Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 5669 - 5682
Received 25 April 2021; Accepted 08 May 2021.

Micrarray Image Segmentation Using Protracted K-Means Net Algorithm in Enhancement of Accuracy And Robustness

Leta Tesfaye Jule[1], A. Sampath Kumar[2], Krishnaraj Ramaswamy[3]
[1]Centre for Excellence in Indigenous Knowledge, Innovative Technology Transfer and Enterpreneurship and, Department of Physics, College of Natural and Computational Science, Dambi Dollo University, Ethiopia.
[2]Department of Computer Science and Engineering, Dambi Dollo University, Ethiopia..
[3]Centre for Excellence in Indigenous Knowledge, Innovative Technology Transfer and Enterpreneurship and, Department of Mechanical Engineering, College of Natural and Computational Science, Dambi Dollo University, Ethiopia.

Abstract

Microarray is the technology, which is used in various biological studies mainly in the field of image segmentation. The analysis of data in microarray has provide the result in the terms of accuracy and robustness.In this paper presents theprotracted K-means NeT an efficient technique for improving robustness and accuracy in images.The metrics which are consider for analysis are SNR, coefficient of variation, coefficient of determination and standard deviation.Protracted K-means NeT algorithm, takes advantage of the spatial information around each spot of interest, calculating the of the found foreground F through existing K-means algorithm. Subsequently the method analyses the neighbors at the boundary of F carefully, before making the final decision of including some of the noisy pixels to F. Here, Jeffrey's divergence metric is projected for dinding the intensity values. Experimental results and analysis, and comparison of the proposed method namely protracted K-means NeT Algorithm with existing methods such as K-Means , GMM and Multifeature shows the promising results such as 95%of SNR 1029.4 SSD 90% of CV, 521sec of MAE.

Keywords: SNR signal to noise ratio,cDNA Complementary DNA, GMM - GaussianMixture Model

1. Introduction

Microarray innovation permits concurrent estimation of thousands of qualities in a solitary investigation [1]. This gives a helpful apparatus to assessing the declaration of qualities and extraction of the portrayal and chromosomal primary data about these qualities. Microarrays are varieties of glass magnifying lens slides, in which a great many discrete DNA successions are

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 5669 - 5682
Received 25 April 2021; Accepted 08 May 2021.

printed by a mechanical cluster, in this way, shaping round spots of known width [2]. Each spot in the microarray picture contains the hybridization level of a solitary quality [3]. Any place, the measure of the fluorescence hybridization is influenced by things that occur during the assembling of cDNA microarray pictures [4], the productivity of the exploratory planning of the microarray pictures straightforwardly influences the accuracy of the microarray information investigation [5]. Microarray pictures handling consistently go through three stages: [6] (I) gridding to distinguish the situation of the spot focus of the picture and recognizes their directions, (ii) division, which sections, each microarray spot into closer view and foundation pixels, and (iii) force extraction to ascertain the frontal area fluorescence power and foundation powers [7]. The microarray picture division strategies can be ordered into four classes (I) Fixed and versatile circle, considers the spots with circle shape [8] , which is utilized in ScanAlyze and GenePix, (ii) Histogram-based strategy, it utilizes a circle target veil to cover all the closer view pixels, and figures a limit utilizing the Mann-Whitney test [9], (iii) Adaptive shape technique, performs picture division dependent on spatial likeness among pixels [10] (iv) Clustering strategy, as a most normal method, enjoys the benefit that they are not confined to a specific shape and size for the spots [11]. Since division is utilized for partitioning the picture into the districts of forefront and foundation, the quantity of group focuses k is set to two. As the underlying group communities, the pixels with least and greatest powers are chosen. All information focuses are then relegated to the closest bunch habitats as indicated by a distance measure (e.g., Euclidean distance). From there on, new group places are set to the mean of the pixel esteems in each bunch. At last, the calculation is iteratively rehashed until the bunch habitats stay unaltered. Portion thickness assessment KDE can be applied to discover their assessed densities in the wake of utilizing Gaussian combination model to establish the forefront and foundation. At that point, a remove point for fragmenting a spot into two groups is controlled by the balance of two assessed densities [12].Sampathkumar et al. Hybrid Cuckoo[13], Modified Honey Bee [14], Parallel Lion Optimizations [15] are Search algorithm has beenused to found the better accuracy by eliminating the localminimum optimal minimum problem. The Hill-Climbing technique for programmed gridding, communicated in [16], can perform gridding appropriately just if the ideal frameworks are available. However it is hard to fulfill this ideal circumstance practically speaking. Furthermore, the gridding techniques dependent on design grouping or knowledge calculation, for example, K-mean bunch, Fuzzy examination and Genetic

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 5669 - 5682
Received 25 April 2021; Accepted 08 May 2021.

calculation, be that as it may, are too computational to ever be applied [17]–[18][19]. Other notable methodologies depend on histogram division of pixel power data. What's more, they enjoy appealing benefits of computational proficiency and few info boundaries to programmed gridding, albeit the ideal edge isn't not difficult to be looked to recognize the fluorescence spots from the picture foundation impeccably [20], [21]. Particularly, the most extreme between-class fluctuation (i.e., the Otsu strategy) could give an exceptionally basic approach to programmed gridding of microarray pictures [22], [23]. It is by all accounts a straw for the gridding of microarray naturally. Tragically, the Otsu strategy is just an ideal method to accomplish the limit for histogram with bimodal or multimodal circulation, yet it would fall flat in affirming the division if the histogram is unimodal or near unimodal conveyance [24].

To overcome the issues in proposed methodology spot image is divided into regions of two classes: -foreground, background. Protracted K-means NeT algorithm, takes advantage of the spatial information around each spot of interest, calculating the of the found foreground F through existing K-means algorithm. are used to switch the spot for segmentation to the appropriate methods to effectively segment the foreground signals from the background in presence of noise, artifacts and weakly expressed spots in DNA microarray images.Also, the proposed work makes no assumption on the shape and the size of the spot.

Refinement of protracted K-means NeT algorithm, takes into account, intensity of pixels, which is the output of K-means method, as well as spatial information as in multifeature method. Pixels that are neighbors to foreground are analyzed carefully from the two angles: intensity and spatial neighborhood, before making final decision.Segmentation is the most crucial stage, which has to be as accurate as possible because this is the stage which affects the quantification of gene expression levels, from which biological conclusions are drawn.

2. Methdology

2.1 Protracted K-Means Net Algorithm

The idea of extending conventional K-means to protracted K-means NeT is particularly useful for noisy spot image. For such images, it is difficult to identify the foreground expression region (low SNR). Since the signal is weak and there is no marked transition between the foreground

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 5669 - 5682
Received 25 April 2021; Accepted 08 May 2021.

and background, the K-means algorithm fails to accurately segment the foreground region which leads to wrong estimation of intensities leading to wrong biological conclusion. Accurate gene signal intensity is essential, for its use in biological analysis. So, a method which works by gathering pixel by pixel information of components at the boundary is proposed, which accurately classifies the foreground pixels from the background.

Input : Spot image M

Output: Foreground image F and background image B

1: Find minimum intensity of the pixel in the obtained foreground F=a. (after applying K-means algorithm)

2: Threshold T=a*0.9.

3: Find (μ/σ) of the already found foreground F.

4: Apply K-means with new threshold (T) to find new foreground N and adjoining connected components

C1…….Cn (part of F) which are a adjacent to F.

5. Find (μ/σ) of each Ci.

6. Find average (μ/σ) of all the components, (μ/σ)avg

7. Join components Ci to F if(μ/σ) of Ci>=(μ/σ)average and get N_F.

8. Find (μ/σ) of N_F.

9. If (μ/σ)N_F> (μ/σ) $_F$ then N_F=F and continue steps 1 to 9 until (μ/σ)NF <= (μ/σ) $_F$

2.2Performance Evaluationof protracted K-means NET Algorithm

- To evaluate the clustering ability of the proposed protracted K-means NeT method versus the existing method, a simulated microarray image is created, consisting of 324 spots(subgrid), corrupted with Gaussian noise of five different signal to noise ratio levels 1,3,5,7 and 9 db.
- Additionally, the performances of the segmentation algorithms are estimated on five real microarray images with replicates drawn at different distinct time intervals of the same experiment.
- Metrics calculated from the simulated images.
- Metrics calculated from the real images.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 5669 - 5682
Received 25 April 2021; Accepted 08 May 2021.

2.3Refinement of Protracted K-Means NET Technique

The algorithm given below shows the steps for segmentation of image into foreground and background using Refinment of Protracted K-Means NET Technique

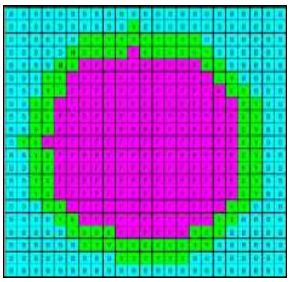


Figure 1: Refinement of NeT technique

Input: Spot image S

Output: Foreground F and background B image.

1:	Convert the input image to gray image G
2:	Use K-means on G and identify Foreground (F) and background(B) pixels.
3:	Find μF , σF , μB , σB
4:	Find the neighbors *x* of F on the boundary.
5:	**For** each pixel *p* Є x
6:	**If***Ip* (Intensity of i) Є $(\mu_F - \sigma_F, \mu_F + \sigma_F)$ **then***p* becomes F.
7:	Else **If***Ip* Є $(\mu_B - 3*\sigma_B, \mu_B + 3*\sigma_B)$ **then***p* becomes B
8:	**endif**
9:	**end for**
10:	**If** there is no pixel *x***then** stop

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 5669 - 5682
Received 25 April 2021; Accepted 08 May 2021.

11: Find pixels *N* neighbors of *x* and background pixel

12: **For** each pixel $p \in x$

13: $d_1 = | I_p - \mu_F |$

14: $d_2 = | I_p - I_N |$

15: **If** $d_1 < (0.5 * d_2)$

16: **then** *p* becomes F

17: **else** *p* becomes B

18: **endif**

19: **end for**

20: **end** Refined K-means

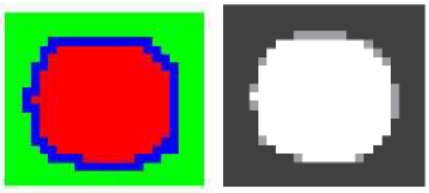


Figure 2: Foreground and Background Images

[Foreground (F) and its neighbors] [Some neighbors classified as background and some yet (shaded) to be decided }

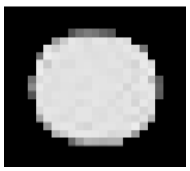


Figure 3: Final refined foreground

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 5669 - 5682
Received 25 April 2021; Accepted 08 May 2021.

3. Experimental Results and Analysis

3.1 Results for protracted K-means NeT Algorithm

Jeffrey's divergence metric values (in bits) between spot and background intensity values for the green channel, for the five evaluated cDNA images.

Table 1: Results for protracted K-means NeT Algorithm

Image	K-Means	GMM	Multifeature	Protracted K-means net
1c7b060rex2	8.82	7.45	5.13	11.42
1c4b064rex2	4.65	5.43	6.12	9.35
62919	6.78	5.64	7.35	10.56
40031	8.76	7.85	6.98	12.78
44004	9.32	7.45	6.42	11.32
17931	7.34	6.34	5.32	10.45
39119	9.78	7.23	6.53	12.32

Comparative results for three different spots randomly chosen from SMD using protracted K-means NeT, conventional K-means, GMM, and multifeature Techniques.

Table 2: Comparative results for three different spots

Spot Type	Original Spot	K-Mean	GMM	Multifeature	Extended K-Mean
Typical Spot					
Donut Spot					
Spot with Artifacts					

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 5669 - 5682
Received 25 April 2021; Accepted 08 May 2021.

r^2 Results on simulated microarray images

Tabel 3:Comparative results for SNR

SNR (dB)	K-Mean	GMM	Multifeature	Protracted K-means NeT
$\mathbf{r^2}$				
1	0.83	0.86	0.88	**0.93**
3	0.85	0.89	0.90	**0.95**
5	0.90	0.93	0.92	**0.97**
7	0.93	0.95	0.94	**0.98**
9	0.94	0.96	0.97	**0.99**
$\mathbf{p_c}$				
1	0.80	0.83	0.84	**0.87**
3	0.82	0.84	0.85	**0.88**
5	0.83	0.85	0.87	**0.91**
7	0.85	0.85	0.89	**0.93**
9	0.85	0.86	0.90	**0.95**

Table 4: A comparison of SSD values using different approaches for a real array spot image

Methods	SSD
K-Means	244.65
GMM	254.62
Multifeatured	751.04
protracted K-means NeT	**1029.4**

Table 5: MAE and CV results on real microarray images.

Methods	MAE	CV
K-Means	600	0.96
GMM	1110	0.98
Multifeatured	1213	1.10
protracted K-means NeT	**521**	**0.90**

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 5669 - 5682
Received 25 April 2021; Accepted 08 May 2021.

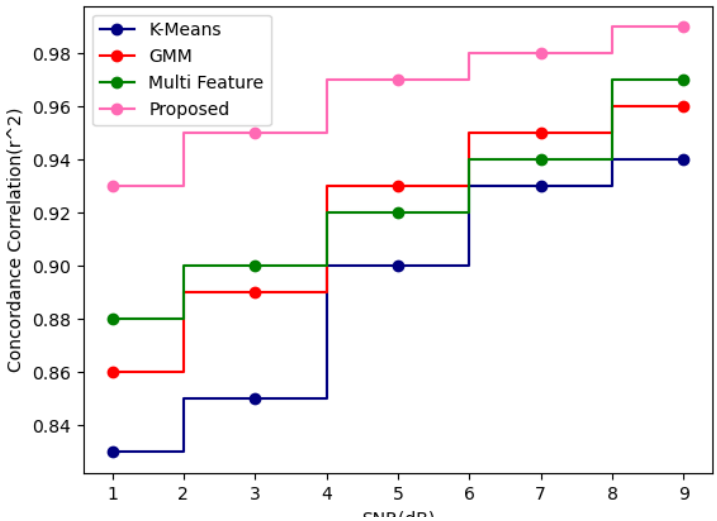


Figure 4: Comparison of Concordance Correlation(r^2)

concordance correlation calculated for GMM, K-means, GMM, K-means, Multifeature and protracted K-means NeT methods with varying noise levels

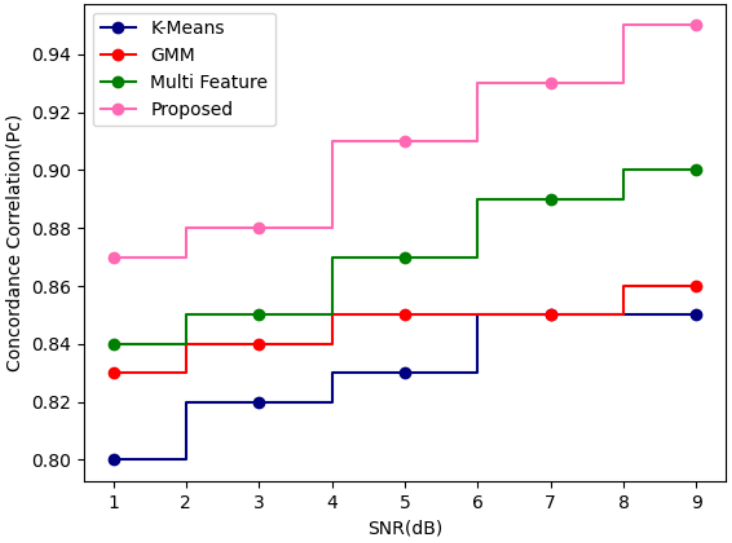


Figure 5: Comparison of Concordance Correlation(Pc)

Coefficient of determination, calculated for GMM,K-means, Multifeature and protracted K-means NeT methods with varying noise levels

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 5669 - 5682
Received 25 April 2021; Accepted 08 May 2021.

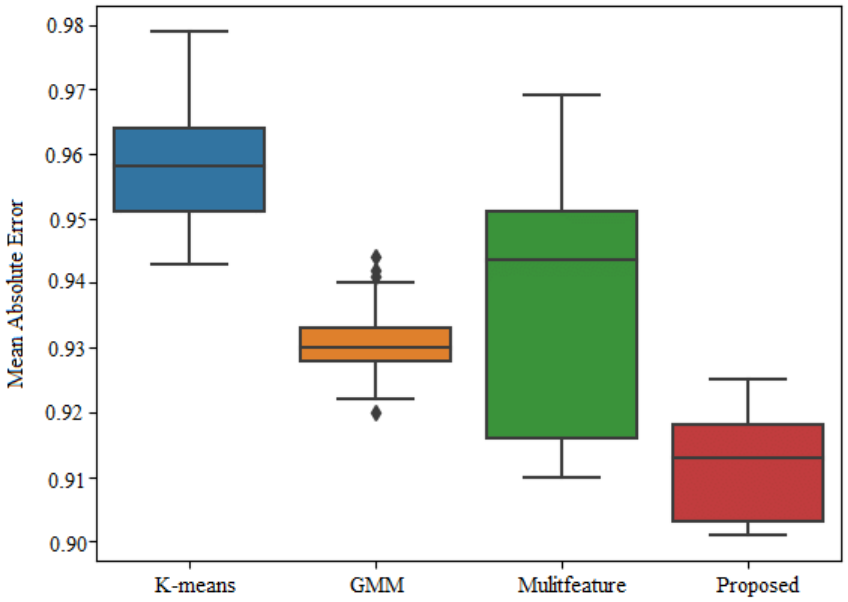


Figure 6: Comparison of Mean Absolute Error

Box plots that illuatrate the MAE using K-means, GMM,Multifeatured and proposed protracted K-means NeT method applied on five real microarray images

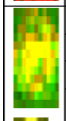
Spot Type	Methods	Mean FG	Mean BG	Mean /S.D FG)	SD FG	FG Pixels	BG Pixels
	K-means	218.02	28.43	**7.78**	28.42	30942	50698
	GMM	219.42	32.83	**6.85**	32.03	31992	49648
	Multifeature	219.34	30.98	**7.08**	30.98	31880	49760
	Proposed	218.14	37.01	**10.27**	21.24	32330	49310
	K-means	103.70	20.62	**5.02**	20.66	38183	32041
	GMM	116.39	34.63	**3.36**	34.64	33700	36524
	Multifeature	101.60	24.48	**4.15**	24.48	32852	37372
	Proposed	95.88	15.2	**6.30**	15.29	41978	28246
	K-means	213.58	23.24	**9.19**	23.26	27057	54583
	GMM	213.79	28.74	**7.43**	28.77	27646	53994
	Multifeature	212.25	26.70	**7.94**	26.73	27244	54396
	Proposed	214.15	22.28	**9.61**	22.28	28638	53002
	K-means	174.09	37.60	**4.63**	37.58	42577	39063
	GMM	190.63	46.99	**4.05**	47.06	54100	27540
	Multifeature	167.12	40.09	**4.17**	40.07	28444	53192
	Proposed	179.09	36.60	**4.89**	36.62	43493	38147
	K-means	212.18	84.43	**8.16**	26.00	38418	68656
	GMM	210.43	48.93	**5.76**	36.45	58573	48501
	Multifeature	208.57	83.38	**6.78**	30.76	40102	66972
	Proposed	213.68	85.10	**8.85**	24.14	37612	69462

Figure7: Comparison of Segmentation Performances

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 5669 - 5682
Received 25 April 2021; Accepted 08 May 2021.

Comparison of segmentation performances of proposed protracted K-means NeT, K-means, GMM, and Multifeature on real microarray images drawn from SMD.

3.2Experimental Results and Analysis of Refinement protracted K-means NeT Algorithm:

Probability Density Functions (PDF) of the Coefficient of Variations for all from 5 replications of common reference channel. Black line corresponds to the results obtained using the proposed protracted K-means NeT method. Blue, red and green line corresponds to the results obtained using K-means, GMM and Multi feature methods respectively.

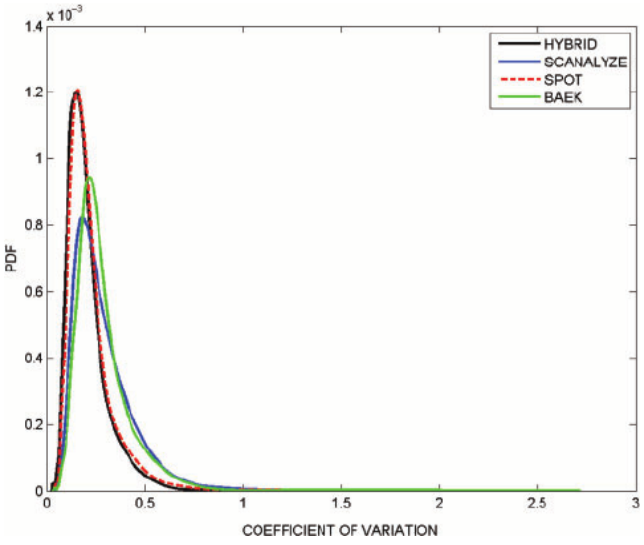


Figure 8: Comparison of Probability Density Function.

Results of these studies have indicated a superior quality of the enhanced images, without however examining whether enhancement leads to more accurate spot segmentation compared to the all existing methods.

4. Conclusion

Predicting and effective classification of images will solve the issue of complex biological processes. Across fields like gene identification, cancer detection and disease diagnosis, prediction and treatment, there has been a widespread application of the microarray data. This can in turn trigger the development of medicines at a later stage. As the sample size id extremely small and the data is of high dimensionality, the classification problem is time

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 5669 - 5682
Received 25 April 2021; Accepted 08 May 2021.

consuming. The running time is reduced and the precision of forecast improves when the feature selection is performed before classification. The proposed methodology has two segmentation techniques of microarray image analysis facilitate efficient intensity extraction results and the validity of gene expression levels, thus allowing biologists to draw meaningful conclusions. Experimental results and analysis, and comparison of the proposed method namely protracted K-means NeT Algorithm with existing methods such as K-Means , GMM and Multifeature shows the promising results such as 95%of SNR 1029.4 SSD 90% of CV, 521sec of MAE. Therefore, the results obtained in the experimental analysis shows the protracted K-means NeT method has 3 to 4 percent accuracy deviations when compared to the previous methodology.

References

[1] Shao, Guifang, et al. "Automatic microarray image segmentation with clustering-based algorithms." *PloS one* 14.1 (2019): e0210075.

[2] Sivalakshmi, B., and N. Naga Malleswara Rao. "Gridding and Segmentation Method for DNA Microarray Images." *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 12.5 (2021): 618-628.

[3] Farouk, R. M., and M. A. SayedElahl. "Microarray spot segmentation algorithm based on integro-differential operator." *Egyptian Informatics Journal* 20.3 (2019): 173-178.

[4] Song, Yujing, et al. "Machine learning-based cytokine microarray digital immunoassay analysis." *Biosensors and Bioelectronics* 180 (2021): 113088.

[5] Farouk, R. M., and M. A. SayedElahl. "Robust cDNA microarray image segmentation and analysis technique based on Hough circle transform." *arXivpreprint arXiv:1603.07123* (2016).

[6] Reddy, T. Srinivas. "Technical Investigation, Analysis and cDNA Microarray Image Segmentation Based on Hough Circle Transform."

[7] Li, Tiejun, et al. "Contrast enhancement for cDNA microarray image based on fourth-order moment." *Signal, Image and Video Processing* 12.6 (2018): 1069-1077.

[8] Kashyap, Ramgopal, and Vivek Tiwari. "Energy-based active contour method for image segmentation." *International Journal of Electronic Healthcare* 9.2-3 (2017): 210-225.

[9] Carreras, Joaquim, et al. "A Combination of Multilayer Perceptron, Radial Basis Function Artificial Neural Networks and Machine Learning Image Segmentation for the

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 5669 - 5682
Received 25 April 2021; Accepted 08 May 2021.

Dimension Reduction and the Prognosis Assessment of Diffuse Large B-Cell Lymphoma." *AI* 2.1 (2021): 106-134.

[10] Guo, Zhiqing, et al. "Image processing of porous silicon microarray in refractive index change detection." *Sensors* 17.6 (2017): 1335.

[11] Li, Yang, Andrei Păun, and Mihaela Păun. "Improvements on contours based segmentation for DNA microarray image processing." *Theoretical Computer Science* 701 (2017): 174-189.

[12] A. Sampathkumar, Mulerikkal, J. & Sivaram, M. Glowworm swarm optimization for effectual load balancing and routing strategies in wireless sensor networks. Wireless Networks, vol. 26, no. 6, 4227–4238 (2020). https://doi.org/10.1007/s11276-020-02336-w

[13] A. Sampathkumar, Rastogi, R., Arukonda, S. Achyut Shankar, Sandeep Kautish & M. Sivaram "An efficient hybrid methodology for detection of cancer-causing gene using CSC for micro array data" Journal of Ambient Intelligence and Human Comput (2020),Springer. https://doi.org/10.1007/s12652-020-01731-7.

[14] A. Sampathkumar, Vivekanandan. P "Gene Selection Using PLOA Method In Microarray Data For Cancer Classification" Journal of Medical Imaging and Health Informatics (2019). 9, 1294-1300.

[15] Sampathkumar. A, Vivekanandan. P "Gene Selection Using Multiple Queen Colonies In Large Scale Machine Learning" Journal of Electrical Engineering (2018). 9 (6), 97-111.

[16] L. Rueda and V. Vidyadharan, "A hill-climbing approach for automatic gridding of cDNA microarray images", *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 3, no. 1, pp. 72-83, Jan./Mar. 2006.

[17] E. Zacharia and D. Maroulis, "An original genetic approach to the fully automatic gridding of microarray images", *IEEE Trans. Med. Imag.*, vol. 27, no. 6, pp. 805-813, Jun. 2008.

[18] A. K. Helmy and G. S. El-Taweel, "Regular gridding and segmentation for microarray images", *Comput. Electr. Eng.*, vol. 39, no. 7, pp. 2173-2182, Oct. 2013.

[19] N. Zeng, Z. Wang and H. Zhang, "Inferring nonlinear lateral flow immunoassay state-space models via an unscented Kalman filter", *Sci. China Inf. Sci.*, vol. 59, no. 11, Nov. 2016.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 5669 - 5682
Received 25 April 2021; Accepted 08 May 2021.

[20] J. C. Liu and T. M. Lin, "Location and image-based plant recognition and recording system", *J Inform Hiding Multimedia Signal Process.*, vol. 6, no. 5, pp. 898-910, Sep. 2015.

[21] E. Küçükkülahli, Pakize Erdoğmuş and K. Polat, "Histogram-based automatic segmentation of images", *Neural Comput. Appl.*, vol. 27, no. 5, pp. 1445-1450, Apr. 2016.

[22] C. C. Charalambous and G. K. Matsopoulos, "A new method for gridding DNA microarrays", *Comput. Biol. Med.*, vol. 43, no. 10, pp. 1303-1312, Oct. 2013.

[23] J.-S. Pan, Q. Feng, L. Yan and L.-F. Yang, "Neighborhood feature line segment for image classification", *IEEE Trans. Circuits Sys. Video Technol.*, vol. 25, no. 3, pp. 387-398, Mar. 2015.

[24] J.-L. Fan and B. Lei, "A modified valley-emphasis method for automatic thresholding", *Pattern Recognit. Lett.*, vol. 33, no. 6, pp. 703-708, Apr. 2012.