IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received April 8, 2021, accepted May 2, 2021, date of publication May 11, 2021, date of current version May 20, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3079119

An Improved SFLA-Kmeans Algorithm Based on Approximate Backbone and Its Application in Retinal Fundus Image

WEIPING DING, (Senior Member, IEEE), YI ZHANG, YING SUN, AND TINGZHEN QIN
School of Information Science and Technology, Nantong University, Nantong 226019, China

Corresponding author: Weiping Ding (dwp9988@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61976120, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20191445, in part by the Six Talent Peaks Project of Jiangsu Province under Grant XYDXXJS-048, in part by the Qing Lan Project of Jiangsu Province, in part by the Postgraduate Innovation Practice Project in Jiangsu Province under Grant SJCX20_1150, and in part by the College Students' Innovative Entrepreneurial Training Plan Program under Grant 202010304023Z.

ABSTRACT In order to improve the global search ability of K-means algorithm and the clustering effect, a K-means method based on the approximate backbone and the shuffled frog leaping algorithm was proposed. Firstly, the classic iterative formula of the K-means algorithm is replaced by the classic shuffled frog leaping algorithm to obtain better clustering results. Secondly, the K-means algorithm based on the approximate backbone and the shuffled frog leaping algorithm is used for the obtained clustering results. Instead of searching for cluster centers, the cluster division is directly modified. Finally, the experimental results on the UCI dataset show that, the running time of the improved clustering algorithm is shorter than that based on the shuffled frog leaping algorithm only, and clustering results obtained by using the improved clustering algorithm are better than those of other algorithms. In addition, the paper uses the improved clustering algorithm to preprocess medical fundus images to optimize the effect of vascular cutting.

INDEX TERMS K-means algorithm, shuffled frog leaping algorithm, approximate backbone.

I. INTRODUCTION

With the advent of the big data era, how to find the meaning behind the data through a large amount of complex data has become an important topic. As a result, data mining has developed rapidly. Clustering is one of the important data analysis techniques in data mining. It finds the similarity and difference between data which could facilitate people to snoop the internal laws of things.

K-means algorithm is a classic clustering algorithm based on distance division. It is easy to understand, and implement. Meanwhile, it has a good clustering effect. Based on the above advantages, the K-means algorithm is currently widely used [1]–[3].

However, the K-means algorithm also has many problems. For example, the selection of k value lacks theoretical basis [4]; it has a strong dependence on the initial cluster center selection; it is easy to converge to the local optimal and excessive noise sensitive, etc [5]. Therefore, many people have improved the K-means algorithm [6], [7].

The swarm intelligence algorithm comes from the simulation of biological groups such as birds, ants, frogs, and fish in nature. These groups complete the foraging behavior through the independent work of individuals and the mutual cooperation between individuals. The individuals in the swarm intelligence algorithm achieve a powerful overall search function under the interaction through a single, limited intelligence and behavior [8].

Aiming at the problem that the K-means algorithm exists, many people try to optimize with the help of intelligent algorithms [9]. An improved K-means algorithm is proposed with hybridizes Chaos Optimization and Flower Pollination to minimize the cluster integrity. Because the efficiency of K-Means depends on its initialization of cluster centers [10], a hybrid clustering approach based on K-means and Ant Lion Optimization is proposed for optimal cluster analysis [11]. A hybrid version of the artificial chemical reaction optimization algorithm (HACRO) is proposed to optimize clustering problems, considering the artificial chemical reaction suffers

The associate editor coordinating the review of this manuscript and approving it for publication was P. K. Gupta.

from slower convergence speed due to its poor exploitation mechanism [12].

In this paper, the shuffled frog leaping algorithm is used to improve the K-means algorithm, modifying the global search ability of the K-means algorithm. After fully converging, the local optimal solution is further used by the shuffled frog leaping algorithm improved by the approximate backbone.

II. THE CORE IDEA OF K-MEANS ALGORITH

The K-means algorithm is a clustering operation on unlabeled data sets, and is an unsupervised learning algorithm. The algorithm is described as follows: firstly, k-samples are randomly selected from the data set $D = \{x_1, x_2, \ldots, x_n\}$ as the initial clustering center. Then, the cluster centers are used as the basis for the cluster $C = \{C_1, C_2, \ldots, C_k\}$, comparing the data points with each cluster center, and dividing the data points into the clusters corresponding to the nearest cluster center. Finally, the center of each cluster as the new cluster center is calculated, and the above operation is repeated, until the objective function (1) converges.

$$E = \sum\nolimits_{i=1}^{k} \sum\nolimits_{x \in C_i} \|x - \mu_i\|_2^2 \tag{1}$$

In formula (1), E is the sum of the distance differences between all data points and the cluster center to which they belong. μ_i is the mean vector of cluster C_i, and μ_i is expressed as $\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$. Intuitively, the tightness of the samples within the cluster around the mean vector of the cluster, to a certain extent, is characterized by equation (1), that means, the smaller the E value, the higher the tightness within the cluster.

III. IMPROVED SFLA-KMEANS ALGORITHM

A. ALGORITHM OVERVIEW

The shuffled frog leaping algorithm (SFLA) is a heuristic swarm intelligence algorithm [13]. This algorithm is proposed by Eusuff and Lansey, which is inspired by the foraging behavior of frogs. It combines the advantages of particle swarm algorithm and memetic algorithm to conduct modeling and simulation research. The shuffled frog leaping algorithm has the characteristics of simple structure, fast convergence speed and strong global optimization ability. The following are the specific implementation steps of the improved K-means algorithm using the shuffled frog leaping algorithm (hereinafter referred to as SFLA-Kmeans algorithm):

1) GENERATE INITIALIZATION GROUPS

In the data set $D = \{d_1, d_2, \ldots, d_l\}$, k-samples are randomly selected as the value of the i-th frog μ_i, that means, $\mu_i = [d_{\lambda_1}, d_{\lambda_2}, \ldots, d_{\lambda_k}]$, in which λ_i is the subscript of the randomly selected data point, and $\lambda_i \in \{1, 2, \ldots, l\}$, in addition, if $i \neq j$, $\lambda_i \neq \lambda_j$. A total of N frogs are generated as the initial group. Perform step 2 on the initial group.

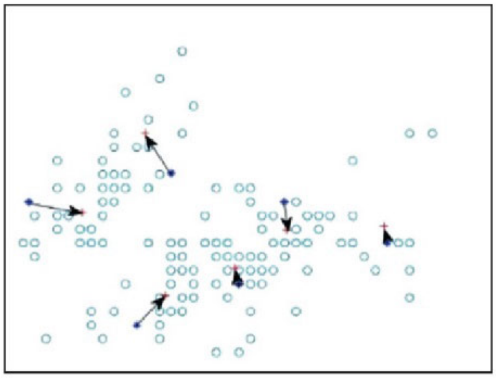


FIGURE 1. Update schematic diagram.

2) GROUP THE FROGS

a. Arrange all frog individuals in the group in descending order of suitability $f(X) = 1/E$, where X is the X-th frog. In addition, the frog with the highest fitness value is recorded as μ^g.

b. Divide the whole group into m-groups, then each group contains n-frogs, that means, n = N/m. After the grouping is finished, go to step 3 for each group.

c. Repeat step 2 until the maximum number of external iterations G_m^{out} is reached.

3) UPDATE THE WORST FROGS IN THE GROUP, AS SHOWN IN FIGURE 1

a. The individual with the highest fitness value in the group is marked as μ^b, and the one with the lowest fitness is marked as μ^w.

b. Calculate the Euclidean distance of each feature of μ^b and μ^w.

$$d_{ij} = \left\| \mu_i^b - \mu_j^w \right\|_2 \tag{2}$$

c. Find the shortest distance and determine the label $\lambda_i = argmin_{j \in \{1,2,\ldots,k\}} d_{ij}$ of the update object of each feature of μ^w.

d. Calculate the new frog individual μ^{new}, and μ_i^{new} is:

$$\begin{cases} l = \left(\mu_{\lambda_i}^b - \mu_i^w\right) \\ \mu_i^{new} = \mu_i^w + rand \times l \\ -D_m \leq \sqrt{l_1^2 + l_2^2 + \ldots + l_k^2} \leq D_m \end{cases} \tag{3}$$

where *rand* is a uniformly distributed random number of [0, 1], and D_m is the maximum allowed moving distance.

e. If $(\mu^{new}) > f(\mu^w)$, replace μ^w with μ^{new}, update the worst frog individual, and reorder; otherwise, repeat the 3 steps with μ^g instead of μ^b to obtain μ^{new}. Similarly, if $f(\mu^{new}) > f(\mu^w)$ is satisfied, replace μ^w with μ^{new} and reorder. If none of the above processes can improve the fitness, perform step 4 and then reorder.

f. Repeat step 3 to the maximum number of internal iterations G_m^{in}.

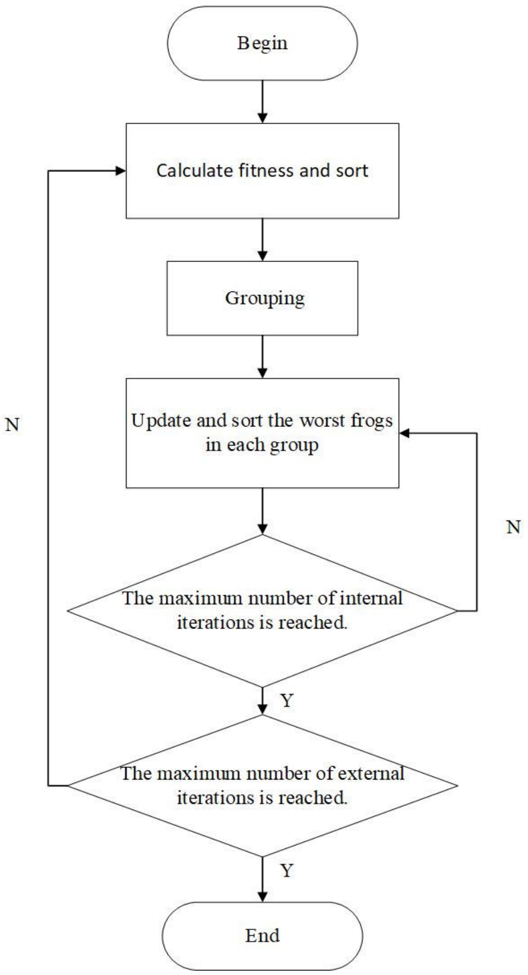


FIGURE 2. SFLA-Kmeans algorithm flowchart.

4) SELECT NEW FROGS

a. When the number of iterations is less than $\sqrt{G_m^{in}}$, randomly select k-samples from the data set D to form a frog μ^{new} and replace μ^w.
b. When the number of iterations is greater than or equal to $\sqrt{G_m^{in}}$, k-vectors are randomly generated in the sample space to constitute the frog μ^{new} and replace μ^w.

The algorithm flow chart is shown in Figure 2.

B. ALGORITHM ANALYSIS OF SFLA-KMEANS

In the above SFLA-Kmeans algorithm, let $d_i = [d_{i1}d_{i2}, , \ldots, d_{id}]$ $, i \in \{1, 2, \ldots, l\}$, which means that each data point in the data set has d attribute features. In this algorithm, the calculation amount for calculating the cluster number of each data point is $O(dkl)$, the calculation amount for obtaining the mean vector is $O(dl)$, the calculation amount required for calculating the suitability is also $O(dl)$, the calculation required for grouping is $O(l)$, the calculation amount required for sorting is $O(l\ logl)$, the calculation amount for updating the frog is $O(d)$, and the calculation amount for sorting the updated frog is $O(n)$. Then the algorithm time complexity is:

$$O(n) = O(G_m^{out}(dkl + dl + l\log l + l + dmG_m^{in} + NG_m^{in})) \tag{4}$$

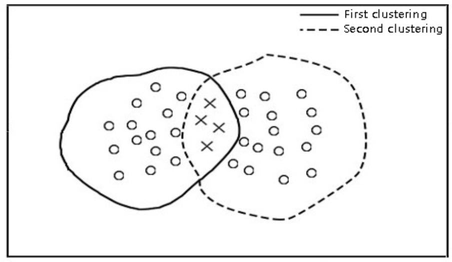


FIGURE 3. Approximate backbone schematic diagram.

IV. AN IMPROVED SFLA-KMEANS ALGORITHM BASED ON APPROXIMATE BACKBONE

A. ALGORITHM OVERVIEW

The cluster clustering algorithm is an operation that divides all data points with a certain target. Among them, the K-means algorithm divides samples with the target of the smallest E value. The classic K-means algorithm compares and replaces the clustering center with the mean vector of the group, to change the clustering center and achieve the goal of reducing the E value; the improved SFLA-Kmeans algorithm uses a linear method to update the cluster center to make the E value decreased. Both of the above algorithms need the help of clustering centers. Obviously, with the help of clustering centers, the division difficulty is greatly reduced and the convergence rate is improved. However, for the SFLA-Kmeans algorithm, as the algorithm converges, the clustering center becomes yoke instead, which mainly reflected in:

(1) Different clustering centers may have the same clustering effect, that is, the division of clusters is the same.

(2) The "movement" of small steps in the clustering center may lead to the change of most points, which in turn leads to a large change in fitness.

In addition, after the SFLA-Kmeans algorithm has fully converged, most data points should have found the correct cluster, and only a few data points have problems with their attribution. In this paper, after the SFLA-Kmeans algorithm has converged, it will no longer try to update the clustering center, instead of which, it directly changes the attribution of each point, and uses the approximate backbone to further improve the SFLA-Kmeans algorithm (hereinafter referred to as BSFLA-Kmeans algorithm).

The concept of the approximate backbone [14] in the clustering problem is introduced here, as shown in Figure 3. For a clustering problem, there are multiple local approximate solution $\pi_1, \pi_2, \pi_3, \ldots, \pi_M$, and the approximate backbone clusters B_i is defined as which satisfies (1)$|B_i| \geq 2$; (2) any $x_i, x_j \in B_i$ belong to the same cluster to any local optimal solution π_l. The set of all the approximate backbone clusters is called the approximate backbone of the problem and is denoted as $Bone(\pi_1, \pi_2, \ldots, \pi_M)$.

The approximate backbone considers data points clustered in the same cluster multiple times as a class. The approximate

Algorithm 1 Possibility *C*alculation

Input : $C_{i-1}^b, C_{i-1}^w, C_i^b, C_i^w$
Output : P(x)
$if\ (f(C_{i-1}^b) - f(C_{i-1}^w) \neq 0)$
{
$if\ (\frac{f(C_i^b)-f(C_i^w)}{f(C_{i-1}^b)-f(C_{i-1}^w)} \leq 1)$
$P(x) = \frac{f(C_i^b)-f(C_i^w)}{f(C_{i-1}^b)-f(C_{i-1}^w)}$
$else\ P(x) = 1$
}
else
$P(x) = rand(0, 1)$
Return $P(x)$

backbone contains the common part of the local optimal solution, and these common parts play an important role in guiding the heuristic clustering algorithm.

Therefore, the common part of the approximate backbone is used to guide the update of the SFLA-Kmeans algorithm. For this, the following improvements are made in this paper:

(1) Change the definition of frog from a group of cluster centers μ_i to a set of cluster numbers $[c_{i_1}, c_{i_2}, \ldots, c_{i_l}]$, where c_{i_j} is the cluster number of the j-th data of the i-th frog, then, $c_{i_j} \in [1, k], c_{i_j} \in Z$. The initial solution of the BSFLA-Kmeans algorithm is initialized with the approximate optimal solution obtained as the result of running the SFLA-Kmeans algorithm repeatedly.

(2) When updating within the group, use the global optimal frog, the optimal frog within the group and the worst frog within the group as the approximate backbone, and divide it according to the optimal frog in the group to select the largest approximate backbone cluster in each cluster number, which means the cluster contains the most point. Every point in the selected largest approximate backbone cluster, decide whether to keep the cluster number, according to the possibility calculated by the following steps,

$f(C_i^b)$ is the fitness of the optimal within the group in the i-th operation. $f(C_i^w)$ is the fitness of the worst within the group in the i-th operation. $rand(0, 1)$ is the random number from 0 to 1. Besides, the remaining points randomly select clusters.

(3) When updating outside the group, select the global optimal and the global worst as the approximate backbone, and the rest is the same as the internal update.

(4) When both the in-group update and the out-of-group update cannot produce an effect, randomly select k-points as the clustering center and relabel each point with the cluster number.

The algorithm flow is shown in Figure 4.

B. ALGORITHM ANALYSIS OF BSFLA-KMEANS

In the above BSFLA-Kmeans algorithm, let $d_i = [d_{i1}d_{i2}, \ldots, d_{id}]$, $i \in \{1, 2, \ldots, l\}$. In this algorithm, the SFLA-Kmeans algorithm needs to be run multiple times to obtain the initial solution. The following only analyzes the

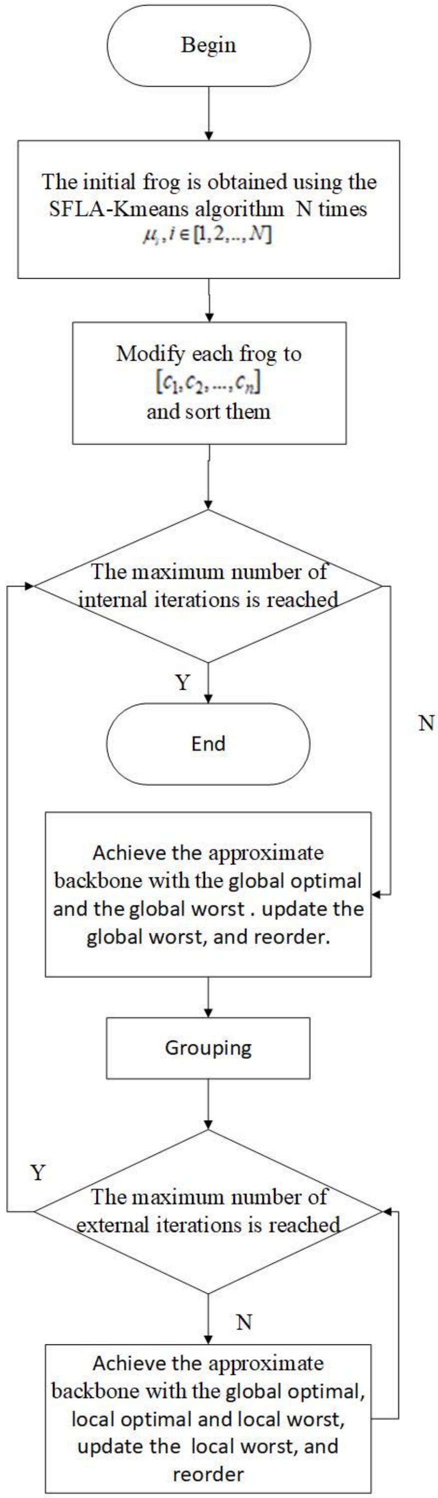


FIGURE 4. BSFLA-Kmeans algorithm flowchart.

time complexity after obtaining the initial solution. Since the algorithm does not need to calculate the cluster number of the data point, there is no need to calculate the amount of $O(dkl)$. The rest is roughly the same as the SFLA-Kmeans algorithm, then the time complexity of this algorithm is as follows:

$$O(G_m^{out}(dl + l\log l + l + dmG_m^{in} + NG_m^{in})) \quad (5)$$

Algorithm 2 Update steps of BSFLA- Kmeans

Input :, N, m, $Bone$, Gr

Output : λ^{new}

$t = N/m$

For i = 1 To t

{

$if(i \in Bone \&\& rand(1) \geq P(x))$

$\lambda_i^{new} = Gr_{1i}$

$else$

$\lambda_i^{new} = x$

}

Return λ^{new}

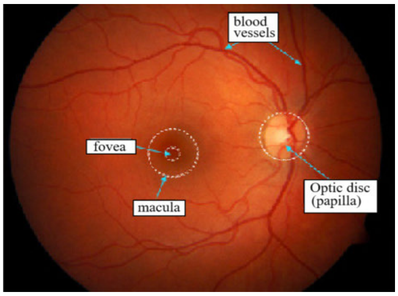


FIGURE 5. Fundus image composition.

V. APPLICATION OF BSFLA-KMEANS ALGORITHM IN MEDICAL FUNDUS IMAGE

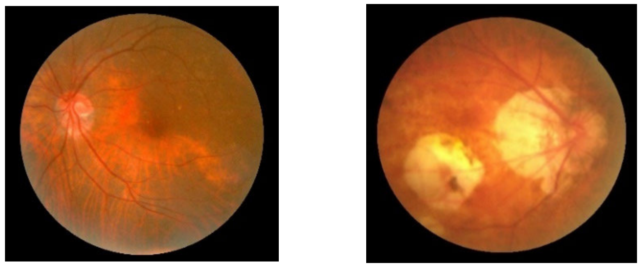
A. OVERVIEW OF FUNDUS IMAGES

The retinal fundus image is mainly composed of fovea, macular, blood vessels, and optic disc, as shown in Figure 5. A variety of diseases can induce retinal lesions, in the case of diabetes, in which 93 million people worldwide suffer from diabetic retinopathy [15]. This image is of great significance for the diagnosis of diseases including diabetes, glaucoma, coronary heart disease, and arteriosclerosis [16]. Therefore, the processing and analysis of the retinal fundus image is very important. The application of machine vision in the retinal fundus image has attracted the attention of many researchers, and a certain effect is achieved [17]–[19]. Intelligent algorithm, as a kind of optimization algorithm, is also applied to the processing of the retinal fundus image, such as the use of fireflies algorithm to locate the disc [20].

B. RELATED TECHNIQUES AND PROBLEMS OF VESSEL SEGMENTATION

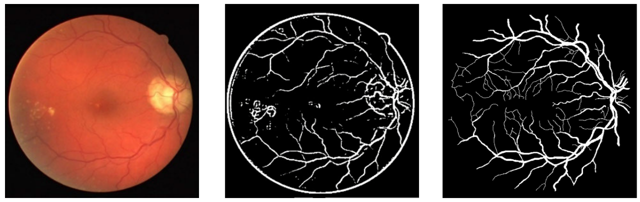
In the fundus image, the blood vessel is an important component, and its structural and morphological changes can effectively point out the development of many systemic diseases including hypertension and diabetes [21].

Since the 1970s, many vessel segmentation algorithms including tracking methods, multi-scale methods, statistical inference, local models, and matched filtering have been proposed [22]. Among them, the segmentation of blood vessels by means of matched filtering is a popular method at present, which has the characteristics of simple, fast, effective and accurate.



(a) Drusen (b) Pathological myopia

FIGURE 6. Fundus disease.



(a) Original draft (b) Result (c) Expectation

FIGURE 7. Vessel segmentation.

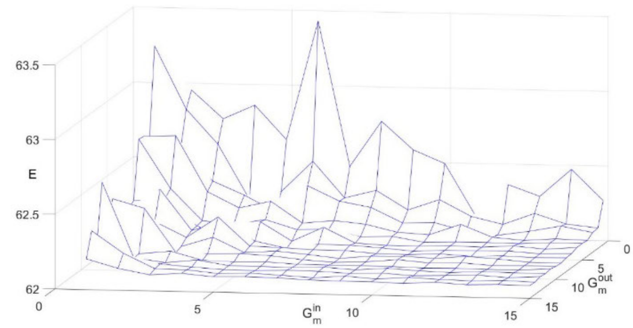


FIGURE 8. The relationship between the number of updates and *E*.

TABLE 1. Data set.

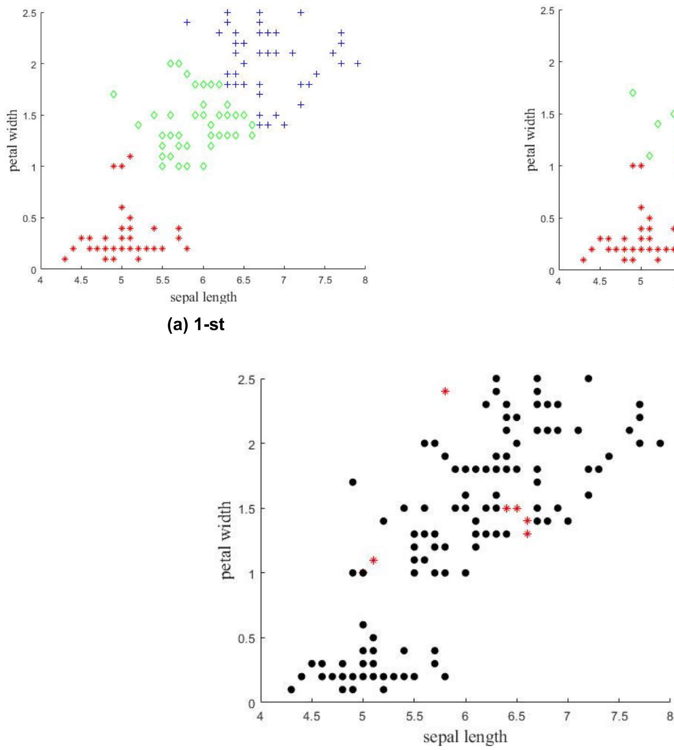
Name	Instances	Attributes	Class
Iris	150	4	3
Wine	178	13	3
Glass	214	9	6

However, many blood vessel segmentation algorithms, including matched filtering, rely more on the difference in color between pixel blocks, so it is easy to treat some large bright lesions caused by disease including the drusen, the pathological myopia, and so on, which are depicted in Figure 6, as blood vessels, which reduces the accuracy of blood vessel cutting, and waste a lot of resources, as shown in Figure 7.

VI. EXPERIMENTAL RESULTS AND ANALYSIS

A. EXPERIMENTAL ENVIRONMENT

The experimental environment of this article is: Matlab 2018a, Windows10 64bit, CPU 2.60HZ. The data was selected from the UCI dataset, as shown in TABLE 1, and images containing lesions in the STARE library.



(a) 1-st

(b) 2-nd

(c) Comparison results

FIGURE 9. Comparison results.

B. EXPERIMENTAL RESULTS

1) EVALUATION INDEX

This article selects the optimization object of the K-means algorithm, within-group error sum E, to compared. In parameter setting, according to our experiments about the parameter relationship in Figure 8, we set the number of frogs $N = 15$, the number of groups $m = 5$, the number of updates outside the group G_{max}^{out} =15, and the number of updates inside the group G_{max}^{in} =15. At the same time, in order to further demonstrate the effect of the BSFLA-Kmeans algorithm, in this paper, in addition to the original K-means algorithm and SFLA-Kmeans algorithm, the BSFLA-Kmeans algorithm is compared to the PSO, KPSO algorithm [23] and the MBCO [24] and other algorithms. In the fundus image processing, this paper introduces three indicators such as sensitivity Sn, specific Sp and accuracy rate ACC for evaluation:

$$Sn = \frac{TP}{TP + FN} \tag{6a}$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{6b}$$

where TP is a correctly classified blood vessel pixel, TN is a correctly classified background pixel, FP is a wrongly classified blood vessel pixel, and FN is a wrongly classified background pixel.

2) VERIFICATION

In order to visualize the effect, we select attribute 1 and attribute 4 of the Iris data set to run the SFLA-kmeans algorithm.

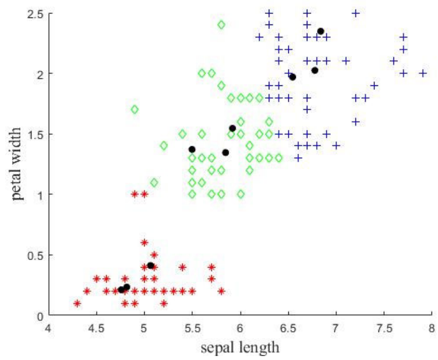


FIGURE 10. Cluster center.

As shown in the Figure 9, comparing two independent running algorithms, only a few data points belong to different cluster centers.

As shown in the TABLE 2 and Figure 10, although the clustering center is constantly changing, the effect of clustering is exactly the same.

3) EXPERIMENTAL RESULTS

This paper shows the running time difference between the PSO, BSFLA-Kmeans and SFLA-Kmeans algorithm, without considering the time of initialization, as shown in TABLE 3.

This paper compares the BSFLA-Kmeans algorithm with the original K-means algorithm and SFLA-Kmeans algorithm, but also compares it with the PSO, KPSO algorithm,

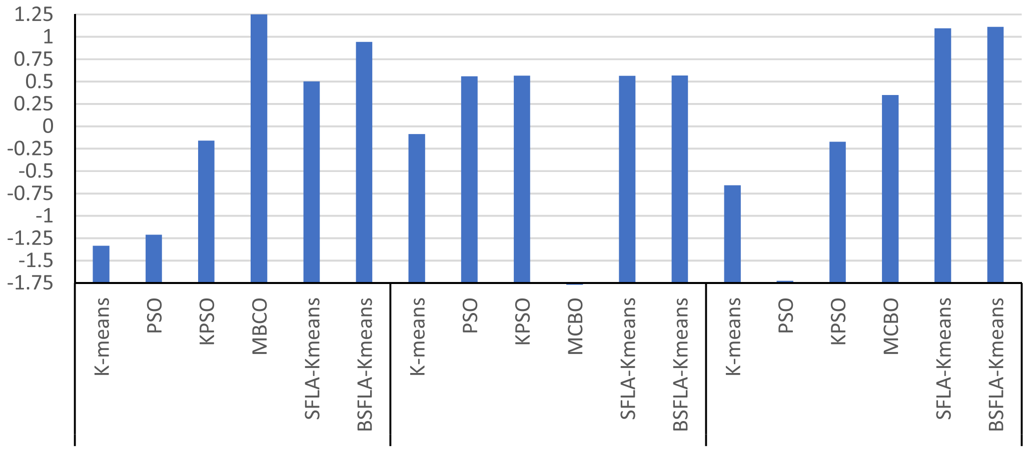
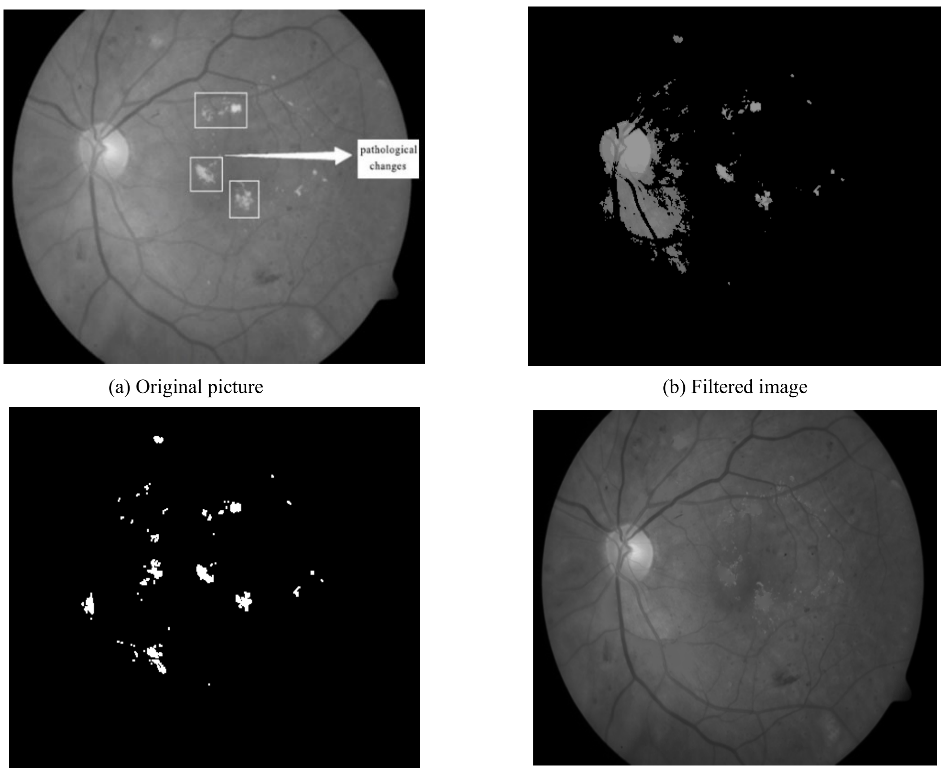


FIGURE 11. **Comparison of various algorithm.**



(a) Original picture

(b) Filtered image

(c) Image after removing the optic disc

(d) Image after process

FIGURE 12. **Image processing.**

and MBCO colony K-means algorithm to highlight the clustering effect of the BSFLA-Kmeans algorithm. The comparison results of each algorithm in the UCI data set are shown in TABLE 4. To visual display, the result is normalized by function (7), as shown in Figure 11.

$$z = -\frac{x-\mu}{\sigma} \tag{7}$$

where x_{min} is the minimum value, x_{max} is the maximum value.

This paper uses the BSFLA-Kmeans algorithm to segment the blood vessels in the fundus medical record as shown in Figure 12(a), and locate and remove the lesions according to the characteristics of the lesions. The specific steps are as follows:

(1) After the image is processed with gray scale, the pixels are clustered with the aid of the BSFLA-Kmeans algorithm.

(2) Filter out the clusters with the higher average gray value within the cluster, that is, the top 10% of the gray average value, and the pixels included, as shown in Figure 12(b).

(3) The highlighted part is composed of the optic disc and the lesion. The optic disc can be removed with the help of existing algorithms including blood vessel, shape and brightness characteristics, as shown in Figure 12(c).

TABLE 2. Cluster center.

Times	Result		
	μ_1	μ_2	μ_3
1st	(6.545,1.969)	(4.758,0.211)	(5.918,1.546)
2nd	(6.838,2.348)	(5.063,0.412)	(5.496,1.371)
3rd	(6.776,2.024)	(4.815,0.233)	(5.846,1.345)
4th	(4.858,0.319)	(6.242,1.471)	(6.536,1.731)

TABLE 3. Comparison of running time(s).

Way	DataSet		
	Iris	Wine	Glass
PSO	2.3424	3.5245	5.1309
SFLA-Kmeans	2.8616	3.4585	6.0764
BSFLA-Kmeans	2.2788	3.2629	4.8072

TABLE 4. Comparison of various algorithm values *E*.

Name	k	Methods	E(Mean)
Iris	3	K-means	102.57
		PSO	102.26
		KPSO	99.61
		MBCO	96.03
		SFLA-Kmeans	97.94
		BSFLA-Kmeans	96.83
Wine	3	K-means	1.7662($\times 10^4$)
		PSO	1.6320($\times 10^4$)
		KPSO	1.6307($\times 10^4$)
		MCBO	2.1989 ($\times 10^4$)
		SFLA-Kmeans	1.6312($\times 10^4$)
		BSFLA-Kmeans	1.6302($\times 10^4$)
Glass	6	K-means	241.03
		PSO	258.02
		KPSO	233.28
		MCBO	225.00
		SFLA-Kmeans	213.10
		BSFLA-Kmeans	212.86

(4) Cover the remaining pixel blocks after screening, and the processed image is shown in Figure 12(d).

According to the above steps, use the BSFLA-Kmeans algorithm to preprocess the fundus medical record image to obtain the processed image as shown in Figure 12(d). The original image, the image processed by the orthodox method, like histogram equalization, and the image processed by the BSFLA-Kmeans algorithm are respectively used to match the filter algorithm. It performs blood vessel cutting, and evaluates the effect of blood vessel cutting in the fundus medical record image according to the three evaluation indicators proposed in the previous section. The results are shown in TABLE 5 and Figure 13.

4) ANALYSIS OF RESULTS

It can be seen from TABLE 3 that the BSFLA-Kmeans algorithm has a shorter running time than the original SFLA-Kmeans and PSO. This is because the BSFLA-Kmeans algorithm uses numbering instead of comparing clustering centers with data, which saves much time. Of course, the above comparison is not completely fair,

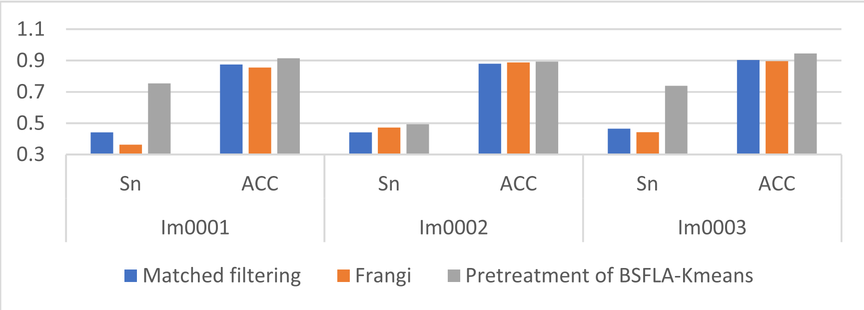


FIGURE 13. Image processing results.

TABLE 5. Image processing results.

Way	Im0001		Im0002		Im0003	
	Sn	ACC	Sn	ACC	Sn	ACC
Matched filtering	0.4410	0.8747	0.4413	0.8799	0.4654	0.9030
Frangi	0.3636	0.8548	0.4721	0.8880	0.4436	0.8969
Pretreatment of BSFLA-Kmeans	0.7547	0.9149	0.4945	0.8922	0.7385	0.9452

because a certain accuracy of the initial solution is required by the BSFLA-Kmeans algorithm, and the above comparison does not take the time to obtain the initial solution into consideration.

TABLE 4 and Figure 11 show the value of E obtained by running 6 clustering algorithms including K-means algorithm, PSO algorithm, KPSO algorithm, MBCO algorithm, SFLA-Kmeans algorithm, and BSFLA-Kmeans algorithm.

From them, it can be seen that the E value of the BSFLA-Kmeans algorithm is significantly smaller than other algorithms.

Therefore, it can be concluded that the adoption of intelligent algorithms to improve the K-means algorithm has achieved certain results. At the same time, the BSFLA-Kmeans algorithm has significant advantages over other algorithms.

It can be seen from TABLE 5 and Figure 13 that with the help of BSFLA-Kmeans algorithm to segment the image, the sensitivity Sn of the matched filter algorithm are greatly improved. Accuracy ACC of the matched filter algorithm are improved

Based on the above analysis, it can be inferred that the image segmented by the BSFLA-Kmeans algorithm effectively removes the lesion points, thus successfully reducing the misclassified background pixels caused by the lesion, but at the same time sacrifices the part too close to the lesion Blood vessel pixels. In general, preprocessing by BSFLA-Kmeans algorithm is effective for improving accuracy.

VII. SUMMARY

The K-means method based on the approximate backbone and the shuffled frog leaping algorithm is proposed. the iterative formula of the K-means method to obtain clustering results is discarded, and the classic shuffled frog leaping algorithm is used to obtain the clustering results, instead. The clustering results obtained are modified directly by using the proposed K-means algorithm based on the approximate backbone and the shuffled frog leaping algorithm, instead of relying on the cluster centers to optimize the clustering results. Experimental results show that, the running time of the improved clustering algorithm is shorter than that based on the shuffled frog leaping algorithm only, and clustering results obtained by using the improved clustering algorithm are better than those of other algorithms. The improved clustering algorithm is applied to medical fundus record images, and it has a better effect of optimizing blood vessel cutting. Next, the automatic setting of some hyperparameters in the algorithm will be further studied, such as the number of frogs, the number of groups, and the number of iterations. In addition, multiple runs of the original SFLA-Kmeans algorithm are required to the initial solution of the BSFLA-Kmeans algorithm, which make the response time too long, simplifying the SFLA-Kmeans algorithm or using the classic Kmeans algorithm to obtain a better initial solution is reckoned.

REFERENCES

[1] P. Govender and V. Sivakumar, "Application of K-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019)," *Atmos. Pollut. Res.*, vol. 11, no. 1, pp. 40–56, Jan. 2020.

[2] S. Khanmohammadi, N. Adibeig, and S. Shanehbandy, "An improved overlapping K-means clustering method for medical applications," *Expert Syst. Appl.*, vol. 67, pp. 12–18, Jan. 2017.

[3] P. Chévez, D. Barbero, I. Martini, and C. Discoli, "Application of the K-means clustering method for the detection and analysis of areas of homogeneous residential electricity consumption at the Great La Plata region, Buenos Aires, Argentina," *Sustain. Cities Soc.*, vol. 32, pp. 115–129, Jul. 2017.

[4] D. K. Kotary and S. J. Nanda, "Automatic determination of K in distributed K-means clustering," *Procedia Comput. Sci.*, vol. 165, pp. 556–564, Jan. 2019.

[5] S.-S. Yu, S.-W. Chu, C.-M. Wang, Y.-K. Chan, and T.-C. Chang, "Two improved K-means algorithms," *Appl. Soft Comput.*, vol. 68, pp. 747–755, Jul. 2018.

[6] S. Khanmohammadi, N. Adibeig, and S. Shanehbandy, "An improved overlapping k-means clustering method for medical applications," *Expert Syst. Appl.*, vol. 67, pp. 12–18, Jan. 2017.

[7] C. Yin and S. Zhang, "Parallel implementing improved K-means applied for image retrieval and anomaly detection," *Multimedia Tools Appl.*, vol. 76, no. 16, pp. 16911–16927, Aug. 2017.

[8] X.-S. Yang, "Nature-inspired optimization algorithms: Challenges and open problems," *J. Comput. Sci.*, vol. 46, Oct. 2020, Art. no. 101104.

[9] R. Tang, S. Fong, X.-S. Yang, and S. Deb, "Integrating nature-inspired optimization algorithms to K-means clustering," in *Proc. 7th Int. Conf. Digit. Inf. Manage. (ICDIM)*, Macau, China, Aug. 2012, pp. 116–123.

[10] A. Kaur, S. K. Pal, and A. P. Singh, "Hybridization of chaos and flower pollination algorithm over K-means for data clustering," *Appl. Soft Comput.*, vol. 97, Dec. 2020, Art. no. 105523.

[11] S. K. Majhi and S. Biswal, "Optimal cluster analysis using hybrid K-means and ant lion optimizer," *Karbala Int. J. Mod. Sci.*, vol. 4, no. 4, pp. 347–360, Dec. 2018.

[12] H. Singh and Y. Kumar, "Hybrid artificial chemical reaction optimization algorithm for cluster analysis," *Procedia Comput. Sci.*, vol. 167, pp. 531–540, Jan. 2020.

[13] T. Venkatesan and M. Y. Sanavullah, "SFLA approach to solve PBUC problem with emission limitation," *Int. J. Electr. Power Energy Syst.*, vol. 46, pp. 1–9, Mar. 2013.

[14] Y. Zong, G. Xu, P. Jin, Y. Zhang, and E. Chen, "HC_AB: A new heuristic clustering algorithm based on approximate backbone," *Inf. Process. Lett.*, vol. 111, no. 17, pp. 857–863, Sep. 2011.

[15] J. Lee, B. G. Moon, A. R. Cho, and Y. H. Yoon, "Optical coherence tomography angiography of DME and its association with anti-VEGF treatment response," *Ophthalmology*, vol. 123, no. 11, pp. 2368–2375, Nov. 2016.

[16] M. Hassan, M. Amin, I. Murtza, A. Khan, and A. Chaudhry, "Robust hidden Markov model based intelligent blood vessel detection of fundus images," *Comput. Methods Programs Biomed.*, vol. 151, pp. 193–201, Nov. 2017.

[17] S. J. G. Shoba and A. B. Therese, "Detection of glaucoma disease in fundus images based on morphological operation and finite element method," *Biomed. Signal Process. Control*, vol. 62, Sep. 2020, Art. no. 101986.

[18] C. Zhou, X. Zhang, and H. Chen, "A new robust method for blood vessel segmentation in retinal fundus images based on weighted line detector and hidden Markov model," *Comput. Methods Programs Biomed.*, vol. 187, Apr. 2020, Art. no. 105231.

[19] K. Chang, N. Balachandar, C. Lam, D. Yi, J. Brown, A. Beers, B. Rosen, D. L. Rubin, and J. Kalpathy-Cramer, "Distributed deep learning networks among institutions for medical imaging," *J. Amer. Med. Inform. Assoc.*, vol. 25, no. 8, pp. 945–954, Aug. 2018.

[20] J. Rahebi and F. Hardalaç, "A new approach to optic disc detection in human retinal images using the firefly algorithm," *Med. Biol. Eng. Comput.*, vol. 54, nos. 2–3, pp. 453–461, Mar. 2016.

[21] S. Ugrenović, I. Jovanović, B. Kundalić, V. Stojanović, M. Pavlović, A. Antović, M. Milić, and J. Čukuranović-Kokoris, "Morphometric analysis of the epineurial and endoneurial blood vessels of the human sciatic nerve in relation to aging," *Tissue Cell*, vol. 66, Oct. 2020, Art. no. 101389.

[22] S. Moccia, E. De Momi, S. El Hadji, and L. S. Mattos, "Blood vessel segmentation algorithms—Review of methods, datasets and evaluation metrics," *Comput. Methods Programs Biomed.*, vol. 158, pp. 71–91, May 2018.

[23] T. Hassanzadeh and M. R. Meybodi, "A new hybrid approach for data clustering using firefly algorithm and K-means," in *Proc. 16th CSI Int. Symp. Artif. Intell. Signal Process. (AISP)*, Shiraz, Iran, May 2012, pp. 7–11.

[24] P. Das, D. K. Das, and S. Dey, "A modified bee colony optimization (MBCO) and its hybridization with k-means for an application to data clustering," *Appl. Soft Comput.*, vol. 70, pp. 590–603, Sep. 2018.

WEIPING DING (Senior Member, IEEE) received the Ph.D. degree in computation application from the Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China, in 2013. He was a Visiting Scholar with the University of Lethbridge (UL), Lethbridge, AB, Canada, in 2011. From 2014 to 2015, he was a Postdoctoral Researcher with the Brain Research Center, National Chiao Tung University (NCTU), Hsinchu, Taiwan. In 2016, he was a Visiting Scholar with the National University of Singapore (NUS), Singapore. From 2017 to 2018, he was a Visiting Professor with the University of Technology Sydney (UTS), Sydney, NSW, Australia. He has published more than 100 research peer-reviewed journal articles and conference papers in these fields, including IEEE Transactions on Fuzzy Systems (TFS), IEEE Transactions on Neural Networks and Learning Systems (TNNLS), IEEE Transactions on Evolutionary Computation (TEVC), IEEE Transactions on Cybernetics (TCYB), IEEE Transactions on Industrial Informatics (TII), IEEE Transactions on Systems, Man, and Cybernetics: Systems (TSMC), IEEE Transactions on Biomedical Engineering (TBME), and IEEE Transactions on Emerging Topics in Computational Intelligence (TETCI), and has held 15 approved invention patents till now. His research interests include deep neural networks, multimodal machine learning, granular data mining, uncertainty modeling in big data, co-evolutionary algorithm, medical images analysis, and big data analysis. He has served on the Editorial Advisory Board for *Knowledge-Based Systems*, and serves on the Editorial Board for *Information Fusion*, *Engineering Applications of Artificial Intelligence*, and *Applied Soft Computing*. He also serves/served as an Associate Editor for IEEE Transactions on Fuzzy Systems, IEEE/CAA Journal of Automatica Sinica, *Information Sciences*, *Neurocomputing*, *Swarm and Evolutionary Computation*, IEEE Access, and *Journal of Intelligent & Fuzzy Systems*, and the Co-Editor-in-Chief for *Journal of Artificial Intelligence and System*. He is also the Leading Guest Editor of Special Issues in several prestigious journals, including IEEE Transactions on Evolutionary Computation and IEEE Transactions on Fuzzy Systems.

YI ZHANG is currently pursuing the bachelor's degree with the School of Information Science and Technology, Nantong University. His research interests include data mining and machine learning.

YING SUN is currently pursuing the master's degree with the School of Information Science and Technology, Nantong University. Her research interests include rough set and deep learning.

TINGZHEN QIN is currently pursuing the master's degree with the School of Information Science and Technology, Nantong University. His research interests include data mining, rough set, and distributed computing.