HSO Based K-Means for Automatic Segmentation of Disease and Deficiency in Tomato Plant

S. Sivagami, S. Mohanapriya

Abstract: Image segmentation is an important part in image processing because it split the given whole image into parts, among the parts we can choose which one is most important. There are wide range of image processing algorithms available among them K-means algorithm is very simple to understand is one of the simplest and Produce accurate result, even though it is simple and accurate it has some limitations that is we need to guess the value for K and randomly select K initial centroids among the given data points. To overcome this initialization problem a new method for image segmentation based on Harmonic search optimization (HSO) and K-means for deficiency detection was proposed. The performance of K-means algorithm is mainly based on the k value and k initial centroids of the clusters. Random initialization is followed in normal K-means algorithm due to this normal k-means take lots of time to produce correct. A new method called HSO based K-means is proposed to speed up the initialization process. The proposed algorithms exploit an initial step derived from the HSO, considering Otsu method as the objective function. After finding the cluster centers using HSO, K-means algorithm is initialised with these cluster centers. Finally, segmentation result is compared with normal K-Means and EM segmentation algorithm our proposed HSO based K-means algorithm gives better result than others.

Keywords: Image segmentation, K-Means, EM, Improved K-Means

I. INTRODUCTION

Now a days, Image processing is broadly in all areas like medical science, Vehicle control and more among them farming is very important. In agriculture image processing used in the following field early identification of disease in leaf, automatic identification of deficiency in soil using leaves, weed detection, maturity identification etc., among all these deficiency identification and disease identification are very important because these two decides the yield. Early days disease and deficiency are identified manually it is experience and expertise people directly come and visit the field predict what kind of disease or what is the nutrient that is lacking in soil. To do this process manually means lots of time and money has to spend by the farmer. Image processing is boon to farmers that it automatically identifies the disease and deficiency.

Nutrients and causes of its deficiency

Plant needs nutrients to grow healthy, there are two type of nutrients Primary and Secondary nutrients, Primary nutrients are very important for the plant, the quality, colour and shape of the fruits is mostly depending upon the primary nutrient. Primary nutrients are Nitrogen, Phosphorous, Potassium. Nitrogen deficiency can result in stunted growth of the plant and plant leaves becomes yellow and die. It decreases the size and number of the fruit, colour and taste also get decreased. After seeding or transplanting phosphorus initiate root growth, it also increases the taste, colour and vitamin C content. Tomatoes will get stunted growth with the deficiency of phosphorus. Potassium helps to increase the yield of tomato production by stimulating early flowering and setting of fruits. Uneven pigmentation, vascular browning are the symptoms of Potassium deficiency. When the soil pH is less than 4.5 then calcium deficiency occurs. When tomatoes grown in greenhouse then Magnesium deficiency is most frequently occurred. Interveinal chlorosis in leaves, yellowing or orange of older leaves are the symptoms of Magnesium deficiency. Purpling of Veins and petioles, interveinal chlorosis are the symptoms of Sulphur deficiency.

Image segmentation algorithms

There are lot of segmentation algorithms available for image segmentation based on the following five categories threshold based segmentation method, regional growth method, segmentation based on edge detection, clustering method and CNN. In Clustering based segmentation lots of algorithms available among them K-Means is one of the important algorithm. In this research paper we have chosen clustering based method that is K-Means for image segmentation and to improve the efficiency of hybrid algorithm that is k-means with HSO was used for better result. Advantage of K-means algorithms is it is simple to understand and fast to evaluate. It is very much efficient and we can enlarge it for large data sets. Disadvantage of this algorithm is to initially tell the number of cluster that is value for K , it is very difficult to tell the value for k, and initially selecting centroids. In this paper we proposed HSO metaheuristic methods for initializing value for k as well as initial selection of centroids

The following sections are scheduled as follows: Literature survey followed by proposed system followed by result and discussion and finally conclusion.

II. LITERATURE SURVEY

Md. Rakib Hassan et al., [4] have proposed a new method based on HSV and RGB color space, in this research work they use color space like HSV and RGB are used to find the value for k in k-means. Zubair Khan et al., [5] presented a new mehtod to k value initialization in k-means they used grey level histogram, from this histogram they identify the highest peak and then vertical scan and horizontal scan methods are used to find the value for k.

Revised Manuscript Received on April 15, 2020.

S. Sivagami, Assistant Professor in Computer Science department at Adhiyamaan College of Agriculture and Research, Hosur, Tamil Nadu.

Dr. S. Mohanapriya, Assistant Professor of the Computer Science Department at KSR College of Arts and Science for Women, Tiruchengode, Tamil Nadu.

PasiFränti et al.,[7] have find a new technique for improving k-means, in this study they use Maxmin method to find out the value for k. Lahouaoui et al.,[3] have proposed modified Expectation maximization algorithm for image segmentation using pruning strategy with maximum likelihood and they have compare their new algorithms result with other algorithms. Compare to traditional EM algorithm computation time was reduced since number of iteration will be reduced. Eman Abdel-Maksoud et al., have suggested a new method for Brain tumor segmentation based on a hybrid clustering technique[6], in their research they integrate both FCM (Fuzzy c-means) and K-means, which is followed by thresholding and level set segmentation method for accurately segmenting the brain tumor image

III. PROPOSED SYSTEM

Harmony Searching Algorithm: The Harmony searching algorithm was proposed from the observation of musical instrument harmony. The musician makes different tones that can be consider as vector values. if fine tune was selected then that tone is considered as base to create a better harmony in the consequent attempt. To find better tune in Musical harmony is similar to the optimization problem, HS algorithm uses very few mathematical expressions and less parameters for the random search compared with early heuristic optimization algorithms. This algorithm can be used easily in various engineering problems to optimize the solution. Harmony memory consists of harmony which is also called as solutions. The harmonies are generated by the random selection of vectors, After the generation of the Harmony Memory a novel candidate is created for each of the harmony. Harmony vector consists of worst and best solution. By comparing newly created value with worst solution the harmony will be uptated, if better solution is found then the values should be swapped. It consists of three steps:

- Harmony memory initialization
- New harmony vectors
- Updating the harmony memory

a) Harmony memory Initialization: The harmony memory initialization consists objective function which is used to optimize the problem and it is described by the following function:

$$\min \; f(x), x = \big(x(1), x(2), \dots . x(n)\big) \in R^n$$
$$Subject\ to: x(j) \in [l(j), u(j)]\ j = 1,2 \dots, n \quad (1)$$

Here f(x) is an objective function, n is total number of variables, l(j) is the lower bound value and u(j) is the upper bound value of x(j). We have different parameters to be initialized that are NI-Number of Improvisation, HMCR - Harmony memory Consideration Rate and HM - Harmony Memory. The whole Harmony search algorithms performance was depends on the initial assignment of these parameters. Initial value to this variable was assigned based the application domain.

The initial vector components at HM are calculated as follows, $x_i=\{x_i(1), x_i(2),..., x_i(n)\}$ where $x_i(k)=l(k)+(u(k)-l(k)).rand(0,1)$ for k=1,2,...,n and i=1,2,..,HMS. Here the value between 0 and 1 was produced using the function rand (0,1). The Harmony memory matrix will be filled as follows:

$$HM=[x_1,x_2,x_3,......,x_{HMS}] \quad (2)$$

Improvisation of new harmony vectors: Improvisation means to generate new harmony. By taking into account of pitch adjustment, random re-initialization the new harmony vector x_{new} is built. $x_{new}(1)$ is chosen randomly from the harmony memory and then compared with the HMCR if the r1(0,1) is smaller than the HMCR or not if it is lesser means then the decision variable $x_{new}(1)$ is calculated, else it is calculated from the lower bound and upper bound values, Like that all the other variables $x_{new}(2)$, $x_{new}(3)$,...., $x_{new}(n)$ are also calculated. The processes are carried out as follows:

$$X_{new} = \begin{cases} X_i(j) \in \{X_i(j), \dots, X_{HMS}(j)\}\ with\ probability\ HMCR \\ l(j) + u(j) - l(j).rand(0,1)\ with\ probability\ l - HMCR \end{cases} \quad (3)$$

All the variables calculated from the formula (3) was tested whether to adjust pitch or not. To do the pitch adjustment PAR - Pitch adjusting rate was initialized to adjust the frequency as well as control local search around Harmonic memory. The process of pitch adjusting is carried out as follows:

$$X_{new}(j) = \begin{cases} X_{new}(j) = X_{new}(j) +/-rand(0,1) \\ BW\ with\ probability\ PAR \\ X_{new}(j)\ with\ probability\ (1-PAR) \end{cases} \quad (4)$$

To generate new harmony as an alternat for the original in memory pitch adjusting is very important.

Updating the harmony memory: When new harmony is calculated it is compared with worst elements in harmony. During comparison process if newly calculated is better than worst element means the worst element was replaced by the newly generated one. Harmonic memory contains only best element after few attempts.

Algorithm 1 for generation of the best vector:

1. Initialized the elements NI, BW, HMS, PAR and HMCR.
2. The value of each vector in the harmony memory was calculated using the formula 3.
3. Harmony value to be improved as $x_{new,}$ Where the random value generated should be less than the HMCR then $x_{new}(k)= x_{new}(k)$ where k=1,2,...,HMS, then it should be verified with the PAR value if r_2 is less than the PAR value then $x_{new}(k)= x_{new}(k)$ +/- r_3.BW. Here r_1, r_2, r_3 are generated by the operation rand(0,1) which selects a value between 0 and 1. The newly generate harmony is then compared with the lower bound value if it is <l(j) then the new harmony is assigned l(j), If the value is greater than the u(j) then the new harmony is assigned with the u(j). If every condition as stated as above does not satisfied then the value is calculated according to the following condition that calculated new element is less than the worst than the worst element is replaced with new element.
4. Then the worst harmony in the harmony memory is updated with the newly generated harmony.
5. Check how many number of times the above process was calculated if it is equal to number of iteration that is NI

then return the best harmony vector else go back to step 3.

Algorithm 2 for the Thresholding according to the generated vector:

1. Import an image I, and store it into I_{Gr} and c=1
2. Using the imhist function generate the histogram h^{Gr}.
3. The probability is calculated using the distribution obtained in the histograms.
4. Initialize a Harmony memory X^c_j of HMS random elements with k dimension.
5. Initialize HSA parameters: K, HMS, PAR, HMCR, PAR, NI ,BW and the limits of l and u.
6. The values w_i^c of u^c_1 are calculated using the objective function in Otsu.
7. The improvisation of the new harmony is performed as follows: The random value generated should be less than the HMCR then $x^c_{new}(j)=x_j^c$ where j=1,2,3,...,HMS Then it should be verified with the PAR value if r_2 is less than the PAR value then $x^c_{new}(j)=x^c_{i(j)}+/-r_3.BW$. Here r_1,r_2,r_3 are generated by the operation rand(0,1) which selects a value between 0 and 1. The newly generate harmony is then compared with the lower bound value if it is <l(j) then the new harmony is assigned l(j). If the value is greater than the u(j) then the new harmony is assigned with the u(j). If every condition as stated as above does not satisfied then the value is calculated as follows:
   $x^c_{new}(k)=l(k)+r.u(k)-l(k)$ where r ϵ rand(0,1)
8. Update the HM as $x^c_{worst}= x^c_{new=}$ if $f(x^c_{new}) > f(x^c_{worst})$
9. If NI is completed, then go to next step, otherwise go to step 6.
10. Harmony was selected that has the best x^c_{best} objective function value
11. Apply the thresholds values contained x^c_{best}in to the image.

The Proposed Segmentation System: The proposed detecting disease and deficiency system consists of four steps: first is image pre-processing, second one is finding cluster centers using HSO, Segmentation using Modified K-means, third is extraction of features using Grey-Level Co-occurrence Matrix(GLCM) and last step is classification. The main objective of the proposed HSO based K-means is by choosing correct value for cluster centres for K-means clustering.

Pre-processing: Acquired images are in different size and various color combinations and sometimes it has some noise that to be removed to do the following segmentation process. By using filtering algorithm, the noise has to be removed. In our proposed method we used median filter, which is used to remove noise effectively and preserve noise. Each pixel value is replaced by median pixel value of neighbouring pixels.

HSO for cluster center initialization:

Step 1: Optimization algorithm uses k different variables as decision variables for candidate solution. These decision variables are used as threshold point and that indirectly used for center selection.

$$HM = [X_1^C, X_2^C, \dots\dots, X_{HMS}^C, HMS]$$

$$X_i^c = [th_1^c, th_2^c, \dots\dots, th_k^c]^T \quad (5)$$

Here: T is the transpose operator which is used to take transpose of the given matrix, Size of harmony memory was represented by HMS, ith element of HM is represented by X_i and c=1. We have to set boundaries based on our problem on hand. In our problem of identifying deficiency and decease we have set lower bound as zero and upper bound as 255 that is based on intensity level of the image.

Step 2: Implementation of HMA: The proposed algorithm has been implemented considering Otsu method as an objective function:

$$J(t) = max\left(\sigma^{2c}(t)\right), 0 \le t_i \le L-1, i = 1,2,3,\dots\dots,k \quad (6)$$

Where t={t_1, t_2, t_3,…,t_{k-1}} are various threshold, based on these variance are calculated using the following formula:

$$\sigma^{2c} = \sum_{i=1}^{k} \sigma_i^c = \sum_{i=1}^{k} w_i^c\,(u_i^c - u_T^c)^2 \quad (7)$$

In the above formula I represents class I, w_i^c represents the probability of occurrence and u_i^c represents the mean of a class. These values are obtained in MT as follows:

$$w_i^c(t) = \sum_{i=1}^{t_i} ph_i^c \text{ and } u_i^c = \sum_{i=1}^{t_i} \frac{iph_i^c}{w_i^c(t_i)}$$

Step 3: with the best possible configuration of HAS using image histograms and the Otsu functions cluster centers are generated [3]

Step 4: Based on the values obtained from the above three steps K- Means Algorithm was executed. It is one of the unsupervised algorithms, that means it does not require any prior knowledge about the problem in hand like how many classes will be produced as a final result. K-means mainly used to split the given data into k number of clusters.

K-Means(I, n, k)

Input: n= Total number of pixels of the given image to be clustered

I={I1,I2,….,In}: Image pixels

k = Numbers of segments and centroids are estimated by HSO

Output: segmented image

Begin:

Step1: K-no of cluster and c_i represent initial cluster for all i=1,2,…,n.

Step2: Compute the euclidian distance between all the data points and all the segment centers.

Step3: Among all the distance find out minimum distance and put the data point to that minimum distance cluster.

Step4: Cluster center was updated using the following formula $c_i = \frac{\sum_{j=1}^{v_j} x_j}{v_j}$

where, v_j is the number of data points in the i^{th} cluster.

Step5: Once again calculate the distance between all data points and updated cluster centers.

Step6: If there is no change between data point assignment then we arrive the solution so we can stop the algorithm, otherwise we have to repeat the steps 3-5.

Feature Extraction: Feature is bit of data which is used to solve the computational task. It may be points, lines, small object or anything in an image. All the features in the image are not important for doing classification, few features are less important and some are more important for classification. Selecting most important feature for doing efficient classification is the next important process in image processing other than segmentation. Color, shape and texture are the three important types of features, among these three in this paper we used texture feature. In texture feature also there are two approach structural and statistical, here we used statistical features. statistics is branch of mathematics deal with data. statistics have the following five function that is collection of data, grouping or organization of data, analysis of data, presentation of data and interpretation of data. Intensity based features are based on pixel values individual pixels. Here in our disease and deficiency detection problem intensity means gray level distribution of the disease and deficiency part of image. The following features that is average, energy, standard deviation, entropy and skewness are used to measure the intensities and its variations. Average can be calculated by adding all the pixel intensity value and divide the sum by no of pixels, it can be defined as follows:

$$\mu = \sum_{i=1}^{m}\sum_{j=1}^{n} p(i,j)$$

Where p(i, j) represent pixel value at point (i, j) of the given image of size M*N. The standard deviation is calculated based on the value of variance, first calculate variance and square root of the variance will give standard deviation. Standard deviation mainly used to show the contrast of the gray level intensities. In this research to extract correct features from the segmented image Gray-Level Co-Occurrence Matrix (GLCM) is used. GLCM used to extract texture features from image, this GLCM measure how frequently pairs of pixels with specific values and specified relationship occur in an image.

Classification: In this research work we used Support Vector Machine for classification of deficiency and diseased affected plant image. SVM is a supervised learning algorithm so we first train SVM with known deficiency image and it later used for testing.

IV. RESULT AND DISCUSSION

In our research, comparison of the performance of HSO based KM with other algorithms like K-Means, Fuzzy C-Means are carried out. We present the experimental results on the plant disease and deficiency dataset. The dataset consists of 100 images of tomato leaf along with expert suggestion about their disease and deficiency. From the result it is obvious that the proposed method produces better result than K-Means and Fuzzy C-Means.

Quantitative results: The following parameters are used to compare the performance of different image segmentation algorithm

True Positive (TP): Result is positive and our proposed method is also produced positive result.

True Negative (TN): Result is negative and our proposed method is also produced negative result.

False Positive (FP): Actual result is negative our Proposed segmentation algorithm Produce result as positive.

False Negative (FN): Actual result is positive our Proposed segmentation algorithm Produce result as negative.

$$Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative} X\ 100$$

$$Specificity = \frac{True\ Negative}{True\ Negative + False\ Positive} X\ 100$$

Accuracy=

$$\frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negativ} X\ 100$$

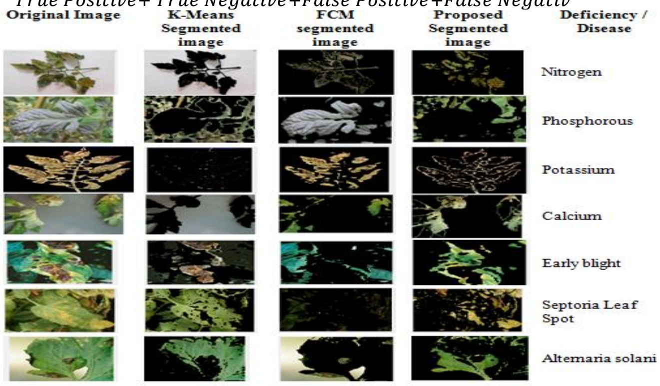

Fig1. Various Tomato leaf deficiency and disease and segmentation by using K-means, FCM and Proposed method

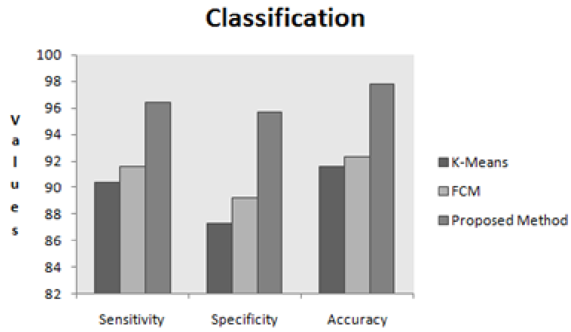

Fig 2. Comparison of average classification performance

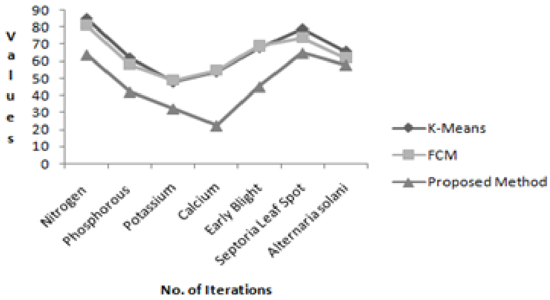

Fig 3. Comparison in number of Iterations

V. CONCLUSION

Image segmentation is one of the important parts of image processing. In this research, we proposed a hybrid K-Means algorithm with harmony search optimization technique for cluster center initialization. K-Means can easily predict disease and deficiency. To improve the performance of K-means we propose a hybrid approach that integrates Harmony search optimisation algorithm with K-means to detect the disease and deficiency accurately and minimum execution time. From the experimental results it is proved that our proposed HSO based K-means produce better result than traditional FCM and K-means algorithms. The proposed system determined the initial cluster centers from HSO which is used to minimize the execution time and gets the optimal solution.

REFERENCES:

1. Sona Kajla, Rajiv Bansal"Efficient Improved K means Clustering for Image Segmentation" International Journal of Innovative Research in Computer and Communication Engineering vol. 4, Issue 6, June 2016 pp 10625-69
2. K. A. Abdul Nazeer, M. P. Sebastian "Improving the Accuracy and Efficiency of the k-Means Clustering Algorithm" Proceedings of the World Congress on Engineering 2009 vol. I WCE 2009.
3. LahouaouiLalaouia, Tayeb. Mohamadib , and AbdelhakDjaalab " New Method for Image Segmentation" World Conference on Technology, Innovation and Entrepreneurship 2015 pp 1972-80
4. Md. Rakib Hassan, Romana Rahman Ema&Tajul Islam "Color Image Segmentation using Automated K-Means Clustering with RGB and HSV Color Spaces " Global journal of Computer Science and Technology: F Graphics and Vision 2017vol. 17, issue : 2, version:1.0 pp. 33-41
5. Zubair khan, jianjunni, xinnan fan and pengfeishi " An improved *k*-means clustering algorithm based onAn adaptive initial parameter estimation procedure For image segmentation" *International Journal of Innovative Computing, Information and Control* 2017 ISSN 1349-4198 Vol**13**, pp 1509-1525
6. Eman Abdel-Maksoud , Mohammed Elmogy, Rashid Al-Awadi " Brain tumor segmentation based on a hybrid clustering technique" Egyptian Informatics Journal 2015.
7. PasiFränti, Sami Sieranoja" How much can k- means be improved by using better initialization and repeats" Pattern recognition - Elsevier 2019 pp. 95-112.
8. Guiling Sun , Xinglong Jia , and Tianyu Geng " Plant Diseases Recognition Based on Image Processing Technology" Hindawi Journal of Electrical and Computer Engineering vol. 2018 pp. 1- 7
9. X. Z. Gao, V. Govindasamy, H. Xu, X. Wang, and K. Zenger "Harmony Search
10. Method: Theory and Applications" Hindawi Publishing Corporation Computational Intelligence and Neuroscience vol. 2015 pp. 1-10.
11. K. Preetha and Dr. S.K. Jeyanthi "GLCM and GLRLM based feature extraction t echnique in mammogram image" International Journal of Engineering & Technology
12. 2018, pp. 266-270
13. Osama Moh'd Alia & Rajeswari Mandava " the variants of the harmony search algorithm: an Overview" Artificial Intellignece Review, 2011 pp. 49-68
14. Cuevas, E., Ortega-Sánchez, N., Zaldivar, D., Pérez-Cisneros, M. Circle detection by Harmony Search Optimization, Journal of Intelligent and Robotic Systems*: Theory and Applications* 66 (3), (2013), pp. 359-376.
15. Zong Woo Geem, Kang Seok Lee and Yongjin Park " Application of Harmony Search To Vehicle Routing" American Journal of Applied Science, 2005 pp. 1552- 1557
16. Joong Hoon Kim, Zong Woo Geem, Eung Seok Kim "Parameter Estimation of the Nonlinear Muskingum Model using Harmony Search" Journal of the American Water Resources Association, vol.37 Issue.5 pp. 1131-1138
17. Kang Seok Lee, Zong Woo Geem, Sang- ho Lee and Kyu-Woong Bae "The harmony search heuristic algorithm for discrete
18. structural optimization" Engineering Optimization, 2005 vol.37 issue.7 pp. 663– 684.

ACKNOWLEDGEMENT:

The authors would like to thank Dr. H. Vijayaraghavan Principal and Professor (Crop Physiology) and Dr. S. Sathiyaraj AP (Crop Physiology) of Adhiyamaan College of Agriculture and Research for their valuable support for identifying disease and deficiency.

AUTHORS PROFILE

S.Sivagami, Assitant Professor of Computer Science Department at Adhiyamaan College of Agriculture and Research. She is Pursuing Ph.D in Periyar University Salem and her area of interest is Image Processing.

Dr. S. Mohanapriya, Head and Assistant Professor of Computer Science Department, KSR College of Arts and Science for Women, Tiruchngode. She has completed her Ph.D at Anna University, Chennai and her area of interest is MANET and Image Processing