



Identifying the presence of bacteria on digital images by using asymmetric distribution with k-means clustering algorithm

K. V. Satyanarayana¹ · N. Thirupathi Rao² · Debnath Bhattacharyya³ · Yu-Chen Hu⁴ 

Received: 3 February 2021 / Revised: 8 September 2021 / Accepted: 19 September 2021 /
Published online: 9 October 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

This paper is mainly aimed at the decomposition of image quality assessment study by using Three Parameter Logistic Mixture Model and k-means clustering (TPLMM-k). This method is mainly used for the analysis of various images which were related to several real time applications and for medical disease detection and diagnosis with the help of the digital images which were generated by digital microscopic camera. Several algorithms and distribution models had been developed and proposed for the segmentation of the images. Among several methods developed and proposed, the Gaussian Mixture Model (GMM) was one of the highly used models. One can say that almost the GMM was playing the key role in most of the image segmentation research works so far noticed in the literature. The main drawback with the distribution model was that this GMM model will be best fitted with a kind of data in the dataset. To overcome this problem, the TPLMM-k algorithm is proposed. The image decomposition process used in the proposed algorithm had been analyzed and its performance was analyzed with the help of various performance metrics

✉ Yu-Chen Hu
ychu@pu.edu.tw

K. V. Satyanarayana
vsatyanarayana.kalahasthi@gmail.com

N. Thirupathi Rao
nakkathiru@gmail.com

Debnath Bhattacharyya
debnathb@gmail.com

¹ Department of Computer Science and Engineering, RAGHU Engineering College (A), Visakhapatnam, AP, India

² Department of Computer Science and Engineering, Vignan's Institute of Information Technology (A), Visakhapatnam 530049, India

³ Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Greenfield, Vaddeswaram, Guntur 522502, India

⁴ Department of Computer Science and Information Management, Providence University, 200, Sec. 7, Taiwan Boulevard, Shalu Dist, Taichung City 43301, Taiwan, Republic of China

like the Variance of Information (VOI), Global Consistency Error (GCE) and Probabilistic Rand Index (PRI). According to the results, it is shown that the proposed algorithm achieves the better performance when compared with the previous results of the previous techniques. In addition, the decomposition of the images had been improved in the proposed algorithm.

Keywords Probabilistic Framework · Bacterial Culture · Mixture Density Estimation · Expectation · Maximization · Segmentation Quality Metrics

1 Introduction

The present-day scenario of world was very fearful due to sudden raises of various viruses from various locations of the globe. The new viruses are being identified at various locations with various symptoms while some are easily being identified whereas some other viruses are still very difficult to identify and to understand better to identify the cure for such viruses. The sudden growth and development of such viruses are creating the medical emergency in various countries. The existence of medical emergencies is creating a panic in the public, the recent outbreak of Novel Corona Virus outbreak was one of the best examples for this case. With the outbreak and identification of such serious viruses, the impact of these viruses is making total life of the human being in vain. The development activities are being slowed down are stopped at several places and the serious harm to be happened to supply chain mechanism around the world. Now, due to the globalization of the entire world the spreading of viruses had become faster as the people move in a greater number from one country to the other country. People without their knowledge, they were becoming the carriers for such viruses to spread from one country to the other countries very faster.

In order to identify the viruses in a good and golden time period, it is always better to identify the bacteria through which the viruses are being generating and being spreading in and around the countries and locations. The best possible solution available for us was the identification of the culture of the bacteria that was creating the virus at a proper interval of time to reduce the spread of such viruses. Very few methods were available in the literature for identifying such culture of the bacteria with respect to the decomposition of digital images which were being collected from the digital microscopes. With the help of image processing and its related techniques, these sorts of problems can resolve to some extent. With the combination of mathematical modeling to the image decomposition techniques available can be developed and used for the better and fast identification of the culture of the bacteria such that these critical issues and problems can be resolved to some extent in a faster manner. This model of mixing the mathematical models with the image processing and its related techniques will yield a new results and new set of understandings which may give the new findings and new ideas of identification bacteria culture on digital images. These methods may give the new ideas and new strategies for the future research in this area (Guang & Chen, 2012; Li et al., 2013; Nahar & Ali, 2014). In general, the edge detection and the image decomposition methods are mostly used for the identification of similarity in local neighborhood and the observation of intensity values of the images.

One of the important mostly used techniques in the process of image processing was the decomposition of the images. It may also be considered that this technique is a mandatory one in the image processing techniques. It is widely used with the other techniques in the processing are like the computational mechanism of the images and the computer vision techniques. With the usage of these models, the decomposition of images can be used in

several areas of the applications like smoothening of the images, enhancing the size of the images, abstracting the data of the images and also the compression of images. The major work that was completed in decomposition of images was separating the structure from an image. This separation of structure forms the images could be implemented by removing the fine scale pixel details and saving the components of the edges and its structures. All these processes could be implemented without the knowledge or the prior information. The major work carried out in the decomposition of the images was to identify similarity in the local neighborhood and extend this process to the entire rest of the image. This process can be completed with the help of various methods like growing of the regions, probability distribution models and thresholding methods. In order to understand the performance of the decomposition method used in the works, several performance measure metrics have been designed and used. In the current work also, we had used three performance metrics to understand the performance of the currently considered model. Those performance metrics are the VOI, GCE and PRI.

Several authors had considered several problems related to the identification of bacteria on the digital images by decomposing the images. Some of those works were discussed in the current section are, W.X, Guang had developed and proposed an improved segmentation algorithm (Guang & Chen, 2012) such that to decompose the data. The process of decomposition had done with the help of two-dimensional Otsu method. The results obtained were discussed in detail and this article can be used for the segmentation of data for better understanding of the segmentation process of data. L. Zhengzhou, et al. had modified and proposed a new algorithm (Li et al., 2013) based on the mean shift method to solve the problem of inaccurate segmentation of the grey scale images. The new proposed algorithm works on the mechanism of combining both information of the position and the information of the grey scale. At the next phase the mean shift algorithm had used for the better segmentation of the given images.

M. Nahar et al. had proposed a new method (Nahar & Ali, 2014) for the identification of edges with respect to the image processing techniques. They had used the canny edge detector model to identify the edges of the images. This detection of edges will be very useful to the researchers which will reduce the size and shape of the images through which the further processing of the images will be easier for the researchers. Jyothirmayi et al. had discussed about the segmentation of the images by using a new method called the Doubly Truncated Generalized Laplace Mixture Model and k-means clustering (DTGLMM-k) (Jyothirmayi et al., 2016) by which the segmentation of the images had processed. The performance of the currently considered model had verified with the help of various performance metrics. The results were discussed in detail and the performance of the model had considered better when compared with the previous models. The current model developed in the article was useful in solving the problems like medical diagnostics, security problems, surveillance problems etc. The results generated by using the current model gives the better confirmation and better results in segmentation of the images.

Akhavan an Faez (Akhavan & Faez, 2014) had discussed about the importance of analyzing the blood vessel details in the images related to the retina. The identification and analysis of blood vessel details from time to time was very important. The diseases like macular degeneration, hypertensive retinopathy, diabetic retinopathy and glaucoma can be identified earlier stages with the utilization of these methods by dealing with vessels of the retina. The authors had used the method called the iterative region growing method through which the segmentation had processed and the results had obtained. The blood vessels considered here are enhanced by using various operations and the noise was removed by using the Adaptive

Fuzzy Switching Median filter. The current considered model had applied on two datasets and results were discussed in detail.

Srinivas and Srinivas Rao. had proposed a new image segmentation method (Srinivas & Srinivas Rao, 2007) based on finite truncated Gaussian mixture model. The advantage of their method lies in its efficiency and the entire process had completed in an unsupervised manner. The results show that the considered method can provide the better results with the other existing models. Lei and Udupa presented a new framework (Lie & Udupa, 1993) to assess the performance of the image segmentation models. In order to identify the performance, they had used the techniques like under detection model and the over detection model to detect the models. By using these methods, they had defined the number of regions on the images, quality of the images and model parameters to be considered on those selected images. In order to understand the performance of the classification for the selected images, they had used the Bayesian classifier. They derived a new formula to estimate the parameters and also to evaluate the errors being generated during the segmentation process.

Zhang et al. (Zhang et al., 2003) had tried to solve the problem of local convergence with respect to the Expectation and Maximization (EM) algorithm. They had introduced a new technique known as the split and merge process in the EM algorithm for Gaussian Mixtures. They had also used the two methods for solving the problems like Cholesky decomposition method and singular value decomposition model. The results show that the proposed methods are working fine and yielding good results in solving the segmentation of the color images used in unsupervised models of the machine learning.

Unnikrishnan et al. (Unnikrishnan et al., 2007) have discussed and developed an unsupervised based image segmentation model and results were discussed in detail. In their article, they demonstrated the similarity measure between the images with the new technique called Normalized Probabilistic Rand index (NPR). This technique can be used to compare the image segmentation techniques and also their performance for segmenting the images. They had presented the segmentation results carried out with various algorithms on different images. Some of the algorithms used for segmenting the images are like the hybrid algorithm which was a combination of expectation maximization (Bilmes et al., 1997; Lie & Sewehand, 1992; McLachlan & Krishnan, 1997) and other graph-based algorithms. The results had discussed in detail in the results section for all types of algorithms for various images. Results are presented on the 300 images in the publicly available Berkeley segmentation data set.

Wang et al. (Wang et al., 2020) have discussed about the early identification of the bacteria on various items like the food, water and fluids. They presented a computationally live bacteria detection model through which the bacteria can be identified with the help of neural network models. This model would be a cost-effective model and results were discussed in detail. Dias et al. (Dias et al., 2016) have discussed about the monitoring of the light microscopy which was commonly combined with staining techniques. The main drawback here for these methods was that these methods were susceptible to human errors which may yield to very bad decisions. Hence, the authors had proposed a new modified algorithm for the automated recognition and measurement of filamentous images and objects with very little errors.

Wahid, Ahmed and Habib (Wahid et al., 2018) have discussed about the classification of bacteria which was very important and keen in identifying the bacteria through various images. But it is time taking process for identifying the images manually. With the technology advancement, convolutional neural networks are being using for the same identification process in an automated manner. They had used transfer of learning method to identify the bacteria on around 500 number of images with 5 different types of bacteria. They got around 95% efficiency and accuracy in terms of identification of the bacteria on the images. Hay and Parthasarathy (Hay & Parthasarathy, 2018) have discussed on neural network models

about the identification of bacteria on the 3D model of images. With the experimentation they had done, they found that the neural network is accurate as human beings in terms of identifying the bacteria on images. The utilization of support vector machine and random forest algorithms had given a good result when compared with the other neural network model for the identification of bacteria presence on images (Lee et al., 2019; Preetha & Pandi Selvi, 2018). They had also discussed and given various considerations to be considered by the other researchers when they were working on these neural network models to identify the species or the bacteria on 3D digital images (Cimpoi et al., 2016; Zieliński et al., 2017).

Mohamed and Afify (Mohamed & Afify, 2018) have discussed about the importance of recognizing the bacteria cells on images on time for effective treatment of various diseases caused by the bacteria. Several algorithms had been developed for the same purpose to identify the bacteria on images with the help of large-scale microscopic image bacteria (Sharma et al., 2019). They had used the feature selection and extraction methods with the help of the support vector machine algorithms to identify the content on images. By using the same method, they had achieved around 97% of the accuracy and efficiency for the same which shows with the help of results.

From the above various works developed and implemented by various authors for the segmentation of the images, it is observed that very few works had worked on the identification of the bacteria on images with the utilization of Asymmetric distribution with k-means clustering algorithm. Hence, in the current article an attempt has been made to use such technique and to analyze the performance of the models with respect to various performance measure metrics and the results had analyzed in the results section and a comparative study had been presented for the better understanding of the currently developed model.

2 Methodology used for the current proposed model

The proposed model with the methodology for obtaining estimates of the parameters involved in the model through Expectation and Maximization algorithm (McLachlan & Krishnan, 1997). The pixel intensities of the image region are considered as features of the image. Here the three-parameter logistic type distribution is assumed for modeling the pixel intensities of the image regions. As a result of it, the whole image can be characterized with a three-parameter logistic type mixture model. The probability density function of the pixel intensity is of the form as,

$$f(x, \mu, \sigma^2) = \frac{\left[\frac{3}{(3k+\pi^2)}\right] \left[k + \left(\frac{x-\mu}{\sigma}\right)^2\right] e^{-\left(\frac{x-\mu}{\sigma}\right)}}{\sigma \left[1 + e^{-\left(\frac{x-\mu}{\sigma}\right)}\right]^2} \tag{1}$$

where $-\infty < x < \infty, -\infty < \mu < \infty, k \geq 4$.

Here, x denotes the pixel values. μ represents the mean of pixel intensities. k is the number of segments where $k \geq 4$ is chosen since it is a univariate distribution model used for gray scale model images.

For various values of the parameters generate the corresponding shapes of probability curves associated with the three-parameter logistic type distribution are shown in Fig. 1. Each value of the shape parameter k (4,5,6,7, ...) gives a bell-shaped distribution.

This distribution is symmetric about μ and the distribution function is

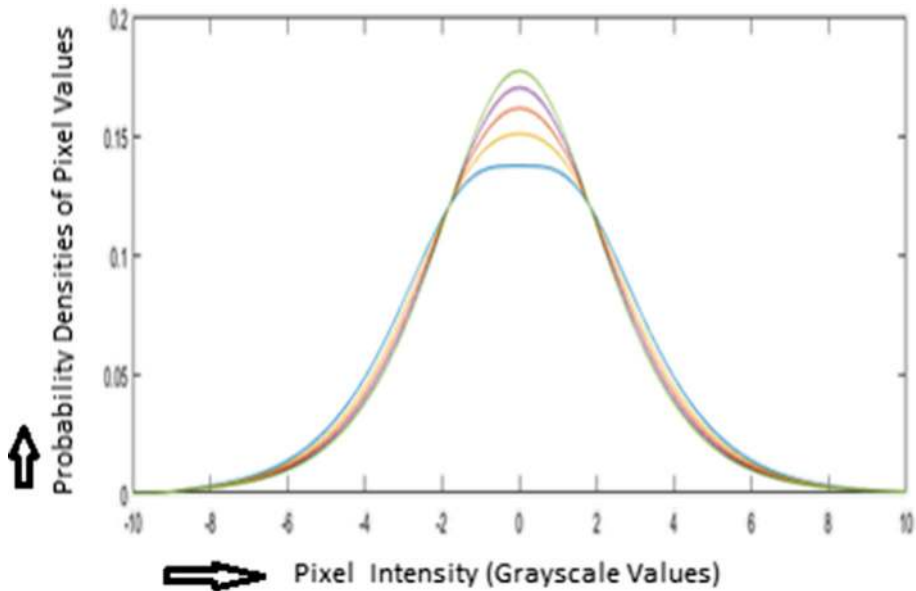


Fig. 1 Frequency curve of two-parameter logistic type distribution

$$F(X) = \frac{\left[\frac{3}{3k+\pi^2} \right] \left[2\left(\frac{x-\mu}{\sigma}\right) \left[1 + e^{-\left(\frac{x-\mu}{\sigma}\right)} \right] + \left[k + \left(\frac{x-\mu}{\sigma}\right)^2 \right] \left[e^{-\left(\frac{x-\mu}{\sigma}\right)} - 1 \right] \right]}{\sigma^2 e^{\left(\frac{x-\mu}{\sigma}\right)} \left[1 + e^{-\left(\frac{x-\mu}{\sigma}\right)} \right]^3} \tag{2}$$

The three-parameter logistic type distribution characterizes the entire image where it is the collection of regions. Hence, in the current work the pixel intensities of full image track the three-parameter logistic type distribution which has k -component mixtures. The probability density function of the current model considered as,

$$p(x) = \sum_{i=1}^k \alpha_i f_i(x, \mu, \sigma^2) \tag{3}$$

Even though the neighboring pixel intensities are correlated. The correlation can be made insignificant by considering spatial sampling proposed by Sewehand and Lei (Lie & Sewehand, 1992). After reducing the correlations, the pixel intensities are to be considered as independent.

For the entire image, the pixel intensity had been calculated and the mean for the same intensity will be calculated by using the following model as $E(X) = \sum_{i=1}^K \alpha_i \mu_i$. The architecture of entire model and the entire process had depicted in Fig. 2.

The proposed model is shown is an end to end approach for segmentation of the bacteria. It has taken inspiration from the GLMM-H, GLMM-K and DTGLMM-k for the semantic bacterial segmentation with the greater efficiency. The pixel intensities of the image region are considered image characteristics. A three-parameter logistic type distribution is assumed

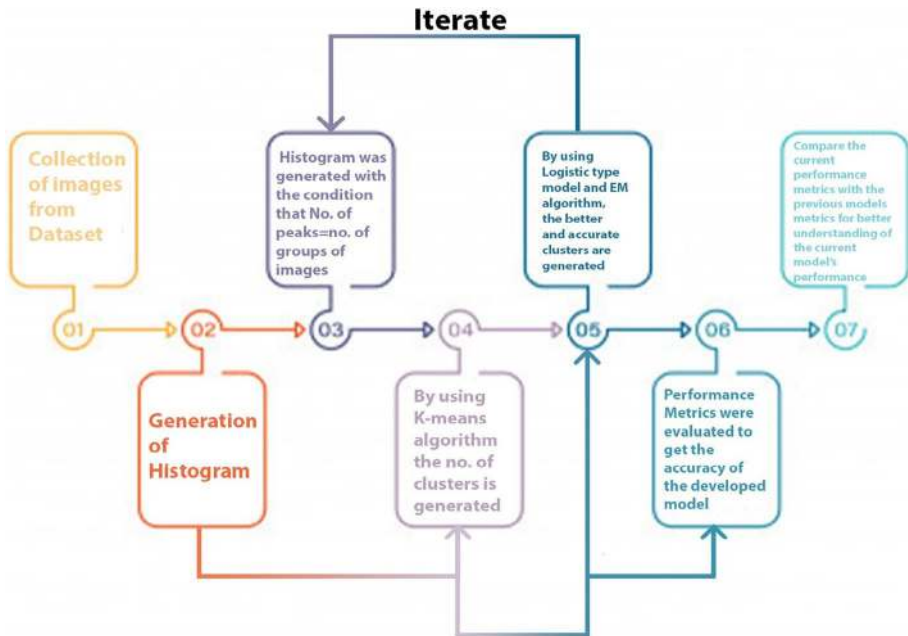


Fig. 2 Illustration of the proposed working model on COG dataset

for modelling the pixel intensities of the image regions. As a result, the entire image can be characterized using a three-parameter logistic type mixture model.

2.1 Parameter estimation by using EM algorithm

The Expectation and Maximization (EM) algorithm can be utilized for obtaining the estimates of the parameters involved in the model (Jeckel & Drescher, 2021; Liang et al., 2017; Xin et al., 2017; Zhang et al., 2021). The major consideration for EM algorithm is expectation of the likelihood function and then maximization of it with respect to its parameters as follows:

The likelihood of the function of model parameters is,

$$L(\theta) = \prod_{s=1}^N p(x_s, \theta^{(l)}) \tag{4}$$

$$L(\theta) = \prod_{s=1}^N \left(\sum_{i=1}^k \alpha_i f_i(x_s, \theta^{(l)}) \right). \tag{5}$$

This implies

$$\log L(\theta) = \sum_{s=1}^N \log \left(\sum_{i=1}^k \alpha_i f_i(x_s, \theta^{(l)}) \right) \tag{6}$$

$$\theta = (\mu_i, \sigma_i^2, \alpha_i) \text{ where } i = 1, 2, 3, \dots, n \tag{7}$$

$$\log L(\theta) = \sum_{S=1}^N \log \left[\sum_{i=1}^m \alpha_i \frac{\left[\frac{3}{(3k+\pi^2)} \right] \left[k + \left(\frac{x_s - \mu_i}{\sigma_i} \right)^2 \right] e^{-\left(\frac{x_s - \mu_i}{\sigma_i} \right)}}{\sigma_i \left[1 + e^{-\left(\frac{x_s - \mu_i}{\sigma_i} \right)} \right]^2} \right] \tag{8}$$

2.2 Expectation step

In the current step, the expectation calculated as,

$$Q(\theta, \theta^{(0)}) = E_{\theta^{(0)}} [\log L(\theta) / \bar{x}] \tag{9}$$

Given the initial parameters $\theta^{(0)}$. One can compute the density of pixel intensity X as

$$P(x_s, \theta^{(l)}) = \sum_{i=1}^k \alpha_i f_i(x_s, \theta^{(l)}) \tag{10}$$

$$L(\theta) = \prod_{S=1}^N p(x_s, \theta^{(l)}) \tag{11}$$

$$\log L(\theta) = \sum_{S=1}^N \log \left(\sum_{i=1}^k \alpha_i^{(l)} f_i(x_s, \theta^{(l)}) \right) \tag{12}$$

The conditional probability of any observations x_s , belongs to any region K is

$$P_k(x_s, \theta^{(l)}) = \left[\frac{\alpha_k^{(l)} f_k(x_s, \theta^{(l)})}{P_i(x_s, \theta^{(l)})} \right] \tag{13}$$

$$p_k(x_s, \theta^{(l)}) = \left[\frac{\alpha_k^{(l)} f_k(x_s, \theta^{(l)})}{\sum_{i=1}^k \alpha_i^{(l)} f_i(x_s, \theta^{(l)})} \right] \tag{14}$$

The Expectation of the log likelihood function of the sample is $Q(\theta, \theta^{(l)}) = E_{\theta^{(l)}} [\log L(\theta) / \bar{x}]$.

But we have

$$f_i(x_s, \theta^{(l)}) = \frac{\left[\frac{3}{(3k+\pi^2)} \right] \left[k + \left(\frac{x_s - \mu_i^{(l)}}{\sigma_i^{(l)}} \right)^2 \right] e^{-\left(\frac{x_s - \mu_i^{(l)}}{\sigma_i^{(l)}} \right)}}{\sigma_i^{(l)} \left[1 + e^{-\left(\frac{x_s - \mu_i^{(l)}}{\sigma_i^{(l)}} \right)} \right]^2} \tag{15}$$

Following the heuristic arguments of Bilmes et al. (Bilmes et al., 1997) we have

$$Q(\theta, \theta^{(l)}) = \sum_{i=1}^k \sum_{s=1}^N (P_i(x_s, \theta^{(l)})) (\log f_i(x_s, \theta^{(l)}) + \log \alpha_i^{(l)}) \tag{16}$$

2.3 Maximization step

For obtaining the estimation of model parameters one has to maximize $Q(\theta, \theta^{(l)})$ such that $\sum \alpha_i = 1$. This can be solved by applying the standard solution method for constrained maximum by constructing the first order Lagrange type function as,

$$F = \left[E(\log L(\theta^{(l)})) + \lambda \left(1 - \sum_{i=1}^k \alpha_i^{(l)} \right) \right], \tag{17}$$

where λ is Lagrangian multiplier combining the constraint with the log likelihood functions to be maximized.

Analytical Expressions for α_i [Pixels Intensity]:

To find the expression for α_i , we solve that following equation. Hence, $\frac{\partial F}{\partial \alpha_i} = 0$. This implies

$$\frac{\partial}{\partial \alpha_i} \left[\sum_{i=1}^N \sum_{s=1}^K P_i(x_s, \theta^{(l)}) \log \left[\frac{\left[\frac{3}{(12+\pi^2)} \right] \left[4 + \left(\frac{x_s - \mu_i^{(l)}}{\sigma_i^{(l)}} \right)^2 \right] e^{-\left(\frac{x_s - \mu_i^{(l)}}{\sigma_i^{(l)}} \right)^2}}{\sigma_i^{(l)} \left[1 + e^{-\left(\frac{x_s - \mu_i^{(l)}}{\sigma_i^{(l)}} \right)^2} \right]} \right] + \log \alpha_i + \lambda \left(1 - \sum_{i=1}^k \alpha_i \right) \right] = 0 \tag{18}$$

This implies $\sum_{i=1}^N \frac{1}{\alpha_i} P_i(x_s, \theta^{(l)}) + \lambda = 0$. Summing both sides over all observations, we get $\lambda = -N$. Therefore, $\alpha_i = \frac{1}{N} \sum_{s=1}^N P_i(x_s, \theta^{(l)})$. The updated equations of α_i for $(l + 1)^{th}$ iteration is $\alpha_i^{(l+1)} = \frac{1}{N} \sum_{s=1}^N P_i(x_s, \theta^{(l)})$. In other words,

$$\alpha_i^{(l+1)} = \frac{1}{N} \sum_{s=1}^N \left[\frac{\alpha_i^{(l)} f_i(x_s, \theta^{(l)})}{\sum_{i=1}^k \alpha_i^{(l)} f_i(x_s, \theta^{(l)})} \right]. \tag{19}$$

2.4 Derivation of Expressions for μ_i [Mean of Pixel Values]

For updating the parameter $\mu_i, i = 1, 2, 3, \dots, k$ we consider the derivatives of $Q(\theta, \theta^{(l)})$ with respect to μ_i and equal to zero, We have $Q(\theta, \theta^{(l)}) = E[\log L(\theta, \theta^{(l)})]$. Therefore, $\frac{\partial}{\partial \mu_i} (Q(\theta, \theta^{(l)})) = 0$. It then implies that $E\left[\frac{\partial}{\partial \mu_i} (\log L(\theta, \theta^{(l)}))\right] = 0$. Taking the partial derivative with respect to μ_i , we have

$$\frac{\partial}{\partial \mu_i} \left[\sum_{s=1}^N \sum_{i=1}^K P_i(x_s, \theta^l) \log \left[\alpha_i \frac{\left[\frac{3}{3k+\pi^2} \right] \left[k + \left(\frac{x_s - \mu_i}{\sigma_i} \right)^2 \right] e^{-\left(\frac{x_s - \mu_i}{\sigma_i} \right)^2}}{\sigma_i \left[1 + e^{-\left(\frac{x_s - \mu_i}{\sigma_i} \right)^2} \right]} \right] \right] = 0 \tag{20}$$

$$\frac{\partial}{\partial \mu_i} \left[\sum_{s=1}^N \sum_{i=1}^K P_i(x_{s.}, \theta^l) \left[\log \alpha_i + \log \left[\frac{\left[\frac{3}{3k+\pi^2} \right] \left[k + \left(\frac{x_s - \mu_i}{\sigma_i} \right)^2 \right] e^{-\left(\frac{x_s - \mu_i}{\sigma_i} \right)}}{\sigma_i \left[1 + e^{-\left(\frac{x_s - \mu_i}{\sigma_i} \right)} \right]^2} \right] \right] \right] = 0 \tag{21}$$

$$\frac{\partial}{\partial \mu_i} \left[\sum_{s=1}^N \sum_{i=1}^K P_i(x_{s.}, \theta^l) \left[\log \alpha_i + \log \left[\left[\frac{3}{3k+\pi^2} \right] \left[k + \left(\frac{x_s - \mu_i}{\sigma_i} \right)^2 \right] e^{-\left(\frac{x_s - \mu_i}{\sigma_i} \right)} \right] - \log \left[\sigma_i \left[1 + e^{-\left(\frac{x_s - \mu_i}{\sigma_i} \right)} \right]^2 \right] \right] \right] = 0 \tag{22}$$

$$\frac{\partial}{\partial \mu_i} \left[\sum_{s=1}^N \sum_{i=1}^K P_i(x_{s.}, \theta^l) \left[\log \alpha_i + \log \left[\frac{3}{3k+\pi^2} \right] + \log \left[k + \left(\frac{x_s - \mu_i}{\sigma_i} \right)^2 \right] + \log e^{-\left(\frac{x_s - \mu_i}{\sigma_i} \right)} - \log \sigma_i - \log \left[1 + e^{-\left(\frac{x_s - \mu_i}{\sigma_i} \right)} \right]^2 \right] \right] = 0 \tag{23}$$

$$\sum_{s=1}^N P_i(x_{s.}, \theta^l) \left[\frac{\left[2 \left(\frac{x_s - \mu_i}{\sigma_i} \right) \left(\frac{-1}{\sigma_i} \right) \right]}{\left[k + \left(\frac{x_s - \mu_i}{\sigma_i} \right)^2 \right]} + \left[\frac{1}{\sigma_i} \right] - \frac{2e^{-\left(\frac{x_s - \mu_i}{\sigma_i} \right)} \left[-\left(\frac{-1}{\sigma_i} \right) \right]}{\left[1 + e^{-\left(\frac{x_s - \mu_i}{\sigma_i} \right)} \right]} \right] = 0 \tag{24}$$

$$\sum_{s=1}^N P_i(x_{s.}, \theta^l) \left[\frac{\left[-2 \left(\frac{x_s - \mu_i}{\sigma_i} \right) \left(\frac{1}{\sigma_i} \right) \right]}{\left[k + \left(\frac{x_s - \mu_i}{\sigma_i} \right)^2 \right]} + \left[\frac{1}{\sigma_i} \right] - \frac{2e^{-\left(\frac{x_s - \mu_i}{\sigma_i} \right)}}{\sigma_i \left[1 + e^{-\left(\frac{x_s - \mu_i}{\sigma_i} \right)} \right]} \right] = 0 \tag{25}$$

$$\sum_{s=1}^N P_i(x_{s.}, \theta^l) \left[\frac{\left[-2(x_s - \mu_i) \right]}{\sigma_i^2 \left[k + \left(\frac{x_s - \mu_i}{\sigma_i} \right)^2 \right]} + \left[\frac{1}{\sigma_i} \right] - \frac{2}{\sigma_i \left[1 + e^{-\left(\frac{x_s - \mu_i}{\sigma_i} \right)} \right]} \right] = 0 \tag{26}$$

$$2 \sum_{s=1}^n \frac{P_i(x_{s.}, \theta^{(l)}) \mu_i}{\sigma_i^2 \left[k + \left(\frac{x_s - \mu_i}{\sigma_i} \right)^2 \right]} = \sum_{s=1}^n \frac{P_i(x_{s.}, \theta^{(l)}) (2x_s)}{\sigma_i^2 \left[k + \left(\frac{x_s - \mu_i}{\sigma_i} \right)^2 \right]} - \sum_{s=1}^n \frac{P_i(x_{s.}, \theta^{(l)})}{\sigma_i} + \sum_{s=1}^n \frac{2P_i(x_{s.}, \theta^{(l)})}{\sigma_i \left[1 + e^{-\left(\frac{x_s - \mu_i}{\sigma_i} \right)} \right]} \tag{27}$$

$$\mu_i = \frac{\sum_{s=1}^n \frac{P_i(x_{s.}, \theta^{(l)}) (2x_s)}{\sigma_i^2 \left[k + \left(\frac{x_s - \mu_i}{\sigma_i} \right)^2 \right]} - \sum_{s=1}^n \frac{P_i(x_{s.}, \theta^{(l)})}{\sigma_i} + \sum_{s=1}^n \frac{2P_i(x_{s.}, \theta^{(l)})}{\sigma_i \left[1 + e^{-\left(\frac{x_s - \mu_i}{\sigma_i} \right)} \right]}}{2 \sum_{s=1}^n \frac{P_i(x_{s.}, \theta^{(l)})}{\sigma_i^2 \left[k + \left(\frac{x_s - \mu_i}{\sigma_i} \right)^2 \right]}} \tag{28}$$

For updating σ_i^2 [Variance of Segmented Image’s Pixels] we differentiate $Q(\theta, \theta^{(l)})$ with respect to σ_i^2 and equate it to zero. That is $\frac{\partial}{\partial \sigma^2} (Q(\theta, \theta^{(l)})) = 0$. This implies, $E \left[\frac{\partial}{\partial \sigma^2} (\log L(\theta, \theta^{(l)})) \right] = 0$.

Taking the partial derivative with respect to σ_i^2 .

$$\frac{\partial}{\partial \sigma_i^2} \left[\sum_{s=1}^N \sum_{i=1}^K P_i(x_{s.}, \theta^l) \log \alpha_i \frac{\left[\frac{3}{3k+\pi^2} \right] \left[k + \left(\frac{x_s - \mu_i}{\sigma_i} \right)^2 \right] e^{-\left(\frac{x_s - \mu_i}{\sigma_i} \right)}}{\sigma_i \left[1 + e^{-\left(\frac{x_s - \mu_i}{\sigma_i} \right)} \right]^2} \right] = 0 \tag{29}$$

$$\frac{\partial}{\partial \sigma_i^2} \left[\sum_{s=1}^N \sum_{i=1}^K P_i(x_{s..}, \theta^l) \left[\log \alpha_i + \log \left[\frac{3}{3k + \pi^2} \right] + \log \left[k + \left(\frac{x_s - \mu_i}{\sigma_i} \right)^2 \right] + \log e^{-\left(\frac{x_s - \mu_i}{\sigma_i} \right)} - \log \left[\sigma_i \left[1 + e^{-\left(\frac{x_s - \mu_i}{\sigma_i} \right)} \right]^2 \right] \right] \right] = 0 \tag{30}$$

$$\frac{\partial}{\partial \sigma_i^2} \left[\sum_{s=1}^N \sum_{i=1}^K P_i(x_{s..}, \theta^l) \left[\log \alpha_i + \log \left[\frac{3}{3k + \pi^2} \right] + \log \left[k + \left(\frac{x_s - \mu_i}{\sigma_i} \right)^2 \right] + \log e^{-\left(\frac{x_s - \mu_i}{\sigma_i} \right)} - \log \sigma_i - \log \left[1 + e^{-\left(\frac{x_s - \mu_i}{\sigma_i} \right)} \right]^2 \right] \right] = 0 \tag{31}$$

$$\frac{\partial}{\partial \sigma_i^2} \left[\sum_{s=1}^N \sum_{i=1}^K P_i(x_{s..}, \theta^l) \left[\log \alpha_i + \log \left[\frac{3}{3k + \pi^2} \right] + \log \left[k + \left(\frac{x_s - \mu_i}{\sigma_i} \right)^2 \right] + \log e^{-\left(\frac{x_s - \mu_i}{\sigma_i} \right)} - \log \sigma_i - 2 \log \left[1 + e^{-\left(\frac{x_s - \mu_i}{\sigma_i} \right)} \right] \right] \right] = 0 \tag{32}$$

$$\frac{\partial}{\partial \sigma_i^2} \left[\sum_{s=1}^N \sum_{i=1}^K P_i(x_{s..}, \theta^l) \left[\log \alpha_i + \log \left[\frac{3}{3k + \pi^2} \right] + \log \left[k + \left(\frac{x_s - \mu_i}{\sigma_i} \right)^2 \right] - \left(\frac{x_s - \mu_i}{\sigma_i} \right) \log e - \log \sigma_i - 2 \log \left[1 + e^{-\left(\frac{x_s - \mu_i}{\sigma_i} \right)} \right] \right] \right] = 0 \tag{33}$$

$$\frac{\partial}{\partial \sigma_i^2} \left[\sum_{s=1}^N \sum_{i=1}^K P_i(x_{s..}, \theta^l) \left[\log \alpha_i + \log \left[\frac{3}{3k + \pi^2} \right] + \log \left[k + \frac{(x_s - \mu_i)^2}{\sigma_i^2} \right] - \frac{(x_s - \mu_i)}{(\sigma_i^2)^{\frac{1}{2}}} - \log(\sigma_i^2)^{\frac{1}{2}} - 2 \log \left[1 + e^{-\left(\frac{x_s - \mu_i}{(\sigma_i^2)^{\frac{1}{2}}} \right)} \right] \right] \right] = 0 \tag{34}$$

$$\sum_{s=1}^N P_i(x_{s..}, \theta^{(l)}) \left[\frac{-\frac{(x_s - \mu_i)^2}{\sigma_i^4}}{\left[k + \frac{(x_s - \mu_i)^2}{\sigma_i^2} \right]} + \frac{(x_s - \mu_i)}{2\sigma_i^3} - \frac{1}{2\sigma_i^2} - \frac{-\frac{(x_s - \mu_i)}{(\sigma_i^2)^{\frac{1}{2}}}}{\left[1 + e^{-\frac{(x_s - \mu_i)}{(\sigma_i^2)^{\frac{1}{2}}} \right]} \right] \left[-\frac{(x_s - \mu_i)}{2\sigma_i^3} \right] \right] = 0 \tag{35}$$

$$\sum_{s=1}^N P_i(x_{s..}, \theta^{(l)}) \left[\frac{-\frac{(x_s - \mu_i)^2 \sigma_i^2}{\sigma_i^4 [k\sigma_i^2 + (x_s - \mu_i)^2]} + \frac{(x_s - \mu_i)}{2\sigma_i^3} - \frac{1}{\sigma_i^2} - \frac{(x_s - \mu_i)}{\sigma_i^3 \left[1 + e^{-\frac{(x_s - \mu_i)}{\sigma_i}} \right]} \right] = 0 \tag{36}$$

$$\begin{aligned} \sum_{s=1}^N \frac{P_i(x_{s..}, \theta^{(l)}) (x_s - \mu_i)^2 \sigma_i^2}{\sigma_i^4 [k\sigma_i^2 + (x_s - \mu_i)^2]} &= - \sum_{s=1}^N \frac{P_i(x_{s..}, \theta^{(l)}) (x_s - \mu_i)}{\sigma_i^3 \left[1 + e^{-\frac{(x_s - \mu_i)}{\sigma_i}} \right]} + \sum_{s=1}^N \frac{P_i(x_{s..}, \theta^{(l)}) (x_s - \mu_i)}{2\sigma_i^3} - \sum_{s=1}^N \frac{P_i(x_{s..}, \theta^{(l)})}{2\sigma_i^2} \\ \sigma_i^{2(l+1)} &= \frac{\sum_{s=1}^N \frac{P_i(x_s, \theta^{(l)}) (x_s - \mu_i^{(l+1)})}{2\sigma_i^{3(l)}} - \sum_{s=1}^N \frac{P_i(x_s, \theta^{(l)}) (x_s - \mu_i^{(l+1)})}{\sigma_i^{3(l)} \left[1 + e^{-\frac{(x_s - \mu_i)}{\sigma_i}} \right]} - \sum_{s=1}^N \frac{P_i(x_s, \theta^{(l)})}{2\sigma_i^{2(l)}}}{\sum_{s=1}^N \frac{P_i(x_s, \theta^{(l)}) (x_s - \mu_i^{(l+1)})^2}{\sigma_i^{4(l)} \left[\kappa \sigma_i^{2(l)} + (x_s - \mu_i^{(l+1)})^2 \right]}} \end{aligned} \tag{37}$$

where,

$$P_i(x_s, \theta^{(l)}) = \left[\frac{\alpha_i^{(l+1)} f_i(x_s, \mu_i^{(l+1)}, \sigma_i^{2(l)})}{\sum_{i=1}^k \alpha_i^{(l+1)} f_i(x_s, \mu_i^{(l+1)}, \sigma_i^{(l)})} \right] \tag{38}$$

3 Initialization of parameters by k-means algorithm

One of the best algorithms for analyzing and estimating the parameters is the expectation and maximization algorithm. The number of regions present in an image is going to play a key role in this model. One should get the idea that the EM algorithm performance and its efficiency mainly depends on these factors (Khamparia et al., 2021; Zhong et al., 2021). By implementing the same method, one can get the mixture components number. Therefor by drawing the histogram with the help of the pixel intensities considered for the whole image. The number of tops in the histogram can be considered as the initial value of the image pixels.

In some cases, it may not be possible always to identify the number of tops and also to identify the initial values of the image pixels. Hence, the k-means clustering algorithm is mainly used for avoiding such problems. Hence, by using this model, the major image is divided into several pixels and the images can be identified easily. The entire major image had been classified to various regions in the form of homogeneous form. This k-means clustering algorithm is one of the best algorithms in the list of clustering algorithms. This algorithm is mainly used for identifying the parts or bits of which had reduced or decreased the size of the data and also reduced the between several points which were selected form the large set of points in the image data.

The k-means clustering algorithm used in the form of an iterative algorithm. The procedure followed was the iterative model through which the distance between the objects could be reduced or minimized from its point of centers. This process of selection can be considered from the overall set of pixel's data stored as a large set of pixels. This processing steps are discussed in detail as follows,

Step 1 At first, the k data points were selected from the dataset as random model from the whole dataset as clusters selected as initial stages. The selected data points could be considered as the centroids of the clusters at initial stages of the data.

Step 2 In the next step, the Euclidean distance was calculated for each data point from the data cluster center and the data point was reassigned to another cluster of data point which were near to the old cluster center.

Step 3 The center to the new cluster to be calculated such to calculate the squared error distance. This calculated squared error distance should be always minimum when compared with the other data points and the center of the data appoints selected from the pool of large data points.

Step 4 The steps two and three were repeated until the centers of the cluster might not change with changes in the data points selections.

Step 5 The process will be stopped.

From the above algorithm, it is understood that the centers of the clusters are being updated at only at a time of when all data points are being allocated or reached to their closest closed cluster centers. The major thing that happens in the k-means clustering algorithm was the value of the parameter k , which denotes the number of clusters in an image. Here the k value was 4 which was taken from the histogram where the peaks in the histogram are 4.

After determining the final values of k (number of regions), we obtain the initial estimates of μ_i , σ_i^2 and α_i for the i^{th} region using the segmented region pixel intensities with the method given by Srinivasa and Srinivas Rao (Srinivas & Srinivas Rao, 2007) for three parameter logistic distribution.

The initial estimate as $\alpha_i = \frac{1}{k}$, where $i = 1, 2, 3, \dots, k$. The parameter μ_i [Mean of Pixels] and σ_i^2 [Variance of Segmented Image's Pixels] are estimated by the method of moments as

$\hat{\mu}_i = X$ and $\sigma_i^2 = \frac{4n_i}{3(n_i-1)} S^2$, where S^2 is sample variance, n_i is the number of observations in the i^{th} segmentation.

4 Segmentation algorithm

The image segmentation algorithm is proposed in this section. The model parameters are estimated with the help of properties which are discussed in Sects. 2 and 3. We had to segment the pixels to the respective image regions. The major steps in image segmentation are as follows:

Step 1 k-means algorithm is utilized for dividing pixel intensities of the whole image into K-image regions. where k is number of image regions or clusters.

Step 2 Compute the initial estimates for the parameters of the model using moment method of estimates for each image region as discussed in Sect. 3.

Step 3 The Expectation and Maximization algorithm with the updated equations of parameters given in Sect. 2 is to be utilized for computing the final parameters of model for k-image regions.

Step 4 The allocation of each pixel in the whole image into its corresponding j^{th} image region done by computing component maximum likelihood of each image region as follows.

That is x_s is assigned to the j^{th} region for which L_j is maximum, where

$$L_j = \text{MAX} \left[\frac{\left[\frac{3}{(3k+\pi^2)} \right] \left[k + \left(\frac{x_s - \mu_i}{\sigma_i} \right)^2 \right] e^{-\left(\frac{x_s - \mu_i}{\sigma_i} \right)^2}}{\sigma_i \left[1 + e^{-\left(\frac{x_s - \mu_i}{\sigma_i} \right)^2} \right]^2} \right], -\infty < \mu_i < \infty, -\infty < \sigma_i^2 < \infty, k \geq 4 \tag{39}$$

5 Data and experiments

This section deals with the data, experimentation, implementation details and evaluation metrics. For the experimentation and training of the proposed model, the approach makes use of a publicly available dataset from the COG Database for experimentation and training of the proposed model, which is described by Yin and Ding (Yin & Ding, 2009). This dataset was created by combining the public dataset with other data. COGs, or Clusters of Orthologous Genes, are a type of bacteria cluster. Originally developed in 1997, the database has been updated several times since then, with the most recent update occurring in 2019 (Tatusov et al., 1997). Table 1 lists the details of the database. As of this writing, the database contains the complete 1,187 bacteria and 122 archaea, which correspond to 1,234 different bacteria. Among the new features are 250 updated COG annotations, which are included in the price. There are a total of 888 CT scans in this dataset. A total of 1166 images were created with ground truth masks and divided into 922 and 244 training and test subsets, respectively, following the pre-processing.

With the help of suggested image segmentation algorithm, the experiment is carried with three randomly taken sample images namely bacteria 1, bacteria 2, bacteria 3 from the bacteria dataset (Tatusov et al., 1997; Yin & Ding, 2009). The features of the images are obtained

Table 1 Images details of dataset

| S. No | Characteristics | Train Set (n = 922) | Test Set (n = 244) |
|-------|-----------------|---------------------|--------------------|
| 1 | Diameter | 8.16 ± 4.69 | 9.89 ± 5.26 |
| 2 | Margin | 4.85 ± 0.89 | 4.65 ± 0.69 |
| 3 | Spiculation | 1.89 ± 0.89 | 1.78 ± 0.98 |
| 4 | Lobulation | 1.89 ± .56 | 1.56 ± 0.56 |
| 5 | Subtlety | 2.96 ± 0.99 | 3.56 ± 1.56 |

by considering pixel intensities. The pixel intensities are obtaining by using MATLAB. Assuming that pixel intensities of image regions follow mixture of Three parameter logistic type distribution, the model characterizing the whole image is developed with the help of k-means algorithm. The three parameters considered here are, pixel intensity (α_i), mean of pixels (μ_i), and variance of the behavior of segmented images of pixels (σ^2).

After generating the histogram, the k value was identified by histogram as 4 which means four peaks were observed at the histogram of each input images. As histogram was used to grouping of individual clusters of images, the images may not be perfectly segmented. Hence, we use the k-means clustering algorithm for better clustering of image. For generating the better results, we had used the proposed model of Three Parameter Logistic Mixture Model and k-means clustering (TPLMM-k) for better segmentation results.

The output values obtained for the several parameters like the intensity of the pixels, mean values of the pixels considered in the images and the variance of the behavior of segmented image pixels had considered in the current model are shown at Table 2. The sample image of bacteria 1 chosen for converting to gray scale and processing the image with the current proposed model can be observed at Fig. 3. The grayscale converted image after using the current process was shown at Fig. 4.

The histogram generated for the current sample input image 1 was observed at Fig. 5.

The histogram representation of the current model with respect to bacteria 1 can be shown at Fig. 5 and the plots of density of probability with respect to the TPLMM-k Means algorithm had shown at Fig. 6. The plots had calculated and presented at Fig. 6 are for Pixel Intensity (α_i), Mean of Pixels (μ_i), and Variance of the behavior of Segmented Images of Pixels (σ^2).

The output values obtained for the several plots considered in the current model are shown at Table 3. The sample image of bacteria 2 chosen for converting to gray scale and processing

Table 2 Estimates of the parameters for sample bacteria 1 image

| Parameters | Initial Parameters | | | Refined estimates | | |
|------------|--------------------|----------|----------|-------------------|---------|---------|
| | Image Region | | | Image Region | | |
| | 1 | 2 | 3 | 3 | 2 | 3 |
| α_i | 0.555 | 0.555 | 0.555 | 0.1205 | 0.6177 | 0.2618 |
| μ_i | 221.6327 | 105.5619 | 71.7264 | 54.25 | 48.14 | 160.23 |
| σ^2 | 936.5615 | 406.2072 | 5738.391 | 355.458 | 498.258 | 1958.21 |

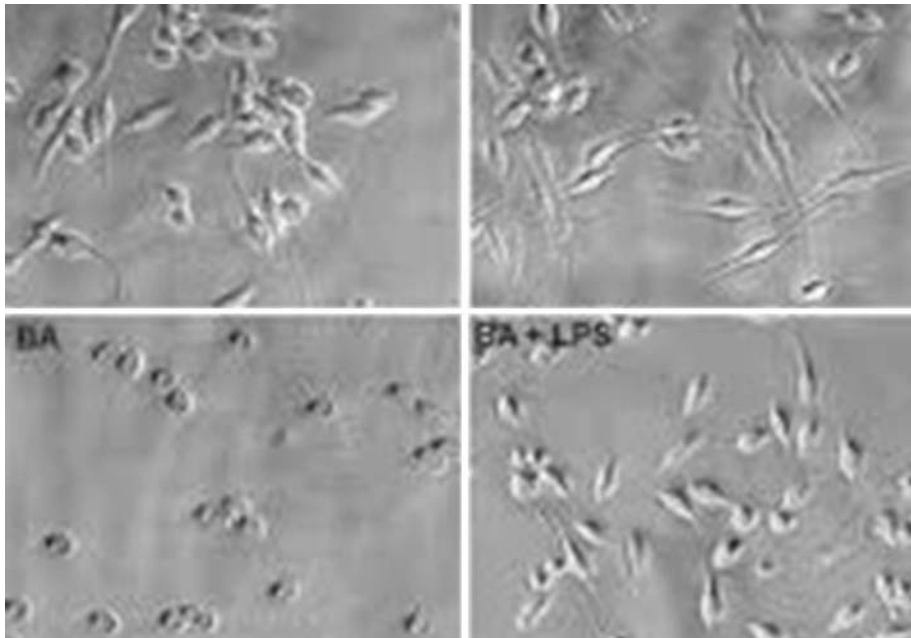


Fig. 3 Input image of sample bacteria 1 image

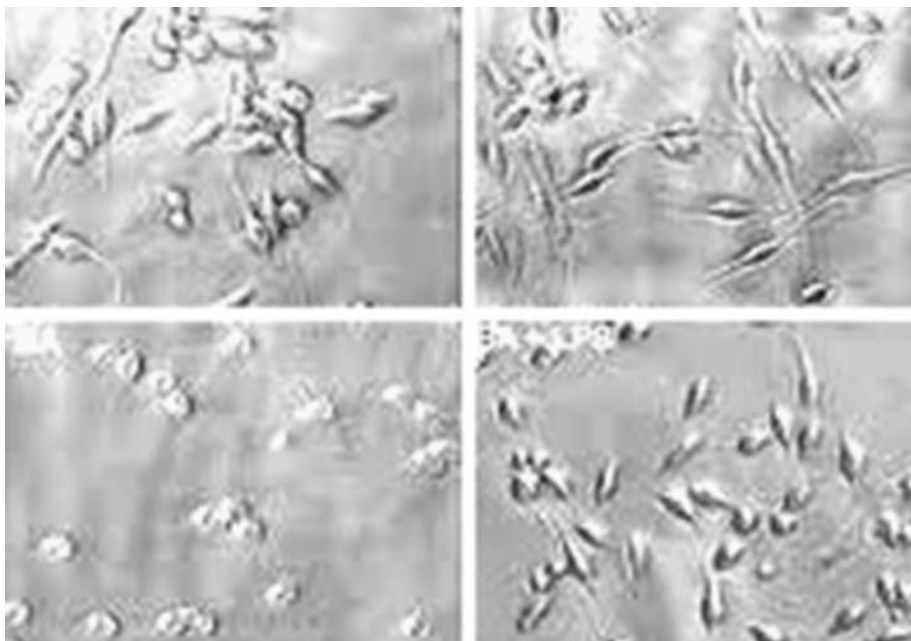


Fig. 4 Segmented image for sample bacteria 1 image

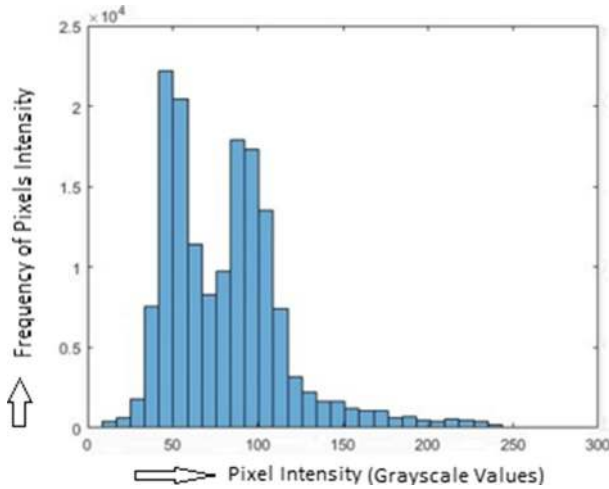


Fig. 5 Histogram for sample bacteria 1 image

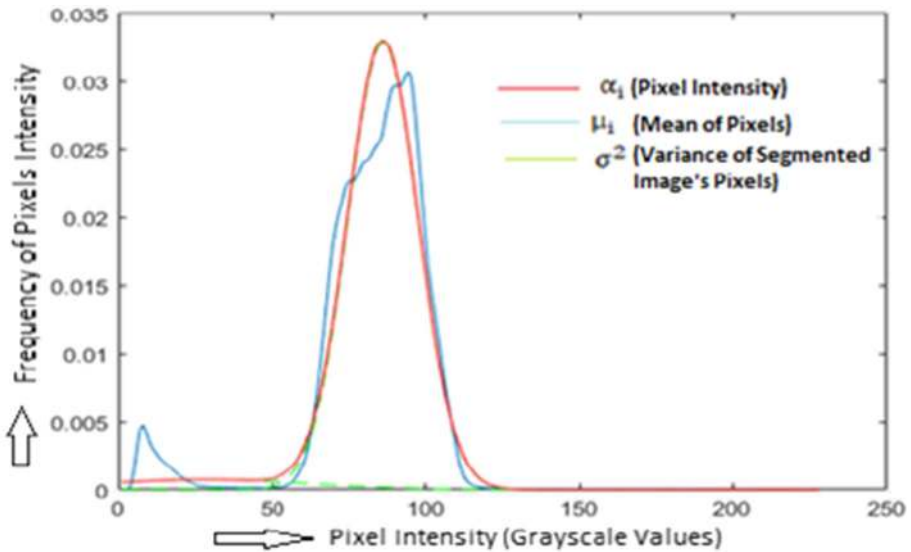


Fig. 6 Plots of sample bacteria 1 image's probability density and TPLMM-k Means algorithm with EM [pixel intensity (α_i), mean of pixels (μ_i), and variance of the behavior of segmented images of pixels (σ^2)]

the image with the current proposed model can be observed at Fig. 7. The grayscale converted image after using the current process was shown at Fig. 8.

The histogram representation of the current model with respect to bacteria 2 can be shown at Fig. 9 and the plots of density of probability with respect to the TPLMM-k Means algorithm had shown at Fig. 10.

The output values obtained for the several parameters considered in the current model are shown at Table 4. The sample input image of bacteria 3 can be observed at Fig. 11.

Table 3 Estimates of the parameters for sample bacteria 2 image

| Parameters | Initial Parameters | | | Refined estimates | | |
|------------|--------------------|----------|----------|-------------------|--------|--------|
| | Image Region | | | Image Region | | |
| | 1 | 2 | 3 | 1 | 2 | 3 |
| α_i | 0.500 | 0.500 | 0.500 | 0.2402 | 0.5711 | 0.1887 |
| μ_i | 83.1978 | 126.4605 | 185.7868 | 82.99 | 161.91 | 181.25 |
| σ^2 | 277.4086 | 167.946 | 139.257 | 240.2154 | 571.45 | 188.27 |

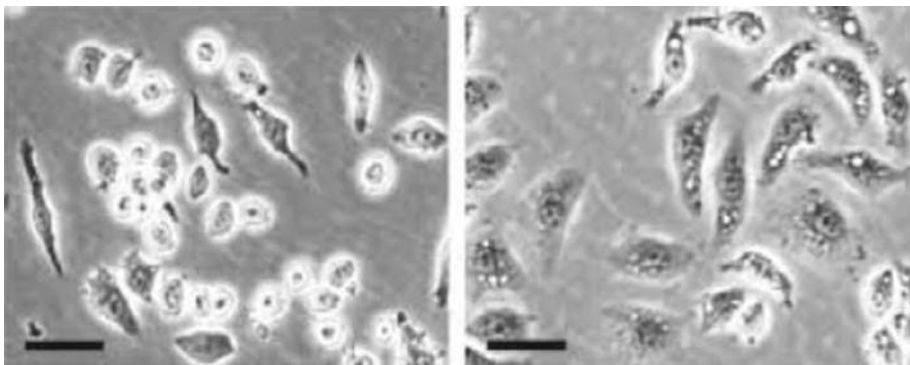


Fig. 7 Input image of sample bacteria 2 image

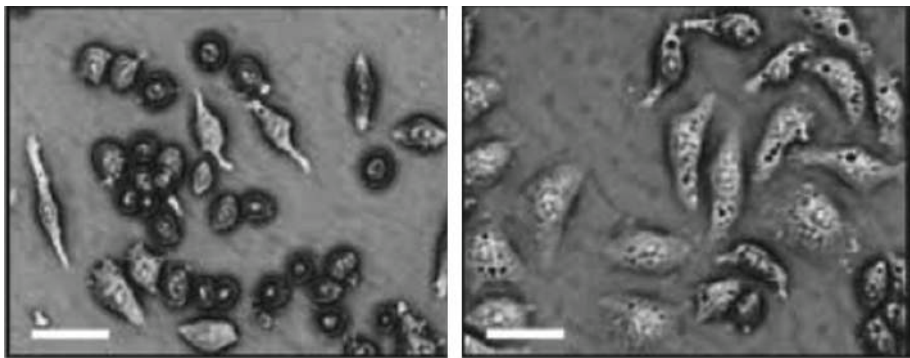


Fig. 8 Segmented image for sample bacteria 2 image

The histogram representation of the current model with respect to bacteria 3 can be shown at Fig. 13 and the plots of density of probability with respect to the TPLMM-k Means algorithm had shown at Fig. 14.

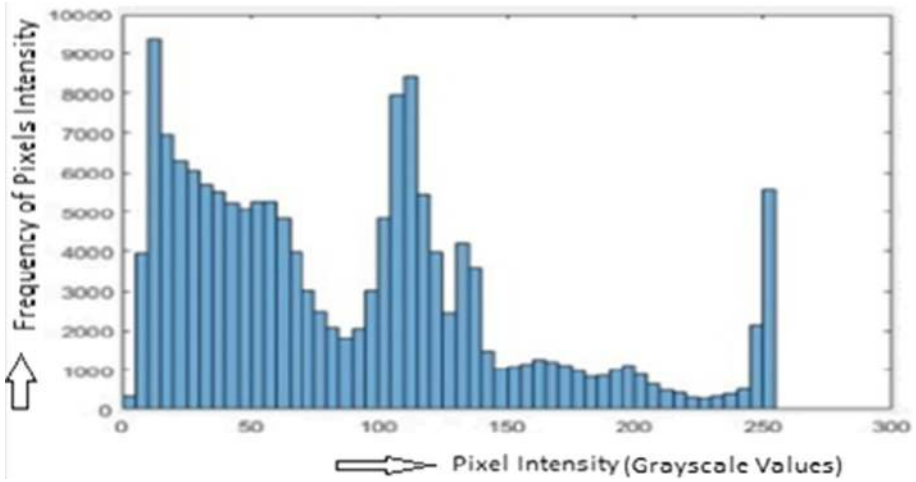


Fig. 9 Histogram for sample bacteria 2 image

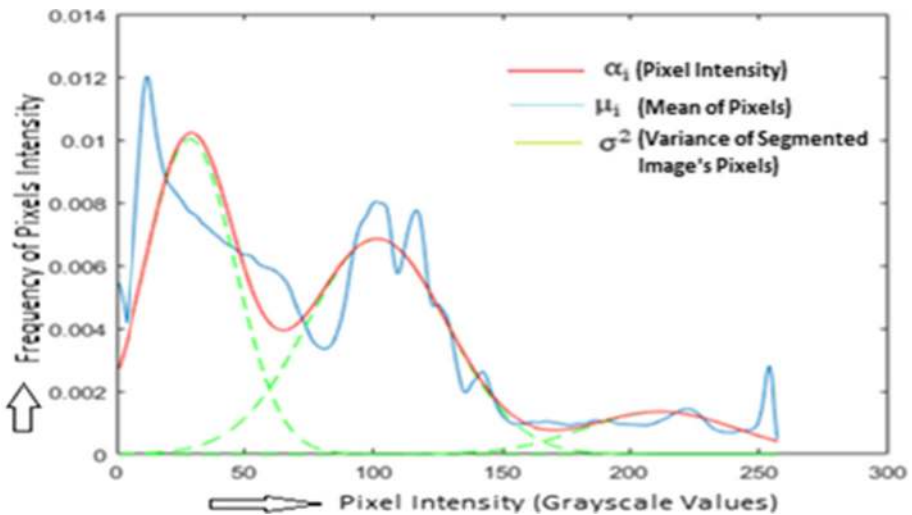


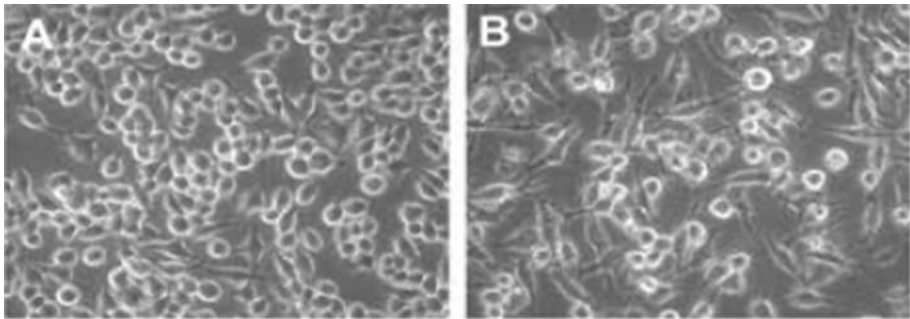
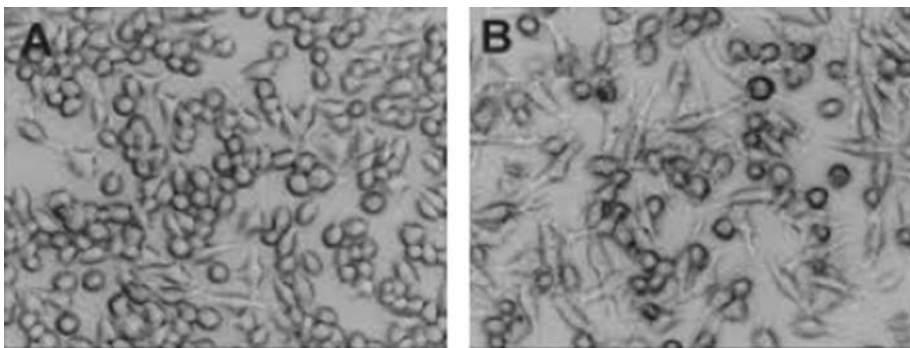
Fig. 10 Plots of sample input bacteria 2 image's probability density and TPLMM-k Means algorithm with EM [pixel intensity (α_i), mean of pixels (μ_i), and variance of the behavior of segmented images of pixels (σ^2)]

5.1 Performance evaluation

With the help of the chosen segmentation algorithm in the current work, the experimentation process had been completed. The performance evaluation of the current considered algorithm with the segmentation-based model had studied and analyzed. The performance of the current model had analyzed with the help of three performance evaluation metrics like GCE, PRI and VOI. The performance of the current considered model had analyzed with the help of three performance evaluation metrics. The model used in the current work was the three parameter Logistic mixture model with k means model (TPLMM-k Means Algorithm) is studied by

Table 4 Estimates of the parameters for bacteria 3 image

| Parameters | Initial Parameters | | | Refined estimates | | |
|--------------|--------------------|----------|--------|-------------------|---------|--------|
| | Image Region | | | Image Region | | |
| | 1 | 2 | 3 | 1 | 2 | 3 |
| α_i | 0.500 | 0.500 | 0.500 | 0.0635 | 0.9365 | 0.188 |
| μ_i | 40.5146 | 113.2603 | 184.78 | 23.11 | 99.33 | 181.25 |
| σ_i^2 | 64.09627 | 141.798 | 147.27 | 63.254 | 181.257 | 188.27 |

**Fig. 11** Input image of sample bacteria 3 image**Fig. 12** Segmented image for sample bacteria 3 image

computing the segmentation performance measures namely PRI, GCE, and VOI for the five images under study.

The performance of the developed algorithm using Three-Parameter Logistic mixture model with k means model (TPLMM-k means Algorithm) is analyzed by calculating the separation presentation procedures explicitly PRI, GCE, and VOI for the three bacteria images considered in the current work. The experimental values were calculated for both the earlier existing model of Gaussian Mixture Models and also the current model o had calculated

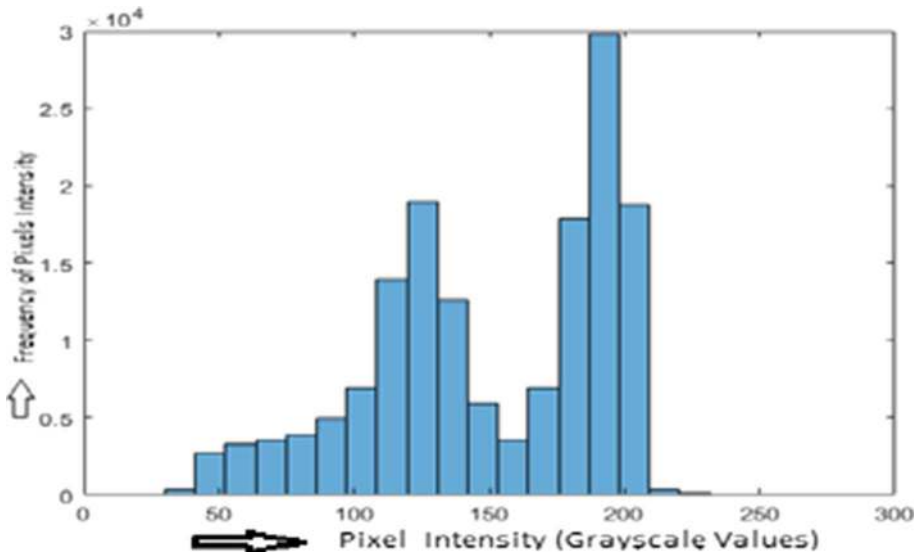


Fig. 13 Histogram for sample bacteria 3 image

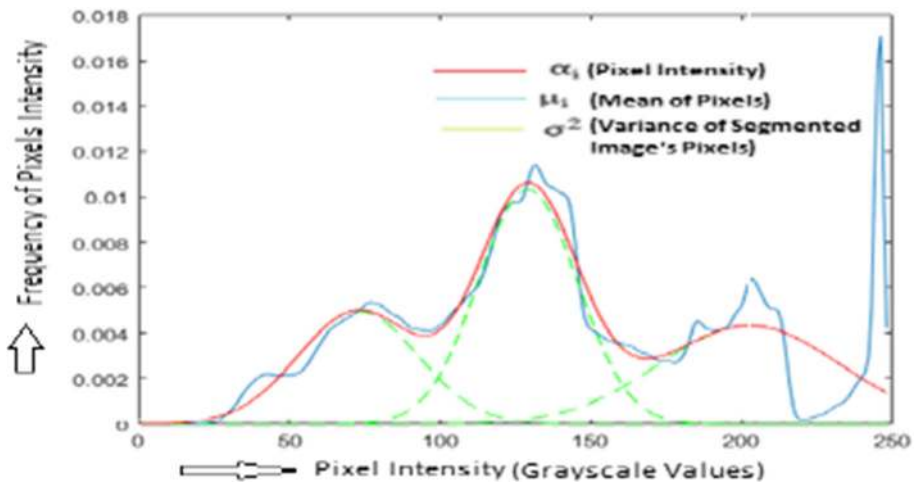
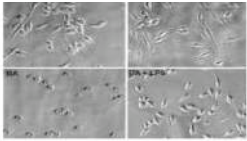
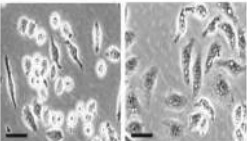
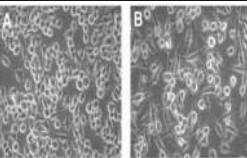


Fig. 14 Plots of sample bacteria 3 image's probability density and TPLMM-k Means algorithm with EM [pixel intensity (α_i), mean of pixels (μ_i), and variance of the behavior of segmented images of pixels (σ^2)]

and presented as tabular form in Table 5. The values can be understood easily and also can compared with the old existing models very easily.

In order to understand the performance of current developed model, the performance of the model had verified with respect to the performance evaluation metrics and had calculated these metric values for the three bacteria images and the values are represented in the above Table 4. From the results of these metrics, the values which had obtained from the considered model are better than that of the base model that is Gaussian Mixture Model with k-means

Table 5 Overall performance measures of segmentation for all sample input images

| Input Images | Methods | Performance Measures | | |
|---|--|----------------------|--------|--------|
| | | PRI | GCE | VOI |
|  | GMM (Gaussian Mixture Model) | 0.9147 | 0.2785 | 0.4273 |
| | TPLMM-k Algorithm (Three Parameter Logistic Mixture Model and k-means clustering) | 0.9190 | 0.2266 | 0.1763 |
|  | GMM (Gaussian Mixture Model) | 0.8845 | 0.0225 | 0.1448 |
| | TPLMM-k Algorithm (Three Parameter Logistic Mixture Model and k-means clustering) | 0.8968 | 0.0197 | 0.1421 |
|  | GMM (Gaussian Mixture Model) | 0.8852 | 0.0339 | 0.1927 |
| | TPLMM-k Algorithm (Three Parameter Logistic Mixture Model and k-means clustering) | 0.8898 | 0.0246 | 0.1539 |

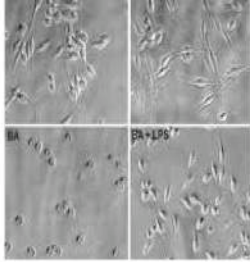
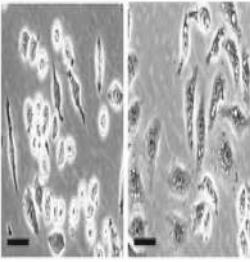
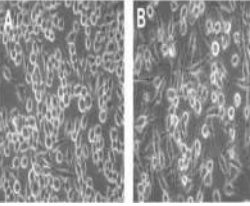
algorithm. The other two parameters to evaluate the performance of the model are considered and the values calculated will show as that the model is working perfectly as the output values of the GCE and VOI parameters are less than that of the GMM with k-means model.

5.2 Comparative study

A comparative study of the proposed Three-parameter Logistic mixture model with k-means model (TPLMM-k Means Algorithm) algorithm with that of the algorithms based on finite Gaussian mixture model reveals that the MSE of the proposed model is less than that of finite Gaussian mixture model. Based on all other quality metrics it is also observed that the performance of the proposed model in retrieving the images is better than the Gaussian mixture model. For better understanding of the working of the proposed model in the current paper, the results obtained in the current work had compared with the various other models considered by various authors had studied. The results obtained by those authors in their works also compared in Table 5 for better understanding of the work.

The other models developed by other authors and considered here in the current paper are, Unnikrishnan et al. (Unnikrishnan et al., 2007) and Jyothirmayi et al. (Jyothirmayi et al., 2016). The models they had developed and considered the performance are Gaussian Mixture Models with Hierarchical Clustering (GLMM-H), Gaussian Mixture Models with k-Means Clustering (GLMM-k), Double Truncated Gaussian Mixture Models with k-Means Clustering (DTGLMM-k). A comparison had presented in the below Table 6 with all these three models with the currently developed Three-Parameter Logistic Mixture Model with k means model (TPLMM-k Means Algorithm) algorithm. The results are displayed in the following Table 6 as follows,

Table 6 Comparative study of GMM (Gaussian Mixture Model) with TPLMM-k-means algorithm (Three Parameter Logistic Mixture Model and k-means clustering)

| Input Images | Quality Metrics | GLMM-H | GLMM-k | DTGLMM-k | Proposed TPLMM-k means Algorithm |
|---|-----------------------|---------|--------|----------|----------------------------------|
|  | Average Difference | 0.5315 | 0.5320 | 0.5290 | 0.5104 |
| | Maximum Distance | 0.4763 | 0.4866 | 0.4987 | 0.5273 |
| | Image Fidelity | 0.8124 | 0.9900 | 0.9900 | 0.8247 |
| | Mean Square Error | 0.0770 | 0.0685 | 0.0631 | 0.0781 |
| | Signal to Noise Ratio | 14.080 | 0.4190 | 0.3000 | 14.323 |
| | Image Quality Index | 0.8460 | 0.9880 | 0.9900 | 0.9950 |
|  | Average Difference | 0.3211 | 0.3125 | 0.3024 | 0.2101 |
| | Maximum Distance | 0.6810 | 0.5687 | 0.5987 | 0.6941 |
| | Image Fidelity | 0.6885 | 0.9900 | 0.9900 | 0.6954 |
| | Mean Square Error | 0.0645 | 0.0547 | 0.0512 | 0.0451 |
| | Signal to Noise Ratio | 4.0802 | 0.4100 | 4.2000 | 4.2540 |
|  | Image Quality Index | 0.7763 | 0.9480 | 0.9000 | 0.9514 |
| | Average Difference | 0.2350 | 0.3254 | 0.2568 | 0.4581 |
| | Maximum Distance | 0.5925 | 0.5847 | 0.5774 | 0.6547 |
| | Image Fidelity | 0.9882 | 0.9900 | 0.9900 | 0.9981 |
| | Mean Square Error | 0.0038 | 0.0045 | 0.0041 | 0.0012 |
| | Signal to Noise Ratio | 11.1494 | 1.1300 | 11.354 | 14.521 |
| Image Quality Index | 0.9869 | 0.9800 | 0.9900 | 0.9914 | |

From Table 6, It is observed that all the image quality measures for the three images are meeting the standard criteria. This implies that using the proposed algorithm, the images are retrieved accurately. The metrics considered in Table 6 are average difference, maximum distance, image fidelity, mean square error and the signal to noise ratio, image quality index. The current considered model is working perfectly for segmenting the images and identifying the content on the images perfectly with resp to the values obtained from the models.

5.3 Applications and major contributions of the proposed model

The current model has vast areas of applications in industry and for individuals in their day-to-day life. Some of them are,

- This model helps in medical images
 - This model helps in filming and video
 - This model helps in security surveillance
 - This model helps in face recognition
 - This model helps in gesture analysis
 - This model helps in computer vision
 - This model helps in human computer interactions
- Some of the major contributions of the proposed model on COG dataset are as follows,
- TPLMM-K Means is used for the automated distinction of bronchi blemishes. Reviewed with the GLMM-H, GLMM-K, DTGLMM-K, TPLMM-K has been effortlessly encoding richer longitudinal information to disregard supplementary single segmentation.
 - Multiview places remain in point of fact applied in our proposals. Our staff use the Multiview-one-network strategy that contrasts and arises from the one-view-one-network technique used in the study. Completion leads to the fact that our technique may achieve a decreased mistake rate than the one-view-one-network approach while using far fewer requirements. Note that while the layout used a similar approach and GLMM-H, GLMM-K, DTGLMM-K, TPLMM-K is simply utilized, our group used TPLMM-K Means in this paper.
 - (To the best of our knowledge, this is the first study to use TPLMM-K Means substitutes of GLMM-H as well as GLMM-K and DTGLMM-K to categorize bacteria imperfections.
 - Our model attained a much better outcome than the remaining models associated with the differences in Bio-medical Images.

We utilize TPLMM-K Means for the automatic detection of bacteria. Our team used the improvised proposed algorithm.

6 Conclusions

This paper deals with a novel application of three-parameter logistic type distribution in image decomposition. The whole image is characterized by a finite mixture of three-parameter logistic type distribution. The three-parameter logistic type distribution includes a family of platykurtic distributions. The updated equations of the model parameters are derived and solved using MATLAB code. The initialization of parameters is done through k-means algorithm and moment method of estimation. An experimentation which is generates digital micro scope camera images taken as database revealed that the proposed algorithm performs better with respect to image decomposition metric that image decomposition metrics method of GMM. The hybridization of model-based approach with k-means clustering improves the efficiency of decomposition. The proposed algorithm can be further extended to the other method of estimation of the model parameters such as Monto-Carlo methods and Bootstrapping methods which will be taken elsewhere.

References

- Akhavan, R., & Faez, K. (2014). A novel retinal blood vessel segmentation algorithm using fuzzy segmentation. *International Journal of Electrical and Computer Engineering*, 4(4), 561–572.
- Bechar, M. E., Settouti, N., Barra, V., et al. (2018). Semi-supervised superpixel classification for medical images segmentation: Application to detection of glaucoma disease. *Multidimensional Systems and Signal Processing*, 29, 979–998.
- J. Bilmes, A. Vahdat, W. Hsu, E.-J. Im, (1997). Empirical observations of probabilistic heuristics for the clustering problem. *Technical Report ICSI-TR097–018, ICSI*.
- Cimpoi, M., Maji, S., Kokkinos, I., & A. Vedaldi A. . (2016). Deep filter banks for texture recognition, description, and segmentation. *International Journal of Computer Vision*, 118(1), 65–94.
- Dias, P. A., Dunkel, T., Fajado, D. A. S., et al. (2016). Image processing for identification and quantification of filamentous bacteria in in situ acquired images. *BioMedical Engineering OnLine*, 15(64), 1–21.
- Guang, W. X., & Chen, S. H. (2012). An improved image segmentation algorithm based on two-dimensional otsu method. *Information Science Letters*, 1(2), 77–83.
- Hay, E. A., & Parthasarathy, R. (2018). Performance of convolutional neural networks for identification of bacteria in 3D microscopy datasets. *PLoS Computational Biology*, 14(12), 1–24.
- Jeckel, H., & Drescher, K. (2021). Advances and opportunities in image analysis of bacterial cells and communities. *FEMS Microbiology Reviews*. <https://doi.org/10.1093/femsre/fuaa062>
- Jyothirmayi, T., Srinivasa Rao, K., & Srinivasa Rao RaoSatyanarayana Ch, PCh. (2016). Image segmentation based on doubly truncated generalized laplace mixture model and k means clustering. *International Journal of Electrical and Computer Engineering*, 6(5), 2188–2196.
- Khamparia, A., Bharati, S., Podder, P., P, et al. (2021). Diagnosis of breast cancer based on modern mammography using hybrid transfer learning. *Multidimensional Systems and Signal Processing*, 32, 747–765. <https://doi.org/10.1007/s11045-020-00756-7>
- Lee, D., Lee, J., Ko, J., Yoon, J., Ryu, K., & Nam, Yo. (2019). Deep learning in MR image processing. *Investigate Magnetic Resonance Imaging*, 23(2), 81–99.
- Li, Z., Liu, M., Wang, H., Yang, Y., Chen, J., & Jin, G. (2013). Gray-scale edge detection and image segmentation algorithm based on mean shift. *Indonesian Journal of Electrical Engineering and Computer Science*, 11(3), 1414–1421.
- Liang, Y., Sun, L., Ser, W., et al. (2017). Hybrid threshold optimization between global image and local regions in image segmentation for melasma severity assessment. *Multidimensional Systems and Signal Processing*, 28, 977–994.
- Lie, T., & Sewehand, W. (1992). Statistical approach to X -ray CT imaging and its applications in image analysis. *IEEE Transactions on Medical Imaging*, 11(1), 53–61.
- Lie, T., & Udupa, J. K. (1993). Performance evaluation of finite normal mixture model based image segmentation. *IEEE Transactions on Image Processing*, 12(10), 1153–1169.
- McLachlan, G. J., & Krishnan, T. (1997). *The EM Algorithm and Extensions*. Wiley.
- B. A. Mohamed, H. M. Afify, (2018). Automated classification of bacterial images extracted from digital microscope via bag of words model. *9th Cairo International Biomedical Engineering Conference (CIBEC), Cairo, Egypt*, 86–89.
- Nahar, M., & Ali, M. S. (2014). An improved approach for digital image edge detection. *International Journal of Recent Development in Engineering and Technology*, 2(3), 14–20.
- Preetha, V., & Pandi Selvi, P. (2018). Identification of bacteria using digital image processing. *International Journal of Engineering Research in Computer Science and Engineering*, 5(3), 254–258.
- Sharma, A., Kumar, S., & Singh, S. N. (2019). Brain tumor segmentation using DE embedded OTSU method and neural network. *Multidimensional Systems and Signal Processing*, 30, 1263–1291.
- Srinivas, Y., & Srinivas Rao, K. (2007). Unsupervised image segmentation using finite doubly truncated Gaussian mixture model and hierarchical clustering. *Journal of Current Science*, 93(4), 507–514.
- Tatusov, R. L., Koonin, E. V., & Lipman, D. J. (1997). A genomic perspective on protein families. *Science*, 278(5338), 631–637. <https://doi.org/10.1126/science.278.5338.631> PMID: 9381173.
- Unnikrishnan, R., Pantofaru, C., & Hebert, M. (2007). Toward objective evaluation of image segmentation algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6), 929–944.
- M. F. Wahid, T. Ahmed, M. A. Habib, (2018). Classification of microscopic images of bacteria using deep convolutional neural network. *10th International Conference on Electrical and Computer Engineering (ICECE), Dhaka, Bangladesh*, 217–220.
- Wang, H., Kodmir, H. C., Qiu, Y., et al. (2020). Early detection and classification of live bacteria using time-lapse coherent imaging and deep learning. *Light Science Applications*, 9(118), 1–17.
- Xin, J., Wang, Z., Z., S. Tian, , et al. (2017). NMR image segmentation based on unsupervised extreme learning machine. *Multidimensional Systems and Signal Processing*, 28, 1013–1030.

- Yin, Y., & Ding, Y. (2009). A close to real-time prediction method of total coliform bacteria in foods based on image identification technology and artificial neural network. *Food Research International*, 42(1), 191–199. <https://doi.org/10.1016/j.foodres.2008.10.006>
- Zhang, Y., Jiang, H., Ye, T., & Juhas, M. (2021). Deep learning for imaging and detection of microorganisms. *Trends in Microbiology*, 29(7), 569–572.
- Zhang, Z. H., Chen, C., Sun, J., & Chan, K. L. (2003). EM algorithms for gaussian mixtures with split-and-merge operation. *Pattern Recognition*, 36(9), 1973–1983.
- Zhong, Z., Wang, M., Gao, W., et al. (2021). A novel multisource pig-body multifeature fusion method based on gabor features. *Multidimensional Systems and Signal Processing*, 32, 381–404. <https://doi.org/10.1007/s11045-020-00744-x>
- Zieliński, B., Plichta, A., Misztal, K., Spurek, P., Brzywczy-Włoch, M., & Ochońska, D. (2017). Deep learning approach to bacterial colony classification. *PLoS ONE*, 12(9), 1–18.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



K.V. Satyanarayana received Ph.D. from JNTUK Kakinada, Kakinada, India. Currently, Dr. K. V. Satyanarayana Rao associated with Raghu Engineering College (A), Visakhapatnam-531162, India, as Associate Professor of Computer Science and Engineering of the Institute year 2020. His research areas include Queuing Models, Stochastic Modeling, Image Processing, Pattern Recognition. He published 25+ research papers in various reputed International Journals and Conferences. He is a member of ACM, CSI, and ISPS.



N. Thirupathi Rao received Ph.D. from Andhra University, Visakhapatnam, India. Currently, Dr. N. Thirupathi Rao associated with Vignan's Institute of Information Technology, Visakhapatnam-530049, India, as Associate Professor of Computer Science and Engineering of the Institute year 2015. His research areas include Communication Networks, Queuing Models, Stochastic Modeling, Image Processing, Pattern Recognition, Bio-Informatics, Evolutionary Computing, and Security. He published 55+ research papers in various reputed International Journals and Conferences. He is a member of ACM, IE, CSI, and ISPS.



Debnath Bhattacharyya received Ph.D. (Tech., CSE) from the University of Calcutta, Kolkata, India. Currently, Dr. Bhattacharyya associated with K L Deemed to be University, KLEF, Guntur, AP, India as Professor of Computer Science and Engineering. His research areas include Image Processing, Pattern Recognition, Bio-Informatics, Computational Biology, Evolutionary Computing, and Security. He published 200+ research papers in various reputed International Journals and Conferences. He published six textbooks for Computer Science as well. He is a member of IEEE, ACM, ACM SIGKDD, IAENG, and IACSIT.



Yu-Chen Hu received his PhD. degree in computer science and information engineering from the Department of Computer Science and Information Engineering, National Chung Cheng University, Taiwan. Currently, Dr. Hu is a professor in the Department of Computer Science and Information Management, Providence University, Taiwan. He is a senior member of IEEE. He is also a member of Computer Vision, Graphics, and Image Processing, Chinese Cryptology and Information Security Association, Computer Science and Information Management and Phi Tau Phi Society of the Republic of China. He joints the editorial boards of several other journals. His research interests include image and signal processing, data compression, information hiding, information security, computer network, and machine learning.