

## Research Article

### A New Optimized Data Clustering Technique using Cellular Automata and Adaptive Central Force Optimization (ACFO)

<sup>1</sup>G. Srinivasa Rao, <sup>2</sup>Vakulabharanam Vijaya Kumar and <sup>3</sup>Penmesta Suresh Varma

<sup>1</sup>Pragati Engineering College, Kakinada, India

<sup>2</sup>Department of CSE and IT, Anurag Group of Institutions, Hyderabad, India

<sup>3</sup>Department of Computer Science, College Development Council, Adikavi Nannaya University, Rajahmundry, India

**Abstract:** As clustering techniques are gaining more important today, we propose a new clustering technique by means of ACFO and cellular automata. The cellular automata uniquely characterizes the condition of a cell at a specific moment by employing the data like the conditions of a reference cell together with its adjoining cell, total number of cells, restraint, transition function and neighbourhood calculation. With an eye on explaining the condition of the cell, morphological functions are executed on the image. In accordance with the four stages of the morphological process, the rural and the urban areas are grouped separately. In order to steer clear of the stochastic turbulences, the threshold is optimized by means of the ACFO. The test outcomes obtained vouchsafe superb performance of the innovative technique. The accomplishment of the new-fangled technique is assessed by using additional number of images and is contrasted with the traditional methods like CFO (Central Force Optimization) and PSO (Particle Swarm Optimization).

**Keywords:** ACFO (Adaptive Central Force Optimization), cellular automata, convolution, correlation, morphological operation

## INTRODUCTION

Now, the Cellular Automata (CA) are the arithmetical brands meant for mega systems housing huge numbers of straightforward matching modules with limited interactions (Sharma *et al.*, 2012). They are non-linear dynamical systems in which space and time are distinct and are termed as cellular, on account of the fact that they are made up of cells such as points in the lattice or squares of the checker boards and are referred to as 'automata' (Sree and Devi, 2013). They encompass a vast number of comparatively trouble-free individual units, or "cells". Each cell, in turn, is a straightforward predetermined automation which continually refreshes its own status, where the fresh cell state is dependent on the existing state of the cell and its immediate (local) neighbours. On account of these glistening qualities, CAs have been employed widely to assess intricate systems in environment (Sree and Devi, 2013). The spectacular nature of the Cellular automata can be characterized as follows: (Singh and Lal, 2012):

- A cellular automaton is discrete time space.
- Each and every cell comprises a number of restricted states.
- The entire cells are located in the identical location.

- Each and every cell is rationalized simultaneously.
- The regulation in each and every locality is dependent on the value of the locality around its neighbours.
- The regulation for fresh value of each and every locality is also based on value of restricted number of preceding conditions.

In cellular automata each and every cell has got a specific state (Zuhdi *et al.*, 2011). The cell adapts its state in line with those of its neighbourhood and also that of its preceding time stage (Tripathy and Nandi, 2009). CA, in fact, is characterized by two parameters such as  $k$  and  $r$ , where  $k$  refers to a number of states and  $r$  specifies a radius of its neighbourhood (Sree and Babu, 2010). In the case of a cellular automata vector, each cell is endowed with 2 states and 1 radius. The relative regulations are known as transition rules (Gamal *et al.*, 2013). The foremost Application domain of Cellular automata is the data mining (Silwattananusarn and Tuamsuk, 2012). The modus operandi of cellular automata in data mining method flows as follows: At the outset the input data is encapsulated from a definite time interval of inspection and transition rules are obtained by means of application of data mining methods (Esmaeilpour *et al.*,

2012). Predominantly, three data mining methods are employed, such as locating sequential models in the data, grouping investigation of the model and categorization of a fresh data grid into specific type of group, which is related to the preferred transition rule (Zuhdi *et al.*, 2011). CAs are endowed with several merits for modelling, together with their decentralized method, straightforward to the intricacy rule, the association of form with task and model with procedure, the comparative easiness in visualizing the model outcomes, their elasticity, their vibrant technique and also their kinship with geographical data systems and remotely sensed data. Of all, the most noteworthy quality is its effortlessness (Shanthi and Rajan, 2012).

It is pertinent to note, here, that mammoth quantity of data are generated by various applications such as high-speed networking, finance logs, sensor networks and web tracking (Milea *et al.*, 2011). The enormous amount of data thus collected from various sources is developed as an unrestricted data sequence arriving at the port of the system (Kozak *et al.*, 2009). Usually, it is a herculean task to amass the entire data stream in main memory for online processing as the magnitude of a data stream is sensibly gigantic (Mohamed *et al.*, 2010). The data stream processing has to be executed in conformity with the ensuing stipulations:

- Restricted usage of memory
- Linear time consideration of the incessantly produced fresh elements
- Unfeasibility of executing blocking functions
- Impracticality to examine data for manifold occasions (Sree and Devi, 2013)

There is a feast of methods designed for the mining of items or models from data streams (Nirkhi *et al.*, 2012). Nevertheless, the provision of proficient and high-speed techniques without infringing the limitations of the data stream atmosphere has emerged as the most important challenge. This has ultimately and inevitably led to the origin of the procedure of data mining in cellular automata (Zuhdi *et al.*, 2011). The data mining methods are highly essential in several application domains such as online photo and video streaming services, economic analysis, concurrent manufacturing process control, search engines, spam filters, security and medical services (Javadzadeh *et al.*, 2011).

## LITERATURE REVIEW

Wang *et al.* (2011) have fabulously formulated a potential Rough Set Theory (RST) to direct the factor selection. And the proposed data mining technique has been experimented for the calibration of a CA model to replicate land use. The factors chosen by RST were different for different land uses. The qualifying merit of RST is evident from the fact that it preserves the original factors in the classification of the transition regulations. Moreover, the calculation interval

necessary for the replication employing the RST factors has been observed to be significantly lesser than the time interval essential to engender the outcomes by means of employing the original set of factors. Nevertheless, the data mining method itself is found to be computationally exhaustive. The promising outcomes have clearly exhibited the fact that RST is capable of directing the choice of the leading factors necessary for the calibration of a CA brand, though it remains a challenging fact that its potential is in further need of additional analysis.

Lope and Maravall (2013) have jubilantly advocated a data clustering algorithm founded on the concept of deeming the individual data items as cells forming part of a one-dimensional cellular automaton. It integrates the insights into both social segregation brands rooted on Cellular Automata Theory, where the data items themselves have the capability to travel freely in lattices and also from Ants clustering algorithms. They also deem an automatic technique for estimating the number of clusters in the dataset by evaluating the intra-cluster variations. A sequence of tests with both synthetic and real datasets has been offered with the intent to examine empirically the convergence and performance outcomes. The gathered test outcomes have been analyzed and contrasted with those achieved by the traditional clustering algorithms.

Perez and Dragicevic (2012) have legitimately launched the incorporation of the ABM with CA technique to successfully tackle modelling at both fine and large spatial scales. The distinct nature of CA facilitates incorporation with raster-based geospatial datasets in GIS and is also advantageous during the course of modelling intricate natural procedures that develop over time. The renovated model encompasses various factors like the wind directions and altitude to illustrate their sway in the spread models of the outbursts at a landscape spatial scale. The consequential outcomes have resulted in a superb comprehension of all the factors and variables that influence and play a part in the occurrence of MPB and by and large forest insect turbulences.

Gorsevski *et al.* (2012) have proficiently put forward a straightforward technique of edge recognition in accordance with the cellular automata by means of using a digital image. The recognition process is generally applicable to both monochromatic and color images. The preliminary work offers a 2-D image technique which can further be renovated to pre stack and 3-D images. The choice of transition regulations is restricted to a small number of exploratory regulations, but the technique shines with the superb quality of elasticity for conceiving and experimenting diverse regulations and neighbourhoods which have been useful for edge recognition. The charismatic outcomes achieved from the sample investigated have been found to have ushered in superb results when the technique is contrasted against a conventional manual digitization method and a modern GIS-based technique.

Wang *et al.* (2013) have heftily launched a cloud-based CA to characterize ambiguity proliferation and reliance of replication outcomes on diverse scales of ambiguity hyper-entropy. Immediately after building the proper parameter settings for the cloud-CA model, an evaluation of cloud-CA with the fuzzy-set-based CA (fuzzy-CA) model and the hybrid CA model based on Fuzzy Set and the Monte Carlo method (FSMC-CA) is conducted by replicating spatial models of urban development. The test outcomes have come out with the success stories of the cloud-CA model leading the other two CA brands by a clear edge, ushering in superb performance, beating them hands down, with superior kappa indices and figure of merit, highlighting and upholding the efficiency of the cloud-CA model.

### PROPOSED METHODOLOGY

In our innovative method, at the outset, the images are obtained from the database and thereafter, they are subjected to four stages of morphological functions like opening by convolution, opening by reconstruction, closing by correlation and closing by reconstruction. The consequential image gathered after the morphological process is known as the clustered image. However, it is unfortunate that it is haunted by certain stochastic turbulences also. With a view to successfully tackle this phenomenon, the threshold is optimized by means of the ACFO. The efficiency of the innovative technique is analyzed and contrasted with the peer optimization algorithms like CFO and PSO. The architecture of the proposed clustering technique is given in Fig. 1.

At the outset, Let us consider a database D with various several urban and rural images represented by  $\{I_{1i}, I_{2i}, \dots, I_{ni}\}$ . Thereafter, the urban part 'u' is physically segregated from the images and the segregated components are amassed in a different database known as T ( $T = t_{i1}, t_{i2}, \dots, t_{im}$ ) where  $m > n$ . The four phases of morphological function embrace the

fundamental functions like dilation and erosion. However, in our technique, we propose to employ convolution and correlation in place of dilation and erosion.

**Convolution:** Convolution function has emerged as the most widespread among several image processing operators. It is fundamentally dependent on a straightforward mathematical function. It is the technique of multiplying two assortments of diverse dimensions to generate a new third assortment of numbers. In the course of image processing, Convolution is employed to execute operators which are the linear blend of definite input pixel value of image and to generate the output pixel value. It is based on a kind of algorithm which is termed as spatial filters. These filters employ an extensive gamut of masks or kernels, to determine diverse outcome, based on the preferred function. 2-D convolution has established itself as the most crucial one to current image processing. The crucial concept behind this is to scrutinize a window of certain set dimension over an image. The output pixel value is the weighted sum of input pixels within the window where the weights are the values of the filter allocated to each and every pixel of the window. The window with its weights is known as either convolution mask or as kernels. The mathematical equation of the convolution for the image is given by the expression:

$$C(f, k) = \sum_{mn} f(i - m, j - n) k(m, n) \quad (1)$$

where,  
 f = The input image  
 k = The kernel

From the equation, it is crystal clear that convolution is analogous to dilation in morphological function with the employment of configuration factor.

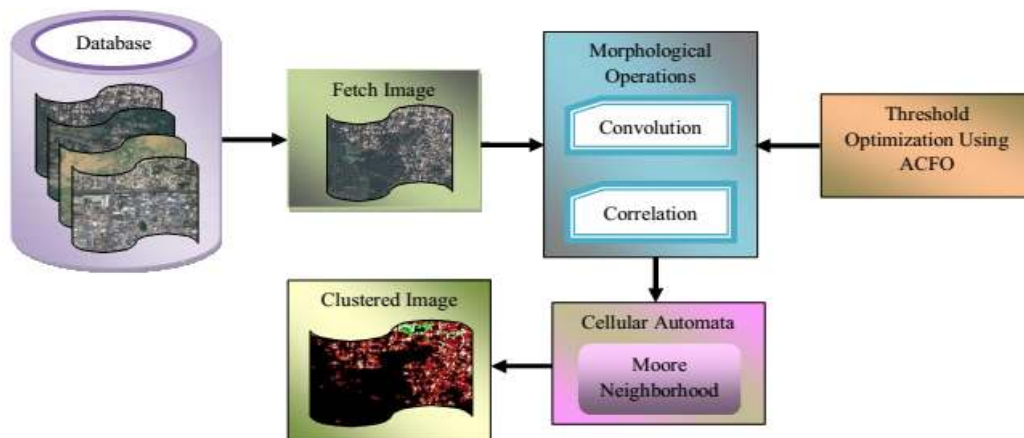


Fig. 1: Architecture of the proposed technology

The different phases of convolution in filtering procedure are detailed below:

- Each and every pixel in the image neighbourhood is multiplied by the contents of the matching component in the filtering kernel.
- The outcomes from the multiplication are added together and then got divided by the sum of the kernel.
- The outcome is listed and improved and employed to reinstate the center pixel in the image neighbourhood.

**Correlation:** The correlation function is intimately linked to the convolution. Identical to the convolution process, correlation estimates the output pixels as a weighted sum of neighbouring pixels. The only disparity is that the matrix of weights here is termed as correlation kernels, which amounts to 180° rotation of convolution kernel. Correlation is aptly defined as:

$$C(f, k) = \sum_{mn} f(i + m, j + n)k(m, n) \quad (2)$$

where,  
 f = The input image  
 k = The kernel

Here, both correlation and convolution function as Erosion and Dilation in Morphological function. In grey-scale Images, Convolution tends to enhance the brightness of object by taking the neighbourhood maximum in the course of passing with filter, as correlation exhibits the tendency to trim down the brightness of object by taking the neighbourhood minimum while passing with filter.

Four phases of morphological function are executed on the image. When the morphological filter is executed on the image, the clustered image is gathered in accordance with the threshold value. However, on account of the stochastic turbulences haunting the clustered image, optimization process is executed. The threshold is optimized by means of the ACFO where the threshold is fixed according to the degree of intensity. If the pixel value is lesser than the threshold, it is labelled as rural otherwise it is represented as urban.

**Adaptive Central Force Optimization (ACFO):** CFO employs probes as its fundamental population, which are spread all through the search space and as time passes, they slowly shift towards the probe that has obtained the maximum mass or fitness. This motion of the probes is constructed based on the mathematical theories which have been evolved with a view to successfully portray the force between two objects.

Each and every individual probe possesses an initialized position vector  $W_i = (w_i^1, w_i^2, \dots, w_i^d)$  where

$w_i^d$  being its position in the  $d^{\text{th}}$  dimension, acceleration,  $A_i = (A_i^1, A_i^2, \dots, A_i^d)$  where  $A_i^d$  being the acceleration in the  $d^{\text{th}}$  dimension and  $C_i = (c_i^1, c_i^2, \dots, c_i^d)$  where  $c_i^d$  being fitness value in the  $d^{\text{th}}$  dimension. The position vector is a symbol of the probes current coordinates in respect of each and every dimension of the search space. The acceleration vector is similar to the position vector with the exception that it amasses values in respect of modifications in location and as of now it possesses no minimum or maximum value. Moreover, each and every particle possesses a fitness value that is chased:

$$A_{j-1}^p = G \sum_{k=1, k \neq p}^{N_p} I(F) F^x \frac{(W_{j-1}^k - W_{j-1}^p)}{(W_{j-1}^k - W_{j-1}^p)^\beta} \quad (3)$$

$$F = C_{j-1}^k - C_{j-1}^p \quad (4)$$

$$W_j^p = W_{j-1}^p + \frac{1}{2} A_{j-1}^p \Delta t^2 \quad (5)$$

where,  
 $N_p$  = The number of probe  
 $p$  = The current probe  
 $G$  = The gravitational constant  
 $\alpha$  and  $\beta$  = Constants  
 $C^k$  = The fitness of a probe  
 $k, j$  = The current time step  
 $I$  = The unit step function

**Disadvantages of CFO:** Time complication is a fundamental challenge as it consumes an extensive amount of time to find keys to the specified issues. CFO has a tendency to consume a larger slice of time to finish a run than is usually preferred. Many of the challenges may be assigned to the computation of fresh acceleration values. Novel techniques for both making things simpler and optimizing the computations have to be scrupulously followed.

It is with these intentions that adaptive central force optimization is launched by means of genetic operators, such as crossover and mutation.

**Threshold optimization by ACFO:**

- Generate the probe initialization in an arbitrary manner. Initialize position and acceleration vector of each and every probe to zero.
- **Estimate initial probe distribution:** Initial probe distribution should be determined by placing all probes in a way that is homogeneously spaced across the axes of each and every dimension, homogeneously spaced on the diagonals of a definite dilemma, or in an arbitrary way.

- **Define the fitness function:** The fitness function selected must be used for the constraints in accordance with the current population:

$$C^k = \text{Max (Accuracy)} \quad (6)$$

The accuracy is determined by comparing the clustered image with the physically segmented image.

- **Calculation of the acceleration vector:** The acceleration factors are determined by means of the Eq. (3).
- **Updating probe position:** The probe position may be revised by determining the acceleration and fitness function of each and every probe. If the fresh value is superior to the existing one, replace it with fresh value:

$$W_j^p = W_{j-1}^p + \frac{1}{2} A_{j-1}^p \Delta t^2 \quad (7)$$

- **Determining the fitness value:**  $C^k$  is the fitness of probe k, j is the current time step and  $I(F)$  is the unit step function. It is to be noted that the equivalent to mass in these equations is the same as the difference between fitness values:

$$FF = C_{j-1}^k - C_{j-1}^p \quad (8)$$

- **Estimate new acceleration:** After revising the current probe position, fresh acceleration factors are determined.
- **Errant probe retrieval:** Errant Probe retrieval is a different topic in respect of PSO. This initial algorithm employs Eq. (8) to relocate dimensions of probes that are greater than their minimum value whereas Eq. (9) is utilized to relocate dimensions of probes which are greater than their maximum value:

$$W(p, i, j) = B_{\min}(i) + (W(p, i, j-1) - B_{\min}(i)) \quad (9)$$

$$W(p, i, j) = B_{\max}(i) - F_{rep} * (B_{\max}(i) - (W(p, i, j-1))) \quad (10)$$

In Eq. (9),  $F_{rep}$  is the repositioning factor which defined by the user. Typically it is set as 0.5,  $B_{\min}(i)$  is the lower bound of the solution space and  $B_{\max}(i)$  is the upper bound of the solution space:

$$B_{\min}(i) = B_{\min}(i) + 0.5 \times (W(p, i, j)_{best} - B_{\min}(i)) \quad (11)$$

$$B_{\max}(i) = B_{\max}(i) + 0.5 \times (B_{\max}(i) - W(p, i, j)_{best}) \quad (12)$$

In Eq. (10) and (11),  $W(p, i, j)_{best}$ -best position vector.

In addition, in the case of ACFO technique, the obtained probes are revised by the genetic operators such as crossover and mutation. The genetic operators in the probe generation procedure will enhance the system performance in relation to the conventional CFO technique.

- **Crossover:** Crossover is a genetic operator and is the procedure of taking multiple parent solutions and generating a child solution from them. It is employed to produce offspring by swapping bits in a couple of individuals.
- **Mutation:** It is a genetic operator which modifies one or more gene values in a chromosome from its preliminary state. This has the effect of supplementing completely fresh gene values to the gene pool. With the aid of the fresh gene values thus added, the genetic algorithm is capable of realizing a superior solution than is formerly feasible. Mutation, in essence, is an essential segment of the genetic exploration as it extends a helping hand in thwarting the population from languishing at any limited optima.
- **Decisive factor to stop the operation:** Go on with the step till the solution arrived at is superb or till maximum criterion is attained.

The optimized threshold is used in Eq. (7) to forecast the condition of the cell. Thus the relative optimization of threshold by means of ACFO paves the way for attaining superb clustering outcome.

**Cellular automata:** Finally, by applying the Moore neighborhood of the cellular automata on the obtained image from the above process, the resultant is the clustered image. In cellular automata, the Moore neighborhood involves the eight cells encompassing a focal cell on a two-dimensional square grid. The area is named after Edward F. Moore, a pioneer of cell automata hypothesis. It is one of the two most normally utilized neighborhood sorts, the other one being the 4-cell von Neumann neighborhood. The well known Conway's Game of Life, for instance, utilizes the Moore neighborhood. It is like the idea of 8-joined pixels in workstation representation. The idea might be stretched out to higher measurements, for instance structuring a 26-phone cubic neighborhood for a phone machine in three measurements, as utilized by 3 day Life. The Moore neighborhood of a point is the focuses at a Chebyshev separation of 1. The amount of cells in a Moore neighborhood, provided for its extend r is:

$$(2r + 1)^2 - 1$$

The thought behind the detailing of Moore neighborhood is to discover the form of a given diagram. This thought was an extraordinary test for

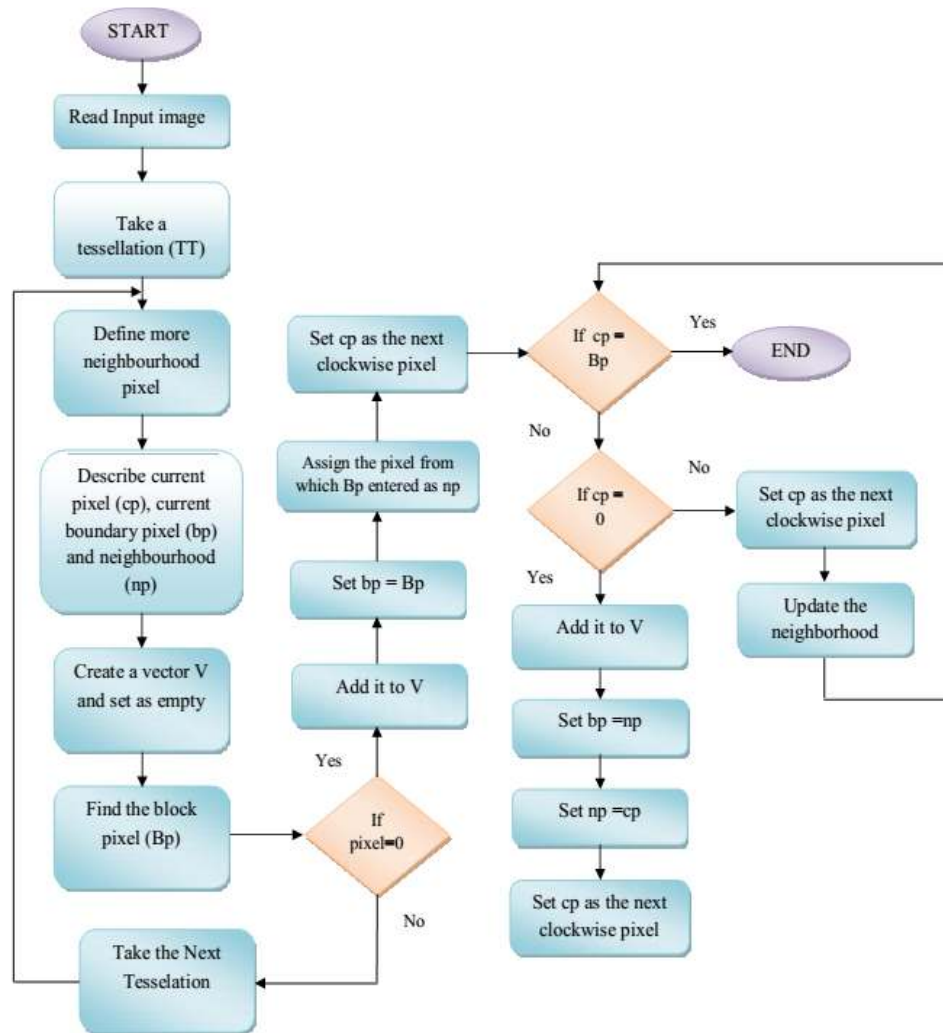


Fig. 2: Flowchart of the Moore neighborhood process

most experts of the eighteenth century and therefore a calculation was determined from the Moore diagram which was later called the Moore Neighborhood calculation. The Moore neighborhood process is depicted in Fig. 2.

Explanation of the flowchart given in Fig. 2 follows as: The image obtained as the resultant morphological operation is fed as the input for processing the clustering with the help of the cellular automaton. Since we are using the Moore neighborhood, the Connected component (Cc) in a Tessellation (TT) is 8. A Moore neighborhood is defined where the number of pixels in the Moore neighborhood is 8. After that, current boundary pixel (bp), current pixel (cp) and neighborhood pixel (np) are defined. Create a vector (V) and set as  $V = \text{null}$  and examine the Tessellation (TT) in all the direction in order to find whether any Black pixel (Bp) (i.e., pixel value = 0) exists in the Tessellation (TT). If any Black pixel (Bp) exists in the tessellation, add it in to the Vector (V). Assign the found out Black pixel (Bp) as

the current Boundary point (bp) and the pixel from which the Black pixel (Bp) entered while the examining process is assigned as the neighborhood of the current boundary pixel (np). Now the current pixel (cp) is set as the next clockwise pixel from the neighborhood of the boundary pixel (np). Then check whether the current pixel (cp) is equal to the Black pixel (Bp) which is already detected and if it is not then check whether current pixel (cp) is a black pixel. If cp is a black pixel, then add it to the vector V. Then assign the neighborhood of the current boundary point (np) as the new boundary point (i.e.,  $bp = np$ ) and also the current pixel is now set as the new neighborhood (i.e.,  $np = cp$ ). After that, the current pixel (cp) is assigned as the next clockwise pixel from the neighborhood of the boundary pixel (np) and update the neighborhood. This process is continued until, the current boundary pixel is equal to the begin pixel for the second time. Finally the clustered image is obtained.

This restricts the set of shapes the calculation will walk totally. An enhanced ceasing condition proposed by Jacob Eliosoff is to stop in the wake of entering the begin pixel for the second time in the same course you initially entered it.

### EXPERIMENTAL RESULTS AND DISCUSSION

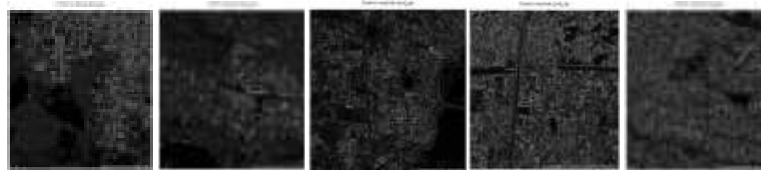
Our proposed ACFO based Optimized Cellular Automata in data mining is executed in the working platform of MATLAB (version 7.13).

The performance of the proposed ACFO based Optimized Cellular Automata is estimated by means of the satellite images extracted from the database. The sample images extracted from the satellite database are exhibited in Fig. 3.

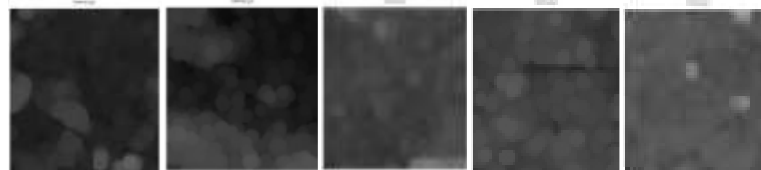
**Performance analysis:** By applying the statistical measures, the performance of our proposed ACFO

based Optimized Cellular Automata in data mining is examined. The images taken from the satellite databases are utilized to analyze the performance of our proposed ACFO based Optimized Cellular Automata in data mining technique. Then the promising results obtained in this regard are compared with those of the existing CFO (Central Force Optimization) and PSO (Particle Swarm Optimization) Techniques.

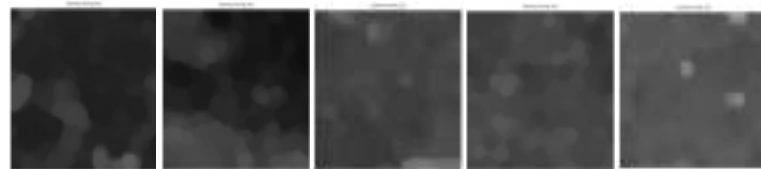
**Discussion:** Table 1 shows the performance of our Proposed ACFO based Optimized Cellular Automata in data mining technique in terms of Accuracy, sensitivity and specificity. For the 5 different images the accuracy values of the suggested method are found to be (95%), (95%), (87%), (85%) and (89%) respectively, though Existing Systems such as PSO and CFO techniques offer only (91%, (72%), (77%), (82%), (89%) and (88%), (72%), (86%), (82%), (89%) of accuracy correspondingly.



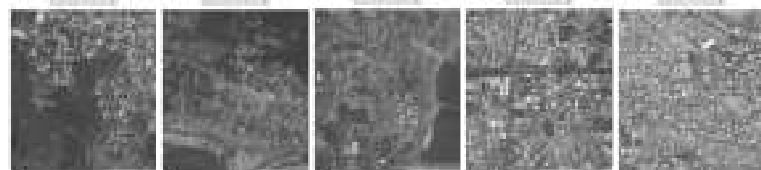
(a)



(b)



(c)



(d)



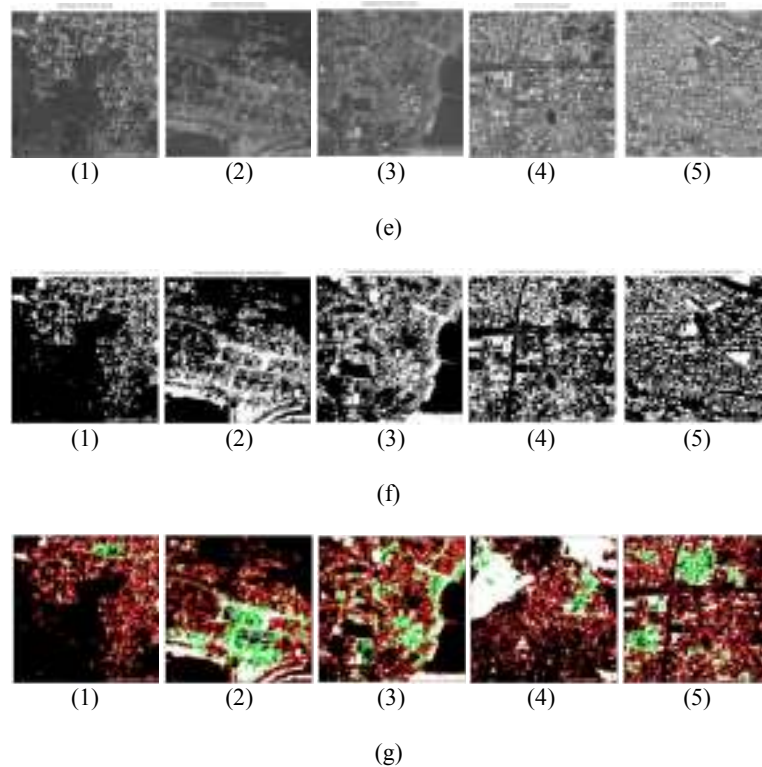







Fig. 3: (a) Gray-level image, (b) opening operation, (c) opening by reconstruction, (d) closing, (e) closing by reconstruction, (f) image after threshold optimization, (g) output of cellular automata

Table 1: Performance of our proposed ACFO based optimized cellular automata in data mining technique and the existing techniques such as CFO and PSO for the satellite database in terms of accuracy sensitivity and specificity

Images	ACFO			CFO			PSO		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
	0.954296	0.574123	0.996499	0.910229	0.402670	1	0.884895	0.344585	1
	0.952171	0.641312	0.955496	0.725971	0.156440	1	0.725971	0.156440	1
	0.875187	0.846113	0.875835	0.773842	0.382174	1	0.868683	0.517120	0.986226
	0.858822	0.604059	0.946149	0.826180	0.527842	0.999801	0.826180	0.527842	0.999801
	0.890774	0.699801	1	0.890774	0.699801	1	0.890774	0.699801	1

When compared to the existing systems, the proposed ACFO based Optimized Cellular Automata technique has given higher accuracy rate. Similarly, for the 5 different images the sensitivity values of the proposed ACFO method are observed to be (57%), (64%), (84%), (60%) and (69%), respectively. However, the Existing Systems such as PSO and CFO techniques offer only (40%), (15%), (38%), (52%), (69%) and (34%), (15%), (51%), (52%), (69%) of sensitivity, respectively. Thus when compared to the existing systems, the proposed ACFO based Optimized

Cellular Automata technique has amazingly yielded higher sensitivity rate. Even though the specificity of our proposed method is moderately lower than that of the existing system, it ushers in higher sensitivity and accuracy when compared to the existing systems. Hence, we are glad to declare that our innovative technique has been able to yield superb performance results.

**Discussion:** Here, we observe that the graphs effectively show the fitness function of an image



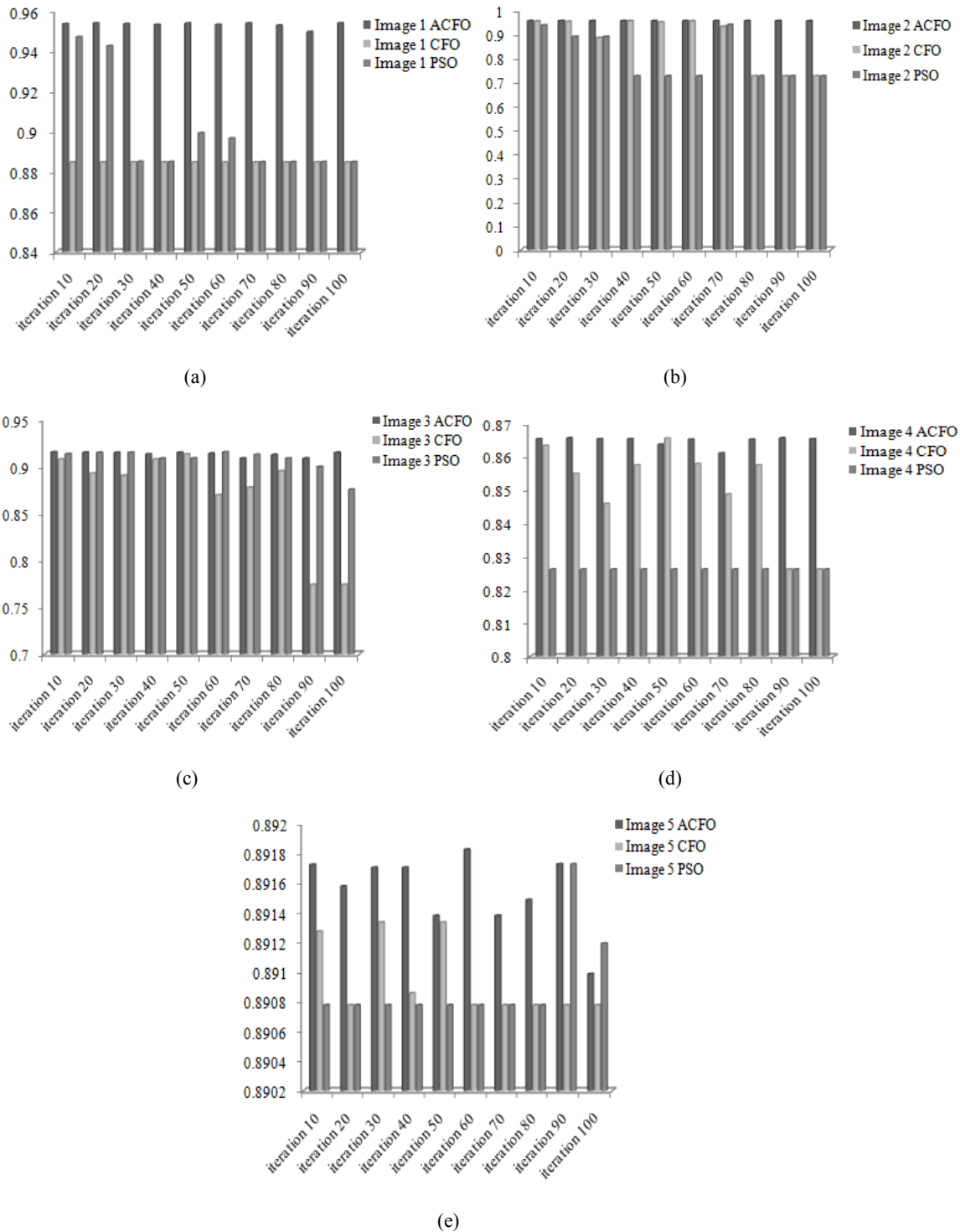


Fig. 4: The graphs have shown the performance of the proposed ACFO based optimized cellular automata technique and the existing techniques such as CFO and PSO

calculated based on the accuracy. For the 5 different images the iteration is varied and thus the graph is plotted. Consider the Fig. 4a. Here, for each and every iteration, our proposed ACFO method attains high fitness value consistently. But in the case of PSO the

fitness value continues to be low and there is no change even when the iteration is varied. Let us consider Fig. 4b and c. Here for the initial iterations the ACFO, CFO and the PSO fitness remain the same and continue to be high. But when the iteration is changed CFO and

PSO decrease to very low values. In Fig. 4d and e also the PSO remains very low and is consistent for all the iterations. In case of CFO it is more or less equal to the ACFO but as the iteration is changed, it shows a lot of variations and at some point of time it stoops to very low values. However, ACFO remains constant for all iterations. Thus the performance metrics and the comparison effect demonstrate that the suggested method clusters the data more accurately than the other methods.

### CONCLUSION

In this study, we have taken earnest efforts to successfully propose an ACFO based Optimized Cellular Automata technique. The performance analysis has clearly proved that our proposed system has been able to attain higher accuracy results than those of the existing system. The process has succeeded in offering a remarkable rate of accuracy. The comparison table and the graphs have shown that the clustering based on ACFO based Optimized Data Clustering has given higher performance than those of the existing PSO and CFO techniques, Hence it is established that our proposed system is highly competent to efficiently perform the clustering process when compared to the other existing methods.

### REFERENCES

- Esmailpour, M., V. Naderifar and Z. Shukur, 2012. Cellular learning automata approach for data classification. *Int. J. Innov. Comput. I.*, 8(12): 8063-8076.
- Gamal, M., A.A. El-Fetouh and S. Barakat, 2013. A fuzzy rough rule based system enhanced by fuzzy cellular automata. *Int. J. Adv. Comput. Sci. Appl.*, 4(5): 1-11.
- Gorsevski, P., C. Onasch, J. Farver and X. Ye, 2012. Detecting grain boundaries in deformed rocks using a cellular automata approach. *Comput. Geosci.*, 42(1): 136-142.
- Javadzadeh, R., Z. Afsahi and M.R. Meybodi, 2011. Hybrid model based on artificial immune system and cellular automata. *World Acad. Sci. Eng. Technol.*, 59(1): 658-661.
- Kozak, K., A. Agrawal, N. Machuy and G. Csucs, 2009. Data mining techniques in high content screening: A survey. *J. Comput. Sci. Syst. Biol.*, 2: 219-239.
- Lope, J.D. and D. Maravall, 2013. Data clustering using a linear cellular automata-based algorithm. *Neurocomputing*, 114(2): 86-91.
- Milea, P.L., M. Teodorescu, R. Muller, M. Dragulinescu, O. Oltu, G. Tiplea, G. Stefan and S. Pompilian, 2011. Cellular automata applications for renewable energy monitoring. *Int. J. Energ. Environ.*, 5(3): 418-425.
- Mohamed, H.R., L.A. Chaouki, A. Lehireche and R. Mohamed, 2010. Text clustering based on the N-grams by bio inspired method. *J. Art. Sci. Commer.*, 1(1): 56-68.
- Nirkhi, S.M., R.V. Dharaskar and V.M. Thakre, 2012. Data mining: A prospective approach for digital forensics. *Int. J. Data Min. Knowl. Manage. Process*, 2(6): 41-48.
- Perez, L. and S. Dragicevic, 2012. Landscape level simulation of forest insect disturbance Coupling swarm intelligent agents with GIS-based cellular automata model. *Ecol. Model.*, 231: 53-64.
- Shanthi, M. and E.G. Rajan, 2012. Agent based cellular automata: A novel approach for modeling spatiotemporal growth processes. *Int. J. Appl. Innov. Eng. Manage.*, 1(3): 56-61.
- Sharma, V., A. Dev and S. Rai, 2012. A comprehensive study of cellular automata. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, 2(10): 340-344.
- Silwattanusarn, T. and K. Tuamsuk, 2012. Data mining and its applications for knowledge management. *Int. J. Data Min. Knowl. Manage. Process*, 2(5): 13-24.
- Singh, S. and J. Lal, 2012. Role of cellular automata in budding fields of computer science. *Int. J. Adv. Comput. Math. Sci.*, 3(4): 550-555.
- Sree, K. and R. Babu, 2010. Identification of promoter region in genomic DNA using cellular automata based text clustering. *Int. Arab J. Inf. Technol.*, 7(1): 75-78.
- Sree, P.K. and N.U. Devi, 2013. Achieving efficient file compression with linear cellular automata pattern classifier. *Int. J. Hybrid Inform. Technol.*, 6(2): 15-26.
- Tripathy, S. and S. Nandi, 2009. Lightweight cellular automata-based symmetric-key encryption. *Int. J. Netw. Secur.*, 8(2): 243-252.
- Wang, F., J. Hasbani, X. Wang and D.J. Marceau, 2011. Identifying dominant factors for the calibration of a land-use cellular automata model using rough set theory. *Comput. Environ. Urban*, 35(2): 116-125.
- Wang, H., S. He, X. Liu, L. Dai, P. Pan, S. Hong and W. Zhang, 2013. Simulating urban expansion using a cloud-based cellular automata model. *Landscape Urban Plan.*, 110: 99-112.
- Zuhdi, A., A.M. Arymurthy and H. Suhartanto, 2011. Geographic spatiotemporal dynamic model using cellular automata and data mining techniques. *Int. J. Comput. Sci. Issues*, 8(2): 53-61.