Analysis and Identification of Dermatological Diseases Using Gaussian Mixture Modeling

Chaahat Gupta¹ (Student Member, IEEE), Naveen Kumar Gondhi² and Parveen Kumar Lehana³

¹ Department of Computer Science and Engineering, Shri Mata Vaishno Devi University & MIET Jammu, INDIA
² Department of Computer Science and Engineering, Shri Mata Vaishno Devi University, INDIA
³ Dept of Electronics, University of Jammu, INDIA

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

ABSTRACT Skin diseases are common and are mainly caused by virus, bacteria, fungus, or chemical disturbances. Timely analysis and identification is 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 lack of expertise in primary health centers. Hence, there is a need of providing self-assisting and innovative measures for appropriate diagnosis of the skin diseases. Use of mobile applications may provide an inexpensive, simple, and efficient solutions for early diagnosis and treatment. This paper investigates the application of Gaussian mixture model (GMM) based for the analysis and classification of skin diseases from their visual images using Mahalanobis distance measure. GMM has been preferred over 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 has been used as 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 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]. Dermal 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 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
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. Nicholas et al. [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, naïve 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 multidimensional 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 analysed from sparse matrix decomposition. MD has also been used in Ribonucleic acid (RNA) sequencing to analyse 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, E, and H) using 72 weights, 8 priors, 8x4 centers and 8x4 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.

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 \(Y_i\) and \(Y_j\) 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:

\[
P(\mathbf{x} | \lambda) = \sum_{k=1}^{K} w_k g(\mathbf{x} | \mathbf{\mu}_k, \Sigma_k)\]

(1)

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

\[
g(\mathbf{x} | \mathbf{\mu}, \Sigma) = \frac{1}{D} \exp\left(-\frac{1}{2} (\mathbf{x} - \mathbf{\mu})^T \Sigma^{-1} (\mathbf{x} - \mathbf{\mu})\right)\]

\[
(2\pi)^{\frac{D}{2}}
\]

Clustering can be improved by using GMM algorithm which can be used to estimate GMM parameters, i.e. mean \((\mathbf{\mu}_k)\), weight \((w_k)\), and covariance \((\Sigma_k)\) [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]. Parameters are updated on the basis of the probabilities from the previous step using:

\[
P_{ik}^t = \frac{w_k^t g(x_i | \mu_k^t, \Sigma_k^t)}{\sum_{k=1}^{K} w_k^t g(x_i | \mu_k^t, \Sigma_k^t)}\]

(3)

\[
w_{k}^{t+1} = \frac{1}{N} \sum_{i=1}^{N} P_{ik}^{t}\]

(4)

\[
\mu_{k}^{t+1} = \frac{\sum_{i=1}^{N} P_{ik}^{t} x_i}{\sum_{i=1}^{N} P_{ik}^{t}}\]

(5)
Mostly, the clustering algorithms use Euclidean distance for classification assuming the data to be isotropically gaussian [59]. In multivariate modeling, the feature vectors do not 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]:

\[
D_M(\bar{x}) = \sqrt{(\bar{x} - \bar{\mu})^T \Sigma^{-1}(\bar{x} - \bar{\mu})}
\]

(7)

where, \(\bar{x} = (x_1, x_2, x_3, x_4, \ldots, x_N)^T\) form a set of observations with mean \(\bar{\mu} = (\mu_1, \mu_2, \mu_3, \mu_4, \ldots, \mu_N)^T\) and covariance matrix \(\Sigma\). 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 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 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, 0.06 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 Discolor Erythematous 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 0.00, 0.28, 0.20 for correlation, at 1.00, 0.42, 0.67 for energy, and at 0.00, 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, 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 criticalt 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, -0.02 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, -0.01 for 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.

IV. CONCLUSION AND FUTURE 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] https://www.shutterstock.com/search/anatomy.
[3] A.A. Amaratunga, E.P. Ellawala, G.N. Abeysekara, and C.R. Amalraj, “Expert system for diagnosis of skin diseases,” Journal of the American Academy of Dermatology Research, 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,” 2017 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON), Aug. 2017.
[5] M. W. Tsang and J. S. Resnek, “Even patients with changing moles face long dermatology appointment wait-times: A study of simulated patient calls to dermatologists,” Journal of the American Academy of Dermatology, 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,” Archives of Dermatology, vol. 137, no. 10, Oct. 2001.
[7] L. Mieras, A. Taal, E. Post, A. Ndeve, and C. van Hees, “The Development of a Mobile Application to Support Peripherial Health Workers to Diagnose and Treat People with Skin Diseases in Resource-Poor Settings,” Tropical Medicine and Infectious Disease, vol. 3, no. 3, p. 102, Sep. 2018.
[8] University of Colorado Denver, “Study shows over 200 mobile apps related to dermatology,” Science Daily, Sep. 2013.[Online]. Available: http://www.sciencedaily.com/releases/2013/09/130925185603.htm [Accessed March 24, 2019].
[9] H.K., Flaten, C.C. St, E. Schlager, C.A. Dunnick, and R.P. Dellavalle, “Growth of mobile applications in dermatology-2017 update,” Dermatology online journal, vo.24, no.2, 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 Dermatology, vol. 149, no. 11, p. 1300, Nov. 2013.
[11] M. Deveau and S. Chilukuri, “Mobile Applications for Dermatology,” Seminars in Cutaneous Medicine and Surgery, 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 Speciality Collaboration,” Pediatric Dermatology, vol. 34, no. 2, pp. e89–e92, Mar. 2017.
[13] “Introduction to Telemedicine, second edition,” Apr. 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,” British Journal of Dermatology, vol. 165, no. 5, pp. 1058–1065, Oct. 2011.
[15] J. D. White, “Teledermatology research review,” International Journal of Dermatology, vol. 45, no. 3, pp. 220–229, Mar. 2006.
[16] A. M. M. Oakley, F. Reeves, J. Bennett, S. H. Holmes, and H. Wickham, “Diagnostic value of written referral and/or images for skin lesions,” Journal of Telemedicine and Telecare, vol. 12, no. 3, pp. 151–158, Apr. 2006.
[17] R. B. Mallett, “Teledermatology in practice,” Clinical and Experimental Dermatology, 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 multistory RC buildings,” Neural Computing and Applications, vol. 28, no. 8, pp. 2005–2016, Jan. 2016.
[19] N. Yadav, V. Kumar, and U. Shrivastava, “Skin Diseases Detection Models using Image Processing: A Survey,” International Journal of Computer Applications, vol. 137, no. 12, pp. 34–39, Mar. 2016.
[20] https://www.medicaldevice-network.com/comment/ai-dermatology.
[21] L. Oakden-Rayner, “Reply to ‘Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists’ by H. A. Haenssle et al.,” Annals of Oncology, Dec. 2018.
[22] T. Celik and T. Tjahjadi, “Automatic Image Equalization and Contrast Enhancement Using Gaussian Mixture Modeling,” IEEE Transactions on Image Processing, vol. 21, no. 1, pp. 145–156, Jan. 2012.
[23] H. Ibrahim and N. Pik Kong, “Image sharpening using sub-regions histogram equalization,” IEEE Transactions on Consumer Electronics, 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,” 2015 International Conference on Computer, Communication and Control (IC4), Sep. 2015.
[25] G. Kaur, and N. Aggarwal, “Organized Classification of Melanoma Images Using Gaussian Mixture Model and Artificial Neural Network,” Imperial Journal of Interdisciplinary Research, vol. 3, no. 8, 2017.
[26] S. N. Ramteke, and S.V. Jain, “Analysis of Skin Cancer Using Fuzzy and Wavelet Technique –Review & Proposed New Algorithm,” International Journal of Engineering Trends and Technology (IJETT), vol. 4, no. 6, 2013.

[27] A. Pal, U. Garain, R. Chatterjee, and S. Senapati, “Psoriatic plaque segmentation in skin images,” 2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIGP), Dec. 2015.

[28] N. Alamdari, K. Tavakolian, M. Alhashim, and R. Fazel-Rezai, “Detection and classification of acne lesions in acne patients: A mobile application,” 2016 IEEE International Conference on Electro Information Technology (EIT), May 2016.

[29] M. M. Yusof, R. A. Aziz, and C. S. Fei, “The Development of Online Children Skin Diseases Diagnosis System,” International Journal of Information and Education Technology, pp. 231–234, 2013.

[30] Carl Louie Aruta, Collin Razen Calaguas, Jan Kyss Gameng, Marc Vennos Prudentino, August Anthony Chestel J. Lubaton, “Mobile-based Medical Assistance for Diagnosing Different Types of Skin Diseases Using Case-based Reasoning with Image Processing,” International Journal of Conceptions on Computing and Information Technology Vol. 3, Issue. 3, October’ 2015; ISSN: 2345 – 9808.

[31] R. Hassanpour, A. Shahbahrami, and S. Wong, “Adaptive Gaussian mixture model for skin color segmentation,” World Academy of Science, Engineering and Technology, vol. 41, pp.1-6, 2008.

[32] C. Barata, M. A. T. Figueiredo, M. E. Celebi, and J. S. Marques, “Color identification in dermoscopy images using Gaussian mixture models,” 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2014.

[33] S. L. Varma and V. Behera, “Human skin detection using histogram processing and Gaussian Mixture Model based on color spaces,” 2017 International Conference on Intelligent Sustainable Systems (ICISS), Dec. 2017.

[34] A. Amelio and C. Pizzuti, “Skin lesion image segmentation using a color genetic algorithm,” Proceeding of the fifteenth annual conference companion on Genetic and evolutionary computation conference companion - GECCO ’13 Companion, 2013.

[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 in Neuroinformatics, vol. 12, Jun. 2018.

[36] R. B. Patil and S. Ramakrishnan, “A Gaussian Mixture Model Based Diagnosis of Alzheimer’s Using Diffusion Tensor Imaging,” 2013 39th Annual Northeast Bioengineering Conference, Syracuse, NY, 2013, pp. 137-138.

[37] J. Trabelsi, R. Amami, and N. Ellouze, “Automatic emotion recognition using generative and discriminative classifiers in the GMM mean space,” 2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Mar. 2016.

[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 Transactions on Biomedical Engineering, vol. 52, no. 11, pp. 1801–1811, Nov. 2005.

[39] M. G. FORERO, G. CRISTOBAL, and M. DESCO, “Automatic identification of Mycobacterium tuberculosis by Gaussian mixture models,” Journal of Microscopy, vol. 223, no. 2, pp. 120–132, Aug. 2006.

[40] D. Zangenhein and M. Yazdi, “Automatic segmentation of multiple sclerosis lesions in brain MRI using constrained GMM and genetic algorithm,” 2016 24th Iranian Conference on Electrical Engineering (ICEE), May 2016.

[41] M. Iscan, F. Yigit, and C. Yilmaz, “Heartbeat pattern classification algorithm based on Gaussian mixture model,” 2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA), May 2016.

[42] Y. Jiang and H. F. Frank Leung, “Gaussian Mixture Model and Gaussian Supervisor for Image Classification,” 2018 IEEE 23rd International Conference on Digital Signal Processing (DSP), Nov. 2018.

[43] J. Molina-Mora and R. Mora-Rodriguez, “Identification of cancer chemo sensitivity by ODE and GMM modeling of heterogeneous cellular response to perturbations in fluorescent sphingolipid metabolism,” 2016 IEEE 36th Central American and Panama Convention (CONCAPAN XXXVI), Nov. 2016.

[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,” 2016 IEEE International Ultrasonics Symposium (IUS), Sep. 2016.

[45] A. Poonsri and W. Chiracharit, “Fall detection using Gaussian mixture model and principle component analysis,” 2017 9th International Conference on Information Technology and Electrical Engineering (ICITEE), Oct. 2017.

[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 Transactions on Geoscience and Remote Sensing, 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, Feb. 2013.

[48] A. Ali, N. A. H. Haldar, F. A. Khan, and S. Ullah, “ECG Arrhythmia Classification Using Mahalanobis-Taguchi System in a Body Area Network Environment,” 2015 IEEE Global Communications Conference (GLOBECOM), Dec. 2015.

[49] Y. Gu, J. Zhou, and B. Qian, “Melanoma Detection Based on Mahalanobis Distance Learning and Constrained Graph Regularized Nonnegative Matrix Factorization,” 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), Mar. 2017.

[50] S. Zhou, Q. Wang, Y. Fang, and Q. Liu, “An Extreme Learning Machine Method for Multi-Classification with Mahalanobis Distance,” 2018 2nd IEEE Advanced Information Management,Communicates,Electronic and Automation Control Conference (IMCEC), May 2018.

[51] H. M. Mashaly, N. A. Masood, and A. S. A. Mohamed, “Classification of Papulo-Squamous skin diseases using image analysis,” Skin Research and Technology, 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 Processing*, 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,” *Automatic Target Recognition XXVII*, May 2017.

[54] https://www.dermnetnz.org/image-library/.

[55] K. P. Truong and D. A. van Leeuwen, “Automatic discrimination between laughter and speech,” *Speech Communication*, 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 Computing Surveys*, vol. 31, no. 3, pp. 264–323, Sep. 1999.

[57] J. S. Vitter, “External memory algorithms and data structures: dealing with massive data,” *ACM Computing Surveys*, vol. 33, no. 2, pp. 209–271, Jun. 2001.

[58] R. Li, R. Pernecky, I. Yakushev, S. Förster, A. Kurz, A. Drzezga, and S. Kramer, “Gaussian Mixture Models and Model Selection for [18F] Fluorodeoxyglucose Positron Emission Tomography Classification in Alzheimer’s Disease,” *PLOS ONE*, vol. 10, no. 4, p. e0122731, Apr. 2015.

[59] A. 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, p. e0144059, Dec. 2015.

[60] G. Gan, C. Ma, and J. Wu, “Data Clustering: Theory, Algorithms, and Applications,” Jan. 2007.

[61] “Cluster Analysis for Data Mining and System Identification,” 2007. Available from: http://dx.doi.org/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,” 2014 IEEE Intl Conf on High Performance Computing and Communications, 2014 IEEE 6th Intl Symp on Cyberspace Safety and Security, 2014 IEEE 11th Intl Conf on Embedded Software and Syst (HPCC, CSS, ICCESS), Aug. 2014.

[63] S. . Nielsen, “Numerical Ecology. Legendre P. and Legendre L., second ed. Elsevier, Amsterdam, p. 853, 1998.,” *Ecological Modelling*, vol. 132, no. 3, pp. 303–304, Aug. 2000.

**FIRST AUTHOR**, has over 5 years of work experience working as an Assistant Professor in MIET, Jammu, India. She did her B.E. in CSE from MIET, Jammu and her M.Tech in (CSE) from Chandigarh Engineering College, Landran. She is pursuing her Ph. D. from Shri Mata Vaishno Devi University, Katra, India. Her research interest areas are Digital Image Processing and Software Engineering. She has to her credit, research papers published in various journals of national and international repute and also has a patent filled. Also, she is a student member of IEEE. Corresponding email id is chaahatgupta249@gmail.com

**SECOND AUTHOR**, did his Ph.D in Computer Sciences. He is working as an Assistant Professor in SMVDU, Katra, India. His research areas are Computer Network Management, Expert System, Mobile Computing, Cluster & Grid Computing. He has several research papers in national/ international journals to his credit. Corresponding email id is naveen.gondhi@smvdu.ac.in.

**THIRD AUTHOR**, did his post-graduation from Kurukshetra University and Ph.D from IIT Bombay in Signal Processing. Presently, he is working as Professor of Electronics in University of Jammu. He has a wide experience of teaching, research, and guiding M.Phil., M.Tech., and PhD students. He has more than 200 research papers in national/ international journals to his credit. Also, he has filled several patents and authored several books. He has given hundreds of invited talks and is also a life member of several professional bodies like IETE, IAPT, and ISTE. Presently, he is Hon Chairman of IETE, J&K. Corresponding email id is pklehana@gmail.com.
Fig. 1. Layers of human skin [1]-[2]

Fig. 2. Multivariate Gaussian densities over two variables $Y_1$ and $Y_2$
Fig. 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.
Fig. 4. GMM distribution of the proposed framework for Normal skin
Fig. 5. GMM distribution of the proposed framework for Molluscum Contagiosum
Fig. 6. GMM distribution of the proposed framework for Milia
Fig. 7. GMM distribution of the proposed framework for Discoid Lupus Erythematosus
Fig. 8. GMM distribution of the proposed framework for Tinea Corporis
Fig. 9. GMM distribution of the proposed framework for Warts
Fig. 10. GMM distribution of the proposed framework for Acne-Blackhead
Fig. 11. GMM distribution of the proposed framework for Psoriasis
Fig. 12. GMM distribution of the proposed framework for Discoid Eczema
Fig. 13. GMM distribution of the proposed framework for Chromoblastomycosis
Fig. 14. GMM distribution of the proposed framework for Athletes Foot
Fig. 15. GMM distribution of the proposed framework for Melanoma
Fig. 16. Three-dimensional scatter plot of Euclidean distance for 11 skin diseases taken for investigation.
Fig. 17. Three-dimensional scatter plot for various instances of Melanoma disease taken for investigation using Euclidean distance measure approach
Fig. 18. Three-dimensional scatter plot of Mahalanobis distance for 11 skin diseases taken for investigation.
Fig. 19. Three-dimensional scatter plot for various instances of Melanoma disease taken for investigation using Mahalanobis distance approach