International Journal of Computer Vision and Signal Processing, 14(1), 43-53(2024) ORIGINAL ARTICLE

In-Depth Analysis of Automated Acne Disease Recognition and Classification

Afsana Ahsan Jeny*, Masum Shah Junayed

Department of Computer Science & Engineering, University of Connecticut, Storrs, CT 06269, USA

Md Robel Mia

Department of Computer Science & Engineering, Daffodil International University, Dhaka, Bangladesh

Md Baharul Islam**

Deptment of Computing and Software Engineering, Florida Gulf Coast University, Fort Myers, FL 33965, USA

Abstract

Facial acne is a common disease, especially among adolescents, negatively affecting individuals both physically and psychologically. Classifying acne is vital for providing appropriate treatment. Traditional visual inspection or expert scanning is time-consuming and challenging to differentiate acne types. This paper introduces an automated expert system for acne recognition and classification. The proposed method employs a machine learning-based technique to classify and evaluate six types of acne diseases, facilitating the diagnosis process for dermatologists. The preprocessing phase includes contrast improvement, smoothing filter application, and RGB to *Lab* color conversion to eliminate noise and improve classification accuracy. Next, a clustering-based segmentation method, k-means clustering, is applied to segment the disease-affected regions, which then proceed to the feature extraction step. Characteristics of these disease-affected regions are extracted using a combination of gray-level co-occurrence matrix (GLCM) and statistical features. Finally, five different machine learning classifiers are employed to classify acne diseases. The experimental results show that Random Forest (RF) achieves the highest accuracy of 98.50%, which is promising compared to state-of-the-art methods.

Contribution of the Paper: Introduce a machine learning-based system for accurately classifying six types of acne, achieving 98.50% accuracy with the Random Forest classifier.

Keywords: acne disease classification, segmentation, GLCM, expert system, machine vision

IJCVSP

ISSN: 2186-1390 (Online)
http://cennser.org/IJCVSP

Article History:
Received: 10 August 2023
Revised: 23 October 2023
Accepted: 30 November 2013
Published Online: 10 February 2024

1. INTRODUCTION

Acne is a prevalent dermatological condition affecting individuals irrespective of gender. Its onset is primarily attributed to bacterial presence, obstruction of hair follicles by oil, accumulation of dead skin cells, and the overproduction of sebum. Areas rich in sebaceous glands, such as the face, forehead, chest, back, and shoulders, are commonly afflicted. The heterogeneity in acne lesion morphology necessitates accurate differentiation for effective assessment and intervention. Moreover, lesion location may provide insights into underlying conditions, emphasizing the importance of precise acne identification [1].

According to the global acne market study [2], over 90% of the global population is affected by acne. In 2006, acne afflicted 612 million individuals, a figure that rose by 10% over the subsequent decade. Projections suggest that by 2026, nearly 23 million people in India alone will suffer from acne, representing a compound annual growth rate of 0.5%. Consequently, global treatment costs are ris-

*Corresponding author
**Principal Corresponding author
Email addresses: afsana.jeny@uconn.edu (Afsana Ahsan Jeny), masumshah.junayed@uconn.edu (Masum Shah Junayed), mia15-5538@diu.edu.bd (Md Robel Mia), bislam.eng@gmail.com (Md Baharul Islam)

ing. Dermatologists often diagnose specific acne conditions and enumerate acne lesions through visual inspection, a method that can be laborious and imprecise. Two leading skin analysis tools in cosmetic surgery are VISIA from Canfield [3] and ANTERA 3D from Miravex [4]. They aid in treating skin pores, acne scars, and offer anti-aging solutions. However, these tools are expensive and demand specialized knowledge for effective utilization [5]. Several diagnostic methods, such as fluorescence light-based photography [6] and multi-spectral imaging [7], strive to offer dermatologists clearer insights into acne lesions and their characteristics. Despite their precision, these methods require significant manual effort from dermatologists, especially considering the varied scales, shapes, and locations of acne lesions and the diversity in skin tones and lesion types.

Researchers have recently proposed machine and deep learning-based techniques to solve these issues and assist dermatologists. Some of these methods are for acne severity[8, 9], acne grading[10, 11] and only acne detection (acne has or not) [12]. However, limited studies have been performed for the acne classification [13, 1, 14, 15, 16]. The most recent acne detection method, e.g., [12], utilized a human visual system (HSV) based model for segmentation and support vector machine (SVM) and two CNN-based models for detection. Although this method's accuracy is higher than other methods, they utilized only 120 Acne images for training and a total of 200 images for the experiments. In addition, they only detect acne, which can be divided into two classes: normal skin and acne; they do not categorize the numerous types of acne. In another study, Hameed et al. [1] used image processing techniques for segmentation and a Naive Bayes classifier for four types of Acne classification. In their experiments, they used only 40 images and obtained 93.42% accuracy. Therefore, the main challenges are a proper acne recognition system and a large-scale dataset containing different acne types from the above observations. In addition, it is also significant to improve the performance of the automated acne disease classification clinical trial.

Contributions. Motivated by the abovementioned challenges, we propose an automated acne disease recognition and classification method for six classes and a comprehensive analysis using a large-scale comparative dataset. Firstly, the proposed system used some pre-processing before applying K-means clustering to segment different acne classes. For feature extraction, the GLCM and statistical feature selection methods are employed. Five machine learning classifiers, including decision tree (DT), k-nearest neighbors (KNN), support vector machine (SVM), random forest (RF), and logistic regression (LR), are used to classify six acne classes and compare the acquired results. Our main contributions are given below.

- We introduce an automated expert system to recognize acne disorders through segmentation and categorize them into six different classes from acne images.
- We applied contrast enhancement, smoothing filter, and RGB to $L*a*b$ color conversion to better represent the acne regions and reduce the noise in the preprocessing stage. Then, the acne-affected areas are segmented using k-means clustering.
- Two standard feature extraction methods, i.e., GLCM, and statistical features, are utilized together for acne-segmented images to improve classification accuracy in every classifier.
- To demonstrate the system's robustness, five machine-learning classifiers are employed to classify acne disease images with high accuracy.

The rest of this paper is organized as follows: In Section 2, some previous related works are briefly reviewed. Section 3 describes our proposed system and its subsequent parts, such as dataset, pre-processing, segmentation, feature extraction, and classification. The experimental setup, training details, and evaluation matrix are briefly described in Section 4. Section 5 provides the experimental results in detail. Finally, we conclude the work with future research direction in Section 6.

2. Related Works

In this section, we will describe some previous Acne-related studies that have been done using machine learning and deep learning-based methods.

Malik et al. [17] built up an imaging strategy for mechanized skin inflammation by reviewing the past approaches. Discrimination between skin and non-skin pixels was carried out utilizing an automated modified cluster of the K-means on the CIELab color space and four statistical characteristics to differentiate clusters between Acne and non-Acne skin pixels. They used mean, variance, entropy, and energy as the statistical characteristics of the images. At long last, Acne blobs were characterized into various skin break-out classes utilizing shading and measurement as primary features. In their experiments, 70% of the dataset was used for training, and the remaining 30% was used for cross-validation and testing. Their final results show an accuracy of 93.6%. However, they used only 50 images in their experiment with only four features. On the other hand, to figure out the area affected by Acne, Abas et al. [18] used entropy-based filtering and thresholding for feature extraction of skin lesion classes, discrete wavelet transform, and GLCM were used. They used only 17 Acne images to classify six Acne classes, and various classifiers, namely, Binary Classification Tree, Discriminant Analysis Classifier, k-NN, and Naïve Bayesian Classifier. These classifiers were used to detect several Acne lesion classes. They obtained 85.5% accuracy via the Binary Classification Tree, while the accuracy of the other classifiers was comparatively lower.

Alamdari et al. [19] detected and classified Acne lesions using k-means clustering for segmentation; for texture analysis and color-based segmentation, 3D Gaussian and HSV-based models were used, respectively. Moreover, fuzzy-c-means [20] and SVM were utilized to classify Acne lesions. The two-level k-means clustering algorithm was proven to be the dominant performer after analyzing different imaging segmentation methods, and they reached the accuracy of 70%. On the other hand, the accuracies of the fuzzy-c-means and SVM were only 80% and 66.6%, respectively. Their low accuracy value is mainly due to the use of small amount of the images (only 35 images). In another study, Kittigul et al. [21] created an Acne detection system that used a web camera and a Haar Cascade classifier [22] to obtain and extract the patient's anterior face image, then eliminated background pixels with the Grab Cut segmentation algorithm [23] to retrieve only the patient's facial area. Finally, adaptive thresholding [24] and blob detection [25] have been used to mark the observed Acne spot. In addition, a potential solution was indicated using multi-spectral imaging and multi-modal input sources. However, in this method, no feature extraction and classifiers were implemented to detect and classify Acne images.

Maroni et al. [26] proposed a pipeline to detect and extract Acne automatically and this pipeline was able to evaluate Acne lesions and Acne tracking severity as well. The entire body displayed in the picture was classified using Haar Cascade detectors, which were used to identify the frontal face, right portrait, left pattern, and chest, respectively. Body pixels were segmented using a combination of color, structure, pattern, temporal data through unsupervised identifiers. Ten Random Forest models were applied to improve features with rich accuracy on FSD public dataset [26]. However, the participants in this dataset were not affected by Acne diseases, and they were not patients actually. For this reason, they collected 100 images from online websites. The CLElab model has been performed suitably, and Adaptive Thresholding also performed well on Acne lesions and healthy skins. Finally, the algorithm chosen to identify Acne patches and label them in the image was the Laplacian Gaussian filter [27]. Just Acne is detected in this paper but not classified.

Some recent approaches [1, 28] detected and classified Acne, which were based on machine learning models. In [28], the authors utilized some image processing techniques for only Acne detection, and in [1] for Acne classification. Both methods used only Naive Bayes classifier which included four different types of Acne lesions. Their accuracy was about 85.71% and 93.42%, respectively. Zhao et al. [29] developed a grading system based on 4,700 selfie images, ranging from "clear" to "severe", to evaluate the severity of face Acne vulgaris. The image characteristics were retrieved by utilizing a pre-trained model (ResNet 152) and a transfer learning-based approach. The skin patch rolling data augmentation was applied, and the Root Mean Squared Error (RMSE) was 0.482. However, it was detected the Acne severity only, and the dataset was not also made publicly available.

Then, the research in Acne disease classification moved to the deep learning-based methods. For example, Junayed et al. [15] presented a new model based on deep Convolutional Neural Networks (CNNs) to detect Acne classes. In this paper, five Acne classes were used, and each class contained 60 images which means a total of 360 images were used before applying five augmentation methods. They achieved an accuracy of 86.28%, 86.11%, and 95.89% for the training, validation, and testing, respectively. Although overfitting was occurred, accuracy was acceptable. Another CNN-based method was proposed by Shen et al. [14]. They suggested a new approach for detecting Acne vulgaris and other skin diseases. The authors used a public dataset of skin and non-skin photographs in their study. This method used the VGG-16 [30] pre-trained model to compare the performance. They used augmentation techniques to expand their dataset in order to prevent overfitting. The accuracies of the proposed and the VGG-16 pre-trained models were about 89% and 91%, respectively. Therefore, the proposed model performed slightly weaker than the pre-trained model, and some Acne images of papule, cysts, and pustules performed worse in their experiments.

3. Proposed Method

Fig. 1 depicts the step-by-step visualization of the proposed expert system for acne disease recognition and classification using machine learning-based techniques. A preprocessing procedure is first performed on our collected dataset. During the pre-processing stage, the image contrast is enhanced, a smoothing filter is applied, and the color space is transferred from RGB to $L * a * b$. To avoid overfitting, we used some augmentation methods. Next, k-means clustering is applied to segment the input images. Finally, two feature extraction methods are applied together to the segmented areas, and these features are used to classify the acne types. In the following sub-sections, every step of the proposed architecture is explained in detail.

Pre-processing: In the pre-processing step, the medical images are primarily treated to decrease the distortion caused by noises and enhance the essential information in the original image. Three image processing techniques, such as contrast enhancement, smoothing filter, and RGB to $L * a * b$ color conversion, are included in our system to depict acne areas better and eliminate the noises. The guided image filtering [31] is employed in our system as a contrast enhancement technique to reduce the input image noise while maintaining the borders of acne diseases by increasing contrast. Enhancing the images' edge makes it simpler for the machine learning models to learn and discriminate between various types of acne diseases. After that, the blurring and noise reduction in images are accomplished with the help of a smoothing filter [32]. In our

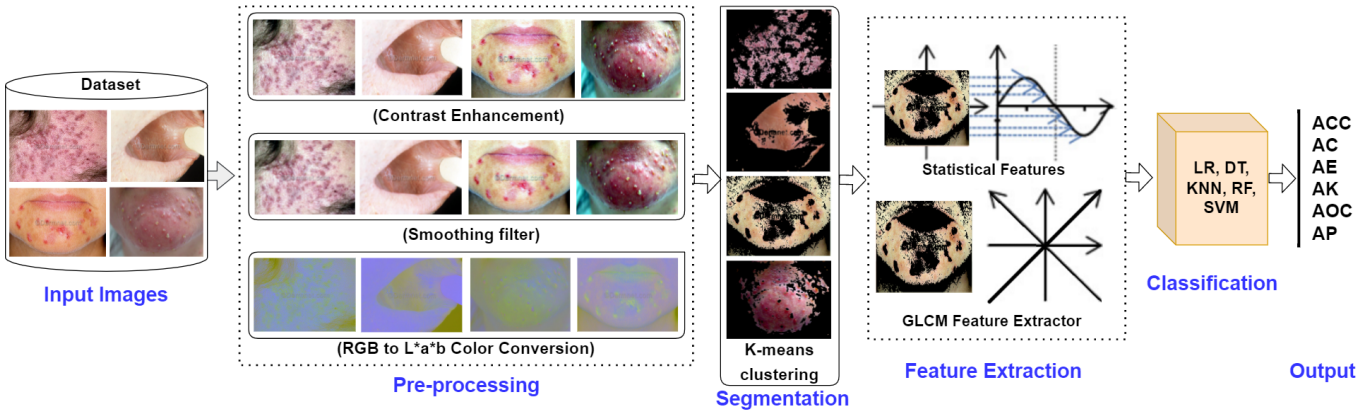

Figure 1: An overview of the proposed expert system for recognizing and classifying acne disease. Input images are pre-processed through contrast enhancement, smoothing filter, and L*a*b color conversion and then segmented using k-means clustering. The GLCM matrix and Statistical feature extraction methods are employed to extract features. The five classifiers, namely logistic regression, decision tree, K-nearest neighbors, random forest, and support vector machine, are utilized to classify acne diseases.

experiment, the Gaussian filter is applied as a smoothing filter that blurs images without removing information and eliminates noises. Some samples of contrast enhancement and smoothing filter are depicted in the pre-processing portion of Fig. 1. Then, the images are converted from RGB to $L*a*b$ color space [33]. It can be written as per color conversion from RGB color space to XYZ color space.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 3.240479 & -1.537150 & -0.498535 \\ -0.969256 & 1.875992 & 0.041556 \\ 0.055648 & -0:204043 & 1.057311 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

To convert XYZ color space into $L*a*b$, we can assume that X_n, Y_n, and Z_n are tri-incentive standards as per the situation. We can redraft it as the following equation 1:

$$f(x) = \begin{cases} x^{\frac{1}{3}}, & \text{if } x > 0.008856 \\ 7.787x + \dfrac{16}{116}, & \text{if } x > 0.008856 \end{cases} \tag{1}$$

Then $L*a*b$ equations (2, 3, 4) can be written as:

$$L = \begin{cases} 116\left(\dfrac{J}{J_n}\right)^{\frac{1}{3}} - 16, & \text{if } \dfrac{J}{J_n} > 0.008856 \\ 903.3\left(\dfrac{J}{J_n}\right), & \text{if } \dfrac{J}{J_n} > 0.008856 \end{cases} \tag{2}$$

$$a* = 500 f(\frac{I}{I_n}) - f(\frac{J}{J_n}) \tag{3}$$

$$b* = 200 f(\frac{J}{J_n}) - f(\frac{K}{K_n}) \tag{4}$$

the color contrast is defined by I, J, and K, and the white colorless specular reflection is described by I_n, J_n, and K_n. This conversion provides a more accurate representation of the input images and the acne disease areas, which improves the final classification accuracy dramatically. The visualization of some $L*a*b$ color conversion images are shown in Fig. 2.

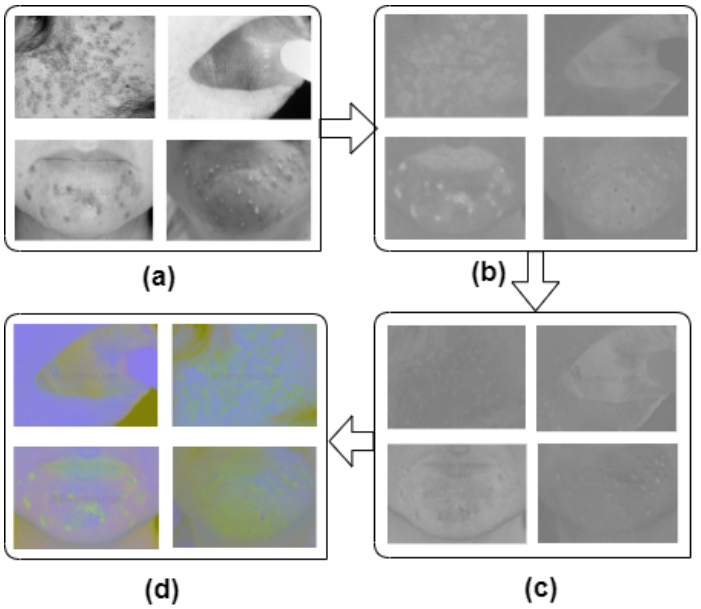

Figure 2: Visualization of the L*a*b color conversion process step by step. Here, (a), (b), (c), and (d) denote the visualization of L, a, b, and L*a*b conversions, respectively.

3.1. Segmentation

In the acne image segmentation process using K-means clustering, statistical features were primarily introduced based on texture analysis. Acne lesions often exhibit distinct textural patterns compared to healthy skin. These features provided a foundation for differentiating acne from unaffected areas, optimizing the segmentation step. This method is used to segment the image into different regions so that we can do further processing. Here, the main target is to define K centers, one for each cluster. Since the other locations cause different results, some repetitive experiments are used in this paper. K-means aims to divide C observations among k clusters. Assume a set of data $X_1, ..., X_C$ with C instances (observations) of a specific

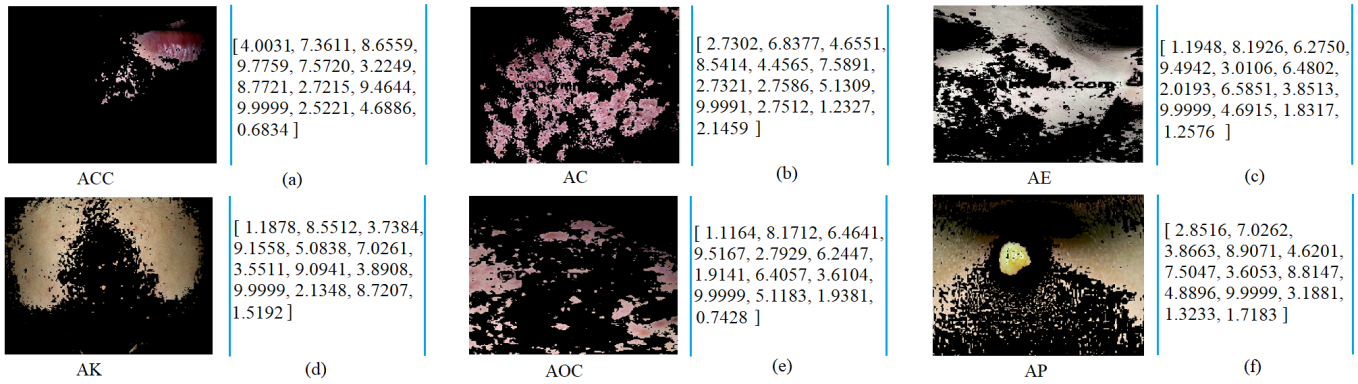

Figure 3: Segmentation and feature extraction results of acne disease are taken as examples. The following list including, (a), (b), (c), (d), (e), and (f) are extracted feature matrics (C, ρ, E, S, H, μ, σ, σ^2, K, RMS, Smoothness, Skewness, and cluster shade (Cs)) of ACC, AC, AE, AK, AOC, and AP, respectively.

D-dimensional Euclidean value of X. This approach clusters data by comparing the inter-point measurements of data points to the measurements between points outside the cluster. To institutionalize this idea, we consider a template, V_k, affiliated with the K^{th} cluster. These experiments represent the clusters' boundaries. The aim is to identify the middle of each region so that square numbers from every value are kept at least to the cluster center V_k nearest to it [34]. The objective function J is generally referred to as distortion measure, is given as follows (5):

$$J = \sum_{i=1}^{C}\sum_{j=1}^{K}(||X_i - V_k||)^2 \tag{5}$$

where $(||X_i - V_k||)^2$ is the Euclidean distance between X_i and V_k. Also, K is the number of data points in k clusters, and C is the number of the cluster centers.

3.2. Feature Extraction

Feature extraction is pivotal in machine learning, guiding the optimal use of variables for specific classifiers. In image classification, the Gray Level Co-occurrence Matrix (GLCM) texture attributes are frequently employed [35]. However, GLCM can exhibit reduced accuracy near class boundaries. To address this, we integrate statistical features with GLCM attributes. This combination not only enhances textural feature extraction from images but also mitigates GLCM's aforementioned limitation. Let, $f(x, y)$ is a two-dimensional image with M×N pixels and L gray levels. In $f(x, y)$, suppose $(x1, y1)$ and $(x2, y2)$ to be two pixels, the length between them is d, and the point among them and the ordinate is h. At that point turns into a GLCM $P(i, j, d, h)$: Let us assume $f(x, y)$ is a digital image in two dimensions of Z by G pixels with gray level numbers L. More specifically, we surmise that $(x1, y1)$ and $(x2, y2)$ are two pixels in $f(x, y)$, the distance is T, and the point between the two and the ordinate is O. At that point, a GLCM $Q(i, j, T, O)$ turns according to [35] is composed as 6:

$$Q(i, j, T, O) = (x1, y1), (x2, y2) \in Z \times G : \\ T, O, f(x1, y1) = i, f(x2, y2) = j \tag{6}$$

We utilized five GLCM features in our experiments, such as Contrast C, Correlation (ρ), Energy E, Entropy S, Homogeneity H, and cluster shade (Cs) Below are their corresponding equations 7 - 11.

$$\mathrm{C} = \sum_{i=1}^{L-1}\sum_{j=0}^{L-1}(i-j)^2 P(i-j) \tag{7}$$

$$\rho = \frac{\sum_{i=1}^{L-1}\sum_{j=0}^{L-1} i.j.P(i,j) - \mu_x.\mu_y}{\sigma_x.\sigma_y} \tag{8}$$

$$\mathrm{E} = \sum_{i=1}^{L-1}\sum_{j=0}^{L-1}(i,j)^2 \tag{9}$$

$$\mathrm{S} = \sum_{i=1}^{L-1}\sum_{j=0}^{L-1} P(i,j)\log P(i,j) \tag{10}$$

$$\mathrm{H} = \sum_{i=1}^{L-1}\sum_{j=0}^{L-1}\frac{P(i,j)}{1+(i-j)^2} \tag{11}$$

$$\mathrm{Cs} = \sum_{i=1}^{L-1}\sum_{j=0}^{L-1}(i+j-i.p(i,j)-j.p(i,j))^3 p(i,j) \tag{12}$$

where μ_x, μ_y, σ_x and σ_y are the sum of estimated and modification standards for the row and column matrix correspondingly.

Based on the findings in [36], we have selected a few statistical features that can be used to identify Acne diseases. Here, we have provided the statistical features which are selected in our system. The related equations for the mentioned statistical features are as equations (13 - 17):

$$\text{Mean},\mu = \frac{\sum_{i=1}^{N} GS_i}{N} \tag{13}$$

Table 1: Description of the five classifiers used in the proposed system.

Classifiers	Description
LR	LR is used to predict the value of the deciding y variable when the y variable is y [0, 1], the negative class is 0, and the positive class is 1. Similarly, it can also be used for multi-classification to identify the value of y when y [0,1,2,3] is provided. Our LR model is Multinomial and the logistic function is: logit(k) = lnk - ln (1-k).
SVM	SVM transforms nonlinear data into a higher-dimensional space in which it is linearly separable, thus improving its classification accuracy. To divide several classes, hyperplanes are used, and the optimum hyperplane is the one that maximizes the margin between the classes. Because of its strong generalization capacity, it is utilized effectively in a wide range of classification areas. In this experiment, we have used the linear kernel function, and the parameter value of C is 55.
RF	The random forest is competent in locating null values from a vast number of datasets and can give a more accurate result than other methods for data mining. The tree's maximum height is an unlimited integer, and the number of attributes is zero, which is selected randomly in this work. The size of each RF bag is equal to the percentage of training data size.
DT	As with a tree, DT shapes include leaves or decision nodes that may be selected. It is composed of internal and external nodes in the system. The offspring nodes that visit the following nodes are among the internal nodes that make decisions. We have utilized the Gini index as a degree of impurity. Entropy is for information gain, the minimum number of the split is 2, and the maximum depth of the tree is infinity. Note that the expansion of nodes continues until all leaves are pure or when the minimum number of split samples is reached for all leaves.
KNN	As an example, when KNN predicts the class label of new input, it is compared with the similarity of new input to the input's samples from the training set. This condition is fulfilled if the new input is identical to samples from a previously trained set. Usually, KNN classification performance is not very excellent for the classification problem. Here, we have set the value of K as 7, and the Manhattan distance is utilized.

$$\text{Standard deviation},\sigma = \sqrt{\frac{\sum_{i=1}^{N}(GS_i-\mu)^2}{N}} \tag{14}$$

$$\text{Variance},\sigma^2 = \frac{\sum_{i=1}^{N}(GS_i-\mu)^2}{N} \tag{15}$$

$$\text{Kurtosis},K = \frac{\frac{1}{N}\sum_{i=1}^{N}(GS_i-\mu)^4}{(\frac{1}{N}\sum_{i=1}^{N}(GS_i-\mu)^2)^2} - 3 \tag{16}$$

$$\text{Skewness},\gamma = \frac{\mu - N_o}{\sigma} \tag{17}$$

we have addressed the N number of pixels in defective districts, where GS represents the grayscale shading power of a pixel, μ, and N_o addressed the mean and mode of the grayscale color intensity of all pixels separately.

Fig. 3 depicts the outcomes of our selected feature extraction techniques (GLCM and Statistical) for six acne disease classes that are absolutely robust to increase the classification accuracy. Moreover, in Table 3, it is observed that we were able to achieve higher accuracy in every classifier by using our combined feature extraction methods. The feature map in Fig. 3 represents the value of GLCM features (Contrast C, Correlation (ρ), Energy E, Entropy S, Homogeneity H, and cluster shade (Cs)), Statistical features (mean μ, standard deviation σ, variance σ^2, kurtosis K, and skewness γ), RMS, and smoothness respectively.

Classification: For the classification of acne diseases, five machine learning classifiers, namely LR, DT, KNN, SVM, and RF, are utilized. These classifiers are shortly explained in Table 1.

4. Experiements

Dataset: We collect 320 images from the public platform Dermnet (`https://dermnet.com/`), and 120 images from New Zealand Dermatologists (`https://dermnetnz.org/`) prior to applying augmentation. A total of 2100 images are used for our experiments. There are $410, 300, 340, 330, 370,$ and 350 images for ACC, AC, AE, AK, AOC, and AP classes, respectively. The dataset is divided into two parts: training and testing. 80% images (1680 images) are used for training and 20% images (420 images) for testing. Utilizing a 5-fold cross-validation [37] technique and averaging the results, the proposed method's performance is assessed.

Experimental Setup: Experiments in this paper are done using the following computing system: Intel Core i9 Central Processing Unit (CPU) operating at 3.60 GHz, 64 GB of RAM, and an NVIDIA Geforce RTX 2080 Super GPU with 8GB of GPU memory. All the detection and classification computations are implemented using MATLAB (R2016a) software.

Training Details: We divided our dataset into two parts: training and testing. We employed the holdout approach [38] to specify the number of data allocated for training and testing [39]. Around 66% of the sample data set (1387 color images) is used for training, and the rest (713 color images) is used for testing. Additionally, the original training dataset is split into two smaller subsets: one for testing, and another one for validation. Next, about 66% of the training set (916 images) is utilized for the classification, and the rest (471 images) is used for error evaluation. To identify the best classifier, holdout techniques

[38] is applied multiple times. After extensive experiments, the performance of all of the classifiers is analyzed.

Evaluation Metrics: For evaluating the performance of each classifier, accuracy cannot be asserted as a thorough measurement for the estimation of the open exhibition. This is because, it would not be ideally suited for evaluating identification characteristics acquired from unequal class distribution datasets, because for example, the quantities of samples in different classes vary greatly. Other assessment matrices for evaluating the output of a classifier are built on the Confusion Matrix [40], as described in [41]. For a two-class scenario, a binary Confusion Matrix (CM) shows the number of true positives (TPs), false negatives (FNs), false positives (FPs), and true negatives (TNs). CM for a multi-class classification (W), can be written as the following equation:

$$W = [b_{ij}]_{nxn} \tag{18}$$

The multi-class CM (W) is a $n \times n (n > 2)$ square matrix. It has n rows and n columns, totally including n^2 entries. There is no simple way to calculate the number of FPs ($\sum_{j=1,j\neq i}^{n} b_{ji}$), FNs ($\sum_{j=1,j\neq i}^{n} b_{ij}$), TPs ($b_{ii}$), and TNs ($\sum_{j=1,j\neq i}^{n} \sum_{k=1,k\neq i}^{n} b_{jk}$) for the multi-class CM. The results of TPs, FNs, FPs, and TNs are determined and as per the regulations for multi-class matrix as described in the CM. The final CM dimension is $2x2$, and the mean values are kept in n confusion matrices for each class. An uncertainty matrix, also known as an error matrix, can be utilized for statistical classification. The is used to measure the accuracy ($\frac{(TP+TN)}{(TP+TN+FP+FN)} \times 100(\%)$), precision ($\frac{(TP)}{(TP+FP)} \times 100(\%)$), specificity ($\frac{(TN)}{(FP+TN)} \times 100(\%)$), sensitivity ($\frac{(TP)}{(FN+TP)} \times 100(\%)$), false positive rate (FPR) ($\frac{(FP)}{(FP+TN)} \times 100(\%)$), and false negative rate (FNR) ($\frac{(FN)}{FN+TP} \times 100(\%)$).

Table 2: The binary Confusion Matrix (CM) for every class.

Class	Matrix				Class	Matrix			
ACC	Actual		Predicted		AC	Actual		Predicted	
			+	-				+	-
		+	77	5			+	57	3
		-	3	335			-	3	357
AE	Actual		Predicted		AK	Actual		Predicted	
			+	-				+	-
		+	63	5			+	64	2
		-	4	348			-	2	352
AOC	Actual		Predicted		AP	Actual		Predicted	
			+	-				+	-
		+	72	2			+	68	5
		-	2	344			-	2	345

5. Results and Discussion

5.1. Performance Analysis

Table 2 shows the binary CM for each of the six acne disease classes of RF utilizing our dataset. As can be seen in this table, the performance of our classifiers is competitive on the acne dataset, as shown by the large number of TP values we obtained in our research and the low misclassification rates in each of the classes.

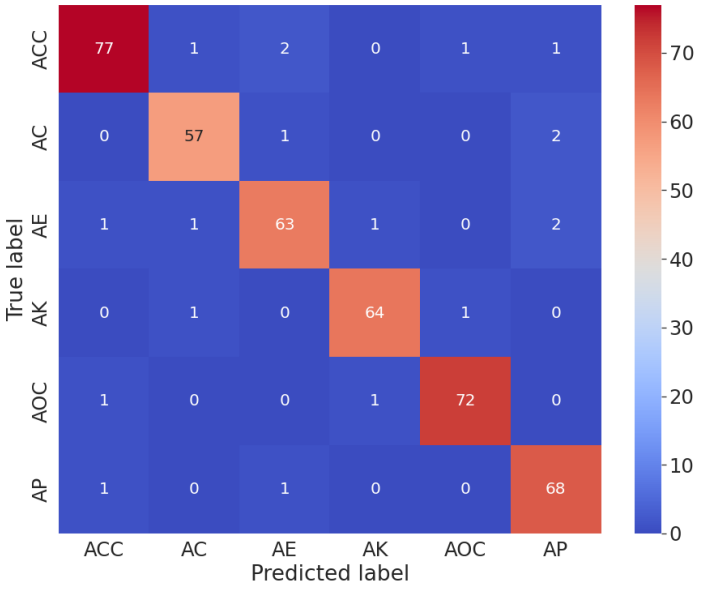

Figure 4: Multi-class confusion matrix for the RF classifier.

Fig. 4 represents the multiclass CM for the RF classifier, where the X and Y axes represent the predicted and true levels, respectively. The misclassification rate for the RF classifier is limited. The number of correctly recognized images in ACC and AOC classes is 77 and 72, respectively, while the misclassification value is only 5 and 2, respectively. Similarly, the misclassification values for AC, AE, AK, and AP are just $3, 5, 2$, and 2 correspondingly, demonstrating the RF's competitive performance.

Table 3 displays the performance of five machine learning models, namely, DT, KNN, SVM, RF, and LR in terms of accuracy, precision, sensitivity, specificity, FPR, and FNR for each of the Acne types (i.e., ACC, AC, AE, AK, AOC, and AP). The average accuracy of LR, DT, KNN, SVM, and RF are 96.87%, 96.99%, 97.22%, 98.06%, and 98.50% respectively, according to the data presented in this table, which again RF shows the best performance. RF exceeds not only in terms of accuracy but also in other evaluation metrics, such as precision (94.83%), sensitivity (96.12%), specificity (98.95%), FPR (1.04%), and FNR (3.88%), respectively. This Table also includes the individual accuracy for each of the diseases. For example, AK got the highest accuracy of 99.07% while AE, has the lowest accuracy, with 97.86% for RF classifier.

To further analyze the results of our method, we utilized the receiver operating characteristic (ROC) curves [42] and calculated the area under it, known as AUC. Determining which classifier is superior on average can be estimated using the area under the ROC curve. A ROC curve depicts the connection between the FPR and the TPR at various thresholds where TPR refers to Y and

Table 3: Classifier's performance on different classes on Accuracy(Acc), Precision(Pre), Sensitivity(Sen), Specificity(Spe), FPR, and FNR.

Classifiers	Classes	Acc	Pre	Sen	Spe	FPR	FNR
LR	ACC	97.14	89.39	89.39	98.35	1.65	10.61
	AC	97.34	94.31	95.08	98.09	1.91	4.92
	AE	95.91	84.62	89.19	97.11	2.89	10.81
	AK	96.73	85.48	88.33	97.90	2.10	11.67
	AOC	98.16	95.40	94.32	99.00	1.00	5.68
	AP	95.91	86.96	90.91	97.01	2.99	9.09
Total Average		96.87	89.36	91.20	97.01	2.99	8.80
DT	ACC	96.31	84.85	87.50	97.64	2.36	12.50
	AC	96.73	93.50	93.50	97.81	2.19	6.50
	AE	96.32	85.90	90.54	97.35	2.65	9.46
	AK	98.36	93.55	93.55	99.06	0.94	6.45
	AOC	97.75	94.25	93.18	98.75	1.25	6.82
	AP	96.52	89.13	92.13	97.50	2.50	7.87
Total Average		96.99	90.20	91.73	98.02	1.92	8.27
KNN	ACC	96.72	86.36	89.06	97.88	2.12	10.94
	AC	96.11	91.87	92.62	97.28	2.72	7.38
	AE	97.55	89.74	94.59	98.07	1.93	5.41
	AK	97.75	88.71	93.22	98.37	1.63	6.78
	AOC	98.16	94.25	95.35	98.76	1.40	4.65
	AP	97.55	92.39	94.44	98.25	1.75	5.56
Total Average		97.22	90.55	93.21	98.10	1.90	6.79
SVM	ACC	98.77	93.94	96.88	99.06	0.94	3.12
	AC	97.96	95.93	95.93	98.63	1.37	4.07
	AE	96.93	89.74	90.91	98.06	1.94	9.09
	AK	98.36	91.94	95.00	98.83	1.17	5.00
	AOC	98.57	96.55	95.45	99.25	0.75	4.75
	AP	97.75	93.48	94.51	98.49	1.51	5.49
Total Average		98.06	93.59	94.78	98.72	1.28	5.22
RF	ACC	98.13	93.91	96.26	98.53	1.47	3.74
	AC	98.57	95.02	95.00	99.17	0.83	5.00
	AE	97.86	92.65	94.03	98.58	1.42	5.97
	AK	99.05	96.97	96.97	99.44	0.56	3.03
	AOC	99.07	97.30	97.31	99.42	0.58	2.69
	AP	98.33	93.15	97.14	98.57	1.43	2.86
Total Average		**98.50**	**94.83**	**96.12**	**98.95**	**1.04**	**3.88**

FPR to X axes. Fig. 5 depicts the ROC curves and area under curves of all five classifiers used in this study. This Fig. shows that RF acquires more AUC than the other four classifiers, with 85.8%. On the other hand, the LR has the lowest AUC (77.9). Furthermore, the AUC for KNN is 2.4% lower than that of the SVM (84.9 vs. 82.5) and 9.3% higher than that of the decision tree (82.5 vs. 73.2).

In Fig. 6, the RF's performance was further validated across all Acne disease classes by the ROC curve, which helped to make our claim in contribution more strong. Based on our observations, we discovered that the AUC of every class is excellent; among them, the Acne Keloidalis class has the most significant amount of AUC (98%), and the Acne Pustular class has the lowest amount of AUC (73%). Finally, the AUC for the RF class as a whole is 85.8% on average.

Table 4: The performance of the feature extraction methods on different classifiers (LR, DT, KNN, SVM, and RF).

Feature Extraction	LR	DT	KNN	SVM	RF
Statistical	87.56%	88.76%	89.93%	90.02%	90.61%
GLCM	92.13%	93.31%	94.67%	95.11%	95.31%
Statistical+GLCM	**96.87%**	**96.99%**	**97.22%**	**98.06%**	**98.50%**

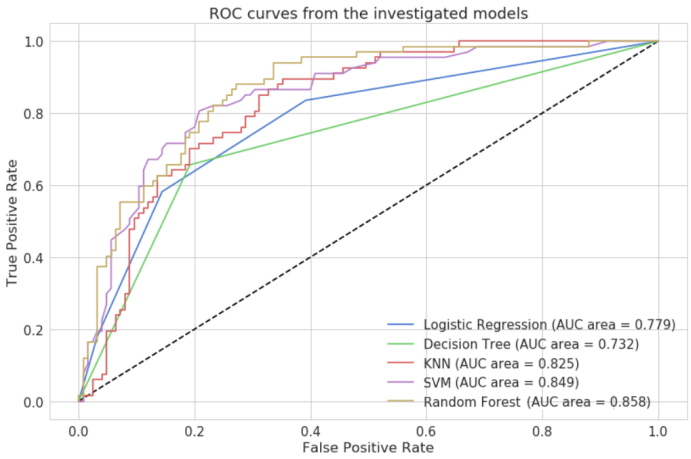

Figure 5: AUC-ROC curves of five different ML classifiers.

5.2. Qualitative Analysis of Feature Extraction

Table 4 shows the performance analysis utilizing feature extraction techniques. When the Statistical and GLCM feature approaches are used individually, the system's performance is worse than when the two methods are used together, as shown in Table 4. The GLCM approach fails to recognize image boundaries, whereas the Statistical method performs well when used for texture classification and edge detection. Consequently, their individual performance is lower, but their combined performance is remarkable in every classifier. For instance, when Statistical and GLCM methods perform separately in RF, the accuracy is 90.61% and 95.31%, respectively. However, during the performance of their combination, the precision is increased by roughly 8% (98.5% vs. 90.61%) and 3% (98.5% vs. 95.31%), respectively.

Fig. 7 portrays the performance investigation of accuracy vs. the number of features we employed. RF is an ensemble model, so it is projected to perform excellently. From Fig. 7, we can see that as the number of features rises, the performance of the RF model does not degrade. Similarly, SVM also performs well on our dataset without showing any fluctuation. On the other hand, in LR, DR, and KNN, during feature analysis, they showed some fluctuations though their accuracy is good. It is most likely due to a violation of the feature's independence. However, it is worth mentioning that utilizing a high number of features may result in a worse performance of the model on the test dataset.

5.3. Comparison

The earlier studies on acne detection and classification are depicted with their employed images in Table 5. In 2016, Abas et al. [18], and Hameed et al.[1] used machine learning classifiers, and their obtained accuracy was 85.5% and 93.42%, respectively. Shen et al.[14], Junayed et al.[15], and Isa et al.[16] employed deep learning classifiers and their obtained accuracy were 91.95% (VGG),

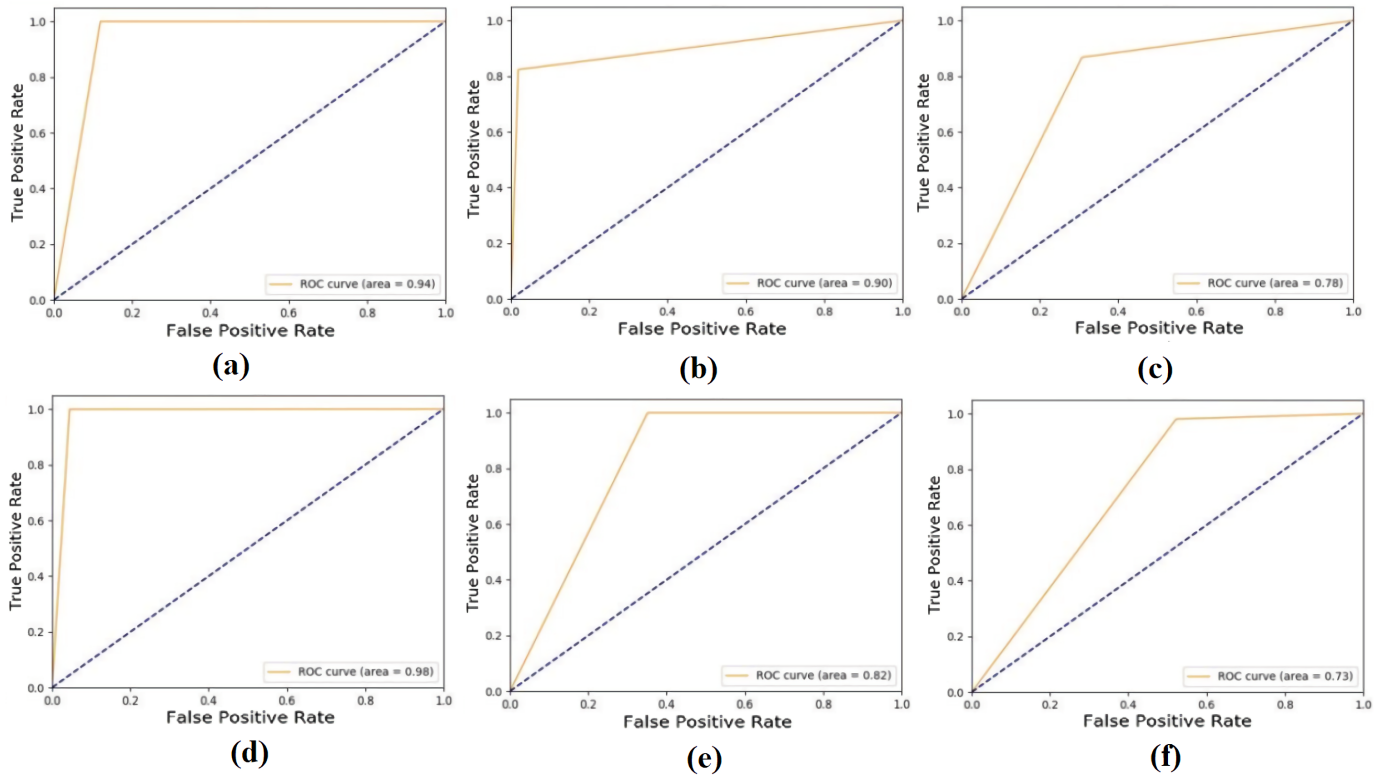

Figure 6: The AUC-ROC curves for six acne classes, (a)-(f): the ROC curves of ACC, AC, AE, AK, AOC, and AP using RF classifier.

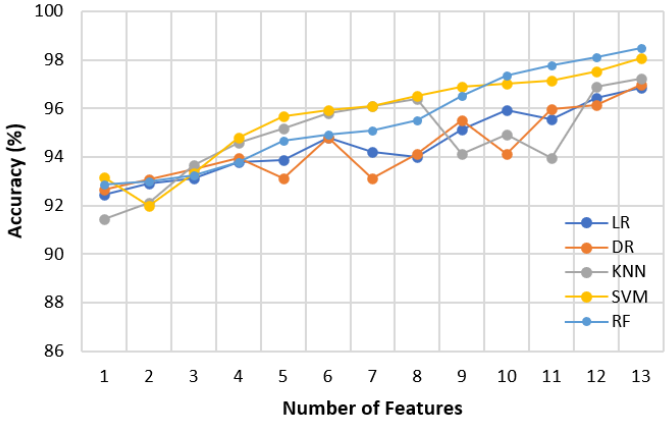

Figure 7: The average accuracy of the images using Statistical and GLCM features together in different classifiers (LR, DT, KNN, SVM, and RF).

Table 5: Comparison with state-of-the-art on different datasets.

Approaches	Year	Dataset size	Accuracy
Abas et al.[18]	2016	17 acne images, 6 classes	85.5%
Shen et al.[14]	2018	3000 skin-nonskin images	91.95% (VGG) 91.35% (CNN)
Junayed et al.[15]	2019	360 images, 5 classes	95.89%
Hameed et al.[1]	2020	40 images, 4 classes	93.42%
Isa et al.[16]	2021	215 images, 4 classes	91.25%
Ours (Proposed)	2022	**440 images, 6 classes (Original)**	**98.50% (RF)**

Table 6: Comparison with state-of-the-art on same dataset.

Datasets	Classes	Size	Approaches	Accuracy
Acne Classes	5	1800	AcneNet [15]	95.89%
			Our System	**97.13%**
Ours	6	2100	AcneNet [15]	96.79%
			Our System	**98.50%**

and 91.35% (Proposed CNN), 95.89%, and 91.25%, respectively. However, it is important to mention that these strategies were developed using private datasets that are not publicly accessible. As a result, this comparison is expressed from different datasets. We attempted to contact these researchers to get the implementation parameters but was not reachable.

To have a fair comparison between our method with state-of-the-art, we must implement state-of-the-art methods on the given dataset. After searching deeply, we got one source code, and one dataset from this related work [15]. Therefore, we only compared the proposed method with the AcneNet[15] and used the same dataset for implementation. Moreover, we utilized their source code on their dataset using our method. Table 6 represents the evaluated result of the comparison. It can be seen that our system, through RF, is applied on the AcneNet data set with 1800 images and five different classes of Acne disease

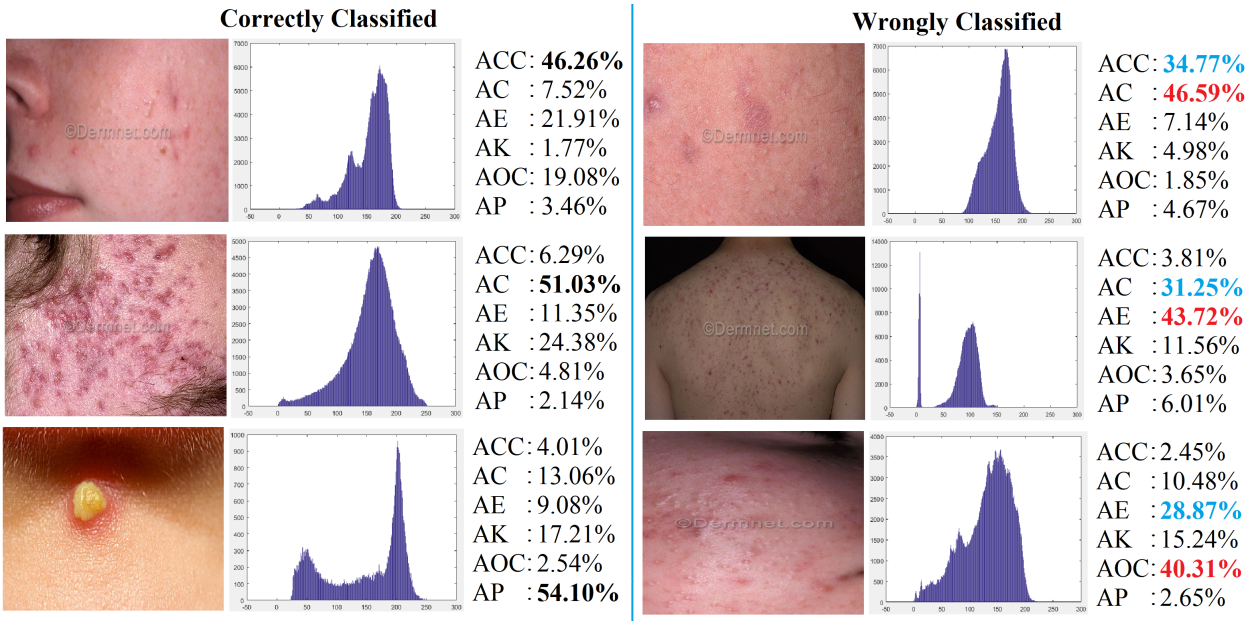

Figure 8: The classification and misclassification results of the proposed system. We demonstrated 6 acne images. Col 1-3: correctly classified images, their histograms, and scores (bold scores for the correct classes). Col 4-6: wrongly classified images, their histograms, and scores (red scores for the misclassified classes and blue scores for the correct classes) respectively.

achieved 1. 24% more precision than AcneNet (97. 13% vs. 95. 89%). On the other hand, while AcneNet is applied to our dataset, our system also showed higher performance in terms of accuracy (98.50%) with 2100 images and six different Acne disease classes.

5.4. Misclassification

Fig. 8 depicts the classification and misclassification performance of our proposed model with the histogram on the acne disease images. Six acne disease images have been taken into consideration, where three of them are predicted correctly (first column), and the rest failed in prediction (fourth column). Still, our system does not identify some of the images with severe and blurry acne. For example, the fourth column's first, second, and last images are a little bit blurry and have severe acne. For this reason, these images are misclassified into AC, AE, and AOC classes, respectively, instead of ACC, AC, and AE classes. However, they are wrongly classified, but our findings are not much different from the scores of the actual images (46.59% vs. 34.77%, 43.72% vs. 31.25%, and 40.31% vs. 28.87%). The severity of acne is readily apparent from the histogram. For example, the first image (col 1) has less acne on its left and right, so we got values between 50 to 200.

6. Conclusion

This article presented an automated system that recognizes and classifies six acne diseases. A pre-processing step that included contrast enhancement, smoothing filter, and $L * a * b$ color conversion was performed to remove noise from the input images and provide better visualization. We extracted the GLCM and statistical features before performing segmentation using k-mean clustering. Finally, extracted features were used for training five classifiers to recognize and classify acne diseases. The RF classifier achieved 98.50% accuracy compared to other classifiers, a promising performance. However, our proposed system is unable to distinguish specific acne disease images. Therefore, possible future works can be focused on reducing the misclassification results and making a more uniform dataset with different acne classes to develop a more effective detection, recognition, and grading system for different types of acne disease.

References

[1] A. Z. Hameed, W. K. Awad, N. A. Irsan, A. S. Abdulbaqi, Hybrid technique for skin pimples image detection and classification.

[2] ReportLinker, Global Acne Market Report for 2016-2026, `https://www.reportlinker.com/p05251482/Global-Acne-Market-Report-for.html`, [Online; accessed 04-October-2022] (2021).

[3] I. Canfield Scientific, Redefining the Vision of Skin Care, `https://www.canfieldsci.com/imaging-systems/visia-complexion-analysis/`, [Online; accessed 18-June-2022] (2021).

[4] A. 3D, Versatile and rich in features, `http://miravex.com/antera-3d/`, [Online; accessed 10-July-2022] (2021).

[5] C. S. Park, J.-H. Park, C. R. Kim, J. H. Lee, Objective analysis of volume restoration in atrophic acne scars and skin pores: a split study using human stem cell-conditioned media, Journal of Dermatological Treatment (2019) 1–5.

[6] L. C. Lucchina, N. Kollias, R. Gillies, S. B. Phillips, J. A. Muccini, M. J. Stiller, R. J. Trancik, L. A. Drake, Fluorescence photography in the evaluation of acne, Journal of the American Academy of Dermatology 35 (1) (1996) 58–63.

[7] H. Fujii, T. Yanagisawa, M. Mitsui, Y. Murakami, M. Yamaguchi, N. Ohyama, T. Abe, I. Yokoi, Y. Matsuoka, Y. Kubota, Extraction of acne lesion in acne patients from multispectral images, in: 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2008, pp. 4078–4081.

[8] A. Nguyen, H. Thai, T. Le, Severity assessment of facial acne, in: International Conference on Computational Collective Intelligence, Springer, 2021, pp. 599–612.

[9] H. Wen, W. Yu, Y. Wu, Z. Jun, X. Liu, Z. Kuang, R. Fan, Acne detection and severity evaluation with interpretable convolutional neural network models, Technology and Health Care (Preprint) (2022) 1–11.

[10] Y. Lin, Y. Guan, Z. Ma, H. You, X. Cheng, J. Jiang, An acne grading framework on face images via skin attention and sfnet, in: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2021, pp. 2407–2414.

[11] S. Alzahrani, B. Al-Bander, W. Al-Nuaimy, Attention mechanism guided deep regression model for acne severity grading, Computers 11 (3) (2022) 31.

[12] N. Yadav, S. M. Alfayeed, A. Khamparia, B. Pandey, D. N. Thanh, S. Pande, Hsv model-based segmentation driven facial acne detection using deep learning, Expert Systems (2021) e12760.

[13] M. S. Junayed, M. B. Islam, N. Anjum, A transformer-based versatile network for acne vulgaris segmentation, in: 2022 Innovations in Intelligent Systems and Applications Conference (ASYU), IEEE, 2022, pp. 1–6.

[14] X. Shen, J. Zhang, C. Yan, H. Zhou, An automatic diagnosis method of facial acne vulgaris based on convolutional neural network, Scientific reports 8 (1) (2018) 1–10.

[15] M. S. Junayed, A. A. Jeny, S. T. Atik, N. Neehal, A. Karim, S. Azam, B. Shanmugam, Acnenet-a deep cnn based classification approach for acne classes, in: 2019 12th International Conference on Information & Communication Technology and System (ICTS), IEEE, 2019, pp. 203–208.

[16] N. A. M. Isa, N. N. A. Mangshor, Acne type recognition for mobile-based application using yolo, in: Journal of Physics: Conference Series, Vol. 1962, IOP Publishing, 2021, p. 012041.

[17] A. S. Malik, R. Ramli, A. F. M. Hani, Y. Salih, F. B.-B. Yap, H. Nisar, Digital assessment of facial acne vulgaris, in: 2014 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings, IEEE, 2014, pp. 546–550.

[18] F. S. Abas, B. Kaffenberger, J. Bikowski, M. N. Gurcan, Acne image analysis: lesion localization and classification, in: Medical Imaging 2016: Computer-Aided Diagnosis, Vol. 9785, International Society for Optics and Photonics, 2016, p. 97850B.

[19] N. Alamdari, K. Tavakolian, M. Alhashim, R. Fazel-Rezai, Detection and classification of acne lesions in acne patients: A mobile application, in: 2016 IEEE International Conference on Electro Information Technology (EIT), IEEE, 2016, pp. 0739–0743.

[20] P. K. SK, M. Sumithra, N. Saranya, Particle swarm optimization (pso) with fuzzy c means (pso-fcm)–based segmentation and machine learning classifier for leaf diseases prediction, Concurrency and Computation: Practice and Experience 33 (3) (2021) e5312.

[21] N. Kittigul, B. Uyyanonvara, Automatic acne detection system for medical treatment progress report, in: 2016 7th International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES), IEEE, 2016, pp. 41–44.

[22] J. Kaur, P. Sinha, R. Shukla, V. Tiwari, Automatic cataract detection using haar cascade classifier, in: Data Intelligence and Cognitive Informatics, Springer, 2021, pp. 543–556.

[23] J. Sivasamy, et al., Classification of healthy control and abnormal lung chest radiography images using cbir and atlas-based graph cut segmentation by transfer learning cnns, Turkish Journal of Computer and Mathematics Education (TURCOMAT) 12 (9) (2021) 1829–1835.

[24] C. Patgiri, A. Ganguly, Adaptive thresholding technique based classification of red blood cell and sickle cell using naïve bayes classifier and k-nearest neighbor classifier, Biomedical Signal Processing and Control 68 (2021) 102745.

[25] M. A. Vidhyavani, P. Malvika, M. Supriya, M. Alekya, Automatic blood cell count using blob detection algorithm, Annals of the Romanian Society for Cell Biology (2021) 13448–13456.

[26] G. Maroni, M. Ermidoro, F. Previdi, G. Bigini, Automated detection, extraction and counting of acne lesions for automatic evaluation and tracking of acne severity, in: 2017 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, 2017, pp. 1–6.

[27] S. Mangal, P. Meshram, Plant disease identification using deep learning classification model: Cnn.

[28] N. P. N. Xuan, T. T. Thi, T. D. Minh, D. T. N. Bao, A multilevel thresholding approach for acne detection in medical treatment, in: 2021 3rd International Conference on Image Processing and Machine Vision (IPMV), 2021, pp. 17–23.

[29] T. Zhao, H. Zhang, J. Spoelstra, A computer vision application for assessing facial acne severity from selfie images, arXiv preprint arXiv:1907.07901.

[30] D. Baby, S. J. Devaraj, et al., Leukocyte classification based on transfer learning of vgg16 features by k-nearest neighbor classifier, in: 2021 3rd International Conference on Signal Processing and Communication (ICPSC), IEEE, 2021, pp. 252–256.

[31] K. He, J. Sun, X. Tang, Guided image filtering, IEEE transactions on pattern analysis and machine intelligence 35 (6) (2012) 1397–1409.

[32] L. Cadena, A. Zotin, F. Cadena, A. Korneeva, A. Legalov, B. Morales, Noise reduction techniques for processing of medical images, in: Proceedings of the World Congress on Engineering, Vol. 1, 2017, pp. 5–9.

[33] I. Hosen, T. Tabassum, J. Akhter, I. Islam, Detection of fruits defects using colour segmentation technique, International Journal of Computer Science & Information Security (IJCSIS) 16 (6) (2018) 215–223.

[34] S. A. Burney, H. Tariq, K-means cluster analysis for image segmentation, International Journal of Computer Applications 96 (4).

[35] Y. Liu, Q. Li, B. Du, M. Farzaneh, Feature extraction and classification of surface discharges on an ice-covered insulator string during ac flashover using gray-level co-occurrence matrix, Scientific Reports 11 (1) (2021) 1–9.

[36] W. K. Mutlag, S. K. Ali, Z. M. Aydam, B. H. Taher, Feature extraction methods: A review, in: Journal of Physics: Conference Series, Vol. 1591, IOP Publishing, 2020, p. 012028.

[37] B. G. Marcot, A. M. Hanea, What is an optimal value of k in k-fold cross-validation in discrete bayesian network analysis?, Computational Statistics 36 (3) (2021) 2009–2031.

[38] S. Raschka, Model evaluation, model selection, and algorithm selection in machine learning, arXiv preprint arXiv:1811.12808.

[39] J. Awwalu, O. Nonyelum, On holdout and cross validation: A comparison between neural network and support vector machine, International Journal of Trend in Research and Development 6 (2) (2019) 2394–9333.

[40] M. S. Junayed, A. N. M. Sakib, N. Anjum, M. B. Islam, A. A. Jeny, Eczemanet: A deep cnn-based eczema diseases classification, in: 2020 IEEE 4th International Conference on Image Processing, Applications and Systems (IPAS), IEEE, 2020, pp. 174–179.

[41] D. M. Tax, R. P. Duin, Using two-class classifiers for multiclass classification, in: Object recognition supported by user interaction for service robots, Vol. 2, IEEE, 2002, pp. 124–127.

[42] K. Feng, H. Hong, K. Tang, J. Wang, Decision making with machine learning and roc curves, Available at SSRN 3382962.