ANALYSIS AND COMPARISON OF K-MEANS CLUSTERING SEGMENTATION TECHNIQUE FOR IMPROVING THE ACCURACY OF BLOOD SMEAR IMAGES OVER THRESHOLD TECHNIQUE

P. Naga Ramya Bhargavi[1], R.Ramadevi[2*]

Article History: Received: 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

Abstract

Aim: Aim of this research work is to improve accuracy of blood smear image by applying novel k-means clustering algorithms and threshold segmentation technique.
Materials and Methods: This investigation made use of a collection of data from Kaggle's Website. Samples were considered as (N=30) for K-means clustering technique and (N=30) for threshold technique with total sample size calculated using clinical.com. As a result the total number of samples was calculated to be 60. Using IBM SPSS Software and a standard data set, the accuracy was obtained. Both segmentation techniques were implemented on blood smear images through Matlab coding and also extracting accuracy values of each image. Then through SPSS software comparison and analysis has been made.
Results: The accuracy (%) value of both segmentation techniques are compared using SPSS software by independent sample t-tests. There is statistical insignificance with an accuracy of (99.8810%) and demonstrated a better outcome in comparison to threshold technique accuracy (58.6514%).
Conclusion: Novel K-means clustering segmentation technique appears to give better accuracy than that of the threshold technique on blood smear images.

Keywords: Image Processing, Image Segmentation, Novel K-means Clustering Technique, Threshold Technique, Clustering, Blood Smear Image.

[1]Research scholar, Department of Biomedical Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India. Pincode: 602105.
[2*]Project Guide, Corresponding Author, Department of Biomedical Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India. Pincode: 602105.

1. Introduction

The blood smear is a drop of blood, innovatively spread thinly onto a glass slide that is then treated with a special stain, and the blood cells on the slide are examined and evaluated (Padma Suresh and Panigrahi 2015). Traditionally, trained laboratorians have examined blood smears manually using a microscope. More recently, automated digital systems have become available to analyze blood smears more efficiently (Saha, Bajger, and Lee 2016). A blood smear is a snapshot of the cells that are present in the blood at the time the sample is obtained. The blood smear analysis allows for the evaluation of these cells (Al-Qudah and Suen 2021; Padma Suresh and Panigrahi 2015; Rahman et al. 2021; Trivedi et al. 2017). These cell populations are produced and mainly mature in the bone marrow and are eventually released into the blood as needed (Melin et al. 2007). This research can be carried out and the type of each cell present in the blood is dynamic but is generally maintained by the body within specific ranges (Bhattacharyya et al. 2017). The drop of blood on the slide used for a blood smear contains millions of RBCs, thousands of WBCs, and hundreds of thousands of platelets (Padma Suresh and Panigrahi 2015).

This study has been linked to 34 google scholar and 150 science direct articles were seen in websites. In recent years, medical image segmentation work is carried out reporting the developed algorithms, models utilizing machine learning such as clustering of images, and access to their performance. In this research k-means clustering algorithm and threshold technique has been encoded with machine learning which is the easiest and simplest method to form the clustering of blood smear image. There have been numerous works that exhausted the zone of picture division by utilizing distinctive strategies by picture segmentation. The K-means algorithm is one of the best clustering techniques (Braiki et al. 2020; "Segmentation of Blood Cell Images Using Hybrid K-Means with Cluster Center Estimation Technique" 2019). These techniques are used in improving imaging, including volume rendered images from computed tomography and magnetic resonance imaging such as locating tumors and other pathologies, which measure tissue volumes.

Our institution is passionate about high quality evidence based research and has excelled in various domains (Vickram et al. 2022; Bharathiraja et al. 2022; Kale et al. 2022; Sumathy et al. 2022; Thanigaivel et al. 2022; Ram et al. 2022; Jothi et al. 2022; Anupong et al. 2022; Yaashikaa, Keerthana Devi, and Senthil Kumar 2022; Palanisamy et al. 2022)(Vickram et al. 2022; Bharathiraja et al. 2022; Kale et al. 2022; Sumathy et al. 2022; Thanigaivel et al. 2022; Ram et al. 2022; Jothi et al. 2022; Anupong et al. 2022; Yaashikaa, Keerthana Devi, and Senthil Kumar 2022; Palanisamy et al. 2022).The methods which were used before have less accuracy on medical image segmentation which includes blood smear image. Hence the aim of this research work is to improve the accuracy of blood smear image segmentation using the novel K-means clustering algorithm and compare with threshold algorithms (Romadhon and Kurniawan 2021).

2. Materials and Methods

This research work was carried out at the department of Biomedical Engineering, Saveetha School of Engineering. This study was implemented using Matlab software. Two groups are required for this study. The accuracy in image segmentation is performed by evaluating two groups. A total of 30 iterations were performed on each group to achieve better accuracy. The study uses a blood smear images dataset downloaded from kaggle website (Varma et al. 2021) It is not necessary to obtain ethical approval. The sample size calculation was done using clinicalc.com by keeping alpha error threshold as 0.05 with 80% g power and value for enrollment ratio is 0.1 with 95% confidence Interval. As per calculation total sample size is 60.

In sample preparation for Group1, 30 blood smear image samples are collected and used for accuracy based image segmentation using thresholding technique. For Group 2 sample preparation, 30 blood smear images were collected and segmentation was done by K-means clustering algorithm. Testing setup is done by installing the Matlab R2018a software. After preparation a Matlab code was implemented for both algorithms and then the performance measure which is accuracy in terms of percentage, also calculated for each image. This parameter was calculated and evaluated to assess the method's efficacy and comparison of results was done in order to find which algorithm performed significantly better results for segmenting images.

Statistical Analysis

The SPSS statistical software was used in the research for statistical analysis. Group statistics and independent sample t-tests were performed on the experimental results and the graph was built for two groups with one parameter under study (Nayak et al. 2021). The independent variable is noise which affects the contrast of an edge. Majorly segmentation depends on the edge of an image and here the dependent variable is segmentation accuracy.

3. Results

Table 1 shows accuracy(%) values of 30 blood smear images obtained using k-means clustering algorithm and thresholding technique. The mean value of accuracy(%) of the k-means algorithm is 99.8810% and for the thresholding technique is 58.65%.

Table 2 represents group statistics that shows comparison of k-means clustering algorithms with thresholding techniques based on accuracy values. Mean value of accuracy is high (99.8810%) for k-means clustering algorithm and low (58.6514%) for thresholding technique. Standard error mean of accuracy is low (0.04937) for k-means clustering algorithm and high (0.19526) for thresholding technique .

Table 3 represents an independent sample test of accuracy based image segmentation using k-means clustering and thresholding technique. The two tailed significance p-value is 0.003. The mean difference is 3.4700 with a standard error difference of 1.10176. In the test confidence interval, the lower value is (1.2396) and the upper value is (5.700). Based on independent sample T test results, k-means clustering based image segmentation method performed better than thresholding technique with 95% confidence interval.

4. Discussion

Figure 1 represents the results of the k-means clustering and thresholding algorithm on blood smear image. (a) Input image (Blood smear image), (b) Output image of k-means clustering algorithm (c) Output image of Thresholding Technique

Figure 2 shows bar chart representation of the comparison between novel k-means clustering technique and thresholding technique on blood smear image. X-Axis represents k-means clustering and thresholding technique and Y-Axis represents mean accuracy of both techniques with (+/-2 SD) and 95% CI.

Many researchers have proposed a model based segmentation using threshold technique algorithms, with the purpose of assessing the model's accuracy and precision (Tran et al. 2018). With a precision of 98 percent and an accuracy of 54 percent, the findings were accomplished (Eliza and K. 2019). Another investigation developed a computer-aided segmentation approach based on the logistic regression and threshold technique (Hassanien and Oliva 2017) used feature selection and a principal component analysis model to correctly identify segmentation with an accuracy of 90%.

Actual (Class 1) and Actual (Class 2) selections can be reduced using clustering approaches (Patgiri and Ganguly 2021). (Thakore and Bhatt, n.d.) carried out a comparison study. The experimental outcomes revealed that k-means had the highest accuracy of 98.9 percent over the clinical threshold technique data set, out of numerous classifiers. In this research paper, Novel k-means clustering technique (99.98%) outperformed threshold technique (58.65%) in terms of accuracy on blood smear images.

Major drawback with segmentation is loss of valuable information especially from the edges of the input image. In this research only blood cells were segmented from smear images. In the future, it can be taken to classify the type of blood cell/normal/abnormal cell using machine learning or deep learning. Thus it may be used in hospitals and testing facilities.

5. Conclusions

In this image segmentation study, Matlab-based K-means clustering algorithm generated superior results (99.94%) than threshold technique (58.61%) in terms of segmentation accuracy performed on blood smear images.

Declaration

Conflict of Interest

In this manuscript, there are no conflicts of interest.

Authors Contribution

Author PNRB was involved in data collection, data analysis, manuscript writing. Author RR was involved in conceptualization, data validation, and critical review of manuscript.

Acknowledgement

The authors would like to thank the Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (formerly known as Saveetha University) management for providing us all the necessary facilities to complete this project successfully.

Funding: We thank the following organizations for providing financial support that enabled us to complete this study.

1. Axia clinics Pvt.Ltd, Hyderabad.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

6. References

Al-Qudah, Rabiah, and Ching Y. Suen. 2021.

"Improving Blood Cells Classification in Peripheral Blood Smears Using Enhanced Incremental Training." *Computers in Biology and Medicine* 131 (April): 104265.

Anupong, Wongchai, Lin Yi-Chia, Mukta Jagdish, Ravi Kumar, P. D. Selvam, R. Saravanakumar, and Dharmesh Dhabliya. 2022. "Hybrid Distributed Energy Sources Providing Climate Security to the Agriculture Environment and Enhancing the Yield." *Sustainable Energy Technologies and Assessments*. https://doi.org/10.1016/j.seta.2022.102142.

Bharathiraja, B., J. Jayamuthunagai, R. Sreejith, J. Iyyappan, and R. Praveenkumar. 2022. "Techno Economic Analysis of Malic Acid Production Using Crude Glycerol Derived from Waste Cooking Oil." *Bioresource Technology* 351 (May): 126956.

Bhattacharyya, Siddhartha, Anirban Mukherjee, Indrajit Pan, Paramartha Dutta, and Arup Kumar Bhaumik. 2017. *Hybrid Intelligent Techniques for Pattern Analysis and Understanding*. CRC Press.

Braiki, Marwa, Abdesslam Benzinou, Kamal Nasreddine, and Nolwenn Hymery. 2020. "Automatic Human Dendritic Cells Segmentation Using K-Means Clustering and Chan-Vese Active Contour Model." *Computer Methods and Programs in Biomedicine* 195 (October): 105520.

Eliza, Thenu, and Sreekumar K. 2019. "Malaria Parasite Detection from Microscopic Blood Smear Image: A Literature Survey." *International Journal of Computer Applications*. https://doi.org/10.5120/ijca2019918741.

Hassanien, Aboul Ella, and Diego Alberto Oliva. 2017. *Advances in Soft Computing and Machine Learning in Image Processing*. Springer.

Jothi, K. Jeeva, K. Jeeva Jothi, S. Balachandran, K. Mohanraj, N. Prakash, A. Subhasri, P. Santhana Gopala Krishnan, and K. Palanivelu. 2022. "Fabrications of Hybrid Polyurethane-Pd Doped ZrO2 Smart Carriers for Self-Healing High Corrosion Protective Coatings." *Environmental Research*. https://doi.org/10.1016/j.envres.2022.113095.

Kale, Vaibhav Namdev, J. Rajesh, T. Maiyalagan, Chang Woo Lee, and R. M. Gnanamuthu. 2022. "Fabrication of Ni–Mg–Ag Alloy Electrodeposited Material on the Aluminium Surface Using Anodizing Technique and Their Enhanced Corrosion Resistance for Engineering Application." *Materials Chemistry and Physics*. https://doi.org/10.1016/j.matchemphys.2022. 125900.

Melin, Patricia, Oscar Castillo, Eduardo G. Ramírez, and Witold Pedrycz. 2007. *Analysis and Design of Intelligent Systems Using Soft Computing Techniques*. Springer Science & Business Media.

Nayak, Janmenjoy, Margarita N. Favorskaya, Seema Jain, Bighnaraj Naik, and Manohar Mishra. 2021. *Advanced Machine Learning Approaches in Cancer Prognosis: Challenges and Applications*. Springer Nature.

Padma Suresh, L., and Bijaya Ketan Panigrahi. 2015. *Proceedings of the International Conference on Soft Computing Systems: ICSCS 2015, Volume 1*. Springer.

Palanisamy, Rajkumar, Diwakar Karuppiah, Subadevi Rengapillai, Mozaffar Abdollahifar, Gnanamuthu Ramasamy, Fu-Ming Wang, Wei-Ren Liu, Kumar Ponnuchamy, Joongpyo Shim, and Sivakumar Marimuthu. 2022. "A Reign of Bio-Mass Derived Carbon with the Synergy of Energy Storage and Biomedical Applications." *Journal of Energy Storage*. https://doi.org/10.1016/j.est.2022.104422.

Patgiri, Chayashree, and Amrita Ganguly. 2021. "Adaptive Thresholding Technique Based Classification of Red Blood Cell and Sickle Cell Using Naïve Bayes Classifier and K-Nearest Neighbor Classifier." *Biomedical Signal Processing and Control*. https://doi.org/10.1016/j.bspc.2021.102745.

Rahman, Samiur, Bakht Azam, Sana Ullah Khan, Muhammad Awais, Ishtiaq Ali, and Raja Jalees Ul Hussen Khan. 2021. "Automatic Identification of Abnormal Blood Smear Images Using Color and Morphology Variation of RBCS and Central Pallor." *Computerized Medical Imaging and Graphics: The Official Journal of the Computerized Medical Imaging Society* 87 (January): 101813.

Ram, G. Dinesh, G. Dinesh Ram, S. Praveen Kumar, T. Yuvaraj, Thanikanti Sudhakar Babu, and Karthik Balasubramanian. 2022. "Simulation and Investigation of MEMS Bilayer Solar Energy Harvester for Smart Wireless Sensor Applications." *Sustainable Energy Technologies and Assessments*. https://doi.org/10.1016/j.seta.2022.102102.

Romadhon, Manzilur Rahman, and Fachrul Kurniawan. 2021. "A Comparison of Naive Bayes Methods, Logistic Regression and KNN for Predicting Healing of Covid-19 Patients in Indonesia." *2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT)*.

https://doi.org/10.1109/eiconcit50028.2021.9431845.

Saha, Ratna, Mariusz Bajger, and Gobert Lee. 2016. "Spatial Shape Constrained Fuzzy C-Means (FCM) Clustering for Nucleus Segmentation in Pap Smear Images." *2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*. https://doi.org/10.1109/dicta.2016.7797086.

"Segmentation of Blood Cell Images Using Hybrid K-Means with Cluster Center Estimation Technique." 2019. *International Journal of Recent Technology and Engineering*. https://doi.org/10.35940/ijrte.b1026.0982s1119.

Sumathy, B., Anand Kumar, D. Sungeetha, Arshad Hashmi, Ankur Saxena, Piyush Kumar Shukla, and Stephen Jeswinde Nuagah. 2022. "Machine Learning Technique to Detect and Classify Mental Illness on Social Media Using Lexicon-Based Recommender System." *Computational Intelligence and Neuroscience* 2022 (February): 5906797.

Thakore, Shuchen B., and Bharat I. Bhatt. n.d. *Introduction to Process Engineering and Design*. McGraw-Hill Education.

Thanigaivel, Sundaram, Sundaram Vickram, Nibedita Dey, Govindarajan Gulothungan, Ramasamy Subbaiya, Muthusamy Govarthanan, Natchimuthu Karmegam, and Woong Kim. 2022. "The Urge of Algal Biomass-Based Fuels for Environmental Sustainability against a Steady Tide of Biofuel Conflict Analysis: Is Third-Generation Algal Biorefinery a Boon?" *Fuel*. https://doi.org/10.1016/j.fuel.2022.123494.

Tran, Thanh, Oh-Heum Kwon, Ki-Ryong Kwon, Suk-Hwan Lee, and Kyung-Won Kang. 2018. "Blood Cell Images Segmentation Using Deep Learning Semantic Segmentation." *2018 IEEE International Conference on Electronics and Communication Engineering (ICECE)*. https://doi.org/10.1109/icecome.2018.8644754.

Trivedi, Sheifalee, Bhumika Nandwana, Dinesh Kumar Khunteta, and Satya Narayan. 2017. "K-Means Clustering with Adaptive Threshold for Segmentation of Hand Images." *2017 7th International Conference on Communication Systems and Network Technologies (CSNT)*. https://doi.org/10.1109/csnt.2017.8418534.

Varma, C. Ganesh, C. Ganesh Varma, P. Nagaraj, V. Muneeswaran, Sudar K. Muthamil, M. Mokshagni, and M. Jaswanth. 2021. "Astute Segmentation and Classification of Leucocytes in Blood Microscopic Smear Images Using Titivated K-Means Clustering and Robust SVM Techniques." *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*. https://doi.org/10.1109/iciccs51141.2021.9432309.

Vickram, Sundaram, Karunakaran Rohini, Krishnan Anbarasu, Nibedita Dey, Palanivelu Jeyanthi, Sundaram Thanigaivel, Praveen Kumar Issac, and Jesu Arockiaraj. 2022. "Semenogelin, a Coagulum Macromolecule Monitoring Factor Involved in the First Step of Fertilization: A Prospective Review." *International Journal of Biological Macromolecules* 209 (Pt A): 951–62.

Yaashikaa, P. R., M. Keerthana Devi, and P. Senthil Kumar. 2022. "Algal Biofuels: Technological Perspective on Cultivation, Fuel Extraction and Engineering Genetic Pathway for Enhancing Productivity." *Fuel*. https://doi.org/10.1016/j.fuel.2022.123814.

Tables and Figures

Table 1. Accuracy(%) values of k-means clustering and thresholding technique of various blood smear images

Samples	K-means clustering	Threshold technique
1	100	58.6
2	100	58.02
3	99.98	57.8

4	99.98	57.3
5	100	57.8
6	100	59.4
7	100	60.1
8	100	57.2
9	99.99	56.7
10	100	57.9
11	100	59.3
12	100	57.5
13	100	58.7
14	100	59.1
15	100	58.7
16	100	58.7
17	100	56.9
18	100	60.2
19	100	60.1
20	100	58.2
21	100	59.6
22	100	57.4
23	98.55	58.7

24	100	60.9
25	100	57.4
26	99.80	58.4
27	99.99	59.2
28	100	58.9
29	100	60.0
30	100	60.6

Table 2. Group statistics shows comparison of k-means clustering algorithms with thresholding techniques based on accuracy. Mean value of accuracy is high 99.8810% for k-means clustering and low 58.6514% for thresholding technique. Standard deviation error mean of accuracy is low for k-means (.04937) and high (.19526) for thresholding technique.

	Group	**N**	**Mean**	**Standard Deviation**	**Standard Error Mean**
Accuracy	K-means Clustering	30	99.8810	.27040	.04937
	Thresholding Technique	30	58.6514	1.06947	.19526

Table 3. Independent sample t-test providing mean difference, significance value (2-tailed), standard error difference and 95% confidence interval of the difference in both lower and upper level for accuracy at equal variances assumed and not assumed.

Parameters	**Levene's test for Equality of variance**		**T-Test for equality of mean**							
			t	**df**	**One-sided p**	**Two sided p**	**Mean difference**	**Std.Error Difference**	**95% confidence of Difference**	
	F	**Sig**							**Lower**	**Upper**

Accuracy	**Equal variances assumed**	6.378	<0.06	3.150	38	0.02	.003	3.4700	1.10176	1.2396	5.700
	Equal Variances not assumed			3.150	30.83	<0.02	.004	3.4700	1.10176	1.2224	5.717

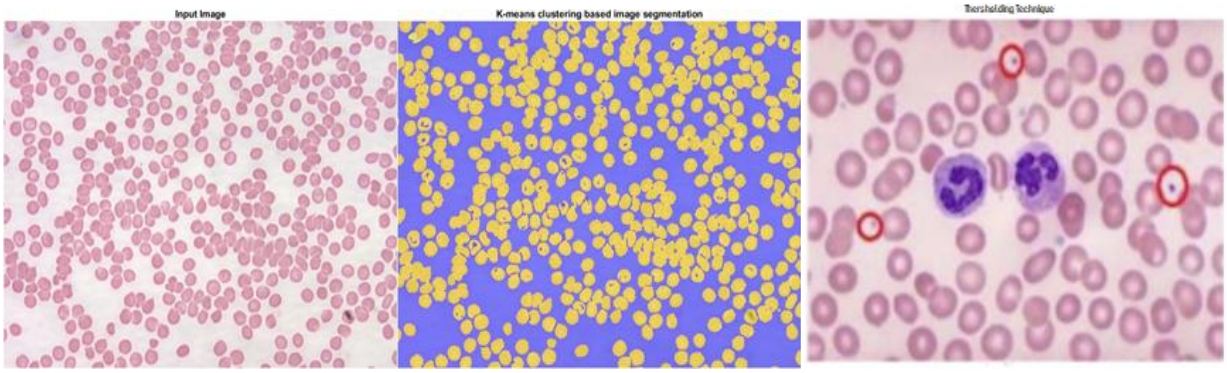

Fig. 1. Simulation results of k-means clustering and thresholding algorithm. (a) Input image (Blood smear image), (b) Output image of k-means clustering algorithm (c) Output image of Thresholding Technique

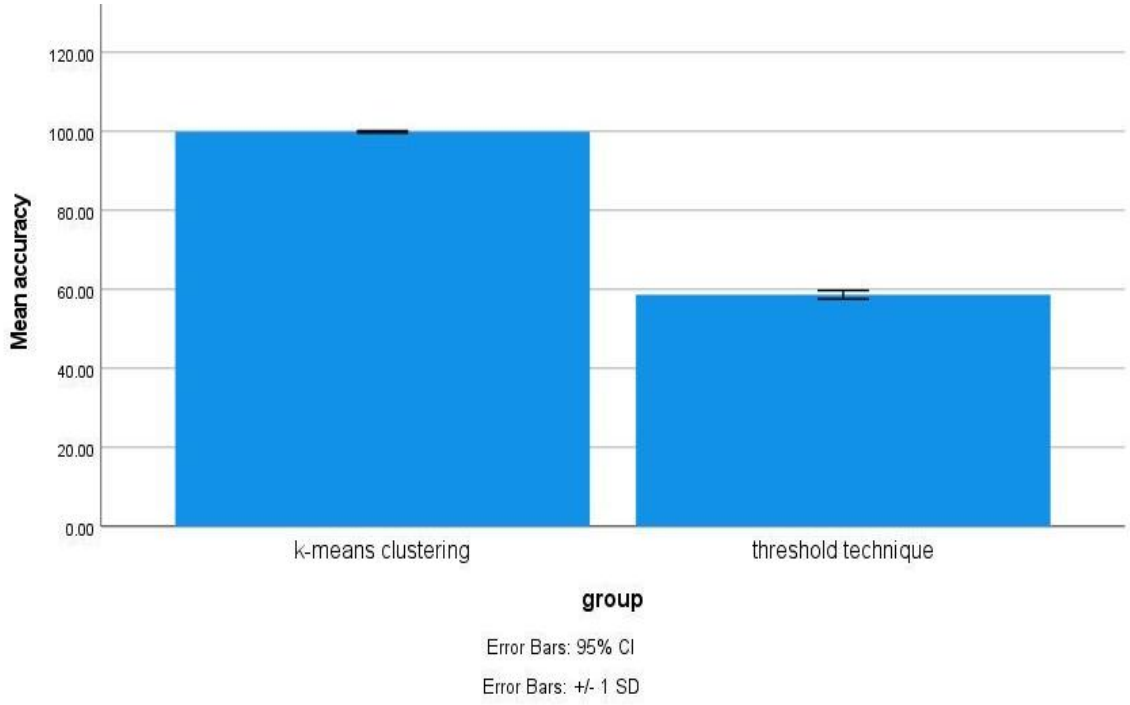

Fig. 2. Bar chart representation of the comparison between novel k-means clustering technique and thresholding technique. X-Axis represents k-means clustering and thresholding technique and Y-Axis represents mean accuracy of both technique with (+/-2 SD) and 95% CI