring{A}_{ij}, G_{ij})$ and $E_T(\mathring{A}_{kl}, G_{kl})$, where $E_T(\mathring{A}_{ij}, G_{ij}) > E_T(\mathring{A}_{kl}, G_{kl})$. Now, if we compute the distances of these two TEs with respect to W_i, then clearly get $Dist[E_T(\mathring{A}_{ij}, G_{ij}), W_i] > Dist[E_T(\mathring{A}_{kl}, G_{kl}), W_i]$. It indicates that as $E_T(\mathring{A}_{ij}, G_{ij})$ increases, $Dist[E_T(\mathring{A}_{ij}, G_{ij}), W_i]$ increases. ■

Property 4. *The ambiguous-entropy distance function follows the symmetry property, i.e., $Dist[E_T(\mathring{A}_{ij}, G_{ij}), W_i] = Dist[W_i, E_T(\mathring{A}_{ij}, G_{ij})]$.*

Proof. It is obvious from Eq. (54). ■

Theorem 3. *For $Dist[E_T(\mathring{A}_{ij}, G_{ij}), W_i] : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, there exists an Euclidean space $\mathbb{E}$ and a distance vector $V : \mathbb{R} \times \mathbb{R} \to \mathbb{E}$ such that $Dist[E_T(\mathring{A}_{ij}, G_{ij}), W_i] = V\left(E_T(\mathring{A}_{ij}, G_{ij}), W_i\right) \cdot V\left(E_T(\mathring{A}_{ij}, G_{ij}), W_i\right)$, where "$\cdot$" denotes the dot product of two vectors.*

Proof. Let $\mathbb{E}$ be the 1-dimensional Euclidean space $\mathbb{R}$. Define $V : \mathbb{R} \times \mathbb{R} \to \mathbb{E}$ by $V(a, b) = a - b$. Then,

$$
\begin{aligned}
&V\left(E_T(\mathring{A}_{ij}, G_{ij}), W_i\right) \cdot V\left(E_T(\mathring{A}_{ij}, G_{ij}), W_i\right) \\
&= [E_T(\mathring{A}_{ij}, G_{ij}) - W_i] \cdot [E_T(\mathring{A}_{ij}, G_{ij}) - W_i] \\
&= [E_T(\mathring{A}_{ij}, G_{ij}) - W_i]^2 \\
&= [E_T(\mathring{A}_{ij}, G_{ij})]^2 + [W_i]^2 - 2 \cdot [E_T(^{\circ}A_{ij}, G_{ij})] \cdot [W_i] \\
&= Dist[E_T(\mathring{A}_{ij}, G_{ij}), W_i]. \quad \blacksquare
\end{aligned}
$$

Theorem 4. *$Dist[E_T(\mathring{A}_{ij}, G_{ij}), W_i] = \|V\left(E_T(\mathring{A}_{ij}, G_{ij}), W_i\right)\|^2$, where $\|\cdot\|$ denotes the standard Euclidean norm on $\mathbb{R}^2$.*

Proof. By the previous theorem, we have

$$
\begin{aligned}
Dist[E_T(\mathring{A}_{ij}, G_{ij}), W_i] &= V\left(E_T(\mathring{A}_{ij}, G_{ij}), W_i\right) \cdot V\left(E_T(\mathring{A}_{ij}, G_{ij}), W_i\right) \\
&= \left[\sqrt{V\left(E_T(\mathring{A}_{ij}, G_{ij}), W_i\right) \cdot V\left(E_T(\mathring{A}_{ij}, G_{ij}), W_i\right)}\right]^2 \\
&= \|V\left(E_T(\mathring{A}_{ij}, G_{ij}), W_i\right)\|^2. \quad \blacksquare
\end{aligned}
$$

6. Experimental results

This section is divided into several subsections, such as: dataset description is provided in Section 6.1. Various performance evaluation metrics are discussed in Section 6.2. The initial experimental setup is presented in Section 6.3. Visual analysis of the clustered images is discussed in Section 6.4. Finally, the comparison with existing models is presented in Section 6.5.

Table 2
Detailed information about the chest CT scan images of COVID-19 patients.

Group	Label	Extracted CT scan image	Image size in KB (before preprocessed)	Image size in KB (after preprocessed)
Group #1	1	#142	169	129
	2	#94	237	172
	3	#105	218	154
	4	#85	202	146
	5	#100	207	155
	6	#110	211	140
	7	#94	212	134
	8	#96	184	126
	9	#109	181	130
	10	#155	173	121
Group #2	1	#118	179	120
	2	#106	231	7.15
	3	#81	216	152
	4	#71	191	144
	5	#76	193	147
	6	#87	211	150
	7	#113	225	149
	8	#120	169	127
	9	#90	173	136
	10	#179	177	130
Group #3	1	#129	188	124
	2	#97	250	164
	3	#92	223	150
	4	#95	206	139
	5	#87	210	143
	6	#98	222	141
	7	#102	222	139
	8	#107	192	129
	9	#100	185	137
	10	#166	179	127
Group #4	1	#136	173	71.0
	2	#86	223	92.5
	3	#100	190	81.6
	4	#77	179	82.2
	5	#80	179	80.9
	6	#92	191	83.0
	7	#89	177	76.5
	8	#114	181	75.3
	9	#104	185	77.8
	10	#171	163	70.6

6.1. Dataset description

The proposed ADMFCA and selected clustering algorithms, namely KMC [23], FCM [29], GFCM [34], IIFCM [36] and NEBCA [38] are applied to different types of chest CT scan images of COVID-19 patients [45]. This dataset contains CT scan images of COVID-19 patients with 20 different labels. In this study, CT scan images with 10 different labels are selected for the experiment. Out of each label, four different CT scan images are selected. Thus, a total of $10 \times 4 = 40$ CT scan images are available with their respective ground truths. These 40 images are split into four different groups, called Group #1, Group #2, Group #3 and Group #4. However, the extracted CT scan images have noise and poor resolution issues. Therefore, these images are preprocessed before carrying out the experiment. The adaptive filtering technique [46] and the histogram equalization method [47] are used for noise removal and resolution improvement, respectively. Eventually, these preprocessed images are used for the experiment. Detailed information on the experimental datasets is available in Table 2.

6.2. Performance evaluation metrics

Clustered images obtained using the proposed ADMFCA and selected clustering algorithms are evaluated using five statistical metrics, namely MSE, PSNR, DSC, JSC and CC. These metrics are

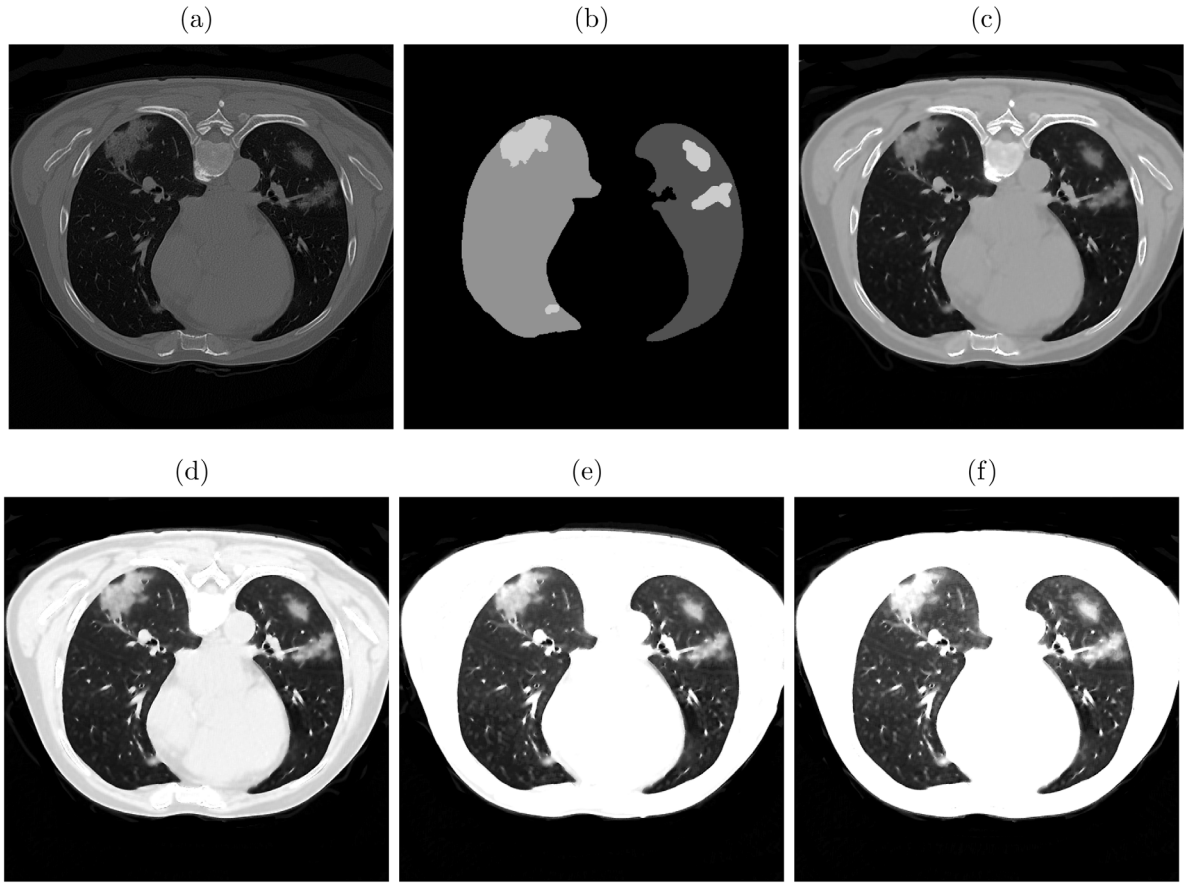

Fig. 2. Example of clustering CT scan image (**Group: #1, Label: 2, Image: #94**) based on the ADMFCA and selection of optimal FCI: (a) original CT scan image, (b) ground truth of (a), (c) preprocessed CT scan image of (a), (d) FCI with $D = 2$ (MSE: 4.03, PSNR: 42.08, DSC: 0.89, JSC: 0.93, CC: 0.94), (e) FCI with $D = 3$ (MSE: 3.04, PSNR: 43.30, DSC: 0.91, JSC: 0.95, CC: 0.96), and (f) final FCI with $D = 4$ (MSE: 4.05, PSNR: 42.06, DSC: 0.90, JSC: 0.94, CC: 0.95). The optimal FCI is shown in (e) obtained with $D = 3$.

defined based on the input grayscale image (I_{GI}), final clustered image (F_{CI}) and corresponding ground truth (G_{TI}) as:

$$MSE(I_{GI}, F_{CI}) = \frac{1}{M \times N} \sum_{m=1}^{M} \sum_{n=1}^{N} (I_{GI} - F_{CI})^2 \tag{55}$$

$$PSNR(I_{GI}, F_{CI}) = 10 \times \log_{10} \left[\frac{(255)^2}{MSE(I_{GI}, F_{CI})} \right] \tag{56}$$

$$DSC(F_{CI}, G_{TI}) = \frac{2|F_{CI} \cap G_{TI}|}{|F_{CI} + G_{TI}|} \tag{57}$$

$$JSC(F_{CI}, G_{TI}) = \frac{|F_{CI} \cap G_{TI}|}{|F_{CI} \cup G_{TI}|} \tag{58}$$

$$CC(F_{CI}, G_{TI}) = \frac{\sum_{m=1}^{M} \sum_{n=1}^{N} (F_{CI} - \bar{F}_{CI})(G_{TI} - \bar{G}_{TI})}{\left(\sum_{m=1}^{M} \sum_{n=1}^{N} (F_{CI} - \bar{F}_{CI})^2 \right) \left(\sum_{m=1}^{M} \sum_{n=1}^{N} (G_{TI} - \bar{G}_{TI})^2 \right)} \tag{59}$$

A lower MSE value (Eq. (55)) indicates a lower intensity loss and produces a robust clustered image. A higher PSNR value (Eq. (56)) shows less distortion in grayscale values and generates a better clustered image. The DSC and JSC values (Eqs. (57) and (58), respectively) always lie in the range of 0–1. Their values close to 1 imply that the region of interest of the clustered image is nearly similar to the corresponding ground truth. The CC value (Eq. (59)) remains within the range $[-1.1]$. In Eq. (59), $\bar{F}_{CI}$ and $\bar{G}_{TI}$ indicate mean values of F_{CI} and G_{TI}, respectively. A CC value close to 1 shows that the region of interest of the clustered image is perfectly similar to its corresponding background truth.

6.3. Initial experimental set-up

The main objective of ADMFCA is to effectively detect infected regions by clustering the grayscale values of the CT scan images of COVID-19 patients. In this algorithm, we first define the grayscale values (G_{ij}) of all preprocessed CT scan images of COVID-19 patients (Table 2) in terms of the four degree of memberships in the ambiguous set as:

- G_{ij} of all white pixels are defined by $T(G_{ij})$,
- G_{ij} of all non-white pixels are defined by $F(G_{ij})$,
- G_{ij} of all white pixels with certain non-white pixels are defined by $TA(G_{ij})$, and
- G_{ij} of all non-white pixels with certain white pixels are represented by $FA(G_{ij})$.

The individual ambiguity of $T(G_{ij})$, $F(G_{ij})$, $TA(G_{ij})$ and $FA(G_{ij})$ is measured using the corresponding TE, FE, TAE and FAE (Eqs. (36)–(39), respectively). Finally, the proposed algorithm is applied sequentially to create clusters $T(G_{ij})$, $F(G_{ij})$, $TA(G_{ij})$ and $FA(G_{ij})$. The proposed algorithm develops four different clustered images, called TECI, FECI, TAECI and FAECI. The final clustered images (FCIs) are obtained by applying the image fusion operation (Eq. (53)) to TECI, FECI, TAECI and FAECI.

The proposed algorithm is simulated by selecting three different cluster numbers as $D = 2, 3, 4$. The main objective of the simulation with different cluster numbers is to determine which cluster number is best to generate the optimal FCIs. The best cluster number for the proposed ADMFCA is determined by

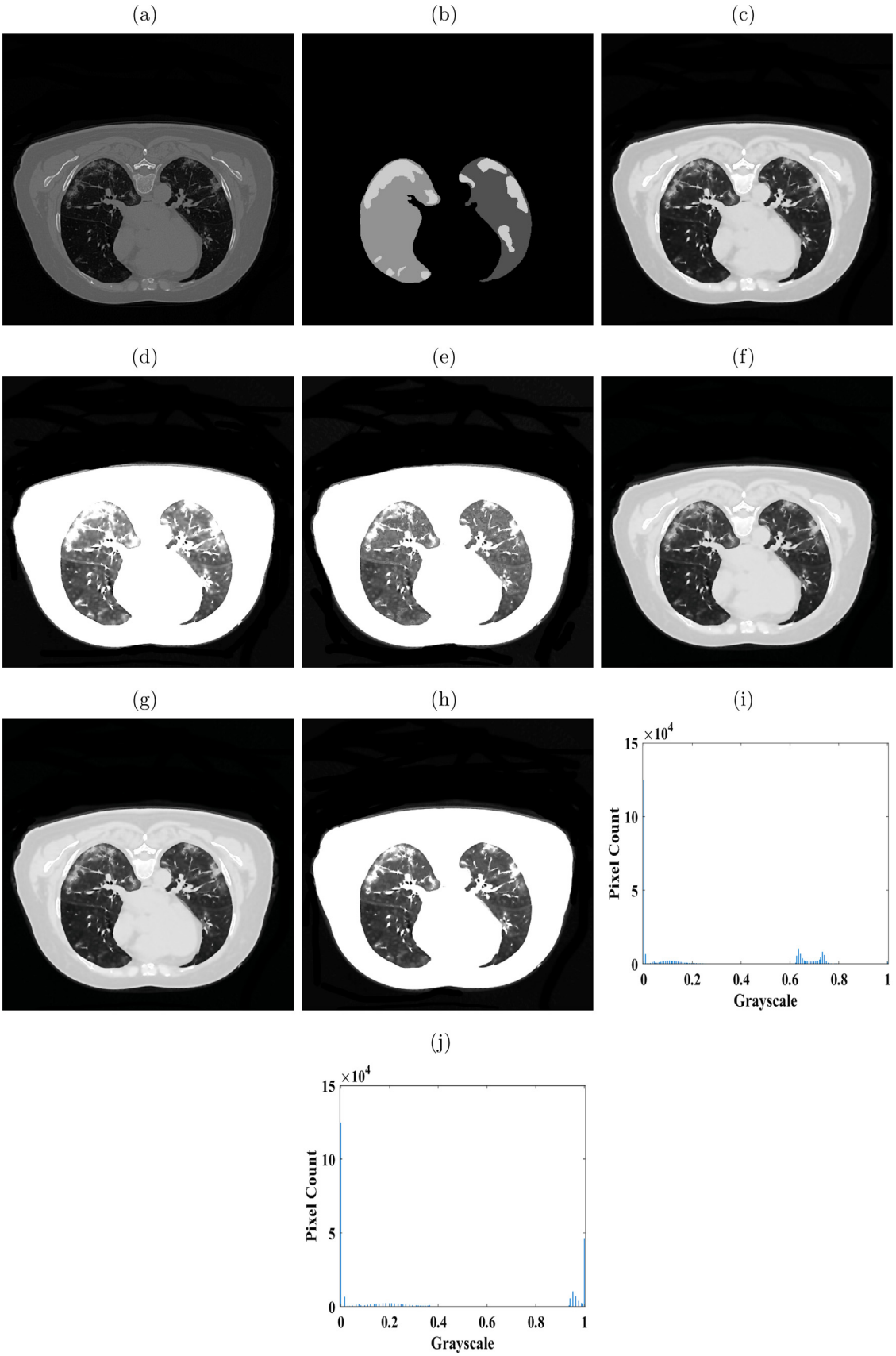

Fig. 3. Visual analysis of clustering the CT scan image (**Group: #1, Label: 1, Image: #142**) based on the ADMFCA: (a) original CT scan image, (b) ground truth of (a), (c) preprocessed CT scan image of (a), (d) TECI, (e) FECI, (f) TAECI, (g) FAECI, (h) FCI, (i) histogram of (c), and (j) histogram of (h).

Table 3
Performance evaluation of existing clustering algorithms with the proposed ADMFCA with respect to MSE for clustering the CT scan images of COVID-19.

Group	Label	CT scan image (preprocessed)	KMC	FCM	GFCM	IIFCM	NEBCA	**ADMFCA**
Group #1	1	#142	405.24	305.60	240.67	179.67	119.34	**2.03**
	2	#94	439.72	307.73	230.24	174.14	118.15	**3.04**
	3	#105	464.27	393.48	238.01	125.01	115.01	**3.04**
	4	#85	493.26	391.21	235.11	126.11	116.11	**1.02**
	5	#100	409.28	394.91	244.44	138.10	114.23	**2.03**
	6	#110	475.81	310.35	248.87	128.23	116.45	**1.03**
	7	#94	513.15	311.36	238.87	130.15	117.12	**1.04**
	8	#96	418.35	318.27	210.30	132.30	115.13	**1.04**
	9	#109	436.93	311.90	217.93	136.93	114.93	**1.05**
	10	#155	463.07	312.73	215.77	136.77	114.77	**2.05**
Group #2	1	#118	497.78	313.26	213.13	135.13	114.13	**1.07**
	2	#106	497.64	312.25	216.40	127.40	116.14	**1.02**
	3	#81	425.53	314.63	243.64	136.64	118.64	**4.03**
	4	#71	420.01	316.22	239.11	136.11	114.11	**1.03**
	5	#76	594.30	317.85	216.13	123.13	116.23	**1.03**
	6	#87	593.02	314.75	236.87	125.87	116.27	**2.01**
	7	#113	438.52	315.23	235.72	128.72	115.72	**1.03**
	8	#120	450.75	317.63	238.28	126.28	115.28	**1.04**
	9	#90	435.72	312.82	232.82	132.82	112.82	**1.05**
	10	#179	493.02	314.75	236.87	126.87	116.87	**1.04**
Group #3	1	#129	416.24	311.10	239.67	135.17	119.54	**1.04**
	2	#97	430.72	312.13	230.24	121.24	110.15	**1.05**
	3	#92	563.17	302.58	238.11	128.11	113.21	**1.06**
	4	#95	583.16	316.24	215.31	124.11	115.11	**1.04**
	5	#87	511.18	339.95	212.42	129.10	114.23	**1.05**
	6	#98	475.31	312.15	266.57	129.23	116.45	**1.03**
	7	#102	411.15	311.16	218.77	127.15	118.12	**2.05**
	8	#107	415.15	324.17	240.32	131.31	111.13	**3.01**
	9	#100	435.13	321.91	236.83	138.93	118.93	**1.04**
	10	#166	413.17	322.13	244.87	136.75	114.77	**2.02**
Group #4	1	#136	425.64	321.20	239.67	129.12	119.14	**2.04**
	2	#86	411.72	332.13	231.14	125.21	110.25	**2.03**
	3	#100	464.17	312.38	248.21	128.21	113.23	**4.05**
	4	#77	583.16	331.14	236.32	139.11	115.21	**1.02**
	5	#80	511.38	341.95	232.45	128.12	114.24	**1.03**
	6	#92	477.36	332.15	246.67	133.43	116.41	**1.06**
	7	#89	414.35	324.66	218.71	139.25	118.22	**1.02**
	8	#114	418.35	314.27	239.31	139.32	111.23	**2.05**
	9	#104	436.23	341.92	236.82	130.82	118.92	**1.03**
	10	#171	413.27	332.23	214.82	136.54	114.17	**2.05**
Average	–	–	464.38	324.11	232.91	133.42	115.50	**1.59**

evaluating the quality of the FCIs using statistical metrics, such as MSE, PSNR, DSC, JSC and CC (Eqs. (55)–(59), respectively). The FCIs are obtained by setting the maximum number of epochs to $Epoch = 100$.

6.4. Visual analysis

Consider a CT scan image #94 (Group #1, Label: 2) shown in Fig. 2(a). The respective ground truth of this image is shown in Fig. 2(b). In Fig. 2(c), the preprocessed image of Fig. 2(a) obtained by the adaptive filtering technique [46] followed by the histogram equalization method [47] is shown. Then, the proposed ADMFCA is applied to the preprocessed image (Fig. 2(c)) to obtain the FCIs with respect to $D = 2, 3, 4$. The FCIs of Fig. 2(c) with respect to $D = 2, 3, 4$ are given in Fig. 2(d)∼(f), respectively.

Finally, the MSE, PSNR, DSC, JSC and CC values of the FCIs are measured with respect to $D = 2, 3, 4$, and their values are mentioned in the caption of Fig. 2(d)∼(f). The MSE, PSNR, DSC, JSC and CC values of the FCI (Fig. 2(e)) obtained by $D = 3$ are 0.01, 68.75, 0.92, 0.96 and 0.97, respectively; these values are significantly better than the FCIs (Fig. 2(d) and (f)) obtained by $D = 2$ and $D = 4$. This way, all preprocessed images listed in Table 2 are grouped using the proposed ADMFCA, and FCIs are generated. From the experiment, it is observed that the proposed ADMFCA generates the optimal FCIs from preprocessed images with $D = 3$. From now on, the rest of our experimental results and discussion on the proposed ADMFCA are based on $D = 3$.

A visual analysis is conducted to assess the quality of FCIs obtained from preprocessed CT scan images of COVID-19 patients. To demonstrate visual analysis, CT scan images #142 (Label: 1), #118 (Label: 1), #166 (Label: 10) and #114 (Label: 8) are selected from Groups #1-#4, respectively. These selected images are shown in Figs. 3–6(a). The ground truths of the selected images are shown in Figs. 3–6(b). The preprocessed images of Figs. 3–6(a) are shown in Figs. 3–6(c). The four different clustered images, namely TECIs, FECIs, TAECIs and FAECIs, obtained by the proposed ADMFCA are depicted in Figs. 3–6(d)∼(g), respectively. By applying the fusion operation to TECI, FECI, TAECI, and FAECI, we obtain the FCIs shown in Figs. 3–6(h). In Figs. 3–6(h), dense white pixels in the lungs denote infection and symptoms of COVID-19. Figs. 3–6(i) present the histograms of the preprocessed images (Figs. 3–6(c)), which indicate the distribution of grayscale values in the pixels before performing the clustering operation. Figs. 3–6(j) depict the histograms of the FCIs (Figs. 3–6(h)), which indicate the distribution of grayscale values in the pixels after performing the clustering operation. From Figs. 3–6(j), it can be seen that the proposed ADMFCA has effectively performed the clustering operation on grayscale values.

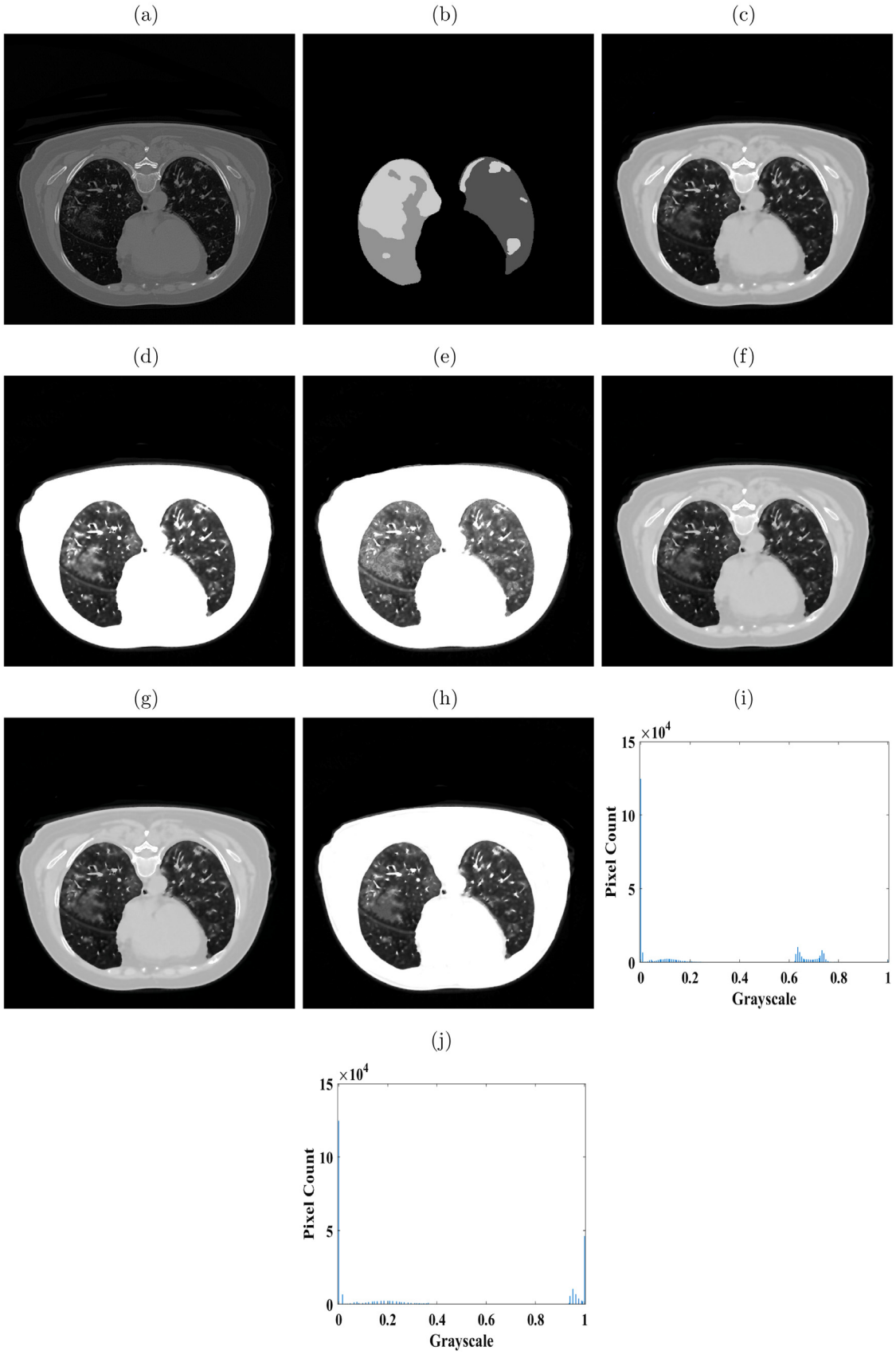

Fig. 4. Visual analysis of clustering the CT scan image (**Group: #2, Label: 1, Image: #118**) based on the ADMFCA: (a) original CT scan image, (b) ground truth of (a), (c) preprocessed CT scan image of (a), (d) TECI, (e) FECI, (f) TAECI, (g) FAECI, (h) FCI, (i) histogram of (c), and (j) histogram of (h).

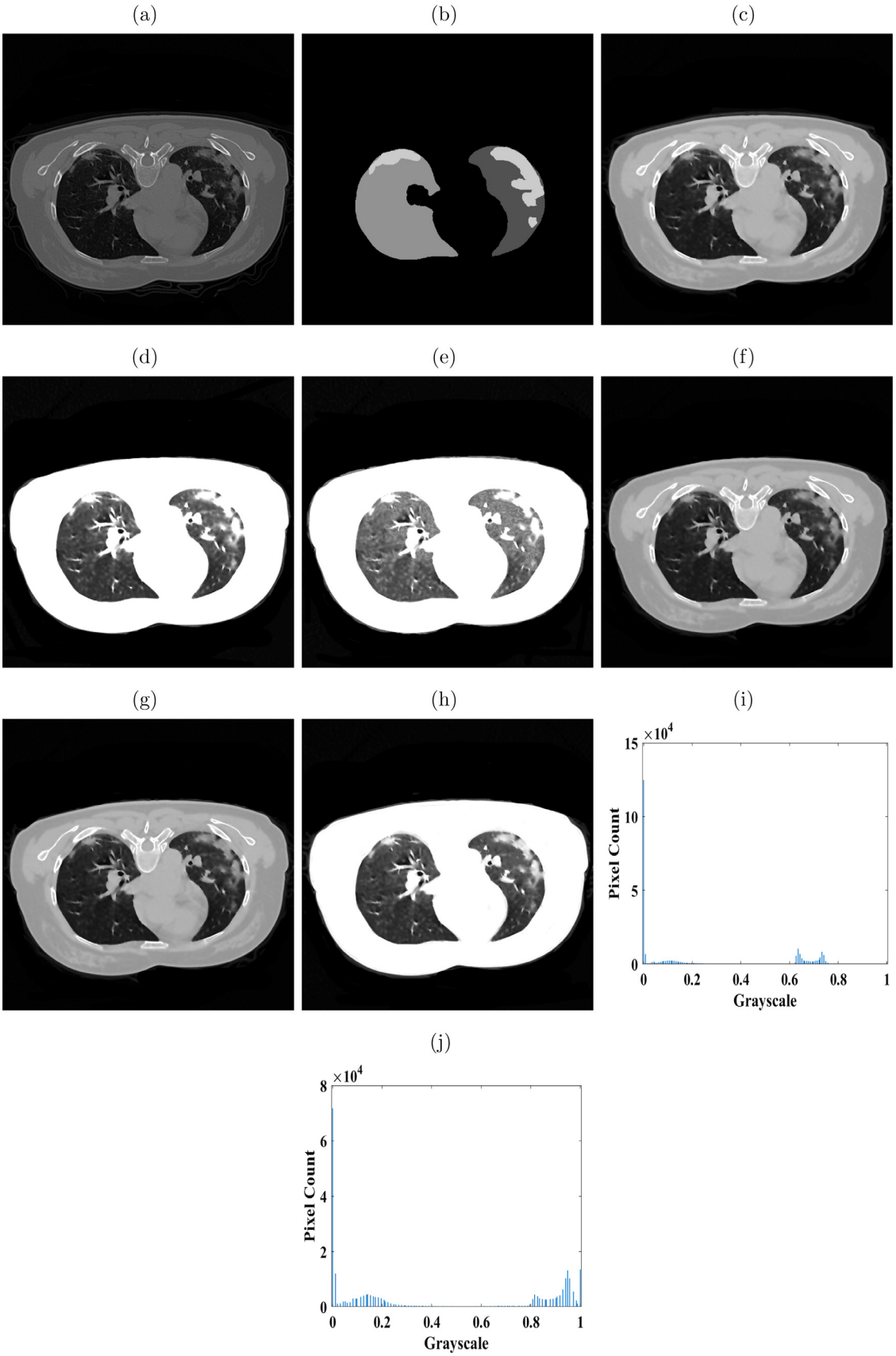

Fig. 5. Visual analysis of clustering the CT scan image (**Group: #3, Label: 10, Image: #166**) based on the ADMFCA: (a) original CT scan image, (b) ground truth of (a), (c) preprocessed CT scan image of (a), (d) TECI, (e) FECI, (f) TAECI, (g) FAECI, (h) FCI, (i) histogram of (c), and (j) histogram of (h).

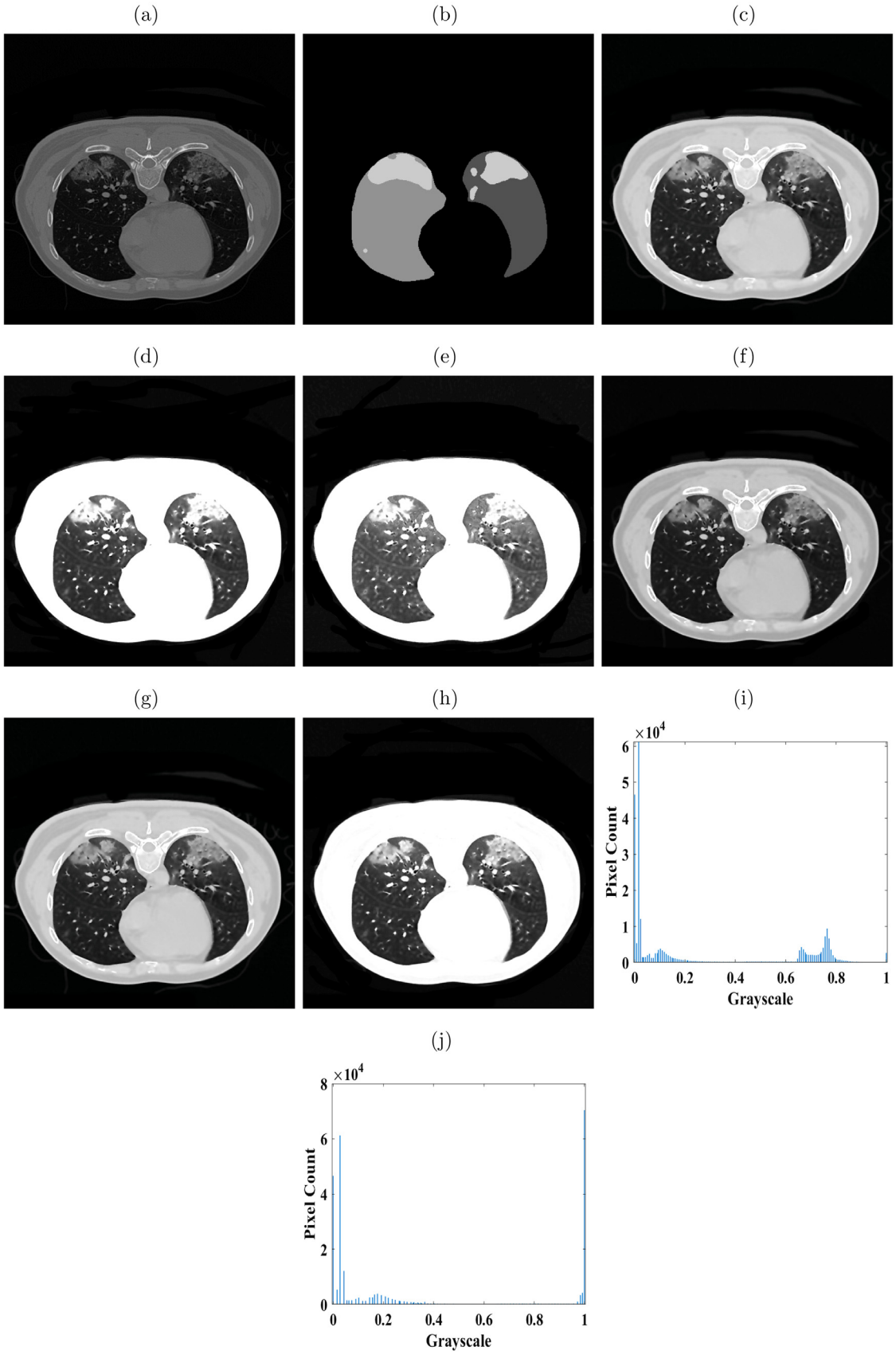

Fig. 6. Visual analysis of clustering the CT scan image (**Group: #4, Label: 8, Image: #114**) based on the ADMFCA: (a) original CT scan image, (b) ground truth of (a), (c) preprocessed CT scan image of (a), (d) TECI, (e) FECI, (f) TAECI, (g) FAECI, (h) FCI, (i) histogram of (c), and (j) histogram of (h).

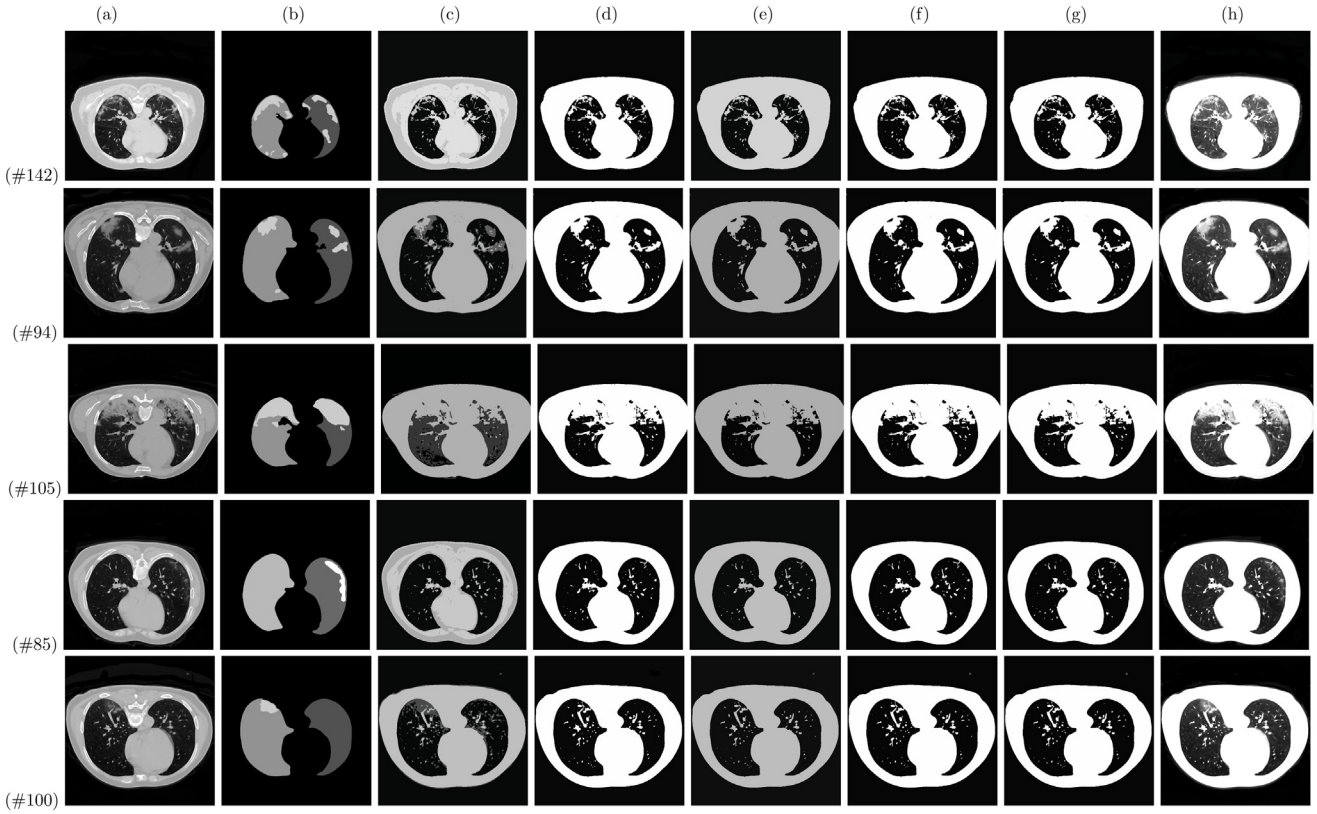

Fig. 7. Clustering of CT scan images (**Group: #1, Labels: 1–5**) of COVID-19 using the existing clustering algorithms and proposed ADMFCA: (a) preprocessed CT scan image, (b) ground truth, (c) KMC (number of clusters: 3), (d) FCM, (e) GFCM, (f) IIFCM, (g) NEBCA, and (h) ADMFCA.

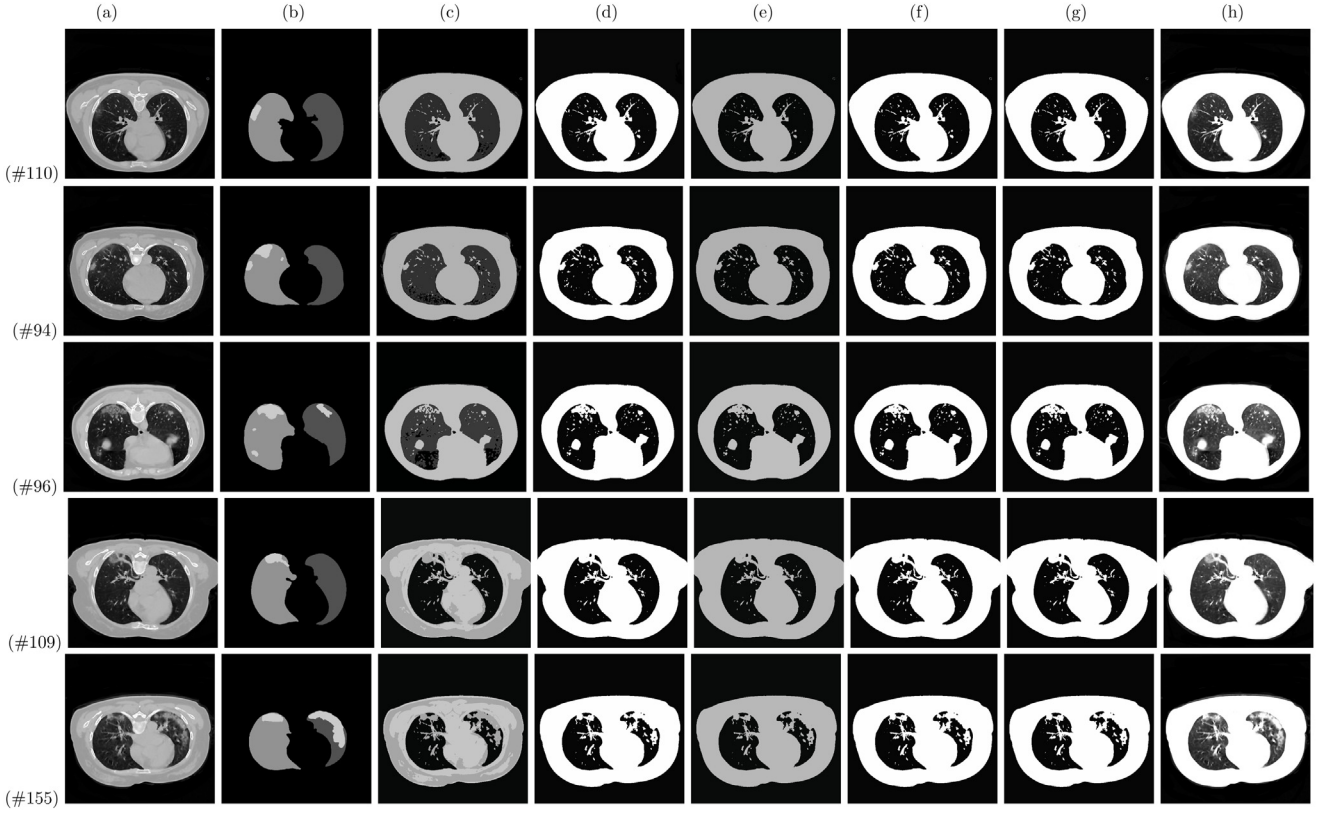

Fig. 8. Clustering of CT scan images (**Group: #1, Labels: 6–10**) of COVID-19 using the existing clustering algorithms and proposed ADMFCA: (a) preprocessed CT scan image, (b) ground truth, (c) KMC (number of clusters: 3), (d) FCM, (e) GFCM, (f) IIFCM, (g) NEBCA, and (h) ADMFCA.

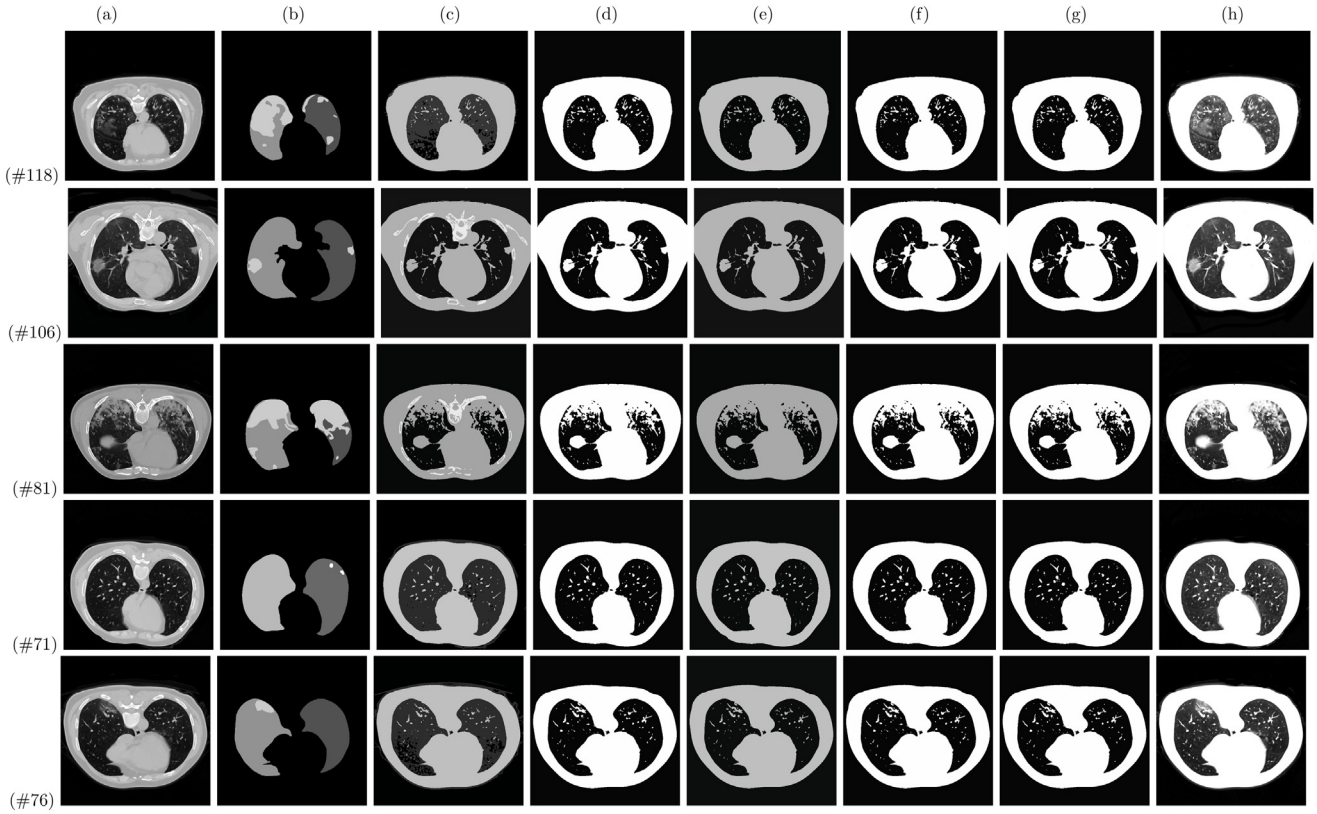

Fig. 9. Clustering of CT scan images (**Group: #2, Labels: 1–5**) of COVID-19 using the existing clustering algorithms and proposed ADMFCA: (a) preprocessed CT scan image, (b) ground truth, (c) KMC (number of clusters: 3), (d) FCM, (e) GFCM, (f) IIFCM, (g) NEBCA, and (h) ADMFCA.

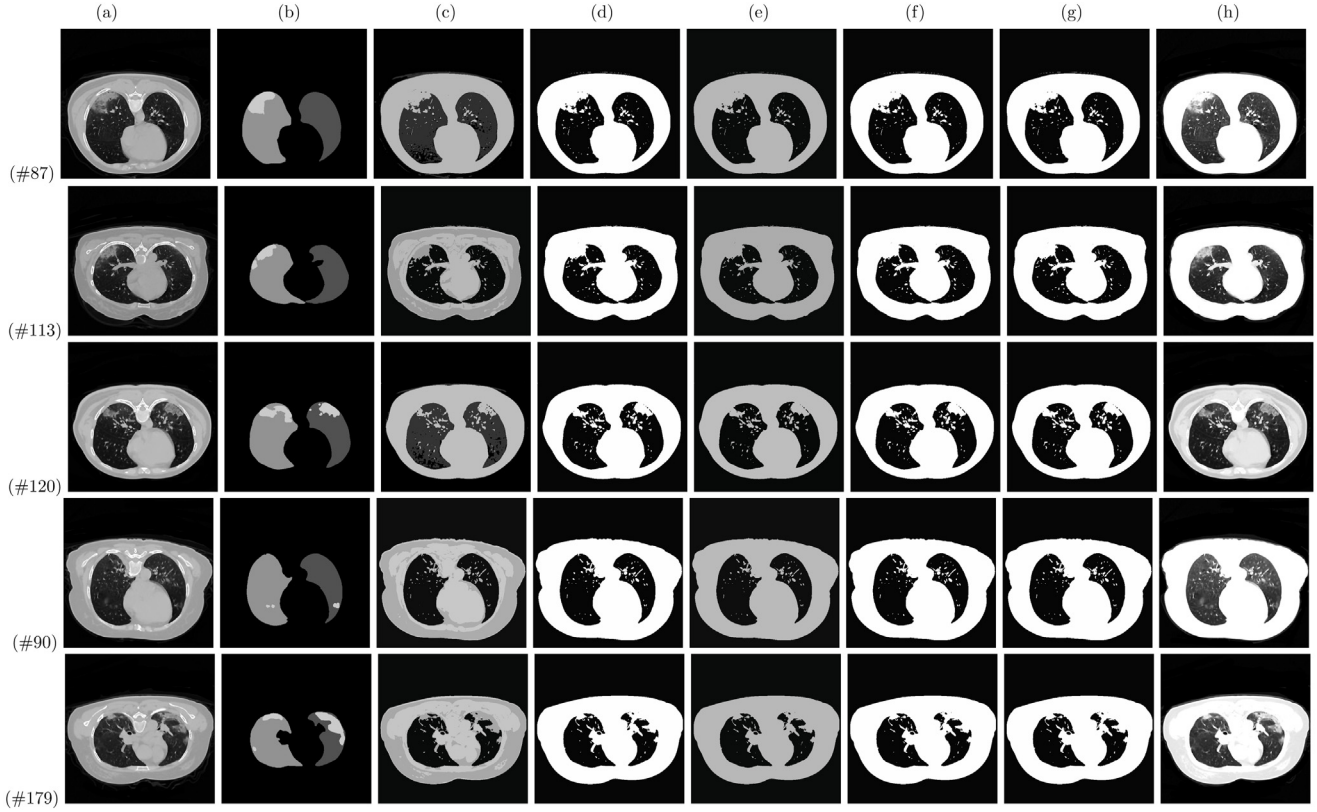

Fig. 10. Clustering of CT scan images (**Group: #2, Labels: 6–10**) of COVID-19 using the existing clustering algorithms and proposed ADMFCA: (a) preprocessed CT scan image, (b) ground truth, (c) KMC (number of clusters: 3), (d) FCM, (e) GFCM, (f) IIFCM, (g) NEBCA, and (h) ADMFCA.

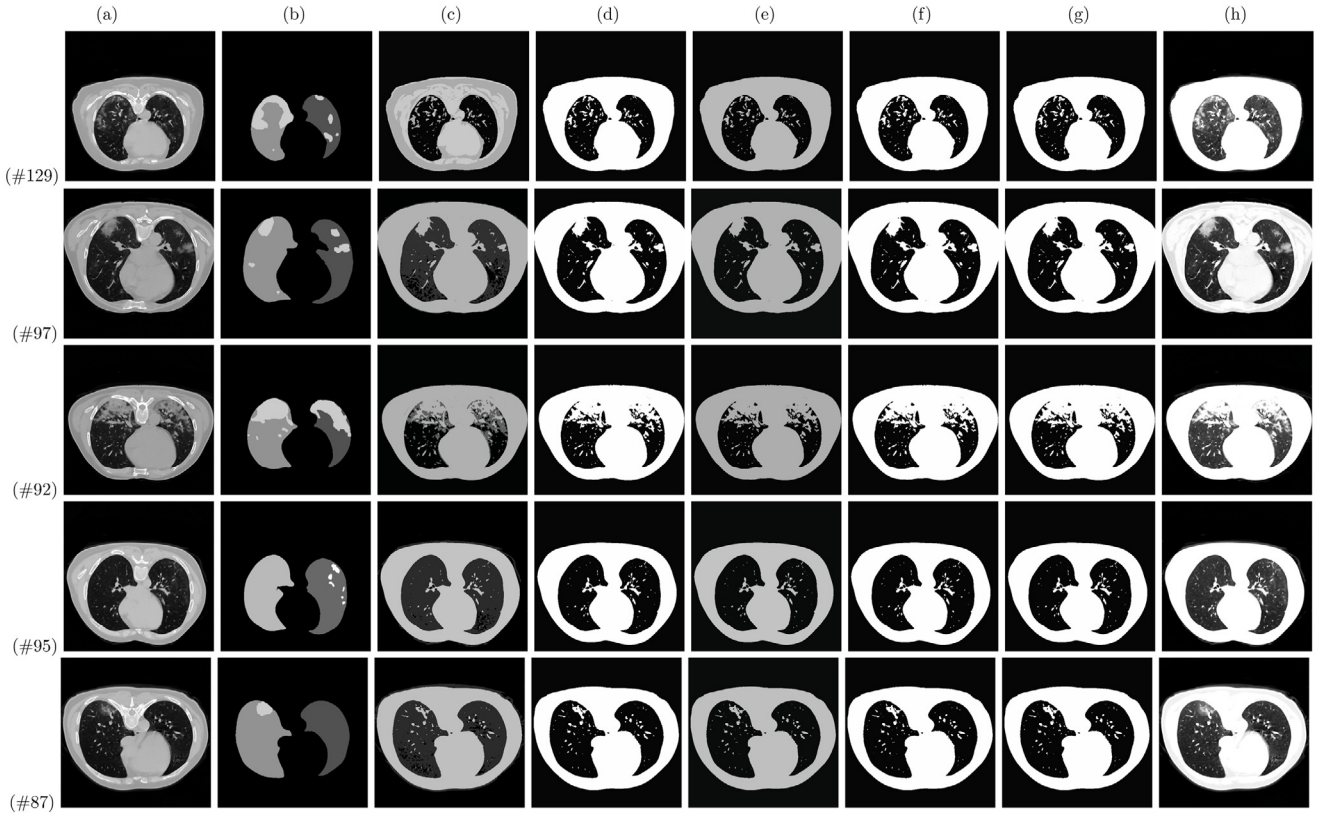

Fig. 11. Clustering of CT scan images (**Group: #3, Labels: 1–5**) of COVID-19 using the existing clustering algorithms and proposed ADMFCA: (a) preprocessed CT scan image, (b) ground truth, (c) KMC (number of clusters: 3), (d) FCM, (e) GFCM, (f) IIFCM, (g) NEBCA, and (h) ADMFCA.

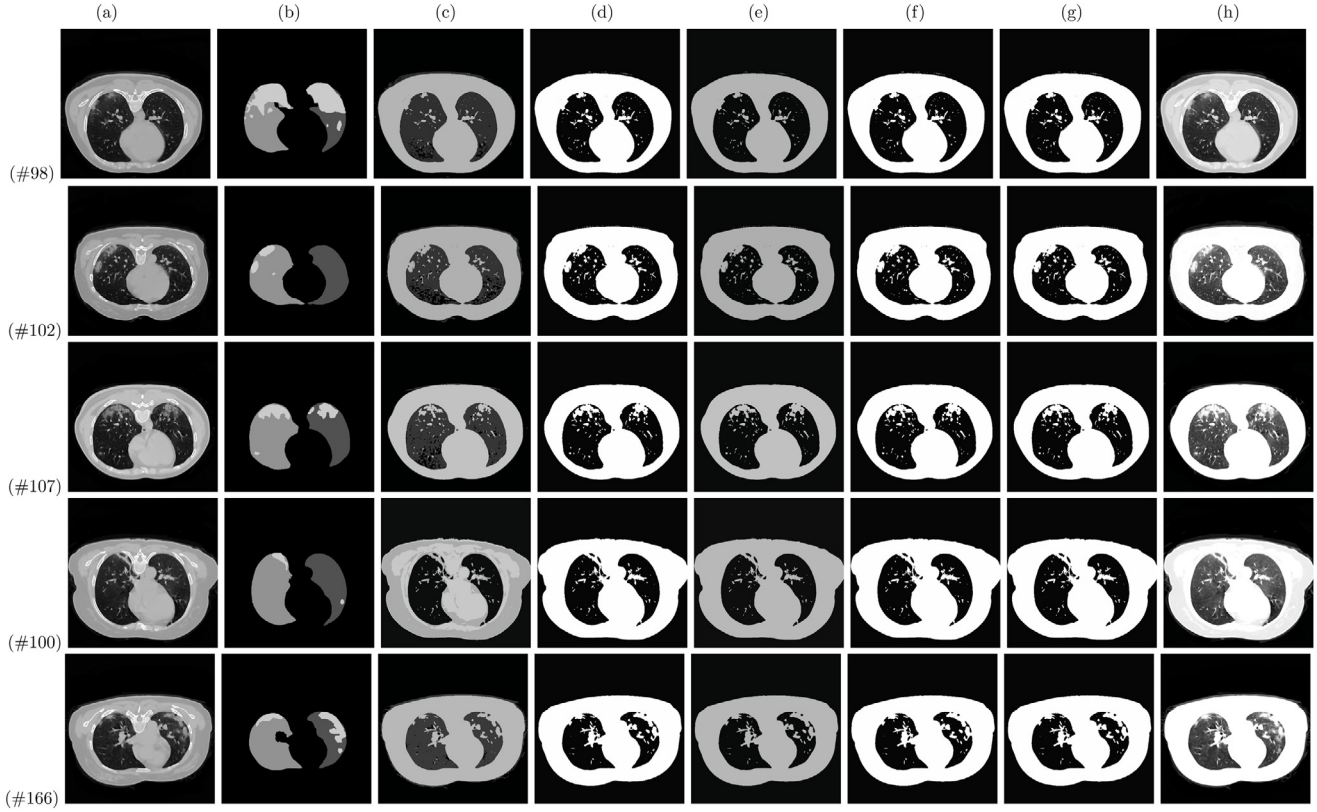

Fig. 12. Clustering of CT scan images (**Group: #3, Labels: 6–10**) of COVID-19 using the existing clustering algorithms and proposed ADMFCA: (a) preprocessed CT scan image, (b) ground truth, (c) KMC (number of clusters: 3), (d) FCM, (e) GFCM, (f) IIFCM, (g) NEBCA, and (h) ADMFCA.

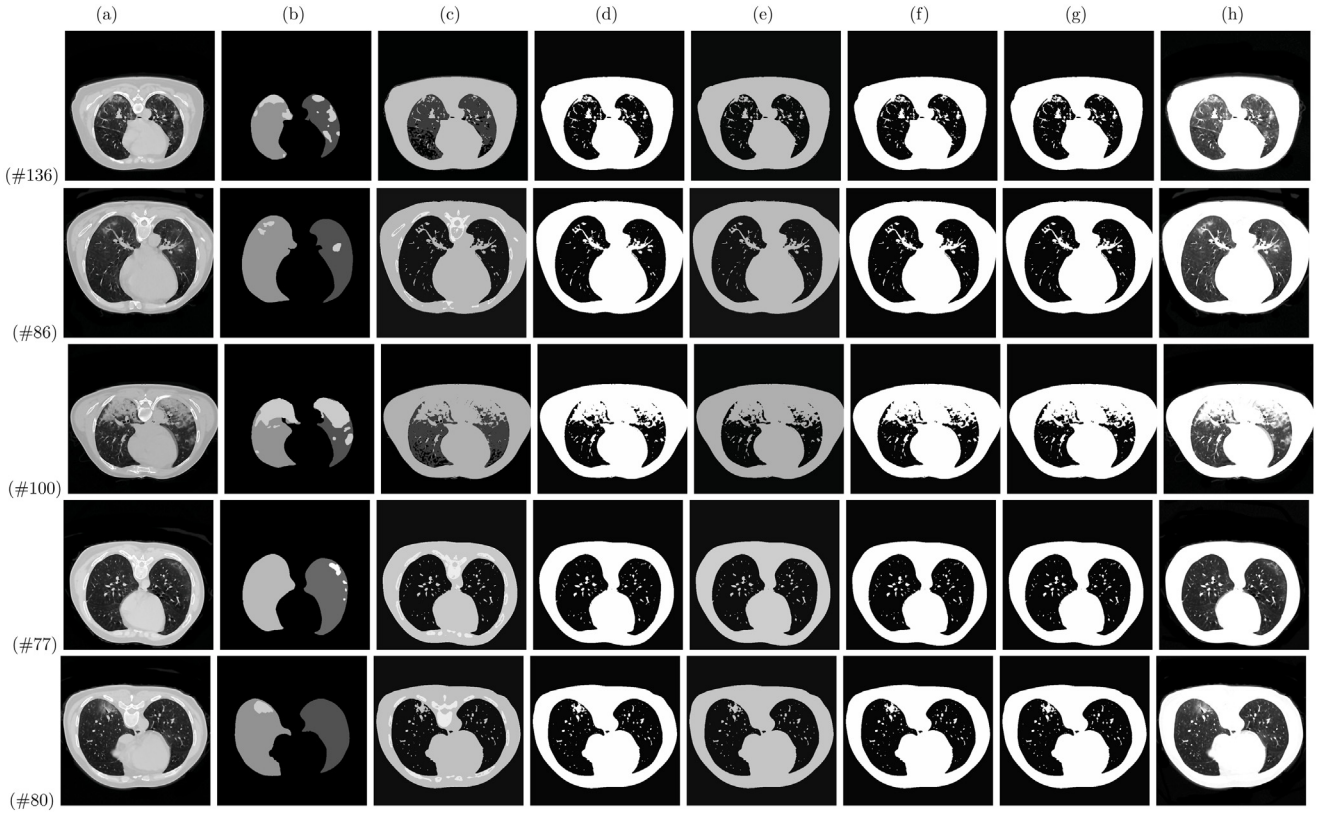

Fig. 13. Clustering of CT scan images (**Group: #4, Labels: 1–5**) of COVID-19 using the existing clustering algorithms and proposed ADMFCA: (a) preprocessed CT scan image, (b) ground truth, (c) KMC (number of clusters: 3), (d) FCM, (e) GFCM, (f) IIFCM, (g) NEBCA, and (h) ADMFCA.

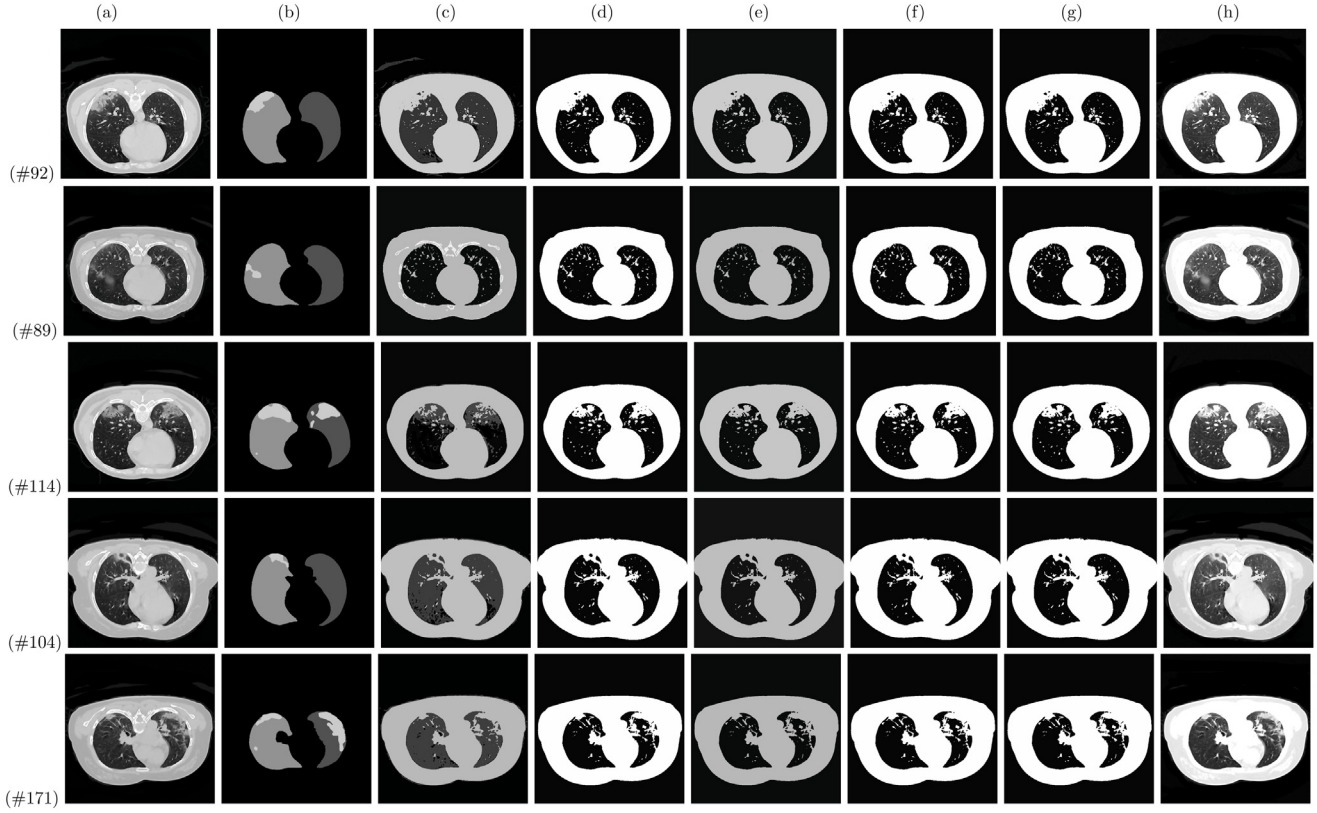

Fig. 14. Clustering of CT scan images (**Group: #4, Labels: 6–10**) of COVID-19 using the existing clustering algorithms and proposed ADMFCA: (a) preprocessed CT scan image, (b) ground truth, (c) KMC (number of clusters: 3), (d) FCM, (e) GFCM, (f) IIFCM, (g) NEBCA, and (h) ADMFCA.

Table 4
Performance evaluation of existing clustering algorithms with the proposed ADMFCA with respect to PSNR for clustering the CT scan images of COVID-19.

Group	Label	CT scan image (preprocessed)	KMC	FCM	GFCM	IIFCM	NEBCA	**ADMFCA**
Group #1	1	#142	22.05	23.28	24.32	25.59	27.36	**45.06**
	2	#94	21.70	23.25	24.51	25.72	27.41	**43.30**
	3	#105	21.46	22.18	24.36	27.16	27.52	**43.30**
	4	#85	21.20	22.21	24.42	27.12	27.48	**48.04**
	5	#100	22.01	22.17	24.25	26.73	27.55	**45.06**
	6	#110	22.01	22.17	24.25	26.73	27.55	**45.06**
	7	#94	21.03	23.20	24.35	26.99	27.44	**47.96**
	8	#96	21.92	23.10	24.90	26.92	27.52	**47.96**
	9	#109	21.73	23.19	24.75	26.77	27.53	**47.92**
	10	#155	21.47	23.18	24.79	26.77	27.53	**45.01**
Group #2	1	#118	21.47	23.18	24.79	26.77	27.53	**45.01**
	2	#106	21.16	23.19	24.78	27.08	27.48	**48.04**
	3	#81	21.84	23.15	24.26	26.78	27.39	**42.08**
	4	#71	21.90	23.13	24.34	26.79	27.56	**48.00**
	5	#76	20.39	23.11	24.78	27.23	27.48	**48.00**
	6	#87	20.40	23.15	24.39	27.13	27.48	**45.10**
	7	#113	20.40	23.15	24.39	27.13	27.48	**45.10**
	8	#120	21.59	23.11	24.36	27.12	27.51	**47.96**
	9	#90	21.74	23.18	24.46	26.90	27.61	**47.92**
	10	#179	21.20	23.15	24.39	27.10	27.45	**47.96**
Group #3	1	#129	21.94	23.20	24.33	26.82	27.36	**47.96**
	2	#97	21.79	23.19	24.51	27.29	27.71	**47.92**
	3	#92	20.62	23.32	24.36	27.05	27.59	**47.88**
	4	#95	20.47	23.13	24.80	27.19	27.52	**47.96**
	5	#87	21.05	22.82	24.86	27.02	27.55	**47.92**
	6	#98	21.36	23.19	23.87	27.02	27.47	**48.00**
	7	#102	21.99	23.20	24.73	27.09	27.41	**45.01**
	8	#107	21.95	23.02	24.32	26.95	27.67	**43.35**
	9	#100	21.74	23.05	24.39	26.70	27.38	**47.96**
	10	#166	21.97	23.05	24.24	26.77	27.53	**45.08**
Group #4	1	#136	21.84	23.06	24.33	27.02	27.37	**45.03**
	2	#86	21.98	22.92	24.49	27.15	27.71	**45.06**
	3	#100	21.46	23.18	24.18	27.05	27.59	**42.06**
	4	#77	20.47	22.93	24.40	26.70	27.52	**48.04**
	5	#80	21.04	22.79	24.47	27.05	27.55	**48.00**
	6	#92	21.34	22.92	24.21	26.88	27.47	**47.88**
	7	#89	21.96	23.02	24.73	26.69	27.40	**48.04**
	8	#114	21.92	23.16	24.34	26.69	27.67	**45.01**
	9	#104	21.73	22.79	24.39	26.96	27.38	**48.00**
	10	#171	21.97	22.92	24.81	26.78	27.56	**45.01**
Average	–	–	21.49	23.03	24.47	26.89	27.51	**46.59**

6.5. Comparison of existing clustering methods with the proposed ADMFCA

The preprocessed CT scan images of COVID-19 patients are shown in Figs. 7–14(a) column-wise. Figs. 7–14(b) show the respective ground truth of Figs. 7–14(a). Various clustered images from the existing clustering methods, such as KMC [23], FCM [29], GFCM [34], IIFCM [36] and NEBCA [38] are shown column-wise in Figs. 7–14(c)∼(g), respectively. Figs. 7–14(h) show the FCIs obtained with the proposed ADMFCA.

It is clear from the clustered images presented in Figs. 7–14(c)∼(g) that CT scan images and infected regions are not properly clustered using the existing clustering methods [23,29,34,36,38]. While comparing these images with the FCIs, it can be seen that the proposed ADMFCA appropriately clusters not only the CT scan images, but also the grayscale values of the infected regions. The clustered images obtained by existing clustering methods, as shown in Figs. 7–14(c)∼(g), demonstrate that these methods are unable to process these images due to the ambiguous and unclear boundaries.

Finally, the performance of the proposed ADMFCA is compared with clustering methods, such as KMC, FCM, GFCM, IIFCM and NEBCA using five statistical metrics, namely MSE, PSNR, DSC, JSC and CC. The MSE, PSNR, DSC, JSC and CC values of the existing clustering methods [23,29,34,36,38] and proposed ADMFCA are presented in Tables 3–7, respectively. The average values of MSE, PSNR, DSC, JSC and CC are obtained and shown in the last row of each table. A discussion is carried out on these statistical values next.

- Table 3 shows comparable results with respect to MSE values. The average MSE values of KMC, FCM, GFCM, IIFCM and NEBCA are 464.38, 324.11, 232.91, 133.42 and 115.50 respectively. The proposed ADMFCA, on the other hand, has an average MSE value of 1.59, which is significantly lower than the existing clustering approaches, such as KMC, FCM, GFCM, IIFCM and NEBCA. This low MSE value for the proposed ADMFCA indicates that it can produce high quality FCIs with minimal intensity loss.
- Table 4 shows the comparison values of PSNR for the existing clustering methods and the proposed ADMFCA. The average PSNR values for KMC, FCM, GFCM, IIFCM and NEBCA are 21.49, 23.03, 24.47, 26.89 and 27.51, respectively. The proposed ADMFCA, on the other hand, has a PSNR value of 46.59, which is significantly higher than the existing clustering methods. This high PSNR value for the proposed ADMFCA demonstrates that it can produce FCIs with very low grayscale deformation.
- Table 5 shows the statistics of DSC values for the existing methods and the proposed ADMFCA. The proposed ADMFCA

Table 5
Performance evaluation of existing clustering algorithms with the proposed ADMFCA with respect to DSC for clustering the CT scan images of COVID-19.

Group	Label	CT scan image (preprocessed)	KMC	FCM	GFCM	IIFCM	NEBCA	**ADMFCA**
Group #1	1	#142	0.44	0.50	0.65	0.78	0.82	**0.90**
	2	#94	0.40	0.51	0.66	0.76	0.81	**0.91**
	3	#105	0.37	0.54	0.68	0.77	0.83	0.89
	4	#85	0.39	0.52	0.65	0.78	0.84	0.93
	5	#100	0.40	0.53	0.63	0.77	0.83	0.94
	6	#110	0.44	0.49	0.67	0.75	0.84	0.92
	7	#94	0.45	0.58	0.64	0.78	0.85	0.93
	8	#96	0.39	0.53	0.63	0.77	0.83	0.90
	9	#109	0.39	0.54	0.64	0.78	0.84	0.94
	10	#155	0.39	0.59	0.63	0.78	0.84	0.87
Group #2	1	#118	0.39	0.49	0.68	0.78	0.90	0.91
	2	#106	0.39	0.49	0.65	0.77	0.85	0.93
	3	#81	0.40	0.51	0.64	0.78	0.83	0.91
	4	#71	0.45	0.50	0.62	0.75	0.87	0.93
	5	#76	0.39	0.49	0.65	0.78	0.83	0.94
	6	#87	0.40	0.51	0.66	0.75	0.89	0.91
	7	#113	0.39	0.52	0.64	0.77	0.86	0.93
	8	#120	0.42	0.58	0.65	0.76	0.81	0.93
	9	#90	0.45	0.47	0.66	0.78	0.84	0.92
	10	#179	0.43	0.46	0.64	0.77	0.85	0.92
Group #3	1	#129	0.37	0.49	0.66	0.79	0.84	0.91
	2	#97	0.39	0.52	0.65	0.78	0.85	0.93
	3	#92	0.42	0.54	0.66	0.74	0.83	0.89
	4	#95	0.45	0.54	0.64	0.76	0.86	0.93
	5	#87	0.39	0.52	0.62	0.75	0.84	0.92
	6	#98	0.39	0.59	0.66	0.77	0.86	0.89
	7	#102	0.38	0.48	0.65	0.75	0.84	0.94
	8	#107	0.37	0.53	0.66	0.77	0.82	0.89
	9	#100	0.37	0.47	0.66	0.74	0.84	0.93
	10	#166	0.37	0.49	0.64	0.77	0.85	0.91
Group #4	1	#136	0.37	0.50	0.63	0.79	0.87	0.92
	2	#86	0.38	0.48	0.65	0.78	0.84	0.92
	3	#100	0.45	0.56	0.64	0.79	0.82	0.86
	4	#77	0.39	0.58	0.63	0.76	0.88	0.93
	5	#80	0.40	0.47	0.66	0.78	0.84	0.92
	6	#92	0.46	0.57	0.65	0.77	0.87	0.92
	7	#89	0.46	0.49	0.64	0.78	0.88	0.94
	8	#114	0.45	0.57	0.63	0.79	0.86	0.87
	9	#104	0.46	0.48	0.64	0.77	0.83	0.94
	10	#171	0.38	0.47	0.65	0.78	0.86	0.92
Average	–	–	0.41	0.52	0.65	0.77	0.85	**0.92**

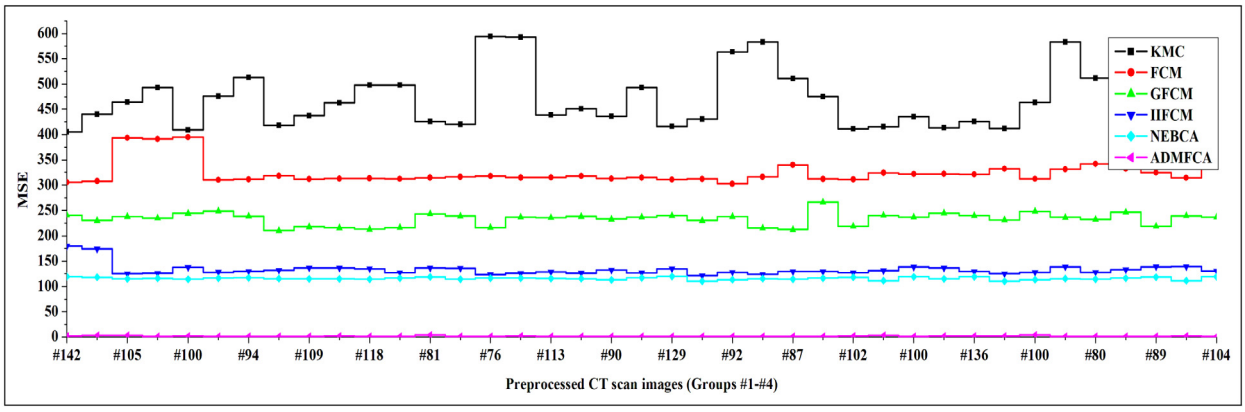

Fig. 15. Comparison of MSE values with the existing clustering algorithms and the proposed ADMFCA.

yields an average DSC value of 0.92, which is significantly higher than the existing clustering methods. This high DSC value indicates that the FCIs produced by the proposed ADMFCA are identical to their respective ground truths.

- The JSC values for the existing competing methods and the proposed ADMFCA are shown in Table 6. The proposed ADMFCA obtains an average JSC value of 0.96 for the FCIs, which is significantly higher than the existing clustering methods. For the proposed ADMFCA, the high JSC value

Table 6
Performance evaluation of existing clustering algorithms with the proposed ADMFCA with respect to JSC for clustering the CT scan images of COVID-19.

Group	Label	CT scan image (preprocessed)	KMC	FCM	GFCM	IIFCM	NEBCA	**ADMFCA**
Group #1	1	#142	0.48	0.54	0.69	0.82	0.86	**0.94**
	2	#94	0.44	0.55	0.70	0.80	0.85	**0.95**
	3	#105	0.41	0.58	0.72	0.81	0.87	**0.93**
	4	#85	0.43	0.56	0.69	0.82	0.88	**0.97**
	5	#100	0.44	0.57	0.67	0.81	0.87	**0.98**
	6	#110	0.48	0.53	0.71	0.79	0.88	**0.96**
	7	#94	0.49	0.62	0.68	0.82	0.89	**0.97**
	8	#96	0.43	0.57	0.67	0.81	0.87	**0.94**
	9	#109	0.43	0.58	0.68	0.82	0.88	**0.98**
	10	#155	0.43	0.63	0.67	0.82	0.88	**0.91**
Group #2	1	#118	0.43	0.53	0.72	0.82	0.94	**0.95**
	2	#106	0.43	0.53	0.69	0.81	0.89	**0.97**
	3	#81	0.40	0.51	0.64	0.78	0.83	**0.91**
	4	#71	0.49	0.54	0.66	0.79	0.91	**0.97**
	5	#76	0.43	0.53	0.69	0.82	0.87	**0.98**
	6	#87	0.44	0.55	0.70	0.79	0.93	**0.95**
	7	#113	0.43	0.56	0.68	0.81	0.90	**0.97**
	8	#120	0.46	0.62	0.69	0.80	0.85	**0.97**
	9	#90	0.49	0.51	0.70	0.82	0.88	**0.96**
	10	#179	0.47	0.50	0.68	0.81	0.89	**0.96**
Group #3	1	#129	0.41	0.53	0.70	0.83	0.88	**0.95**
	2	#97	0.43	0.56	0.69	0.82	0.89	**0.97**
	3	#92	0.46	0.58	0.70	0.78	0.87	**0.93**
	4	#95	0.49	0.58	0.68	0.80	0.90	**0.97**
	5	#87	0.43	0.56	0.66	0.79	0.88	**0.96**
	6	#98	0.43	0.63	0.70	0.81	0.90	**0.93**
	7	#102	0.42	0.52	0.69	0.79	0.88	**0.98**
	8	#107	0.41	0.57	0.70	0.81	0.86	**0.93**
	9	#100	0.41	0.51	0.70	0.78	0.88	**0.97**
	10	#166	0.41	0.53	0.68	0.81	0.89	**0.95**
Group #4	1	#136	0.41	0.54	0.67	0.83	0.91	**0.96**
	2	#86	0.42	0.52	0.69	0.82	0.88	**0.96**
	3	#100	0.49	0.60	0.68	0.83	0.86	**0.90**
	4	#77	0.43	0.62	0.67	0.80	0.92	**0.97**
	5	#80	0.44	0.51	0.70	0.82	0.88	**0.96**
	6	#92	0.50	0.61	0.69	0.81	0.91	**0.96**
	7	#89	0.50	0.53	0.68	0.82	0.92	**0.98**
	8	#114	0.50	0.53	0.68	0.82	0.92	**0.98**
	9	#104	0.50	0.52	0.68	0.81	0.87	**0.98**
	10	#171	0.42	0.51	0.69	0.82	0.90	**0.96**
Average	–	–	0.45	0.56	0.69	0.81	0.89	**0.96**

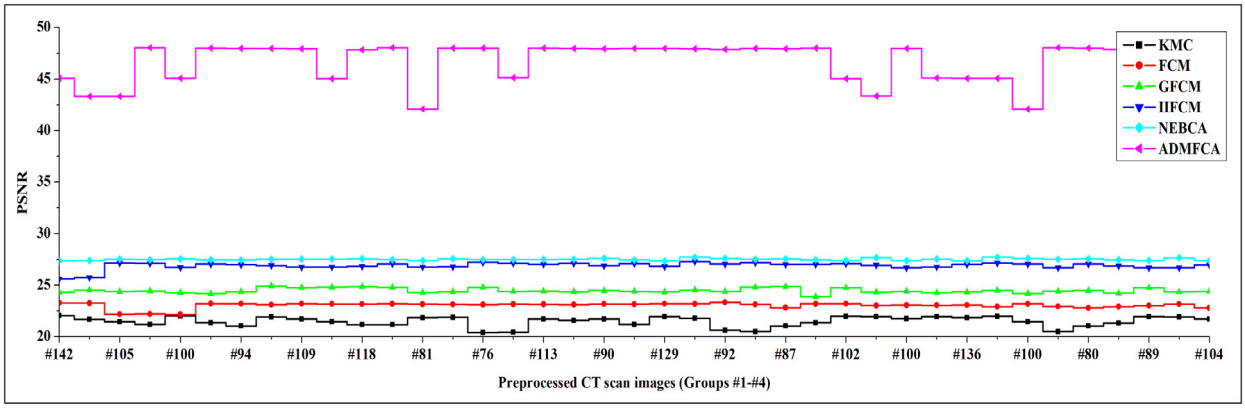

Fig. 16. Comparison of PSNR values with the existing clustering algorithms and the proposed ADMFCA.

indicates that the regions of interest of the FCIs are almost identical to their respective ground truths.

- Table 7 summarizes the CC values obtained from existing methods and the proposed ADMFCA. The average CC values for KMC, FCM, GFCM, IIFCM, NEBCA and the proposed ADMFCA are 0.46, 0.57, 0.70, 0.82, 0.90 and 0.97, respectively. This CC value of the proposed ADMFCA suggests that the clustered grayscale values of the FCIs are strongly similar to their respective ground truths.

Comparison curves for various MSE, PSNR, DSC, JSC and CC (listed in Tables 3–7) are plotted and shown in Figs. 15–19, respectively. These comparison curves show that the proposed ADMFCA outperforms the existing clustering methods (i.e., KMC, FCM, GFCM, IIFCM and NEBCA) in terms of MSE, PSNR, DSC,

Table 7
Performance evaluation of existing clustering algorithms with the proposed ADMFCA with respect to CC for clustering the CT scan images of COVID-19.

Group	Label	CT scan image (preprocessed)	KMC	FCM	GFCM	IIFCM	NEBCA	**ADMFCA**
Group #1	1	#142	0.49	0.55	0.70	0.83	0.87	0.95
	2	#94	0.45	0.56	0.71	0.81	0.86	0.96
	3	#105	0.42	0.59	0.73	0.82	0.88	0.94
	4	#85	0.44	0.57	0.70	0.83	0.89	0.98
	5	#100	0.45	0.58	0.68	0.82	0.88	0.99
	6	#110	0.49	0.54	0.72	0.80	0.89	0.97
	7	#94	0.50	0.63	0.69	0.83	0.90	0.98
	8	#96	0.44	0.58	0.68	0.82	0.88	0.95
	9	#109	0.44	0.59	0.69	0.83	0.89	0.99
	10	#155	0.44	0.64	0.68	0.83	0.89	0.92
Group #2	1	#118	0.44	0.54	0.73	0.83	0.95	0.96
	2	#106	0.44	0.54	0.70	0.82	0.90	0.98
	3	#81	0.45	0.56	0.69	0.83	0.88	0.96
	4	#71	0.50	0.55	0.67	0.80	0.92	0.98
	5	#76	0.44	0.54	0.70	0.83	0.88	0.99
	6	#87	0.45	0.56	0.71	0.80	0.94	0.96
	7	#113	0.44	0.57	0.69	0.82	0.91	0.98
	8	#120	0.47	0.63	0.70	0.81	0.86	0.98
	9	#90	0.50	0.52	0.71	0.83	0.89	0.97
	10	#179	0.48	0.51	0.69	0.82	0.90	0.97
Group #3	1	#129	0.42	0.54	0.71	0.84	0.89	0.96
	2	#97	0.44	0.57	0.70	0.83	0.90	0.98
	3	#92	0.47	0.59	0.71	0.79	0.88	0.94
	4	#95	0.50	0.59	0.69	0.81	0.91	0.98
	5	#87	0.44	0.57	0.67	0.80	0.89	0.97
	6	#98	0.44	0.64	0.71	0.82	0.91	0.94
	7	#102	0.43	0.53	0.70	0.80	0.89	0.99
	8	#107	0.42	0.58	0.71	0.82	0.87	0.94
	9	#100	0.42	0.52	0.71	0.79	0.89	0.98
	10	#166	0.42	0.54	0.69	0.82	0.90	0.96
Group #4	1	#136	0.42	0.55	0.68	0.84	0.92	0.97
	2	#86	0.43	0.53	0.70	0.83	0.89	0.97
	3	#100	0.50	0.61	0.69	0.84	0.87	0.91
	4	#77	0.44	0.63	0.68	0.81	0.93	0.98
	5	#80	0.45	0.52	0.71	0.83	0.89	0.97
	6	#92	0.51	0.62	0.70	0.82	0.92	0.97
	7	#89	0.51	0.54	0.69	0.83	0.93	0.99
	8	#114	0.50	0.62	0.68	0.84	0.91	0.92
	9	#104	0.51	0.53	0.69	0.82	0.88	0.99
	10	#171	0.43	0.52	0.70	0.83	0.91	0.97
Average	–	–	0.46	0.57	0.70	0.82	0.90	**0.97**

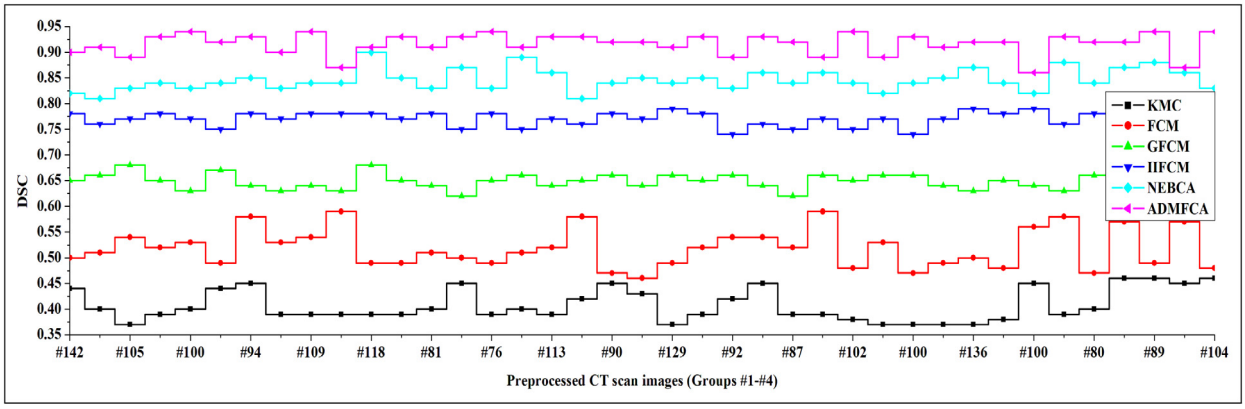

Fig. 17. Comparison of DSC values with the existing clustering algorithms and the proposed ADMFCA.

JSC and CC for clustering the CT scan images of COVID-19 patients. Consequently, the proposed ADMFCA is highly effective at forming clusters of pixels associated with infected regions.

7. Conclusions and future directions

In this study, ambiguous set theory was discussed, which was recently proposed to address inherent uncertainties of events. The ambiguous set theory can be considered as an extension of three existing theories, viz., fuzzy set, intuitionistic fuzzy set and neutrosophic set. The main robustness of this theory was its ability to represent the ambiguity of an uncertain event with four distinct degree of memberships, called true, false, true-ambiguous and false-ambiguous. To endorse this theory, various definitions,

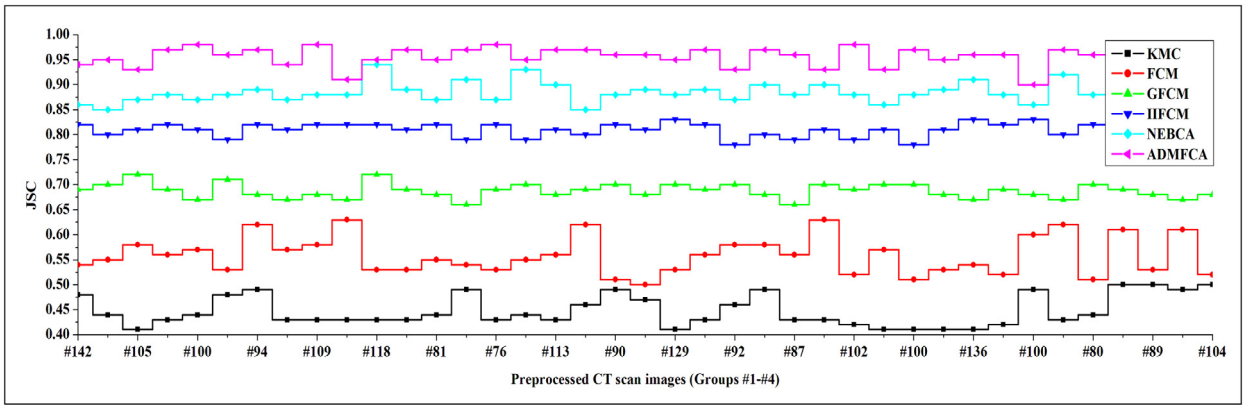

Fig. 18. Comparison of JSC values with the existing clustering algorithms and the proposed ADMFCA.

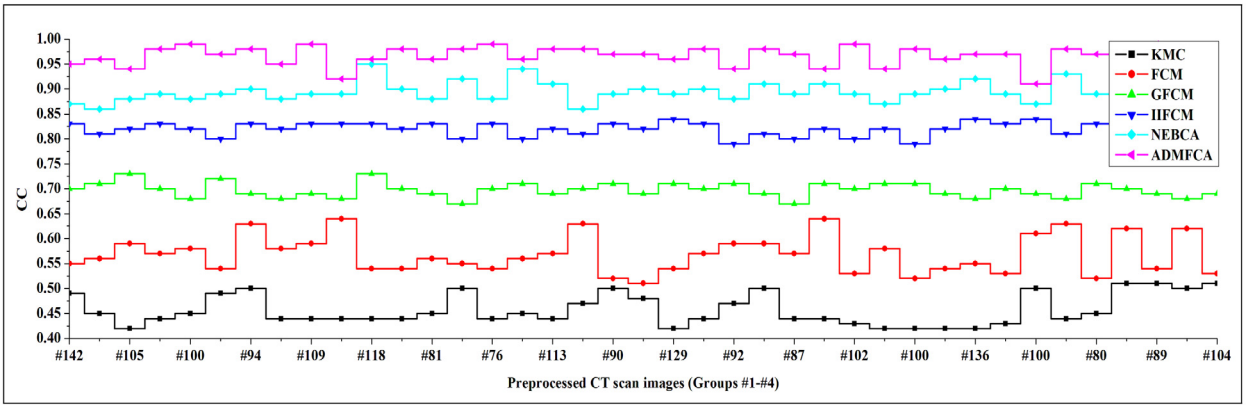

Fig. 19. Comparison of CC values with the existing clustering algorithms and the proposed ADMFCA.

formulas and properties were discussed in this study. The main contributions of this study are summarized as:

- To measure the ambiguity associated with four degree of memberships, four different entropies were defined, called TE, FE, TAE and FAE.
- This study proposed a new image clustering algorithm using the concepts of ambiguous set, entropies (TE, FE, TAE and FAE) and image fusion, called ADMFCA.
- The primary application of the proposed ADMFCA was illustrated in the clustering of chest CT images of COVID-19 patients. This algorithm was allowed to generate four different clustered images based on specified number of clusters. These four clustered images were referred to as TECI, FECI, TAECI and FAECI. Finally, FCIs were generated by combining TECI, FECI, TAECI and FAECI using the image fusion operation. The main purpose of image fusion was to include the best features of TECI, FECI, TAECI and FAECI into FCIs.
- The performance of the proposed ADMFCA was compared against existing clustering methods, including KMC, FCM, FCM, FCM and NEBCA. Various performance evaluation metrics, such as MSE, PSNR, DSC, JSC and CC indicated that the proposed ADFMCA outperformed the existing clustering methods.

It can be concluded that the proposed ADMFCA was proven to be effective in clustering CT scan images of COVID-19 patients. Therefore, the proposed ADMFCA can be considered as a new promising diagnostic method for health professionals. The main limitation of the study was that the proposed ADMFCA was validated only on chest CT scan images of COVID-19 patients. In the future, the proposed ADMFCA can be verified and validated with other forms of digital images, such as X-rays, MRIs [40], remotely sensed high-resolution satellite images [48], and so on. Additionally, the proposed ADMFCA can be used to cluster a variety of numerical data, including meteorological data, financial data, stock market data, and so on.

Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

CRediT authorship contribution statement

Pritpal Singh: Conceptualization, Software, Methodology, Theories and definitions, Mathematical deductions and their proofs, Writing - review & editing. **Surya Sekhar Bose:** Conceptualization, Software, Methodology, Theories and definitions, Mathematical deductions and their proofs, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Y.Y. Zheng, et al., Covid-19 and the cardiovascular system, Nat. Rev. Cardiol. 17 (2020) 259–260.
[2] P. Singh, S.S. Bose, A quantum-clustering optimization method for COVID-19 CT scan image segmentation, Expert Syst. Appl. 185 (2021) 115637.

[3] WHO, Covid-19 weekly epidemiological update, Technical Report, 2021.
[4] S. Boccaletti, W. Ditto, G. Mindlin, A. Atangana, Modeling and forecasting of epidemic spreading: The case of Covid-19 and beyond, Chaos Solitons Fractals 135 (2020) 109794.
[5] S. Flaxman, et al., Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe, Nature 584 (2020) 257–261.
[6] O. Castillo, P. Melin, Forecasting of COVID-19 time series for countries in the world based on a hybrid approach combining the fractal dimension and fuzzy logic, Chaos Solitons Fractals 140 (2020) 110242.
[7] O. Castillo, P. Melin, A novel method for a COVID-19 classification of countries based on an intelligent fuzzy fractal approach, Healthcare 9 (2021) 1–15.
[8] H. Harapan, et al., Coronavirus disease 2019 (COVID-19): A literature review, J. Infect. Public Health 13 (2020) 667–673.
[9] P. Zhou, et al., A pneumonia outbreak associated with a new coronavirus of probable bat origin, Nature 579 (2020) 270–273.
[10] A. Altan, S. Karasu, Recognition of COVID-19 disease from X-ray images by hybrid model consisting of 2D curvelet transform, chaotic salp swarm algorithm and deep learning technique, Chaos Solitons Fractals 140 (2020) 110071.
[11] I.D. Apostolopoulos, T.A. Mpesiana, Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks, Phys. Eng. Sci. Med. 43 (2020) 635–640.
[12] A.A. Ardakani, et al., Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks, Comput. Biol. Med. 121 (2020) 103795.
[13] T. Ozturk, et al., Automated detection of COVID-19 cases using deep neural networks with X-ray images, Comput. Biol. Med. 121 (2020) 103792.
[14] S. Vaid, R. Kalantar, M. Bhandari, Deep learning COVID-19 detection bias: accuracy through artificial intelligence, Int. Orthop. 44 (2020) 1539–1542.
[15] S. Toraman, T.B. Alakus, I. Turkoglu, Convolutional capsnet: A novel artificial neural network approach to detect COVID-19 disease from X-ray images using capsule networks, Chaos Solitons Fractals 140 (2020) 110122.
[16] M. Nour, Z. Cömert, K. Polat, A novel medical diagnosis model for COVID-19 infection detection based on deep features and bayesian optimization, Appl. Soft Comput. 97 (2020) 106580.
[17] H. Kang, et al., Diagnosis of coronavirus disease 2019 (COVID-19) with structured latent multi-view representation learning, IEEE Trans. Med. Imag. 39 (2020) 2606–2614.
[18] X. Wang, et al., A weakly-supervised framework for COVID-19 classification and lesion localization from chest CT, IEEE Trans. Med. Imag. 39 (2020) 2615–2625.
[19] S. Varela-Santos, P. Melin, A new approach for classifying coronavirus COVID-19 based on its manifestation on chest X-rays using texture features and neural networks, Inform. Sci. 545 (2021) 403–414.
[20] M. Tang, H. Liao, Managing information measures for hesitant fuzzy linguistic term sets and their applications in designing clustering algorithms, Inf. Fusion 50 (2019) 30–42.
[21] Z. Deng, K.-S. Choi, F.-L. Chung, S. Wang, Enhanced soft subspace clustering integrating within-cluster and between-cluster information, Pattern Recognit. 43 (2010) 767–781.
[22] A.S. Shirkhorshidi, T.Y. Wah, S.M.R. Shirkhorshidi, S. Aghabozorgi, Evolving fuzzy clustering approach: An epoch clustering that enables heuristic postpruning, IEEE Trans. Fuzzy Syst. 29 (2021) 560–568.
[23] H. Yao, Q. Duan, D. Li, J. Wang, An improved K-means clustering algorithm for fish image segmentation, Math. Comput. Modelling 58 (2013) 790–798.
[24] Y.-P. Huang, P. Singh, H.-C. Kuo, A hybrid fuzzy clustering approach for the recognition and visualization of MRI images of Parkinson's disease, 27, 2020, pp. 25041–25051.
[25] P. Xu, Z. Deng, K.-S. Choi, L. Cao, S. Wang, Multi-view information-theoretic co-clustering for co-occurrence data, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, 2019, pp. 379–386.
[26] L.A. Zadeh, Fuzzy sets, Inf. Control 8 (1965) 338–353.
[27] K.T. Atanassov, Intuitionistic fuzzy sets, in: Proc. VII ITKR's Session, Sofia, Bulgaria, 1983, pp. 16840–1697.
[28] F. Smarandache, Neutrosophy, a new branch of philosophy, Infinite Study (2002).
[29] L. Ma, R. Staunton, A modified fuzzy c-means image segmentation algorithm for use with uneven illumination patterns, Pattern Recognit. 40 (2007) 3005–3011.
[30] P. Qian, et al., Knowledge-leveraged transfer fuzzy C-Means for texture image segmentation with self-adaptive cluster prototype matching, Knowl.-Based Syst. 130 (2017) 33–50.
[31] L. Chen, C.L.P. Chen, M. Lu, A multiple-kernel fuzzy c-means algorithm for image segmentation, IEEE Trans. Syst. Man Cybern. 41 (2011) 1263–1274.
[32] Z. Ji, Y. Xia, Q. Chen, Q. Sun, D. Xia, D.D. Feng, Fuzzy c-means clustering with weighted image patch for image segmentation, Appl. Soft Comput. 12 (2012) 1659–1667.
[33] Z. Wang, Q. Song, Y.C. Soh, K. Sim, An adaptive spatial information-theoretic fuzzy clustering algorithm for image segmentation, Comput. Vis. Image Underst. 117 (2013) 1412–1420.
[34] F. Zhao, L. Jiao, H. Liu, Kernel generalized fuzzy c-means clustering with spatial information for image segmentation, Digit. Signal Process. 23 (2013) 184–199.
[35] T. Chaira, A novel intuitionistic fuzzy c means clustering algorithm and its application to medical images, Appl. Soft Comput. 11 (2011) 1711–1717.
[36] H. Verma, R. Agrawal, A. Sharan, An improved intuitionistic fuzzy c-means clustering algorithm incorporating local information for brain image segmentation, Appl. Soft Comput. 46 (2016) 543–557.
[37] Z. Deng, Y. Jiang, F.-L. Chung, H. Ishibuchi, K.-S. Choi, S. Wang, Transfer prototype-based fuzzy clustering, IEEE Trans. Fuzzy Syst. 24 (2016b) 1210–1232.
[38] P. Singh, A neutrosophic-entropy based clustering algorithm (NEBCA) with HSV color system: A special application in segmentation of parkinson's disease (PD) MR images, Comput. Methods Programs Biomed. 189 (2020b) 105317.
[39] P. Singh, A neutrosophic-entropy based adaptive thresholding segmentation algorithm: A special application in MR images of Parkinson's disease, Artif. Intell. Med. 104 (2020a) 101838.
[40] P. Singh, A type-2 neutrosophic-entropy-fusion based multiple thresholding method for the brain tumor tissue structures segmentation, Appl. Soft Comput. 103 (2021) 107119.
[41] P. Singh, Y.-P. Huang, T.-T. Lee, A novel ambiguous set theory to represent uncertainty and its application to brain MR image segmentation, in: Proc. of IEEE Int. Conf. on Systems, Man and Cybernetics (SMC), Bari, Italy, 2019, pp. 2460–2465.
[42] H. Wang, Y. Zhang, R. Nie, Y. Yang, B. Peng, T. Li, Bayesian image segmentation fusion, Knowl.-Based Syst. 71 (2014) 162–168.
[43] Z. Deng, K.-S. Choi, Y. Jiang, J. Wang, S. Wang, A survey on soft subspace clustering, Inform. Sci. 348 (2016a) 84–106.
[44] H. Wang, F. Smarandache, Y. Zhang, R. Sunderraman, Single valued neutrosophic sets, in: Proceedings of 10th International Conference on Fuzzy Theory and Technology, Salt Lake City, Utah, 2005.
[45] M. Jun, et al., COVID-19 CT Lung and Infection Segmentation Dataset, 2020, http://dx.doi.org/10.5281/zenodo.3757476.
[46] S.C. Douglas, R. Losada, Adaptive filters in Matlab: from novice to expert, in: Proceedings of 2002 IEEE 10th Digital Signal Processing Workshop, 2002 and the 2nd Signal Processing Education Workshop, Pine Mountain, GA, USA, 2002, pp. 168–173.
[47] J. Han, S. Yang, B. Lee, A novel 3-D color histogram equalization method with uniform 1-d gray scale histogram, IEEE Trans. Image Process. 20 (2011) 506–512.
[48] P. Singh, G. Dhiman, Uncertainty representation using fuzzy-entropy approach: Special application in remotely sensed high-resolution satellite images (RSHRSIs), Appl. Soft Comput. 72 (2018) 121–139.