IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 6, 2019, accepted July 8, 2019, date of publication July 18, 2019, date of current version August 7, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2929857

Analysis and Identification of Dermatological Diseases Using Gaussian Mixture Modeling

CHAAHAT GUPTA[1,2], (Student Member, IEEE), NAVEEN KUMAR GONDHI[1], AND PARVEEN KUMAR LEHANA[3]

[1]Department of Computer Science and Engineering, Shri Mata Vaishno Devi University, Katra 182301, India
[2]MIET, Jammu 181122, India
[3]Department of Electronics, University of Jammu, Jammu 180006, India

Corresponding author: Chaahat Gupta (chaahatgupta249@gmail.com)

ABSTRACT Skin diseases are common and are mainly caused by virus, bacteria, fungus, or chemical disturbances. Timely analysis and identification are of utmost importance in order to control the further spread of these diseases. Control of these diseases is even more difficult in rural and resource-poor environments due to a lack of expertise in primary health centers. Hence, there is a need for providing self-assisting and innovative measures for the appropriate diagnosis of skin diseases. Use of mobile applications may provide inexpensive, simple, and efficient solutions for early diagnosis and treatment. This paper investigates the application of the Gaussian mixture model (GMM) based on the analysis and classification of skin diseases from their visual images using a Mahalanobis distance measure. The GMM has been preferred over the convolution neural network (CNN) because of limited resources available within the mobile device. Gray-level co-occurrence matrix (GLCM) parameters contrast, correlation, energy, and homogeneity derived from skin images have been used as the input data for the GMM. The analysis of the results showed that the proposed method is able to predict the classification of skin diseases with satisfactory efficiency. It was also observed that different diseases occupy distinct spatial positions in multidimensional space clustered using the Mahalanobis distance measure.

INDEX TERMS Artificial neural network, convolution neural network, dermatology, dermoscopy, Euclidean distance, Gaussian mixture model, gray level co-occurrence matrix, Mahalanobis distance, melanoma, support vector machine, teledermoscopy, ultra-violet radiations.

I. INTRODUCTION

Human skin performs various functions like Vitamin-D synthesis, internal organs protection, control of water loss, and shielding the body from environmental hazardous. Human skin consists of three layers: epidermis, dermis, and hypodermis as shown in Fig. 1. The external layer is called epidermis. It is the thinnest layer with thickness varying from 0.05 to 0.15 mm. It provides mechanical resistance and acts as barrier against bacteria, harmful chemicals, and ultraviolet (UV) radiations [1]. Dermis is the middle layer whose thickness varies from 1.5 mm to 4 mm. Its primary function is to protect the body from mechanical stress and strain. It is divided in two strata, papillary dermis and reticular dermis. Papillary dermis consists of loose fiber bundles connecting it to the epidermis. Whereas, the reticular dermis is much thicker than the papillary dermis and contains dense networks of elastin, collagen, reticular fibers, capillary vessels, sensory receptors, and hair follicles [2]. Beneath the dermis, is a layer of fat and loose fibers known as hypodermis. It stocks fat and provide thermal insulation. Skin diseases are generally caused by virus, bacteria, fungus, and chemical disorders. Uncontrolled spread of the skin diseases may be dangerous and hence, timely treatment is important and also contagious skin diseases may prove to be even more dangerous [3] for which various methods are in use for automatic identification, classification, and prediction of the necessary precautions to be taken [4]. The problem is more severe in visually similar skin diseases.

In resource-poor environment, where health workers have even lesser expertise specially in dermatology, need more convenient and innovative measures for the proper diagnosis of the skin diseases. In these areas, dermatological services are commonly provided by medical staff and expertise in

The associate editor coordinating the review of this manuscript and approving it for publication was Cristian A. Linte.

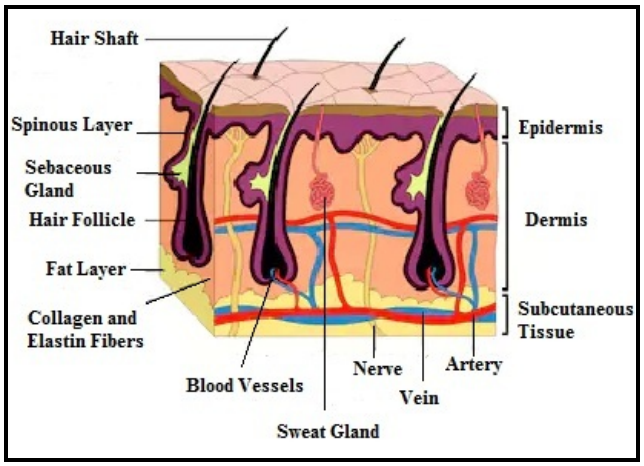

FIGURE 1. Layers of human skin [1], [2].

the dermatology cannot be expected. Therefore, queries are generally sent to specialists and the response takes several days to arrive. Studies conducted in rural areas of countries like Colombia showed that average waiting time for a dermatologist was more than three weeks [5], [6]. Thus, the use of mobile applications to enable real-time dermatological diagnosis in these areas has great applicability. Innovative technical solutions can help in bridging the gap in resource-poor environment [7], [8]. Several dermatology related mobile phone applications are available and leading to the share of teledermatology from 11.0% in 2014 to 20.1% in 2017 [9]. Most of these applications provide only consultation for self-diagnosis using visual, audio, and data services of mobile communication [10]–[15]. Teledermoscopy reduces costs, avoids unnecessary biopsies, and decreases the time to initial therapy [16], [17].

In dermatology, identification of the problem of skin, nails, and hair is carried out. Sometimes, allergies, irritants, genetic structure, and immune system disorder is also responsible for dermatitis, hives, and other skin problems such as acne, cold sore, blisters, hives, actinic keratosis, carbuncle, eczema, psoriasis, measles, etc. Human skin has many types of cancers like melanoma, basal cell carcinoma, and squamous cell carcinoma. Dermoscopy is a non-invasive and visual symptoms based method for the identification of skin abnormalities. Identification carried out by naked eye is generally limited in accuracy, therefore, computer assisted techniques are more effective [18].

Recently, artificial intelligence has also been used in dermoscopy [19]. CNN has shown encouraging results in accurate diagnosing of the skin diseases. These applications under-perform with the images taken in poor lighting conditions and generally lead to wrong diagnoses [20]. The problem is more severe in the diseases with similar symptoms. The main limitations of CNN is the code complexity and the requirement of large amounts of input data for training [20], [21].

Some of the important algorithms used in the area of image processing for estimation and approximation of images include adaptive image equalization algorithm [22] which automatically enhances the contrast of an image using GMM. Contrast equalized image is generated by the preeminent gaussian component and cumulative distribution functions of the input intervals. Ibrahim and Pik Kong [23] introduced sub-regions histogram equalization that partitions the image based on its smoothed intensity values that are obtained by convolving the input image with a gaussian filter. Multilayer feed forward neural network [24] for precise and computationally effective division of components from the dermoscopic image utilizing genetically optimized fuzzy grouping approach is used. The literature reviewed on melanoma skin cancer [25] highlights that various approaches like artificial neural network (ANN) and data mining can be used for classifying skin cancer images. Accuracy obtained by these respective algorithms are 95-98% and 85%. In continuation to the above literature, various skin diseases can be detected and classified with various approaches, some of the important ones include: wavelet transformation and fuzzy inference system [26], support vector machine (SVM) [27] with 65.56% accuracy, k-means clustering and fuzzy-c means clustering [28], rule based and forward chaining inference engine [29], case based reasoning [30] achieving accuracies of 70%, 66.6% and 90%, respectively. For human skin color detection various methods have been used which include statistical modeling (GMM) [31]–[33] and genetic algorithm [34].

GMM has been used in various applications. In this context, GMM for classification of Alzheimer's disease is introduced in [35], [36] leading to the fact that the approach used in [36] is better than statistical hypothesis testing. A combination of GMM and various generative models [37] like k-nearest neighbors, naive bayes, multilayer perceptron and discriminative models (SVM, decision trees) have been reported for emotion recognition. Also, GMM classifier is used for identification of normal and abnormal retinal images of patients suffering from diabetes which attained an accuracy of 97.78%. GMM is used for multiple limb motion classification [38], [39], using continuous myoelectric signals. In continuation, GMM along with genetic algorithm is also used in [40] for auto segmentation of magnetic resonance images (MRI) lesions. Gaussian mixture model and logarithmic linearization algorithms [41] are used for pattern classification of ECG signals achieving an accuracy of 99.21%. In [42], Equal-Variance GMM has been used to model the characteristics of images, where, equal variance is shared by all the GMM variables. It has also been used in identification of cancer chemosensitivity of heterogeneous cellular response to perturbations in fluorescent sphingolipid metabolism [43] for extracting texture and intensity from the cellular images of the flow cytometry assay. GMM-based approach has also been used in [44] to multiparametrically characterize prostate tissue on transrectal dynamic contrast-enhanced ultrasonography giving an accuracy of 81%.

GMM has also found its application [45] for detection of falling positions in human beings. In this context,

authors contribution has been to extract six postures of physically movements of human beings including lying, sitting, standing, getting up, walking, and falling from height captured in video camera. Mixture of gaussian model combined with average filter models have been used in this approach. Although, GMM has widely been used for several classification based applications, its use for analysis and classification of skin diseases has not yet gained much momentum.

Mostly, Euclidean distance is used for multi-dimensional classification and hence leads to limited accuracy for search spaces with different weighted coordinate axis. Zhang *et al.* [46] developed a method called low-rank and sparse matrix decomposition-based mahalanobis distance (MD) method for anomaly detection. Their method used MD for detection of probable anomalies lying in the images analyzed from sparse matrix decomposition. MD has also been used in Ribonucleic acid(RNA) sequencing to analyze molecules for prediction of breast cancer survival rate [47]. In [48], it was reported that MD solved the clustering problems associated with traditional Euclidean Distance (ED) in clustering ECG features by reducing iterations to 50%.

Melanoma detection method based on Mahalanobis distance learning and constrained graph regularized non-negative matrix factorization has been successfully applied in [49] by incorporating global along with the local geometry in supervised learning based training for dimensionality reduction.

In [50] extreme learning machine method for multi-classification with Mahalanobis distance approach has also been investigated. MD was used for inter-class and ED for intra-class distance measurement resulting in about 1% improvement. The same approach has been adopted in the present investigations for distance measurement in four dimensions of the feature vectors (C, C_r, E, and H) using 72 weights, 8 priors, 8×4 centers and 8×4 co-variances for GMM based classification of skin diseases.

The objective of the paper is to investigate the discriminative capabilities of gaussian mixture model based algorithm for the diagnosis of skin diseases from their visual images using Mahalanobis distance measure for mobile platforms where implementation using CNN is difficult because of the limited resources available within the device. GMM is computationally affordable, tractable, and efficient for small datasets in comparison to CNN [51]–[53]. Eleven different types of skin diseases (Molluscum Contagiosum, Milia, Discoid Lupus Erythematosus, Tinea Corporis, Warts, Acne Blackhead, Psoriasis, Discoid Eczema, Chromoblastomycosis, Athletes foot, Melanoma) along with their variants were taken for the investigations. The methodology for the estimation of GMM parameters has been discussed in the following section. The results and discussions are presented in Section III. Conclusions and future work have been discussed in Section IV.

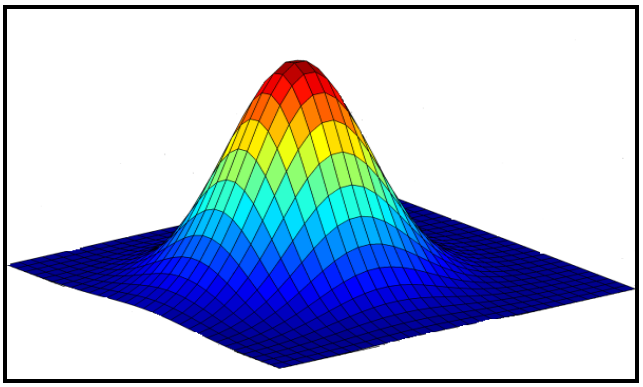

FIGURE 2. Multivariate Gaussian densities over two variables Y_1 and Y_2.

II. METHODOLOGY

This section describes the material and the algorithm used for investigating the discriminative capabilities of GMM based algorithm for skin diseases from their visual images. Multivariate Gaussian density over two variables Y_1 and Y_2 is shown in Fig. 2. The proposed method is shown in Fig. 3. Images of skin patches having different diseases were taken from DermNet Nz database [54]. For investigation, image of the normal human skin is taken as the reference. Images are resized to 256 pixels x256 pixels and RGB components are separated for each image. Each RGB component is segmented into 8x8 blocks. For each block, GLCM parameters (contrast, energy, correlation and homogeneity) are calculated. The distribution of GLCM parameters is approximated using GMM, which is a parametric density estimation approach assuming that input data is to be generated by more than one Gaussian process [55]. GMM may be written as a weighted sum of m components of Gaussian densities:

$$P(\mathbf{x}|\lambda) = \sum_{k=1}^{K} w_k g(\mathbf{x}|\boldsymbol{\mu}_\mathbf{k}, \sum_\mathbf{k}) \tag{1}$$

where, $\mathbf{x}$ is a D-dimensional feature vector, $w_k (k = 1, 2, \ldots K)$ denotes the mixture weights and $g(\mathbf{x}|\boldsymbol{\mu}_\mathbf{k}, \sum_\mathbf{k})(k = 1, 2, \ldots K)$ denotes components of various gaussian densities. Further, $\boldsymbol{\mu}_\mathbf{k}$ is mean vector and $\sum_k$ represents covariance matrix. The sum of all mixture weights is equal to one. Each component density is D-dimensional Gaussian function given by:

$$g(\mathbf{X}|\boldsymbol{\mu}, \sum) = \frac{1}{(2\pi)^{\frac{D}{2}}} \exp(-\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu})^T) \sum^{-1} (\mathbf{X} - \boldsymbol{\mu}) \tag{2}$$

Clustering can be improved by using GMM algorithm which can be used to estimate GMM parameters, i.e. mean ($\boldsymbol{\mu}_\mathbf{k}$), weight ($w_i$), and covariance ($\sum_i$) [56]. The advantages of using GMM based algorithm is that it has low complexity and scalability. It computes the probabilities of cluster memberships by maximizing the log-likelihood of the data generated. GMM is an iterative method in each step of which posterior probability P_{ik}^t at t iteration is given by [57], [58].

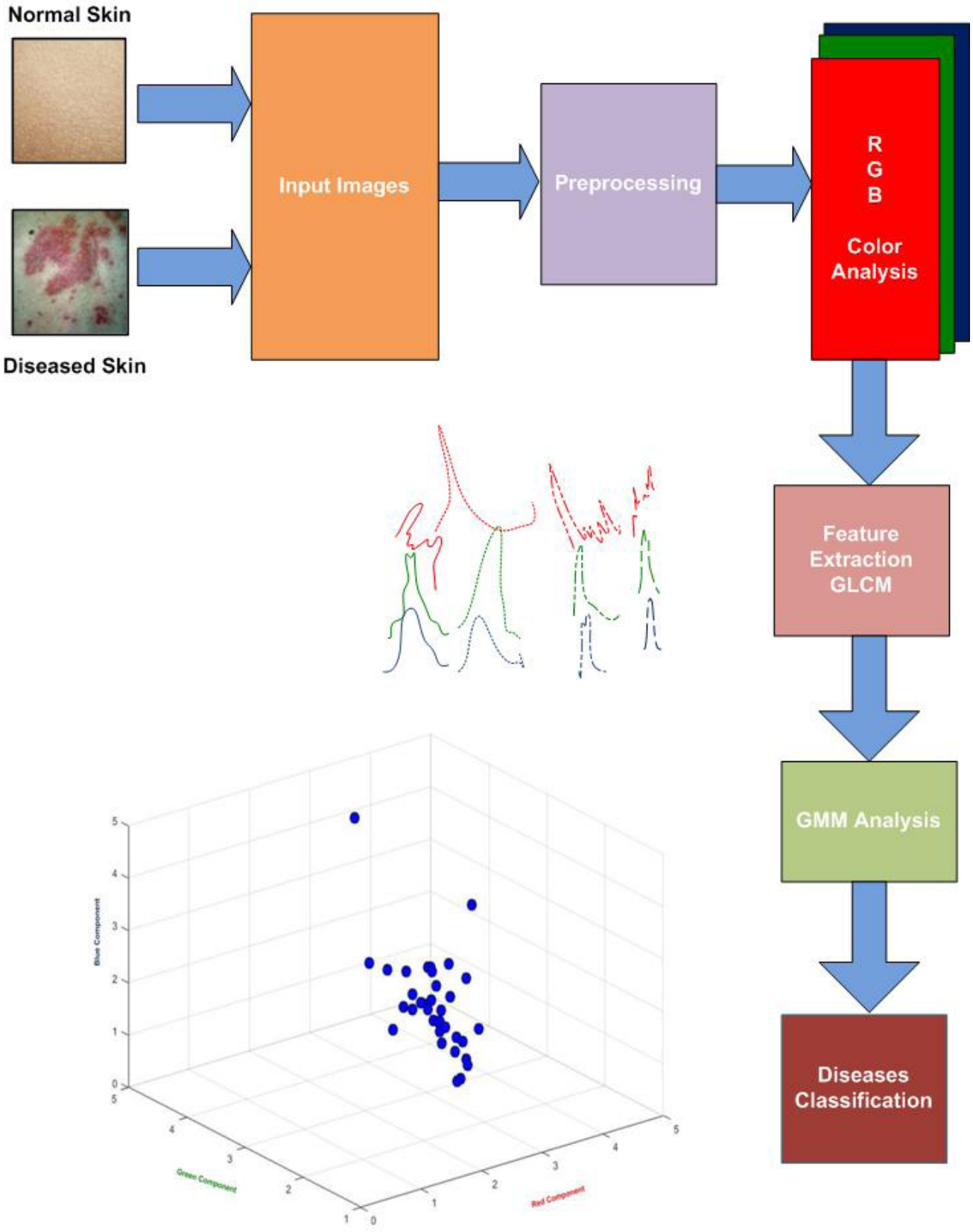

FIGURE 3. Proposed methodology. GLCM parameters of normal and diseased skin are calculated after resizing each color component of the images. Training of GMM is carried out and classification is performed using Euclidean/Mahalanobis distance measure approaches for classification.

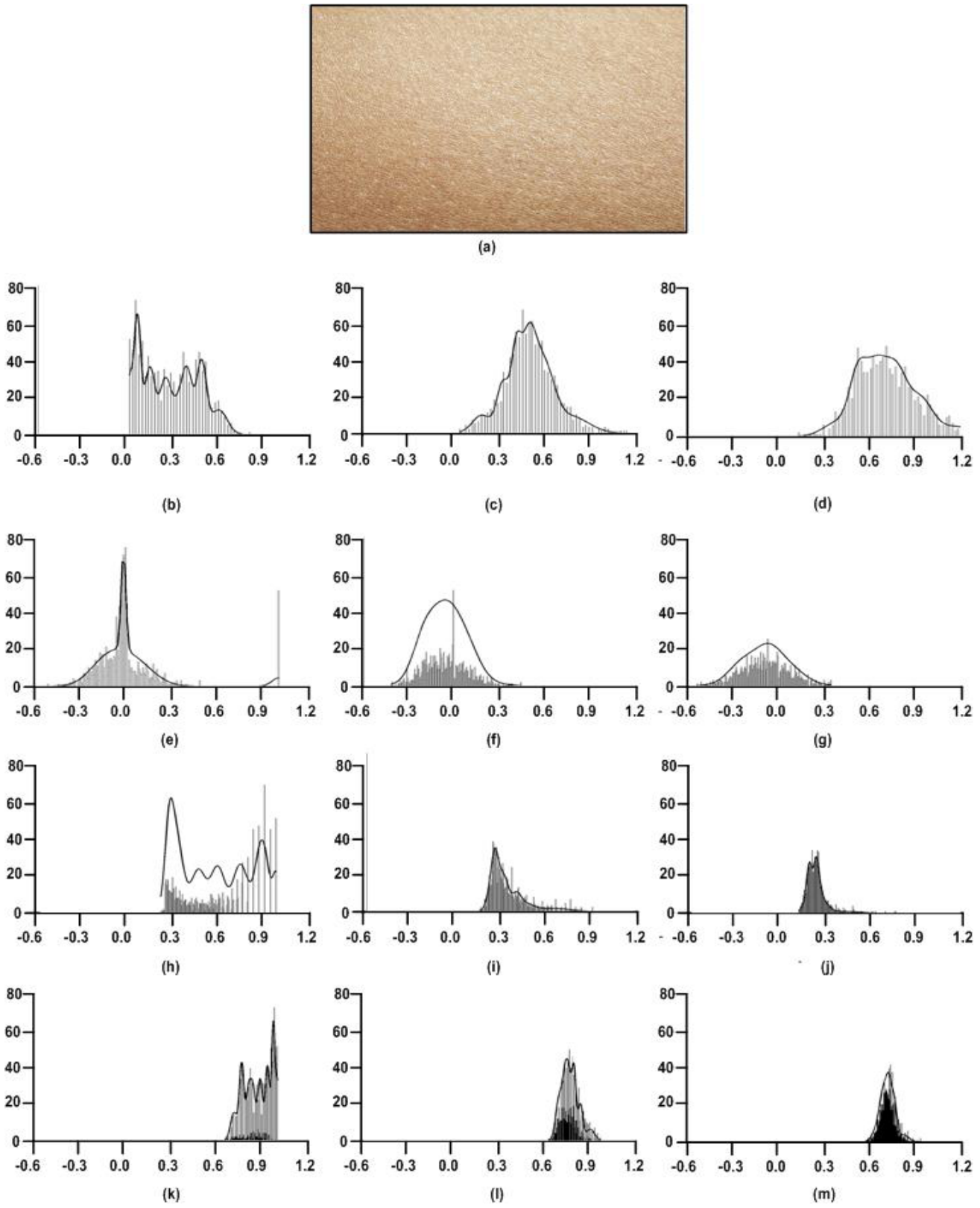

FIGURE 4. GMM distribution of the proposed framework for Normal skin.

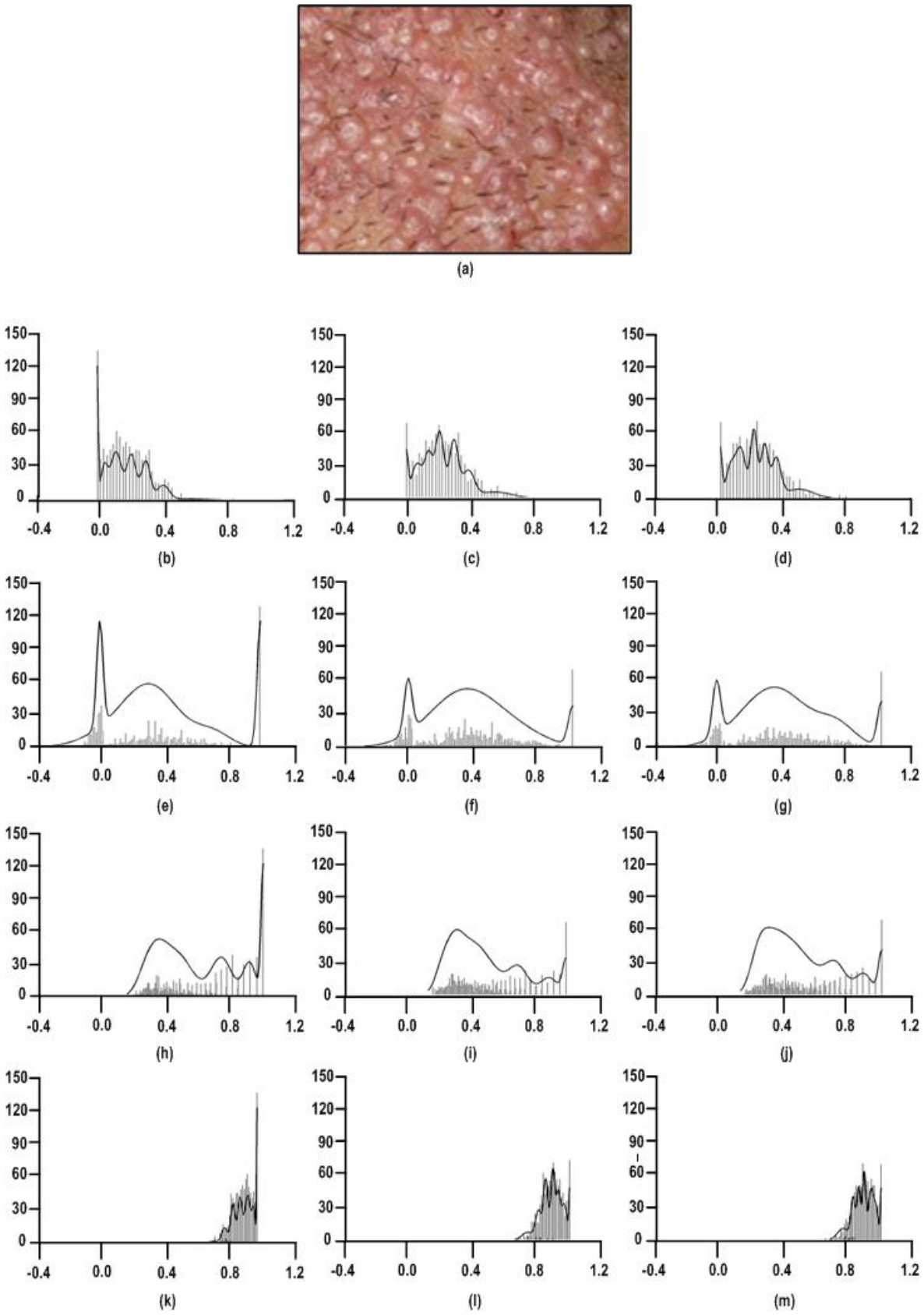

FIGURE 5. GMM distribution of the proposed framework for Molluscum Contagiosum.

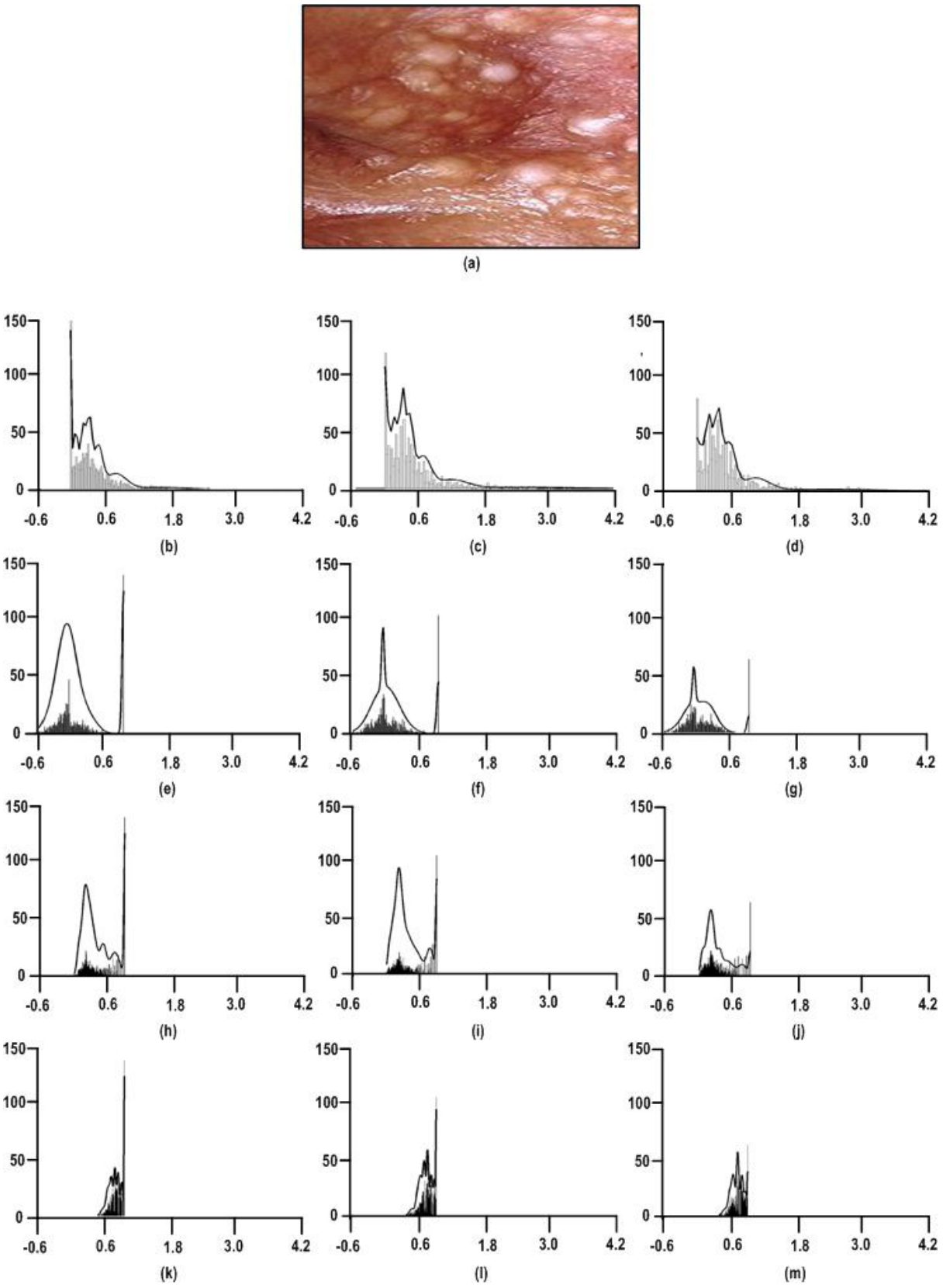

FIGURE 6. GMM distribution of the proposed framework for Milia.

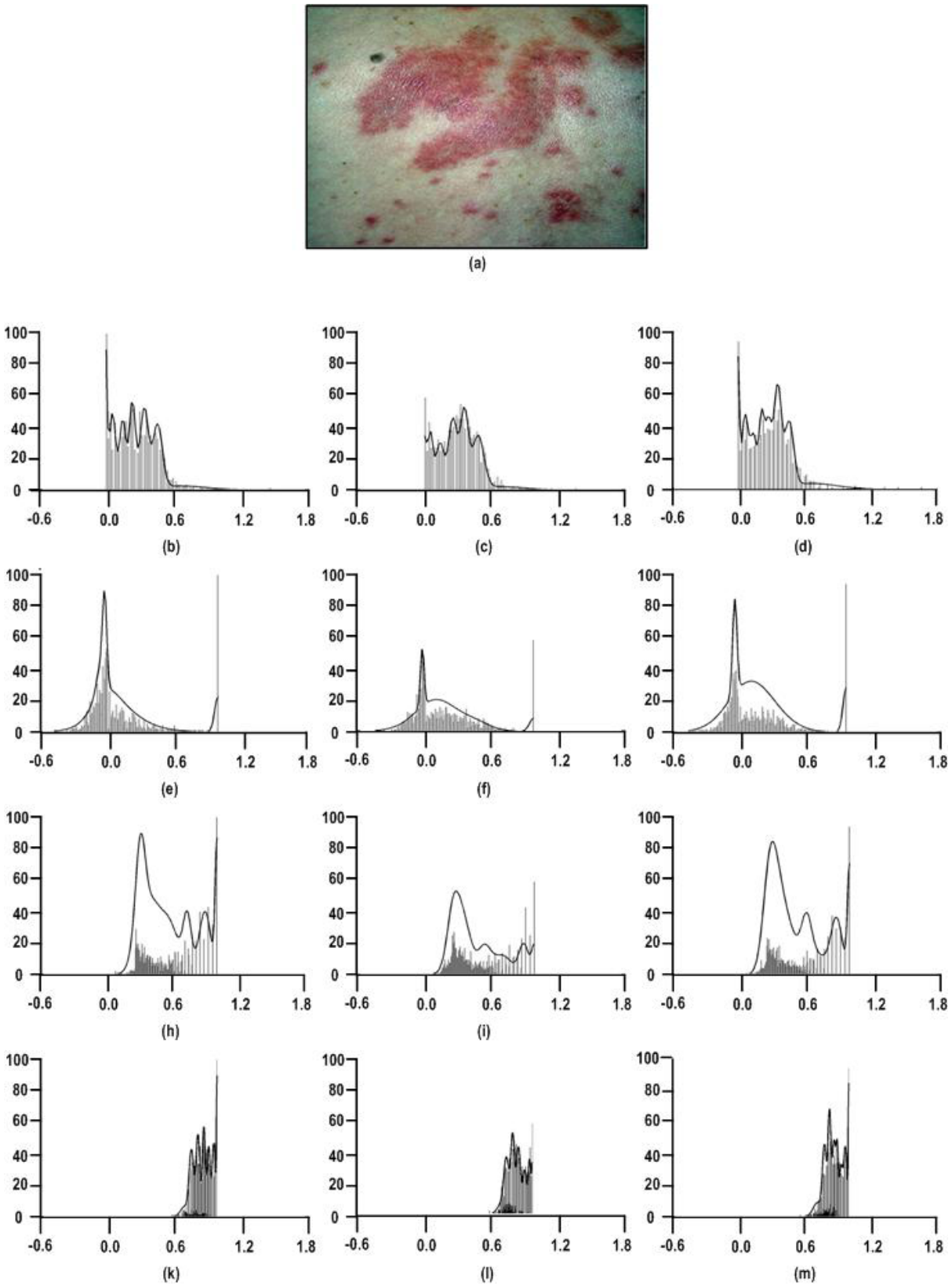

FIGURE 7. GMM distribution of the proposed framework for Discoid Lupus Erythematosus.

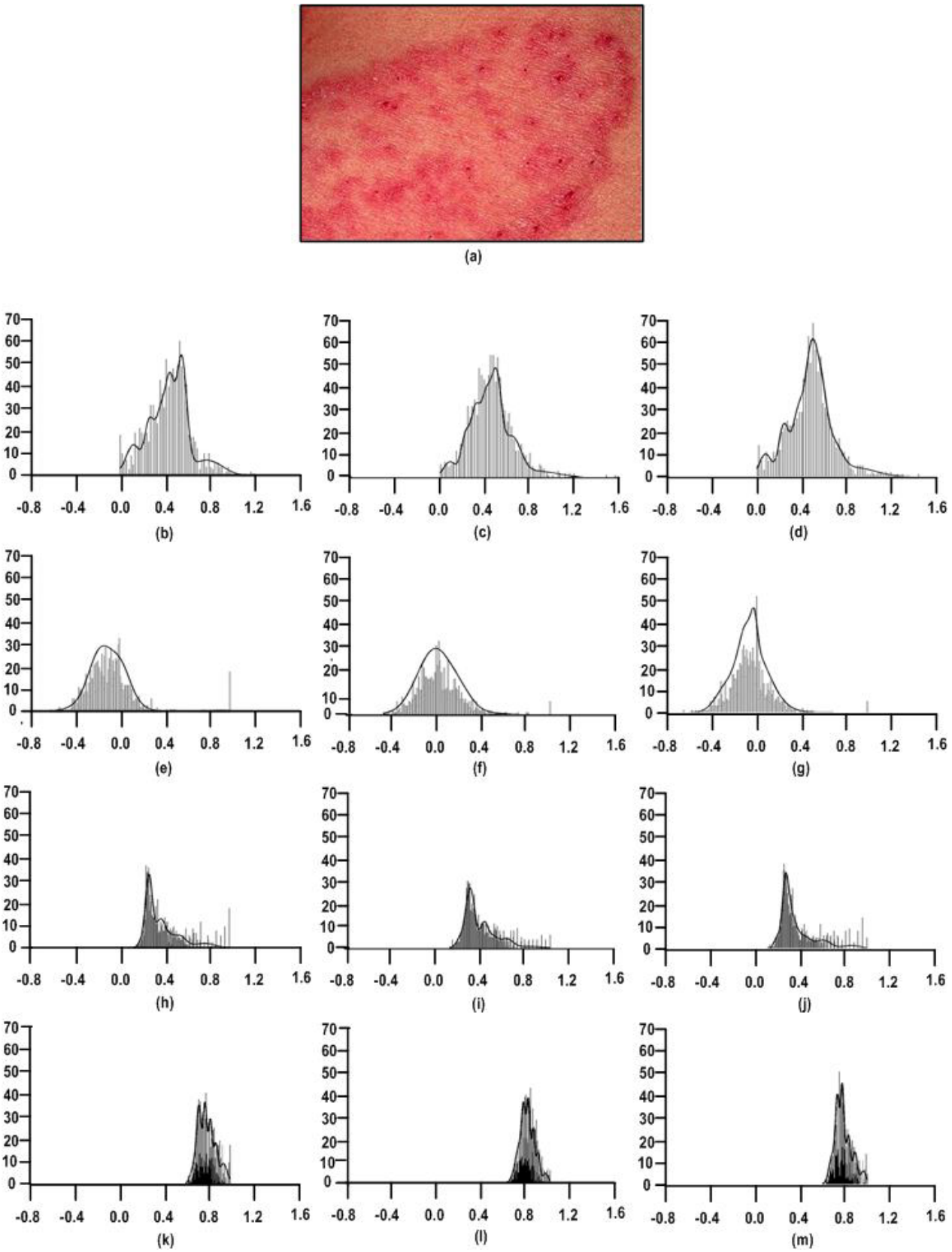

FIGURE 8. GMM distribution of the proposed framework for Tinea Corporis.

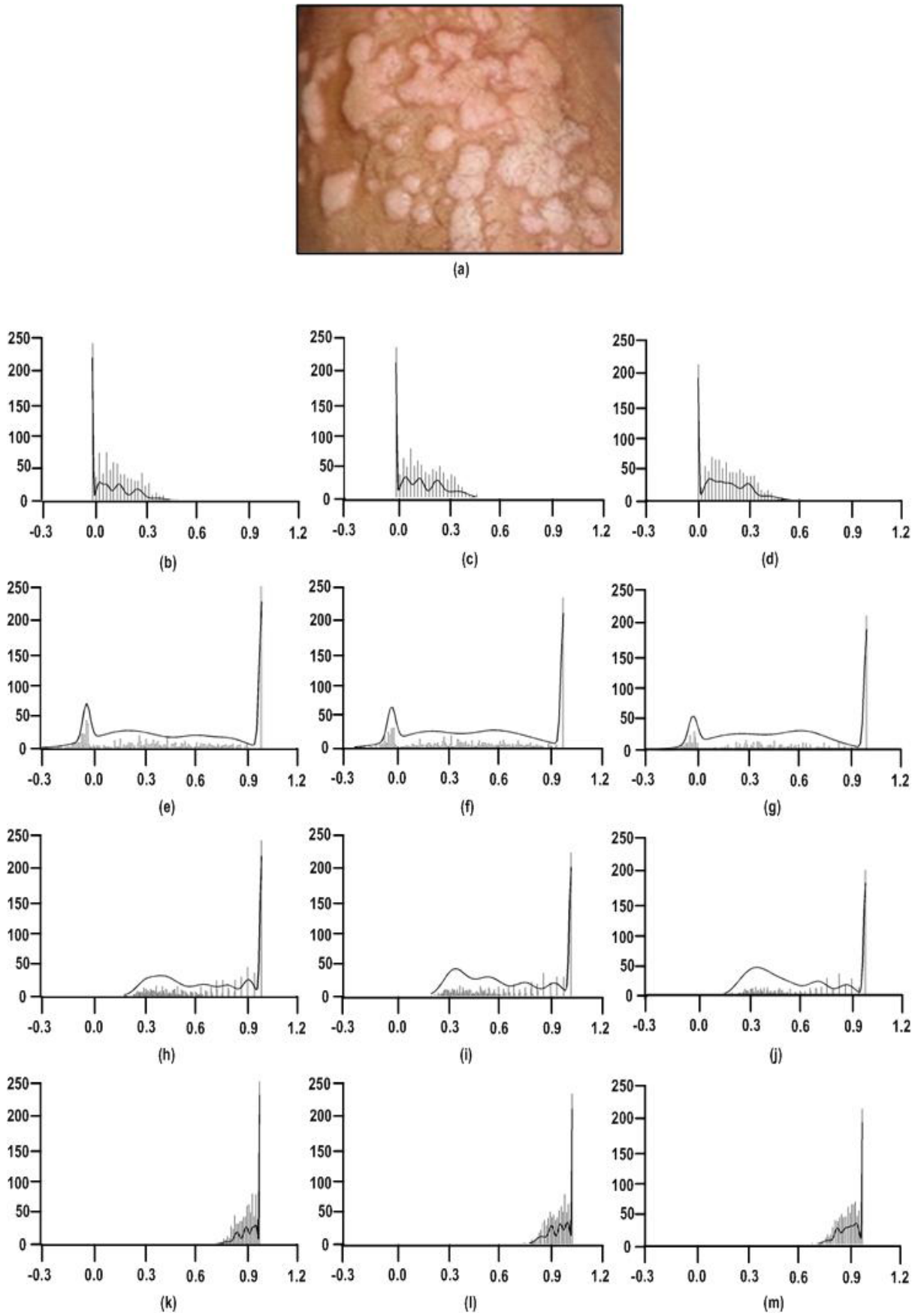

FIGURE 9. GMM distribution of the proposed framework for Warts.

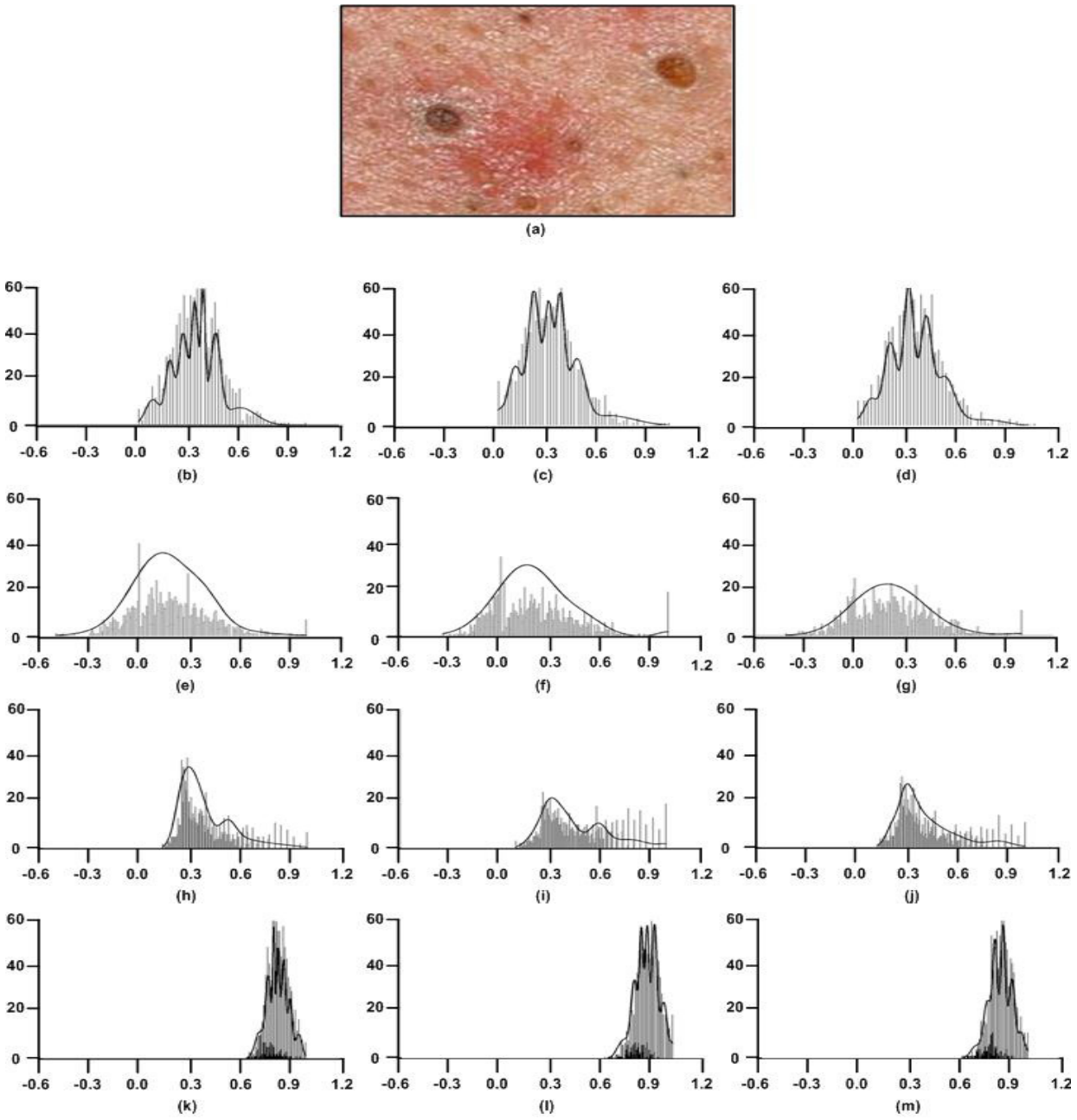

FIGURE 10. GMM distribution of the proposed framework for Acne-Blackhead.

Parameters are updated on the basis of the probabilities from the previous step using:

$$P_{ik}^{t} = \frac{w_k^t g(x_i|\mu_k^t, \sum_i^t)}{\sum_{k=1}^{K} w_k^t g(x_i|\mu_k^t, \sum_i^t)} \tag{3}$$

$$w_k^{t+1} = \frac{1}{N}\sum\nolimits_{i=1}^{N} P_{ik}^t \tag{4}$$

$$\mu_k^{t+1} = \frac{\sum_{i=1}^{N} P_{ik}^t x_i}{\sum_{i=1}^{N} P_{ik}^t} \tag{5}$$

$$\sum_{k}^{t+1} = \mu_k^{t+1} = \frac{\sum_{k=1}^{K} P_{ik}^t (x_i - \mu_k^t)(x_i - \mu_k^t)^T}{\sum_{i=1}^{N} P_{ik}^t} \tag{6}$$

Mostly, the clustering algorithms use Euclidean distance for classification assuming the data to be isotropically gaussian [59]. In multivariate modeling, the feature vectors don't satisfy this condition and, hence, the clustering leads to wrong classification. The solution to this problem may be the use of Mahalanobis distance (MD) defined by [60], [61]:

$$D_M(\vec{\mathbf{x}}) = \sqrt{(\vec{\mathbf{x}} - \vec{\boldsymbol{\mu}})^T \, \Sigma^{-1} (\vec{\mathbf{x}} - \vec{\boldsymbol{\mu}})} \tag{7}$$

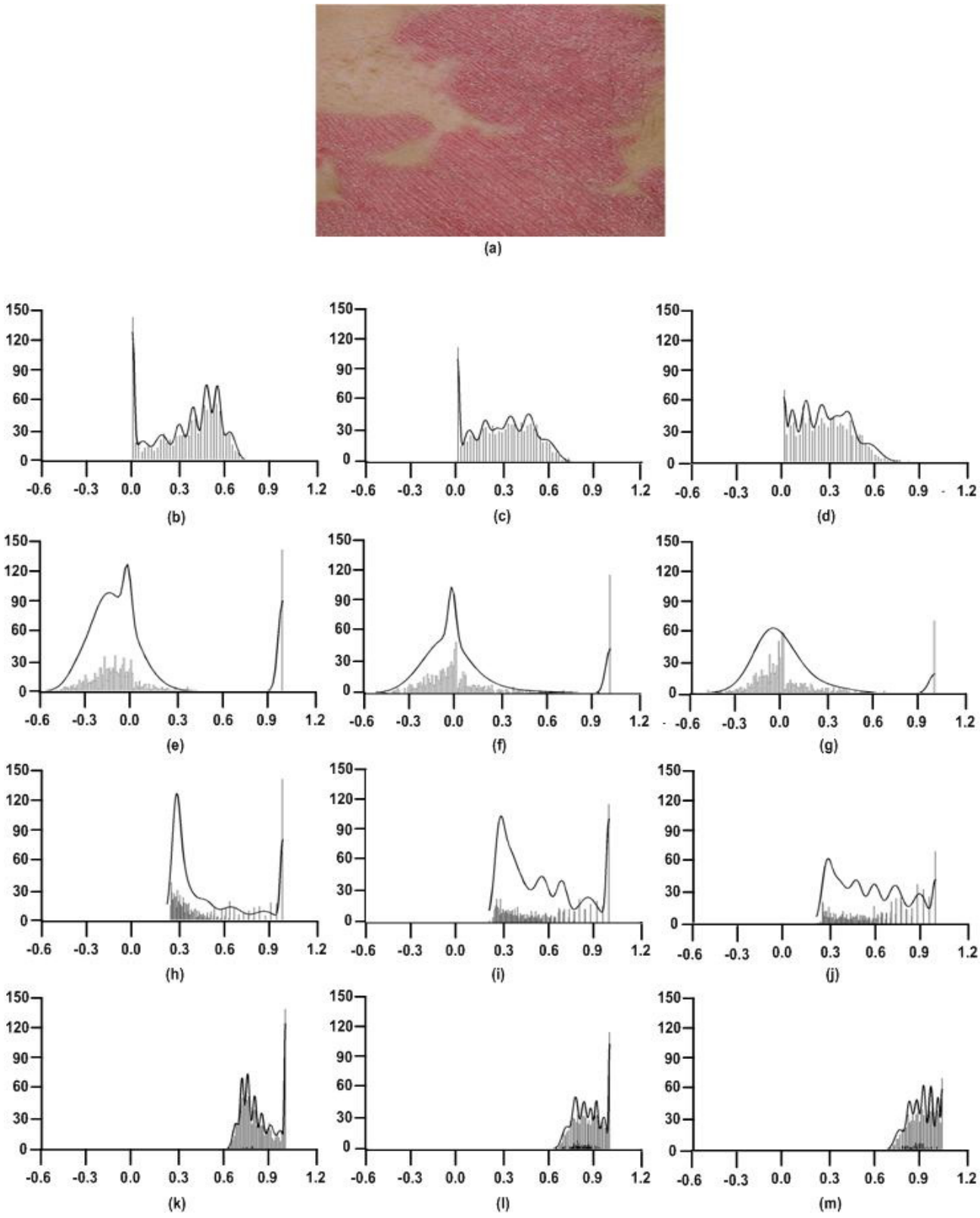

FIGURE 11. GMM distribution of the proposed framework for Psoriasis.

where, $\vec{\mathbf{x}} = (\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}, \mathbf{x_4}, \ldots\ldots, \mathbf{x_N})^T$ form a set of observations with mean $\vec{\boldsymbol{\mu}} = (\boldsymbol{\mu_1}, \boldsymbol{\mu}_2, \boldsymbol{\mu_3}, \boldsymbol{\mu_4}, \ldots\ldots, \boldsymbol{\mu_N})^T$ and covariance matrix $\boldsymbol{\Sigma}$. It is scale-invariant [61]. It is based on correlations between variables leading to efficient identification and analysis of different patterns available in the input feature vectors. MD measures the relative distance between

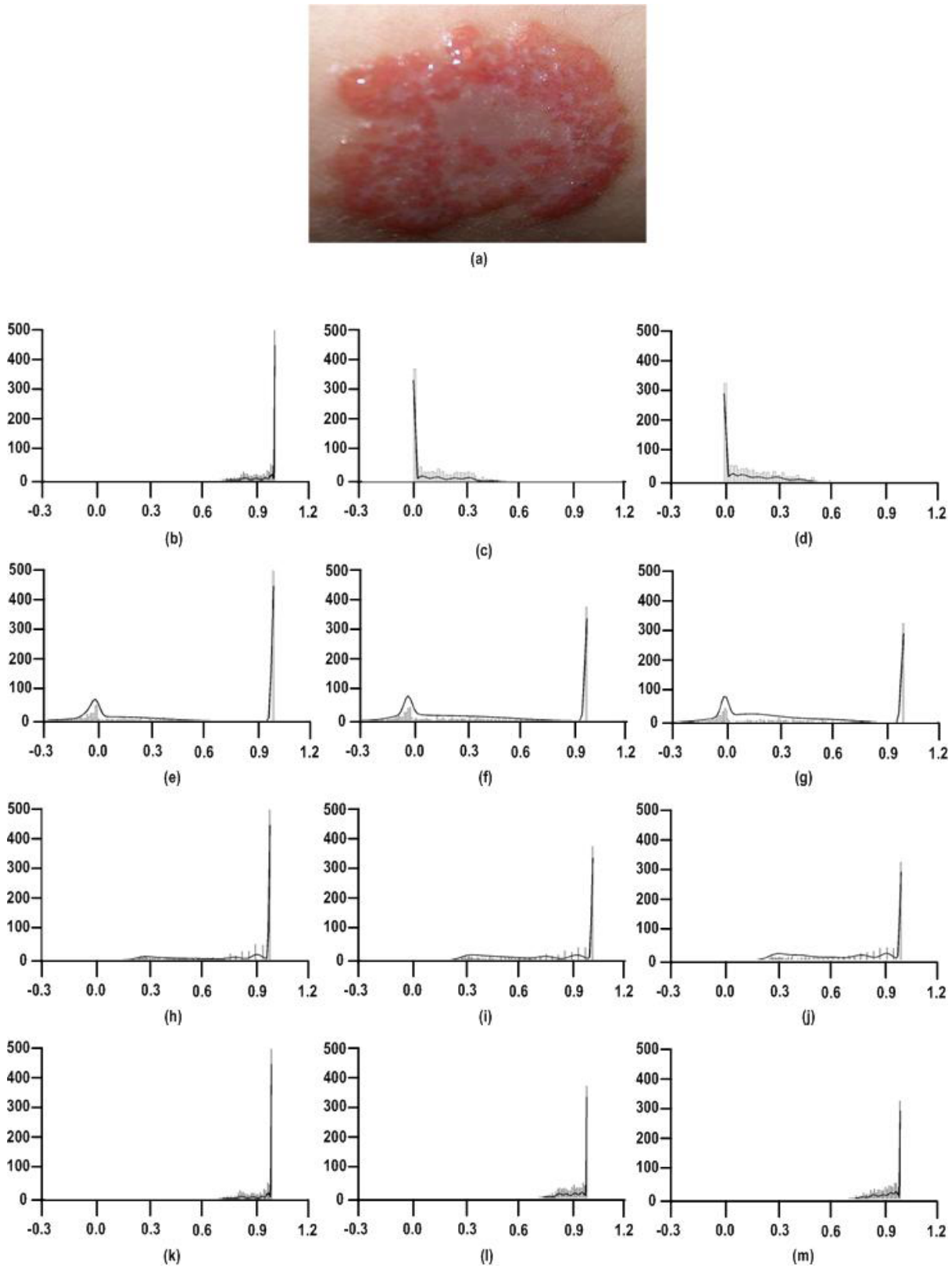

FIGURE 12. GMM distribution of the proposed framework for Discoid Eczema.

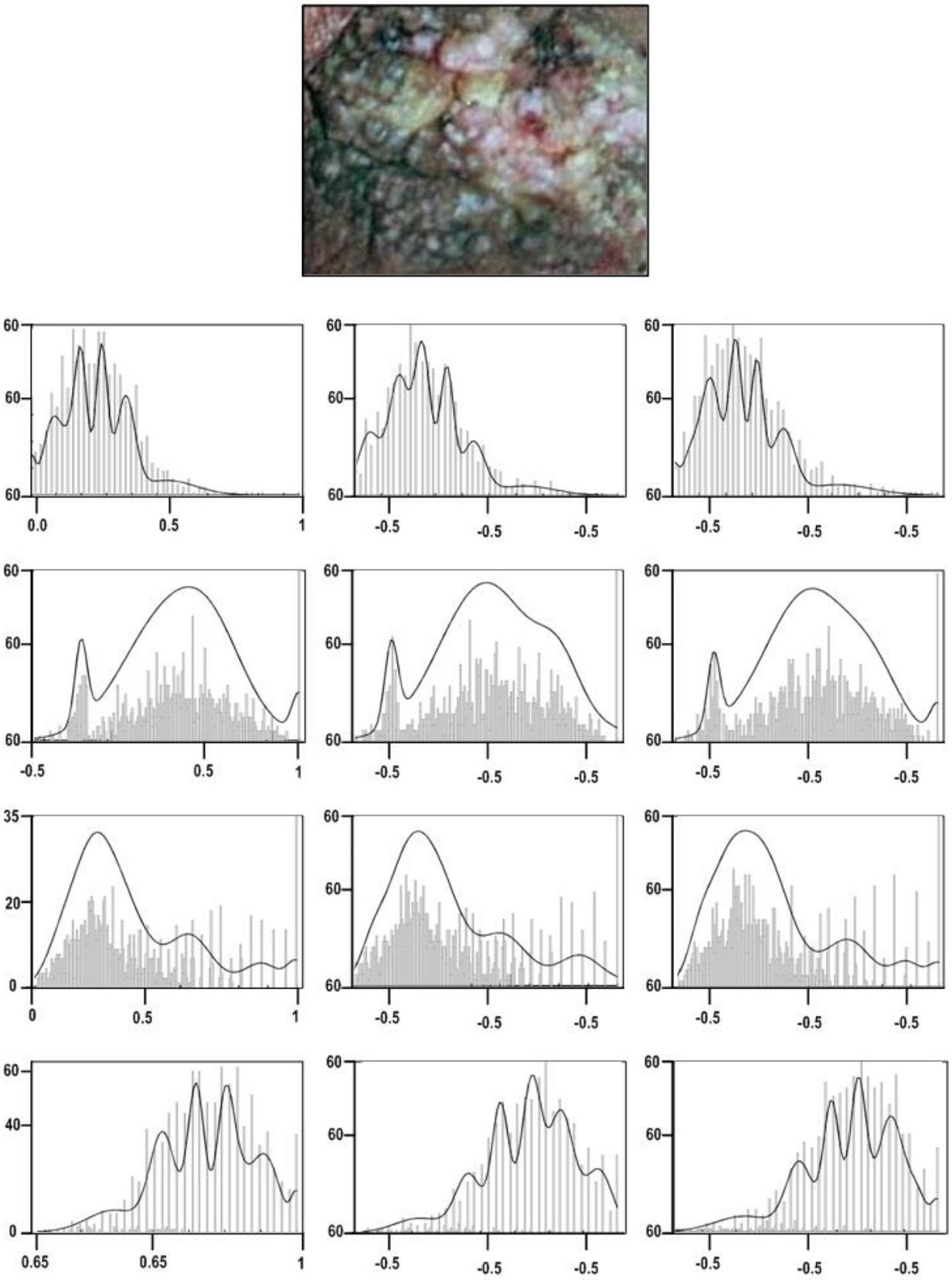

FIGURE 13. GMM distribution of the proposed framework for Chromoblastomycosis.

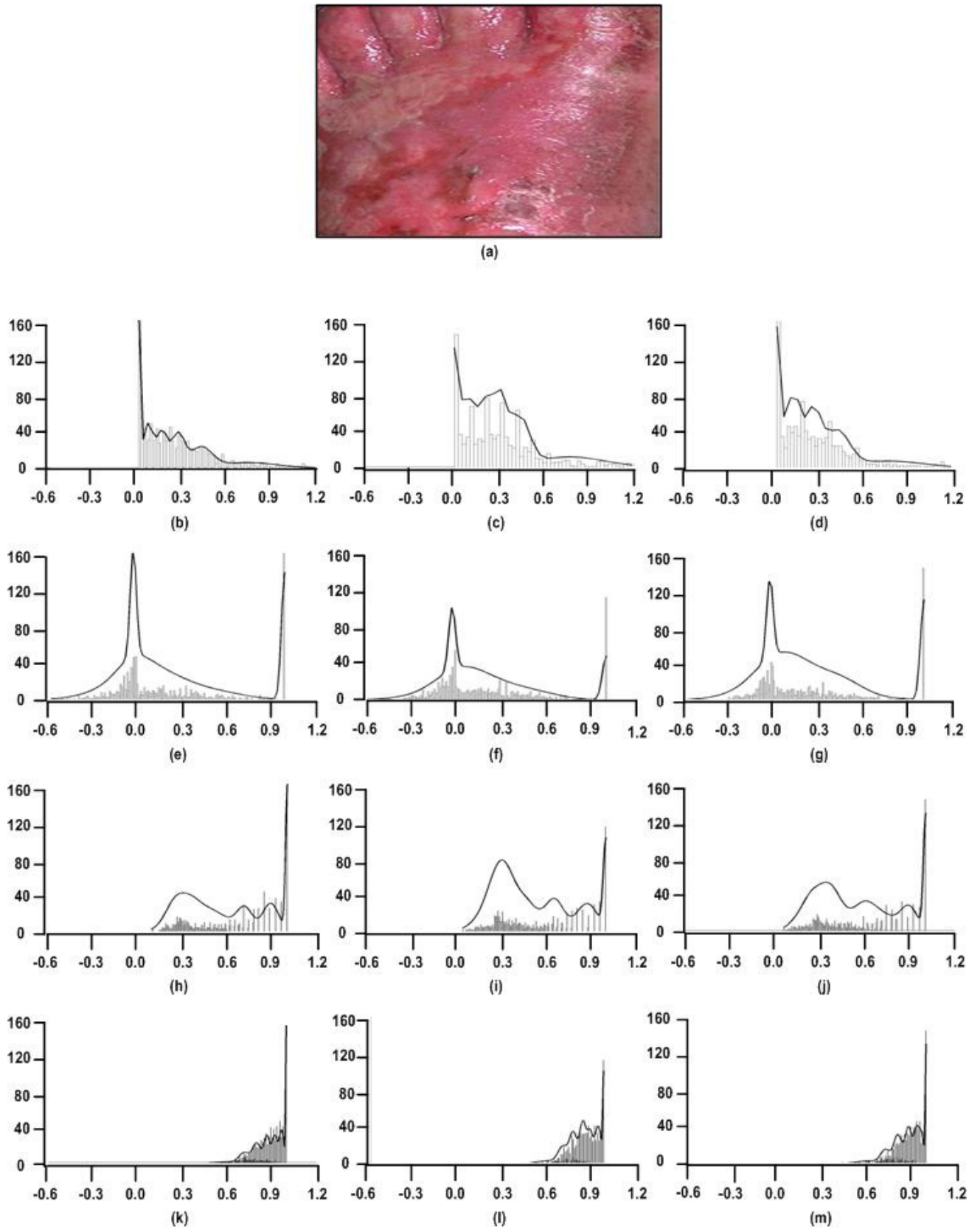

FIGURE 14. GMM distribution of the proposed framework for Athletes Foot.

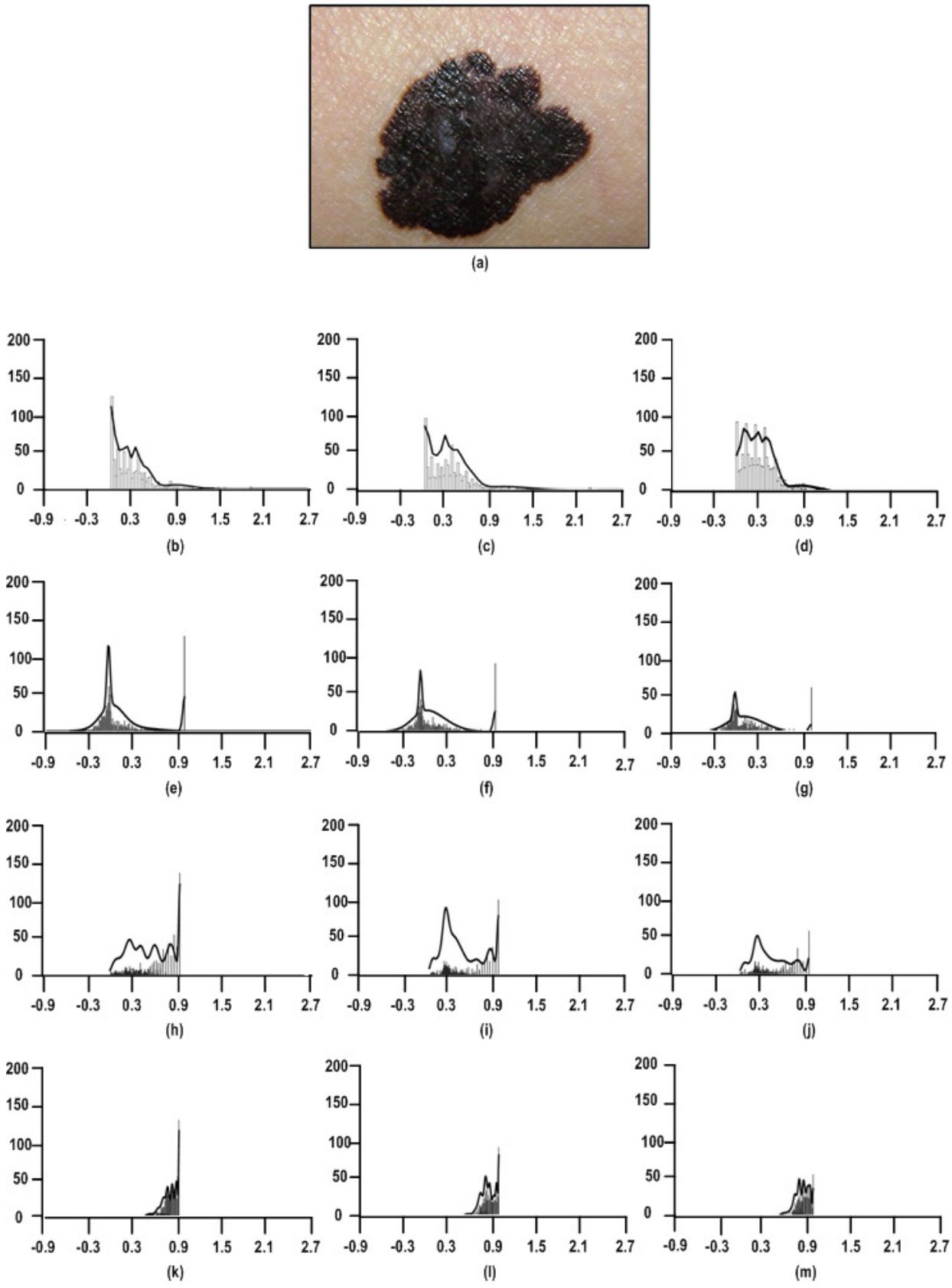

FIGURE 15. GMM distribution of the proposed framework for Melanoma.

two variables with respect to the centroid [62]. It is a data driven measure that can ease the distance distortion caused by a linear combination of the attributes [63].

III. RESULTS AND DISCUSSIONS

The distributions of contrast, correlation, energy and homogeneity of red, green, and blue components of the chosen dermatological diseases were modeled using GMM. About 100 iterations were needed for convergence of the GMM and its approximation of the feature vectors for each type of skin diseases provided 8 priors, 8×4 centers, and 8×4 co-variances for each RGB component, giving a total of 72 valued feature vectors. For classification of the diseases, Euclidean and Mahalanobis distances amongst the diseases were also estimated with respect to the normal skin. Fig. 4 shows the output of GMM modeling of these feature vectors for normal human skin, whereas, Fig. 5 to Fig. 15 represents the output for various dermatological diseases. Row I shows contrast, Row II correlation, Row III energy, and Row IV homogeneity for RGB components. In these figures, bars show the distribution of the actual data of GLCM features while the curve represents the GMM approximation. The four dimensional feature vector set for normal skin, various dermatological diseases and its variants were studied.

Analysis of the GMM modeled four dimensional feature vector set for red component of normal skin showed important peaks at 0.05, 0.47, 0.13 for contrast, at −0.02, −0.06, 0.00 for correlation, at 0.90, 0.31, 0.76 for energy, and at 0.97, 0.77, 0.93 for homogeneity. Green component showed critical peaks at 0.55, 0.50, 0.63 for contrast, at −0.01, −0.08, −0.21 for correlation, at 0.26, 0.33, 0.28 for energy, and at 0.75, 0.77, 0.72 for homogeneity. Blue component of normal skin represented significant peaks at 0.64, 0.77, 0.52 for contrast, at −0.06, −0.02, 0.00 for correlation, at 0.24, 0.20, 0.26 for energy, and at 0.73, 0.70, 0.76 for homogeneity.

Investigations of the GMM modeled four dimensional feature vector set for red component of Molluscum Contagiosum defined critical peaks at 0.22, 0.12, 0.31 for contrast, at 0.35, 0.19, 0.31 for correlation, at 0.48, 0.74, 0.35 for energy, and at 0.89, 0.94, 0.84 for homogeneity. Green component of this disease showed crucial peaks at 0.21, 0.30, 0.40 for contrast, at 0.44, 0.35, 0.28 for correlation, at 0.45, 0.35, 0.29 for energy, and at 0.89, 0.85, 0.80 for homogeneity. Similarly, blue component showed important peaks at 0.21, 0.28, 0.35 for contrast, at 0.36, 0.36, 0.35 for correlation, at 0.48, 0.38, 0.30 for energy, and at 0.89, 0.86, 0.83 for homogeneity.

While studying GMM modeled four dimensional feature vector set for red component of Milia important peaks were observed at 0.53, 0.37, 0.87 for contrast, at −0.05, 0.08, −0.18 for correlation, at 0.30, 0.38,0.23 for energy, and at 0.76, 0.82, 0.69 for homogeneity. Similarly, green component of this disease represented important peaks at 0.72, 0.34, 0.47 for contrast, at −0.12, 0.10,.06 for correlation, at 0.26, 0.41, 0.30 for energy, and at 0.72, 0.84, 0.78 for homogeneity. Also, blue component showed important peaks at 0.22, 0.61, 0.38 for contrast, at 0.15, 0.01, 0.26 for correlation, at 0.56, 0.25, 0.30 for energy, and at 0.89, 0.73, 0.81 for homogeneity.

GMM modeled four dimensional feature vector set for red component of Discoid Lupus Erythematosus showed considerable important peaks at 0.35, 0,48, 0,.25 for contrast, at 0.08, −0.03, 0.00 for correlation, at 0.39, 0.30, 0.54 for energy, and at 0.83, 0.77, 0.88 for homogeneity. Its green component showed critical peaks at 0.37, 0.49, 0.26 for contrast, at 0.17, 0.12, 0.06 for correlation, at 0.35, 0.26, 0.54 for energy, and at 0.82, 0.76, 0.87 for homogeneity. Also, blue component of this disease showed important peaks at 0.36, 0.46, 0.27 for contrast, at 0.19, 0.09, 0.24, for correlation, at 0.35, 0.29, 0.45 for energy, and at 0.82, 0.78, 0.87 for homogeneity.

Analysis of the GMM modeled feature vector set for red component of Tinea Corporis showed significant peaks at 0.58, 0.39, 0.81 for contrast, at −0.17, −0,03, −0.26 for correlation, at 0.28, 0.42, 0.25 for energy, and at 0.72, 0.82, 0.69 for homogeneity. For green component important peaks were shown at 0.51, 0.65, 0.32 for contrast, at 0.01, −0.16, 0.03 for correlation, at 0.28, 0.29, 0.48 for energy, and at 0.76,0.73, 0.85 for homogeneity. Similarly, blue component of this disease showed critical peaks at 0.48, 0.57, 0.37 for contrast, at −0.03, −0.11, −0.01 for correlation, at 0.31, 0.26, 0.43 for energy, and at 0.77, 0.73, 0.83 for homogeneity.

GMM analysis of four dimensional feature vector set for red component of Warts showed important peaks at 0.00, 0.27, 0.15 for contrast, at 1.00, 0.28, 0.20 for correlation, at 1.00, 0.42, 0.67 for energy, and at 1.00, 0.87, 0.92 for homogeneity. The green component of Warts showed important peaks at 0.00, 0.27, 0.15 for contrast, at 1.00, 0.28, 0.20 for correlation, at 1.00, 0.42, 0.67 for energy, and at 1.00, 0.87, 0.92 for homogeneity. Similarly for blue component of this disease essential peaks were observed at 0.00, 0.30, 0.18 for contrast, at 1.00, 0.35, 0.61 for correlation, at 1.00, 0.35, 0.41 for energy, and at 1.00, 0.85, 0.91 for homogeneity.

Similar investigations were carried out for Acne-Blackhead in which the red component of the GMM modeled feature vector set showed important peaks at 0.46, 0.34, 0.39 for contrast, at 0.09, 0.24, 0.14 for correlation, at 0.29, 0.35, 0.33 for energy, and at 0.78, 0.83, 0.81 for homogeneity. Its green component showed important peaks at 0.47, 0.30, 0.37 for contrast, at 0.10, 0.22, 0.20 for correlation, at 0.29, 0.41, 0.34 for energy, and at 0.78, 0.85, 0.82 for homogeneity. Lastly, blue component of this disease showed important peaks at 0.30, 0.40, 0.19 for contrast, at 0.27, 0.18, 0.31 for correlation, at 0.39, 0.31, 0.54 for energy, and at 0.85, 0.80, 0.90 for homogeneity.

Psoriasis showed significant peaks for its red component at 0.49, 0.57, 0.40 for contrast, at −0.12, −0.21, −0.06 for correlation, at 0.31, 0.29, 0.38 for energy, and at 0.75, 0.72, 0.80 for homogeneity. Important peaks for its green component were observed at 0.46, 0.35, 0.59 for contrast, at −0.07, 0.03, −0.19 for correlation, at 0.33, 0.42, 0.28 for energy, and at 0.78, 0.83, 0.72 for homogeneity. Also, blue component of it exhibited important peaks at 0.15, 0.43, 0.25 for contrast,

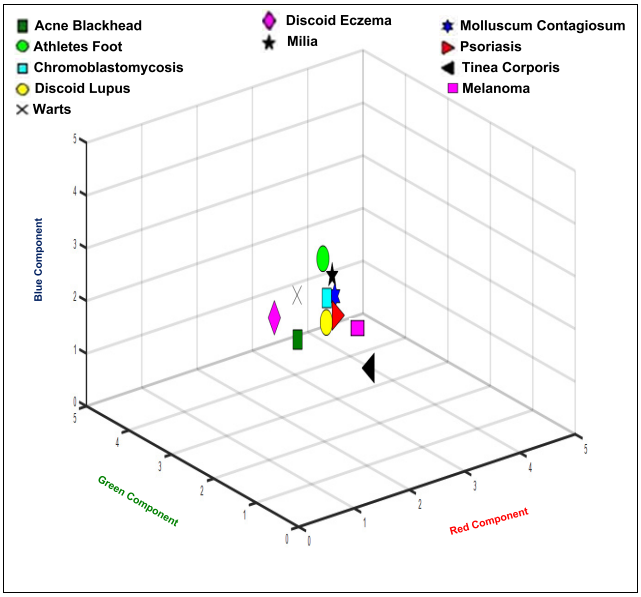

FIGURE 16. Three-dimensional scatter plot of Euclidean distance for 11 skin diseases taken for investigation.

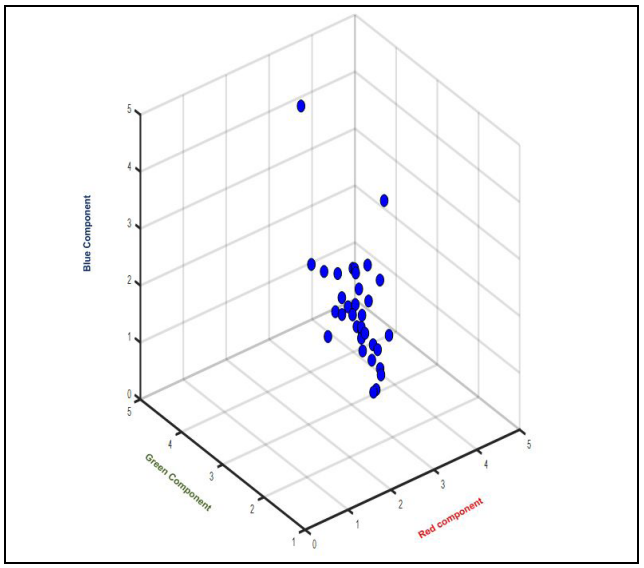

FIGURE 17. Three-dimensional scatter plot for various instances of Melanoma disease taken for investigation using Euclidean distance measure approach.

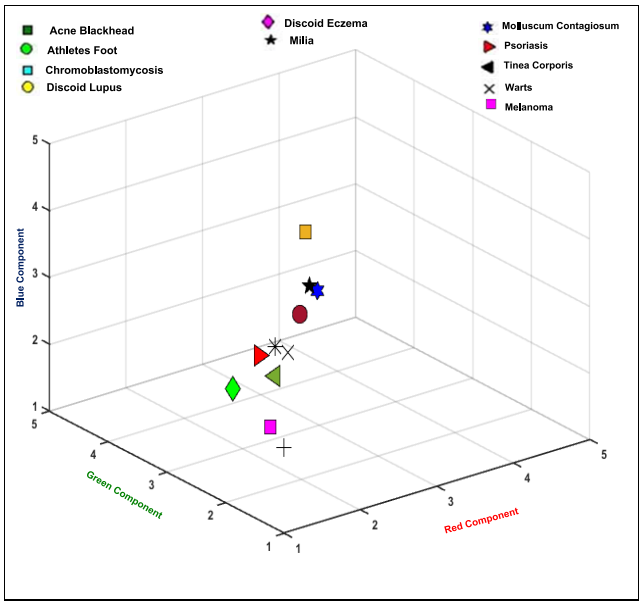

FIGURE 18. Three-dimensional scatter plot of Mahalanobis distance for 11 skin diseases taken for investigation.

at −0.01, −0.03, −0.06 for correlation, at 0.73, 0.35, 0.60 for energy, and at 0.93, 0.79, 0.88 for homogeneity.

Feature vector set for red component of Discoid Eczema produced important peaks at 0.00, 0.34, 0.04 for contrast, at 1.00, 0.14, −0.02 for correlation, at 1.00, 0.38, 0.93 for energy, and at 1.00, 0.83, 0.98 for homogeneity. The green component showed important peaks at 0.00, 0.32, 0.05 for contrast, at 1.00, 0.21, −0.02 for correlation, at 1.00, 0.37, 0.91 for energy, and at 1.00, 0.84, 0.98 for homogeneity. Similarly for blue component critical peaks were observed at 0.00, 0.30, 0.10 for contrast, at 1.00, 0.22, 0.17 for correlation, at 1.00, 0.40, 0.78 for energy, and at 1.00, 0.85, 0.95 for homogeneity.

Analysis of the GMM feature vector set for red component of Chromoblastomycosis exhibited important peaks at 0.20, 0.12, 0.27 for contrast, at 0.50, 0.51, 0.44 for correlation, at 0.44, 0.61, 0.35 for energy, and at 0.90, 0.94, 0.86 for homogeneity. The green component of this disease showed important peaks at 0.25, 0.34, 0.17 for contrast, at 0.48, 0.38, 0.29 for correlation, at 0.37, 0.31, 0.61 for energy, and at 0.87, 0.83, 0.91 for homogeneity. Similarly, its blue component displayed crucial peaks at 0.24, 0.32, 0.42 for contrast, at 0.46, 0.38, 0.39 for correlation, at 0.40, 0.33, 0.25 for energy, and at 0.88, 0.84, 0.79 for homogeneity.

Athletes Foot GMM based feature vector showed important peaks at 0.42, 0.00, 0.26 for contrast, at 0.09, 1.00, 0.18 for correlation, at 0.34, 1.00, 0.48 for energy, and at 0.80, 1.00, 0.87 for homogeneity. While analyzing its green component, important peaks were seen at 0.29, 0.44, 0.78 for contrast, at 0.24, 0.07, −0,.0.02 for correlation, at 0.42, 0.31, 0.24 for energy, and at 0.86, 0.79, 0.72 for homogeneity. For blue component of this disease, significant peaks were observed at 0.42, 0.13, 0.00 for contrast, at 0.13, 0.25, 1.00 for correlation, at 0.34, 0.69, 1.00 for energy, and at 0.81, 0.94, 1.00 for homogeneity.

Lastly, analysis of the GMM feature vector set for red component of Melanoma exhibited important peaks at 0.07, 0.19, 0.46 for contrast, at −0.03, 0.00, 0.05 for correlation, at 0.87, 0.66, 0.32 for energy, and at 0.97, 0.91, 0.79 for homogeneity. The green component of this disease showed important peaks at 0.40, 0.26, 0.57 for contrast, at 0.14, 0.16, −0.01for correlation, at 0.35, 0.49, 0.27 for energy, and at 0.82, 0.87, 0.75 for homogeneity. Similarly, its blue component displayed crucial peaks at −0.03, 0.53, 0.00 for

contrast, at 0.11, 0,.16, 0.18 for correlation, at 0.72, 1.00, 0.96 for energy, and at 0.00, 0.87, 0.71 for homogeneity.

In all the diseases and its variants, maximum wide peak is observed for correlation and minimum wide peak for energy in RGB components. Similar results were observed for normal skin also.

Mathematical and visual analysis of the GMM modeled feature vector of different diseases show that peak structure is disease depended and may be very useful for predicting the dermatological diseases from their visual images. For example, the Fig. 5 (Molluscum Contagiosum) shows different structure than that of Fig. 8 (Tinea Corporis). The comparisons of the classification based on Euclidean and Mahalanobis distance measures of different diseases with respect to normal skin using 72-dimensional feature vectors are visually shown in Fig. 16 to Fig. 19. It may be observed that different diseases occupy different spatial positions in their scatter plots. Mahalanobis based scatter plots (Fig. 17) show better results as dissimilar diseases get relatively more scattered as compared to that of Euclidean based scatter plots (Fig. 16). Further, instances of same disease (e.g. Melanoma in Fig. 19) give close grouping as compared to Euclidean based scatter plots in Fig. 18.

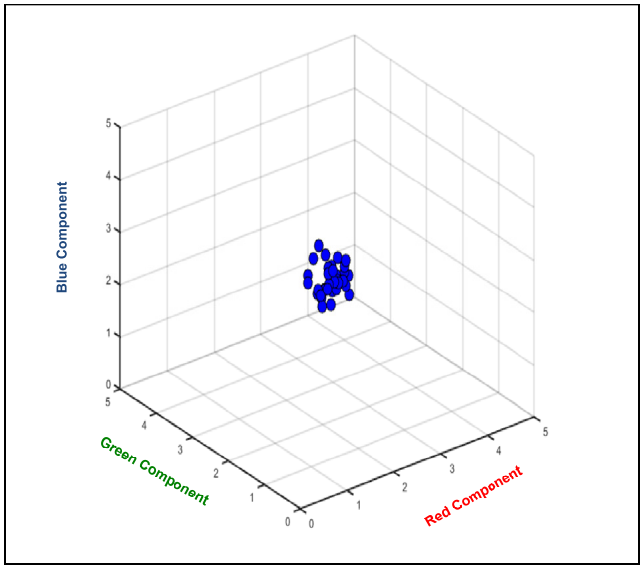

FIGURE 19. Three-dimensional scatter plot for various instances of Melanoma disease taken for investigation using Mahalanobis distance approach.

IV. CONCLUSION AND FURURE WORK

Investigations using GMM based modeling of GLCM parameters (contrast, correlation, energy and homogeneity) showed that different types of dermatological diseases have unique peak structure and, hence, they can be easily predicted only using their colored images. It was also observed that different diseases occupy distinct positions in Mahalanobis based classification. The extension of the work to other skin diseases on larger data sets is in our future plan.

NO CONFLICT STATEMENT

On behalf of all authors, the corresponding author states that there is no conflict of interest.

REFERENCES

[1] S. MacNeil, "Progress and opportunities for tissue-engineered skin," *Nature*, vol. 445, no. 7130, pp. 874–880, Feb. 2007.

[2] [Online]. Available: https://www.shutterstock.com/search/anatomy

[3] A. A. L. C. Amarathunga, E. P. W. C. Ellawala, G. N. Abeysekara, and C. R. J. Amalraj, "Expert system for diagnosis of skin diseases," *Int. J. Sci. Technol. Res.*, vol. 4, no. 01, pp. 174–178, Jan. 2015.

[4] S. Chakraborty, K. Mali, S. Chatterjee, S. Banerjee, K. G. Mazumdar, M. Debnath, P. Basu, S. Bose, and K. Roy, "Detection of skin disease using metaheuristic supported artificial neural networks," in *Proc. 8th Annu. Ind. Automat. Electromechan. Eng. Conf. (IEMECON)*, Aug. 2017, pp. 224–229.

[5] M. W. Tsang and J. S. Resneck, Jr., "Even patients with changing moles face long dermatology appointment wait-times: A study of simulated patient calls to dermatologists," *J. Amer. Acad. Dermatol.*, vol. 55, no. 1, pp. 54–58, Jul. 2006.

[6] T. Suneja, E. D. Smith, G. J. Chen, K. J. Zipperstein, A. B. Fleischer, Jr., and S. R. Feldman, "Waiting times to see a dermatologist are perceived as too long by dermatologists: Implications for the dermatology workforce," *Arch. Dermatol.*, vol. 137, no. 10, pp. 1303–1307, Oct. 2001.

[7] L. F. Mieras, A. T. Taal, E. B. Post, A. G. Z. Ndeve, and C. L. M. van Hees, "The development of a mobile application to support peripheral health workers to diagnose and treat people with skin diseases in resource-poor settings," *Tropical Med. Infectious Disease*, vol. 3, no. 3, p. 102, Sep. 2018.

[8] University of Colorado Denver. (Sep. 2013). *Study Shows Over 200 Mobile apps Related to Dermatology*. Accessed: Mar. 24, 2019. [Online]. Available: http:www.sciencedaily.com/releas es/2013/09/130925185603.htm

[9] H. K. Flaten, C. St Claire, E. Schlager, C. A. Dunnick, and R. P. Dellavalle, "Growth of mobile applications in dermatology—2017 update," *Dermatol. Online J.*, vol. 24, no. 2, pp. 1–5, Feb. 2017.

[10] A. C. Brewer, D. C. Endly, J. Henley, M. Amir, B. P. Sampson, J. F. Moreau, and R. P. Dellavalle, "Mobile applications in dermatology," *JAMA Dermatol.*, vol. 149, no. 11, pp. 1300–1304, Nov. 2013.

[11] M. Deveau and S. Chilukuri, "Mobile applications for dermatology," *Seminars Cutaneous Med. Surg.*, vol. 31, no. 3, pp. 174–182, Sep. 2012.

[12] P. N. Madu, A. Y. Chang, M. K. Kayembe, and C. L. Kovarik, "Teledermatology as a means to provide multispecialty care: A case of global specialty collaboration," *Pediatric Dermatol.*, vol. 34, no. 2, pp. e89–e92, Mar. 2017.

[13] R. Wootton, J. Craig, and V. Patterson, *Introduction to Telemedicine*, 2nd ed. RSM Books, 2006.

[14] J. P. van der Heijden, N. F. de Keizer, J. D. Bos, P. I. Spuls, and L. Witkamp, "Teledermatology applied following patient selection by general practitioners in daily practice improves efficiency and quality of care at lower cost," *Brit. J. Dermatol.*, vol. 165, no. 5, pp. 1058–1065, Nov. 2011.

[15] J. D. Whited, "Teledermatology research review," *Int. J. Dermatol.*, vol. 45, no. 3, pp. 220–229, Mar. 2006.

[16] A. M. Oakley, F. Reeves, J. Bennett, S. H. Holmes, and H. Wickham, "Diagnostic value of written referral and/or images for skin lesions," *J. Telemed. Telecare*, vol. 12, no. 3, pp. 151–158, Apr. 2006.

[17] R. B. Mallett, "Teledermatology in practice," *Clin. Exp. Dermatol.*, vol. 28, no. 4, pp. 356–359, Jul. 2003.

[18] S. Chatterjee, S. Sarkar, S. Hore, N. Dey, A. S. Ashour, and V. E. Balas, "Particle swarm optimization trained neural network for structural failure prediction of multistoried RC buildings," *Neural Comput. Appl.*, vol. 28, no. 8, pp. 2005–2016, Aug. 2017.

[19] N. Yadav, V. Kumar, and U. Shrivastava, "Skin diseases detection models using image processing: A survey," *Int. J. Comput. Appl.*, vol. 137, no. 12, pp. 34–39, Mar. 2016.

[20] [Online]. Available: https://www.medicaldevice-network.com/comment/ai-in-dermatology

[21] L. Oakden-Rayner, "Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists," *Ann. Oncol.*, vol. 30, no. 5, p. 854, May 2019.

[22] T. Celik and T. Tjahjadi, "Automatic image equalization and contrast enhancement using Gaussian mixture modeling," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 145–156, Jan. 2012.
[23] H. Ibrahim and N. S. P. Kong, "Image sharpening using sub-regions histogram equalization," *IEEE Trans. Consum. Electron.*, vol. 55, no. 2, pp. 891–895, May 2009.
[24] A. Mehta, A. S. Parihar, and N. Mehta, "Supervised classification of dermoscopic images using optimized fuzzy clustering based multi-layer feed-forward neural network," in *Proc. Int. Conf. Comput., Commun. Control (IC)*, Sep. 2015, pp. 1–6.
[25] G. Kaur and N. Aggarwal, "Organized classification of melanoma images using Gaussian mixture model and artificial neural network," *Imperial J. Interdiscipl. Res.*, vol. 3, no. 8, pp. 1–8 2017.
[26] N. S. Ramteke and S. V. Jain, "Analysis of skin cancer using fuzzy and wavelet technique—Review & proposed new algorithm," *Int. J. Eng. Trends Technol.*, vol. 4, no. 6, pp. 2555–2566, Jun. 2013.
[27] A. Pal, U. Garain, R. Chatterjee, and S. Senapati, "Psoriatic plaque segmentation in skin images," in *Proc. 5th Nat. Conf. Comput. Vis., Pattern Recognit., Image Process. Graph. (NCVPRIPG)*, Dec. 2015, pp. 1–4.
[28] N. Alamdari, K. Tavakolian, M. Alhashim, and R. Fazel- Rezai, "Detection and classification of acne lesions in acne patients: A mobile application," in *Proc. IEEE Int. Conf. Electro Inf. Technol. (EIT)*, May 2016, pp. 739–743.
[29] M. M. Yusof, R. A. Aziz, and C. S. Fei, "The development of online children skin diseases diagnosis system," *Int. J. Inf. Educ. Technol.*, vol. 3, no. 2, pp. 231–234, Apr. 2013.
[30] C. L. Aruta, C. R. Calaguas, J. K. Gameng, M. V. Prudentino, A. A. Chestel, and J. Lubaton, "Mobile-based medical assistance for diagnosing different types of skin diseases using case- based reasoning with image processing," *Int. J. Conceptions Comput. Inf. Technol.*, vol. 3, no. 3, pp. 115–118, Oct. 2015.
[31] R. Hassanpour, A. Shahbahrami, and S. Wong, "Adaptive Gaussian mixture model for skin color segmentation," *World Acad. Sci., Eng. Technol.*, vol. 41, pp. 1–6, Jul. 2008.
[32] C. Barata, M. A. T. Figueiredo, M. E. Celebi, and J. S. Marques, "Color identification in dermoscopy images using Gaussian mixture models," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 3611–3615.
[33] S. L. Varma and V. Behera, "Human skin detection using histogram processing and Gaussian mixture model based on color spaces," in *Proc. Int. Conf. Intell. Sustain. Syst. (ICISS)*, Dec. 2017, pp. 116–120.
[34] A. Amelio and C. Pizzuti, "Skin lesion image segmentation using a color genetic algorithm," in *Proc. 15th Annu. Conf. Companion Genetic Evol. Comput.*, Jul. 2013, pp. 1471–1478.
[35] M. Liu, D. Cheng, and W. Yan, "Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images," *Frontiers Neuroinform.*, vol. 12, p. 35, Jun. 2018.
[36] R. B. Patil and S. Ramakrishnan, "A Gaussian mixture model based diagnosis of Alzheimer's using diffusion tensor imaging," in *Proc. 39th Annu. Northeast Bioeng. Conf.*, Syracuse, NY, USA, Apr. 2013, pp. 137–138.
[37] I. Trabelsi, R. Amami, and N. Ellouze, "Automatic emotion recognition using generative and discriminative classifiers in the GMM mean space," in *Proc. 2nd Int. Conf. Adv. Technol. Signal Image Process. (ATSIP)*, Mar. 2016, pp. 767–770.
[38] Y. Huang, K. B. Englehart, B. Hudgins, and A. D. C. Chan, "A Gaussian mixture model based classification scheme for myoelectric control of powered upper limb prostheses," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 11, pp. 1801–1811, Nov. 2005.
[39] M. G. Forero, G. Cristóbal, and M. Desco, "Automatic identification of Mycobacterium tuberculosis by Gaussian mixture models," *J. Microsc.*, vol. 223, no. 2, pp. 120–132, Aug. 2006.
[40] D. Zangeneh and M. Yazdi, "Automatic segmentation of multiple sclerosis lesions in brain MRI using constrained GMM and genetic algorithm," in *Proc. 24th Iranian Conf. Elect. Eng. (ICEE)*, May 2016, pp. 832–837.
[41] M. İşcan, F. Yiğit, and C. Yilmaz, "Heartbeat pattern classification algorithm based on Gaussian mixture model," in *Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, May 2016, pp. 1–6.
[42] Y. Jiang and H. F. F. Leung, "Gaussian mixture model and Gaussian supervector for image classification," in *Proc. IEEE 23rd Int. Conf. Digit. Signal Process. (DSP)*, Nov. 2018, pp. 1–5.
[43] J. Molina-Mora and R. Mora-Rodriguez, "Identification of cancer chemosensitivity by ODE and GMM modeling of heterogeneous cellular response to perturbations in fluorescent sphingolipid metabolism," in *Proc. IEEE 36th Central Amer. Panama Conv. (CONCAPAN)*, Nov. 2016, pp. 1–6.
[44] R. R. Wildeboer, A. W. Postema, L. Demi, M. P. J. Kuenen, H. Wijkstra, and M. Mischi, "Multiparametric dynamic contrast-enhanced ultrasound classification of prostate cancer," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Sep. 2016, pp. 1–4.
[45] A. Poonsri and W. Chiracharit, "Fall detection using Gaussian mixture model and principle component analysis," in *Proc. 9th Int. Conf. Inf. Technol. Elect. Eng. (ICITEE)*, Oct. 2017, pp. 1–4.
[46] Y. Zhang, B. Du, L. Zhang, and S. Wang, "A low-rank and sparse matrix decomposition-based Mahalanobis distance method for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1376–1389, Mar. 2016.
[47] V. Kothari, I. Wei, S. Shankar, S. Kalyana-Sundaram, L. Wang, L. W. Ma, P. Vats, C. S. Grasso, D. R. Robinson, Y. M. Wu, X. Cao, D. M. Simeone, A. M. Chinnaiyan, and C. Kumar-Sinha, "Outlier kinase expression by RNA sequencing as targets for precision therapy," *Cancer Discovery*, vol. 3, no. 3, pp. 280–293, Mar. 2013.
[48] A. Ali, N. Al Hasan Haldar, F. A. Khan, and S. Ullah, "ECG arrhythmia classification using mahalanobis-taguchi system in a body area network environment," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–7.
[49] Y. Gu, J. Zhou, and B. Qian, "Melanoma detection based on Mahalanobis distance learning and constrained graph regularized nonnegative matrix factorization," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2017, pp. 797–805.
[50] S. Zhou, Q. Wang, Y. Fang, and Q. Liu, "An extreme learning machine method for multi-classification with mahalanobis distance," in *Proc. 2nd IEEE Adv. Inf. Manage., Commun., Electron. Automat. Control Conf. (IMCEC)*, May 2018.
[51] H. M. Mashaly, N. A. Masood, and A. S. Mohamed, "Classification of Papulo-squamous skin diseases using image analysis," *Skin Res. Technol.*, vol. 18, no. 1, pp. 36–44, Feb. 2011.
[52] M. Abdoli, M. Ghanbari, H. Sarikhani, and P. Brault, "Gaussian mixture model-based contrast enhancement," *IET Image Process.*, vol. 9, no. 7, pp. 569–577, Jul. 2015.
[53] S. Yeom, "Infrared image segmentation based on region of interest extraction with Gaussian mixture modeling," *Proc. SPIE*, vol. 10202, May 2017, Art. no. 102020C.
[54] [Online]. Available: https://www.dermnetnz.org/image-library/
[55] K. P. Truong and D. A. van Leeuwen, "Automatic discrimination between laughter and speech," *Speech Commun.*, vol. 49, no. 2, pp. 144–158, Feb. 2007.
[56] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264–323, Sep. 1999.
[57] J. S. Vitter, "External memory algorithms and data structures: Dealing with massive data," *ACM Comput. Surv.*, vol. 33, no. 2, pp. 209–271, Jun. 2001.
[58] R. Li, R. Perneczky, I. Yakushev, S. Foerster, A. Kurz, A. Drzezga, and A. Kramer, "Gaussian mixture models and model selection for [18F] fluorodeoxyglucose positron emission tomography classification in Alzheimer's disease," *PLoS ONE*, vol. 10, no. 4, Apr. 2015, Art. no. e0122731.
[59] A. S. Shirkhorshidi, S. Aghabozorgi, and T. Y. Wah, "A comparison study on similarity and dissimilarity measures in clustering continuous data," *PLoS ONE*, vol. 10, no. 12, Dec. 2015, Art. no. e0144059.
[60] G. Gan, C. Ma, and J. Wu, *Data Clustering: Theory, Algorithms, and Applications*, vol. 20. Philadelphia, PA, USA: SIAM, 2007.
[61] J. Abonyi and B. Feil, *Cluster Analysis for Data Mining and System Identification*, 22nd ed. Birkhäuser, 2007. doi: 10.1007/978-3-7643-7988-9.
[62] Q. Xie, Y. Zhang, H. Jia, and Y. Lu, "Research on Mahalanobis distance algorithm optimization based on openCL," in *Proc. IEEE Int. Conf. High Perform. Comput. Commun., IEEE 6th Int. Symp. Cyberspace Saf. Secur., IEEE 11th Int. Conf. Embedded Softw. Syst. (HPCC,CSS,ICESS)*, Aug. 2014, pp. 84–91.
[63] S. Nielsen, "Numerical ecology," in *Ecological Modelling*, vol. 132, no. 3, P. Legendre and L. Legendre, 2nd ed. Amsterdam, The Netherlands: Elsevier, Aug. 2000, pp. 303–304.

CHAAHAT GUPTA received the B.E. degree in CSE from MIET, Jammu, India, and the M.Tech. degree in CSE from the Chandigarh Engineering College, Landran. She is currently pursuing the Ph.D. degree with Shri Mata Vaishno Devi University, Katra, India. She has over five years of work experience as an Assistant Professor with MIET. She has research papers published in various journals of national and international repute and also has a patent filled to her credit. Her research interest areas include digital image processing and software engineering.

NAVEEN KUMAR GONDHI received the Ph.D. degree in computer science. He is currently an Assistant Professor with Shri Mata Vaishno Devi University, Katra, India. He has several research papers in national/international journals to his credit. His research areas include computer network management, expert systems, mobile computing, and cluster and grid computing.

PARVEEN KUMAR LEHANA received the master's degree from Kurukshetra University and the Ph.D. degree in signal processing from IIT Bombay. He is currently a Professor of electronics with the University of Jammu. He has a wide experience of teaching, research, and guiding M.Phil., M.Tech., and Ph.D. students. He has more than 200 research papers in national/international journals to his credit. Also, he has filled several patents and has authored several books. He has given hundreds of invited talks and is also a Life Member of several professional bodies, such as IETE, IAPT, and ISTE. He is currently the Honorable Chairman of IETE, Jammu and Kashmir.